Article

Group Structure as a Foundation for Entropies

Henrik Jeldtoft Jensen [1,2,*] and Piergiulio Tempesta [3,4]

[1] Centre for Complexity Science and Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

[2] Department of Computer Science, School of Computing, Tokyo Institute of Technology, 4259, Nagatsuta-cho, Yokohama 226-8502, Japan

[3] Departamento de Fisica Teórica, Universidad Complutense de Madrid, 28040 Madrid, Spain; p.tempesta@fis.ucm.es

[4] Instituto de Ciencias Matemáticas (ICMAT), 28049 Madrid, Spain

[*] Correspondence: h.jensen@imperial.ac.uk

Abstract: Entropy can signify different things. For instance, heat transfer in thermodynamics or a measure of information in data analysis. Many entropies have been introduced, and it can be difficult to ascertain their respective importance and merits. Here, we consider entropy in an abstract sense, as a functional on a probability space, and we review how being able to handle the trivial case of non-interacting systems, together with the subtle requirement of extensivity, allows for a systematic classification of the functional form.

Keywords: entropy; composability; extensivity; information theory; power laws; group theory

1. Introduction

The term "entropy" is used extensively in the modern scientific literature. Originating in the 19th-century theory of thermal dynamics [1], the concept is now, to a near bewildering extent, used widely in one form or another across many sciences. For example, entropy is at the foundation of information theory [2] and is of crucial use in computer science [3,4]. Also, neurosciences make use of entropy both as a tool to characterize and interpret data from brain scans [5] and, more fundamentally, in theories of the dynamics of the brain and mind [6]. Generally speaking, entropy is a fundamental notion in complexity science [7]. Here, we present a brief review of some recent mathematical developments in the theory of entropy.

In mathematical terms, an entropy is a functional $S[p]$ defined on a space of probability distributions $p = (p_1, p_2, \ldots, p_W)$ associated with a W-dimensional event space. Thus, we use the word "entropy" with the same meaning it assumes, for instance, in the case of Rényi's entropy, without direct reference to thermodynamics. From this perspective, the relevance of entropies is clear. They can be considered as analytic tools that can help in the analysis of the inter-dependencies within the system behind a given event space. Similarly, their use in information-theoretic analysis of time series is likewise natural. The connection between entropies as mathematical functionals and the thermodynamic entropy of Clausius defined in terms of heat transfer is much less immediate. Here, we will concentrate on the mathematical aspects of entropies as functionals and only make a few comments on the possible connection to thermodynamics.

The first question to tackle is which functional form of $S[p]$ yields useful entropies. i.e., how to limit the infinite number of choices for $S[p]$. It is well known that the Boltzmann–Gibbs–Shannon form

$$S_{\text{BGS}}[p] = \sum_{i=1}^{W} p_i \log \frac{1}{p_i} \tag{1}$$

(we assume $k_B = 1$) is the unique possibility if one assumes that the entropy must satisfy the four Shannon–Kinchin (SK) axioms [8]:

(SK1) (Continuity). The function $S(p_1, \ldots, p_W)$ is continuous with respect to all its arguments.

(SK2) (Maximum principle). The function $S(p_1, \ldots, p_W)$ takes its maximum value over the uniform distribution $p_i = 1/W, i = 1, \ldots, W$.

(SK3) (Expansibility). Adding an event of zero probability to a probability distribution does not change its entropy: $S(p_1, \ldots, p_W, 0) = S(p_1, \ldots, p_W)$.

(SK4) (Additivity). Given two subsystems A, B of a statistical system, $S(A \cup B) = S(A) + S(B|A)$.

Therefore, to derive entropies of a functional form different from the one in Equation (1), it is necessary to go beyond the four SK axioms. Various strategies in this respect have been adopted in the literature.

Let us start by recalling Constantino Tsallis's elegant observation [9] that the formula

$$S_q[p] = k \frac{1 - \sum_{i=1}^{W} p_i^q}{q - 1}, \quad p = (p_1, p_2, \ldots, p_W) \in [0,1]^W, \quad k \in \mathbb{R}_+ \tag{2}$$

provides a possible generalization of Boltzmann's entropy. This is the case in the sense that S_q is a functional on the space of probability distributions $p : \{1, 2, \ldots, W\} \mapsto [0,1]^W$, and in the limit $q \to 1$, the entropy $S_q[p]$ becomes equal to the Boltzmann–Gibbs–Shannon entropy in Equation (1). Tsallis's 1988 article [9] has inspired a tremendous effort to generalize Boltzmann's entropy in different scenarios, including what we will review in this paper. Tsallis pointed out that the entropy S_q fulfills a specific procedure for combining independent systems which can be seen as a generalization of the additivity property (SK4) of the Boltzmann–Gibbs–Shannon entropy. In particular, Tsallis suggested that the free parameter q should be determined by requiring that S_q for a given system is extensive (for a recent reference to Tsallis's argument, see [10]), i.e., that in the uniform case where the probabilities are $p_i = 1/W$ for all $i = 1, 2, \ldots, W$, the entropy $S_q \propto N$ for $n \to \infty$, where N denotes the number of components in the system under analysis. For clarity, we note that when considering physical systems, the entropy may become volume-dependent; for example, because the number of states available, W, depends on the volume. Volume dependence can also enter through the probabilities p_i^* determined by the maximum entropy principle.

Although the Tsallis entropy does not fulfill axiom SK4 in its original form and hence is *non-additive*, it does satisfy a composition relation different from addition.

