Chaos

A generalized permutation entropy for noisy dynamics and random processes

Cite as: Chaos **31**, 013115 (2021); https://doi.org/10.1063/5.0023419
Submitted: 29 July 2020 . Accepted: 16 December 2020 . Published Online: 06 January 2021

José M. Amigó, Roberto Dale, and Piergiulio Tempesta

A generalized permutation entropy for noisy dynamics and random processes

Cite as: Chaos **31**, 013115 (2021); doi: 10.1063/5.0023419
Submitted: 29 July 2020 · Accepted: 16 December 2020 ·
Published Online: 6 January 2021

José M. Amigó,[1,a] Roberto Dale,[1,b] and Piergiulio Tempesta[2,c]

AFFILIATIONS

[1]Centro de Investigación Operativa, Universidad Miguel Hernández, 03202 Elche, Spain
[2]Departamento de Física Teórica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain and Instituto de Ciencias Matemáticas, C/ Nicolás Cabrera, No. 13–15, 28049 Madrid, Spain

[a]**Author to whom correspondence should be addressed:** jm.amigo@umh.es
[b]**Electronic mail:** rdale@umh.es
[c]**Electronic mail:** p.tempesta@fis.ucm.es and piergiulio.tempesta@icmat.es

ABSTRACT

Permutation entropy measures the complexity of a deterministic time series via a data symbolic quantization consisting of rank vectors called ordinal patterns or simply permutations. Reasons for the increasing popularity of this entropy in time series analysis include that (i) it converges to the Kolmogorov–Sinai entropy of the underlying dynamics in the limit of ever longer permutations and (ii) its computation dispenses with generating and *ad hoc* partitions. However, permutation entropy diverges when the number of allowed permutations grows super-exponentially with their length, as happens when time series are output by dynamical systems with observational or dynamical noise or purely random processes. In this paper, we propose a generalized permutation entropy, belonging to the class of group entropies, that is finite in that situation, which is actually the one found in practice. The theoretical results are illustrated numerically by random processes with short- and long-term dependencies, as well as by noisy deterministic signals.

Measuring the complexity of a time series is usually about counting distinct blocks of a given length along the series. If the entries belong to a finite set of numbers or symbols (called alphabet), then the number of blocks can grow at most exponentially with length, so taking the logarithm is a good idea to distinguish between polynomial and exponential growth. Moreover, the limit of the growth rate with increasing lengths produces a finite number that is independent of length and, hence, intrinsic to the time series. Otherwise, if the alphabet is continuous (think of an interval of real numbers), the situation is more complicated. Such is the case with observations from nonlinear processes and continuous-valued random processes. In this event, one usually divides the alphabet into bins or, as envisaged in this paper, represents each block by the permutation obtained by ranking the numbers in the block. The problem is that the growth rate of the permutations with length becomes super-exponential in the case of noisy and random signals, which prevents a theoretical definition of complexity ("permutation entropy") along the standard lines. In this paper, we borrow ideas from statistical physics and complexity theory to remedy this shortcoming.

I. INTRODUCTION

In general, time series result from observing real-valued random processes or dynamical flows at discrete times. A further step may be the discretization of the data, a procedure called symbolic representation. Such representations simplify the mathematical tools needed for the data analysis and, what is more interesting for practitioners, may be sufficient for the application sought. In this regard, ordinal patterns and permutation entropy have become increasingly popular in nonlinear time series analysis since their introduction by Bandt and Pompe in 2002.[1] The reasons are multiple. Perhaps most importantly from a theoretical point of view, ordinal patterns, which are formally permutations, preserve the temporal structure of a time series and, therefore, its dynamical complexity. In fact, in one-dimensional dynamics, the permutation entropy per symbol converges to the Kolmogorov–Sinai (KS) as the pattern length grows,[2–4] which makes it a proxy of dynamical entropy.

From a practical point of view, the computation of permutation entropy dispenses with *ad hoc* partitions, not to mention the search for generating ones.[5] But even with real-world series, which are finite and usually rather noisy, tools such as permutation entropies of finite order,[6,7] the decay rate of missing ordinal patterns,[8] complexity-entropy causality planes,[9,10] ordinal networks,[11] or simply empirical distributions of ordinal patterns[12] have proved very handy. The potential advantages of the ordinal approach in the analysis of time series include speedy calculation, the possibility of multiscale analysis through a varying pattern length, and high discriminatory power in the classification of data, especially in combination with other complexity indicators.[13] Needless to say, practice also shows some limitations, e.g., in the analysis of short time series with low signal-to-noise ratio, as demonstrated by the study of heart period variability in Ref. 14 with permutation entropy of order 3. Currently, ordinal techniques, alone or complemented by other methods, are being applied in plenty of fields, e.g., chaotic dynamics, earth science, computational neuroscience, biomedicine, and econophysics; see Refs. 15–17 for recent surveys.

More generally, the permutation entropy of a real-valued time series, whether deterministic or random, is just the Shannon entropy of its *ordinal representation*, i.e., the symbolic time series that results from replacing data strings of a fixed length $L \geq 2$ by the corresponding ordinal patterns of length L. There is a twist, though. Shannon's entropy was incepted in the setting of finite-state random processes (information sources with finite alphabets) so that the number of states (words) grows exponentially with the length of the output (message). But in the ordinal representation of time series, each word of length L is replaced by a permutation of $\{0, 1, \ldots, L-1\}$; if all permutations are allowed, as happens in general with real-valued random processes (including noisy chaotic signals), then the number of words grows super-exponentially with L because $L! \simeq e^{L \ln L}$. Similarly, the number of microstates grows super-exponentially with the number of particles in some models of statistical mechanics, the realm of the Boltzmann–Gibbs entropy.[18] For this *super-exponential class* of processes and many-particle systems, the Boltzmann–Gibbs–Shannon (BGS) entropy is not *extensive*, meaning that it does not scale linearly over uniform probability distributions. Consequently, the BGS entropy per symbol or particle is unbounded and, in general, diverges. This is the case, in particular, with the permutation entropy for random processes, where here and hereafter we tacitly include noisy deterministic signals for brevity. Therefore, what is missing to close the conceptual gap between deterministic noiseless and random signals is a definition of permutation entropy rate for the latter.

