Communications in Nonlinear Science and Numerical Simulation 105 (2022) 106077

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

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

Research paper

Complexity-based permutation entropies: From deterministic time series to white noise

José M. Amigó [a,*,1], Roberto Dale [a,1], Piergiulio Tempesta [b,c,1]

[a] Centro de Investigación Operativa, Universidad Miguel Hernández, Elche, Spain
[b] Departamento de Física Teórica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain
[c] Instituto de Ciencias Matemáticas, C/ Nicolás Cabrera, No 13–15, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 25 February 2021
Received in revised form 24 August 2021
Accepted 11 October 2021
Available online 13 October 2021

Keywords:
Time series analysis
Deterministic and random real-valued processes
Metric and topological permutation entropy
Permutation complexity classes
Permutation entropy rate for noisy processes
Discrimination of noisy time series
Numerical simulations

ABSTRACT

This is a paper in the intersection of time series analysis and complexity theory that presents new results on permutation complexity in general and permutation entropy in particular. In this context, permutation complexity refers to the characterization of time series by means of ordinal patterns (permutations), entropic measures, decay rates of missing ordinal patterns, and more. Since the inception of this "ordinal" methodology, its practical application to any type of scalar time series and real-valued processes have proven to be simple and useful. However, the theoretical aspects have remained limited to noiseless deterministic series and dynamical systems, the main obstacle being the super-exponential growth of allowed permutations with length when randomness (also in form of observational noise) is present in the data. To overcome this difficulty, we take a new approach through complexity classes, which are precisely defined by the growth of allowed permutations with length, regardless of the deterministic or noisy nature of the data. We consider three major classes: exponential, sub-factorial and factorial. The next step is to adapt the concept of Z-entropy to each of those classes, which we call permutation entropy because it coincides with the conventional permutation entropy on the exponential class. Z-entropies are a family of group entropies, each of them extensive on a given complexity class. The result is a unified approach to the ordinal analysis of deterministic and random processes, from dynamical systems to white noise, with new concepts and tools. Numerical simulations show that permutation entropy discriminates time series from all complexity classes.

1. Introduction

Complexity in symbolic times series, symbols being taken from a finite alphabet $\mathcal{A}$, has to do with the number of different sequences (strings, words, blocks, …) of a given length n and how this number increases with n. The perhaps simplest approach consists in counting the number of such sequences. In this case, the complexity of periodic sequences is a bounded function of n [1], while the complexity of arbitrary sequences grows as $|\mathcal{A}|^n$ ($|\cdot|$ denotes cardinality). Take the logarithmic growth rate, namely $\log|\mathcal{A}|$, to obtain the Shannon entropy of a memoryless process that outputs the symbols of $\mathcal{A}$ with equal probabilities. The positivity of the entropy differentiates then exponential from sub-exponential growth. Other approaches to the concept of complexity of sequences and the processes producing them have been proposed in

* Corresponding author.
E-mail addresses: jm.amigo@umh.es (J.M. Amigó), rdale@umh.es (R. Dale), p.tempesta@fis.ucm.es, piergiulio.tempesta@icmat.es (P. Tempesta).
[1] All authors have equally contributed to this paper.

https://doi.org/10.1016/j.cnsns.2021.106077

different fields. Thus, in information theory complexity is usually related to compression [2,3]. Here one counts the number of new words arising as one parses the whole message (ideally, a one-sided infinite binary sequence). In dynamical systems and symbolic dynamics, the main tool is the dynamic entropy, both in its metric and topological versions [4]. In computer science, algorithmic (or Kolmogorov) complexity refers to the shortest computer code that generates the sequence at hand, while computational problems are grouped into (polynomial, exponential, …) complexity classes according to how the amount of resources (time, memory, …) needed to solve them using a computation model (Turing machine, probabilistic Turing machine, quantum computer, …) depends on the "size" of the input (usually, the number of bits) [5]. In number theory and cryptography there are also several proposals, some of them going deep into the concepts of randomness, compressibility and typicality [6–8].

This paper deals with the concept of permutation complexity of real-valued time series introduced in [9–11], so our symbols will be ordinal patterns or permutations of length L [12]. As we will see more precisely in the following sections, the count of permutations grows exponentially with L in the case of (noiseless) deterministic signals, while it grows super-exponentially for noisy deterministic and random signals, sometimes called *noisy signals* hereafter for brevity. This different growth behavior of the ordinal patterns and, therefore, of the permutation complexity makes possible to distinguish deterministic signals from noisy signals but, at the same time, it poses a challenge for a unified quantification of permutation complexity for a simple reason: the usual tools for measuring complexity (say, Shannon and Kolmogorov–Sinai entropies) are designed for exponential growths of the symbols they are defined upon, thus diverging when applied to super-exponential growths. This occurs, in particular, with permutation entropy, which is the Shannon entropy of a time series in its *ordinal representation*, i.e., its symbolic representation via ordinal patterns.

As a result, the tools of permutation complexity are applied to time series analysis in different ways. In theoretical applications, where the time series are deterministic and may be assumed to be infinitely long, one uses entropic measures such as metric and topological permutation entropy, or the like. In practical applications, where the time series are noisy and finite, one typically resorts to permutations entropies of finite order (Section 2.1), causality–complexity planes [13,14], the decay rate of the missing patterns [15,16], and ordinal networks [17], to mention some typical techniques. This being the case, the objective of the present paper is to propose an integrating and overarching approach as follows.

The main character of this new approach is the logarithmic growth of allowed (or visible) ordinal patterns with increasing length. Depending on that growth, processes are collected in the exponential, sub-factorial and factorial complexity classes, whether they are deterministic or random. This procedure was inspired by similar ideas in complexity theory, where systems are usually classified according to the state growth rates of the states with the number of constituents N. For each of those classes there is a particular *group entropy*, called Z-entropy, that is *extensive* for the systems in the class, meaning that it is finite over the uniform probability distributions in the limit $N \to \infty$ [18,19]. In our context, system translates into process, the extensive parameter N into the length L of the ordinal patterns, Z-entropy into (generalized) permutation entropy, and extensivity into the convergence of the corresponding topological permutation entropy rate. Nevertheless, the introduction of the key concepts will be self-contained and will concentrate on time series and processes, so that the reader can understand their rationale and properties without further reference. The result is a characterization of time series in the ordinal representation that focuses on complexity rather than data generation. This way we extend the realm of the standard permutation entropy from deterministic processes (dynamical systems) to random processes, thus filling a conceptual gap in permutation complexity. A first step in this direction was taken in [20], where we used the Z-entropy of the factorial complexity class to define a generalized permutation entropy for noisy signals without forbidden patterns, i.e., noisy dynamics and random processes such that all ordinal patterns of any length are allowed; these are the kind of signals encountered in practice. Our approach here is more general and comprehensive.

Regarding the notion of group entropy mentioned above, it was introduced in [21] and discussed, e.g., in [18,19,22–25]. Essentially, a group entropy is a functional defined on a probability space which satisfies several important properties, such as the first three Shannon–Khinchin axioms (Section 4) and a so-called composability axiom: the entropy of a system compound by two statistically independent systems is expressed by a formal group law [20]. By construction, group entropies have a direct interpretation as information measures [18,25]. In particular, they can be used to define divergences and Riemannian structures over statistical manifolds.

The ordinal approach, where the information contained in the ordinal patterns is exploited via probability distributions, entropies, etc., is quite popular in time series analysis for a number of reasons, including its computational simplicity and speed. Applications to biomedicine where among the first and include epilepsy [26], cardiopathies [27], heart rate variability [28], and more [29]. Further applications include dynamical change detection [30], signal characterization [16,31,32], and image processing [33,34]. Currently, ordinal techniques, alone or complemented by other methods, are being applied in plenty of fields, e.g., chaotic dynamics, earth science, computational neuroscience, and econophysics; see [35] for examples, and [36] for a recent survey.

The rest of this paper is organized as follows. Section 2 contains the mathematical setting for the subsequent discussion, in particular, metric and topological permutation entropies as well as the concepts of allowed and forbidden patterns. In doing so, we cover the full range of discrete-time, real-valued time series envisaged in this paper, namely: noiseless deterministic, noisy deterministic, and random signals. This section is partially based on our paper [20]. Section 3 is devoted to the permutation complexity function and classes. Here we introduce the exponential, sub-factorial and factorial permutation complexity classes that are further analyzed in the subsequent sections. In Section 4 we briefly review the

general concept of entropy (based on the Shannon–Khinchin axioms), before extending permutation entropy from the exponential class to the factorial and sub-factorial classes. Numerical simulations is the subject of Section 5. In this section, the discriminatory power of the permutation entropy (Section 5.1) and the permutation complexity function (Section 5.2) is put to the test with a battery of seven noisy processes from the factorial class. In Section 5.3 we study numerically and analytically a toy model for sub-factorial processes. The conclusions are summarized in Section 6 .

2. Permutation complexity

Real-valued time series typically result from sampling analog signals or observing dynamical flows at discrete times. A further step in the analysis of such series can be the discretization of the data, a procedure that is usually called symbolic representation. The information provided by a symbolic representation may be sufficient for the intended application while simplifying the mathematical tools needed for the analysis. In this regard, ordinal patterns [12] are becoming increasingly popular to represent symbolically real-valued time series. Some reasons for this is their mathematically sound relation to Kolmogorov–Sinai entropy via permutation entropy [37–40] and their ease of computation. Ordinal patterns and permutation entropies are the main ingredients of permutation complexity.

2.1. Ordinal representations and permutation entropy

Given a (finite or infinite) time series $(x_t)_{t\geq 0} = x_0, x_1, \ldots, x_t, \ldots$, where $t = 0, 1, \ldots, N \leq \infty$ is discrete time and $x_t \in \mathbb{R}$, its symbolic representation by *ordinal patterns of length* $L \geq 2$ is $\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_t, \ldots$, where $\mathbf{r}_t$ is the rank vector of the string $x_t^L := x_t, x_{t+1}, \ldots, x_{t+L-1}$ $(t \leq N-L+1)$, i.e., $\mathbf{r}_t = (\rho_0, \rho_1, \ldots, \rho_{L-1})$ where $\{\rho_0, \rho_1, \ldots, \rho_{L-1}\} \in \{0, 1, \ldots, L-1\}$ are such that

$$x_{t+\rho_0} < x_{t+\rho_1} < \cdots < x_{t+\rho_{L-1}} \tag{1}$$

(other rules can also be found in the literature). In case of two or more ties, one can adopt some convention, e.g., the earlier entry is smaller. Sometimes we say that x_t^L defines the *ordinal L-pattern* $\mathbf{r}_t$ or that it is of *type* $\mathbf{r}_t$. Ordinal L-patterns can be identified with permutations of $\{0, 1, \ldots, L-1\}$, i.e., with elements of the symmetric group of degree L, $\mathcal{S}_L$; the cardinality of $\mathcal{S}_L$, $|\mathcal{S}_L|$, is $L!$. Symbolic representations of time series by means of ordinal patterns are called *ordinal representations*. The algebraic structure of $\mathcal{S}_L$ was exploited in [11] , which led to the more general concept of algebraic representations.

Furthermore, the time series $(x_t)_{t\geq 0}$ is assumed to be output by a discrete-time deterministic or random process $\mathbf{X}$ taking values on an interval $I \subset \mathbb{R}$. By deterministic process we mean a one dimensional dynamical system $(I, \mathcal{B}, \mu, f)$, where I (the state space) is a bounded interval of $\mathbb{R}$, $\mathcal{B}$ is the Borel σ-algebra of I, μ is a measure over the measurable space $(I, \mathcal{B})$ such that $\mu(I) = 1$ (i.e., $(I, \mathcal{B}, \mu)$ is a probability space) and, for the time being, $f : I \to I$ is any μ -invariant map (i.e., $\mu(f^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$); alternatively, we say that μ is f-invariant. In this case, the output $(x_t)_{t\geq 0}$ of $\mathbf{X}$ is the orbit of x_0, i.e., $(x_t)_{t\geq 0} = (f^t(x_0))_{t\geq 0}$, where $f^0(x_0) = x_0 \in I$ and $f^t(x_0) = f(f^{t-1}(x_0))$. An ordinal representation of the orbits of f by ordinal L-patterns partitions the state space I into the $L!$ bins

$$P_{\mathbf{r}} = \{x \in I : \ (x, f(x), \ldots, f^{L-1}(x)) \text{ is of type } \mathbf{r} \in \mathcal{S}_L\}. \tag{2}$$

Therefore, the probability $p(\mathbf{r})$ of the ordinal pattern $\mathbf{r} \in \mathcal{S}_L$ to occur in an output of the deterministic process $\mathbf{X}$ generated by the map f is

$$p(\mathbf{r}) = \mu(P_{\mathbf{r}}). \tag{3}$$

Note that, although the outputs $(x_t)_{t\geq 0}$ are deterministic ("sharp" orbits), their ordinal representations $(\mathbf{r}_t)_{t\geq 0}$ are random sequences ("pixelated" orbits), as occurs with any symbolic dynamics of a map with respect to a partition of its state space [9].