Another set of non-Boltzmann–Gibbs–Shannon entropies was derived by Hanel and Thurner [11] by simply discarding axiom SK4 and then determining the functional form from the asymptotic behavior of the number of events, or states, W, as a function of the number of components in the system. However, in this approach, there is no systematic rule for handling the computation of the entropy of a system consisting of independent parts.

It is well known that in physics, the investigation of required symmetries has often been helpful. Think of Einstein's use of the symmetry between different reference frames to derive special and general relativity theory. Consider also the eightfold way and the derivation of QCD. Additionally, consider the application of symmetry and group theory to atomic spectra. Therefore, it seems natural to, rather than discarding the fourth SK axiom, replace it in a way informed by the symmetry properties (and related group-theoretic restrictions) that an entropy must necessarily satisfy. Consequently, the question is, which symmetry cannot be ignored when dealing with entropies?

The fourth SK axiom addresses how the entropy of a system AB, consisting of two independent parts A and B, can be expressed as a sum of the entropy of A and the entropy of B. Tsallis entropy also allows the entropy of AB to be expressed in terms of the entropy of the two parts, not as a sum but as a generalized combination of the entropies of the parts.

The notion of group entropy, introduced in [12,13], goes one step further and exploits the idea that the process of combining two independent systems can be seen as a group operation. The group entropies satisfy the three first SK axioms, as well as a fourth

one, which consists of a generalized composition procedure making use of formal group theory [14,15].

This approach leads to axiomatically defined entropies, whose composition rule is defined in terms of the generator $G(t)$ of a suitable formal group law, namely

$$S[p] = \frac{G\left(\ln \sum_{i=1}^{W} p_i^{\alpha}\right)}{1 - \alpha} \quad \text{with } \alpha > 0 \text{ and } \alpha \neq 1. \tag{3}$$

Although this restricts the allowed functional forms available for an entropy, it does not uniquely determine the entropy as the four SK axioms do. Below, we will discuss how the analysis of combining independent systems using formal group theory, together with requiring extensivity, allows for a systematic classification of entropies in "universality classes". These classes are defined by taking into account how fast the number of available states W grows with the number of components N. We accomplish this by starting with a more general functional non-trace form than the one given in Equation (2). For details, see [12].

By generalizing the Tsallis functional form and requiring composability together with extensivity, we are able to regard Tsallis entropy as the composable entropy associated with systems where interdependence between their components forces W to grow slowly, namely as a power of N. Below, we will discuss why composability on the entire probability space is an indispensable property of an entropy. We will also address the need for extensivity in a general sense. We will point out that extensivity can be very relevant even beyond the thermodynamic need for a well-defined limit as the number of components approaches infinity. For example, extensivity is essential for using an entropy as a measure of the complexity of a time series or for handling power-law probability distributions.

2. Why Composability

The need for composability does not arise because real systems can always be regarded as a simple combination of subsystems. The requirement is, in a sense, a logical necessity [16]. When we consider two independent systems with state spaces A and B, we should obtain the same result if we compute the entropy of the Cartesian combined system $A \times B$ as if we first compute the entropy of A and B separately and then afterward decide to consider them as one combined system. By Cartesian combination, we mean that the system $A \times B$ is given by the set of states $\{(a,b)|a \in A, b \in B\}$, with the probabilities for the individual states given by $p(a,b) = p(a)p(b)$. This Cartesian combination immediately suggests familiar properties from group theory. The composition is as follows:

Commutative: Since we have state spaces in mind, we consider $A \times B = B \times A$. The ordering is immaterial.
Associative: $A \times (B \times C) = (A \times B) \times C$.
"Neutral" element: $A \times B \sim A$ if $B = \{b\}$. In other words, $A \times B$ is essentially the same set as A if B consists of one element only. In terms of state spaces, all sets with one state only are considered to be identical, that is, indistinguishable. In this sense, a unique "neutral" element exists in our composition process. Accordingly, we want the entropy of a probability distribution on a set containing a single element to be zero: indeed, it would correspond to a certainty configuration. Moreover, we want the entropy of $A \times B$ to be equal to the entropy of A if the entropy of B is zero.

The group structure of the Cartesian combination of the event spaces for systems must also be satisfied by the entropy functional operating on the corresponding probability spaces. This can be ensured by employing formal group theory [16]. Define the entropy using the expression in Equation (3), where the function $G(t) = t + \sum_{k=2}^{\infty} \beta_k t^k$ is the "group generator". Here, the formal power series G(t) is said to be the group exponential in the formal group literature [15]. The combination of independent systems is now expressed as

$$S(A \times B) = \phi(S(A), S(B)), \tag{4}$$

where the function $\phi(x,y)$ is given by $\phi(x,y) = G(G^{-1}(x) + G^{-1}(y))$. Given a formal group law, namely when $\phi(x,y)$ is a formal power series in two variables, it is possible to prove that there exists a one-variable formal power series $\psi(x)$ such that $\phi(x, \psi(x)) = 0$. This series represents the formal inverse and completes the group structure.

