Chaos, Solitons and Fractals 193 (2025) 116085

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Asymptotic distribution of the statistical complexity under the multinomial law☆

Andrea Rey [a], Alejandro C. Frery [b],*, Juliana Gambini [a]

[a] LIDEC, Universidad Nacional de Hurlingham (UNAHUR), Hurlingham, Pcia. de Buenos Aires, Argentina and CPSI, Universidad TecnológicaNacional, FRBA, Buenos Aires, Argentina
[b] School of Mathematics and Statistics, Victoria University of Wellington, New Zealand

ARTICLE INFO

Keywords:
Statistical complexity
Asymptotic distribution
Multinomial distribution

ABSTRACT

The Statistical Complexity is a feature computed from a probability function that aims to quantify the structure of the system that produced the observations. It is the product between the normalized Shannon Entropy and the normalized Jensen–Shannon distance between the probability function and the uniform law. We obtain the Statistical Complexity asymptotic distribution under the Multinomial model, and we validate this result with numerical experiments. We present examples where this asymptotic result provides a good approximation, even in scenarios where the Multinomial model is not strictly valid, such as in applications to Bandt and Pompe ordinal patterns. We provide the R code that implements these functions.

1. Introduction

The concept of Entropy first appeared in the field of thermodynamics as a way to measure the amount of disorder in a system. While Clausius [1] introduced this idea in 1850, it was not until 1877 that Boltzmann [2] provided a mathematical expression concerning a probability distribution. Later, Shannon [3] proposed the use of Entropy in the context of Information Theory to study the uncertainty associated with the distribution of a random variable. However, Huberman and Hogg [4] exposed the limitation of the Entropy as a quantifier. They identified the need of measuring systems' diversity in a computational way that vanishes at the extremes of minimum and maximum Entropy (a constant output and white noise, respectively.) Grassberg [5], using the grammar of formal languages with probabilities, reinforced that such measures should stem from a stochastic process. Crutchfield and Young [6] obtained a measure of complexity using the Rényi entropy on the spectral decomposition of transition matrices of the symbols emitted by the system. These ideas were further developed in Ref. [7]. López-Ruiz et al. [8] proposed that the complexity cannot be defined only in terms of order or information and used the disequilibrium, a distance between the system's distribution to the equiprobable model, to describe the complexity as a combination of two concepts: the Information and the Disequilibrium. used the pair "Entropy – Complexity" to describe the Logistic and Lorenz maps.

In this framework, Feldman and Crutchfield [9] verified that a quadratic measure of disequilibrium yields a non-extensive quantity that vanishes for a large variety of structured processes. Thus, they proposed to replace it with the Kullback–Leibler divergence or relative entropy, with the drawback this quantity is not symmetric.

As an attempt to solve this asymmetry, Lamberti et al. [10] used the Jensen–Shannon divergence, which is a distance, instead of the Kullback–Leibler divergence and obtained an intensive measure of Disequilibrium. Along with the normalized Shannon Entropy, the authors used this redefined "Entropy – Statistical Complexity" feature to analyze the Logistic map.

This "Entropy – Statistical Complexity" approach to signal analysis has been widely used in the literature for characterizing complex systems. See, for instance, the review papers by Zanin et al. [11] and Amigó and Rosso [12].

Despite its deserved success, the approach lacks a complete statistical characterization. Chagas et al. [13] studied the empirical distribution of the points in the Entropy – Statistical Complexity representation using true white noise sequences from physical devices. They verified that such joint distribution is heavily skewed and, thus, cannot be directly described by a bivariate Normal law. Rey et al. [14] provided an asymptotic distribution of several forms of entropy and the Fisher information measure under the Multinomial model. These results were extended by Rey et al. [15] to incorporate the correlation inherent to

---

☆ This work has been partially supported by GRANTS 80020240200003HU, Universidad Nacional de Hurlingham (UNAHUR), and 410695, Victoria University of Wellington.

* Corresponding author.
E-mail addresses: andrea.rey@unahur.edu.ar (A. Rey), alejandro.frery@vuw.ac.nz (A.C. Frery), juliana.gambini@unahur.edu.ar (J. Gambini).

https://doi.org/10.1016/j.chaos.2025.116085
Received 25 December 2024; Received in revised form 12 January 2025; Accepted 29 January 2025

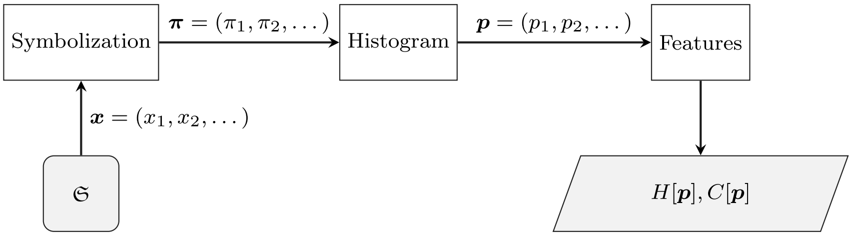

Fig. 1. From a system to its analysis through features from the histogram of symbols.

the overlapped sampling of time series realizations.

The stochastic properties of the Statistical Complexity have not been fully characterized yet. We fill this gap obtaining its asymptotic distribution under the Multinomial hypothesis. Then, we show that this is a good approximation for ordinal patterns from time series for which the Multinomial model does not hold.

The remainder of this article is as follows. Section 2 outlines the pipeline for system analysis with the Entropy and the Statistical Complexity. In Section 3, we formulate the Statistical Complexity according to Ref. [8]. Section 4 presents the main properties of the Multinomial distribution. Our main result is presented in Section 5: first, we recall the asymptotic distribution of the Shannon Entropy, and we then obtain the asymptotic distribution of the Statistical Complexity under the Multinomial distribution. We extended the `StatOrdPattHxC` package available at https://github.com/arey1911/StatOrdPattHxC/ with these functions. In Section 6, we present experiments to validate the theoretical results: first, under the Multinomial model (Section 6.1), and then to patterns that result from the Bandt & Pompe symbolization strategy (Section 6.2). We conclude our article in Section 7.

2. From systems to entropy and complexity

Fig. 1 outlines a pipeline of transformations from a system ($\mathfrak{S}$) up to its analysis using symbolic information. A non-isolated system interacts with the environment and produces various types of observations along time $\boldsymbol{x} = (x_1, x_2, \dots)$. Such observations may be textual, numerical, interval, ordinal, categorical, and graphs, among other types. Their nature dictates the choice of appropriate analytical methods to apply.

For example, Parsons [16] studied over 10 000 tunes (the system $\mathfrak{S}$). The author encodes the tunes (the "Symbolization" step) into sequences of three letters: "R" (repeats the previous note), "U" (up), and "D" (down). He finds that the first note and up to fifteen letters are enough to differentiate any number of tunes. If the probability of each note does not change, and if the tunes are independent, the 3^{15} symbols follow a Multinomial law characterized by a probability vector of dimension fifteen.

The books by Agresti [17,18] are classical references for analyzing categorical and ordinal observations. A different approach, outlined in Fig. 1, consists of obtaining features, e.g., the Shannon entropy $H[\boldsymbol{p}]$, from the probability function ($\boldsymbol{p}$) of symbols. This approach has been particularly successful when the symbols $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots)$ result from a symbolization operation of real-valued measurements.

Transforming observations into symbols is a unified approach for several types of data. On the one hand, the transformation may be performed using disjoints blocks of data, e.g., using windows of size D, and $\pi_t = \pi_t(x_t, x_{t+1}, \dots, x_{t+D-1})$, $\pi_{t+1} = \pi_{t+1}(x_{t+D}, x_{t+D+1}, \dots, x_{t+2D-1})$, and so on. On the other hand, the symbols may use overlapping sets of data, for instance, $\pi_t = \pi_t(x_t, x_{t+1}, \dots, x_{t+D-1})$ and $\pi_{t+1} = \pi_{t+1}(x_{t+1}, x_{t+2}, \dots, x_{t+D})$, as in Ref. [19]. For the same underlying process, the symbols obtained with the first approach will be typically less correlated than those from the second, but in both cases the Multinomial distribution is a simplification.