In this paper, we propose a generalization of permutation entropy that is finite for random processes. To this end, we resort in Sec. IV to a new entropy belonging to the class of group entropies,[19] which is extensive and has several interesting properties (Theorem 2). But before reaching that point, we need to delve into *permutation complexity* in Sec. II, which stands for the complexity of discrete-time, continuous-state deterministic, or random processes and their realizations in ordinal representations,[20,21] as well as deepen into the very concept of group entropy and related universality classes in Sec. III. The theoretical results will be illustrated numerically with random processes of different natures.

II. PERMUTATION COMPLEXITY

Given a time series $(x_t)_{t\geq 0} = x_0, x_1, \ldots, x_t, \ldots$, with t being discrete time and $x_t \in \mathbb{R}$, let $L \geq 2$ and denote by $\mathbf{r}_t$ the rank vector of the string (word, block,...) $x_t^L := x_t, x_{t+1}, \ldots, x_{t+L-1}$. That is,

$$\mathbf{r}_t = (\rho_0, \rho_1, \ldots, \rho_{L-1}), \tag{1}$$

where $\rho_0, \rho_1, \ldots, \rho_{L-1}$ is the permutation of $0, 1, \ldots, L-1$ such that

$$x_{t+\rho_0} < x_{t+\rho_1} < \ldots < x_{t+\rho_{L-1}} \tag{2}$$

(other rules can be found in the literature). The rank vectors $\mathbf{r}_t$ are called *ordinal patterns* or *permutations of length L*, as well as *ordinal L-patterns* for short; the string x_t^L is said to be of *type* $\mathbf{r}_t$. In the case of two or more ties in x_t^L, one can adopt some convention, e.g., the earlier entry is smaller. We suppose tacitly that such occurrences are rare. As a result, the alphabet (set of symbols) of $(\mathbf{r}_t)_{t\geq 0}$, the ordinal representation of the original time series $(x_t)_{t\geq 0}$, is the group of the $L!$ permutations of $0, 1, \ldots, L-1$, which will be denoted by $\mathcal{S}_L$.

Consider a stationary, discrete-time deterministic, or random process $\mathbf{X} = (X_t)_{t\geq 0}$ taking values on a closed interval $I \subset \mathbb{R}$. By a deterministic process, we mean that every output $(x_t)_{t\geq 0}$ of $\mathbf{X}$ is the orbit of x_0 generated by the same mapping $F: I \to I$, i.e., $x_{t+1} = F(x_t) = F^t(x_0)$ for $t \geq 0$. Therefore, random processes include deterministic ones with observational or dynamical noise. Let $p(\mathbf{r})$ be the probability that a string x_t^L output by $\mathbf{X}$ is of type $\mathbf{r}$ and $p = \{p(\mathbf{r}) : \mathbf{r} \in \mathcal{S}_L\}$ the corresponding probability distribution. If $p(\mathbf{r}) > 0$, then $\mathbf{r}$ is an *allowed pattern* for $\mathbf{X}$; otherwise, $\mathbf{r}$ is a *forbidden* pattern. The Shannon entropy (or the BGS entropy for that matter) of p is called the (metric) *permutation entropy of order L*:

$$H^*(X_0^L) = -\sum_{\mathbf{r}\in\mathcal{S}_L} p(\mathbf{r}) \ln p(\mathbf{r}), \tag{3}$$

where $X_0^L := X_0, X_1, \ldots, X_{L-1}$ and $0 \cdot \ln 0 := 0$ by continuity. In the event that $\mathbf{X}$ is a deterministic process, $p(\mathbf{r}) = \mu(\{x_t \in I : x_t^L \text{ is of type } \mathbf{r}\})$, where μ is the *physical measure* of $\mathbf{X}$, which is an F-invariant measure that coincides with the empirical probability distribution.[22] If $\mathbf{X}$ is otherwise a random process, then the probabilities $p(\mathbf{r})$ can only exceptionally be derived from the probability distribution of X_t^L, see e.g., Ref. 23. This means that, in general, the probabilities $p(\mathbf{r})$ of random processes have to be estimated, e.g., by the relative frequencies of each $\mathbf{r} \in \mathcal{S}_L$ in a finite time series $x_0, x_1, \ldots, x_T$,

$$\nu(\mathbf{r}) = \frac{\#\{x_t^L \text{ of type } \mathbf{r} \in \mathcal{S}_L : 0 \leq t \leq T-L+1\}}{T-L+2}, \tag{4}$$

where # stands for "number of" and $T \gg L!$ (maximum likelihood estimator). Then, $p(\mathbf{r}) = \lim_{T\to\infty} \nu(\mathbf{r})$, where this limit exists with probability 1 when the underlying random process fulfills the following weak condition.[1]

Stationarity Condition. *For* $k \leq L-1$, *the probability for* $x_t < x_{t+k}$ *should not depend on t.*

Random processes that meet this condition include, in addition to stationary ones, non-stationary processes with stationary increments such as fractional Brownian motion.[24] Notice for future reference that fractional Gaussian noise, being defined as the increments of a fractional Brownian motion,[24] is, therefore, a stationary random process. From now on, we assume the Stationarity Condition so that the estimations of $p(\mathbf{r})$ converge as the amount of data increases.

The *topological permutation entropy of order L* is the tight upper bound of $H^*(X_0^L)$. It is formally obtained by assuming that all allowed L-patterns are equiprobable,

$$H_0^*(X_0^L) = \ln \mathcal{N}_L(\mathbf{X}), \tag{5}$$

where $\mathcal{N}_L(\mathbf{X})$ is the number of *allowed* patterns of length L for $\mathbf{X}$. In turn, the *metric* and *topological permutation entropies* of a process $\mathbf{X}$, $h^*(\mathbf{X})$, and $h_0^*(\mathbf{X})$ respectively, are obtained by taking the corresponding entropies of order L per variable and letting $L \to \infty$,

$$h^*(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} H^*(X_0^L), \quad h_0^*(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} H_0^*(X_0^L), \tag{6}$$

so that dependencies of any length among variables are taken into account. To ensure that these and the forthcoming limits converge or otherwise diverge to $+\infty$, one can use "lim sup" (limit superior) instead of "lim." We elaborate next on the fact that permutation entropy is finite for deterministic processes while diverging for random processes, in general.