The *metric permutation entropy* (*rate*) of the process $\mathbf{X}$ is defined as

$$h^*(\mathbf{X}) = \limsup_{L\to\infty} \frac{1}{L} H^*(X_0^L), \tag{4}$$

where $X_0^L = X_0, X_1, \ldots, X_{L-1}$ and

$$H^*(X_0^L) = -\sum_{\mathbf{r}\in\mathcal{S}_L} p(\mathbf{r}) \ln p(\mathbf{r}) \tag{5}$$

is the *metric permutation entropy of* $\mathbf{X}$ *of order* L.

In other words, $H^*(X_0^L)$ is the Shannon entropy of the probability distribution $\{p(\mathbf{r}) : \mathbf{r} \in \mathcal{S}_L\}$. If $\mathbf{X}$ is a deterministic process (and μ is known), then $p(\mathbf{r})$ is given as in Eqs. (2)–(3). If $\mathbf{X}$ is a random process, the probabilities $p(\mathbf{r})$ can only exceptionally be derived from the probability distributions of $\mathbf{X}$ [41] so, in general, they have to be estimated, e.g. by relative frequencies:

$$\hat{p}(\mathbf{r}) = \frac{\left|\{x_t^L \text{ of type } \mathbf{r} \in \mathcal{S}_L : 0 \leq t \leq N-L+1\}\right|}{N-L+2}. \tag{6}$$

In the theoretical case of an infinite time series, take the limit $N \to \infty$ in (6). In nonlinear time series analysis, the ergodic invariant measure defined by $\mu(P_{\mathbf{r}}) = \hat{p}(\mathbf{r})$ is called the physical or natural measure because it is the only relevant measure for physical systems and numerical simulations [42]. More about this in Section 5.1.

Remark 1. The limit $\lim_{N\to\infty} \hat{p}(\mathbf{r})$ exists with probability 1 when the underlying stochastic process fulfills the following weak stationarity condition: for $k \leq L-1$, the probability for $x_t < x_{t+k}$ should not depend on t [12] . This is the case for stationary processes but also for non-stationary processes with stationary increments such as the fractional Brownian motion [43] and its increments, that is, the fractional Gaussian noise. We will use these random processes, which have long range dependencies, in the numerical simulations.

Let $\mathbf{X}$ be a deterministic or random process that takes values on an interval $I \subset \mathbb{R}$. We say that an ordinal pattern $\mathbf{r} \in \mathcal{S}_L$ is *allowed* for $\mathbf{X}$ if the probability that a string x_t^L of type $\mathbf{r}$ is output by $\mathbf{X}$ is positive. That is, an L-pattern $\mathbf{r}$ is allowed if there are strings $x_t, \ldots, x_{t+L-1}$ in some outputs or orbits of $\mathbf{X}$ such that the type of those strings is $\mathbf{r}$. Otherwise, the ordinal L-pattern $\mathbf{r}$ is *forbidden* for $\mathbf{X}$. For example, the ordinal 3-pattern $\mathbf{r} = (2, 1, 0)$ is forbidden for the logistic map $f(x) = 4x(1-x)$, $0 \leq x \leq 1$, because there is no string x_t, x_{t+1}, x_{t+2} in any orbit of f such that $x_{t+2} < x_{t+1} < x_t$; all other 3-patterns $\mathbf{r} = (\rho_0, \rho_1, \rho_2)$, where $\rho_0, \rho_1, \rho_2 \in \{0, 1, 2\}$ and $\mathbf{r} \neq (2, 1, 0)$, are allowed for the logistic map, that is, $x_{t+\rho_0} < x_{t+\rho_1} < x_{t+\rho_2}$ for x_t in a suitable subinterval of [0, 1], see [9]. Since we do not consider patterns other than ordinal patterns in this paper, we speak of allowed and forbidden patterns for brevity.

If $\mathcal{A}_L(\mathbf{X})$ denotes the number of allowed patterns of length L for $\mathbf{X}$, the *topological permutation entropy* (*rate*) of the process $\mathbf{X}$ is then defined as

$$h_0^*(\mathbf{X}) = \limsup_{L\to\infty} \frac{1}{L} H_0^*(X_0^L), \tag{7}$$

where

$$H_0^*(X_0^L) = \ln \mathcal{A}_L(\mathbf{X}) \tag{8}$$

is the *topological permutation entropy of* $\mathbf{X}$ *of order* L. Moreover,

$$H^*(X_0^L) \leq H_0^*(X_0^L) \leq \ln L!, \tag{9}$$

where $H^*(X_0^L) = H_0^*(X_0^L)$ for flat probability distributions of the allowed L-patterns, and $H_0^*(X_0^L) = \ln L!$ if all L-patterns are allowed.

2.2. Allowed pattern growths for deterministic and random processes

The map $f : I \to I$ is called *piecewise monotone* if there is a finite partition of I such that f is continuous and strictly monotone on each subinterval of the partition. If the graph of f has n humps, then f is called unimodal ($n = 1$) or multimodal ($n > 1$). Most of the one-dimensional maps encountered in practice are piecewise monotone, so this condition does not imply any strong restriction for practical purposes. Let $h_0(f)$ denote the topological entropy of f, and $h(f)$ its metric (or Kolmogorov–Sinai) entropy [44]. The following theorem holds [37].

Theorem 2. *If f is piecewise monotone, then (a) $h^*(f) = h(f)$, and (b) $h_0^*(f) = h_0(f)$.*

Theorem 2(a) was generalized to countably piecewise monotone maps in [40]. Generalizations to higher dimensional intervals can be found in [45].

From Theorem 2(b) and Eqs. (7)–(8) it follows that

$$\ln \mathcal{A}_L(\mathbf{X}) = \ln |\{\text{allowed } L\text{ -patterns for deterministic } \mathbf{X}\}| \sim h_0(f)L, \tag{10}$$

where f is the map generating the outputs of $\mathbf{X}$, the symbol $\sim$ stands for "asymptotically when $L \to \infty$" (i.e., $\lim_{L\to\infty} \ln \mathcal{A}_L(\mathbf{X})/(h_0(f)L) = 1$) and, for the sake of this paper, we assume $h_0(f) > 0$ throughout. Therefore, the number of allowed L-patterns for a piecewise monotone map f grows exponentially with L. To be more precise, according to the proof of Proposition 1(b) in [37], $\mathcal{A}_L(\mathbf{X}) \sim \exp[h_0(f)L + \ln L + const]$.

Remark 3. More generally, it is easy to show that $\ln f(L) \sim \phi(L)$ if and only if $f(L) = \exp[\phi(L) + o(\phi(L))]$, where $o(\phi(L))$ denotes a function such that $o(\phi(L))/\phi(L) \to 0$ when $L \to \infty$. Exponential growth corresponds to $\phi(L)$ linear, as in Eq. (10). In this case, $f(L) = \mathcal{A}_L(\mathbf{X})$, $\phi(L) = h_0(f)L$ and, hence, $o(\phi(L)) = \ln L + const = o(L)$.

Since, on the other hand, the number of possible L-patterns is $L!$ and

$$\ln L! \sim L \ln L \tag{11}$$

by Stirling's formula $\ln L! \simeq L(\ln L - 1) + \frac{1}{2}\ln(2\pi L)$, we conclude from Eq. (10) that deterministic processes necessarily have forbidden L-patterns for L large enough and, in fact, the number of forbidden L-patterns grows super-exponentially

with L. By *deterministic process* we mean here and hereafter the dynamics generated by (the iteration of) a piecewise monotone map so that Theorem 2 is applicable and Eq. (10) holds with $h_0(f) > 0$. Sometimes we write $\mathbf{X} = f$ in this case.

As mentioned before, $\mathbf{r} = (2, 1, 0)$ is the only forbidden 3-pattern for the logistic parabola, while all 3-patterns are allowed for the shift map $x \mapsto 2x \bmod 1$ [9]. The respective number of forbidden 4-patterns is 12 and 6 [9]. Each forbidden pattern of length L_0 in a deterministic dynamic is the seed of an infinitely long trail of "outgrowth forbidden patterns" of lengths $L > L_0$ whose structure can be found in [9]. Let us mention in passing that forbidden patterns there exist also in higher dimensional dynamics (at least) for expansive maps, ordinal patterns being defined lexicographically [46]. Therefore, projections of higher dimensional dynamics are expected to have forbidden patterns and exponential growths of allowed patterns as well.

At the other extreme are random processes without forbidden patterns, that is, processes for which all ordinal patterns of any length are allowed and, hence, their growth is factorial: $\mathcal{A}_L(\mathbf{X}) = |\mathcal{S}_L| = L!$. A trivial example of a random process without forbidden patterns is white noise.

Also noisy deterministic time series may not have forbidden patterns (for sufficiently long series). Indeed, when the dynamics takes place on a nontrivial attractor so that the orbits are dense, then the observational (white) noise will "destroy" all forbidden patterns in the long run, no matter how small the noise. For this reason, we sometimes call noisy deterministic processes and other random processes without forbidden patterns just *forbidden-pattern-free* (FPF) *processes* or signals. Unlike Eq. (10) for deterministic processes, for FPF processes we have

$$\ln \mathcal{A}_L(\mathbf{X}) = \ln |\{\text{allowed } L \text{ -patterns for FPF } \mathbf{X}\}| = \ln L! \sim L \ln L, \tag{12}$$

where we used the asymptotic equivalence (11).

To complete the picture, let us point out that random processes can have forbidden patterns too. A conceptually simple (though impractical) way of constructing such a process is to repeatedly draw x_t until the type of the block $x_0, x_1, \ldots, x_t$ is allowed, for $t = 1, 2, \ldots$ By controlling the number of allowed L-patterns, this constrained random process outputs time series with any feasible growth of allowed L-patterns, in particular, an exponential one (as in the deterministic case). A more realistic example of a random process with a sub-factorial growth of allowed pattern is the following.

Example 4 (*Not-So-Noisy Measurement of a Periodic Signal*)**.** Suppose that a periodic time series $(y_t)_{t\geq 0} = (f^t(y_0))_{t\geq 0}$ of prime period $\mathfrak{p} \geq 2$ is observed; so

$$y_t = f^k(y_0) = y_k$$

for every $t = k \bmod \mathfrak{p}$, $k = 0, 1, \ldots, \mathfrak{p} - 1$, where $y_0 < y_1 < \cdots < y_{\mathfrak{p}-1}$ for simplicity. Furthermore, suppose that the points y_k are measured with a device whose precision is value dependent, so that only the measurement of, say, $y_{\mathfrak{p}-1}$ is noiseless and, otherwise, the uncertainty intervals of $y_0, \ldots, y_{\mathfrak{p}-2}$ do not overlap. To model this situation, let $\delta > 0$ be the minimum separation between the points of the periodic cycle $(y_0, y_1, \ldots, y_{\mathfrak{p}-1})$ and add white noise to y_t with amplitude less than $\delta/2$, except when $t = \mathfrak{p} - 1 \bmod \mathfrak{p}$. That is, the noisy observations are $x_t = y_t + \zeta_t$ for $t = 0, 1, \ldots, \mathfrak{p} - 2 \bmod \mathfrak{p}$, where ζ_t are independent and uniformly distributed random variables in $(-\delta/2, \delta/2)$, and $x_t = y_t$ for $t = \mathfrak{p} - 1 \bmod \mathfrak{p}$, so that $x_{\nu\mathfrak{p}} < x_{\nu\mathfrak{p}+1} < \cdots < x_{(\nu+1)\mathfrak{p}-1} = y_{\mathfrak{p}-1}$ for all $\nu \in \mathbb{N}$. Choose $L = \nu\mathfrak{p}$ for simplicity. Then the number of allowed L-patterns is given by

$$\mathcal{A}_L(\mathbf{X}_\mathfrak{p}) = \mathfrak{p}\,(\nu!)^{\mathfrak{p}-1} = \mathfrak{p}[(L/\mathfrak{p})!]^{\mathfrak{p}-1}, \tag{13}$$

where $\mathbf{X}_\mathfrak{p}$ is the noisy process that outputs the time series $(x_t)_{t\geq 0}$. The factor $\mathfrak{p}$ in Eq. (13) comes from the $\mathfrak{p}$ different values of the time index t modulus $\mathfrak{p}$. For each such t (say, $t = 0, 1, \ldots, \mathfrak{p} - 1$), the window $x_t^L = x_t, x_{t+1}, \ldots, x_{t+L-1}$ splits in $\mathfrak{p}$ disjoint groups of ν points each as follows:

$$\{x_{t+j} : t + j = 0 \bmod \mathfrak{p}\} < \{x_{t+j} : t + j = 1 \bmod \mathfrak{p}\} < \cdots . < \{x_{t+j} : t + j = \mathfrak{p} - 1 \bmod \mathfrak{p}\}, \tag{14}$$

where $0 \leq j \leq L - 1$ and $x_{t+j} = y_{\mathfrak{p}-1}$ for all $t + j = \mathfrak{p} - 1 \bmod \mathfrak{p}$ (last group). As a result, the time indices of the ν noisy points x_{t+j} in each of the first $\mathfrak{p} - 1$ groups of the splitting (14) can be ordered in any of the $\nu!$ permutations (ν-patterns) possible, while the time indices of the noiseless points $x_{t+j} = y_\mathfrak{p}$ in the last group leads to only one ν-pattern, namely: the permutation consisting of the corresponding time indices in increasing order (according to the convention for repeated values). This explains the second factor $(\nu!)^{\mathfrak{p}-1}$ in Eq. (13). Therefore, as $\nu = L/\mathfrak{p}$ increases,

$$\ln \mathcal{A}_L(\mathbf{X}_\mathfrak{p}) = \ln \mathfrak{p} + (\mathfrak{p} - 1) \ln[(L/\mathfrak{p})!] \sim (\mathfrak{p} - 1)\frac{L}{\mathfrak{p}} \ln \frac{L}{\mathfrak{p}} = \frac{\mathfrak{p} - 1}{\mathfrak{p}} L(\ln L - \ln \mathfrak{p}) \sim cL \ln L \tag{15}$$

where $c = (\mathfrak{p} - 1)/\mathfrak{p} < 1$ and $L = \nu\mathfrak{p}$. Obviously, if the number of "noiseless" measurements of the periodic cycle $(y_0, y_1 ..., y_{\mathfrak{p}-1})$ is generalized to m, then, $c = (\mathfrak{p} - m)/\mathfrak{p}$.