In the following section, we recall the definitions of the Shannon Entropy and Statistical Complexity from the probability function $\boldsymbol{p}$.

3. Statistical complexity

In this section, the Statistical Complexity is defined according to [8]. Consider a system with $\{o^1, o^2, \dots, o^k\}$ possible states as outcomes, where o^ℓ occurs with probability $p_\ell \geq 0$, for $\ell = 1, 2, \dots, k$, and $\sum_{\ell=1}^{k} p_\ell = 1$. The Shannon information or Entropy of the probability function $\boldsymbol{p} = (p_1, p_2, \dots, p_k)$ [3] is defined by

$$H_S[\boldsymbol{p}] = -\sum_{\ell=1}^{k} p_\ell \ln p_\ell. \tag{1}$$

An isolated system is in equilibrium when all the states are equiprobable; i.e. the associated probability function is $\boldsymbol{u} = (1/k, 1/k, \dots, 1/k)$. In this case, $H_S[\boldsymbol{u}] = \ln k$ is the maximum entropy. The normalized Shannon Entropy is defined by

$$H[\boldsymbol{p}] = \frac{1}{\ln k} H_S[\boldsymbol{p}]. \tag{2}$$

It holds that $0 \leq H[\boldsymbol{p}] \leq 1$. The concept of Disequilibrium arises as a kind of distance to a reference probability density function. In this work, we consider $\boldsymbol{u}$ as the reference distribution. Let $\mathscr{D}$ be the Jensen–Shannon divergence [20] between $\boldsymbol{p}$ and $\boldsymbol{u}$, given by

$$\mathscr{D}[\boldsymbol{p}, \boldsymbol{u}] = H_S\left[\frac{\boldsymbol{p}+\boldsymbol{u}}{2}\right] - \frac{1}{2} H_S[\boldsymbol{p}] - \frac{1}{2}\ln k. \tag{3}$$

Expanding this expression, we have the following equivalent formula:

$$\begin{aligned}\mathscr{D}[\boldsymbol{p}, \boldsymbol{u}] &= -\sum_{\ell=1}^{k} \frac{p_\ell + u_\ell}{2} \ln \frac{p_\ell + u_\ell}{2} + \frac{1}{2}\sum_{\ell=1}^{k} p_\ell \ln p_\ell - \frac{1}{2}\ln k \\ &= \frac{1}{2}\left(\sum_{\ell=1}^{k} p_\ell \ln p_\ell - \sum_{\ell=1}^{k} (p_\ell + u_\ell) \ln \frac{p_\ell + u_\ell}{2} - \ln k\right).\end{aligned} \tag{4}$$

Then, the Disequilibrium can be defined by

$$Q[\boldsymbol{p}] = Q_0 \mathscr{D}[\boldsymbol{p}, \boldsymbol{u}], \tag{5}$$

where $Q_0 = -2\left[(k+1)k^{-1}\ln(k+1) - 2\ln(2k) + \ln k\right]^{-1}$ is a normalization constant [21, Eq. (24)].

Finally, following the approach introduced in [21], the Statistical Complexity of the vector of probability $\boldsymbol{p}$ can be defined by

$$C[\boldsymbol{p}] = Q[\boldsymbol{p}] H[\boldsymbol{p}]. \tag{6}$$

4. Multinomial distribution

Meanwhile the Binomial distribution counts the number of success in a sequence of n independent trials of Bernoulli's essays with probability of success equal to p (i.e., there are only two possible outcomes), the Multinomial distribution is a generalization for $k > 2$ outcomes. Thus, the Multinomial distribution is adequate to model the probability of each outcome occurring in an experiment that has multiple categories.

In this work, we consider a process that has k accessible states with the associated vector of probability $\boldsymbol{p} = (p_1, p_2, \dots, p_k)$. Let $\boldsymbol{N} = (N_1, N_2, \dots, N_k)$ be the random vector that counts the number of occurrences of these states in n independent trials, with $\sum_{\ell=1}^{k} N_\ell = n$. Then, the joint distribution of $\boldsymbol{N}$ is

$$\Pr\big(\boldsymbol{N} = (n_1, n_2, \dots, n_k)\big) = n! \prod_{\ell=1}^{k} \frac{p_\ell^{n_\ell}}{n_\ell!}, \tag{7}$$

where $n_\ell \geq 0$ and $\sum_{\ell=1}^{k} n_\ell = n$. In other words, $\boldsymbol{N}$ has a Multinomial distribution with parameters n and $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$, which is denoted by $\boldsymbol{N} \sim \text{Mult}(n, \boldsymbol{p})$. In this case, the maximum likelihood estimator (MLE) of $\boldsymbol{p}$ is $\widehat{\boldsymbol{p}}_n = \boldsymbol{N}/n$.

It is known that

$$\sqrt{n}(\widehat{\boldsymbol{p}}_n - \boldsymbol{p}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{p}}), \tag{8}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{p}} = \mathbf{D}_{\boldsymbol{p}} - \boldsymbol{p}'\boldsymbol{p}$, $\boldsymbol{p}'$ denotes the transpose of $\boldsymbol{p}$, and $\mathbf{D}_{\boldsymbol{p}} = \text{Diag}(p_1, p_2, \ldots, p_k)$. Therefore:

$$(\boldsymbol{\Sigma}_{\boldsymbol{p}})_{\ell j} = \begin{cases} p_\ell(1 - p_\ell) & \text{if } \ell = j, \\ -p_\ell p_j & \text{if } \ell \neq j. \end{cases} \tag{9}$$

5. Asymptotic results

In this section we compute the asymptotic distribution of the Statistical Complexity. We need the asymptotic distribution of its two components: the Entropy (Section 5.1) and the Disequilibrium (Section 5.2). The Entropy and the Disequilibrium are asymptotically normally distributed, but they are not independent, so we rely on a result about the product of two correlated Normal random variables to obtain the asymptotic distribution of the Statistical Complexity (Section 5.3). We obtain their exact correlation with a variables transformation (Appendix, Lemma 1). The resulting distribution is numerically intractable, therefore, we apply another approximation and obtain a useful asymptotic distribution for the Statistical Complexity: a Normal law with mean and variance given in (43) and (44), respectively.

The obtained results are based on the multivariate version of the Delta Method [22, Theorem 8.22]. This technique requires the definition of real-functions continuously differentiable in a neighborhood of the parameter point (in our case, $\boldsymbol{p}$) and whose matrix of partial derivatives is non-singular in this neighborhood.

Due to our interest in finding the asymptotic distribution of unidimensional variables, we recall the following result [23, Theorem 5.2]. Let $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a k-dimensional multivariate Normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^k$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{\ell j}) \in \mathbb{R}^{k\times k}$. For $\boldsymbol{a} \in \mathbb{R}^k$, define the unidimensional random variable U as the inner product $U = \langle \boldsymbol{a}, \boldsymbol{Z}\rangle$, then

$$U \sim \mathcal{N}\left(\langle \boldsymbol{a}, \boldsymbol{\mu}\rangle, \sum_{\ell=1}^{k} a_\ell^2 \sigma_{\ell\ell} + 2\sum_{\ell=1}^{k-1}\sum_{j=\ell+1}^{k} a_\ell a_j \sigma_{\ell j}\right). \tag{10}$$

5.1. Asymptotic distribution of the entropy

The Entropy is crucial in the treatment of the Statistical Complexity since it is one of the factors in (6).

Ref. [14] proves that the Shannon entropy of the sequence $\widehat{\boldsymbol{p}}_n$ converges in distribution since

$$\sqrt{n}\left(H[\widehat{\boldsymbol{p}}_n] - H[\boldsymbol{p}]\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma_{\boldsymbol{p}}^2\right), \tag{11}$$

where

$$\sigma_{\boldsymbol{p}}^2 = \frac{1}{\ln^2 k}\left[\sum_{\ell=1}^{k} p_\ell(1-p_\ell)(\ln p_\ell + 1)^2 - 2\sum_{j=1}^{k-1}\sum_{\ell=j+1}^{k} p_j p_\ell(\ln p_j + 1)(\ln p_\ell + 1)\right]. \tag{12}$$

The proof relies on the multivariate Delta Method applied to the sequence $\widehat{\boldsymbol{p}}_n$.

