

Statistical Complexity as a Criterion for the Useful Signal Detection Problem

A. A. Galyaev^{*,a}, P. V. Lysenko^{*,b}, and L. M. Berlin^{*,c}

**Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
e-mail: ^agalyaev@ipu.ru, ^bpavellysen@ipu.ru, ^cberlin.lm@phystech.edu*

Received March 21, 2023

Revised April 19, 2023

Accepted May 11, 2023

Abstract—Three variants of the statistical complexity function, which is used as a criterion in the problem of detection of a useful signal in the signal-noise mixture, are considered. The probability distributions maximizing the considered variants of statistical complexity are obtained analytically and conclusions about the efficiency of using one or another variant for detection problem are made. The comparison of considered information characteristics is shown and analytical results are illustrated on an example of synthesized signals. A method is proposed for selecting the threshold of the information criterion, which can be used in decision rule for useful signal detection in the signal-noise mixture. The choice of the threshold depends a priori on the analytically obtained maximum values. As a result, the complexity based on the total variation demonstrates the best ability of useful signal detection.

Keywords: statistical complexity, signal detection, information divergence

DOI: 10.25728/arcRAS.2023.27.32.001

1. INTRODUCTION

The concept of information entropy was firstly introduced in Claude Shannon's article [1] in 1948. This work marked the beginning of a new field of science called information theory [2]. The development of information theory made possible an analytical and practical research in many applied fields of science and technology. Such terms as Gibbs and von Neumann entropies, Kullback–Leibler distance, Jensen–Shannon divergence, information divergences and some others were introduced and interpreted and later began to serve as criteria for various optimization problems of recognition [3], classification [4] and filtering.

By the end of the last century various information criteria, mainly Shannon information entropy, had began to be actively applied in the tasks of digital signal processing, in particular in the problem of detection of a useful signal in a noise environment [5]. The concept of spectral entropy [6], associated with the Fourier spectrum of the considered signal, has appeared and proved to be especially relevant in the analysis of acoustic signals [7]. In addition, the entropic approach has been successfully applied in the analysis of time series in the medical field, such as ECG or EEG [8]. Later, a statistical complexity function was proposed as a development of the entropy concept [9–11]. However, the articles mostly do not provide an analytical study of the properties of these functions, which turns out to be especially important when solving the problem of hypothesis testing.

It should be noted that there are several classical ways of solving the detection problem. The first of them is based on solving the problem of optimal filtering and requires knowledge of the properties of the signal: periodicity, bandwidth, etc. [12]. The second way is based on the Neumann–Pearson lemma, solves the problem of hypothesis testing, and determines the fact of exceeding the optimal threshold at a given false alarm probability and requires estimation of statistical properties of sample distributions of noise and mixture of signal and noise [13]. The third way is equivalent to solving the changepoint detection problem when the unknown statistical characteristics of the signal distributions change. The anomaly detection problem [14] has a similar formulation. All these methods demonstrate qualitative and reliable performance when the signal exceeds the noise, but for small signal-to-noise ratios often give the wrong answer.

The article is devoted to the problem of detection of useful signal in the signal-noise mixture and combines all three previously listed ways of solving the detection problem. We propose to use a variant of the Neumann–Pearson lemma for the problem of hypothesis testing [13], which is indeed valid when the error probability is close to one and depends on the total variation of the measure of two distributions of the null and alternative hypotheses. Based on the analytical expression of this error function, the criterion of the signal detection problem is formalized as one of the variants of the statistical complexity [15], which takes into account the deterministic nature of the signal mixed in with the noise. The peculiarity of the statistical complexity is that it is multiplicative and consists of two multipliers, one of which is zero on deterministic sinusoidal signals of the same frequency (in physics these are objects of a given structure, such as crystals [9]) and the other is zero on uniform distribution functions [10], corresponding, for example, to white noise. Then the introduced criterion is compared with the already known two variants of statistical complexity based on Euclidean distance square and Jensen–Shannon divergence, their properties are established, and optimization as a function of many variables on a set of discrete distributions is performed. As a result, families of optimal distributions are identified and maxima of statistical complexity functions are calculated.

The article has the following structure. Section 1 provides a literature review and highlights the current state of research on the topic of the article. Section 2 is devoted to the connection of the considered information criteria with the classical criterion of the signal detection problem. Section 3 investigates the properties of the three types of statistical complexity. In 4 the analytical results of the previous section are supported by numerical simulations for synthesized signals. Section 5 summarizes the results obtained in the paper and lists plans for the future.

2. NEYMAN-PEARSON LEMMA AND STATISTICAL COMPLEXITY

The problem of signal detection $s(n)$ is traditionally reduced to the problem of hypothesis testing

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

Hypothesis Γ_0 corresponds to the decision of receiving only noise, and hypothesis Γ_1 — of receiving a mixture of useful signal and noise, where the sequences $\{x(n)\}$, $n = 1, \dots, N$ are time series of the received data, $\{s(n)\}$ — useful signal, $\{w(n)\}$ — additive white Gaussian noise, N — the length of the time series of data.

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

In order to obtain an analytical expression for estimating the error probability in hypothesis testing, we can apply a variant of the Neyman–Pearson lemma [13, 16].

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

$$d(x_1, \dots, x_N) = \begin{cases} 1, & \text{hypothesis } \Gamma_0 \text{ is true,} \\ 0, & \text{hypothesis } \Gamma_1 \text{ is true,} \end{cases}$$

by which the following probabilities can be determined:

$$\begin{aligned} \alpha(d) &= \text{Probability (accept } \Gamma_0 | \Gamma_1 \text{ is true),} \\ \beta(d) &= \text{Probability (accept } \Gamma_1 | \Gamma_0 \text{ is true).} \end{aligned}$$

Then the decision rule d^* is optimal if

$$\alpha(d^*) + \beta(d^*) = \inf_d [\alpha(d) + \beta(d)] = \mathcal{E}r(N; \Gamma_0, \Gamma_1) \text{ — error function,} \tag{1}$$

where the infimum is taken for all tests.

Here $\alpha(\cdot)$ is the probability of a false alarm, and $\beta(\cdot)$ is the probability of a useful signal missing. The exact formula for the error function is as follows:

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

where $P_0^{(N)}$ is the multivariate distribution function of the observation statistics by hypothesis Γ_0 , $P_1^{(N)}$ is the multivariate distribution function of the observation statistics by hypothesis Γ_1 , and $TV(P_0, P_1)$ is the total variation of the signed measure, $\|Q\| = 2 \sup_A |Q(A)|$. Thus, if the supports of measures P_0, P_1 do not overlap, then error-free distinguishing of hypotheses is possible. If the measures $P_0^{(N)}$ and $P_1^{(N)}$ are close, then $\|P_0^{(N)} - P_1^{(N)}\| \approx 0$, leading to $\mathcal{E}r(N; \Gamma_0, \Gamma_1) \approx 1$.

For the problem of detecting a deterministic useful signal, for example, at a small signal-to-noise ratio, the case $\|P_0^{(N)} - P_1^{(N)}\| = 2TV(P_0, P_1) \approx 0$ is of interest and the possibility of reasonable estimation of this value. Therefore, when the probability of the total error of distinguishing two hypotheses is close to one, it becomes possible to use the analytical expression $TV(P_0, P_1)$ to design a criterion in the problem of detecting a useful signal in a mixture. But first let us turn to already known criteria and establish their properties.

Most often, for the convenience of mathematical investigation, both of the useful signal and noise are modeled by Gaussian random processes with different parameters. In that case the problem of finding the moment of appearance of the signal $s(n)$ in the received sequence of samples is called the the problem of changepoint detection [13].

Here and below, we consider discrete probability distributions $p = (p_1, \dots, p_i, \dots, p_N)$, that by definition have the following properties:

$$\forall p_i \in [0, 1], \quad \sum_{i=1}^N p_i = 1. \tag{3}$$

To formalize criterion that takes into account the deterministic component of the signal as well as the random one, let us explore the concepts of disequilibrium function D and statistical complexity C of the distribution. The simplest example of the disequilibrium function is the square of Euclidean distance in the space of discrete probability distributions [10].