The noisy process presented in Example 4 will be discussed with greater detail in Section 5.3. In particular, the asymptotic growth of $\ln \mathcal{A}_L(\mathbf{X}_\mathfrak{p})$ depends on L modulus $\mathfrak{p}$.

Since real-world data is noisy, one certainly expects super-exponentially growing numbers of allowed patterns in empirical observations, although sub-factorial growths such as in Eq. (15) seem elusive.

3. Permutation complexity functions and classes

Next we wish to associate the notion of permutation complexity to processes ranging from deterministically generated signals to white noise. Unfortunately, the metric and topological permutation entropies are not up to the job. For instance, $h_0^*(\mathbf{X})$ converges for deterministic processes (Theorem 2) but diverges for forbidden-pattern-free (FPF) signals:

$$h_0^*(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} \ln \mathcal{A}_L(\mathbf{X}) = \lim_{L\to\infty} \ln L = \infty \tag{16}$$

by (12).

This being the case, we shall rather focus on the *permutation complexity* (PC) *class* of the process $\mathbf{X}$, which we define by the asymptotic growth of $\ln \mathcal{A}_L(\mathbf{X})$ with respect to L. In view of Eq. (10) for the deterministic processes and Eq. (12) for the FPF processes, we propose the following definition.

Definition 5. Let $g(t)$ be a positive, invertible and sufficiently regular function of the real variable $t \geq 0$. A process $\mathbf{X}$ is said to belong to the PC class g if

$$\ln \mathcal{A}_L(\mathbf{X}) \sim g(L) \tag{17}$$

as $L \to \infty$.

The function $g(t)$ will be called the *permutation complexity* (PC) *function* of the process $\mathbf{X}$. The name of $g(t)$ is suggested by Eq. (10) with $L = \lfloor t \rfloor$, since the topological entropy $h_0(f)$ measures the dynamical complexity of the deterministic dynamic generated by f. In some cases, for convenience or economy, we will group a family of classes under a single "super-class", although we will also call them classes.

Remark 6. Two important observations on the PC function of a process:

(1) Regarding regularity, we will assume henceforth that $g(t)$ is bicontinuous, i.e., both $g(t)$ and its inverse $g^{-1}(s)$ are continuous. The bicontinuity and invertibility of $g(t)$ imply that $g(t)$ and, hence, $g^{-1}(t)$ are strictly monotonic [47], in fact, strictly increasing in our case.
(2) Regarding uniqueness, the complexity class g depends only on the asymptotic behavior of $g(t)$; any other function $\tilde{g}(t) \sim g(t)$ (i.e., $\tilde{g}(t) = g(t) + o(g(t))$) will work out as well. Put in other terms, PC classes are defined up to asymptotic equivalence.

Considering the growth of $\mathcal{A}_L(\mathbf{X})$, there is a first clear-cut division of processes: deterministic processes, for which $\mathcal{A}_L(\mathbf{X})$ grows exponentially, and FPF processes, for which $\mathcal{A}_L(\mathbf{X})$ grows factorially. Data analysis and numerical simulations show that the latter are ubiquitous in practice. Processes with super-exponential but sub-factorial growths will be grouped in a third class. Specifically, we are going to turn our attention to the following three PC classes.

(C1) *Exponential class*: $\ln \mathcal{A}_L(\mathbf{X}) \sim cL$ ($c > 0$), i.e.,

$$g(t) = ct =: g_{\text{exp}}(t). \tag{18}$$

Thus, the exponential class is actually a class of classes, one for each c. Each class with a given constant c includes all deterministic processes $\mathbf{X} = f$ with topological entropy $h_0(f) = c$; maps with the same $h_0(f)$ are said to be topologically conjugate. Therefore, deterministic processes with different topological entropies have different permutation complexities, in line with the concept of dynamical complexity.

Moreover, for each $c > 0$ the corresponding class is non-empty. Indeed, for every $\sigma > 1$ there exists a piecewise monotone map f with $h_0(f) = \ln \sigma > 0$, namely, the piecewise linear selfmap of the interval $[0, 1]$ with constant slopes $\pm\sigma$. Therefore, any function of the form $g(t) = ct$ is the PC function of a deterministic processes generated by a piecewise linear map with $\sigma = e^c$.

(C2) *Factorial class*: $\ln \mathcal{A}_L(\mathbf{X}) \sim L \ln L$, i.e.,

$$g(t) = t \ln t =: g_{\text{fac}}(t). \tag{19}$$

Regarding the applications, the factorial class is the most interesting since virtually all random processes in practice are FPF.

(C3) *Sub-factorial class*: $\ln \mathcal{A}_L(\mathbf{X}) \sim g(t)$, where (i) $g_{\text{exp}}(t) = o(g(t))$ and $g(t) = o(g_{\text{fac}}(t))$ or, else, (ii)

$$g(t) := ct \ln t \quad \text{with } 0 < c < 1. \tag{20}$$

Unlike the exponential and factorial classes, whose PC functions are defined explicitly, the PC functions of the sub-factorial class are defined both implicitly (condition C3(i)) and explicitly (Eq. (20)).

The sub-factorial class is also a class of classes. This class is potentially the largest since it fills the gap between the exponential and the factorial class, although practical examples are hard to find. Examples of functions $g(t)$ such that $ct = o(g(t))$ and $g(t) = o(t \ln t)$ (condition C3(i)) are

$$g(t) = t \ln^{(n)} t \quad (n \geq 2), \tag{21}$$

where $\ln^{(n)} t$ denotes the composition of the logarithmic function n times. Toy models with PC functions of the form (20) were presented in Example 4. Prompted by this example, in the forthcoming theorems we will use $g_{\text{sub}}(t) := ct \ln t$, $0 < c < 1$, as a prototypical PC function of the sub-factorial class, although the other representatives in Eq. (21) will also be considered alongside.

Let us mention in passing that $g(t) = ct \ln t$ with $c > 1$ is not the PC function of any random process in the ordinal representation. However, statistical complex systems may have such super-factorial growth rates of the state space as the number of constituents increases [24].

Of course, the exponential and sub-factorial classes can be thought of as refined in smaller classes whenever convenient.

To conclude this section, let us return to the asymmetry between deterministic and FPF processes regarding their PC functions. As already mentioned, $g_{\text{exp}}(t) = ct$ distinguishes deterministic processes from each other up to topological conjugacy, since $c = h_0(f)$ in this case. On the contrary, all FPF processes have the same PC function, namely, $g_{\text{fac}}(t) = t \ln t$, the reason being that $\mathcal{A}_L(\mathbf{X})$ counts the number of allowed L-patterns for $L \gg 1$, and this number is $L!$ for all FPF processes. The result is that $g_{\text{fac}}(t)$ is useless in distinguishing FPF processes from each other. A possible way out of this shortcoming is to take into account the probability distribution of the allowed L-patterns, e.g., through permutation entropies tailored to each PC class, as we do in the next section. A different approach, based on the convergence rate of $\ln \mathcal{A}_L(\mathbf{X})$ to $g_{\text{fac}}(t) = t \ln t$ with the length of the time series, will be presented in Section 5.2, when discussing numerical simulations.

4. Generalized permutation entropy

Let $p = (p_1, p_2, \ldots, p_W)$ be a discrete probability distribution; we denote by $\mathcal{P}_W$ the set of all discrete probability distributions with W entries. From the point of view of information theory, an entropy is a positive functional $S(p)$ defined on $\cup_{W \geq 2} \mathcal{P}_W$ that satisfies certain properties required by Shannon [48,49] and Khinchin [50] in their uniqueness theorem for $S(p)$, and nowadays known as the *Shannon–Khinchin* (SK) *axioms*. The first three (SK) axioms are:

(SK1) *Continuity*: S is continuous on $\mathcal{P}_W$ for each W.

(SK2) *Maximality*: For each $(p_1, p_2, \ldots, p_W) \in \mathcal{P}_W$,

$$S(p_1, p_2, \ldots, p_W) \leq S\left(\tfrac{1}{W}, \tfrac{1}{W}, \ldots, \tfrac{1}{W}\right).$$

(SK3) *Expansibility*: For each $(p_1, p_2, \ldots, p_W) \in \mathcal{P}_W$ and $i \in \{0, 1, \ldots, n-1\}$,

$$S(p_1, \ldots, p_i, 0, p_{i+1}, \ldots, p_n) = S(p_1, p_2, \ldots, p_n).$$

If $S(p)$ satisfies (SK1)-(SK3) and a fourth axiom called *separability* or *strong additivity* (SK4), then $S(p)$ must be the *Boltzmann–Gibbs–Shannon entropy* (usually called Shannon entropy in information theory):

$$S(p) = -k \sum_{i=1}^{W} p_i \ln p_i =: S_{BGS}(p), \tag{22}$$

where k is an arbitrary positive constant that can be interpreted as the freedom in the choice of the logarithm base. If, otherwise, $S(p)$ only satisfies the first three SK axioms, then $S(p)$ is called a *generalized entropy* and its form is only known under additional assumptions [51,52].

Remark 7. In the case of *group entropies*, of interest in this work, the strong additivity axiom (SK4) is replaced by the *composability axiom*, namely, the requirement that there exists a suitable function of the form $\Phi(x, y) = x + y +$ higher order terms, which takes care of the composition process of two independent systems that are described by probability distributions. Specifically,

$$S(p \times q) = \Phi(S(p), S(q)), \tag{23}$$

where p, q are any two probability distributions and $p \times q$ is their product distribution. Here Φ is supposed to satisfy three properties: *(i)* $\Phi(x, y) = \Phi(y, x)$ (symmetry), (ii) $\Phi(x, \Phi(y, z)) = \Phi(\Phi(x, y), z)$ (associativity), and *(iii)* $\Phi(x, 0) = x$ (*null-composability*), which coincide with those of a formal group law [18]. Thus, a group entropy is a functional satisfying the first three SK axioms and the composability axiom. Property (23) is actually crucial to generalize the standard notion of entropy. The entropies of Shannon (22), Rényi (24), and Tsallis [53] belong to this class. A multivariate extension of the notion of group entropy has been proposed in [19]. An independent axiomatic approach to composable entropies, the pseudoadditive entropies, has been discussed in [52] (see also the references therein).

As it turns out, $S_{BGS}(p)$ is not well suited to deal with the diversity of complex systems, including the thermodynamical ones. In complexity theory, systems are usually classified in sub-exponential, exponential and super-exponential "complexity classes", according to the state growth rates of the states with the number of constituents N. For each of such classes there is a specific group entropy, called Z-entropy, that is *extensive* for the systems in the class, meaning that it is finite over uniform probability distributions in the limit $N \to \infty$ [18,19].

In this section we capitalize on the similarities between this approach and ours to extend the concept of permutation entropy from deterministic processes to random processes via the Z-entropies for the exponential, sub-factorial and factorial complexity classes.

4.1. Permutation entropy of finite order

Given a probability distribution $p = (p_1, \ldots, p_W)$ and $\alpha \in \mathbb{R}$, $\alpha > 0$, the Rényi entropy $R_\alpha(p)$ is defined as [54]

$$R_\alpha(p) = \frac{k}{1-\alpha} \ln \left(\sum_{i=1}^{W} p_i^\alpha \right) \tag{24}$$

($k > 0$) for $\alpha \neq 1$, and

$$R_1(p) := \lim_{\alpha \to 1} R_\alpha(p) = -k \sum_{i=1}^{W} p_i \ln p_i = S_{BGS}(p), \tag{25}$$

see Eq. (22). In statistical mechanics, $k = 1.380649 \times 10^{-23}$ JK^{-1} is the Boltzmann constant; in information theory, k is usually set equal to 1, as we do from now on.

The following definition is an adaptation to our context of the concept of Z-entropy [18,19]. Remember that, according to Remark 6 on the PC function $g(t)$ of a process, its inverse $g^{-1}(s)$ is continuous and strictly increasing.

Definition 8. Let $g(t)$ be the PC function of a process **X**. The *(metric) permutation entropy of order L* of **X** is defined as

$$Z^*_{g,\alpha}(X_0^L) \equiv Z^*_{g,\alpha}(p) = g^{-1}(R_\alpha(p)) - g^{-1}(0), \tag{26}$$

where $\alpha > 0$, p is the probability distribution of the ordinal L -patterns of $X_0^L = X_0, X_1, \ldots, X_{L-1}$, and $R_\alpha(p)$ is Rényi's entropy.