A mapping $F : I \to I$ is said to be piecewise monotone if there is a finite partition of I such that F is continuous and monotone on each subinterval of the partition. Let $h(F)$ be the KS entropy of F, and $h_0(F)$ its topological entropy.[25] The following theorem holds.[2]

Theorem 1. *If F is piecewise monotone, then (i) $h^*(F) = h(F)$ and (ii) $h_0^*(F) = h_0(F)$.*

All one-dimensional mappings encountered in practice are piecewise monotone, so we may assume this property for the mappings underlying deterministic processes. Therefore, $h^*(\mathbf{X}) \le h_0^*(\mathbf{X}) < \infty$ for deterministic processes since $h_0(F) < \infty$ for piecewise monotone mappings.[26] Incidentally, Theorem 1(ii) implies $\mathcal{N}_L(\mathbf{X}) \sim e^{h_0(F)L}$ ($\sim$ stands for "asymptotically"), meaning that such processes have only exponentially many allowed L-patterns for ever larger L's, despite the fact that there are $L! \sim e^{L \ln L} = L^L$ (Stirling's formula) possible ordinal L-patterns. The upshot is that the number of forbidden patterns for deterministic processes grows super-exponentially with L, see Ref. 27. Also, higher dimensional dynamics along with their lower dimensional projections may have forbidden patterns.[28] However, if the dynamics takes place on an attractor so that the orbits are dense, then the observational or dynamical noise will "destroy" all forbidden patterns in the long run, no matter how small the noise. Theorem 1 has been generalized to include countably many monotonicity intervals.[29]

On the other hand, random processes may have forbidden patterns too. For the sake of our analysis, though, we will consider the general or "worse" scenario in which all ordinal patterns of any length are allowed. A necessary and sufficient condition for this is that, for $k \le L - 1$, the probability for $x_t < x_{t+k}$ is neither 0 nor 1 (so that the same holds for $x_t > x_{t+k}$), which amounts to a mild addendum to the Stationarity Condition. With this proviso, we may assume hereafter $\mathcal{N}_L(\mathbf{X}) = L!$ for all random processes $\mathbf{X}$. Then,

$$h_0^*(\mathbf{X}) = \lim_{L\to\infty} \frac{1}{L} \ln L! = \lim_{L\to\infty} \ln L = \infty \tag{7}$$

by Stirling's formula. We conclude from (7) that permutation entropy, unlike Shannon's entropy, cannot be applied to random processes in general. In particular, $H^*(X_0^L)$ does not scale linearly when $L \to \infty$ over flat probability distributions.

Numerical evidence is shown in Fig. 1. Here, we have numerically generated 10 realizations of size $T > 50L!$ [see (4)] of the following processes: (i) *white noise* (WN) in the form of an independent and uniformly distributed process on $[0, 1]$; (ii) fractional Gaussian noise (fGn) with Hurst exponent $H = 0.5$ (Gaussian white

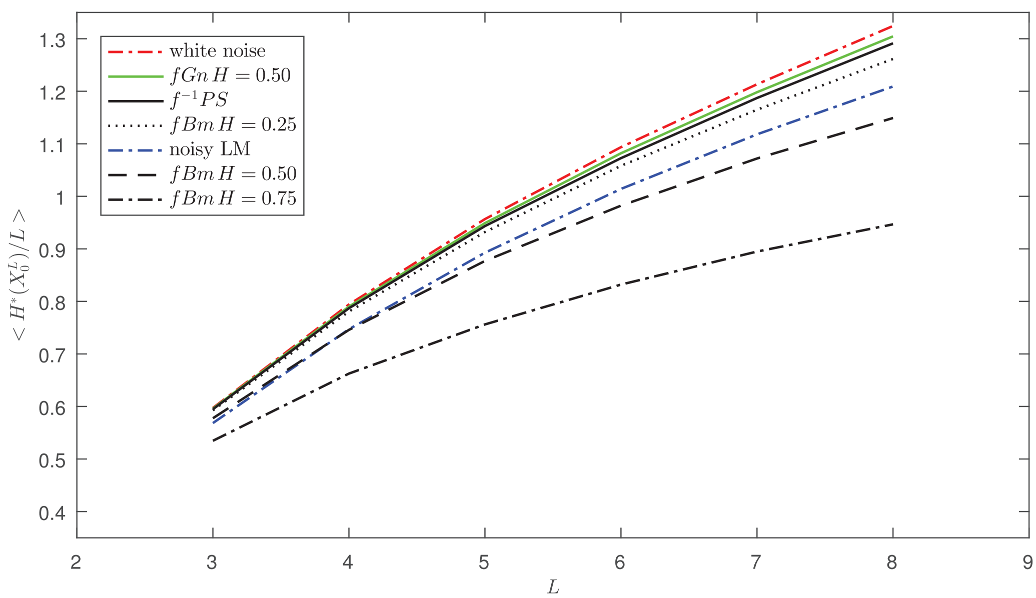

FIG. 1. The average of $H^*(X_0^L)/L$ over 10 realizations, $\langle H^*(X_0^L)/L\rangle$, is plotted vs L for $1 \le L \le 8$ and the random processes listed in the inset. See the text for detail.

noise);[24] (iii) noise with an f^{-1} power spectrum (PS); (iv) fractional Brownian motion (fBm) with $H = 0.25$ (anti-persistent process), $H = 0.5$ (classical Brownian motion), and $H = 0.75$ (persistent process);[24] (v) logistic map iterations with additive white noise of amplitude 0.25 (noisy LM), i.e., $x_t = y_t + z_t$, where $y_t = 4y_{t-1}(1 - y_{t-1})$, $0 < y_0 < 1$, and $(z_t)_{t\geq 0}$ is WN with $-0.25 \leq z_t \leq 0.25$. Computations were done with MatLab.[30] The average of $H^*(X_0^L)/L$ over the 10 realizations of each process, denoted $\langle H^*(X_0^L)/L\rangle$, is then plotted against L, $3 \leq L \leq 8$. We see in all cases that $\langle H^*(X_0^L)/L\rangle$ follows a seemingly divergent trajectory as L grows.

III. GROUP ENTROPIES

The theory of group entropies[31–34] is an axiomatic approach, which allows us to construct information measures with mathematical properties that make them suitable to describe specific universality classes of complex systems.[35] We recap here some basic definitions.

