entropy

MDPI

Review

A Brief Review of Generalized Entropies

José M. Amigó [1,*], Sámuel G. Balogh [2] and Sergio Hernández [3]

[1] Centro de Investigación Operativa, Universidad Miguel Hernández, Avda. de la Universidad s/n, 03202 Elche, Spain
[2] Department of Biological Physics, Eötvös University, H-1117 Budapest, Hungary; balogh@hal.elte.hu
[3] HCSoft Programación S.L., 30007 Murcia, Spain; sergio@hcsoft.net
* Correspondence: jm.amigo@umh.es

Received: 22 September 2018; Accepted: 19 October 2018; Published: 23 October 2018

Abstract: Entropy appears in many contexts (thermodynamics, statistical mechanics, information theory, measure-preserving dynamical systems, topological dynamics, etc.) as a measure of different properties (energy that cannot produce work, disorder, uncertainty, randomness, complexity, etc.). In this review, we focus on the so-called generalized entropies, which from a mathematical point of view are nonnegative functions defined on probability distributions that satisfy the first three Shannon–Khinchin axioms: continuity, maximality and expansibility. While these three axioms are expected to be satisfied by all macroscopic physical systems, the fourth axiom (separability or strong additivity) is in general violated by non-ergodic systems with long range forces, this having been the main reason for exploring weaker axiomatic settings. Currently, non-additive generalized entropies are being used also to study new phenomena in complex dynamics (multifractality), quantum systems (entanglement), soft sciences, and more. Besides going through the axiomatic framework, we review the characterization of generalized entropies via two scaling exponents introduced by Hanel and Thurner. In turn, the first of these exponents is related to the diffusion scaling exponent of diffusion processes, as we also discuss. Applications are addressed as the description of the main generalized entropies advances.

Keywords: generalized entropy; Tsallis; Rényi; Hanel–Thurner exponents; non-stationary regime

1. Introduction

The concept of entropy was introduced by Clausius [1] in thermodynamics to measure the amount of energy in a system that cannot produce work, and given an atomic interpretation in the foundational works of statistical mechanics and gas dynamics by Boltzmann [2,3], Gibbs [4], and others. Since then, entropy has played a central role in many-particle physics, notoriously in the description of non-equilibrium processes through the second principle of thermodynamics and the principle of maximum entropy production [5,6]. Moreover, Shannon made of entropy the cornerstone on which he built his theory of information and communication [7]. Entropy and the associated entropic forces are also the main character in recent innovative approaches to artificial intelligence and collective behavior [8,9]. Our formalism is information-theoretic (i.e., entropic forms are functions of probability distributions) owing to the mathematical properties that we discuss along the way, but can be translated to a physical context through the concept of microstate.

The prototype of entropy that we are going to consider below is the Boltzmann–Gibbs–Shannon (BGS) entropy,

$$S_{BGS}(p_1,...,p_W) = k\sum_{i=1}^{W} p_i \ln \frac{1}{p_i} = -k\sum_{i=1}^{W} p_i \ln p_i. \quad (1)$$

In its physical interpretation, $k = 1.3807 \times 10^{-23}$ J/K is the Boltzmann constant, W is the number of microstates consistent with the macroscopic constraints of a given thermodynamical system, and p_i is the probability (i.e., the asymptotic fraction of time) that the system is in the microstate i. In information theory, k is set equal to 1 for mathematical convenience, as we do hereafter, and S_{BGS} measures the average information conveyed by the outcomes of a random variable with probability distribution $\{p_1, ..., p_W\}$. We use natural logarithms unless otherwise stated, although logarithms to base 2 is the natural choice in binary communications (the difference being the units, nats or bits, respectively). Remarkably enough, Shannon proved in Appendix B of his seminal paper [7] that Equation (1) follows necessarily from three properties or axioms (actually, four are needed; more on this below).

BGS entropy was later on generalized by other "entropy-like" quantities in dynamical systems (Kolmogorov–Sinai entropy [10], etc.), information theory (Rényi entropy [11], etc.), and statistical physics (Tsallis entropy [12], etc.), to mention the most familiar ones (see, e.g., [13] for an account of some entropy-like quantities and their applications, especially in time series analysis). Similar to with S_{BGS}, the essence of these new entropic forms was distilled into a small number of properties that allow sorting them out in a more systematic way [13,14]. Currently, the uniqueness of S_{BGS} is derived from the four *Khinchin–Shannon axioms* (Section 2). However, the fourth axiom, called the *separability* or *strong additivity axiom* (which implies additivity, i.e., $S(A_1 + A_2) = S(A_1) + S(A_2)$, where $A_1 + A_2$ stands for a system composed of any two probabilistically independent subsystems A_1 and A_2), is violated by physical systems with long-range interactions [15,16]. This poses the question of what mathematical properties have the "generalized entropies" satisfying only the other three axioms. These are the primary candidates for *extensive* entropic forms, i.e., functions S such that $S(B_1 \cup B_2) = S(B_1) + S(B_2)$, the shorthand $B_1 \cup B_2$ standing for the physical system composed of the subsystems B_1 and B_2. Note that $B_1 \cup B_2 \neq B_1 + B_2$ in non-ergodic interacting systems just because the number of states in $B_1 \cup B_2$ is different from the number of states in $B_1 + B_2$. A related though different question is how to weaken the separability axiom to identify the extensive generalized entropies; we come back briefly to this point in Section 2 when speaking of the composability property.