Definition 1. The disequilibrium D_{SQ} has the meaning of the variance of a distribution relative to a uniform distribution

$$D_{SQ}(p) = \sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 = \sum_{i=1}^N p_i^2 - \frac{1}{N}. \tag{4}$$

Definition 2. The statistical complexity, defined through the expression of disequilibrium by the Definition 1, is equal to

$$C_{SQ}(p) = H(p) \times D_{SQ}(p), \tag{5}$$

where

$$H(p) = \frac{1}{\log N} \left(- \sum_i^N p_i \log p_i \right) \tag{6}$$

— Shannon’s normalized entropy [1].

In evaluating the sum (6), it is assumed that $\frac{0}{\log 0} = 0$ by continuity, and this assumption holds for all subsequent equations.

It follows from the Definition 1 that disequilibrium of the form (4) and complexity of the form (5) are convenient to apply in estimation and comparison of signals having spectral distribution close to uniform. In general, instead of a uniform distribution $q_i = 1/N$ at $i = 1, \dots, N$, the formula (4) may include an arbitrary discrete distribution.

The formula (4) is proposed in [10] for computing the disequilibrium with respect to a uniform distribution, but most studies use the Jensen–Shannon divergence $JSD(p||q)$ [17] instead.

Definition 3. The Jensen–Shannon disequilibrium equals

$$D_{JSD}(p) = JSD(p||q), \tag{7}$$

where $q = (1/N, \dots, 1/N)$ is the uniform distribution.

Definition 4. Statistical complexity defined through the expression of disequilibrium from Definition 3, is expressed as

$$C_{JSD}(p) = H(p) \times D_{JSD}(p). \tag{8}$$

Remark 1. It was noted above that $\sqrt{D_{SQ}}$ is a Euclidean metric on the space of discrete distributions. At the same time $\sqrt{D_{JSD}}$ is also a metric which is proportional to the Fisher metric.

Since the error function of distinguishing between two hypotheses depends on the total variation $TV(p, q)$, which is obtained in the Neyman–Pearson lemma 1, we introduce another notion of disequilibrium.

Definition 5. The disequilibrium based on the total variation of signed measure is equal to

$$D_{TV}(p) = TV^2(p, q), \tag{9}$$

where $q = (1/N, \dots, 1/N)$.

Definition 6. The statistical complexity, defined through the disequilibrium expression according to the Definition 5, is equal to

$$C_{TV}(p) = H(p) \times D_{TV}(p). \tag{10}$$

The information divergence functions presented above, which define different variants of the disequilibrium function, can be unified by the general concept of *f-divergence* [18]:

$$D_f(p||q) = \sum_{x \in \mathbb{R}^N} q(x) f \left(\frac{p(x)}{q(x)} \right). \tag{11}$$

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

- The Kulback–Leibler divergence $D_{KL}(p, q)$ is obtained from (11) by choosing $f(x) = x \log(x)$, $x > 0$.
- The Jensen–Shannon divergence is obtained from (11) by choosing

$$f(x) = x \log \frac{2x}{x+1} + \log \frac{2}{x+1}, \quad x > 0. \quad (12)$$

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

$$TV(p, q) = \frac{1}{2} \sum_{x \in \mathbb{R}^N} |p(x) - q(x)|; \quad (13)$$

$TV(p, q)$ is also a metric on the space of probability distributions. The total variation is related to the Jensen–Shannon divergence by the following relation:

$$JSD(p||q) \leq TV(p, q). \quad (14)$$

It follows from the inequality (14) that the total variation is the upper bound of Jensen–Shannon divergence.

Next, let us investigate the possibility of using each variant of statistical complexity as a criterion for indicating the appearance of a signal, but at first their properties must be established.

3. STATISTICAL COMPLEXITY OPTIMIZATION

3.1. Optimization of C_{SQ}

Let us formulate the problem of maximizing the statistical complexity function on the set of discrete distributions $p = (p_1, \dots, p_N)$

$$C_{SQ}(p) = \frac{1}{\log N} \left(- \sum_{i=1}^N p_i \log p_i \right) \left(\sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 \right) \longrightarrow \max_p \quad (15)$$

with the condition

$$\sum_{i=1}^N p_i = 1. \quad (16)$$

An auxiliary result will be needed to formulate the lemma about the maximum value of statistical complexity.

Lemma 2. *Let $0 < x \leq y \leq z \leq 1$, then $f(x, y, z) = x^y y^{-x} z^x x^{-z} y^z z^{-y} \geq 1$, with equality possible only when either $x = y$ or $y = z$.*

Proof. Let us introduce a new function $g(x, y, z) = \ln f(x, y, z)$,

$$g(x, y, z) = y \ln x - x \ln y + x \ln z - z \ln x + z \ln y - y \ln z.$$

Then it is required to prove that $g(x, y, z) \geq 0$, $0 < x \leq y \leq z \leq 1$.

By the Kuhn–Tucker theorem, the solution of the conditional optimization problem of a function of three variables is either at the interior point of the constraint manifold or at its boundary. The

necessary conditions for the unconditional extremum of the function $g(x, y, z)$ take the following form

$$\begin{aligned} \frac{\partial g}{\partial x} &= \ln z - \ln y + \frac{y - z}{x} = 0, \\ \frac{\partial g}{\partial y} &= \ln x - \ln z + \frac{z - x}{y} = 0, \\ \frac{\partial g}{\partial z} &= \ln y - \ln x + \frac{x - y}{z} = 0. \end{aligned} \tag{17}$$

Let us summarize all the equations of the last system:

$$\frac{y - z}{x} + \frac{z - x}{y} + \frac{x - y}{z} = 0,$$

which can be rewritten as

$$\frac{(y - z)(x - y)(z - x)}{xyz} = 0.$$

This means that when one of the equalities either $x = y$ or $y = z$ is satisfied, the function $g(x, y, z)$ possibly has a minimum. Let $x = y$, then the third equation from (17) is fulfilled identically, and the first and second equations are identical and can be written as

$$\ln \eta = \eta - 1, \quad \eta = \frac{z}{y}.$$

The last equation has only one root $\eta = 1$, i.e., $y = z$.

Let us calculate the second derivatives and write the Hesse matrix:

$$G(x, y, z) = \begin{pmatrix} \frac{z - y}{x^2} & \frac{1}{x} - \frac{1}{y} & \frac{1}{z} - \frac{1}{x} \\ \frac{1}{x} - \frac{1}{y} & \frac{x - z}{y^2} & \frac{1}{y} - \frac{1}{z} \\ \frac{1}{z} - \frac{1}{x} & \frac{1}{y} - \frac{1}{z} & \frac{y - x}{z^2} \end{pmatrix}. \tag{18}$$

Minors of the Hesse matrix are equal to

$$\begin{aligned} M_1(x, y, z) &= \frac{z - y}{x^2}, \\ M_2(x, y, z) &= -\frac{(x - y)^2 + (z - x)^2 + (y - z)^2}{2x^2y^2}, \\ M_3(x, y, z) &= 0. \end{aligned} \tag{19}$$

The Hesse matrix is not sign-defined, moreover, its determinant equals zero. Therefore, let us consider a small vicinity of the extremum point.

In a small vicinity $x = y = z$, provided that $\delta x \leq \delta y \leq \delta z$, the variation δg of the function $g(x, y, z)$ is written in the form of

$$\begin{aligned} \delta g &= (x + \delta y) \ln(x + \delta x) - (x + \delta x) \ln(x + \delta y) + (x + \delta x) \ln(x + \delta z) \\ &\quad - (x + \delta z) \ln(x + \delta x) + (x + \delta z) \ln(x + \delta y) - (x + \delta y) \ln(x + \delta z) \\ &= (\delta z - \delta x)(\delta y - \delta x)(\delta z - \delta y) + o(((\delta x)^2 + (\delta y)^2 + (\delta z)^2)^{3/2}) \geq 0, \end{aligned}$$

where values in the cubes of variations of the independent variables are nonzero, and the variation of the function $g(x, y, z)$ itself is positive by virtue of the lemma conditions. In the case when, for example, $\delta y = \delta x$, we have $g(x, y, z) \equiv 0$, and $f(x, y, z) \equiv 1$. Therefore, the extremum of the function is its non-strict minimum. ■

Lemma 3. *The maximum statistical complexity (15) is achieved on the distribution of the form*

$$\begin{cases} p_i = \frac{1 - p_{\max}}{N - 1}, & i = \overline{1, N} \setminus k, \\ p_k = p_{\max}, \end{cases} \tag{20}$$

where $p_{\max} = \text{const}$, i.e., at the appearance of a single component of an arbitrary index k over the uniform distribution.