The term $-g^{-1}(0)$ in (26) ensures that $Z^*_{g,\alpha}(X_0^L) = 0$ for *singular* probability distributions, i.e., when $p_{i_0} = 1$ and $p_i = 0$ for $i \neq i_0$. By the continuity and strictly increasing monotonicity of $g^{-1}(s)$, $Z^*_{g,\alpha}(X_0^L)$ fulfills the axioms (SK1)-(SK3), i.e., $Z^*_{g,\alpha}(X_0^L)$ is a generalized entropy. In addition, $Z^*_{g,\alpha}(X_0^L)$ satisfies the composability axiom (23) with $\Phi(x, y) = \chi^{-1}(\chi(x) + \chi(y))$, where $\chi(t) = g(t + g^{-1}(0))$ and, hence, $\chi^{-1}(s) = g^{-1}(s) - g^{-1}(0)$.

By its definition (and the increasing monotonicity of $g^{-1}(s)$), $Z^*_{g,\alpha}(X_0^L)$ inherits some of the properties of $R_\alpha(p)$. For instance, $Z^*_{g,\alpha}(X_0^L)$ is monotone decreasing with respect to the parameter α [51],

$$Z^*_{g,\alpha}(X_0^L) \geq Z^*_{g,\beta}(X_0^L) \text{ for } \alpha < \beta \tag{27}$$

and each $L \geq 2$.

To formulate the next theorem, we need to introduce the special function $\mathcal{L}(x)$, by which we denote the *principal* branch of the real *W-Lambert function*. This is a smooth function, defined as the solution of $ye^y = x$, i.e., $W(x)e^{W(x)} = x$, for $x \geq -e^{-1}$. $\mathcal{L}(x)$ is the unique solution for $x \geq 0$, while for $-e^{-1} \leq x < 0$ there is another solution belonging to a second branch. Some basic properties of $\mathcal{L}(x)$ are the following [55]: (i) $\mathcal{L}(x)$ is strictly increasing and $\cap$-convex; (ii) $\mathcal{L}(-e^{-1}) = -1$ and $\mathcal{L}(0) = 0$; (iii) $\mathcal{L}(x) > 0$ for $x > 0$; (iv) $\mathcal{L}(x) > 1$ for $x > e$; and (v) $\mathcal{L}(x) \to \infty$ as $x \to \infty$. Moreover, $\mathcal{L}(x)$ satisfies the identity

$$\mathcal{L}(x \ln x) = \ln x \tag{28}$$

for $x \geq e^{-1}$.

Theorem 9. *Given a process* **X**, *let p be the probability distribution of the ordinal L-patterns of* X_0^L. *For the PC classes (C1)–(C3) of Section 3, the following holds.*

(a) *For* $g_{exp}(t) = ct$:

$$Z^*_{g_{exp},\alpha}(X_0^L) = \frac{1}{c} R_\alpha(p) =: Z^*_{exp,\alpha}(X_0^L). \tag{29}$$

(b) *For* $g_{fac}(t) = t \ln t$:

$$Z^*_{g_{fac},\alpha}(X_0^L) = e^{\mathcal{L}[R_\alpha(p)]} - 1 =: Z^*_{fac,\alpha}(X_0^L). \tag{30}$$

(c) *For* $g_{sub}(t) = ct\ln t$ $(0 < c < 1)$*:*

$$Z^*_{g_{sub},\alpha}(X_0^L) = e^{\mathcal{L}[R_\alpha(p)/c]} - 1 =: Z^*_{sub,\alpha}(X_0^L). \tag{31}$$

Proof. Eq. (29) follows readily from the definition (26) and

$$g^{-1}_{\exp}(s) = \frac{s}{c}.$$

If $s = ct\ln t$ $(0 < c \leq 1)$, then $\mathcal{L}(s/c) = \mathcal{L}(t\ln t) = \ln t$ by the identity (28). Hence

$$g^{-1}_{\text{fac}}(s) = e^{\mathcal{L}(s)}$$

for $c = 1$,

$$g^{-1}_{\text{sub}}(s) = e^{\mathcal{L}(s/c)}$$

for $c < 1$, and $g^{-1}_{\text{fac}}(0) = g^{-1}_{\text{sub}}(0) = e^{\mathcal{L}(0)} = 1$. Eqs. (30) and (31) follow.

Remark 10. Regarding Theorem 9, let us highlight the following points.

(1) $Z^*_{\exp,\alpha}(p) = R_\alpha(p)$ for $c = 1$, that is, the sub-class that includes the maps with topological entropy 1. Since $R_\alpha(p)$ is defined anyway up to a positive constant k, see Eqs. (24)–(25), we may conclude that Rényi's entropy (of the probability distribution of the ordinal L-patterns) is the permutation entropy when dealing with deterministic processes, regardless of their topological entropy.

(2) In particular (see Eq. (25)),

$$Z^*_{\exp,1}(X_0^L) = S_{BGS}(p) = H^*(X_0^L), \tag{32}$$

where $H^*(X_0^L)$ is the conventional metric permutation entropy of order L, Eq. (5). In other words, $Z^*_{g,\alpha}(X_0^L)$ reduces to the conventional permutation entropy under the right assumptions. This justifies calling it a (generalized) permutation entropy.

(3) $Z^*_{\text{fac},\alpha}(X_0^L)$ was used in [20] (with the notation $Z^*_\alpha(X_0^L)$) to generalize $H^*(X_0^L)$ to FPF processes. There it is proved that

$$Z^*_{\text{fac},\alpha}(p) = R_\alpha(p) - \frac{1}{2}R_\alpha(p)^2 + O(3) \tag{33}$$

if $R_\alpha(p) < 1/e$. Therefore, when $R_\alpha(p)$ is small, it is a good approximation of $Z^*_{\text{fac},\alpha}(p)$.

As anticipated in Section 3, we have chosen $g_{\text{sub}}(t) = ct\ln t$ $(0 < c < 1)$ in Theorem 9(c) mainly because of Example 4. For the choice $g_{\text{sub}}(t) = t\ln^{(n)} t$ $(n \geq 2)$, the other examples of sub-factorial PC functions given in Eq. (21), we need to generalize the Lambert function $\mathcal{L}(x)$. We define the *generalized Lambert function* $\mathcal{L}^{(n)}(x)$ ($n \geq 1$, with $\mathcal{L}^{(1)}(x) = \mathcal{L}(x)$) by the functional equation

$$\mathcal{L}^{(n)}(x)\exp^{(n)}[\mathcal{L}^{(n)}(x)] = x \tag{34}$$

for $x \geq -\exp^{(n)}(-1)$, where $\exp^{(n)}(x)$ denotes the composition of the exponential function n times. Hence, $\mathcal{L}^{(n)}(x) \geq 0$ for $x \geq 0$, $\mathcal{L}^{(n)}(0) = 0$, and the identity (28) generalizes to

$$\mathcal{L}^{(n)}[x\ln^{(n)} x] = \ln^{(n)} x \tag{35}$$

for $x \geq \exp^{(n)}(-1)$ (since $\mathcal{L}^{(n)}[-\exp^{(n)}(-1)] = -1$). It follows that the inverse of $g(t) = t\ln^{(n)} t$ is

$$g^{-1}(s) = \exp^{(n)}[\mathcal{L}^{(n)}(s)], \tag{36}$$

so that $g^{-1}(0) = 1$ since $\mathcal{L}^{(n)}(0) = 0$. Therefore, the permutation entropy of order L of the sub-factorial class defined by $g(t) = t\ln^{(n)} t$ is

$$Z^*_{g,\alpha}(X_0^L) = \exp^{(n)}[\mathcal{L}^{(n)}(R_\alpha(p))] - 1 \quad (n \geq 2). \tag{37}$$

For $n = 1$ we recover $Z^*_{\text{fac},\alpha}(X_0^L)$, Eq. (30).

4.2. Permutation entropy rate

According to axiom SK2, entropies reach their maxima over uniform probability distributions. Sometimes these upper bounds are called the topological versions of the corresponding entropies or simply topological entropies. Thus, the topological version of $Z^*_{g,\alpha}(X_0^L)$ is its tight upper bound, which is obtained over the uniform distribution p_u of the *allowed* ordinal L-patterns for $\mathbf{X}$. This means that $p_u = (p_1, \ldots, p_{L!})$ with

$$p_i = \begin{cases} 1/\mathcal{A}_L(\mathbf{X}) & \text{if the } i\text{th ordinal } L\text{ -pattern is allowed for } \mathbf{X} \\ 0 & \text{if the } i\text{th ordinal } L\text{-pattern is forbidden for } \mathbf{X} \end{cases} \tag{38}$$

for $i = 1, \ldots, L!$. Note that

$$R_\alpha(p_u) = \ln \mathcal{A}_L(\mathbf{X}) \tag{39}$$

for $\alpha > 0$. Plugging Eq. (39) into (26), we are led to the following definition.

Definition 11. The *topological permutation entropy of order L* of a process $\mathbf{X}$ of class g is defined as

$$Z^*_{g,0}(X_0^L) \equiv Z^*_{g,0}(p_u) = g^{-1}(\ln \mathcal{A}_L(\mathbf{X})) - g^{-1}(0), \tag{40}$$

where p_u is the uniform probability distribution of allowed L -patterns for $\mathbf{X}$ as defined in Eq. (38).

The notation $Z^*_{g,0}$ for the topological permutation entropy is justified because $\ln \mathcal{A}_L(\mathbf{X})$ is formally obtained from Eq. (24) by setting $\alpha = 0$; indeed,

$$R_0(p_1, \ldots, p_{L!}) = \ln |\{p_i : p_i > 0,\ 1 \le i \le L!\ \}| = \ln \mathcal{A}_L(\mathbf{X}) \tag{41}$$

for all $p = \{p_1, \ldots, p_{L!}\}$. It follows,

$$R_0(p) \ge R_\alpha(p) \tag{42}$$

for all $\alpha > 0$, so that

$$Z^*_{g,0}(X_0^L) \ge Z^*_{g,\alpha}(X_0^L) \tag{43}$$

for all $\alpha > 0$ since $g^{-1}(s)$ is a strictly increasing function (see Remark 6).

Uniform probability distributions are special for several reasons. From the viewpoint of statistical mechanics, they correspond to the most disordered state, hence to equilibrium in the microcanonical ensemble. From the point of view of information theory, they amount to the principle of insufficient reason or maximum entropy principle [56] under null knowledge (maximum ignorance). Most important for us, the concept of extensivity (inherited from thermodynamics) also refers to such probability distributions: we say that an entropy $S(p)$ is *extensive* if it scales linearly with the number of constituents (degrees of freedom, etc.) N of the system over the uniform probability distribution $p = (1/N, \ldots, 1/N)$, i.e.,

$$\lim_{N\to\infty} \frac{S(\frac{1}{N}, \ldots, \frac{1}{N})}{N} \sim \ const > 0. \tag{44}$$

Therefore, extensivity depends on how the number of states grows with N, the sub-exponential, exponential and super-exponential regimes (or classes) being the most important ones.

Theorem 12. *The permutation entropy $Z^*_{g,\alpha}(X_0^L)$ is extensive with respect to the parameter L. In fact, for all $\alpha > 0$,*

$$\frac{Z^*_{g,\alpha}(p_u)}{L} = \frac{Z^*_{g,0}(X_0^L)}{L} \sim 1. \tag{45}$$

Proof. From definition (40) and $\ln \mathcal{A}_L(\mathbf{X}) \sim g(L)$, Eq. (17), we obtain

$$Z^*_{g,0}(X_0^L) = g^{-1}(\ln \mathcal{A}_L(\mathbf{X})) - g^{-1}(0) \sim g^{-1}(g(L)) - g^{-1}(0) = L - g^{-1}(0) \sim L.$$

This proves Eq. (45).

To get rid of the dependence of $Z^*_{g,\alpha}(X_0^L)$ on L, we turn to the entropy rates per variable, $Z^*_{g,\alpha}(X_0^L)/L$, and take the limit when $L \to \infty$.

Definition 13. The *permutation entropy rate (or just permutation entropy) of a process* $\mathbf{X}$ of class g is defined as

$$z^*_{g,\alpha}(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} Z^*_{g,\alpha}(X_0^L), \tag{46}$$

where $\alpha \ge 0$: $z^*_0(\mathbf{X})$ is the *topological* permutation entropy of $\boldsymbol{X}$, and $z^*_\alpha(\mathbf{X})$ with $\alpha > 0$ is the *metric* permutation entropy of $\mathbf{X}$.

The permutation entropy rate $z^*_{g,\alpha}(\mathbf{X})$ quantifies an intrinsic property of the process $\mathbf{X}$. The existence of the limit (46) follows from Theorem 12. As a matter of fact, the existence of $z^*_{g,\alpha}(\mathbf{X})$ amounts to the extensivity of $Z^*_{g,\alpha}(X_0^L)$. Note that

$$z^*_{g,0}(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} Z^*_{g,0}(X_0^L) = 1 \tag{47}$$

by Theorem 12, hence,

$$z^*_{g,\alpha}(\mathbf{X}) \le z^*_{g,0}(\mathbf{X}) = 1 \tag{48}$$

for each complexity class g and $\alpha > 0$, by Eq. (43).