Along with S_{BGS}, typical examples of generalized entropies are the *Tsallis entropy* [12],

$$T_q(p_1, ..., p_W) = \frac{1}{1-q}\left(\sum_{i=1}^{W} p_i^q - 1\right) \tag{2}$$

($q \in \mathbb{R}$, $q \neq 1$, with the proviso that for $q < 0$ terms with $p_i = 0$ are omitted), and the *Rényi entropy* [11],

$$R_q(p_1, ..., p_W) = \frac{1}{1-q} \ln\left(\sum_{i=1}^{W} p_i^q\right) \tag{3}$$

($q \geq 0$, $q \neq 1$). The Tsallis and Rényi entropies are related to the BGS entropy through the limits

$$\lim_{q\to 1} T_q(p_1, ..., p_W) = \lim_{q\to 1} R_q(p_1, ..., p_W) = S_{BGS}(p_1, ..., p_W),$$

this being one of the reasons they are considered generalizations of the BGS entropy. Both T_q and R_q have found interesting applications [15,17]; in particular, the parametric weighting of the probabilities in their definitions endows data analysis with additional flexibility. Other generalized entropies that we consider in this paper are related to ongoing work on graphs [18]. Further instances of generalized entropies are also referred to below.

Let us remark at this point that S_{BGS}, T_q, R_q and other generalized entropies considered in this review can be viewed as special cases of the (h, ϕ)-entropies introduced in [19] for the study of asymptotic probability distributions. In turn, (h, ϕ)-entropies were generalized to quantum information theory in [20]. Quantum (h, ϕ)-entropies, which include von Neumann's entropy [21] as well as the quantum versions of Tsallis' and Rényi's entropies, have been applied, for example, to the detection of quantum entanglement (see [20] and references therein). In this review, we do not consider quantum

entropies, which would require advanced mathematical concepts, but only entropies defined on classical, discrete and finite probability distributions. If necessary, the transition to continuous distributions is done by formally replacing probability mass functions by densities and sums by integrals. For other approaches to the concept of entropy in more general settings, see [22–25].

Generalized entropies can be characterized by two scaling exponents in the limit $W \to \infty$, which we call Hanel–Thurner exponents [16]. For the simplest generalized entropies, which include T_q but not R_q (see Section 2), these exponents allow establishing a relationship between the abstract concept of generalized entropy and the physical properties of the system they describe through their asymptotic scaling behavior in the thermodynamic limit. That is, the two exponents label equivalence classes of systems which are universal in that the corresponding entropies have the same thermodynamic limit. In this regard, it is interesting to mention that, for any pair of Hanel–Thurner exponents (at least within certain ranges), there is a generalized entropy with those exponents, i.e., systems with the sought asymptotic behavior. Furthermore, the first Hanel–Thurner exponent allows also establishing a second relation with physical properties, namely, with the diffusion scaling exponents of diffusion processes, under some additional assumptions.

The rest of this review is organized as follows. The concept of generalized entropy along with some formal preliminaries and its basic properties are discussed in Section 1. As way of illustration, we discuss in Section 3 the Tsallis and Renyi entropies, as well as more recent entropic forms. The choice of the former ones is justified by their uniqueness properties under quite natural axiomatic formulations. The Hanel–Thurner exponents are introduced in Section 4, where their computation is also exemplified. Their aforementioned relation to diffusion scaling exponents is explained in Section 5. The main messages are recapped in Section 6. There is no section devoted to the applications but, rather, these are progressively addressed as the different generalized entropies are presented. The main text has been supplemented with three appendices at the end of the paper.

2. Generalized Entropies

Let $\mathcal{P}$ be the set of probability mass distributions $\{p_1, ..., p_W\}$ for all $W \geq 2$. For any function $H : \mathcal{P} \to \mathbb{R}^+$ ($\mathbb{R}^+$ being the nonnegative real numbers), the *Shannon–Khinchin axioms* for an entropic form H are the following.

SK1 *Continuity*. $H(p_1, ..., p_W)$ depends continuously on all variables for each W.

SK2 *Maximality*. For all W,

$$H(p_1, ..., p_W) \leq H(\tfrac{1}{W}, ..., \tfrac{1}{W}).$$

SK3 *Expansibility*: For all W and $1 \leq i \leq W$,

$$H(0, p_1, ..., p_W) = H(p_1, ..., p_i, 0, p_{i+1}, ..., p_W) = H(p_1, ..., p_i, p_{i+1}, ..., p_W).$$

SK4 *Separability* (or *strong additivity*): For all W, U,

$$\begin{aligned} &H(p_{11}, \ldots, p_{1U}, p_{21}, \ldots p_{2U}, \ldots, p_{W1}, \ldots, p_{WU}) \\ &= H(p_{1\cdot}, p_{2\cdot}, \ldots, p_{W\cdot}) + \sum_{i=1}^{W} p_{i\cdot} H\left(\frac{p_{i1}}{p_{i\cdot}}, \frac{p_{i2}}{p_{i\cdot}}, \ldots, \frac{p_{iU}}{p_{i\cdot}}\right), \end{aligned} \tag{4}$$

where $p_{i\cdot} = \sum_{j=1}^{U} p_{ij}$.