Let $\mathscr{P}_W$ be the set of all discrete probability distributions with W entries, i.e., $\mathscr{P}_W = \{p = (p_i)_{i=1,\ldots,W} : 0 \leq p_i \leq 1, \sum_{i=1}^W p_i = 1\}$. Let S be a non-negative function on $\mathscr{P} := \cup_{W=1}^{\infty} \mathscr{P}_W$ so that S is defined on any probability distribution p and $S(p) \geq 0$. The Shannon–Khinchin (SK) axioms are a set of requirements first considered in Refs. 36–38 to uniquely characterize the BGS entropy. The first three SK axioms amount to the following properties:

(SR1) $S(p)$ is continuous with respect to all variables $p_1, \ldots, p_W$.
(SR2) $S(p)$ takes its maximum value over the uniform distribution.
(SR3) $S(p)$ is expansible: adding an event of zero probability does not affect the value of $S(p)$.

These axioms represent a minimal set of "non-negotiable" requirements that such functions $S(p)$ should satisfy necessarily to be meaningful, both from a physical and information-theoretical point of view. Non-negative functions on $\mathscr{P}$ that verify axioms (SK1)–(SK3) are called generalized entropies and their structure is only known under additional conditions.[19,39,40] Thus, the fourth SK axiom, requiring specifically additivity on conditional distributions, leads to the BGS entropy,[38]

$$S_{BGS}(p) = -k \sum_{i=1}^{W} p_i \ln p_i, \tag{8}$$

where k is a positive constant that we equate to 1 for definiteness [as in Eq. (3)]. Instead, the more general axiom of composability (see below) leads to the concept of group entropies. As we will discuss shortly, this new class of entropies, which includes $S_{BGS}(p)$, is better suited to deal with the diversity of themodynamical and complex systems. Another independent approach is based on the concept of *pseudo-additive entropy*.[41]

An entropy $S(p)$ is said to be *composable* if there exists a (sufficiently regular) function $\Phi(x, y)$ such that

$$S(p_A \times p_B) = \Phi(S(p_A), S(p_B)) \tag{9}$$

for any probability distributions p_A and p_B, where $p_A \times p_B$ is the product probability distribution of both. Equivalently, (9) can be written as $S(A \cup B) = \Phi(S(A), S(B))$, where A and B are two statistically *independent* subsystems of a complex system, defined over *any arbitrary probability distributions* p_A and p_B, respectively, and $A \cup B$ is the system composed of A and B. All quantities are assumed to be dimensionless.

In addition to Eq. (9), we shall also require the following properties for the *composition law* Φ:

(C1) Symmetry: $\Phi(x, y) = \Phi(y, x)$.
(C2) Associativity: $\Phi(x, \Phi(y, z)) = \Phi(\Phi(x, y), z)$.
(C3) Null-composability: $\Phi(x, 0) = x$.

We shall say that an entropy satisfies the *composability axiom* if it fulfills Eq. (9) and the requirements (C1)–(C3). Observe that, indeed, requirements (C1)–(C3) are crucial: they impose the independence of the composition process with respect to the order of A and B, the possibility of composing three independent subsystems in an arbitrary way, and the requirement that, when composing a system with another one having zero entropy, the total entropy remains unchanged. In our opinion, these properties are also fundamental: no thermodynamic or information-theoretic applications would be easily conceivable without these properties. For $\Phi(x, y) = x + y$, we obtain from (9) the additivity of the BGS entropy (8) with respect to the composition of two statistically independent subsystems.

From an algebraic point of view, the requirements (C1)–(C3) define a *formal group law* for a function (infinite series) of the form $\Phi(x, y) = x + y + O(2)$, where $O(n)$ stands for terms of degree $\geq n$.

Definition 1. A group entropy is a function $S : \mathscr{P} \to [0, \infty)$, which satisfies the Shannon–Khinchin axioms (SK1)–(SK3) and the composability axiom.

A well-known group entropy, introduced by Tsallis,[42] is

$$S_\alpha(p) = \frac{1}{1-\alpha}\left(\sum_{i=1}^{W} p_i^\alpha - 1\right) \tag{10}$$

for $\alpha > 0, \alpha \neq 1$, and $S_1(p) := \lim_{\alpha\to 1} S_\alpha(p) = S_{BGS}(p)$, whose composition law is

$$\Phi(x, y) = x + y + (1 - \alpha)xy \tag{11}$$

so that $S_\alpha(p_A \times p_B) = S_\alpha(p_A) + S_\alpha(p_B) + (1 - \alpha)S_\alpha(p_A)S_\alpha(p_B)$. Except for the Tsallis entropy, group entropies are, in general, non-trace functions,[43] that is, they cannot be written as $\sum_{i=1}^W g(p_i)$, where $g : [0, 1] \to [0, \infty)$ is a mapping with suitable properties, usually[39] continuity, $\cap$-convexity, and $g(0) = 0$.

As has been shown,[19,35] one can classify complex systems according to their state space growth rate $\mathscr{W}(N)$, which counts the number of microstates allowed as a function of the number N of particles or constituents of a given system, for large N. Generally speaking, we distinguish sub-exponential, exponential, and super-exponential regimes with regard to the state space growth rate (which can be further discriminated if necessary). All systems that are characterized by the same asymptotic behavior of $\mathscr{W}$ define a *universality class*. According to Theorem 1 of Ref. 19, under mild hypotheses one can explicitly construct a suitable group entropy associated with a given universality class of systems, which would play the role of information or complexity measure for the class considered. This specific entropy (actually, a one-parametric family of entropies) is called a Z-entropy[32] and is denoted by $Z_{G,\alpha}(p)$, where G refers to the group-theoretical structure associated with it, $\alpha > 0$, $p \in \mathscr{P}_W$, and $W = \lfloor \mathscr{W}(N) \rfloor$.