Theorem 14. *Given a process* **X**, *let p be the probability distribution of the ordinal L-patterns of* X_0^L. *The permutation entropy rate of* **X** *is given as follows.*

(a) *For the exponential class:*

$$z^*_{exp,\alpha}(\mathbf{X}) := z^*_{g_{exp},\alpha}(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{cL} R_\alpha(p). \tag{49}$$

In particular, if $\mathbf{X} = f$ *and* $\alpha = 1$, *then*

$$z^*_{exp,1}(\mathbf{X}) = \frac{h(f)}{h_0(f)} \leq 1, \tag{50}$$

where $h(f)$ *is the Kolmogorov–Sinai entropy of* f *and* $h_0(f)$ *is its topological entropy.*

(b) *For the factorial class:*

$$z^*_{fac,\alpha}(\mathbf{X}) := z^*_{g_{fac},\alpha}(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} e^{\mathcal{L}[R_\alpha(p)]}. \tag{51}$$

(c) *For the sub-factorial class (defined by* $g_{sub}(t) = ct \ln t$, $0 < c < 1$*):*

$$z^*_{sub,\alpha}(\mathbf{X}) := z^*_{g_{sub},\alpha}(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} e^{\mathcal{L}[R_\alpha(p)/c]}. \tag{52}$$

Proof. Use the definition (46) and

$$Z^*_{g,\alpha}(X_0^L) \sim Z^*_{g_{\rm exp},\alpha}(X_0^L) = \frac{1}{c} R_\alpha(p)$$

for **X** in the exponential class (Theorem 9(a)),

$$Z^*_{g,\alpha}(X_0^L) \sim Z^*_{g_{\rm fac},\alpha}(X_0^L) = e^{\mathcal{L}[R_\alpha(p)]} - 1$$

for **X** in the factorial class (Theorem 9(b)), and

$$Z^*_{g,\alpha}(X_0^L) \sim Z^*_{g_{\rm sub},\alpha}(X_0^L) = e^{\mathcal{L}[R_\alpha(p)/c]} - 1$$

for **X** in the sub-factorial class defined by $g_{\rm sub}(t) = ct \ln t$, $0 < c < 1$ (Theorem 9(c)), to derive Eqs. (49), (51) and (52), respectively.

As for Eq. (50), use (i) $R_1(p) = S_{BGS}(p) = H^*(X_0^L)$, (ii)

$$\lim_{L\to\infty} \frac{1}{L} R_1(p) = \lim_{L\to\infty} \frac{1}{L} H^*(X_0^L) = h(f)$$

(Theorem 2(a)), and (iii) $g_{\rm exp}(t) = ct = h_0(f)t$ in Eq. (49). Moreover $h(f) \leq h_0(f)$.

Remark 15. A few closing observations on permutation entropy rates.

(1) The hierarchical order (27) and (43) carries on trivially to permutation entropy rates:

$$z^*_{g,\alpha}(\mathbf{X}) \geq z^*_{g,\beta}(\mathbf{X}) \tag{53}$$

for $0 \leq \alpha < \beta$.

(2) For the sub-factorial subclasses defined by $g(t) = t \ln^{(n)} t$, $n \geq 2$, use Eq. (37) to obtain

$$z^*_{g,\alpha}(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} \exp^{(n)}[\mathcal{L}^{(n)}[R_\alpha(p)]]. \tag{54}$$

(3) For white noise (WN), the probability distribution of the L-patterns is uniform for every L. Therefore,

$$z^*_{\rm fac,\alpha}({\rm WN}) = z^*_{\rm fac,0}({\rm WN}) = 1 \tag{55}$$

for every $\alpha > 0$ by Eq. (47).

5. Numerical simulations

One of the most important applications of permutation complexity to data analysis is the characterization and, hence, discrimination of time series. In this section we illustrate the discrimination power of permutation complexity in time series analysis. For this purpose, we are going to use permutation entropy in Section 5.1 and (perhaps surprisingly) the PC function in Section 5.2. Both tools are used in Section 5.3 to further dissect the sub-factorial "not-so-noisy measurement of a periodic signal" introduced in Example 4. Numerical receipts for computing ordinal patterns and permutation entropies can be found, e.g., in Refs. [57,58].

Since real time series analysis have finite length, some allowed ordinal patterns can be missing in random time series simply for statistical reasons. Therefore, practitioners prefer to speak of *visible patterns* and *missing patterns* rather than allowed patterns and forbidden patterns, respectively, as we will sometimes do as well.

5.1. Time series discrimination using permutation entropies

In Section 4 we have explicitly constructed a permutation entropy $Z^*_{g,\alpha}(X_0^L)$ for each PC class g such that the corresponding entropy rate $z_{g,\alpha}(\mathbf{X})$ is finite. However, real-world data is noisy, which seems to exclude the exponential class — but not quite.

In nonlinear time series analysis, it is good practice to test for determinism first. Underlying determinism in noisy time series can be unveiled by several techniques [59], including forbidden ordinal patterns [9]. If the noise to signal ratio is sufficiently small, then the data can be denoised, which allows the analyst to work with time series as good as noiseless deterministic. This is the exponential class, and the realm of the conventional permutation entropy $Z^*_{\exp,\alpha}(X_0^L) = R_\alpha(p)$ (or $H^*(X_0^L)$ for $\alpha = 1$, Eq. (32)) and its rate $z^*_{\exp,\alpha}(\mathbf{X})$. Since real-world time series are finite, the entropy rate $z^*_{\exp,\alpha}(\mathbf{X})$ can only be estimated if the convergence of $Z^*_{\exp,\alpha}(X_0^L)/L$ is sufficiently fast. This can be checked, e.g., by plotting $Z^*_{\exp,\alpha}(X_0^L)/L = R_\alpha(p)/L$ vs $1/L$; if there is an interval where the curve is linear (before undersampling sets in), then fit a straight line to the linear segment of the curve and the sought limit $z^*_{\exp,\alpha}(\mathbf{X}) = \lim_{L\to\infty} Z^*_{\exp,\alpha}(X_0^L)/L$ is the intercept of the straight line with the vertical axis [10, Sect. 2.1]. If desired, the parameter c that appears in Eq. (29) can be estimated by $H^*_0(X_0^L)/L = \ln \mathcal{A}_L(\mathbf{X})/L$ (see Eqs. (7) and (8)), because $H^*_0(X_0^L)/L$ is a proxy of $c = h_0(f)$ for L large enough by Theorem 2(b) with $f = \mathbf{X}$. For the purpose of time series discrimination, however, c can be dispensed with, which amounts to setting $c = 1$. The estimation of $\mathcal{A}_L(\mathbf{X})$ is usually done by just counting visible patterns in a sample of time series or even in a single, sufficiently long time series. This procedure can be justified if the orbits densely visit the state space, a property that goes by the name of transitivity. By the way, this is the first property in Devaney's definition of chaos and, in fact, it implies the other two properties (density of periodic points and sensitivity to initial conditions) for interval maps [60].

Furthermore, in nonlinear time series analysis, the dynamics of the (often unknown) system under observation is assumed to settle down on a low dimensional attractor, where it is transitive. However, for the asymptotic dynamics to be accessible to finite precision observations and numerical simulations, it is necessary that the physical measure (see Eq. (6)) is smooth (or absolutely continuous in technical terminology [44]). Typically, the physical measure of chaotic attractors has a smooth density in the stretching, or unstable, directions of the dynamics, while it has a discontinuous (e.g., Cantor set-like) structure transversally to those directions [42]; think of the Hénon attractor. Finite precision smooths out the physical measure when the attractor is viewed transversally to the stretching directions.

Otherwise, if there is no good reason to assume determinism, the data are handled as random. Moreover, numerical simulations and empirical observations show that virtually all random time series encountered in practice are FPF. This entails that $Z^*_{\mathrm{fac},\alpha}(X_0^L)$ and $z^*_{\mathrm{fac},\alpha}(\mathbf{X})$ are the appropriate tools to characterize random time series in the absence of more information. However, the factorial growth of the L-patterns and the associated computational cost restrict L to moderate values (say, $L \lesssim 7$) in practice, which makes the numerical estimation of $z^*_{\mathrm{fac},\alpha}(\mathbf{X})$ an open question in general.

For the above reasons, we have selected seven random processes from the factorial class to illustrate the discrimination power of permutation complexity with numerical simulations. This is also a particularly difficult case because all processes belong to the same complexity class in strict sense (there are no parameters in $g_{\mathrm{fac}}(t) = t \ln t$). Those seven processes are the following.

(Fac1) *White noise* (WN) in the form of an independent and uniformly distributed process on [0, 1];

(Fac2) *Fractional Gaussian noise* (fGn) with Hurst exponent $H = 0.2$ [43];

(Fac3-5) *Fractional Brownian motion* (fBm) with $H = 0.2$, 0.4 (anti-persistent processes), and $H = 0.6$ (persistent process) [43];

(Fac6) *Logistic map with additive white noise* of amplitude 0.30 (noisy LM), i.e., $x_t = y_t + z_t$, where $y_t = 4y_{t-1}(1 - y_{t-1})$ with $y_0 = 0.2002$, and $(z_t)_{t\geq 0}$ is WN with $-0.30 \leq z_t \leq 0.30$;

(Fac7) *Schuster map* with exponent 2 [61] and *additive white noise* of amplitude 0.25 (noisy SM), i.e., $x_t = y_t + z_t$, where $y_t = y_{t-1} + y_{t-1}^2 \bmod 1$ with $y_0 = 0.2002$, and $(z_t)_{t\geq 0}$ is WN with $-0.25 \leq z_t \leq 0.25$.

Arguments for this specific pick include that (i) the processes (Fac1)-(Fac7) cover a diversity of interesting cases (white noise, random processes with long dependence ranges, deterministic dynamics contaminated with observational noise); (ii) they are relevant to time series analysis and familiar to the analysts; and (iii) there are well-tested numerical routines available for simulations.

Fig. 1 shows $\langle Z^*_{\mathrm{fac},\alpha}(X_0^L)/L\rangle$, the average of $Z^*_{\mathrm{fac},\alpha}(X_0^L)/L$ over 35 time series for the random processes (Fac1)-(Fac7) and $3 \leq L \leq 7$, where $\alpha = 0.5$ (a), $\alpha = 1$ (b) and $\alpha = 1.5$ (c). For calculation purposes, the maximal length of the time series was set at $T_{\max} = 50{,}000$ ($\gg 7! = 5040$), but the computational loop is actually exited as soon as the probability distribution of the L -patterns stabilizes, so that the numerical routine is the same for all $3 \leq L \leq 7$. We see in all panels of Fig. 1 that $\langle Z^*_{\mathrm{fac},\alpha}(X_0^L)/L\rangle$ follows a distinct and seemingly convergent trajectory for each process as L grows, upper bounded by the white noise. In agreement with (27), $\langle Z^*_{\mathrm{fac},0.5}(X_0^L)/L\rangle \geq \langle Z^*_{\mathrm{fac},1}(X_0^L)/L\rangle \geq \langle Z^*_{\mathrm{fac},1.5}(X_0^L)/L\rangle$ for each process. For white noise, $\langle Z^*_{\mathrm{fac},\alpha}(X_0^L)/L\rangle \to 1$ as L grows by Eq. (55).

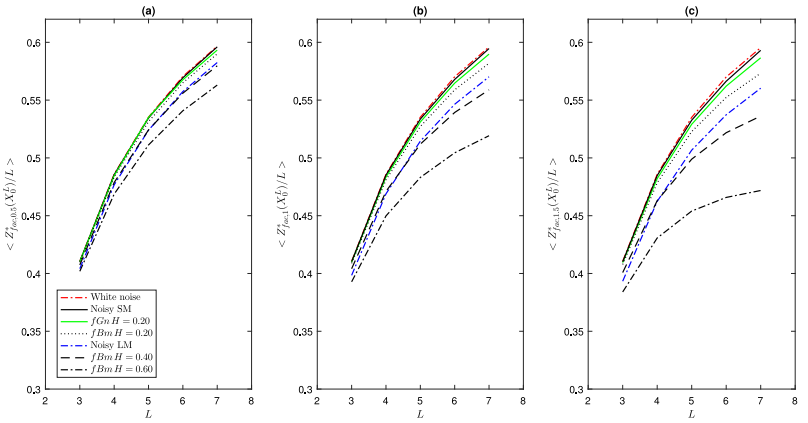

Fig. 1. The averages $\langle Z^*_{\mathrm{fac},\alpha}(X_0^L)/L\rangle$ of $Z^*_{\mathrm{fac},0.5}(X_0^L)/L$ (a), $Z^*_{\mathrm{fac},1}(X_0^L)/L$ (b) and $Z^*_{\mathrm{fac},1.5}(X_0^L)/L$ (c) over 35 realizations of the random processes listed in the inset are plotted vs L for $3 \le L \le 7$.