Proof. Without loss of generality, let us assume $k = N$. From Eq. (16) one variable p_N from the set p_i can be expressed through all the others:

$$p_N = 1 - \sum_{i=1}^{N-1} p_i. \tag{21}$$

Let us rewrite the Eq. (15) in the form

$$C_{SQ} = -\frac{1}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \left(\sum_{i=1}^{N-1} \left(p_i - \frac{1}{N} \right)^2 + \left(p_N - \frac{1}{N} \right)^2 \right). \tag{22}$$

A necessary condition for the extremum of a function at an interior point of the domain (simplex 3) is that all partial derivatives of p_i are equal to zero:

$$\frac{\partial C_{SQ}}{\partial p_i} = 0, \quad i = 1, \dots, N - 1. \tag{23}$$

Substituting the function (22) into (23) gives (provided that $\frac{\partial p_N}{\partial p_i} = -1$):

$$\begin{aligned} \frac{\partial C_{SQ}}{\partial p_i} &= -\frac{1}{\log N} (\log p_i - \log p_N) \times \left(\sum_{i=1}^{N-1} \left(p_i - \frac{1}{N} \right)^2 + \left(p_N - \frac{1}{N} \right)^2 \right) \\ &- \frac{2}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \times (p_i - p_N) = 0, \quad i = 1, \dots, N - 1. \end{aligned} \tag{24}$$

In a more convenient form the equations can be rewritten as

$$\frac{\partial C_{SQ}}{\partial p_i} = \frac{1}{\log N} (-\log p_i + \log p_N) \times D + 2H \times (p_i - p_N) = 0, \quad i = 1, \dots, N - 1. \tag{25}$$

Let us write the difference of any two equations from the system above for indices i and j :

$$\frac{\partial C_{SQ}}{\partial p_i} - \frac{\partial C_{SQ}}{\partial p_j} = \frac{1}{\log N} (-\log p_i + \log p_j) \times D + 2H \times (p_i - p_j) = 0. \tag{26}$$

Given that the values of D and H are positive, the following equations can be constructed from the Eqs. (25) and (26), provided that the considered probabilities p_j , $j = 1, \dots, N - 1$ are not equal to p_N :

$$\frac{\log p_i - \log p_j}{\log p_N - \log p_j} - \frac{p_i - p_j}{p_N - p_j} = 0, \tag{27}$$

$$(p_N - p_j) \log p_i + (p_i - p_N) \log p_j + (p_j - p_i) \log p_N = 0, \tag{28}$$

$$p_i^{p_N - p_j} \times p_j^{p_i - p_N} \times p_N^{p_j - p_i} = 1. \tag{29}$$

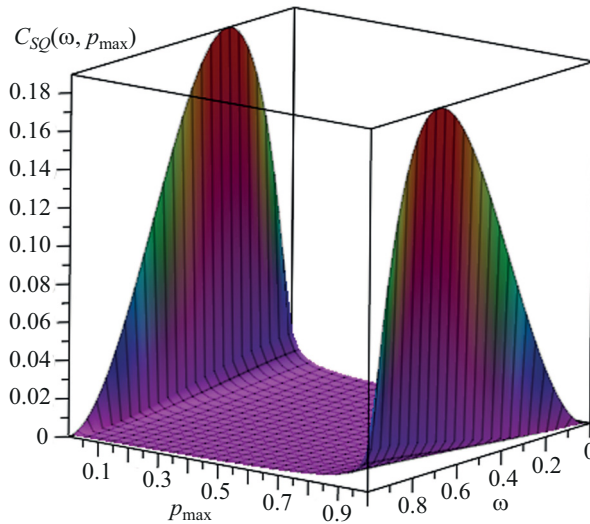


Fig. 1. Level surfaces of statistical complexity $C_{SQ}(\omega, p_{\max})$.

After applying the Lemma 2 we conclude that the last equation can be satisfied when $p_i = p_j$.

Thus, it is obtained that each of the probabilities p_i can take one of two different values, which define a distribution of the form

$$\begin{cases} p_i = \frac{1 - p_{\max}}{K}, & \forall i = 1, \dots, K, \\ p_i = p_N = \frac{p_{\max}}{N - K}, & \forall i = K + 1, \dots, N. \end{cases} \quad (30)$$

Now we need to show that the maximum complexity corresponds to values $K = 1$ and $K = N - 1$. For this purpose, let us calculate the value of the disequilibrium (4) on the distribution (30), which we denote by $D^{(K)}(\omega, p_{\max})$:

$$D^{(K)}(\omega, p_{\max}) = \frac{1}{N} \frac{(p_{\max} + \omega - 1)^2}{\omega(1 - \omega)}, \quad \omega = \frac{K}{N}. \quad (31)$$

In turn, entropy is equal to

$$H^{(K)}(\omega, p_{\max}) = 1 - \frac{1}{\log N} \left((1 - p_{\max}) \log \frac{1 - p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1 - \omega} \right). \quad (32)$$

The maximum of $C_{SQ}(\omega, p_{\max})$ at $N \leq 100$ was investigated numerically, and it was reached at $K = 1$. From the expression for $D^{(K)}(\omega, p_{\max})$ (31), it can be seen that at $N \geq 101$ and when changing from $K = 1$ to $K = 2$ or from $K = N - 1$ to $K = N - 2$, its value changes by almost a factor of two, while the entropy (32) changes only slightly. Thus, the probability distribution (30) that delivers the complexity function to the extremum value at $K = 1$ or $K = N - 1$ is of the form (20). ■

For clarity, Fig. 1 shows the graph $C_{SQ} = C_{SQ}(\omega, p_{\max})$ at $N = 1024$, where ω is changing continuously (although K is changing discretely).

Corollary 1. *Let us substitute the values of p_i and $p_N = p_{\max}$ from (20) into (22) and consider the complexity C_{SQ} as a function of p_{\max} . For sufficiently large values of N , it will take the following form*

$$C_{SQ} \approx (1 - p_{\max}) \times p_{\max}^2.$$

Whence it follows that this function takes the maximum value $C_{SQ}^* \approx 4/27$ when $p_{\max} = 2/3$.

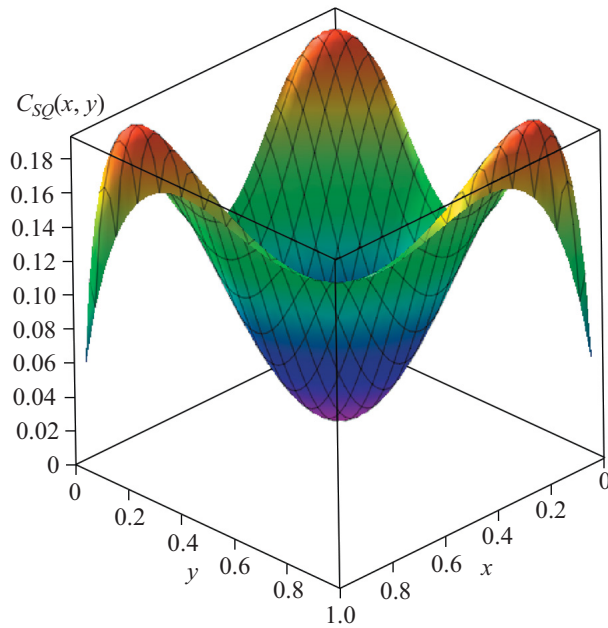


Fig. 2. Level surfaces of statistical complexity $C_{SQ}(x, y)$ for $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Corollary 2. *The minimum value of $C_{SQ} = 0$ is achieved on a uniform distribution $p = (1/N, \dots, 1/N)$.*

The validity of the Lemma 3 for the case when the discrete distribution $p = \{p_1, p_2, p_3\}$ consists of three samples is demonstrated in Fig. 2. The complexity depends on two variables, since one of the probabilities can be expressed through the others. Here C_{SQ} has three identical pronounced maxima and three identical local minima pertaining to the cases $p_1 = p_2, p_2 = p_3, p_1 = p_3$ when the necessary extremum conditions are met, and a global minimum when $p_1 = p_2 = p_3$.