To be more precise, one can construct a suitable $Z_{G,\alpha}(p)$ entropy which is *extensive* for the systems of a given class, that is, if $Z_{G,\alpha}(N) := Z_{G,\alpha}\left(\frac{1}{W}, \ldots, \frac{1}{W}\right)$ is the Z-entropy over the uniform distribution (the most "disordered" situation), then

$$\lim_{N\to\infty} \frac{Z_{G,\alpha}(N)}{N} = \text{const.} \tag{12}$$

In other words, $Z_{G,\alpha}(N)$, the topological version of $Z_{G,\alpha}(p)$, scales linearly with N, at least for N sufficiently large. According to (SK2), $Z_{G,\alpha}(p) \leq Z_{G,\alpha}(N)$ for all $p \in \mathscr{P}_W$.

Prototypical examples of Z-entropies are (i) the Tsallis entropy $S_\alpha(p)$, Eq. (10), for the sub-exponential class and (ii) the Rényi entropy[44]

$$R_\alpha(p) = \frac{1}{1-\alpha} \ln\left(\sum_{i=1}^{W} p_i^\alpha\right) \tag{13}$$

for $\alpha > 0$, $\alpha \neq 1$, and $R_1(p) := \lim_{\alpha\to 1} R_\alpha(p) = S_{BGS}(p)$, for the exponential class. Notice that $S_\alpha(p) = \frac{1}{1-\alpha}(\exp[(1-\alpha)R_\alpha(p)] - 1)$. The Z-entropy for the super-exponential class is our next concern.

IV. A GENERALIZED PERMUTATION ENTROPY

In our context, where random processes are real-valued and blocks x_t^L of size $L \geq 2$ are quantized by means of ordinal L-patterns $\mathbf{r}_t$, discrete probability distributions p refer necessarily to the symbols $\mathbf{r} \in \mathscr{S}_L$ and hence the growth function is $\mathscr{W}(L) = L! \sim e^{L\ln L}$ under very weak conditions. This being the case, we propose the Z-entropy for the super-exponential class to measure permutation complexity. Such an entropy was introduced[18] to describe the thermodynamic properties of the so-called pairing model, which represents an example of a Hamiltonian system possessing a super-exponential state space growth rate. Precisely, we propose the following:

Definition 2. The *permutation Z-entropy of order L* of a process $\mathbf{X} = (X_t)_{t\geq 0}$ is the function

$$Z_\alpha^*(X_t^L) := Z_\alpha(p) = \exp\left[\mathscr{L}\left(R_\alpha(p)\right)\right] - 1 \tag{14}$$

for $\alpha > 0$. Here, $p \in \mathscr{P}_{L!}$ is the probability distribution of the ordinal L-patterns of X_t^L, $R_\alpha(p)$ is Rényi's entropy (13) with $W = L!$, and $\mathscr{L}(x)$ denotes the principal branch of the real Lambert function.

$\mathscr{L}(x)$ is a smooth function that is defined for $x \geq -1/e$ and satisfies the equation $\mathscr{L}(x)\, e^{\mathscr{L}(x)} = x$, hence $\mathscr{L}(0) = 0$ and $\mathscr{L}(x) > 0$ for $x > 0$, see Ref. 45. The term -1 in (14) renders $Z_\alpha(p) = 0$ in situations without uncertainty, i.e., when $p_{i_0} = 1$ and $p_i = 0$ for $i \neq i_0$.

From a conceptual point of view, $Z_\alpha^*(X_t^L)$ can be interpreted to be a suitable, extensive deformation of $R_\alpha(p)$, sharing with it many fundamental properties, except additivity. For example, $Z_\alpha^*(X_t^L)$ inherits from $R_\alpha(p)$ its $\cap$-convexity for $0 < \alpha \leq 1$ and decreasing monotonicity with respect to α, see Ref. 40; that is,

$$Z_\alpha^*(X_t^L) \geq Z_\beta^*(X_t^L) \quad \text{for} \quad \alpha < \beta, \tag{15}$$

because the function $e^{\mathscr{L}(x)}$ is strictly increasing and $\cap$-convex.

Remark. According to Eq. (4.13.5) of Ref. 45,

$$\mathscr{L}(x) = x - x^2 + O(3) \tag{16}$$

for $|x| < 1/e$. Therefore,

$$e^{\mathscr{L}(x)} = 1 + x - \tfrac{1}{2}x^2 + O(3). \tag{17}$$

In view of Eqs. (17) and (14),

$$Z_\alpha(p) = R_\alpha(p) - \tfrac{1}{2}R_\alpha(p)^2 + O(3) \simeq R_\alpha(p)$$

if $R_\alpha(p) < 1/e$. Small values of the permutation Rényi entropy $R_\alpha(p)$, $0 \leq R_\alpha(p) \leq \ln L!$ [see Eq. (13) with $p_i = 1/W$ and $W = L!$] occur for probability distributions that peak around a single ordinal pattern or a few ordinal patterns, i.e., in situations where the uncertainty is low. We conclude that when the Rényi entropy of the probability distribution of the ordinal L-patterns of X_t^L is small, it is a good approximation of the permutation Z-entropy of order L.

It is clear that $Z_\alpha(p)$ verifies the axioms (SK1)–(SK3) since $R_\alpha(p)$ is a group entropy and $e^{\mathscr{L}(x)}$ is strictly increasing. The composability of $Z_\alpha(p)$ for the growth function $\mathscr{W}(L) = e^{L\ln L}$ follows from Proposition 1 of Ref. 19 (with $\mathscr{W}^{-1}(\xi) = \exp[\mathscr{L}(\ln\xi)]$). Alternatively, one can directly check that if

$$\begin{aligned}\Phi(x,y) &= e^{\mathscr{L}[(x+1)\ln(x+1)+(y+1)\ln(y+1)]} - 1 \\ &= x + y - \tfrac{1}{2}x^2 - 2xy - \tfrac{1}{2}y^2 + O(3),\end{aligned} \tag{18}$$

then the composition law $Z_\alpha(p_A \times p_B) = \Phi(Z_\alpha(p_A), Z_\alpha(p_B))$ holds for any probability distributions p_A and p_B.

As with conventional permutation entropy, we can introduce a topological version of $Z_\alpha^*(X_t^L)$.