Let $\{p_{11}, \ldots, p_{1U}, p_{21}, \ldots p_{2U}, \ldots, p_{W1}, \ldots, p_{WU}\}$ be the joint probability distribution of the random variables X and Y, with marginal distributions $\{p_{i\cdot} : 1 \leq i \leq W\}$ and $\{p_{\cdot j} = \sum_{i=1}^{W} p_{ij} : 1 \leq j \leq U\}$, respectively. Then, axiom SK4 can be written as

$$H(X, Y) = H(X) + H(Y|X),$$

where $H(Y|X)$ is the entropy of Y conditional on X. In particular, if X and Y are *independent* (i.e., $p_{ij} = p_{i.}p_{.j}$), then $H(Y|X) = H(Y)$ and

$$H(X,Y) = H(X) + H(Y). \tag{5}$$

A function H such that Equation (5) holds (for independent random variables X and Y) is called *additive*. Physicists prefer writing $X+Y$ for composed systems with microstate probabilities $p_{ij} = p_{i.}p_{.j}$; this condition holds approximately only for weakly interacting systems X and Y.

With regard to Equation (5), let us remind that, for two general random variables X and Y, the difference $I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0$ is the mutual information of X and Y. It holds $I(X;Y) = 0$ if and only if X and Y are independent [26].

More generally, a function H such that

$$\begin{aligned} &H(p_1q_1,\ldots,p_1q_U,p_2q_1,\ldots,p_2q_U,\ldots,p_Wq_1,\ldots,p_Wq_U) \\ &= H(p_1,\ldots,p_W) + H(q_1,\ldots,q_U) + (1-\alpha)H(p_1,\ldots,p_W)H(q_1,\ldots,q_U), \end{aligned} \tag{6}$$

($\alpha > 0$) is called *α-additive*. With the same notation as above, we can write this property as

$$H(X,Y) = H(X) + H(Y) + (1-\alpha)H(X)H(Y), \tag{7}$$

where, again, X and Y are independent random variables. In a statistical mechanical context, X and Y may stand also for two probabilistically independent (or weakly interacting) physical systems. If $\alpha = 1$, we recover additivity (Equation (5)).

In turn, additivity and α-additivity are special cases of *composability* [15,27]:

$$H(X,Y) = \Phi(H(X),H(Y)), \tag{8}$$

with the same caveats for X and Y. Here, Φ is a symmetric function of two variables. Composability was proposed in [15] to replace axiom SK4. Interestingly, it has been proved in [27] that, under some technical assumptions, the only composable generalized entropy of the form in Equation (10) is T_q, up to a multiplicative constant.

As mentioned in Section 1, a function $F : \mathcal{P} \to \mathbb{R}^+$ satisfying axioms SK1–SK4 is necessarily of the form $F(p_1,...,p_W) = kS_{BGS}(p_1,...,p_W)$ for every W, where k is a positive constant ([28], Theorem 1). The same conclusion can be derived using other equivalent axioms [14,29]. For instance, Shannon used continuity, the property that $H(1/n,...,1/n)$ increases with n, and a property called *grouping* [29] or *decomposibility* [30], which he defined graphically in Figure 6 of [7]:

$$\begin{aligned} H(p_1,...,p_W) \;=\; & H((p_1+...+p_r),(p_{r+1}+...+p_W)) \\ & +(p_1+...+p_r)H\left(\frac{p_1}{\sum_{i=1}^r p_i},...,\frac{p_r}{\sum_{i=1}^r p_i}\right) \\ & +(p_{r+1}+...+p_W)H\left(\frac{p_{r+1}}{\sum_{i=r+1}^W p_i},...,\frac{p_W}{\sum_{i=r+1}^W p_i}\right) \end{aligned} \tag{9}$$

($1 \leq r \leq W-1$). This property allows reducing the computation of $H(p_1,...,p_W)$ to the computation of the entropy of dichotomic random variables. According to ([15], Section 2.1.2.7), Shannon missed in his uniqueness theorem to formulate the condition in Equation (5), X and Y being independent random variables.

Nonnegative functions defined on $\mathcal{P}$ that satisfy axioms SK1–SK3 are called *generalized entropies* [16]. In the simplest situation, a generalized entropy has the *sum property* [14], i.e., the algebraic form

$$F_g(p_1, ..., p_W) = \sum_{i=1}^{W} g(p_i), \tag{10}$$

with $g : [0, 1] \to \mathbb{R}^+$.

The following propositions are immediate.

(i) *Symmetry*: $F_g(p_1, ..., p_W)$ is invariant under permutation of $p_1, ..., p_W$.

(ii) F_g satisfies axiom SK1 if and only if g is continuous.

(iii) If F_g satisfies axiom SK2, then

$$\sum_{i=1}^{W} g(p_i) \leq W g(\tfrac{1}{W})$$

for all $W \geq 2$ and $p_1, ..., p_W$ with $p_1 + ... + p_W = 1$.

(iv) If g is concave (i.e., $\cap$-convex), then F_g satisfies axiom SK2.

(v) F_g satisfies axiom SK3 if and only if $g(0) = 0$.

Note that Proposition (iv) follows from the symmetry and concavity of F_g (since the unique maximum of F_g must occur at equal probabilities).

We conclude from Propositions (ii), (iv) and (v) that, for F_g to be a generalized entropy, the following three condition suffice:

(C1) g is continuous.