5.2. Asymptotic distribution of the disequilibrium

We apply the multivariate Delta Method to the sequence $\widehat{\boldsymbol{p}}_n$ to find the asymptotic distribution of the Jensen–Shannon divergence $\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]$, using the following functions:

$$h_\ell(p_1, p_2, \ldots, p_k) = p_\ell \ln p_\ell - (p_\ell + 1/k)\ln\frac{p_\ell + 1/k}{2}, \tag{13}$$

for $\ell = 1, 2, \ldots, k$. Direct computations show that

$$\frac{\partial h_\ell}{\partial p_\ell} = (1 + \ln p_\ell) - \left(1 + \ln\frac{p_\ell + 1/k}{2}\right) = \ln p_\ell - \ln\frac{p_\ell + 1/k}{2}, \text{ and}$$
$$\frac{\partial h_j}{\partial p_\ell} = 0 \text{ if } j \neq \ell. \tag{14}$$

Thus, using (8), the covariance matrix of the multivariate Normal limit distribution is

$$\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\Delta} = \left(\frac{\partial h_\ell}{\partial p_j}\right)\boldsymbol{\Sigma}_{\boldsymbol{p}}\left(\frac{\partial h_\ell}{\partial p_j}\right)'. \tag{15}$$

After matrix product calculation, we obtain that

$$(\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\Delta})_{\ell j} = \begin{cases} (p_\ell - p_\ell^2)\left(\ln p_\ell - \ln\frac{p_\ell+1/k}{2}\right)^2 & \text{if } \ell = j, \\ -p_\ell p_j\left(\ln p_\ell - \ln\frac{p_\ell+1/k}{2}\right)\left(\ln p_j - \ln\frac{p_j+1/k}{2}\right) & \text{if } \ell \neq j. \end{cases} \tag{16}$$

On the one hand, the conditions for the multivariate Delta Method are verified if $p_\ell \neq 0$ for all $\ell = 1, 2, \ldots, k$. On the other hand, suppose that there exists an unique $\ell_0 \in \{1, 2, \ldots, k\}$ such that $p_{\ell_0} = 0$, then the ℓ_0th row and the ℓ_0th column of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{p}}$ are null. We consider $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{p}}$ the matrix obtained from $\boldsymbol{\Sigma}_{\boldsymbol{p}}$ by removing the ℓ_0th row and the ℓ_0th column. In this case, the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\Delta}$ is computed as in (15) replacing $\boldsymbol{\Sigma}_{\boldsymbol{p}}$ by $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{p}}$ and using the partial derivatives $\partial h_\ell/\partial p_j$ for all $\ell, j \in \{1, 2, \ldots, k\}$, such that $\ell \neq \ell_0$ and $j \neq \ell_0$. The procedure is analogous if more than one probability vanishes.

Hence, we conclude that

$$\sqrt{n}\left[h_1(\widehat{p}_1) - h_1(p_1), h_2(\widehat{p}_2) - h_2(p_2), \ldots, h_k(\widehat{p}_k) - h_k(p_k)\right] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{p}}^{\Delta}\right). \tag{17}$$

Applying (10) to (17) with $\boldsymbol{a} = (1/2, 1/2, \ldots, 1/2)$, we directly have the asymptotic distribution:

$$\begin{aligned} &\sqrt{n}\left(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}] - \mathscr{D}[\boldsymbol{p}, \boldsymbol{u}]\right) = \\ &\quad \frac{\sqrt{n}}{2}\left(\sum_{\ell=1}^{k} \widehat{p}_\ell \ln \widehat{p}_\ell - \sum_{\ell=1}^{k}(\widehat{p}_\ell + u_\ell)\ln\frac{\widehat{p}_\ell + u_\ell}{2} - \sum_{\ell=1}^{k} p_\ell \ln p_\ell \right. \\ &\quad \left. + \sum_{\ell=1}^{k}(p_\ell + u_\ell)\ln\frac{p_\ell + u_\ell}{2}\right) \\ &\quad \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma_{\mathscr{D}\boldsymbol{p}}^2\right), \end{aligned} \tag{18}$$

where

$$\begin{aligned} \sigma_{\mathscr{D}\boldsymbol{p}}^2 = {} & \frac{1}{4}\sum_{\ell=1}^{k}(p_\ell - p_\ell^2)\left(\ln p_\ell - \ln\frac{p_\ell + 1/k}{2}\right)^2 - \\ & \frac{1}{2}\sum_{j=1}^{k-1}\sum_{\ell=j+1}^{k} p_\ell p_j\left(\ln p_\ell - \ln\frac{p_\ell + 1/k}{2}\right)\left(\ln p_j - \ln\frac{p_j + 1/k}{2}\right). \end{aligned}$$

Hence, the asymptotic distribution for the Disequilibrium is

$$\sqrt{n}\left(Q[\widehat{\boldsymbol{p}}] - Q[\boldsymbol{p}]\right) = \sqrt{n}\left(Q_0\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}] - Q_0\mathscr{D}[\boldsymbol{p}, \boldsymbol{u}]\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, Q_0^2\sigma_{\mathscr{D}\boldsymbol{p}}^2\right). \tag{19}$$

5.3. Asymptotic distribution of the statistical complexity

We denote the normalized Shannon Entropy asymptotic mean by

$$\mu_{\boldsymbol{p}} = -\frac{1}{\ln k}\sum_{\ell=1}^{k} p_\ell \ln p_\ell = H[\boldsymbol{p}], \tag{20}$$

and the Jensen–Shannon divergence asymptotic mean by

$$\mu_{\mathscr{D}\boldsymbol{p}} = \frac{1}{2}\sum_{\ell=1}^{k}\left[p_\ell \ln p_\ell - (p_\ell + 1/k)\ln\frac{p_\ell + 1/k}{2}\right] - \frac{\ln k}{2}. \tag{21}$$

For n sufficiently large, by (11) and (18), we can assume that $H[\widehat{\boldsymbol{p}}] \sim \mathcal{N}(\mu_{\boldsymbol{p}}, {\sigma_{\boldsymbol{p}}}^2/n)$ and $\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}] \sim \mathcal{N}(\mu_{\mathscr{D}\boldsymbol{p}}, \sigma^2_{\mathscr{D}\boldsymbol{p}}/n)$.

The correlation coefficient between the Jensen–Shannon divergence and the normalized Shannon Entropy is defined by

$$\rho = \frac{\operatorname{cov}\big(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}], H[\widehat{\boldsymbol{p}}]\big)}{\sqrt{\operatorname{Var}(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}])\operatorname{Var}\big(H[\widehat{\boldsymbol{p}}]\big)}} = \frac{\mathrm{E}\big(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]H[\widehat{\boldsymbol{p}}]\big) - \mu_{\mathscr{D}\boldsymbol{p}}\mu_{\boldsymbol{p}}}{\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}/n}. \tag{22}$$

Notice that this result coincides with the one obtained by Corollary 1 from Appendix. To compute the expected value $\mathrm{E}(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]H[\widehat{\boldsymbol{p}}])$, we use the results presented in Appendix. First, notice that

$$\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}] \pm H[\widehat{\boldsymbol{p}}] = H_S\left[\frac{\widehat{\boldsymbol{p}} + \boldsymbol{u}}{2}\right] - \frac{1}{2}H_S[\widehat{\boldsymbol{p}}] - \frac{1}{2}\ln k \pm \frac{1}{\ln k}H_S[\widehat{\boldsymbol{p}}]. \tag{23}$$