Table 1 shows the change of optimal parameters $C_{SQ}(w, p_{\max})$ with increasing N .

Table 1. Optimal parameters $C_{SQ}(\omega, p_{\max})$ for different values of N

N	$C_{SQ}(\omega^*, p_{\max}^*)$	p_{\max}^*	ω^*	$N - K^*$
3	0.1932	0.8315	0.6666	1
256	0.1994	0.7044	0.9960	1
512	0.1942	0.7008	0.9980	1
1024	0.1898	0.6979	0.9990	1
2048	0.1861	0.6955	0.9995	1

The necessary extremum conditions C_{SQ} for the discrete distribution $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$ are written out according to (25) as follows:

$$\begin{cases} (-\log x + \log(1 - x - y)) \left(\left(x - \frac{1}{3}\right)^2 + \left(y - \frac{1}{3}\right)^2 + \left(1 - x - y - \frac{1}{3}\right)^2 \right) \\ - 2(x \log x + y \log y + (1 - x - y) \log(1 - x - y)) (-1 + y) = 0, \\ (-\log y + \log(1 - x - y)) \left(\left(x - \frac{1}{3}\right)^2 + \left(y - \frac{1}{3}\right)^2 + \left(1 - x - y - \frac{1}{3}\right)^2 \right) \\ - 2(x \log x + y \log y + (1 - x - y) \log(1 - x - y)) (-1 + x) = 0. \end{cases} \tag{33}$$

The implicit equations of the system (33) describe the curves shown in Fig. 3.

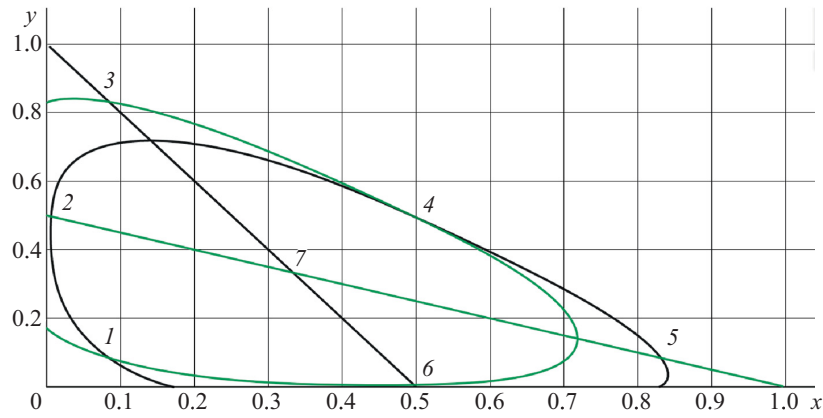


Fig. 3. Curves corresponding to the necessary conditions of extremum C_{SQ} for $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

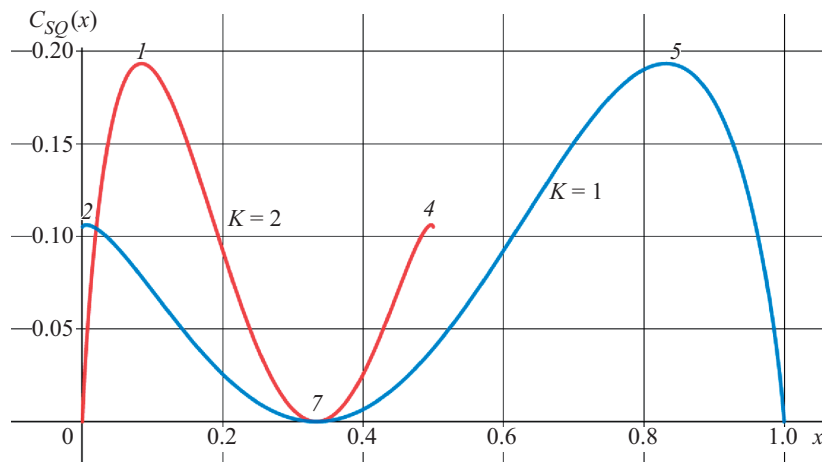


Fig. 4. Statistical complexity C_{SQ} for $p = \{p_1 = x, p_2 = x, p_3 = 1 - 2x\}$.

The black and green curves correspond to the first and second implicit equations of the (33) system, respectively. In Fig. 3 seven extremum points are marked, for which the value of statistical complexity is calculated. All the obtained data is summarized in Table 2.

The first, third and fifth points of maximum correspond to the same value of the maximum of the function. The second, fourth and sixth minimum points correspond to the same value of the minimum of the function. It is worth noting that the extremum points correspond to the case $K = N - 1 = 2$ except for the global minimum, where all probabilities are equal to each other, and thus describe three local minima, one global minimum, and three equal maxima of statistical complexity in Fig. 2. We can separately plot the statistical complexity in the case $p = \{p_1 = x, p_2 = x, p_3 = 1 - 2x\}$.

In Fig. 4 the extremum points are marked according to Table 2, which cover all cases $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Table 2. Extremum points of statistical complexity at $N = 3$

p	1	2	3	4	5	6	7
p_1	0.08425	0.006	0.08425	0.497	0.8315	0.497	0, (3)
p_2	0.08425	0.497	0.8315	0.497	0.08425	0.006	0, (3)
p_3	0.8315	0.497	0.08425	0.006	0.08425	0.497	0, (3)
C_{SQ}	0.1932	0.1062	0.1932	0.1062	0.1932	0.1062	0

3.2. Optimization of C_{JSD}

Let us apply a similar approach using the Jensen–Shannon divergence as the disequilibrium to the statistical complexity C_{JSD}

$$C_{JSD}(p) = H(p) \times JSD(p||q), \quad q_j = 1/N, \quad j = 1, \dots, N, \quad (34)$$

which can be written by considering the expression $JSD(p||q)$ through entropy

$$C_{JSD}(p) = H(p) \times \left(H(m) - \frac{1}{2}(H(p) + H(q)) \right) \times \log N, \quad m = \frac{p+q}{2}. \quad (35)$$

It is possible to write out the necessary conditions of extremum for the statistical complexity of the form (35), but at the same time a lemma similar to the Lemma 3 cannot be proved.

Equation (35) is written in variables p_i and p_N in accordance with the approach of the Lemma 3

$$C_{JSD}(p) = - \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \times \left(H(m) - \frac{1}{2} \left(-\frac{1}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) + 1 \right) \right), \quad (36)$$

where

$$H(m) = -\frac{1}{\log N} \left(\sum_{i=1}^{N-1} \frac{p_i + \frac{1}{N}}{2} \log \frac{p_i + \frac{1}{N}}{2} + \frac{p_N + \frac{1}{N}}{2} \log \frac{p_N + \frac{1}{N}}{2} \right). \quad (37)$$

Then taking into account (21)

$$\frac{\partial H(m)}{\partial p_i} = -\frac{1}{\log N} \left(\frac{1}{2} \log \frac{p_i + \frac{1}{N}}{2} - \frac{1}{2} \log \frac{p_N + \frac{1}{N}}{2} \right), \quad i = 1, \dots, N-1. \quad (38)$$

The necessary conditions of extremum are obtained in the following form after combining all partial derivatives:

$$\begin{aligned} \frac{\partial C_{JSD}(p)}{\partial p_i} &= \frac{H(p)}{2} \times \left(- \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_N + \frac{1}{N}}{2} \right) + (\log p_i - \log p_N) \right) \\ &- (\log p_i - \log p_N) \times JSD(p||q) = 0, \quad i = 1, \dots, N-1. \end{aligned} \quad (39)$$

The equations after simplification are given:

$$\begin{aligned} H(p) \times \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_N + \frac{1}{N}}{2} \right) \\ + (\log p_i - \log p_N) \times (2JSD(p||q) - H(p)) = 0, \quad i = 1, \dots, N-1. \end{aligned} \quad (40)$$

Then the difference of Eqs. (40) for indices i, j takes the following form

$$(\log p_i - \log p_j) \times (2JSD(p||q) - H(p)) + H(p) \times \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_j + \frac{1}{N}}{2} \right) = 0. \quad (41)$$

Remark 2. It follows from the form of the system of Eqs. (41) that the system is satisfied if $p_i = p_j$, which is one of the necessary conditions for the extremum of the function (35). Due to the nonlinearity of the system consisting of Eqs. (41), it may have other roots.