The effect of α on the discriminatory power of $Z_{\mathrm{fac},\alpha}(X_0^L)$ clearly depends on the probability distribution p of the L -patterns through the Rényi entropy $R_\alpha(p)$. In particular, for $\alpha < 1$ the central part of the distribution is flattened, i.e., high-probability events are suppressed, and low-probability events are enhanced. This effect is more pronounced for smaller α. The opposite happens when $\alpha > 1$: low-probability events are suppressed while high-probability events are enhanced. Our choices $\alpha = 0.5$, 1, 1.5 are meant to include both situations $\alpha < 1$ and $\alpha > 1$, along with the Shannonian case $\alpha = 1$. In Fig. 1, this results in a higher discriminatory power of $Z_{\mathrm{fac},\alpha}(X_0^L)$ with increasing α, i.e., the larger α the further apart the curves $\langle Z^*_{\mathrm{fac},\alpha}(X_0^L)/L\rangle$ are, which shows that the parameter α is an asset in applications. Similar choices of α give similar results (not shown).

Note also that the curves of different processes may cross. The reason is that $Z^*_{\mathrm{fac},\alpha}(X_0^L)$ can only capture ranges of interdependence up to L. Put another way, larger window sizes L unveil dependencies between farther variables that can be measured by $Z^*_{\mathrm{fac},\alpha}(X_0^L)$. Therefore, as L grows, $Z^*_{\mathrm{fac},\alpha}(X_0^L)$ can become larger for a noisy chaotic signal, such as the noisy logistic map, than for a process with a longer, or an infinite, span of interdependence between its increments, such as the fractional Brownian motion with $H = 0.40$; this can be see more clearly in panel (c).

In particular, Fig. 1 shows that, although all seven processes (Fac1)-(Fac7) belong to the same PC class, namely, the factorial class, the finite rates $Z^*_{\mathrm{fac},\alpha}(X_0^L)/L$ can distinguish them, evidencing as expected that $Z^*_{\mathrm{fac},\alpha}(X_0^L)$ is a finer measure of permutation complexity than $g(t)$. This does not mean that the growth rate of allowed patterns cannot be utilized for that objective, as we explain in Section 5.2.

5.2. Time series discrimination using permutation complexity functions

It was noticed in [15] that the number of missing L-patterns for a *white noise* series of length $T \ge L$ decreases exponentially with T. This result was generalized in [16] to different sorts of *FPF processes* (f^{-k} power spectrum (PS), fractional Brownian motion (fBm), fractional Gaussian noise (fGn)) by setting

$$\mathcal{M}_{L,T}(\mathbf{X}) := \left|\{\text{missing } L\text{-patterns in } x_0^{T-1}\}\right| = Ce^{-RT}, \tag{56}$$

where x_0^{T-1} is a typical realization of $\mathbf{X}$ of length $T \ge L$, and C and the *decay rate* R are constants that depend on L and the parameters of the random process (k for f^{-k} PS, or the Hurst exponent H for BM and fGn). From Eq. (56) and $\mathcal{M}_{L,L}(\mathbf{X}) = L! - 1$, it follows $C = (L! - 1)e^{RT}$, hence

$$\mathcal{M}_{L,T}(\mathbf{X}) = (L! - 1)e^{-R(T-L)}. \tag{57}$$

Clearly, the decay rate of missing patterns in random time series is not indifferent to the dependencies between the variables of the process, either due to an underlying functional dependence (noisy deterministic signal) or to a statistical correlation.

This being the case, we let the permutation complexity function g depend on T too, i.e., $g = g(L, T)$, and generalize Eq. (17) to

$$\ln \mathcal{A}_{L,T}(\mathbf{X}) := \ln\left|\{\text{visible } L\text{ -patterns in } x_0^{T-1}\}\right| = g(L, T) \tag{58}$$

so that $g(L) \sim g(L, T)$, where now $L \gg 1$ implies $T \gg 1$ because $T \ge L$. Therefore,

$$\mathcal{A}_{L,T}(\mathbf{X}) + \mathcal{M}_{L,T}(\mathbf{X}) = L! \tag{59}$$

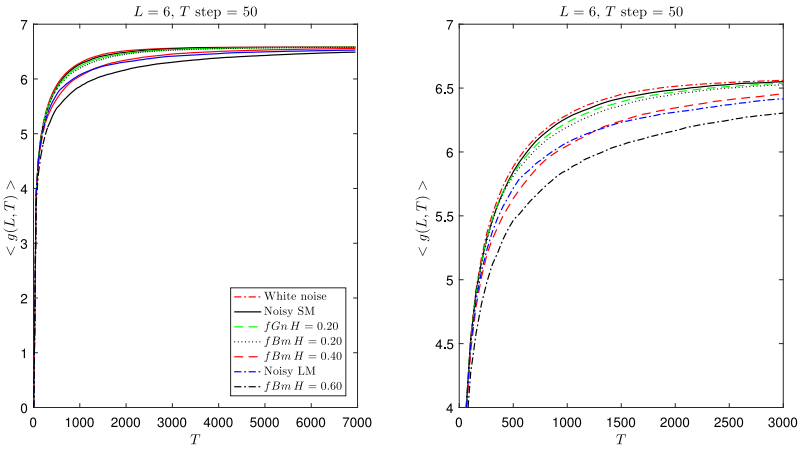

Fig. 2. The averages $\langle g(L,T)\rangle$ of the permutation complexity function $g(L,T)$ over 35 realizations of the random processes listed in the inset (the same as in Fig. 1) are plotted vs T (the time series length) for $L = 6$. The right panel ($6 \le T \le 3000$) is a zoom of the left panel ($6 \le T \le 7000$).

Table 1
The decay exponent R in Eqs. (61)–(62) for $L = 4, 5, 6$, $T = 7000$, and the random processes plotted in Fig. 2.

Series	$L = 4$	$L = 5$	$L = 6$
White Noise	4.43×10^{-2}	8.47×10^{-3}	1.40×10^{-3}
$fGnH = 0.20$	4.93×10^{-2}	8.17×10^{-3}	1.21×10^{-3}
$fBmH = 0.20$	4.20×10^{-2}	7.52×10^{-3}	1.13×10^{-3}
$fBmH = 0.40$	3.76×10^{-2}	6.32×10^{-3}	8.12×10^{-4}
$fBmH = 0.60$	3.24×10^{-2}	4.07×10^{-3}	5.05×10^{-4}
Noisy LM	3.36×10^{-2}	5.55×10^{-3}	7.54×10^{-4}
Noisy SM	4.43×10^{-2}	8.09×10^{-3}	1.30×10^{-3}

for every $2 \le L \le T$. Since there are two parameters L and T, we can fix $L\ (= L_0)$ and vary T so that the allowed patterns have a chance to become visible. Correspondingly, we can distinguish (i) a *transient phase* (T "small"), where the allowed patterns become progressively visible, and (ii) a *stationary phase* (T "large"), where all the allowed patterns are visible. Generally speaking, the transient phase discriminates random signals while the stationary phase discriminates deterministic signals, so both phases complement each other in the analysis of permutation complexity.

Indeed, in the deterministic case $\mathbf{X} = f$,

$$g(L_0,T) = \ln \mathcal{A}_{L_0,T}(\mathbf{X}) \nearrow \ln \mathcal{A}_{L_0}(\mathbf{X}) = h^*_{0,L_0}(f) L_0 \tag{60}$$

as T increases, see Eqs. (7) and (8). Here, the symbol $\nearrow$ means that the convergence is monotone increasing. The limit is achieved in finite time, namely, once all allowed L_0-patterns are visible. For FPF processes, though,

$$\mathcal{A}_{L,T}(\mathbf{X}) = L! - \mathcal{M}_{L,T}(\mathbf{X}) \simeq L! \left(1 - e^{-R(T-L)}\right) \tag{61}$$

for every L by Eqs. (57) and (59). Instead of the limit (60) for $\mathbf{X} = f$, for FPF processes we have

$$g(L_0,T) \simeq \ln L_0! + \ln\left(1 - e^{-R(T-L_0)}\right) \nearrow \ln L_0! \tag{62}$$

as T increases and it remains constant once all L_0-patterns are visible. Therefore, contrarily to the deterministic case (60), the limit of $g(L_0,T)$ as T grows is the same for all FPF processes, namely, $\ln L_0!$.

A direct application of the "finite length" PC function $g(L,T)$ is to discriminate different FPF processes by the different convergence rates of $g(L,T)$ to $\ln L!$ as T increases. Numerical evidence is shown in Fig. 2 for $L = 6$ and $6 \le T \le 7000$ in the left panel; the right panel is a zoom of the left panel with $6 \le T \le 3000$. Here we used the same 35 time series, random processes (Fac1)-(Fac7) and numerical results as in Fig. 1, and plotted $\langle g(6,T)\rangle = \left\langle \ln \mathcal{A}_{6,T}(\mathbf{X})\right\rangle$ every $\Delta T = 50$ points, where $\langle\cdot\rangle$ denotes again the average over the 35 samples. For all those processes (and any other FPF process for that matter), $\langle g(6,T)\rangle$ converges to $\ln 6! \simeq 6.579\,3$ as T grows.

The decay exponents R in the approximation (62) are listed in Table 1 for $L = 4, 5, 6$, $T = 7000$ and the random processes (Fac1)-(Fac7) shown in Fig. 2. Due to the different correlation lengths, the curves $g(L_0,T)$ may intersect as a result of the different decay rates of the missing patterns.

Let us finally mention that Eq. (56) has recently been generalized to [31,32]

$$\mathcal{M}_{L,N}(\mathbf{X}) = C \exp\left(-RN^{\beta}\right). \tag{63}$$

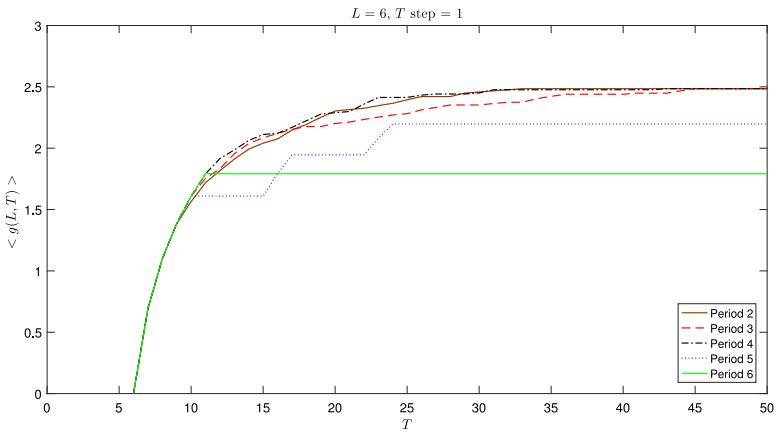

Fig. 3. The averages $\langle g(6,T)\rangle$ of the permutation complexity function $g(6,T)$ over 35 realizations of the sub-factorial processes $\mathbf{X}_\mathfrak{p}$ of periods $2 \le \mathfrak{p} \le 6$ are plotted vs T (the time series length) for $6 \le T \le 50$.

The *stretching exponent* $0 < \beta \le 1$ depends on L as well as on the underlying random process; for $\beta = 1$ one recovers (56).

5.3. The sub-factorial class

To wrap up this numerical section, we consider the sub-factorial class as well. To this end, we revisit $\mathbf{X}_\mathfrak{p} = (X_t)_{t\ge 0}$, the "not-so-noisy measurement of a periodic signal" of period $\mathfrak{p} \ge 2$ introduced in Example 4, since this is the only instance of a sub-factorial process that we know of. As compared to the factorial processes studied in Sections 5.1 and 5.2 , $\mathbf{X}_\mathfrak{p}$ has two peculiarities: (i) its conceptual simplicity makes most details amenable to analytical scrutiny, and (ii) it is a cyclostationary process, i.e., its statistical properties vary periodically with time, so that it can be viewed as a random process composed of $\mathfrak{p}$ interleaved stationary processes $(X_{\nu\mathfrak{p}+\mu})_{\nu\ge 0}$, $0 \le \mu \le \mathfrak{p}-1$. Likewise, it turns out that the ordinal representations of $\mathbf{X}_\mathfrak{p}$ decompose into $\mathfrak{p}$ sequences of representations by means of $(\nu\mathfrak{p}+\mu)$-patterns, $\nu \in \mathbb{N}$, such that $\mathbf{X}_\mathfrak{p}$ has a PC function $g(L)$ for each of those sequences.

We start our numerical analysis of $\mathbf{X}_\mathfrak{p}$ with the finite length PC function $g(L,T)$ for $L = 6$. Fig. 3 depicts $g(6,T)$ vs T for $6 \le T \le 50$ and $2 \le \mathfrak{p} \le 6$. Here we already recognize that these processes are not factorial because $g(6,T) < 2.5$ in all cases, whereas $g(6,T) \nearrow \ln 6! \simeq 6.58$ for factorial processes (see Section 5.2). Since they are not deterministic either (which can be simply checked by a return map), then they must be sub-factorial. The curves $g(6,T)$ in Fig. 3 are visibly distinct, although the ones corresponding to the periods $\mathfrak{p} = 2,3,4$ run close to each other. The reason for this is that $g(6,50) = \ln \mathcal{A}_6(\mathbf{X}_\mathfrak{p})$, where, according to Table 2, $\ln \mathcal{A}_6(\mathbf{X}_\mathfrak{p}) = \ln 12 \simeq 2.485$ for $\mathfrak{p} = 2,3,4$, while $\ln \mathcal{A}_6(\mathbf{X}_5) = \ln 9 \simeq 2.197$ and $\ln \mathcal{A}_6(\mathbf{X}_6) = \ln 6 \simeq 1.792$. Also, contrarily to Fig. 2 for the factorial processes (Fac1)-(Fac7), the curves $g(6,T)$ in Fig. 3 do not follow the exponential ansatz (62) but rather oscillate before leveling off. This is due to the periodicity of the probability distributions of the L-patterns (in particular, of the 6-patterns) as the window $x_t, x_{t+1}, \ldots, x_{t+5}$ slides from $t = 1$ to $t = T-5$.