Thus,

$$\begin{aligned}\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}] \pm H[\widehat{\boldsymbol{p}}] &= -\sum_{\ell=1}^{k}\frac{\widehat{p}_\ell + u_\ell}{2}\ln\frac{\widehat{p}_\ell + u_\ell}{2} + \frac{1}{2}\sum_{\ell=1}^{k}\widehat{p}_\ell \ln \widehat{p}_\ell - \frac{1}{2}\ln k \\ &\quad \mp \frac{1}{\ln k}\sum_{\ell=1}^{k}\widehat{p}_\ell \ln \widehat{p}_\ell \\ &= \frac{1}{2}\left(\frac{\ln k \mp 2}{\ln k}\sum_{\ell=1}^{k}\widehat{p}_\ell \ln \widehat{p}_\ell - \sum_{\ell=1}^{k}(\widehat{p}_\ell + u_\ell)\ln\frac{\widehat{p}_\ell + u_\ell}{2} - \ln k\right).\end{aligned} \tag{24}$$

We consider the following functions:

$$g_\ell^{\pm}(p_1, p_2, \dots, p_k) = \frac{\ln k \mp 2}{\ln k}p_\ell \ln p_\ell - (p_\ell + 1/k)\ln\frac{p_\ell + 1/k}{2}, \tag{25}$$

for $\ell = 1, 2, \dots, k$. The partial derivatives are

$$\begin{aligned}\frac{\partial g_\ell^{\pm}}{\partial p_\ell} &= \frac{\ln k \mp 2}{\ln k}(1 + \ln p_\ell) - 1 - \ln\frac{p_\ell + 1/k}{2}, \text{ and} \\ \frac{\partial g_j^{\pm}}{\partial p_\ell} &= 0 \text{ if } j \neq \ell.\end{aligned} \tag{26}$$

Thus, the covariance matrix of the multivariate Normal limit distribution is

$$\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\pm} = \left(\frac{\partial g_\ell^{\pm}}{\partial p_j}\right)\boldsymbol{\Sigma}_{\boldsymbol{p}}\left(\frac{\partial g_\ell^{\pm}}{\partial p_j}\right)', \tag{27}$$

where

$$\begin{aligned}(\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\pm})_{\ell\ell} &= (p_\ell - p_\ell^2)\left[\frac{\ln k \mp 2}{\ln k}(1 + \ln p_\ell) - 1 - \ln\frac{p_\ell + 1/k}{2}\right]^2, \text{ and} \\ (\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\pm})_{\ell j} &= -\,p_\ell p_j\left[\frac{\ln k \mp 2}{\ln k}(1 + \ln p_\ell) - 1 - \ln\frac{p_\ell + 1/k}{2}\right] \\ &\qquad \left[\frac{\ln k \mp 2}{\ln k}(1 + \ln p_j) - 1 - \ln\frac{p_j + 1/k}{2}\right] \text{ if } \ell \neq j.\end{aligned} \tag{28}$$

By the multivariate Delta Method applied to the sequence $\widehat{\boldsymbol{p}}_n$, we conclude that

$$\sqrt{n}\big[g_1^{\pm}(\widehat{p}_1) - g_1^{\pm}(p_1), g_2^{\pm}(\widehat{p}_2) - g_2^{\pm}(p_2), \dots, g_k^{\pm}(\widehat{p}_k) - g_k^{\pm}(p_k)\big] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \Sigma_{\boldsymbol{p}}^{\pm}\right). \tag{29}$$

Hence, applying (10) with $\boldsymbol{a} = (1/2, 1/2, \dots, 1/2)$, it holds that

$$\sqrt{n}\big(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}] \pm H[\widehat{\boldsymbol{p}}] - \mathscr{D}[\boldsymbol{p}, \boldsymbol{u}] \mp H[\boldsymbol{p}]\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma^2_{\boldsymbol{p}^{\pm}}\right), \tag{30}$$

where

$$\begin{aligned}\sigma^2_{\boldsymbol{p}^{\pm}} \;=\; &\frac{1}{4}\sum_{\ell=1}^{k}(p_\ell - p_\ell^2)\left[\frac{\ln k \mp 2}{\ln k}(1 + \ln p_\ell) - 1 - \ln\frac{p_\ell + 1/k}{2}\right]^2 - \\ &\frac{1}{2}\sum_{j=1}^{k-1}\sum_{\ell=j+1}^{k}p_\ell p_j\left[\frac{\ln k \mp 2}{\ln k}(1 + \ln p_\ell) - 1 - \ln\frac{p_\ell + 1/k}{2}\right] \\ &\left[\frac{\ln k \mp 2}{\ln k}(1 + \ln p_j) - 1 - \ln\frac{p_j + 1/k}{2}\right].\end{aligned} \tag{31}$$

Let

$$\mu_{\boldsymbol{p}^{\pm}} = \frac{1}{2}\sum_{\ell=1}^{k}\left[\frac{\ln k \mp 2}{\ln k}p_\ell \ln p_\ell - (p_\ell + 1/k)\ln\frac{p_\ell + 1/k}{2}\right] - \frac{\ln k}{2}. \tag{32}$$

By Lemma 1, Appendix, with $X = \mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]$ and $Y = H[\widehat{\boldsymbol{p}}]$, we have that

$$\mathrm{E}(\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]H[\widehat{\boldsymbol{p}}]) = \frac{1}{4}(\sigma^2_{\boldsymbol{p}^+} + \mu^2_{\boldsymbol{p}^+} - \sigma^2_{\boldsymbol{p}^-} - \mu^2_{\boldsymbol{p}^-}). \tag{33}$$

Replacing in (22), the correlation coefficient between $\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]$ and $H[\widehat{\boldsymbol{p}}]$ is

$$\rho = \frac{\sigma^2_{\boldsymbol{p}^+} + n\mu^2_{\boldsymbol{p}^+} - \sigma^2_{\boldsymbol{p}^-} - n\mu^2_{\boldsymbol{p}^-} - 4\mu_{\mathscr{D}\boldsymbol{p}}\mu_{\boldsymbol{p}}}{4\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}}. \tag{34}$$

Let $K_\nu(\cdot)$ denote the modified Bessel function of the second kind and order ν, defined in Ref. [24, formula 8.485] as:

$$K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\nu x)}, \tag{35}$$

with

$$I_\nu(x) = \sum_{a=0}^{+\infty}\frac{1}{a!\Gamma(a + \nu + 1)}\left(\frac{x}{2}\right)^{2a+\nu}, \tag{36}$$

as defined in Ref. [24, formula 8.445]. Applying [25, Theorem 2.1], the distribution of the product $W[\widehat{\boldsymbol{p}}] = \mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]H[\widehat{\boldsymbol{p}}]$ with support $w \geq 0$, can be expressed as:

$$f_W(w) = \exp[R(w)]\sum_{a=0}^{+\infty}\sum_{b=0}^{2a}A_{a,b}(w)K_{b-a}[S(w)], \tag{37}$$

where

$$R(w) = -\frac{n}{2(1 - \rho^2)}\left(\frac{\mu^2_{\mathscr{D}\boldsymbol{p}}}{\sigma^2_{\mathscr{D}\boldsymbol{p}}} + \frac{{\mu_{\boldsymbol{p}}}^2}{{\sigma_{\boldsymbol{p}}}^2} - \frac{2\rho(w + \mu_{\mathscr{D}\boldsymbol{p}}\mu_{\boldsymbol{p}})}{\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}}\right), \tag{38}$$

$$S(w) = \frac{nw}{(1 - \rho^2)\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}}, \tag{39}$$

$$\begin{aligned}A_{a,b}(w) = &\frac{n^{2a+1}w^a\sigma_{\mathscr{D}\boldsymbol{p}}^{b-a-1}}{\pi(2a)!(1 - \rho^2)^{2a+1/2}\sigma_{\boldsymbol{p}}^{b-a+1}}\binom{2a}{b}\left(\frac{\mu_{\mathscr{D}\boldsymbol{p}}}{\sigma^2_{\mathscr{D}\boldsymbol{p}}} - \frac{\rho\mu_{\boldsymbol{p}}}{\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}}\right)^b \\ &\times\left(\frac{\mu_{\boldsymbol{p}}}{{\sigma_{\boldsymbol{p}}}^2} - \frac{\rho\mu_{\mathscr{D}\boldsymbol{p}}}{\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}}\right)^{2a-b}.\end{aligned} \tag{40}$$