(C2) g is concave.

(C3) $g(0) = 0$.

As in [16], we say that a macroscopic statistical system is *admissible* if it is described by a generalized entropy F_g of the form in Equation (10) such that g verifies Conditions (C1)–(C3). By extension, we say also that the generalized entropy F_g is admissible. Admissible systems and generalized entropies are the central subject of this review. Clearly, S_{BGS} is admissible because

$$g(x) = -x \log x, \tag{11}$$

$0 \leq x \leq 1$. On the other hand, T_q corresponds to

$$g(x) = \frac{1}{1-q}(x^q - x). \tag{12}$$

For T_q to be admissible, Condition (C1) requires $q \geq 0$ and Condition (C3) requires $q > 0$.

An example of a function $F : \mathcal{P} \to \mathbb{R}^+$ with the sum property that does not qualify for admissible generalized entropy is

$$F(p_1, ..., p_W) = \sum_{i=1}^{W} \left(p_i - \frac{1}{W} \right)^2 = \sum_{i=1}^{W} p_i^2 - \frac{1}{W}. \tag{13}$$

Indeed, $g(x) = (x - \frac{1}{W})^2$ is not $\cap$-convex but $\cup$-convex and $g(0) = \frac{1}{W^2} \neq 0$. This probability functional was used in [31] to classify sleep stages.

Other generalized entropies that are considered below have the form

$$F_{G,g}(p_1, ..., p_W) = G\left(\sum_{i=1}^{W} g(p_i) \right), \tag{14}$$

where G is a continuous *monotonic* function, and g is continuous with $g(0) = 0$. By definition, $F_{G,g}$ is also symmetric, and Proposition (iii) holds with the obvious changes. However, the concavity of g is not a sufficient condition any more for $F_{G,g}$ to be a generalized entropy. Such is the case of the Rényi entropy R_q (Equation (3)); here

$$G(u) = \frac{1}{1-q} \ln u \ \text{ and } \ g(x) = x^q, \tag{15}$$

but $g(x)$ (and, hence, $\sum_{i=1}^{W} g(p_i)$) is not $\cap$-convex for $q > 1$. Furthermore, note that axiom SK3 requires $q > 0$ for R_q to be a generalized entropy.

Since Equation (10) is a special case of Equation (14) (set G to be the identity map $id(u) = u$), we can refer to both cases just by using the notation $F_{G,g}$, as we do hereafter.

We say that two probability distributions $\{p_i\}$ and $\{p'_i\}$, $1 \leq i \leq W$, are close if

$$\left\|\{p_i\} - \{p'_i\}\right\| = \sum_{i=1}^{W} |p_i - p'_i| \leq \delta,$$

where $0 < \delta \ll 1$; other norms, such as the two-norm and the max-norm, will do as well since they are all equivalent in the metric sense. A function $F : \mathcal{P} \to \mathbb{R}^+$ is said to be *Lesche-stable* if for all W and $\epsilon > 0$ there exists $\delta > 0$ such that

$$\left\|\{p_i\} - \{p'_i\}\right\| \leq \delta \ \Rightarrow \ \left|\frac{F(\{p_i\}) - F(\{p'_i\})}{F_{\max}}\right| < \epsilon, \tag{16}$$

where $F_{\max} = \max_{\{p_i\} \in \mathcal{P}} F(\{p_i\})$. It follows that

$$\lim_{\delta \to 0} \lim_{W \to \infty} \left|\frac{F(\{p_i\}) - F(\{p'_i\})}{F_{\max}}\right| = 0.$$

Lesche stability is called *experimental robustness* in [15] because it guarantees that similar experiments performed on similar physical systems provide similar results for the function F. According to [16], all admissible systems are Lesche stable.

3. Examples of Generalized Entropies

As way of illustration, we put the focus in this section on two classical generalized entropies as well as on some newer ones. The classical examples are the Tsallis entropy and the Rényi entropy because they have extensively been studied in the literature from an axiomatic point of view too. As it turns out, they are unique under some natural assumptions, such as additivity, α-additivity or composability (see below for details). The newer entropies are related to potential applications of the concept of entropy to graph theory [18]. Other examples of generalized entropies are listed in Appendix A for further references.

3.1. Tsallis Entropy

A simple way to introduce Tsallis' entropy as a generalization of the BGS entropy is the following [15]. Given $q \in \mathbb{R}$, define the q-logarithm of a real number $x > 0$ as

$$\ln_q x = \begin{cases} \ln x & \text{if } q = 1, \\ \frac{x^{1-q}-1}{1-q} & \text{otherwise.} \end{cases}$$

Note that $\ln_1 x$ is defined by continuity since $\lim_{q\to 1} \ln_q x = \ln x$. If the logarithm in the definition of S_{BGS}, Equation (1), is replaced by $\ln_q$, then we obtain the Tsallis entropy:

$$T_q(p_1, ..., p_W) = \sum_{i=1}^{W} p_i \ln_q(1/p_i) = \frac{1}{1-q}\left(\sum_{i=1}^{W} p_i^q - 1\right). \tag{17}$$

As noted before, $q > 0$ for T_q to be an admissible generalized entropy.