According to Eq. (31), to compute the generalized permutation entropy $Z^*_{\mathrm{sub},\alpha}(X_0^L)$ of a sub-factorial process $\mathbf{X} = (X_t)_{t\ge 0}$ with PC function $g(L) = cL\ln L$, $0 < c < 1$, we need to know the constant c. With this objective, let $L = \nu\mathfrak{p}+\mu$, where $\nu = \lfloor L/\mathfrak{p} \rfloor \in \mathbb{N}$ and $\mu = L \bmod \mathfrak{p} \in \{0,1,\ldots,\mathfrak{p}-1\}$; $\mu = 0$ (i.e., $L = \nu\mathfrak{p}$) is the case considered in Example 4. An argument similar to the one used there shows that Eq. (13) generalizes to

$$\mathcal{A}_L(\mathbf{X}_\mathfrak{p}) = N_1(\nu,\mu) + N_2(\nu,\mu) \tag{64}$$

where

$$N_1(\nu,\mu) = (\mathfrak{p}-\mu)[(\nu+1)!]^\mu[\nu!]^{\mathfrak{p}-\mu-1},\ N_2(\nu,\mu) = \mu[(\nu+1)!]^{\mu-1}[\nu!]^{\mathfrak{p}-\mu}. \tag{65}$$

Note that $N_2(\nu,0) = 0$ for all ν, so that we recover Eq. (13),

$$\mathcal{A}_{\nu\mathfrak{p}}(\mathbf{X}_\mathfrak{p}) = N_1(\nu,0) = \mathfrak{p}(\nu!)^{\mathfrak{p}-1}, \tag{66}$$

from Eq. (64). Table 2 summarizes $\mathcal{A}_L(\mathbf{X}_\mathfrak{p})$ for moderate values of $\mathfrak{p}$ and L.

Therefore,

$$\begin{aligned}\mathcal{A}_L(\mathbf{X}_\mathfrak{p}) &= [(\nu+1)!]^{\mu-1}[\nu!]^{\mathfrak{p}-\mu-1}\,[(\mathfrak{p}-\mu)(\nu+1)! + \mu\nu!] \\ &= [(\nu+1)!]^{\mu-1}[\nu!]^{\mathfrak{p}-\mu-1}(\mathfrak{p}-\mu)(\nu+1)!\left[1 + \tfrac{\mu}{(\mathfrak{p}-\mu)(\nu+1)}\right]\end{aligned} \tag{67}$$

Table 2
The number of allowed L-patterns $\mathcal{A}_L(\mathbf{X}_\mathfrak{p})$ for periods $2 \le \mathfrak{p} \le 6$ and $\mathfrak{p} \le L \le 14$.

Allowed patterns													
$\mathfrak{p}$	Window width (L)												
	2	3	4	5	6	7	8	9	10	11	12	13	14
2	2	3	4	8	12	30	48	144	240	840	1,440	5,760	10,080
3	–	3	5	8	12	28	60	108	324	864	1,728	6,336	20,160
4	–	–	4	7	12	20	32	80	192	432	864	2,808	8,640
5	–	–	–	5	9	16	28	48	80	208	528	1,296	3,024
6	–	–	–	–	6	11	20	36	64	112	192	512	1,344

and

$$\ln \mathcal{A}_L(\mathbf{X}_\mathfrak{p}) = (\mu-1)\ln(\nu+1)! + (\mathfrak{p}-\mu-1)\ln \nu! + \ln(\mathfrak{p}-\mu) + \ln(\nu+1)! + \ln\left[1+\tfrac{\mu}{(\mathfrak{p}-\mu)(\nu+1)}\right]. \tag{68}$$

Use now Stirling's formula (11) to derive

$$\begin{aligned}\ln \mathcal{A}_L(\mathbf{X}_\mathfrak{p}) &\sim (\mu-1)(\nu+1)\ln(\nu+1) + (\mathfrak{p}-\mu-1)\nu\ln\nu + (\nu+1)\ln(\nu+1)\\ &= \mu(\nu+1)\ln(\nu+1) + (\mathfrak{p}-\mu-1)\nu\ln\nu\end{aligned} \tag{69}$$

when $\nu \to \infty$ (μ is bounded by $\mathfrak{p}-1$). Eq. (69) leads to the following two cases.

(1) If $\mu = 0$, i.e., $L = \nu\mathfrak{p}$, then

$$\ln \mathcal{A}_{\nu\mathfrak{p}}(\mathbf{X}_\mathfrak{p}) \sim (\mathfrak{p}-1)\nu\ln\nu = (\mathfrak{p}-1)\tfrac{L}{\mathfrak{p}}\ln\tfrac{L}{\mathfrak{p}} \sim \tfrac{\mathfrak{p}-1}{\mathfrak{p}} L\ln L. \tag{70}$$

This means that

$$g(L=\nu\mathfrak{p}) = cL\ln L \text{ with } c = \tfrac{\mathfrak{p}-1}{\mathfrak{p}}, \tag{71}$$

in accordance with Example 4. Numerical simulations confirm that, as expected, the $N_1(\nu,\mu)$ allowed L-patterns for $\mathbf{X}_\mathfrak{p}$ are equiprobable, hence

$$R_\alpha(p_u) = R_0(p_u) = \ln N_1(L/\mathfrak{p}, 0) = (\mathfrak{p}-1)\ln[\mathfrak{p}(L/\mathfrak{p})!] \tag{72}$$

for all $\alpha > 0$, see Eq. (66).

(2) If $1 \le \mu \le \mathfrak{p}-1$, i.e., $L = \nu\mathfrak{p}+\mu$ with $\mu \ne 0$, then

$$\begin{aligned}\ln \mathcal{A}_{\nu\mathfrak{p}+\mu}(\mathbf{X}_\mathfrak{p}) &\sim \mu(\nu+1)\ln(\nu+1) = \mu\left(\tfrac{L-\mu}{\mathfrak{p}}+1\right)\ln\left(\tfrac{L-\mu}{\mathfrak{p}}+1\right)\\ &\sim \mu\tfrac{L}{\mathfrak{p}}\ln\tfrac{L}{\mathfrak{p}} \sim \tfrac{\mu}{\mathfrak{p}} L\ln L.\end{aligned} \tag{73}$$

This means that

$$g(L=\nu\mathfrak{p}+\mu) = cL\ln L \text{ with } c = \tfrac{\mu}{\mathfrak{p}} \in \left\{\tfrac{1}{\mathfrak{p}}, \tfrac{2}{\mathfrak{p}}, \ldots, \tfrac{\mathfrak{p}-1}{\mathfrak{p}}\right\}. \tag{74}$$

Numerical simulations and theoretical insight show that, in this case, the probability distribution of the allowed L-patterns for $\mathbf{X}_\mathfrak{p}$ is composed of $N_1(\nu,\mu)$ L-patterns of probability $P_1 = (\mathfrak{p}-\mu)/\mathfrak{p}N_1$, and $N_2(\nu,\mu)$ L -patterns of probability $P_2 = \mu/\mathfrak{p}N_2$. It follows,

$$R_1(p) = -N_1P_1\ln P_1 - N_2P_2\ln P_2, \quad \text{and } R_\alpha(p) = \tfrac{1}{1-\alpha}\ln\left(N_1P_1^\alpha + N_2P_2^\alpha\right) \tag{75}$$

for $\alpha > 0$, $\alpha \ne 1$, where p is the probability distribution of the allowed ordinal L-patterns for $\mathbf{X}_\mathfrak{p}$.

From Eqs. (71) and (74) we obtain

$$c = \tfrac{\mathfrak{p}-1}{\mathfrak{p}} \quad \text{for } L = 0, \mathfrak{p}-1 \bmod \mathfrak{p}. \tag{76}$$

In particular, $c = 1/2$ for $\mathfrak{p} = 2$ and $\mu = 0, 1$, i.e., the process $\mathbf{X}_2$ has a unique PC function $g(L) = (1/2)L\ln L$, regardless of whether we use ordinal patterns of even or odd lengths.

In conclusion, $\mathbf{X}_\mathfrak{p}$ generates $\mathfrak{p}$ sub-factorial processes $\mathbf{X}_{\mathfrak{p},\mu}$, $0 \le \mu \le \mathfrak{p}-1$, according to the ordinal representations used to discretize the realizations of $\mathbf{X}_\mathfrak{p}$. Thus, if we use ordinal patterns of lengths $L = \nu\mathfrak{p}$, $\nu \in \mathbb{N}$, the result is a sub-factorial process $\mathbf{X}_{\mathfrak{p},0}$ with

$$\ln \mathcal{A}_{L=\nu\mathfrak{p}}(\mathbf{X}_{\mathfrak{p},0}) \sim \tfrac{\mathfrak{p}-1}{\mathfrak{p}} L\ln L. \tag{77}$$

Otherwise, if we use ordinal patterns of lengths $L = \nu\mathfrak{p}+\mu$, with $\nu \in \mathbb{N}$ and $1 \le \mu \le \mathfrak{p}-1$ fixed, the result is a sub-factorial process $\mathbf{X}_{\mathfrak{p},\mu}$ with

$$\ln \mathcal{A}_{L=\nu\mathfrak{p}+\mu}(\mathbf{X}_{\mathfrak{p},\mu}) \sim \tfrac{\mu}{\mathfrak{p}} L\ln L. \tag{78}$$

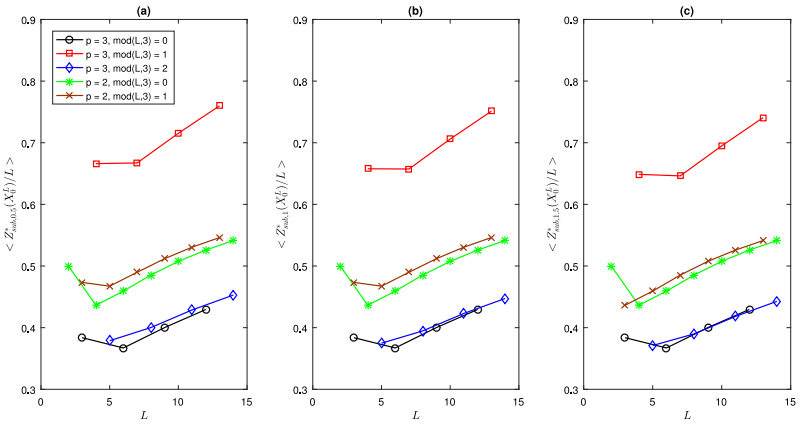

Fig. 4. The averages $\langle Z^*_{\mathrm{sub},\alpha}(X_0^L)/L\rangle$ of $Z^*_{\mathrm{sub},0.5}(X_0^L)/L$ (a), $Z^*_{\mathrm{sub},1}(X_0^L)/L$ (b) and $Z^*_{\mathrm{sub},1}(X_0^L)/L$ (c) over 35 realizations of the sub-factorial processes $\mathbf{X}_{2,0}$, $\mathbf{X}_{2,1}$, $\mathbf{X}_{3,0}$, $\mathbf{X}_{3,1}$ and $\mathbf{X}_{3,2}$ are plotted vs L for $2 \le L \le 14$. By definition, each process $\mathbf{X}_{\mathfrak{p},\mu}$, $0 \le \mu \le \mathfrak{p}-1$, uses ordinal representations $\mathcal{S}_L$ of lengths $L = \nu\mathfrak{p} + \mu$, $\nu \in \mathbb{N}$, to calculate the corresponding complexity function $g(L) = cL\ln L$. Here, $c = 1/2$ for $\mathbf{X}_{2,0}$ and $\mathbf{X}_{2,1}$, $c = 2/3$ for $\mathbf{X}_{3,0}$ and $\mathbf{X}_{3,2}$, and $c = 1/3$ for $\mathbf{X}_{3,1}$.

More formally, we say that the cyclostationary process $\mathbf{X}_{\mathfrak{p}}$ generates the processes $(\mathbf{X}_{\mathfrak{p}}, \{\mathcal{S}_{\nu\mathfrak{p}+\mu}\}_{\nu\ge 1}) =: \mathbf{X}_{\mathfrak{p},\mu}$, $0 \le \mu \le \mathfrak{p}-1$, where $\{\mathcal{S}_{\nu\mathfrak{p}+\mu}\}_{\nu\ge 1}$ is the subsequence of ordinal representations used to compute the PC function $g(L)$. In our case, the choices $\{\mathcal{S}_{\nu\mathfrak{p}+\mu}\}_{\nu\ge 1}$ with μ fixed are justified because then the corresponding PC function exists.