We finally conclude that, for n sufficiently large, the asymptotic distribution of the Statistical Complexity $C[\widehat{\boldsymbol{p}}] = Q[\widehat{\boldsymbol{p}}]H[\widehat{\boldsymbol{p}}] = Q_0\mathscr{D}[\widehat{\boldsymbol{p}}, \boldsymbol{u}]H[\widehat{\boldsymbol{p}}]$ is

$$F_C(x) = F_W\left(Q_0^{-1}x\right). \tag{41}$$

In other words, the probability density function of the Statistical Complexity for $x > 0$ is

$$\begin{aligned}f_C(x) \;=\; &Q_0^{-1}f_W\left(Q_0^{-1}x\right) = \\ &Q_0^{-1}\left\{\exp\left[R(Q_0^{-1}x)\right]\sum_{a=0}^{+\infty}\sum_{b=0}^{2a}A_{a,b}(Q_0^{-1}x)K_{b-a}\left[S(Q_0^{-1}x)\right]\right\}.\end{aligned} \tag{42}$$

Let $\delta_Q = n\mu_{\mathscr{D}\boldsymbol{p}}/\sigma_{\mathscr{D}\boldsymbol{p}}$ and $\delta_H = n\mu_{\boldsymbol{p}}/\sigma_{\boldsymbol{p}}$. For n sufficiently large, δ_Q and δ_H tend to infinity, and the series in (42) converges very slowly [26]. With this in mind, and using Ref. [27, Theorem 2.5], we can approximate the distribution of $C[\widehat{\boldsymbol{p}}]$ by a Normal law with mean μ_C and variance σ_C^2 given by

$$\mu_C = Q_0\sigma_{\mathscr{D}}\sigma_{\boldsymbol{p}}/n\big(\delta_Q\delta_H + \rho\big), \text{ and} \tag{43}$$

$$\sigma_C^2 = Q_0^2\sigma_{\mathscr{D}}^2\sigma_{\boldsymbol{p}}^2/n^2\big(\delta_Q^2 + \delta_H^2 + 2\rho\delta_Q\delta_H + 1 + \rho^2\big). \tag{44}$$

These expressions are available as the functions `meanC` and `varC` in the `StatOrdPattHxC` package available at https://github.com/arey1911/StatOrdPattHxC/. These functions depend on the probability

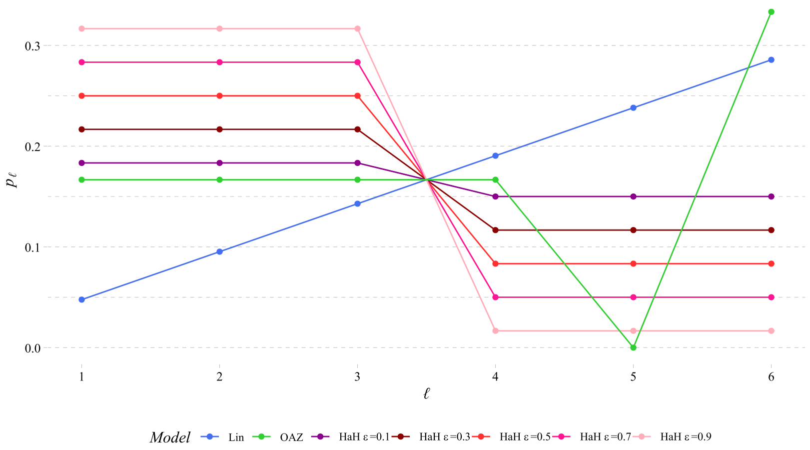

Fig. 2. Probability functions $\boldsymbol{p}$ for different models.

function and on the number of trials.

5.4. Generating moment function of the statistical complexity

In this section we compute the limit generating moment function (GMF) of the Statistical Complexity as an alternative characterization of its distribution.

The asymptotic distributions given by (11) and (18) imply that, for n sufficiently large, $H[\hat{\boldsymbol{p}}] \sim \mathcal{N}(\mu_{\boldsymbol{p}}, {\sigma_{\boldsymbol{p}}}^2/n)$ and $Q[\hat{\boldsymbol{p}}] \sim \mathcal{N}(Q_0\mu_{\mathscr{D}\boldsymbol{p}}, Q_0^2\sigma^2_{\mathscr{D}\boldsymbol{p}}/n)$. Then, since the Statistical Complexity $C[\hat{\boldsymbol{p}}]$ is defined as the product of these two random variables that are asymptotically Gaussian, we can apply Theorem 1, Appendix, to compute its GMF, obtainig the following expression:

$$M_C(t) = \frac{\exp\left\{\dfrac{(\delta_Q^2+\delta_H^2-2\rho\delta_Q\delta_H)t^2Q_0^2\sigma^2_{\mathscr{D}\boldsymbol{p}}\sigma^2_{\boldsymbol{p}}/\ln^2 k + 2tQ_0\mu_{\mathscr{D}\boldsymbol{p}}\mu_{\boldsymbol{p}}/\ln k}{2[1-2t\rho Q_0\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}/\ln k-(1-\rho^2)t^2Q_0^2\sigma^2_{\mathscr{D}\boldsymbol{p}}\sigma^2_{\boldsymbol{p}}/\ln^2 k]}\right\}}{\sqrt{1-2t\rho Q_0\sigma_{\mathscr{D}\boldsymbol{p}}\sigma_{\boldsymbol{p}}/\ln k-(1-\rho^2)t^2Q_0^2\sigma^2_{\mathscr{D}\boldsymbol{p}}\sigma^2_{\boldsymbol{p}}/\ln^2 k}}, \tag{45}$$

where $\delta_Q = n\mu_{\mathscr{D}\boldsymbol{p}}/\sigma_{\mathscr{D}\boldsymbol{p}}$, $\delta_H = n\mu_{\boldsymbol{p}}/\sigma_{\boldsymbol{p}}$ and ρ is the correlation coefficient between the Disequilibrium and the normalized Shannon Entropy, given in (34).

6. Numerical validation

In this section, we conduct experiments to validate the theoretical results proved in Section 5.

6.1. Independent patterns

Since the asymptotic results fail in the case of white noise [28], we consider the following cases described by the vector of probability $\boldsymbol{p} = (p_1, p_2, \dots, p_k)$ from the Multinomial model:

1. Linear (Lin): $p_\ell = 2\ell/[k(k+1)]$, for $\ell = 1, 2, \dots, k$.
2. One-Almost-Zero (OAZ): $p_\ell = 1/k$ for $\ell = 1, 2, \dots, k-2$, $p_{k-1} = \varepsilon_0$, and $p_k = 2/k - \varepsilon_0$ with $\varepsilon_0 = 2.220446 \times 10^{-16}$, machine epsilon in R [29], the smallest positive number that, when added to 1, yields a result different from 1.
3. Half-and-Half (HaH): $p_\ell = 1/k + \varepsilon/k$ for $\ell = 1, 2, \dots, k/2$, and $p_\ell = 1/k - \varepsilon/k$ for $\ell = k/2+1, k/2+2, \dots, k$, with $\varepsilon \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Fig. 2 illustrates these probability function with $k = 6$ and $\varepsilon = 0.1, 0.3, 0.5, 0.7, 0.9$.