Alternatively, the definition

$$S_{BGS}(p_1, ..., p_W) = -\frac{d}{dx}\sum_{i=1}^{W} p_i^x\Big|_{x=1}$$

can also be generalized to provide the Tsallis entropy via the q-derivative,

$$T_q(p_1, ..., p_W) = -D_q \sum_{i=1}^{W} p_i^x\Big|_{x=1},$$

where

$$D_q f(x) := \frac{f(qx) - f(x)}{qx - x}.$$

Set $qx = x + h$, i.e., $h = (q-1)x$, and let $h \to 0$ to check that $D_1 f(x) \equiv \lim_{q\to 1} D_q f(x) = df(x)/dx$.

Although Tsallis proposed his entropy (Equation (17)) in 1988 to go beyond the standard statistical mechanics [12], basically the same formula had already been proposed in 1967 by Havrda and Charvát (with a different multiplying factor) in the realm of cybernetics and control theory [32].

Some basic properties of T_q follow.

- **(T1)** $T_1 = S_{BGS}$ because $\ln_1 p_i = \ln p_i$ (or $D_1 f(x) = df(x)/dx$).
- **(T2)** T_q is (strictly) $\cap$-convex for $q > 0$. Figure 1 plots $T_q(p, 1-p)$ for $q = 0.5$, 1, 2 and 5. Let us mention in passing that T_q is $\cup$-convex for $q < 0$.
- **(T3)** T_q is Lesche-stable for all $q > 0$ [33,34]. Actually, we stated at the end of Section 2 that all admissible systems are Lesche stable.
- **(T4)** T_q is not additive but q-additive (see Equation (6) or (7) with α replaced by q). This property follows from [15]
$$\ln_q xy = \ln_q x + \ln_q y + (1-q)(\ln_q x)(\ln_q y).$$
- **(T5)** Similar to what happens with the BGS entropy, Tsallis entropy can be uniquely determined (except for a multiplicative positive constant) by a small number of axioms. Thus, Abe [35] characterized the Tsallis entropy by: (i) continuity; (ii) the increasing monotonicity of $T_q(1/W, ..., 1/W)$ with respect to W; (iii) expansivity; and (iv) a property involving conditional entropies. Dos Santos [36], on the other hand, used the previous Axioms (i) and (ii), q-additivity, and a generalization of the grouping axiom (Equation (9)). Suyari [37] derived T_q from the first three Shannon–Khinchin axioms and a generalization of the fourth one. The perhaps most economical characterization of T_q was given by Furuichi [38]; it consists of continuity, symmetry under the permutation of $p_1, ..., p_W$, and a property called q-recursivity. As mentioned in Section 2, Tsallis entropy was recently shown [27] to be the only composable generalized entropy of the form in Equation (10) under some technical assumptions. Further axiomatic characterizations of the Tsallis entropy can be found in [39].

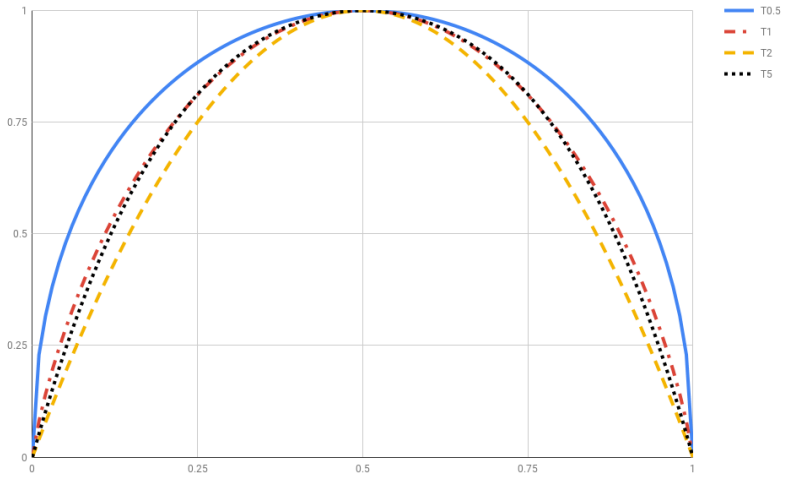

Figure 1. Tsallis entropy $T_q(p, 1-p)$ for $q = 0.5, 1, 2$ and 5.

An observable of a thermodynamical (i.e., many-particle) system, say its energy or entropy, is said to be extensive if (among other characterizations), for a large number N of particles, that observable is (asymptotically) proportional to N. For example, for a system whose particles are weakly interacting (think of a dilute gas), the additive S_{BGS} is extensive, whereas the non-additive T_q ($q \neq 1$) is non-extensive. The same happens with ergodic systems [40]. However, according to [15], for a non-ergodic system with strong correlations, S_{BGS} can be non-extensive while T_q can be extensive for a particular value of q; such is the case of a microcanonical spin system on a network with growing constant connectancy [40]. This is why T_q represents a physically relevant generalization of the traditional S_{BGS}. Axioms SK1–SK3 are expected to hold true also in strongly interacting systems.

Further applications of the Tsallis entropy include astrophysics [41], fractal random walks [42], anomalous diffusion [43,44], time series analysis [45], classification [46,47], and artificial neural networks [48].