Definition 3. The *topological permutation Z-entropy of order L* of a process $\mathbf{X} = (X_t)_{t\geq 0}$ is defined to be the tight upper bound of $Z_\alpha^*(X_t^L)$, which is obtained over the uniform distribution of ordinal L-patterns,

$$Z_0^*(X_t^L) := Z_\alpha(\tfrac{1}{L!}, \ldots, \tfrac{1}{L!}) = \exp\left[\mathscr{L}\left(\ln L!\right)\right] - 1. \tag{19}$$

Here, we took into account that $R_\alpha(\frac{1}{L!}, \ldots, \frac{1}{L!}) = \ln L!$ for all α. The notation Z_0^* is justified because $\ln L!$ is formally obtained from (13) by setting $\alpha = 0$. It follows [use $\mathscr{L}(x\ln x) = \ln x$ for $x \geq 1/e$]

$$\frac{Z_0^*(X_0^L)}{L} = \frac{e^{\mathscr{L}(\ln L!)} - 1}{L} \sim \frac{e^{\mathscr{L}(L\ln L)} - 1}{L} = \frac{L-1}{L} \sim 1 \tag{20}$$

so that $Z_\alpha^*(p)$ is indeed extensive in the regime of factorial growth we are interested in.

Last but not least, we also define the corresponding entropy rate per variable.

Definition 4. The *permutation Z-entropy rate* (or just *permutation Z-entropy*) *of a random process* $\mathbf{X}$ is given as

$$z_\alpha^*(\mathbf{X}) = \limsup_{L\to\infty} \frac{1}{L} Z_\alpha^*(X_0^L), \tag{21}$$

where $\alpha \geq 0$: $z_0^*(\mathbf{X})$ is the topological permutation Z-entropy, and $z_\alpha^*(\mathbf{X})$ with $\alpha > 0$ is the metric permutation Z-entropy.

Next, we prove two basic properties.

Theorem 2. *The permutation Z-entropy rate $z_\alpha^*(\mathbf{X})$ satisfies the following inequalities:*

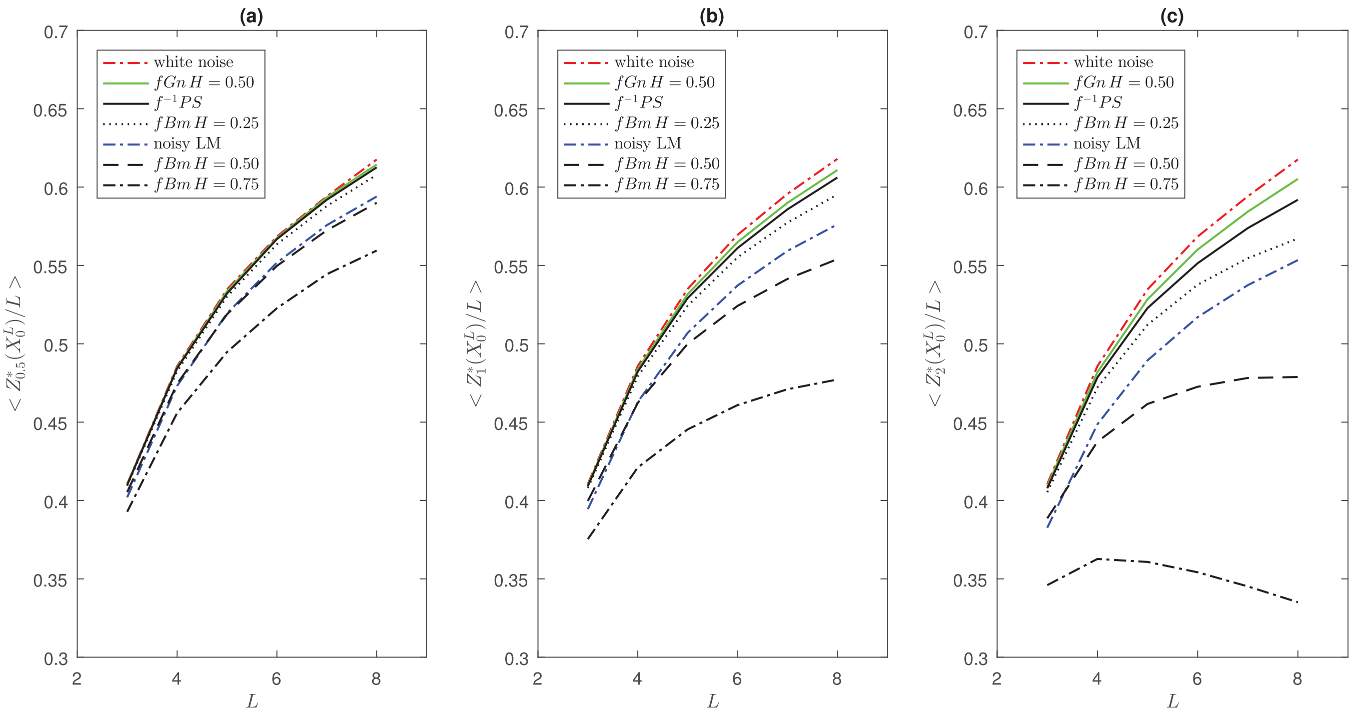

FIG. 2. Same information as in Fig. 1 (but notice the different scale on the Y axis) for $\langle Z^*_{0.5}(X_0^L)/L\rangle$ (a), $\langle Z^*_1(X_0^L)/L\rangle$ (b), and $\langle Z^*_2(X_0^L)/L\rangle$ (c). See the text for detail.

(i) *Normalized range:* $0 \le z^*_\alpha(\mathbf{X}) \le 1$, *where* $z^*_\alpha(\mathbf{X}) = 0$ *for deterministic processes and* $z^*_\alpha(\mathbf{X}) = 1$ *for white noise.*
(ii) *Hierarchical order:* $z^*_\alpha(\mathbf{X}) \ge z^*_\beta(\mathbf{X})$ *for* $\alpha < \beta$.