The parameter values used in our numerical study are $k \in \{6, 12, 18, 24, 30\}$ and $n \in \{10^2k, 10^3k, 10^4k, 10^5k, 10^6k\}$. For each $\boldsymbol{p}$ of each model (Lin, OAZ, and HaH with $\varepsilon \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$), we generated $R = 500$ vector samples of $\boldsymbol{N}$ that record the frequencies of n independent states using the $\mathrm{Mult}(n, \boldsymbol{p})$ distribution. Then, for each vector sample $\underset{\sim}{\hat{\boldsymbol{p}}}_r$ ($r = 1, 2, \dots, R$; used $R = 500$), we computed the Statistical Complexity $c_r = C[\underset{\sim}{\hat{\boldsymbol{p}}}_r]$. For comparison, we generated a sample of size R under the Normal distribution $\mathcal{N}(\mu_C, \sigma_C^2)$ that approximates the asymptotic distribution of $c_1, c_2, \dots, c_R$. Then, we apply the Kolmogorov–Smirnov (KS) test [30] to examine if these two samples were drawn from the same distribution. Due to the randomness of the generated samples, we replicated this procedure $S = 1000$ times. This procedure is illustrated in Algorithm 1. The proposed model fits well in almost all the cases under study. Table 1 shows the twenty-two out of one hundred and seventy-five cases where the percentage of p-values less than 0.01 is greater than 30%. However, we do not recommend approximating the Statistical Complexity distribution by the $\mathcal{N}(\mu_C, \sigma_C^2)$ distribution if $n = 10^2k$. The approximation is not good with a relatively small number of trials, and improves as this value increases.

6.2. Bandt–Pompe symbolization

Let $\boldsymbol{x} = \{x_1, x_2, \dots, x_T\}$ be a time series. To apply the methodology proposed by Bandt and Pompe [19], we consider an embedding dimension $D > 1$ (such that $T \gg D!$) and an embedding time delay $\tau = 1$. For simplicity in the notation, let $k = D!$. Thus, the ordinal pattern symbols are given by the k permutations of the set $\{0, 1, \dots, D-1\}$, labeled by $\boldsymbol{\pi}^1, \boldsymbol{\pi}^2, \dots, \boldsymbol{\pi}^k$. A D-length partition of $\boldsymbol{x}$ is a sequence $x_t, x_{t+1}, \dots, x_{t+D-1}$, with $t \in \{1, 2, \dots, T-D+1\}$, is of type $\boldsymbol{\pi}^j$ ($j = 1, 2, \dots, k$) if $x_{t+\boldsymbol{\pi}^j(0)} < x_{t+\boldsymbol{\pi}^j(1)} < \cdots < x_{t+\boldsymbol{\pi}^j(D-1)}$.

We treat the case of equal values as follows. Suppose that there are indexes $t_0 \in \{1, 2, \dots, T-D+1\}$, $s_0 \in \{0, 1, \dots, D-2\}$, and $j_0 \in \{1, 2, \dots, k\}$ such that

$$x_{t+\boldsymbol{\pi}^{j_0}(0)} < x_{t+\boldsymbol{\pi}^{j_0}(1)} < \cdots < x_{t+\boldsymbol{\pi}^{j_0}(s_0)} = x_{t+\boldsymbol{\pi}^{j_0}(s_0+1)} < \cdots < x_{t+\boldsymbol{\pi}^{j_0}(D-1)}. \tag{46}$$

Define the permutation $\widetilde{\boldsymbol{\pi}}^{j_0}$ as $\widetilde{\boldsymbol{\pi}}^{j_0}(s_0) = \boldsymbol{\pi}^{j_0}(s_0+1)$, $\widetilde{\boldsymbol{\pi}}^{j_0}(s_0+1) = \boldsymbol{\pi}^{j_0}(s_0)$, and $\widetilde{\boldsymbol{\pi}}^{j_0}(s) = \boldsymbol{\pi}^{j_0}(s)$ for all $s \in \{0, 1, \dots, D-2\}$ and $s \neq s_0, s_0+1$. Thus, expression (46) can be rewritten as

$$x_{t+\widetilde{\boldsymbol{\pi}}^{j_0}(0)} < x_{t+\widetilde{\boldsymbol{\pi}}^{j_0}(1)} < \cdots < x_{t+\widetilde{\boldsymbol{\pi}}^{j_0}(s_0)} = x_{t+\widetilde{\boldsymbol{\pi}}^{j_0}(s_0+1)} < \cdots < x_{t+\widetilde{\boldsymbol{\pi}}^{j_0}(D-1)}. \tag{47}$$

Algorithm 1: Assessment of the empirical distribution of the Statistical Complexity under the Multinomial model

```
Input: Number of possible patterns $K = \{6, 12, 18, 24, 30\}$
Input: Models $\mathcal{M} = \{\mathrm{Lin}, \mathrm{OAZ}, \mathrm{HaH}(\varepsilon)\}$, with
       $\varepsilon \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$
Input: Number of experiments $S = 1000$
Input: Number of replications $R = 500$
Input: Threshold $\eta = 0.01$
Output: Proportion of $p$-values smaller than $\eta$ for each number of
        possible patterns and each model
for each number of possible patterns $k \in K$ do
    for each model $M \in \mathcal{M}$ do
        Compute the asymptotic mean $\mu_C$ and variance $\sigma_C^2$ under $k$
          and $M$
        for each experiment $1 \le s \le S$ do
            for each number of trials
              $n \in \{10^2 k, 10^3 k, 10^4 k, 10^5 k, 10^6 k\}$ do
                for each replication $1 \le r \le R$ do
                    Produce the observation $\underset{\sim}{\boldsymbol{n}}_r$ from the
                      Multinomial model Mult$(n, \boldsymbol{p})$
                    Compute the sample probability $\underset{\sim}{\widehat{\boldsymbol{p}}}_r = \underset{\sim}{\boldsymbol{n}}_r / n$
                    Compute the sample statistical Complexity
                      $c_r = C[\underset{\sim}{\widehat{\boldsymbol{p}}}_r]$
                    Produce a sample $z_r$ from the Normal
                      distribution with mean $\mu_C$ and variance $\sigma_C^2$
            Compute $p$-val$_r$ the $p$-value of the Kolmogorov-Smirnov
              test that compares $(c_1, c_2, \dots, c_R)$ and $(z_1, z_2, \dots, z_R)$
        Count the number of $p$-values in $(p\text{-val}_1, p\text{-val}_2, \dots, p\text{-val}_S)$
          that are smaller than $\eta$
```

Table 1
Percentage (greater than 30%) the p-value of the KS test between the sample complexities $c_1, c_2, \dots, c_R$ and the sample $z_1, z_2, \dots, z_R$ from the $\mathcal{N}(\mu_C, \sigma_C^2)$ distribution is smaller than 0.01.

Model	k	n	Percentage of p-values < 0.01
Lin	6	6×10^2	41.7%
	12	12×10^2	42.8%
OAZ	6	6×10^2	52.1%
	12	12×10^3	41.9%
	18	18×10^3	38.1%
	30	30×10^4	31.9%
HaH1	6	6×10^3	48.1%
	18	18×10^4	36.6%
	24	24×10^4	49.8%
	30	30×10^4	51.2%
HaH3	6	6×10^2	44.7%
	18	18×10^3	44.6%
	24	24×10^3	52.6%
	30	30×10^3	47.4%
HaH5	6	6×10^2	43.3%
	12	12×10^2	34.7%
HaH7	6	6×10^2	32.3%
	12	12×10^2	51.9%
	18	18×10^2	37.1%
HaH9	6	6×10^2	38.5%
	12	12×10^2	52.4%
	18	18×10^2	41.0%

The sequence $x_{t_0}, x_{t_0+1}, \dots, x_{t_0+D-1}$ is of type $\boldsymbol{\pi}_{j_0}$ if $\boldsymbol{\pi}_{j_0}(t_0+s_0) < \widetilde{\boldsymbol{\pi}}_{j_0}(t_0+s_0)$, and of type $\widetilde{\boldsymbol{\pi}}_{j_0}$, otherwise.

To illustrate these definitions, consider the realization $\boldsymbol{x} = \{7, 8, 6, 9, 9, 7, 5, 1\}$ of length $T = 8$. If $D = 3$ ($k = 6$), the permutations of the elements $\{0, 1, 2\}$ can be labeled as $\boldsymbol{\pi}^1 = (0, 1, 2)$, $\boldsymbol{\pi}^2 = (0, 2, 1)$, $\boldsymbol{\pi}^3 = (1, 0, 2)$, $\boldsymbol{\pi}^4 = (1, 2, 0)$, $\boldsymbol{\pi}^5 = (2, 0, 1)$, and $\boldsymbol{\pi}^6 = (2, 1, 0)$. Hence, there are six 3-length partitions of $\boldsymbol{x}$ characterized as shown in Table 2.