3.2. Rényi Entropy

A simple way to introduce Rényi's entropy as a generalization of S_{BGS} is the following [17]. By definition, the BGS entropy of the probability distribution $\{p_1, ..., p_W\}$ (or of a random variable X with that probability distribution) is the *linear* average of the information function

$$I(p_i) = \ln \frac{1}{p_i}, \quad 1 \leq i \leq W,$$

or, equivalently, the expected value of the random variable $\ln \frac{1}{p(X)}$:

$$S_{BGS}(p_1, ..., p_W) = \mathbb{E}_p[\ln \tfrac{1}{p(X)}] = \sum_{i=1}^{W} p_i I(p_i).$$

In the general theory of expected values, for any invertible function ϕ and realizations $x_1, ..., x_W$ of X in the definition domain of ϕ, an expected value can be defined as

$$\mathbb{E}_{p,\phi}[X] = \phi^{-1}\left(\sum_{i=1}^{W} p_i \phi(x_i)\right).$$

Applying this definition to $\ln \frac{1}{p(X)}$, we obtain

$$\mathbb{E}_{p,\phi}[\ln \tfrac{1}{p(X)}] = \phi^{-1}\left(\sum_{i=1}^{W} p_i \phi(I(p_i))\right).$$

If this generalized average has to be additive for independent events, i.e., it has to satisfy Equation (6) with $\alpha = 1$, then

$$\phi(x) = c_1 x \;\text{ or }\; \phi(x) = c_2^{(1-q)x}$$

must hold, where c_1, c_2 are positive constants, and $q > 0, q \neq 1$. The first case leads to S_{BGS}, Equation (1), after choosing $c_1 = e$. The second case leads to the Rényi entropy (actually, a one-parameter family of entropies) R_q, Equation (3), after choosing $c_2 = e$ as well.

Next, we summarize some important properties of the Rényi entropy.

(**R1**) R_q is additive by construction.

(**R2**) $R_1 \equiv \lim_{q\to 1} R_q = S_{BGS}$. Indeed, use L'Hôpital's Rule to derive

$$\begin{aligned}\lim_{q\to 1} \frac{1}{1-q} \ln\left(\sum_{i=1}^{W} p_i^q\right) &= -\lim_{q\to 1} \frac{d}{dq} \ln\left(\sum_{i=1}^{W} p_i^q\right)\\ &= -\lim_{q\to 1} \frac{1}{\sum_{i=1}^{W} p_i^q} \sum_{i=1}^{W} p_i^q \ln p_i\\ &= -\sum_{i=1}^{W} p_i \ln p_i.\end{aligned}$$

(**R3**) R_q is $\cap$-convex for $0 < q \leq 1$ and it is neither $\cap$-convex nor $\cup$-convex for $q > 1$. Figure 2 plots $R_q(p, 1-p)$ for $q = 0.5$, 1, 2 and 5.

(**R4**) R_q is Lesche-unstable for all $q > 0$, $q \neq 1$ [49].

(**R5**) The entropies R_q are monotonically decreasing with respect to the parameter q for any distribution of probabilities, i.e.,

$$q < q' \implies R_q \geq R_{q'}.$$

This property follows from the formula

$$-\frac{dR_q}{dq} = \frac{1}{(1-q)^2} \sum_{i=1}^{W} p_i' \ln \frac{p_i'}{p_i} = \frac{1}{(1-q)^2} D(\{p_i'\} \,\|\, \{p_i\}),$$

where $p_i' = p_i^q / \sum_{k=1}^{W} p_k^q$, and $D(\{p_i'\} \,\|\, \{p_i\}) \geq 0$ is the Kullback–Leibler divergence of the probability distributions $\{p_1', ..., p_W'\}$ and $\{p_1, ..., p_W\}$. $D(\{p_i'\} \,\|\, \{p_i\})$ vanishes only in the event that both probability distributions coincide, otherwise is positive [26].

(**R6**) A straightforward relation between Rényi's and Tsallis' entropies is the following [50]:

$$T_q = \frac{1}{1-q}\left(e^{(1-q)R_q} - 1\right) \;\text{ or }\; R_q = \frac{1}{1-q} \ln\left(1 + (1-q)T_q\right).$$

However, the axiomatic characterizations of the Rényi entropy are not as simple as those for the Tsallis entropy. See [27,51,52] for some contributions in this regard.

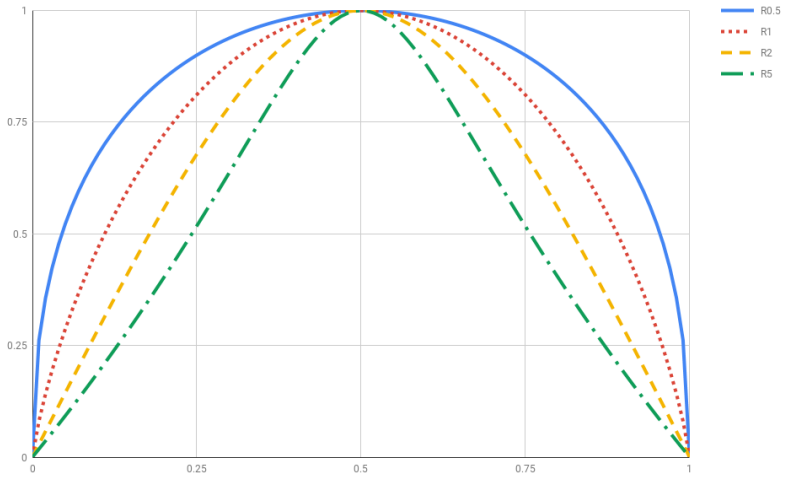

Figure 2. Rényi entropy $R_q(p, 1-p)$ for $q = 0.5, 1, 2$ and 5.