Proof. To prove that $z^*_\alpha(\mathbf{X}) = 0$ for deterministic processes, we recall that, according to Theorem 1(ii), $\mathcal{N}_L(\mathbf{X}) \sim e^{h_0(F)L}$, where $h_0(F)$ is the topological entropy of the mapping F that generates $\mathbf{X}$. Therefore, if p is the probability distribution of the L-patterns, then $R_0(p) = \ln \mathcal{N}_L(\mathbf{X}) \sim h_0(F)L$ and

$$\frac{Z^*_\alpha(X_0^L)}{L} \le \frac{e^{\mathscr{L}[R_0(p)]} - 1}{L} \sim \frac{e^{\mathscr{L}[h_0(F)L]}}{L} = \frac{h_0(F)}{\mathscr{L}[h_0(F)L]} \sim 0, \qquad (22)$$

where we used $e^{\mathscr{L}(x)} = x/\mathscr{L}(x)$. Furthermore, the inequality $z^*_\alpha(\mathbf{X}) \le 1$, with equality for white noise, follows from $Z^*_\alpha(X_0^L) \le Z^*_0(X_0^L)$ and $\limsup_{L\to\infty} \frac{1}{L} Z^*_0(X_0^L) = 1$, see (20).

The hierarchical order of $z^*_\alpha(\mathbf{X})$ is a direct consequence of (15). □

As a way of illustration, Fig. 2 shows $\langle Z^*_\alpha(X_0^L)/L\rangle$, the average of the permutation entropy rate $Z^*_\alpha(X_0^L)/L$ over the same 10 time series and for the same random processes as in Fig. 1, against L, $1 \le L \le 8$, where $\alpha = 0.5$ (a), 1 (b), and 2 (c). Contrarily to Fig. 1, we see in all panels of Fig. 2 that $\langle Z^*_\alpha(X_0^L)/L\rangle$ follows a seemingly convergent trajectory as L grows, upper bounded by the white noise. In agreement with (15), $\langle Z^*_{0.5}(X_0^L)/L\rangle \ge \langle Z^*_1(X_0^L)/L\rangle \ge Z^*_2(X_0^L)/L\rangle$ for each process.

To wrap up, let us point out that the curves of different processes may cross, as happens in the three panels of Fig. 2 with the processes fBm $H = 0.50$ and noisy LM when going from $L = 3$ to $L = 4$. Similar intersections also occur with other parameter settings and processes (not shown). The reason is that $Z^*_\alpha(X_0^L)$ 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 $Z^*_\alpha(X_t^L)$. In particular, as L grows, $Z^*_\alpha(X_t^L)$ can become larger for a noisy chaotic signal, such as the noisy logistic map of Fig. 2, than for a process with a longer, or an infinite, span of interdependence between its increments, such as the fractional Brownian motion with $H = 0.50$.

V. DISCUSSION AND CONCLUSIONS

This paper addresses the divergence of the permutation entropy of finite order $H^*(X_0^L)$ for random processes $\mathbf{X} = (X_t)_{t\ge 0}$ (including noisy deterministic signals) when $L \to \infty$, that is, the lack of a permutation entropy rate for such processes. Therefore, its main scope is to extend the concept of permutation entropy to the realm of random processes. For this purpose, we studied the permutation Z-entropy rate $z^*_\alpha(\mathbf{X})$, a group entropy defined in Eqs. (21) and (14). To be more specific, $z^*_\alpha(\mathbf{X})$ measures the complexity of a real-valued random process $\mathbf{X}$ through permutations, where $\mathbf{X}$ is supposed to fulfill the Stationarity Condition and the mild assumption that all permutations are allowed for each length or, at least, a super-exponentially growing number of them. First and foremost, $z^*_\alpha(\mathbf{X})$ is always finite, contrarily to what happens with the conventional permutation entropy $h^*(\mathbf{X})$, see Eq. (7). Therefore, we may claim that $z^*_\alpha(\mathbf{X})$ extends $h^*(\mathbf{X})$ to the realm of random processes although, needless to say, $z_\alpha^*(\mathbf{X})$ differs from $h^*(\mathbf{X})$ when $\mathbf{X}$ is deterministic. Among the features of $z_\alpha^*(\mathbf{X})$, we singled out in Theorem 2 its normalized range and hierarchical order. Figures 1 and 2 depict the numerical experiments done. On the foregoing grounds, we propose $z_\alpha^*(\mathbf{X})$ as a suitable entropic measure to describe the complexity of real-valued random processes in ordinal representations.

Applications include the analysis of data, in general, and the characterization and classification of noisy signals, in particular. In this regard, the parameter α is an asset because it enhances the discrimination capability of the ordinal approach, as exemplified in Fig. 2. Since real-world series are finite, one has to use permutation Z-entropies of finite order $Z_\alpha^*(X_0^L)$ in that case, where L should be chosen so as to avoid undersampling of the ordinal L-patterns.[46] From its definition, Eq. (14), it follows that $Z_\alpha^*(X_0^L)$ and the permutation Rényi entropy $R_\alpha(p)$, where p is the probability distribution of the ordinal L-patterns, are functions of each other, the difference being that $R_\alpha(p)$ is in general unbounded when $L \to \infty$. Therefore, $Z_\alpha^*(X_0^L)$ and $R_\alpha(p)$ share the same strengths (say, discriminatory power) and weaknesses (say, dependence on L); computation time is virtually the same. As a general rule, it is good practice in time series analysis to use multiple tools and parameter settings to obtain more accurate diagnoses from the data.

In conclusion, it is the concept of the Z-entropy rate of a random process $\mathbf{X}$ that makes the difference. Like the Shannon entropy of a finite-valued process and the Kolmogorov–Sinai entropy of a dynamical system, which are limits on extensive parameters, $z_\alpha^*(\mathbf{X})$ defines an intrinsic characteristic (of $\mathbf{X}$ in this case) because it is also such a limit. Since finite time series are modeled by infinitely long processes, the theoretical study of the latter can provide insights into the properties of the former. For this reason, the numerical evaluation of $z_\alpha^*(\mathbf{X})$, although challenging, deserves further thought and effort and will be the subject of further research.