3. Why Extensivity

Let us first recall the definitions of extensivity and additivity. We say that an entropy is extensive if the entropy per component is finite in the limit of infinitely many components, i.e.,

$$\lim_{N \to \infty} S(N)/N = \text{constant} < \infty \tag{5}$$

An entropy is additive if, for two statistically independent systems A and B, the entropy of the two systems considered as a combined system is equal to the sum of the entropies, i.e.,

$$S(A + B) = S(A) + S(B). \tag{6}$$

When considering the thermodynamics of macroscopic systems with the number of constituents, N, of the order of Avogadro's number, the usual procedure is to compute quantities such as the thermodynamic free energy F for an arbitrary value of N. The limit of large, essentially infinite systems is then handled by considering intensive quantities, e.g., the free energy per constituent. Hence, for an entropy to be thermodynamically useful, it needs to be extensive, given the fundamental thermodynamic relation $F = E - TS$. Since the temperature T is an intensive quantity, the entropy must be extensive. Thus, we need the limit $\lim_{N \to \infty} S(N)/N$ to be well defined.

Outside thermodynamics, entropy finds a significant application within information theory as a tool to characterize the complexity of a deterministic or random process generating a time series. More precisely, we can associate with a time series an ordinal representation formed by all ordinal patterns of length $L \in \mathbb{N}$ assigned [17]. Assuming that all different patterns are allowed for a process, we have $W(L) = L!$, and each pattern i will occur with a probability of $p_i = 1/L!$. The Boltzmann–Gibbs–Shannon entropy in Equation (1) is given by $S_{\text{Shan}}[p] = \ln L! \simeq L \ln L - L$. So, we obtain a diverging entropy rate $S[p]/L$ as the length of the time series increases. As we will see, this is a common situation since random processes exhibit super-exponential growth in the number of permitted patterns. Again, extensivity enters into play. Thus, we would need an entropy that grows proportionally to the number of allowed patterns in the considered time series.

The widespread occurrence of power-law probability distributions in nature, either exact or approximate, has long been the focus of self-organized criticality (for an overview, see [18,19]). It is now clear that power-law distributions with fat tails are common, and for this reason, it seems natural to consider the extent to which the workhorse of information theory, the Shannon entropy, can be used as a meaningful entropic measure for such distributions.

Consider a probability distribution of the following form:

$$P_S(s) = \frac{A}{s^a} \quad \text{for} \quad s = 1, 2, \dots, s(N)_{\text{max}}. \tag{7}$$

Here, A is a normalization factor and a is a positive exponent. The variable s denotes the "size" of some process, e.g., an avalanche or a structure such as a spatial cluster. When $s(N)_{\text{max}}$ grows with N, the usual Boltzmann–Gibbs–Shannon entropy will, in general, not allow a well-defined limit $S[P_S](N)/N$ as $N \to \infty$.

4. The Structure of the Group Entropies

Here, we explain why the expression in Equation (3) is a good starting point for deriving generalized entropies. First, we address why we choose the argument of the

generating function $G(t)$ to be $\ln \sum_i p_i^\alpha$. We also comment on the so-called trace form of the group entropies given by

$$S[p] = \sum_{i=1}^{W} p_i G(\ln \frac{1}{p_i}). \tag{8}$$

Finally, we briefly recapitulate how the functional form of $G(t)$ is determined by reference to formal group theory and the requirement that $S[p]$ is extensive on the uniform (also denoted as the microcanonical) ensemble, given by

$$p_i = \frac{1}{W(N)}, \quad \text{for } i = 1, 2 \ldots, W(N). \tag{9}$$

The structure of Equation (3) is used as the starting point because $G(t)$ being a group generator ensures composability for all distributions p_i, not only the uniform distributions. And, taking the argument to be $\ln \sum_i p_i^\alpha$ enables this functional form to generate a range of well-known entropies, including Boltzmann–Gibbs–Shannon, Rényi, and Tsallis [12]. More specifically, if one chooses

$$G(t) = \frac{e^{at} - e^{bt}}{(a-b)(\alpha - \beta)} \tag{10}$$

one recovers the Boltzmann–Gibbs–Shannon entropy in the limit $\alpha \to 1$; Rényi's entropy in the double limit $a \to 0, b \to 0$; and Tsallis's entropy in the double limit $a \to 1, b \to 0$.

4.1. Extensivity and the Group Law $G(T)$

Let us now briefly describe how the requirement of extensivity determines the group law $G(t)$ in Equation (3). Details can be found in [20,21]. For a given dependence of the number of available states $W(N)$, we want to ensure that the entropy given in Equation (3) is extensive, i.e., that on the uniform ensemble $p_i = 1/W(N)$ for $i = 1, \ldots, W(N)$ we have $\lim_{N \to \infty} S[p]/N = \text{constant}$.

We can express this as

$$S\left(p_i = \frac{1}{W}\right) = \lambda N. \tag{11}$$

Asymptotically, we have

$$S\left(\frac{1}{W}\right) = \frac{G(\ln(W^{1-\alpha})}{1-\alpha} \approx \lambda N. \tag{12}$$

Then, we invert the relation between S and G, which, by Equation (12), amounts to inverting the relation between G and N. For $G(t)$ to generate a group law, we must require $G(0) = 0$ [12,16], so we adjust the expression for $G(t)$ accordingly and conclude that

$$G(t) = \lambda(1-\alpha)\{W^{-1}[\exp(\frac{t}{1-\alpha})] - W^{-1}(1)\}. \tag{13}$$

Hence, given the asymptotic behavior of $W(N)$, we derive different corresponding entropies. In the expressions below, $\lambda \in \mathbb{R}_+$, $\alpha > 0$, and $\alpha \neq 1$ are free parameters.
Non-trace-form case:
(I)   Algebraic, $W(N) = N^a$

$$S[p] = \lambda \left\{ \exp\left[\frac{\ln(\sum_{i=1}^{W(N)} p_i^\alpha)}{a(1-\alpha)} \right] - 1 \right\}. \tag{14}$$

(II)   Exponential, $W(N) = k^N$

$$S[p] = \frac{\lambda}{\ln k} \frac{\ln(\sum_{i=1}^{W(N)} p_i^\alpha)}{1-\alpha}. \tag{15}$$

This is, of course, the Rényi entropy.