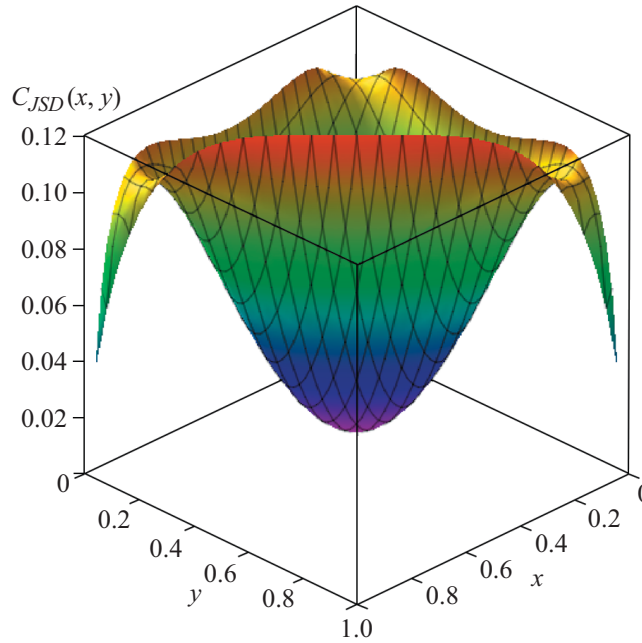


Fig. 5. Level surfaces of statistical complexity C_{JSD} for $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Figure 5 shows a surface plot of the statistical complexity level of the form (35) when the discrete distribution $p = \{p_1, p_2, p_3\}$ consists of three samples to illustrate the Remark 2.

It can be seen from Fig. 5 that the points satisfying $p_1 = p_2$, $p_2 = p_3$ and $p_2 = p_3$ are the saddle points of the surface if the necessary extremum conditions are satisfied.

It was previously established that the distribution (30) delivers an extremum to C_{SQ} at $K = N - 1$. Next, it will be shown that it also delivers the extremum to the complexity based on the total variation of the measure $TV(p, q)$. Therefore, we propose to find the maximum of C_{JSD} on this distribution and compare the obtained optimal distribution parameters at fixed N . Let us write out the complexity explicitly and obtain

$$C_{JSD}^{(K)} = H^{(K)} \times \left(H^{(K)}(m) - \frac{1}{2}(H^{(K)} + 1) \right) \times \log N, \tag{42}$$

where $H^{(K)}$ is corresponding to (32), and $H^{(K)}(m)$ is given by the following formula:

$$H^{(K)}(m) = 1 - \frac{1}{\log N} \left(\frac{(1 - p_{\max} + \omega)}{2} \log \frac{(1 - p_{\max} + \omega)}{2\omega} + \frac{(1 + p_{\max} - \omega)}{2} \log \frac{(1 + p_{\max} - \omega)}{2 - 2\omega} \right). \tag{43}$$

Table 3 shows the change of optimal parameters C_{JSD} with the growth of N .

Table 3. Optimal parameters $C_{JSD}(\omega, p_{\max})$ for different values of N

N	$C_{JSD}(\omega^*, p_{\max}^*)$	p_{\max}^*	ω^*	$N - K^*$
3	0.1266	1	0.4083	1 or 2
256	0.4482	1	0.8703	33
512	0.4790	1	0.8897	56
1024	0.5065	1	0.9051	97
2048	0.5312	1	0.9171	170

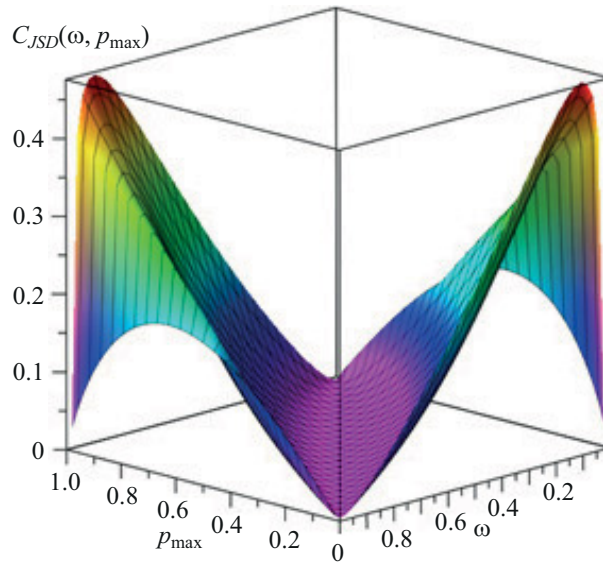


Fig. 6. Level surfaces of statistical complexity $C_{JSD}(\omega, p_{\max})$.

For clarity, Fig. 6 shows the graph $C_{JSD} = C_{JSD}(\omega, p_{\max})$ at $N = 1024$, where ω is changing continuously (although K is changing discretely).

The results shown in Table 3 demonstrate that for the chosen class of distributions (30) the set of N components, where K is equal to each other and the rest are zero, is optimal. It is worth noting that for the resulting distribution C_{JSD} is not zero, as well due to the summand $H^{(K)}(m)$, which corresponds to the already “shifted” distribution consisting of K elements equal to $\frac{\frac{1}{K} + \frac{1}{N}}{2}$, and $N - K$ samples of $\frac{1}{2N}$ each.

3.3. Optimization C_{TV}

Let us proceed to analyze statistical complexity based on total variation

$$C_{TV}(p) = -\frac{1}{4 \log N} \left(\sum_{i=1}^N p_i \log p_i \right) \times \left(\sum_{i=1}^N \left| p_i - \frac{1}{N} \right| \right)^2. \quad (44)$$

Proposition 1. *According to the expression for the error function (2) from the Neyman–Pearson Lemma 1 and the definition of (44), we propose to use C_{TV} as a criterion to solve the problem of hypothesis testing and indicating the appearance of a deterministic component of a useful signal in noise.*

The following lemma is valid.

Lemma 4. *The maximum statistical complexity (44) is achieved on the family of distributions (30).*

Proof. Given the symmetry of the function (44) and the simplex (3), without restriction of generality, we find an integer $K \in \{1, \dots, N - 1\}$ for which the maximum of this lemma is achieved on the part of the simplex (3) defined by the constraints $p_i \leq 1/N$ for $i = 1, \dots, K$ and $p_i \geq 1/N$ for $i = K + 1, \dots, N$. Let us rewrite the Eq. (44) in the form of

$$C_{TV} = -\frac{1}{4 \log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \times \left(\sum_{i=1}^K \left(-p_i + \frac{1}{N} \right) + \sum_{i=K+1}^N \left(p_i - \frac{1}{N} \right) \right)^2. \quad (45)$$

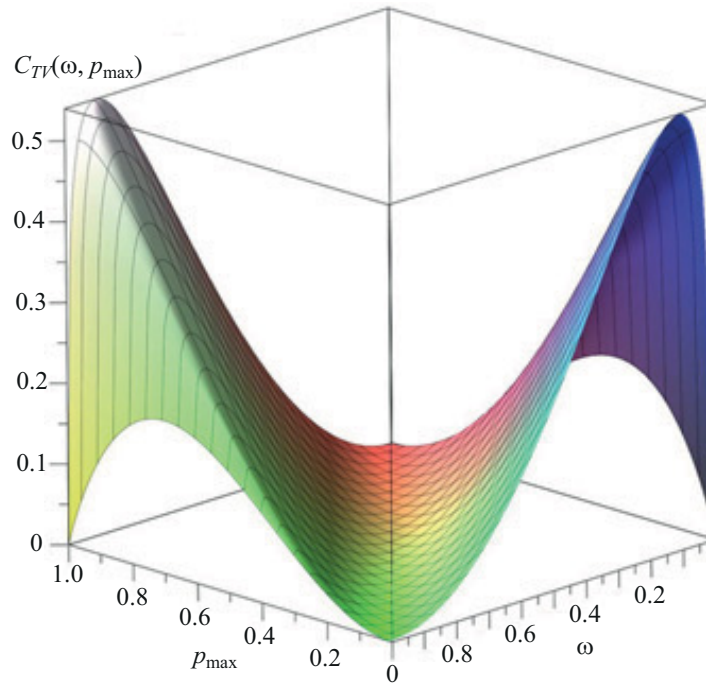


Fig. 7. Level surfaces of statistical complexity $C_{TV}(\omega, p_{\max})$.