Table 2
Type of permutation for the 3-length partitions of the time series $\boldsymbol{x} = \{7, 8, 6, 9, 9, 7, 5, 1\}$.

t	Partition	Ordering	Type of permutation
1	$x_1 = 7, x_2 = 8, x_3 = 6$	$x_3 < x_1 < x_2 \rightarrow x_{t+2} < x_{t+0} < x_{t+1}$	$\boldsymbol{\pi}^5 = (2, 0, 1)$
2	$x_2 = 8, x_3 = 6, x_4 = 9$	$x_3 < x_2 < x_4 \rightarrow x_{t+1} < x_{t+0} < x_{t+2}$	$\boldsymbol{\pi}^3 = (1, 0, 2)$
3	$x_3 = 6, x_4 = 9, x_5 = 9$	$x_3 < x_4 = x_5 \rightarrow x_{t+0} < x_{t+1} = x_{t+2}$	$\boldsymbol{\pi}^1 = (0, 1, 2)$
4	$x_4 = 9, x_5 = 9, x_6 = 7$	$x_6 < x_4 = x_5 \rightarrow x_{t+2} < x_{t+0} = x_{t+1}$	$\boldsymbol{\pi}^5 = (2, 0, 1)$
5	$x_5 = 9, x_6 = 7, x_7 = 5$	$x_7 < x_6 < x_5 \rightarrow x_{t+2} < x_{t+1} < x_{t+0}$	$\boldsymbol{\pi}^6 = (2, 1, 0)$
6	$x_6 = 7, x_7 = 5, x_8 = 1$	$x_8 < x_7 < x_6 \rightarrow x_{t+2} < x_{t+1} < x_{t+0}$	$\boldsymbol{\pi}^6 = (2, 1, 0)$

Table 3
Empirical ($\widehat{\mu}$ and $\widehat{\sigma}^2$) and theoretical results of applying the model $\mathcal{N}(\mu_C, \sigma_C^2)$ to the processes RW and TS ± 1.

Process	$\widehat{\mu}$	μ_C	$\widehat{\sigma}^2$	σ_C^2	p-value KS test
RW	0.0306966	0.0305958	4.644×10^{-6}	4.996×10^{-6}	0.2093
TS±1	0.1829974	0.1829423	3.529×10^{-6}	3.603×10^{-6}	0.6125

We can obtain a sequence of ordinal patterns $\pi_1, \pi_2, \dots, \pi_{T-D+1}$ from $\boldsymbol{x}$. Notice that, due to the overlapping of $D-1$ consecutive partitions, the sequence of ordinal patterns is correlated. However, we assume independence as a simplification and apply our proposal using the Multinomial model, where the probability function $\boldsymbol{p}$ is defined as follows. For $j = 1, 2, \dots, k$, the probability of occurrence of the symbol $\boldsymbol{\pi}^j$ is given by

$$p_j = \Pr(\boldsymbol{\pi}^j) = \frac{\#\{D\text{-length partitions of } \boldsymbol{x} \text{ of type } \boldsymbol{\pi}^j\}}{T - D + 1}. \tag{48}$$

We use two well-known processes to validate the proposed methodology. First, we recall that a random walk (RW) is recursively generated by the formula

$$x_t = x_{t-1} + \xi_t, \tag{49}$$

where ξ_t is a component standard Normal distributed. The expression for the theoretical vector of ordinal pattern probability [31, Eq. (41)] is

$$\boldsymbol{p}(\mathrm{RW}) = \left(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{4}\right). \tag{50}$$

We also consider a time series of equiprobable independent observations in the set $\{-1, 1\}$, denoted by "TS±1". The expression for the theoretical vector of ordinal pattern probability [32, Example 1] is

$$\boldsymbol{p}(\mathrm{TS}{\pm}1) = \left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, 0\right). \tag{51}$$

We conducted a Monte Carlo numerical experiment by generating $R = 10^3$ time series of length $N = D!10^3$ for each one of these two processes using the embedding dimension $D = 3$. Then, for each time series, we computed the Statistical Complexity, C. Table 3 shows the empirical and theoretical asymptotic means and variances related to each process, as well as the p-values of the KS test used to see if the sample of the Statistical Complexity was drawn from the $\mathcal{N}(\mu_C, \sigma_C^2)$ distribution. Besides, the corresponding empirical densities and histograms using the Freedman–Diaconis rule [33], together with the asymptotic distribution, are presented in Fig. 3. These results suggest that the proposed asymptotic model is a good approximation for these processes.

7. Conclusions

We have obtained an approximate distribution for the Statistical Complexity, as defined by López-Ruiz et al. [8]. This asymptotic result relies on the limit distribution of the normalized Shannon Entropy, the limit distribution of the normalized Jensen–Shannon distance to

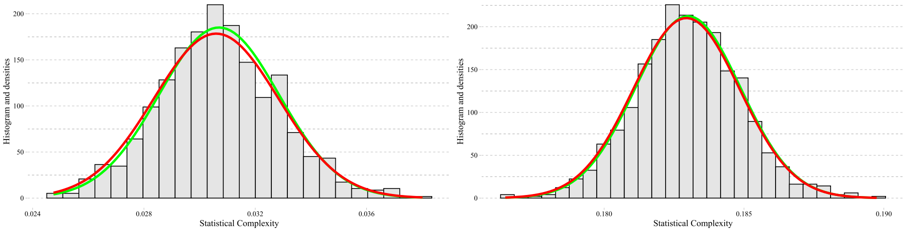

Fig. 3. Histograms, empirical density (in green), and asymptotic model (in red) of the processes RW (left) and TS±1 (right).

the uniform model, the distribution of the product of two correlated normally-distributed random variables, and an approximation for the terms that, for a large number of observations, produce numerical instabilities.

We provided R functions that implement the asymptotic mean and variance of the Statistical Complexity, and verified that the approximate model is a good description in a variety of situations, including independent observations and symbols obtained with the Bandt & Pompe approach.

This asymptotic result can be used to devise tests statistics as in Ref. [15]. Such tests will enhance otherwise qualitative assessments such as in Ref. [34].

CRediT authorship contribution statement

Andrea Rey: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alejandro C. Frery:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Juliana Gambini:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Product of two Gaussian random variables

In this Appendix, we present two results concerning probabilistic properties of the product of two random variables following a Normal distribution.

Lemma 1. *Let X and Y be two random variables. Consider the variables given by their sum $S = X + Y$ and their difference $Z = X - Y$. If $S \sim \mathcal{N}(\mu_S, \sigma_S^2)$ and $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$, then*

$$\mathrm{E}(XY) = \frac{1}{4}\left\{(\sigma_S^2 + \mu_S^2) - (\sigma_Z^2 + \mu_Z^2)\right\}.$$

Proof. A direct computation shows that

$$XY = \frac{1}{4}(X+Y)^2 - \frac{1}{4}(X-Y)^2 = \frac{1}{4}(S^2 - Z^2). \tag{52}$$

Applying the expected value and using well-know properties, it holds that

$$\begin{aligned}\mathrm{E}(XY) &= \frac{1}{4}\left\{\mathrm{E}\big(S^2\big) - \mathrm{E}\big(Z^2\big)\right\} = \frac{1}{4}\left\{\mathrm{Var}(S) + [\mathrm{E}(S)]^2 - \mathrm{Var}(Z) - [\mathrm{E}(Z)]^2\right\} \\ &= \frac{1}{4}\left\{(\sigma_S^2 + \mu_S^2) - (\sigma_Z^2 + \mu_Z^2)\right\}. \quad \square\end{aligned} \tag{53}$$

Theorem 1. *Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be two Normal random variables with correlation coefficient ρ. Denote $\delta_X = \mu_X/\sigma_X$ and $\delta_Y = \mu_Y/\sigma_Y$. The generating moment function of the random variable $Z = XY/(\sigma_X\sigma_Y)$ is:*

$$M_Z(t) = \mathrm{E}(e^{tZ}) = \frac{1}{\sqrt{b(t)}}\exp\frac{a(t)}{2b(t)}, \tag{54}$$

where

$$a(t) = (\delta_X^2 + \delta_Y^2 - 2\rho\delta_X\delta_Y)t^2 + 2\delta_X\delta_Y t, \text{ and} \tag{55}$$

$$b(t) = [1 - (1+\rho)t][1 + (1-\rho)t] = 1 - 2\rho t - (1-\rho^2)t^2. \tag{56}$$

Proof. We refer to [26], where this result was proved. □

Corollary 1. *Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be two Normal random variables. Then,*

$$E(XY) = \mu_X\mu_Y + \rho\sigma_X\sigma_Y, \tag{57}$$

where ρ is the correlation coefficient.