(III) Super-exponential, $W(N) = N^{\gamma N}$

$$S[p] = \lambda \left\{ \exp\left[L\left(\frac{\ln \sum_{i=1}^{W(N)} p_i^{\alpha}}{\gamma(1-\alpha)} \right) \right] - 1 \right\}. \tag{16}$$

This entropy was recently studied in relation to a simple model in which the components can form emergent paired states in addition to the combination of single-particle states [22].

So far, we have only considered the so-called non-trace form of the group entropies given in Equation (3). A set of entropies can be constructed in the same manner, starting with the trace-form ansatz in Equation (8).

4.2. Trace-Form Group Entropies

It is interesting to observe that the ansatz in Equation (8) directly leads to either the Boltzmann, the Tsallis, or an entirely new entropy, depending on the asymptotic behavior of $W(N)$. By applying the procedure described in Section 4.1, we obtain the following three classes corresponding to the ones considered for the non-trace case.

Trace-form case:

(I)   Algebraic, $W(N) = N^a$

$$S[p] \quad = \lambda \sum_{i=1}^{W(N)} p_i \left[\left(\tfrac{1}{p_i} \right)^{\frac{1}{a}} - 1 \right] \tag{17}$$

$$= \tfrac{1}{q-1} \left(1 - \sum_{i=1}^{W(N)} p_i^q \right). \tag{18}$$

To emphasize the relation with the Tsallis q-entropy, we have introduced $q = 1 - 1/a$ and $\lambda = 1/(1-q)$. Note that the parameter q is determined by the exponent a, so it is controlled entirely by $W(N)$.

(II) Exponential, $W(N) = k^N, k > 0$

$$S[p] = \frac{\lambda}{\ln k} \sum_{i=1}^{W(N)} p_i \ln \frac{1}{p_i}. \tag{19}$$

This is the Boltzmann–Gibbs–Shannon entropy.

(III) Super-exponential, $W(N) = N^{\gamma N}, \gamma > 0$

$$S[p] = \lambda \sum_{i=1}^{W(N)} p_i \left\{ \exp\left[L\left(-\frac{\ln p_i}{\gamma} \right) \right] - 1 \right\}. \tag{20}$$

4.3. Examples of Systems and Corresponding Group Entropies

To illustrate the classification of group entropies based on the asymptotic behavior of $W(N)$, we consider three Ising-type models:

(a) The Ising model on a random network [11].

(b) The usual Ising model, for example, with nearest-neighbor interaction on a hyper-cubical lattice.

(c) The so-called pairing model in which Ising spins can form paired states [22].

Let E denote the total energy of the system. We are interested in the asymptotic behavior of the number of possible states for the three models as a function of N for fixed energy per component $\epsilon = E/N$. First, consider (a). As explained in [11], $W(N) \sim N^a$ when the fraction of interaction links, the connectance, in the considered network, is kept constant as the number of nodes N is increased. The exponent a is given by the ratio between the

energy density and the connectance. The entropy corresponding to this functional form of $W(N)$ is, for all values of the exponent a, given by the Tsallis entropy [21].

The entropy corresponding to the standard Ising model (case (b)) with $W(N) = 2^N$ is the Boltzmann–Gibbs–Shannon entropy. The pairing version of the Ising model (case (c)) admits a super-exponential growth in the number of states $W(N) \sim N^{\gamma N}$, leading us to a new functional form of the entropy [22]

$$S_{\gamma,\alpha}[p] = \exp\left[L\left(\frac{\ln \sum_{i=1}^{W} p_i^{\alpha}}{\gamma(1-\alpha)}\right)\right] - 1. \tag{21}$$

5. Group Entropies and the Ubiquity of the Q-Exponential Distribution

It is well known that the q-exponential form relating to the Tsallis q-entropy provides a very good fit to an impressively broad range of data sets (see, e.g., [10]). This may, at first, appear puzzling given that we saw in Section 4.1 that the Tsallis entropy corresponds to one of the three classes considered here, namely systems with strong interdependence between the components that $W(N) \sim N^a$. The reason that the q-exponential appears to be much more pervasive than one would expect, given that the q-entropy is restricted to the case $W(N) \sim N^a$, may be due to the following.

Consider the maximum entropy principle. For all the classes of entropies considered in Section 4, the probability distribution that maximizes the entropy is a q-exponential. The probability distribution for the specific case of $W(N) \sim k^N$ is the usual exponential Boltzmann distribution. But since the Boltzmann distribution is the limiting case of the q-exponential for $q \to 1$, we can say that, independently of the asymptotic behavior of $W(N)$, the maximum entropy principle always leads to q-exponential distributions [21].

How can this be? The reason is the functional form of the argument

$$x \equiv \ln \sum_{i=1}^{N} p_i^{\alpha}$$

of the ansatz in Equation (3). When one applies Lagrange multipliers and extremizes the entropy in Equation (3), the q-exponential functional form will arise from the derivative $\partial x / \partial p_i$. The remaining factors in the expression for the derivative of $S[p]$ will depend on the functional form of the group law $G(t)$ but will formally just be a constant if evaluated on the maximizing distribution p^* and do not depend explicitly on p_i.