Then for $i = 1, \dots, K$ the necessary conditions of extremum take the form

$$\frac{\partial C_{TV}}{\partial p_i} = -\frac{1}{\log N} (\log p_i - \log p_N) \times D_{TV} - 2H(p)\sqrt{D_{TV}} = 0, \quad i = 1, \dots, K, \quad (46)$$

and for $i = K + 1, \dots, N$ the following is true

$$\frac{\partial C_{TV}}{\partial p_i} = -\frac{1}{\log N} (\log p_i - \log p_N) \times D_{TV} = 0, \quad i = K + 1, \dots, N. \quad (47)$$

Let us compose the difference of two equations from (46) for indices i and j . Whence it follows that if $D_{TV} \neq 0$, then $p_i = p_j$ at $i = 1, \dots, K$. Whereas it follows from (47) that $p_i = p_N$ at $i = K + 1, \dots, N$. Again we obtain that the family of distributions (30) delivers the maximum of the complexity function, now C_{TV} . ■

Next, we need to determine the optimal values of K and p_{\max} . For this purpose, let us calculate the disequilibrium value $D_{TV}^{(K)}$ on the distribution (30):

$$D_{TV}^{(K)} = (p_{\max} + \omega - 1)^2, \quad \omega = \frac{K}{N}. \quad (48)$$

In turn, entropy is equal to

$$H^{(K)} = 1 - \frac{1}{\log N} \left((1 - p_{\max}) \log \frac{1 - p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1 - \omega} \right). \quad (49)$$

For clarity, Fig. 7 shows the graph $C_{TV} = C_{TV}(\omega, p_{\max})$ at $N = 1024$, where ω is changing continuously (although K is changing discretely).

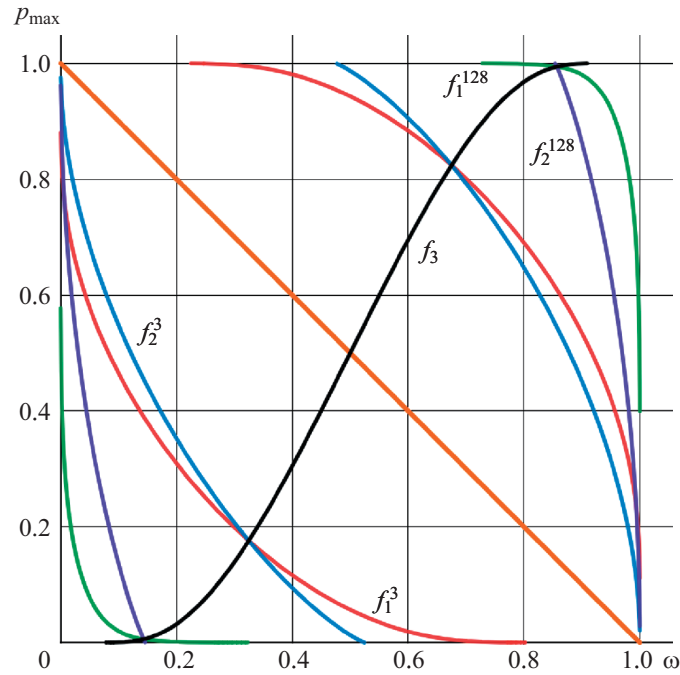


Fig. 8. Curves of necessary conditions of extremum C_{TV} for $N = 3$ and $N = 128$. Equations $f_1^N(p_{\max}, \omega)$ and $f_2^N(p_{\max}, \omega)$ have common solution $(p_{\max} + \omega - 1 = 0)$ (orange curve).

Let us compose the necessary conditions (46) and (47) of the extremum of the statistical complexity C_{TV} written out through the variables p_{\max}, ω .

$$\begin{cases} f_1^N(p_{\max}, \omega) := 2(p_{\max} + \omega - 1) \left(\left(1 - \frac{(1-p_{\max}) \log \frac{1-p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1-p_{\max}}}{\log N} \right) - \frac{(p_{\max} + \omega - 1)}{2 \log N} \left(\log \frac{p_{\max}}{1-\omega} - \log \frac{1-p_{\max}}{\omega} \right) \right) = 0, \\ f_2^N(p_{\max}, \omega) := 2(p_{\max} + \omega - 1) \left(\left(1 - \frac{(1-p_{\max}) \log \frac{1-p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1-p_{\max}}}{\log N} \right) - \frac{(p_{\max} + \omega - 1)}{2 \log N} \left(\frac{p_{\max}}{1-\omega} - \frac{1-p_{\max}}{\omega} \right) \right) = 0. \end{cases} \tag{50}$$

The intersections of the curves corresponding to the implicit Eqs. (50) are related to the extremum points of C_{TV} . Let us compose the difference of two necessary conditions of the extremum

$$\begin{aligned} f_3^N(p_{\max}, \omega) &:= f_1^N(p_{\max}, \omega) - f_2^N(p_{\max}, \omega) \\ &= \frac{(p_{\max} + \omega - 1)^2}{\log N} \left(-\log \frac{p_{\max}}{1-\omega} + \log \frac{1-p_{\max}}{\omega} + \frac{p_{\max}}{1-\omega} - \frac{1-p_{\max}}{\omega} \right) = 0. \end{aligned} \tag{51}$$

Let us construct the implicit curves of equations $f_1^N(p_{\max}, \omega), f_2^N(p_{\max}, \omega), f_3^N(p_{\max}, \omega)$ for some values of N . For convenience, the index N of $f_3^N(p_{\max}, \omega)$ can be omitted since the implicit curve of the Eq. (51) is independent of N .

According to Fig. 8 the statistical complexity has, in addition to the minimum points $(p_{\max} + \omega - 1 = 0 = 0)$, where $C_{TV} = 0$, also two maximum points for each value of N : (p_{\max}^*, ω^*) and $(1 - p_{\max}^*, 1 - \omega^*)$, which lie on the curve $f_3(p_{\max}, \omega)$.

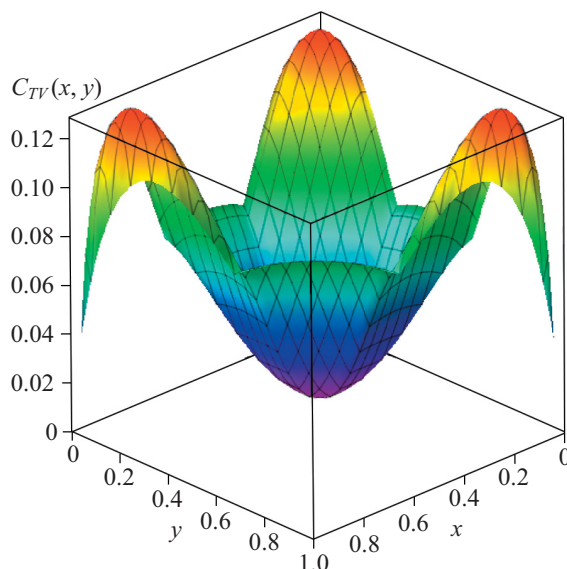


Fig. 9. Level surfaces of statistical complexity $C_{TV}(x, y)$ for $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

In Table 4 the optimal values of the parameters of the formulas (48) and (49) that maximize the statistical complexity of C_{TV} are given.

Table 4. Optimal parameters $C_{TV}(\omega, p_{\max})$ for different values of N

N	$C_{TV}(\omega^*, p_{\max}^*)$	p_{\max}^*	ω^*	$N - K^*$
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

Additionally, the case $N = 3$ is shown in Fig. 9, which shows a graph of the level surface C_{TV} when the discrete distribution $p = \{p_1, p_2, p_3\}$ consists of three samples.