Fig. 4 shows the average $\langle Z^*_{\mathrm{sub},\alpha}(X_0^L)/L\rangle$ of the permutation entropy $Z^*_{\mathrm{sub},\alpha}(X_0^L)/L = (e^{\mathcal{L}[R_\alpha(p)/c]} - 1)/L$ over 35 realizations of the sub-factorial processes $\mathbf{X}_{2,\mu}$, $\mathbf{X}_{3,\mu}$ and $2 \le L \le 14$, where $\alpha = 0.5$ (a), $\alpha = 1$ (b) and $\alpha = 1.5$ (c). According to Eqs. (71) and (74), $c = 1/2$ for $\mathbf{X}_{2,0}$ and $\mathbf{X}_{2,1}$, $c = 2/3$ for $\mathbf{X}_{3,0}$ and $\mathbf{X}_{3,2}$, while $c = 1/3$ for $\mathbf{X}_{3,1}$. At variance with Fig. 1, the curves in Fig. 4 are further apart the lower α is. In particular, we see overlaps of the processes $\mathbf{X}_{3,0}$ and $\mathbf{X}_{3,2}$ in panels (b) and (c) that, however, are resolved in panel (a). This again illustrates how the parameter α can help when it comes to applications.

6. Conclusion

Permutation entropy is a popular tool for characterizing time series that depends on the size of the sliding window used to define the permutations its name refers to. If the data has been output by a dynamical system, the conventional permutation entropy rate (4) converges with increasing window sizes to the Kolmogorov–Sinai entropy of the dynamics (Theorem 2(a)). But if the process is noisy deterministic or random, then that entropy rate diverges in general because the number of visible ordinal patterns (permutations) can grow super-exponentially, see Eq. (12). This different growth behavior of the ordinal patterns and, hence, of the permutation complexity makes possible to distinguish (noiseless) deterministic processes from noisy processes but poses a challenge for a unified formulation of permutation complexity and its measurement across the entire range of processes.

In view of this fact, the objective of this paper was to propose such a unified formulation. Our approach consisted of two steps. First, we introduced the permutation complexity (PC) function $g(t)$ in Section 3. The asymptotic behavior of $g(t)$ defines the exponential, sub-factorial and factorial PC classes, the latter being the most interesting in the application of the ordinal methodology to real-world time series. A "finite-length" version of the PC function, $g(L, T)$, where T is the length of a time series and $L \ll T$ is the length of the ordinal patterns, was used in the numerical simulations, Section 5.2, to discriminate random processes via its convergence rate to $\ln L!$ with satisfactory results (see Fig. 2).

Second, we borrowed the concept of Z-entropy from statistical mechanics and complexity theory to generalize permutation entropy from the exponential PC class (the realm of conventional permutation entropy) to the sub-factorial and factorial PC classes. For this reason, the generalized permutation entropies go by the name $Z^*_{g,\alpha}$ (Eq. (26)), where the PC function $g(t)$ defines the corresponding class. Z-entropies are group entropies [18,19] that are designed to be extensive on the various complexity classes, including classes not considered here such as the sub-exponential and the super-factorial. Precisely, the extensivity of the Z-entropy entails in our context that the rate of the generalized permutation entropy, $z^*_{g,\alpha}$ (Eq. (46)), converges for the processes in the sub-factorial class and, foremost, in the factorial class. The discriminatory power of $Z^*_{g,\alpha}$ was numerically tested in Section 5.1 with seven processes belonging to the factorial class, i.e., $g(t) = t\ln t$. The results, shown in Fig. 1, were satisfactory as well. For completion we also analyzed in Section 5.3 the particularities of a toy model for sub-factorial processes introduced in Example 4. This model is of limited practical interest but it has the virtue of revealing some subtleties such as the role of cyclostationarity.

In conclusion, we have presented in this paper an integrating approach to the study and characterization of real-valued processes, whether deterministic or random, in the ordinal representation. Regarding the methodology of this approach, the processes are sorted into the exponential, sub-factorial and factorial complexity classes. Regarding the tools, the

permutation complexity of the processes in each class is measured by the corresponding permutation entropy, which is the Z-entropy of that class. The result is a "class-wise" generalization of conventional permutation entropy, one per class, whose rate also converges in the sub-factorial and factorial classes. These entropic measures of permutation complexity have both fine theoretical properties and potential in practical applications, in addition to closing the conceptual gap between deterministic and noisy signals in the ordinal analysis of time series.

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.

Acknowledgments

We thank our reviewers for their helpful comments. J.M.A. and R.D. were financially supported by Agencia Estatal de Investigación, Spain, grant PID2019-108654GB-I00. J.M.A. was also supported by Generalitat Valenciana, Spain, grant PROMETEO/2021/063. The research of P.T. has been supported by the research project PGC2018-094898-B-I00, Ministerio de Ciencia, Innovación *y* Universidades, Spain, and by the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000904-S), Ministerio de Ciencia, Innovación *y* Universidades, Spain. P.T. is a member of the Gruppo Nazionale di Fisica Matematica (INDAM), Italy.

References

[1] Morse M, Hedlund GA. Symbolic dynamics II: Sturmian trajectories. Amer J Math 1940;62:1–42.
[2] Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans Inform Theory 1976;22:75–81.
[3] Ziv J, Lempel A. Compression of individual sequences via variable-rate coding. IEEE Trans Inform Theory 1978;24:530–6.
[4] Amigó JM, Keller K, Unakafova V. On entropy, entropy-like quantities and applications. Discrete Contin. Dyn. Syst. B 2015;20:3301–44.
[5] Li M, Vitányi P. An introduction to kolmogorov complexity and its applications. New York: Springer Verlag; 2008.
[6] Volchan SB. What is a random sequence? Amer Math Monthly 2002;109:46–63.
[7] Shen A, Uspensky VA, Vereshchagin N. Kolmogorov complexity and algorithmic randomness. Rhode Island: American Mathematical Society; 2017.
[8] Downey R, Hirschfeldt DR. Computability and randomness, Vol. 66. Notices of the American Mathematical Society; 2019, p. 1001–12.
[9] Amigó JM. Permutation complexity in dynamical systems —Ordinal patterns, permutation entropy and all that. Berlin: Springer Verlag; 2010.
[10] Amigó JM, Zambrano S, Sanjuán MAF. Permutation complexity of spatiotemporal dynamics. Europhys Lett 2010;90:10007, 2010.
[11] Monetti R, Amigó JM, Aschenbrenner T, Bunk W. Permutation complexity of interacting dynamical systems. Eur Phys J Spec Top 2013;222:421–36.
[12] Bandt C, Pompe B. Permutation entropy: A natural complexity measure for time series. Phys Rev Lett 2002;88:174102.
[13] Rosso OA, Larrondo HA, Martin MT, Plastino A, Fuentes MA. Distinguishing noise from chaos. Phys Rev Lett 2007;99:154102.
[14] Zunino L, Soriano MC, Rosso OA. Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. Phys Rev E 2012;86:046210.
[15] Amigó JM, Zambrano S, Sanjuán MAF. True and false forbidden patterns in deterministic and random dynamics. Europhys Lett 2007;79:50001.
[16] Carpi LC, Saco PM, Rosso OA. Missing ordinal patterns in correlated noises. Physica A 2010;389:2020–9.
[17] Pessa AAB, Ribeiro HV. Characterizing stochastic time series with ordinal networks. Phys Rev E 2019;100:042304.
[18] Tempesta P, Jensen H. Universality classes and information-theoretic measures of complexity via group entropies. Sci Rep 2020;10:1–11.
[19] Tempesta P. Multivariate group entropies, super-exponentially growing systems and functional equations. Chaos 2020;30:123119.
[20] Amigó JM, Dale R, Tempesta P. A generalized permutation entropy for noisy dynamics and random processes. Chaos 2021;31:013115.
[21] Tempesta P. Group entropies, correlation laws and zeta functions. Phys Rev E 2011;84:021121.
[22] Tempesta P. Formal groups and Z-entropies. Proc R Soc Lond Ser A Math Phys Eng Sci 2016;472:20160143.
[23] Jensen HJ, Tempesta P. Group entropies: From phase space geometry to entropy functionals via group theory. Entropy 2018;20:804.
[24] Jensen HJ, Pazuki RH, Pruessner G, Tempesta P. Statistical mechanics of exploding phase spaces: Ontic open systems. J. Phys. A Math. Theor. 2018;51:375002.
[25] Rodríguez MA, Romaniega A, Tempesta P. A new class of entropic information measures, formal group theory and information geometry. Proc R Soc Lond Ser A Math Phys Eng Sci 2019;475:20180633.
[26] Keller K, Wittfeld K. Distances of time series components by means of symbolic dynamics. Int J Bifurcation Chaos 2004;14:693–703.
[27] Parlitz U, Berg S, Luther S, Schirdewan A, Kurths J, Wessel N. Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamic. Comput Biol Med 2012;42:319–27.
[28] Graff G, Graff B, Kaczkowska A, Makowiecz D, Amigó JM, Piskorski J, et al. Ordinal pattern stastistics for the assessment of heart rate variability. Eur Phys J Spec Top 2013;222:525–34.
[29] Amigó JM, Keller K, Unakafova V. Ordinal symbolic analysis and its applications to biomedical recordings. Phil Trans R Soc A 2015;373:20140091.
[30] Cao Y, Tung W, Gao JB, Protopopescu VA, Hively LM. Detecting dynamical changes in time series using the permutation entropy. Phys Rev E 2004;70:046217.
[31] Olivares F, Zunino L, Pérez DG. Revisiting the decay of missing ordinal patterns in long-term correlated time series. Physica A 2019;534:122100.
[32] Olivares F, Zunino L, Soriano MC, Pé rez DG. Unraveling the decay of the number of unobserved ordinal patterns in noisy chaotic dynamics. Phys Rev E 2019;100:042215.
[33] Zunino L, Ribeiro HV. Discriminating image textures with the multiscale two-dimensional complexity-entropy causality plane. Chaos Solitons Fractals 2016;91:676–88.
[34] Chagas ETC, Frery AC, Rosso OA, Ramos HS. Analysis and classification of SAR textures using information theory. IEEE J. Sel. Top. Appl. Obs. Remote Sens. 2021;14:663–75.
[35] Amigó JM, Keller K, Kurths J. Recent progress in symbolic dynamics and permutation complexity. Eur Phys J Spec Top 2013;222:241–598.
[36] Zanin M, Zunino L, Rosso OA, Papo D. Permutation entropy and its main biomedical and econophysics applications: A review. Entropy 2012;14:1553–77.

[37] Bandt C, Keller G, Pompe B. Entropy of interval maps via permutations. Nonlinearity 2002;15:1595–602.
[38] Keller K, Sinn M. Kolmogorov sinai entropy from the ordinal viewpoint. Physica D 2010;239:997–1000.
[39] Amigó JM. The equality of Kolmogorov–sinai entropy and metric permutation entropy generalized. Physica D 2012;241:789–93.
[40] Gutjahr T, Keller K. Equality of Kolmogorov–sinai and permutation entropy for one-dimensional maps consisting of countably many monotone parts. Discrete Contin Dyn Syst 2019;39:4207–24.
[41] Bandt C, Shiha F. Order patterns and time series. J Time Series Anal 2007;28:646–65.
[42] Eckmann JP, Ruelle D. Ergodic theory of chaos and strange attractors. Rev Modern Phys 1985;57:617–56.
[43] Mandelbrot BB, Van Ness JW. Fractional brownian motions, fractional noises and applications. SIAM Rev 1968;10:422–37.
[44] Walters P. An introduction to ergodic theory. New York: Springer Verlag; 2000.
[45] Amigó JM, Keller K. Permutation entropy: One concept, two approaches. Eur Phys J Spec Top 2013;222:263–74.
[46] Amigó JM, Kennel MB. Forbidden ordinal patterns in higher dimensional dynamics. Physica D 2008;237:2893–9.
[47] Apostol T. Mathematical analysis. second ed.. Menlo Park CA: Addison Wesley Longman; 1974.
[48] Shannon CE. A mathematical theory of communication. Bell Syst. Tech. J. 1948;27:379–423, 27, 623–653 (1948).
[49] Shannon CE, Weaver W. The mathematical theory of communication. Urbana, IL: University of Illinois Press; 1949.
[50] Khinchin AI. Mathematical foundations of information theory. New York: Dover; 1957.
[51] Amigó JM, Balogh SG, Hernández S. A brief review of generalized entropies. Entropy 2018;20:813, (21 pages).
[52] Ilić VM, Korbel J, Gupta S, Scarfone AM. An overview of generalized entropic forms. Europhys Lett 2021;133:50005.
[53] Tsallis C. Introduction to nonextensive statistical mechanics. New York: Springer; 2009.
[54] Rényi A. On measures of information and entropy, in: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, (1960) 547-561.
[55] Olver FWJ, Lozier DW, Boisvert RF, Clark CW, editors. NIST handbook of mathematical functions. Cambridge UK: Cambridge University Press; 2010.
[56] Jaynes ET. Information theory and statistical mechanics. Phys Rev 1957;106:620–30.
[57] Pessa AAB, Ribeiro HV. Ordpy: A python package for data analysis with permutation entropy and ordinal network methods. Chaos 2021;31:063110.
[58] Unakafova VA, Keller K. Efficiently measuring complexity on the basis of real-world data. Entropy 2013;15:4392–415.
[59] Kantz H, Schreiber T. Nonlinear time series analysis. second ed.. Cambridge, UK: Cambridge University Press; 2004.
[60] Ruette S. Chaos on the interval. Providence: American Mathematical Society; 2017.
[61] Schuster HG. Deterministic chaos. second ed.. Weinheim: VHC; 1988.