6. An Entropic Measure of Complexity

Fully interacting complex systems possess a number of microstates $W(N)$ that may be different from the Cartesian exponential case $W(N) = \Pi_{i=1}^{N} k_i$, where k_i is the number of states available to component number i in isolation. When interactions freeze out states, $W(N)$ can grow slower than exponentially with increasing N. In contrast, when interactions allow for the creation of new states from the combination of components, $W(N)$ can grow faster than exponentially. As an example, think of hydrogen atoms that form hydrogen molecules $H + H \to H_2$. The states of H_2 are not just the Cartesian product of free single hydrogen atomic states.

The possible difference between the cardinality of the state space of the fully interacting system and the state space formed as a Cartesian product of the states available to the individual components can be used to construct a new measure of the degree of emergent interdependence among the components of a complex system. We can think of this as a quantitative measure of the degree of complexity in a given system. We imagine the entire system AB to be divided into two parts, A and B, and compare the entropy of the system $A \times B$, obtained by combining the micro-sates of the two parts as a Cartesian product, with the system AB, obtained by allowing full interaction between the components of A and

those of B. We denote by AB this fully interacting system. The complexity measure is given by [20]

$$\Delta(AB) = S(A \times B) - S(AB) = \phi(S(A), S(B)) - S(AB). \qquad (22)$$

From the dependence of $\Delta(AB)$ on the number of components in the separate systems A and B, one can, in principle, determine the kind of emergence generated by the interactions in a specific complex system. In [20], we conjectured that the number of available states for the brain grows faster than exponentially in the number of brain regions involved. It might, at first, appear impossible to check this conjecture. However, experiments like those conducted on rat brain tissue, such as the famous avalanche experiment by Beggs and Plentz [23], seem to open up the possibility for a study of $\Delta(AB)$ as a function of tissue size. We imagine it would be possible to study a piece of tissue of size N and one of size $2N$, allowing, at least in principle, to determine how $\Delta(AB)$ behaves for such a neuronal system. A different, although related, notion of complexity, the defect entropy, was proposed in [24].

7. Group Entropy Theory and Data Analysis

The theory of group entropies has recently proved to be relevant in data analysis. One important reason for this relevance is extensivity. When the number of patterns that may occur in a given time sequence depends, in a non-exponential way, on the length L of the sequence, the Shannon-based entropy of the sequence $S(L)$ will not permit a well-defined entropy rate $S(L)/L$ because the Shannon entropy will not be extensive in L. This may, for example, pose a problem for the widely used Lempel–Ziv [25] complexity measure. This is similar to the discussion above concerning how the Boltzmann-Gibbs-Shannon entropy fails to be extensive on state spaces that grow non-exponential in the number of constituents. We will see below that time series very often contain a number of patterns that grow super-exponentially in the length of the sequence.

To discuss this fundamental application of group entropies, we start with a brief review of the ordinal analysis of time-series data. We follow the discussion and notations in [17,26,27]. Consider the time series

$$(x_t)_{t \geq 0} = x_0, x_1, \ldots, x_t, \ldots$$

where t represents a discrete time and $x_t \in \mathbb{R}$. Let $L \geq 2$. We introduce the sequence of length L (or L-sequence)

$$x_t^L := x_t, x_{t+1}, \ldots, x_{t+L-1}$$

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

$$x_{t+\rho_0} < x_{t+\rho_1} < \ldots < x_{t+\rho_{L-1}}. \qquad (23)$$

We denote the rank vector of the sequence x_t^L as follows:

$$\mathbf{r}_t := (\rho_0, \rho_1, \ldots, \rho_{L-1}), \qquad (24)$$

The rank vectors $\mathbf{r}_t$ are called *ordinal patterns* of length L (or L-ordinal patterns). The sequence x_t^L is said to be "of type" $\mathbf{r}_t$. In this way, given the original time series $(x_t)_{t \geq 0}$, we have constructed an *ordinal representation* associated with it: the family of all ordinal patterns $(\mathbf{r}_t)_{t \geq 0}$ of length L.

We denote by S_L the group of the $L!$ permutations of $0, 1, \ldots, L-1$, which represents the set of symbols (also called "alphabet") of the ordinal representation.

In the following, we consider discrete-time stationary processes $X = (X_t)_{t \geq 0}$, both deterministic and random, taking values in a closed interval $I \subset \mathbb{R}$. We define a "deterministic process" as a "one-dimensional dynamical system" $(I, \mathcal{B}, \mu, f)$, where I is the state space (a bounded interval of $\mathbb{R}$), $\mathcal{B}$ is the "Borel σ-algebra" of I, μ is a "measure" such that $\mu(I) = 1$, and $f : I \to I$ is a μ-invariant map. In this case, the image $(X_t)_{t \geq 0}$ of 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 $f^t(X_0) = f(f^{t-1}(X_0))$.

First, we associate with an ordinal representation the probability $p(\mathbf{r})$ of finding an ordinal pattern of a given rank $\mathbf{r} \in \mathcal{S}_L$. To this aim, we assume the *stationary condition*: for $k \leq L - 1$, the probability of $X_t < X_{t+k}$ cannot depend on t. This condition ensures that estimates of $p(\mathbf{r})$ converge as data increase. Non-stationary processes with stationary increments, such as the fractional Brownian motion and the fractional Gaussian noise, satisfy the condition above.