ACKNOWLEDGMENTS

We thank very much our referees for their constructive criticism. J.M.A. and R.D. were financially supported by Ministerio de Ciencia e Innovación, PID2019-108654GB-I00, Spain. 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.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," Phys. Rev. Lett. **88**, 174102 (2002).
[2] C. Bandt, G. Keller, and B. Pompe, "Entropy of interval maps via permutations," Nonlinearity **15**, 1595–602 (2002).
[3] K. Keller and M. Sinn, "Kolmogorov–Sinai entropy from the ordinal viewpoint," Physica D **239**, 997–1000 (2010).
[4] J. M. Amigó, "The equality of Kolmogorov–Sinai entropy and metric permutation entropy generalized," Physica D **241**, 789–793 (2012).
[5] Y. Hirata, K. Judd, and D. Kilminster, "Estimating a generating partition from observed time series: Symbolic shadowing," Phys. Rev. E **70**, 016215 (2004).
[6] V. Unakafova and K. Keller, "Efficiently measuring complexity on the basis of real-world data," Entropy **15**, 4392–4415 (2013).
[7] A. M. Unakafov and K. Keller, "Conditional entropy of ordinal patterns," Physica D **269**, 94–102 (2014).
[8] L. C. Carpi, P. M. Saco, and O. A. Rosso, "Missing ordinal patterns in correlated noises," Physica A **389**, 2020–2029 (2010).
[9] O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, "Distinguishing noise from chaos," Phys. Rev. Lett. **99**, 154102 (2007).
[10] L. Zunino, M. C. Soriano, and O. A. Rosso, "Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach," Phys. Rev. E **86**, 046210 (2012).
[11] A. A. B. Pessa and H. V. Ribeiro, "Characterizing stochastic time series with ordinal networks," Phys. Rev. E **100**, 042304 (2019).
[12] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, "Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamic," Comput. Biol. Med. **42**, 319–327 (2012).
[13] L. Zunino and H. V. Ribeiro, "Discriminating image textures with the multiscale two-dimensional complexity-entropy causality plane," Chaos Solitons Fractals **91**, 676–688 (2016).
[14] A. Porta, V. Bari, A. Marchi, B. De Maria, P. Castiglioni, M. di Rienzo, S. Guzzetti, A. Cividjian, and L. Quintin, "Limits of permutation-based entropies in assessing complexity of short heart period variability," Physiol. Meas. **36**, 755–765 (2015).
[15] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: A review," Entropy **14**, 1553–1577 (2012).
[16] J. M. Amigó, K. Keller, and J. Kurths, "Recent progress in symbolic dynamics and permutation complexity," Eur. Phys. J. Spec. Top. **222**, 241–598 (2013).
[17] J. M. Amigó, K. Keller, and V. Unakafova, "Ordinal symbolic analysis and its applications to biomedical recordings," Philos. Trans. R. Soc. A **373**, 20140091 (2015).
[18] H. J. Jensen, R. H. Pazuki, G. Pruessner, and P. Tempesta, "Statistical mechanics of exploding phase spaces: Ontic open systems," J. Phys. A Math. Theor. **51**, 375002 (2018).
[19] P. Tempesta and H. J. Jensen, "Universality classes and information-theoretic measures of complexity via group entropies," Nat. Sci. Rep. **10**, 1–11 (2020).
[20] J. M. Amigó, *Permutation Complexity in Dynamical Systems—Ordinal Patterns, Permutation Entropy and All That* (Springer Verlag, Berlin, 2010), ISBN 978–3642040832.
[21] R. Monetti, J. M. Amigó, T. Aschenbrenner, and W. Bunk, "Permutation complexity of interacting dynamical systems," Eur. Phys. J. Spec. Top. **222**, 421–436 (2013).
[22] J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. **57**, 617–656 (1985).
[23] C. Bandt and F. Shiha, "Order patterns and time series," J. Time Ser. Anal. **28**, 646–665 (2007).
[24] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," SIAM Rev. **10**, 422–437 (1968).
[25] P. Walters, *An Introduction to Ergodic Theory* (Springer Verlag, New York, 2000).
[26] M. Misiurewicz and W. Szlenk, "Entropy of piecewise monotone mappings," Stud. Math. **67**, 45–63 (1980).
[27] J. M. Amigó, L. Kocarev, and J. Szczepanski, "Order patterns and chaos," Phys. Lett. A **355**, 27–31 (2006).
[28] J. M. Amigó and M. B. Kennel, "Forbidden ordinal patterns in higher dimensional dynamics," Physica D **237**, 2893–2899 (2008).
[29] T. Gutjahr and K. Keller, "Equality of Kolmogorov–Sinai and permutation entropy for one-dimensional maps consisting of countably many monotone parts," Disc. Cont. Dyn. Syst. **39**, 4207–4224 (2019).
[30] MATLAB version 9.0.0 (R2016a), The MathWorks Inc., Natick, MA, 2016.
[31] P. Tempesta, "Group entropies, correlation laws and zeta functions," Phys. Rev. E **84**, 021121 (2011).

[32] P. Tempesta, "Formal groups and Z-entropies," Proc. R. Soc. A **472**, 20160143 (2016).
[33] M. A. Rodríguez, A. Romaniega, and P. Tempesta, "A new class of entropic information measures, formal group theory and information geometry," Proc. R. Soc. A **475**, 20180633 (2019).
[34] P. Tempesta, "Multivariate group entropies, super-exponentially growing systems and functional equations," Chaos **30**, 123119 (2020).
[35] H. J. Jensen and P. Tempesta, "Group entropies: From phase space geometry to entropy functionals via group theory," Entropy **20**, 804 (2018).
[36] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948) [**27**, 623–653 (1948)].
[37] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
[38] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).
[39] R. Hanel and S. Thurner, "A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions," Europhys. Lett. **93**, 20006 (2011).
[40] J. M. Amigó, S. G. Balogh, and S. Hernández, "A brief review of generalized entropies," Entropy **20**, 813 (2018).
[41] V. Ilic and M. Stankovic, "Generalized Shannon-Khinchin axioms and uniqueness theorem for pseudo-additive entropies," Physica A **411**, 138–145 (2014).
[42] C. Tsallis, "Possible generalization of the Boltzmann–Gibbs statistics," J. Stat. Phys. **52**, 479–487 (1988).
[43] A. Enciso and P. Tempesta, "Uniqueness and characterization theorems for generalized entropies," J. Stat. Mech. Theory Exp. **2017**, 123101 (2017).
[44] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, CA, 1961), pp. 547–561.
[45] *NIST Handbook of Mathematical Functions*, edited by F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (Cambridge University Press, Cambridge, 2010).
[46] D. Cuesta-Frau, J. P. Murillo-Escobar, D. A. Orrego, and E. Delgado-Trejos, "Embedded dimension and time series length: Practical influence on permutation entropy and its applications," Entropy **21**, 385 (2019).