4. STATISTICAL COMPLEXITY MODELING AND COMPARISON

We analyze the optimal parameters that maximize different types of statistical complexity and compare the values in Tables 1, 3, and 4. Of main interest are the maximum complexity values and the optimal values of K . The maximum values of $C_{TV}(\omega^*, p_{\max}^*) \in [0, 1]$, $C_{JSD}(\omega^*, p_{\max}^*) \in [0, 1]$ are close to each other and grow with increasing N . The optimal values of K for these two types of complexity are also close.

To demonstrate the analytical results obtained in the previous sections, the application of three variants of statistical complexity in the problem of useful signal indication in a noise mixture for synthesized signals is shown. An algorithm from [15] based on the computation of discrete distributions p from the spectral representation of time series is applied.

The synthesized 10-second signal is the sum of a finite number of cosine oscillations mixed with white noise:

$$x(t) = I(t) \sum_{i=1}^K A_i \cos(2\pi f_i t + \Delta\phi_i) + w(t), \quad t \in [0, 10], \tag{52}$$

where A_i , f_i , $\Delta\phi_i$ are the amplitudes, frequencies, and random phases of the harmonic oscillations, respectively, $w(t)$ is the white noise, and $I(t)$ is the indicator function for the presence of the useful signal in the signal-noise mixture.

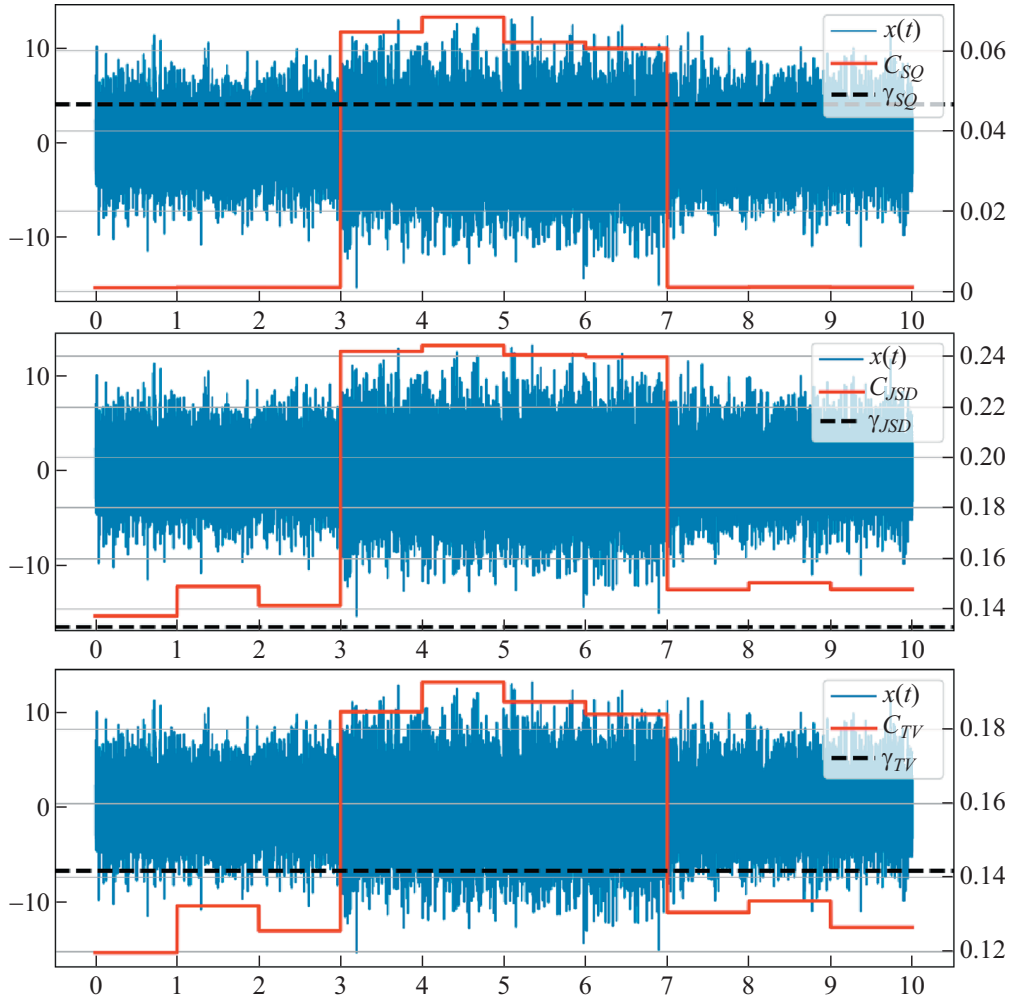


Fig. 10. Three components, $K = 3$.

$I(t)$ is chosen so that the harmonic signals are present in the middle of the final sequence $x(t)$.

$$I(t) = \begin{cases} 0, & t \in [0, 3), \\ 1, & t \in [3, 7], \\ 0, & t \in (7, 10]. \end{cases} \quad (53)$$

The algorithm has the following structure:

1. The signal synthesized with sampling frequency f_s is divided into short windows containing $N = 2048$ samples each.
2. Next, the spectrum for each window is calculated using the FFT algorithm.
3. Based on the spectrum, the discrete densities $p_i, i = 1, \dots, N$ are calculated by normalizing it.
4. The information characteristics $C_{SQ}(p), C_{JSD}(p), C_{TV}(p)$ are calculated for the obtained set p_i .
5. The obtained sequence of information characteristic values is shown along with the signal on the time axis.

It should be noted that the parameters f_s and N were chosen to exclude the effect of spectrum spreading, i.e., to obtain clear spectral components corresponding to K harmonic functions from the formula (52). The signal-to-noise ration is chosen to be close to one.

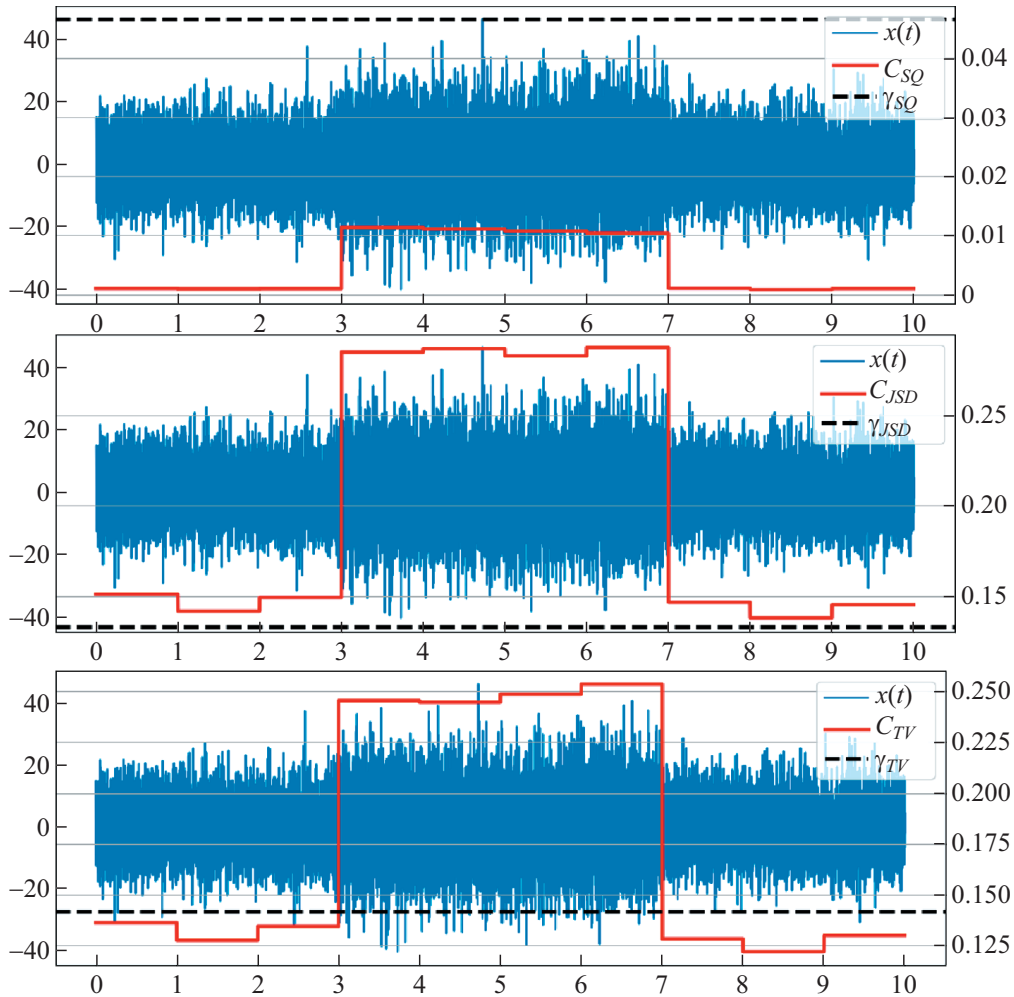


Fig. 11. Thirty components, $K = 30$.