7.1. Metric and Topological Permutation Entropy

Let $\mathbf{X}$ be a deterministic or random process taking real values. Let $p(\mathbf{r})$ be the probability of a sequence X_t^L generated by $\mathbf{X}$ being of type $\mathbf{r}$, and let $p = p(\mathbf{r})$ be the corresponding probability distribution. We define the following:

(i)   If $p(\mathbf{r}) > 0$, then $\mathbf{r}$ is a permitted pattern for X.
(ii)   If $p(\mathbf{r}) = 0$, then $\mathbf{r}$ is a forbidden pattern.

The *permutation metric entropy of order L* of p is defined as

$$H^*(X_0^L) = - \sum_{\mathbf{r} \in \mathcal{S}_L} p(\mathbf{r}) \ln p(\mathbf{r}). \tag{25}$$

The topological entropy of order L of the finite process X_t^L, $H_0^*(X_t^L)$ is the upper limit of the values of the permutation metric entropy of order L. Formally, we obtain it by assuming that all allowed patterns of length L are equiprobable:

$$H_0^*(X_t^L) := \ln \mathcal{A}_L(\mathbf{X}), \tag{26}$$

where $\mathcal{A}_L(\mathbf{X})$ is the number of allowed patterns of length L for $\mathbf{X}$. It is evident that the following inequalities hold:

$$H^*(X_0^L) \leq \ln \mathcal{A}_L(\mathbf{X}) \leq \ln L!$$

We observe that $\mathcal{A}_L(\mathbf{X}) = L!$ if all L-ordinal patterns are allowed.

7.2. Bandt–Pompe Permutation Entropy

In their seminal paper [28], Bandt and Pompe introduced the following notions:

- Permutation metric entropy of $\mathbf{X}$:

$$h_M(\mathbf{X}) := \limsup_{L \to \infty} \frac{1}{L} H^*(X_0^L) = - \limsup_{L \to \infty} \frac{1}{L} \sum_{\mathbf{r} \in SL} p(\mathbf{r}) \ln p(\mathbf{r}),$$

  where $X_0^L = X_0, \ldots, X_{L-1}$.
- Topological permutation entropy of $\mathbf{X}$:

$$h_T(\mathbf{X}) := \limsup_{L \to \infty} \frac{1}{L} H_0^*(X_0^L) = \limsup_{L \to \infty} \frac{1}{L} \ln \mathcal{N}_L(\mathbf{X}).$$

An important question is, what is the relationship between the permutation metric entropy and the standard *Kolmogorov–Sinai (KS) entropy* of a map?

Let f be a strictly piecewise monotone map on a closed interval $I \subset \mathbb{R}$. The vast majority of, if not all, one-dimensional maps used in concrete applications belong to the class of piecewise monotone maps. In [29], it was proved that

$$h_M(f) = h_{\mathrm{KS}}(f).$$

The same relation holds for the topological versions of the two entropies. This is a fundamental result since it allows us to compute the KS entropy using the ordinal analysis

approach. The above theorem and its generalizations imply that the number of permitted patterns of length L for a deterministic process grows exponentially as L increases.

$$|\text{L-permitted patterns for deterministic } \mathbf{X} = f| \sim e^{\tau(f)L},$$

where $\tau(f)$ is the topological KS entropy. In turn, this implies that the number of prohibited patterns grows super-exponentially.

At the other extreme, we have random processes without prohibited patterns. An elementary example is *white noise*. According to Stirling's formula,

$$|L\text{-possible patterns }| = L! \sim e^{L \ln L}.$$

It is also worth noting that noisy deterministic time series may not have prohibited patterns.

For example, in the case of dynamics on a non-trivial attractor where the orbits are dense, observational white noise will "destroy" all prohibited patterns, regardless of how little the noise is.

In general, *random processes exhibit super-exponential growth in permitted patterns*.

Random processes can also have prohibited patterns. In this case, the growth will be "intermediate," meaning it is still super-exponential but subfactorial.

7.3. A Fundamental Problem

For random processes without prohibited patterns, the permutation entropy diverges in the limit as $L \to \infty$:

$$h_T(\mathbf{X}) = \lim_{L \to \infty} \frac{1}{L} \ln|L\text{-possible permitted patterns}| = \lim_{L \to \infty} \frac{1}{L} \ln L! = \lim_{L \to \infty} \ln L = \infty$$

Also, in general, $h^M(\mathbf{X}) = \infty$. Therefore, it is natural to consider the problem of extending the notion of permutation entropy to make it an intrinsically finite quantity. We assume that for a random process $\mathbf{X}$,

$$|L\text{-possible permitted patterns for } \mathbf{X}| \sim e^{g(L)}.$$

where $g(L)$ is a certain function that depends on the type of process considered.

Can we find a suitable, generalized permutation entropy that converges as $L \to \infty$?

7.4. Group Entropies and Ordinal Analysis

We can obtain a new solution to this problem through the theory of group entropies.

Philosophy: Instead of using the Shannon-type permutation entropy introduced by Bandt and Pompe as a universal entropy valid for all random processes, we will adapt our entropic measure to the specific problem we wish to address:

- We will classify our processes into *complexity classes*, defined by complexity functions $g(t)$. These classes, in ordinal analysis, represent a notion entirely analogous to the *universality class* described earlier (inspired by statistical mechanics).
- Each complexity class will correspond to a *group permutation entropy*, i.e., a specific information measure designed for the class under consideration.
- This measure will be convergent as $L \to \infty$.