For some values of q, R_q has particular names. Thus, $R_0 = \ln W$ is called *Hartley* or *max-entropy*, which coincides numerically with S_{BGS} for an even probability distribution. We saw in (R2) that R_q converges to the BGS entropy in the limit $q \to 1$. $R_2 = -\sum_{i=1}^{W} p_i^2$ is called *collision entropy*. In the limit $q \to \infty$, R_q converges to the *min-entropy*

$$R_\infty(p_1, ..., p_W) = \min_{1 \le i \le W}(-\ln p_i) = -\max_{1 \le i \le W} \ln p_i = -\ln \max_{1 \le i \le W} p_i.$$

The name of R_∞ is due to property (R5).

Rényi entropy has found interesting applications in random search [53], information theory (especially in source coding [54,55]), cryptography [56], time series analysis [57], and classification [46,58], as well as in statistical signal processing and machine learning [17].

3.3. Graph Related Entropies

As part of ongoing work on graph entropy [18], the following generalized entropies are defined:

$$H_1(p_1, ..., p_W) = \sum_{i=1}^{W} \left(1 - (p_i)^{p_i}\right), \tag{18}$$

$$H_2(p_1, ..., p_W) = \prod_{i=1}^{W} \left(2 - (p_i)^{p_i}\right) = \exp\left(\sum_{i=1}^{W} \ln\left(2 - (p_i)^{p_i}\right)\right), \tag{19}$$

and

$$H_3(p_1, ..., p_W) = 1 + \ln H_2(p_1, ..., p_W) = \sum_{i=1}^{W} \left(p_i + \ln\left(2 - (p_i)^{p_i}\right)\right). \tag{20}$$

Note that $H_1(..., 0, 1, 0, etc.) = 0$, while $H_2(..., 0, 1, 0, ...) = H_3(..., 0, 1, 0, ...) = 1$. Other oddities of the above entropies include the terms $(p_i)^{p_i}$ in their definitions, as well as the presence of products instead of sums in the definition of H_2.

First, H_1 is of the type in Equation (10) with

$$g_1(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1 - x^x & \text{if } 0 < x \le 1. \end{cases} \tag{21}$$

By definition, $g(x)$ is continuous (even smooth), concave on the interval $[0,1]$, and $g_1(0) = 0$. Therefore (see Conditions (C1)–(C3) in Section 2), H1 satisfies the axioms SK1–SK3, hence it is a generalized entropy.

As for H_2, this probability functional is of the type in Equation (14) with

$$g_2(x) = \begin{cases} 0 & \text{if } x = 0, \\ \ln(2 - x^x) & \text{if } 0 < x \leq 1, \end{cases} \tag{22}$$

and $G(u) = e^u$. To prove that H_2 is a generalized entropy, note that

$$\ln H_2(p_1, ..., p_W) = \sum_{i=1}^{W} \ln(2 - (p_i)^{p_i})$$

satisfies axioms SK1–SK3 for the same reasons as H_1 does. Therefore, the same happens with H_2 on account of the exponential function being continuous (SK1), increasingly monotonic (SK2), and univalued (SK3).

Finally, H_3 is of the type in Equation (10) with

$$g_3(x) = \begin{cases} 0 & \text{if } x = 0, \\ x + g_2(x) & \text{if } 0 < x \leq 1. \end{cases} \tag{23}$$

Since $H_3 = 1 + \ln H_2$, it is a generalized entropy because, as shown above, $\ln H_2$ satisfies axioms SK1–SK3.

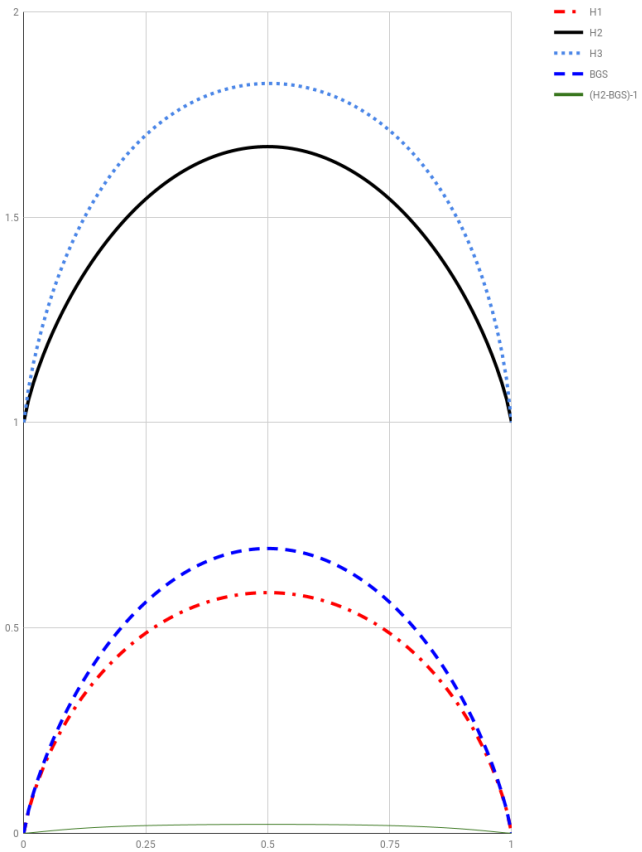

Figure 3. Entropies $H_i(p, 1-p)$, $i = 1, 2, 3$, along with $S_{BGS}(p, 1-p)$ and $H_2 - S_{BGS} - 1$ for comparison.