Proof. Denote $c(t) = a(t)/[2b(t)]$, where $a(t)$ and $b(t)$ are defined by Eqs. (55) and (56), respectively. Notice that $a(0) = 0$, $b(0) = 1$, and $c(0) = 1$. On the other hand, the derivative of Eq. (54) is

$$M_Z'(t) = \frac{\exp[c(t)]c'(t)\sqrt{b(t)} - \exp[c(t)]b'(t)/[2\sqrt{b(t)}]}{b(t)}. \tag{58}$$

Direct computations show that

$$a'(t) = 2(\delta_X^2 + \delta_Y^2 - 2\rho\delta_X\delta_Y)t + 2\delta_X\delta_Y \implies a'(0) = 2\delta_X\delta_Y \tag{59}$$

$$b'(t) = -2\rho - 2(1-\rho^2)t \implies b'(0) = -2\rho, \text{ and} \tag{60}$$

$$\begin{aligned}c'(t) = \frac{a'(t)b(t) - a(t)b'(t)}{2[b(t)^2]} \implies c'(0) &= \frac{a'(0)b(0) - a(0)b'(0)}{2[b(0)^2]} \\ &= \delta_X\delta_Y.\end{aligned} \tag{61}$$

Since

$$\begin{aligned}\mathrm{E}(Z) = M_Z'(0) &= \frac{\exp[c(0)]c'(0)\sqrt{b(0)} - \exp[c(0)]b'(0)/[2\sqrt{b(0)}]}{b(0)} \\ &= \delta_X\delta_Y + \rho,\end{aligned} \tag{62}$$

we can conclude that

$$M_{XY}(t) = \mathrm{E}(e^{tXY}) = \mathrm{E}(e^{t\sigma_X\sigma_Y Z}) = M_Z(t\sigma_X\sigma_Y). \tag{63}$$

Therefore, since $M_{XY}'(t) = \sigma_X\sigma_Y M_Z'(t\sigma_X\sigma_Y)$,

$$E(XY) = M_{XY}'(0) = \sigma_X\sigma_Y(\delta_X\delta_Y + \rho) = \mu_X\mu_Y + \rho\sigma_X\sigma_Y. \quad \square \tag{64}$$

Data availability

No data was used for the research described in the article.

References

[1] Clausius R. Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen. Ann Phys, Lpz 1850;79(3, 4):368–97, 500–524.

[2] Sharp K, Matschinsky F. Translation of Ludwig Boltzmann's paper on the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium sitzungberichte der kaiserlichen akademie der wissenschaften. Mathematisch-Naturwissen classe. Abt. II LXXVI 1877, pp 373–435 (Wien. Ber. 1877 76:373–435). Reprinted in Wiss. Abhandlungen, vol. ii, reprint 42, pp. 164–223, Barth, Leipzig, 1909. Entropy 2015;17(4):1971–2009.

[3] Shannon CE. A mathematical theory of communications. Bell Syst Technol J 1948;27:379–423.

[4] Huberman BA, Hogg T. Complexity and adaptation. Phys D 1986;22(1–3):376–84.

[5] Grassberger P. Toward a quantitative theory of self-generated complexity. Internat J Theoret Phys 1986;25(9):907–38.

[6] Crutchfield JP, Young K. Inferring statistical complexity. Phys Rev Lett 1989;63(2):105–8.

[7] Crutchfield JP. The calculi of emergence: computation, dynamics and induction. Phys D 1994;75(1):11–54.

[8] López-Ruiz R, Mancini HL, Calbet X. A statistical measure of complexity. Phys Lett A 1995;209(5–6):321–6.

[9] Feldman D, Crutchfield J. Measures of statistical complexity: Why? Phys Lett A 1998;238(4):244–52.

[10] Lamberti PW, Martín MT, Plastino A, Rosso OA. Intensive entropic non-triviality measure. Phys A 2004;334(1–2):119–31.

[11] Zanin M, Zunino L, Rosso OA, Papo D. Permutation entropy and its main biomedical and econophysics applications: A review. Entropy 2012;14(8):1553–77.

[12] Amigó JM, Rosso OA. Ordinal methods: Concepts, applications, new developments, and challenges—in memory of Karsten Keller (1961–2022). Chaos 33(8):2023.

[13] Chagas ETC, Queiroz-Oliveira M, Rosso OA, Ramos HS, Freitas CGS, Frery AC. White noise test from ordinal patterns in the entropy–complexity plane. Int Stat Rev 2022;90(2):374–96.

[14] Rey AA, Frery AC, Lucini M, Gambini J, Chagas ETC, Ramos HS. Asymptotic distribution of certain types of entropy under the Multinomial law. Entropy 2023;25(5):734.

[15] Rey AA, Frery AC, Gambini J, Lucini M. Asymptotic distribution of entropies and Fisher information measure of ordinal patterns with applications. Chaos Solitons Fractals 2024;188:115481.

[16] Parsons D. The directory of tunes and musical themes. Cambridge; 1975.

[17] Agresti A. An introduction to categorical data analysis. Hoboken, NJ: Wiley-Interscience; 2007.

[18] Agresti A. Analysis of ordinal categorical data. Wiley series in probability and statistics, 2nd ed.. Hoboken, N.J: Wiley; 2013.

[19] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series. Phys Rev Lett 2002;88(17):174102.

[20] Lin J. Divergence measures based on the Shannon entropy. IEEE Trans Inform Theory 1991;37(1):141–51.

[21] Martin MT, Plastino A, Rosso OA. Generalized statistical complexity measures: Geometrical and analytical properties. Phys A 2006;369(2):439–62.

[22] Lehmann EL, Casella G. Theory of point estimation. Springer Science & Business Media; 2006.

[23] Härdle WK, Okhrin O, Okhrin Y. Multivariate statistical analysis. In: Basic elements of computational statistics. 2017, p. 219–41.

[24] Gradshteyn IS, Ryzhik IM. Table of integrals, series, and products. 1996, CD version 1.0.

[25] Cui G, Yu X, Iommelli S, Kong L. Exact distribution for the product of two correlated Gaussian random variables. IEEE Signal Process Lett 2016;23(11):1662–6.

[26] Craig CC. On the frequency function of xy. Ann Math Stat 1936;7(1):1–15.

[27] Aroian LA. The probability function of the product of two normally distributed variables. Ann Math Stat 1947;265–71.

[28] Chagas ETC, Frery AC, Gambini J, Lucini MM, Ramos HS, Rey AA. Statistical properties of the entropy from ordinal patterns. Chaos 2022;32:113118.

[29] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.

[30] Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. J Amer Statist Assoc 1951;46(253):68–78.

[31] Yamashita Rios de Sousa AM, Hlinka J. Assessing serial dependence in ordinal patterns processes using chi-squared tests with application to EEG data analysis. Chaos 2022;32(7):073126.

[32] Elsinger H. Independence tests based on symbolic dynamics. Working papers, Oesterreichische Nationalbank (Austrian Central Bank); 2010.

[33] Freedman D, Diaconis P. On the histogram as a density estimator: L2 theory. Z Wahrscheinlichkeitstheor Verwandte Geb 1981;57(4):453–76.

[34] Costa K, Frery AC, Cavalcanti G. Analysis of signals from air conditioner compressors with ordinal patterns and machine learning. J Low Freq Noise Vib Act Control 14613484241287620, 0.