Functions and Complexity Classes

A process $\mathbf{X}$ is said to belong to the complexity class g if

$$\ln \underbrace{|\text{allowed } L\text{-patterns for } \mathbf{X}|}_{\mathcal{A}_L(\mathbf{X})} \sim g(L) \text{ for } L \to \infty.$$

The bi-continuous function $g(t)$ is called the *complexity function* of **X**. The process **X** belongs to the *exponential class* if

$$g(L) = cL \quad (c > 0)$$

X belongs to the *factorial class* if

$$g(L) = L \ln L$$

Example 1. *A deterministic process* **X** *belongs to the exponential class. A random process* **X** *like white noise (***X*** i.i.d.) belongs to the factorial class.*
A process **X** *belongs to the subfactorial class if one of the following conditions holds:*

(i)

$$g(L) = o(L \ln L)$$

(ii)

$$g(L) = cL \ln L; \ with \ ; 0 < c < 1$$

Example 2. *Processes with*

$$g(L) = L \ln^{(k)} L; ; (\ln^{(k)} L \equiv \underbrace{\ln \circ \ln \circ \cdots \circ \ln}_{k \ times}(L)) \ with \ k \geq 2$$

belong to subfactorial class (i). Processes with $g(L) = cL \ln L$, $0 < c < 1$ can also be constructed explicitly.

7.5. Group Permutation Entropy

Main Result: The conventional permutation entropy of Bandt–Pompe can be consistently generalized. According to our philosophy, *the complexity class g "dictates" its associated permutation entropy*, which becomes finite in the limit of large L.

Definition 1. *The group entropy of order L for a process* **X** *of class g is*

$$Z_{g,\alpha}^*(\mathrm{p}L) = g^{-1}(R\alpha(\mathrm{p}_L)) - g^{-1}(0)$$

where p_L is the probability distribution of the L-ordinal patterns of $X0^L = X_0, \ldots, X_{L-1}$ and $R_\alpha(\mathrm{p}_L)$ is the Rényi entropy. The corresponding topological group entropy of order L is

$$Z_{g,0}^*(\mathrm{p}L) = g^{-1}(\ln \mathcal{A}L(\mathbf{X})) - g^{-1}(0)$$

The group metric permutation entropy is

$$z_{g,\alpha}^*(\mathbf{X}) = \lim_{L \to \infty} \frac{1}{L} g^{-1}(R_\alpha(\mathrm{p}_L))$$

The topological group permutation entropy is

$$z_{g,0}^*(\mathbf{X}) = \lim_{L \to \infty} \frac{1}{L} g^{-1}(\ln \mathcal{A}_L(\mathbf{X}))$$

The functions defined in this manner are group entropies. Furthermore, they satisfy the inequalities

$$0 \leq z_{g,\alpha}^*(\mathbf{X}) \leq z_{g,0}^*(\mathbf{X}) = 1 \quad \forall \alpha > 0.$$

The following are various examples of group permutation entropies:

(a) For $g_{exp}(t) = ct:$ $Z^*_{g_{exp},\alpha}(\mathrm{p}L) = \frac{1}{c} R\alpha(\mathrm{p}_L)$

(b) For $g_{fac}(t) = t \ln t:$ $Z^*_{g_{fac},\alpha}(\mathrm{p}L) = e^{\mathcal{L}[R\alpha(\mathrm{p}_L)]} - 1$

(c) For $g_{sub}(t) = ct \ln t$ $(0 < c < 1):$ $Z^*_{g_{sub},\alpha}(\mathrm{p}L) = e^{\mathcal{L}[R\alpha(\mathrm{p}_L)/c]} - 1$ (27)

8. Thermodynamics

The application of non-Boltzmann–Gibbs–Shannon entropies to thermodynamics is subtle. We recall that in standard thermodynamics, it is possible to interpret the Lagrange multiplier corresponding to the constraint on the average energy as the inverse of the physical temperature. However, it is not clear if a similar procedure can be adopted for any generalized entropy.

One can certainly derive the probability weights p_i^* corresponding to the extrema of

$$J = S - \lambda_1 (\sum p_i - \mathcal{N}) - \lambda_2 (\sum_i E_i p_i - E) \tag{28}$$

and we can compute the entropy for these weights $S[p_i^*]$. However, given an arbitrary generalized entropy S, we do not know if, for some physical systems, there exists a relationship between $S[p_i^*]$ and Clausius's thermodynamic entropy defined in terms of heat flow. Hence, to us, the relationship between generalized entropies and thermodynamics in the sense of a theory of heat and energy of physical systems, apart from several interesting analogies, remains an open field of research. Detailed discussions concerning the construction of generalized thermostatistics for the case of the Tsallis entropy S_q are available, e.g., in the monographs in [30,31].

9. Discussion

The group entropy formalism described has the pleasant property that all group entropies arise systematically and transparently from a set of underlying axioms combined with the requirement of extensivity. This approach is in contrast to those adopted to define many of the existing entropies, which, sometimes, are intuitively proposed or justified by axioms that ignore the need for composability. Many of the most commonly used entropies are included and classified within the group theoretic framework.

The use of information measures adapted to the universality classes of systems, which are extensive by construction, looks promising in several application contexts, such as the study of neural interconnections in the human brain, classical and quantum information geometry, and data analysis in a broad sense. We plan to further investigate complex systems with super-exponentially growing state spaces as a paradigmatic class of examples where these new ideas can be fruitfully tested.