The threshold γ for the decision rule is proposed to be chosen as 25% of the maximum criterion value for selected N from the Tables 1, 3, 4:

$$\begin{aligned}
 \gamma_{CQ} &= 0.25 \times 0.1861 = 0.0465; \\
 \gamma_{JS D} &= 0.25 \times 0.5312 = 0.1328; \\
 \gamma_{TV} &= 0.25 \times 0.5667 = 0.1417.
 \end{aligned}
 \tag{54}$$

The convenience of choosing such a threshold is that it does not depend on a particular noise realization and is based on analytically derived maximum values of statistical complexity functions.

In all plots, the blue color indicates the amplitude of the original signal, and the red color indicates the statistical complexity, which is calculated using the algorithm described above. The horizontal axis represents time in seconds, and the vertical axes represent the magnitude of the signal amplitude (left) and the criterion (right). The black dashed line shows the value of the selected threshold γ .

In the first experiment, the number of sinusoidal signals and respectively spectral components is equal to $K = 3$ at $N = 2048$. Figure 10 shows the dependencies of statistical complexities on time for the synthesized signal.

As can be seen, the values of C_{SQ} and C_{TV} exceed the selected threshold for the interval of signal presence, which allows us to confidently conclude that the signal has occurred. As for $C_{JS D}$,

the a priori threshold selection was unsuccessful because the true value of its maximum is unknown, as shown in Section 3.2. If we change the threshold value upwards by 20%, the detection based on C_{JSD} will be as successful as that based on C_{TV} .

In the second experiment, the number of spectral components $K = 30$. In this case, C_{SQ} ceases to show a satisfactory result in the sense of exceeding the chosen threshold, since the function C_{SQ} degrades strongly with increasing of K , but still allows a signal indication, as can be seen in Fig. 11. The complexity function C_{TV} still confidently exceeds the threshold, as in the first experiment, and C_{JSD} exceeds the threshold on the whole signal, as in the first experiment.

Thus, we can conclude that C_{TV} is the most convenient in a practical sense, since it works well on signals with a large number of spectral components and allows one to decide on the appearance of an useful signal, using a fairly simple rule, associated with the choice of threshold based on the theoretical maximum value for the statistical complexity.

5. CONCLUSION

The article provides a theoretical justification for the usage of statistical complexity as a criterion for solving the problem of hypothesis testing when its error probability is close to one. Three variants of statistical complexity for different disequilibrium functions are considered. New notions of disequilibrium and statistical complexity based on the total variation measure are introduced. Information criteria are compared and the classes of discrete distributions which provide maximum for different types of statistical complexity are discovered. The values of maxima for fixed numbers of distribution samples are found. It is shown that the statistical complexity C_{TV} based on total variation is directly related to the problem of hypothesis testing, while the statistical complexity C_{JSD} based on Jensen–Shannon entropy gives a close estimate of C_{TV} on sample distributions. In turn, C_{SQ} is most promising for detecting an individual component over a uniform distribution. We propose a method for selecting the threshold for the decisive rule for the detection of a useful signal, taking into account the maximum values of the criteria obtained and show the effectiveness of this approach on the synthesized signals.

Future work will be devoted to the study of information criteria based on bivariate and multivariate distributions, as well as investigation of typical acoustic signals with realistic background noise.

FUNDING

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

REFERENCES

1. Shannon, C.E., A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, 1948, vol. 27, pp. 379–423.
2. Gray, R.M., *Entropy and Information Theory*, New York: Springer, 2011. <https://doi.org/10.1007/978-1-4419-7970-4>
3. Holub, A., Perona, P., and Burl, M.C., Entropy-based Active Learning for Object Recognition, *Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, *IEEE*, 2008, pp. 1–8. <https://doi.org/10.1109/CVPRW.2008.4563068>
4. Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., et al. Supervised Machine Learning Algorithms: Classification and Comparison, *Int. J. Comput. Trends Technol. (IJCTT)*, 2017, vol. 48, no. 3, pp. 128–138. <https://doi.org/10.14445/22312803/IJCTT-V48P126>

5. Shen, J., Hung, J., and Lee, L., Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments, *Proc. 5th International Conference on Spoken Language Processing (ICSLP)*, 1998. <https://doi.org/10.21437/icslp.1998-527>
6. Ribeiro, M., Henriques, T., Castro, L., Souto, A., Antunes, L., Costa-Santos, C., and Teixeira, A., The Entropy Universe, *Entropy*, 2021, vol. 23, no. 2, art. 222. <https://doi.org/10.3390/e23020222>
7. Ramirez, J., Segura, J.C., Benitez, C., et al., A New Kullback-Leibler VAD for Speech Recognition in Noise, *IEEE Signal Proc. Lett.*, 2004, vol. 11, no. 2, pp. 266–269. <https://doi.org/10.1109/LSP.2003.821762>
8. Horie, T., Burioka, N., Amisaki, T., and Shimizu, E., Sample Entropy in Electrocardiogram During Atrial Fibrillation, *Yonago Acta Medica*, 2018, vol. 61, no. 1, pp. 49–57. <https://doi.org/10.33160/yam.2018.03.007>
9. Lamberti, P.W., Martin, M.T., Plastino, A., and Rosso, O.A., Intensive Entropic Non-Triviality Measure, *Phys. A: Stat. Mech. Appl.*, 2004, vol. 334, no. 1, pp. 119–131. <https://doi.org/10.1016/j.physa.2003.11.005>
10. Lopez-Ruiz, R., Shannon Information, LMC Complexity and Renyi Entropies: A Straightforward Approach, *Biophys. Chem.*, 2005, vol. 115, no. 3, pp. 215–218. <https://doi.org/10.1016/j.bpc.2004.12.035>
11. Zunino, L., Soriano, M.C., and Rosso, O.A., 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. 1–5. <https://doi.org/10.1103/PhysRevE.86.046210>
12. Ronald, L.A. and Duncan, W.M., *Signal Analysis: Time, Frequency, Scale, and Structure*, New Jersey: IEEE Press, 2004.
13. Shiryaev, A.N., *Veroyatnostno-statisticheskie metody v teorii prinyatiya reshenii* (Probabilistic-Statistical Methods in Decision-Making Theory), Moscow: MTSNMO: NMU, 2020.
14. Kishan, G.M., Chilukuri, K.M., and HuaMing Huang, *Anomaly Detection Principles and Algorithms*, Cham: Springer, 2017. <https://doi.org/10.1007/978-3-319-67526-8>
15. 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. 2133. <https://doi.org/10.3390/s23042133>
16. Johnson, P., Moriarty, J., and Peskir, G., Detecting Changes in Real-Time Data: A User's Guide to Optimal Detection, *Philos. Trans. Royal Soc. A*, 2017, vol. 375, p. 16, art. 2100. <https://doi.org/10.1098/rsta.2016.0298>
17. Li, Z., Li, Y., and Zhang, K.A., Feature Extraction Method of Ship-Radiated Noise Based on Fluctuation-Based Dispersion Entropy and Intrinsic Time-Scale Decomposition, *Entropy*, 2019, vol. 21, no. 7, art. 693. <https://doi.org/10.3390/e21070693>
18. Sason, I., On f-Divergences: Integral Representations, Local Behavior, and Inequalities, *Entropy*, 2018, vol. 20, no. 5, art. 383. <https://doi.org/10.3390/e20050383>

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