10. Conclusions

We have reviewed a group-theoretic approach to the classification and characterization of entropies, regarded as functionals on spaces of probability distributions. The theoretical framework proposed is axiomatic and generalizes the set of Shannon–Khinchin axioms by replacing the fourth additivity axiom with a more general composition axiom. Perhaps the most relevant achievement so far is the systematic classification of the multitude of existing entropies in terms of the rate at which the corresponding dimension of the state space grows with the number of components in the system. A related result is a constructive procedure for entropies, which exhibit extensivity on state spaces of any assigned growth rate. In turn, this property triggers the application of group entropies to information geometry and data analysis.

References

1. Reif, F. *Fundamentals of Statistical and Thermal Physics*; Waveland Press: Long Grove, IL, USA, 2010.
2. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; JohnWiley & Sons Ltd.: Hoboken, NJ, USA, 2005.
3. Ziv, J.; Lempel, A. Compression of Individual sequences via Variable-Rate Coding. *IEEE Trans. Inform. Theory* **1978**, *24*, 530–536. [CrossRef]
4. Takaoka, T.; Nakagawa, Y. Entropy as computational complexity. *J. Inf. Process.* **2010**, *18*, 227–241. [CrossRef]
5. Carhart-Harris, R.L.; Leech, R.; Hellyer, P.J.; Shanahan, M.; Feilding, A.; Tagliazucchi, E.; Chialvo, D.R.; Nutt, D. The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* **2014**, *8*, 20. [CrossRef]
6. Oizumi, M.; Albantakis, L.; Tononi, G. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol.* **2014**, *10*, 316. [CrossRef]
7. Jensen, H.J. *Complexity Science: The Study of Emergence*; Cambridge University Press: Cambridge, UK, 2022.
8. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover: New York, NY, USA, 1957.
9. Tsallis, C. Possible Generalization of Boltzmann-Gibbs Statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
10. Tsallis, C. Entropy. *Encyclopedia* **2022**, *2*, 264–300. [CrossRef]
11. Hanel, R.; Thurner, S. When do generalized entropies apply? How phase space volume determines entropy. *Europhys. Lett.* **2011**, *96*, 50003. [CrossRef]
12. Tempesta, P. Formal Groups and Z-Entropies. *Proc. R. Soc. A* **2016**, *472*, 20160143. [CrossRef] [PubMed]
13. Tempesta, P. Group entropies, correlation laws and zeta functions. *Phys. Rev. E* **2011**, *84*, 021121. [CrossRef]
14. Bochner, S. Formal Lie groups. *Ann. Math.* **1946**, 47, 192–201. [CrossRef]
15. Hazewinkel, M. *Formal Groups and Applications*; Academic Press: Cambridge, MA, USA, 1978.
16. Tempesta, P. Beyond the Shannon-Khinchin formulation: The composability axiom and the universal-group entropy. *Ann. Phys.* **2016**, *365*, 180–197. [CrossRef]
17. Amigó, J.M.; Dale, R.; Tempesta, P. Permutation group entropy: A new route to complexity for real-valued processes. *Chaos* **2022**, *32*, 112101. [CrossRef] [PubMed]
18. Jensen, H.J. *Self-Oranized Criticality. Emergent Complex Behavior in Physical and Biological Systems*; Cambridge University Press: Cambridge, UK, 1998.
19. Pruessner, G. *Self-Organised Criticality: Theory, Models and Characterisation*; Cambridge University Press: Cambridge, UK, 2012.
20. Tempesta, P.; Jensen, H.J. Universality Classes and Information-Theoretic Measures of Complexity via Group Entropies. *Nat. Sci. Rep.* **2020**, *10*, 5952. [CrossRef] [PubMed]
21. Jensen, H.J.; Tempesta, P. Group Entropies: From Phase Space Geometry to Entropy Functionals via Group Theory. *Entropy* **2018**, *20*, 804. [CrossRef] [PubMed]
22. Jensen, H.J.; Pazuki, R.H.; Pruessner, G.; Tempesta, P. Statistical mechanics of exploding phase spaces: Ontic open systems. *J. Phys. A Math. Theor.* **2018**, 51, 375002. [CrossRef]
23. Beggs, J.M.; Plenz, D. Neuronal Avalanches in neocortical circuits. *J. Neurosci.* **2003**, *23*, 11167–11177. [CrossRef]
24. Livadiotis, G.; McComas, D.J. Entropy defect in thermodynamics. *Nat.-Sci. Rep.* **2023**, *13*, 9033. [CrossRef]
25. Ziv, J. Coding theorems for individual sequences. *IEEE Trans. Inf. Theory* **1978**, *24*, 405–412. [CrossRef]
26. Amigo, J.M.; Dale, R.; Tempesta, P. Complexity-based permutation entropies: From deterministic time series to white noise. *Commun. Nonlinear Sci. Numer. Simul.* **2022**, *105*, 106077. [CrossRef]
27. Amigó, J.M.; Dale, R.; Tempesta, P. A generalized permutation entropy for noisy dynamics and random processes. *Chaos* **2021**, *31*, 0131115. [CrossRef]
28. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef] [PubMed]
29. Bandt, C.; Keller, K.; Pompe, B. Permutations and the Kolmogorov-Sinai entropy. *Nonlinearity* **2002**, *15*, 1595–1602. [CrossRef]
30. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2023.
31. Naudts, J. *Generalised Thermostatistics*; Springer: Berlin/Heidelberg, Germany, 2011.