Figure 3 depicts $H_1(p,1-p)$, $H_2(p,1-p)$, $H_3(p,1-p)$, along with $S_{BGS}(p,1-p)$ and $H_2 - S_{BGS} - 1$ for comparison. As a curiosity, let us point out that the scaled versions

$$\tilde{H}_i(p,1-p) = \frac{H_i(p,1-p) - H_i(0,1)}{H_i(\frac{1}{2},\frac{1}{2}) - H_i(0,1)}, \tag{24}$$

($i = 1,2,3$), see Figure 4, approximate $S_{BGS}(p,1-p)$ measured in bits very well. In particular, the relative error in the approximation of $S_{BGS}(p,1-p)$ by $\tilde{H}_2(p,1-p)$ is less than 2.9×10^{-4}, so their graphs overlap when plotted.

A further description of the entropies in Equations (18)–(20) is beyond the scope of this section. Let us only mention in this regard that these entropies can be extended into the realm of acyclic directed graphs.

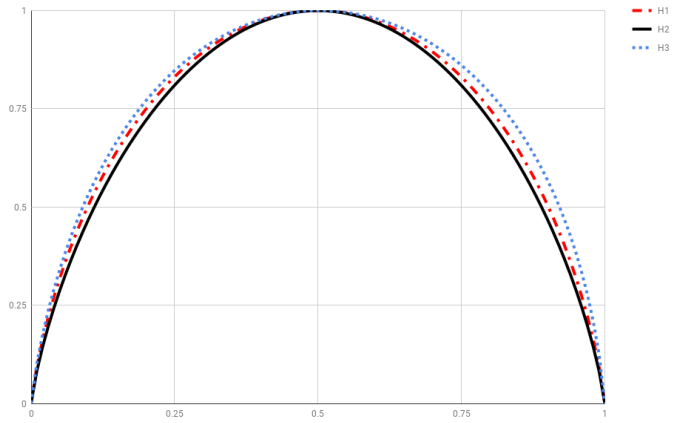

Figure 4. Scaled entropies $\tilde{H}_i(p,1-p)$, $i = 1,2,3$, see Equation (24).

4. Hanel–Thurner Exponents

All generalized entropies $F_{G,g}$ group in classes labeled by two exponents (c,d) introduced by Hanel and Thurner [16], which are determined by the limits

$$\lim_{W\to\infty} \frac{F_{G,g}(p_1,...,p_{\lambda W})}{F_{G,g}(p_1,...,p_W)} = \lambda^{1-c} \tag{25}$$

(W being as before the cardinality of the probability distribution or the total number of microstates in the system, $\lambda > 1$) and

$$\lim_{W\to\infty} \frac{F_{G,g}(p_1,...,p_{W^{1+a}})}{F_{G,g}(p_1,...,p_W)} W^{a(c-1)} = (1+a)^d \tag{26}$$

($a > 0$). Note that the limit in Equation (26) does not depend actually on c. The limits in Equations (25) and (26) can be computed via the asymptotic equipartition property [26]. Thus,

$$F_{G,g}(p_1,...,p_{\lambda W}) \approx G\left(\lambda W g\left(\frac{1}{\lambda W}\right)\right)$$

and

$$F_{G,g}(p_1,...,p_{W^{1+a}}) \approx G\left(W^{1+a} g\left(\frac{1}{W^{1+a}}\right)\right)$$

asymptotically with ever larger W (thermodynamic limit). Set now $x = 1/W$ to derive

$$\lim_{x\to 0+} \frac{G\left(\frac{\lambda}{x} g\left(\frac{x}{\lambda}\right)\right)}{G\left(\frac{1}{x} g\left(x\right)\right)} = \lambda^{1-c} \tag{27}$$

and

$$\lim_{x\to 0+} \frac{G\left(\frac{1}{x^{1+a}} g\left(x^{1+a}\right)\right)}{x^{a(c-1)} G\left(\frac{1}{x} g\left(x\right)\right)} = (1+a)^d. \tag{28}$$

Clearly, the scaling exponents c, d of a generalized entropy $F_{G,g}$ depend on the behavior of g in an infinitesimal neighborhood $(0, \varepsilon]$ of 0 (i.e., $g(\varepsilon)$ with $0 < \varepsilon \ll 1$), as well as on the properties of G if $G \neq id$. We call (c, d) the *Hanel–Thurner* (HT) *exponents* of the generalized entropy $F_{G,g}$.

When $G = id$, Equations (27) and (28) abridge to

$$\lim_{x\to 0+} \frac{g(zx)}{g(x)} = z^c \tag{29}$$

(after replacing λ^{-1} by z), and

$$\lim_{x\to 0+} \frac{g(x^{1+a})}{x^{ac} g(x)} = (1+a)^d, \tag{30}$$

respectively. In this case, $0 < c \leq 1$, while d can be any real number. If $c = 1$, the concavity of g implies $d \geq 0$ [16]. The physical properties of admissible systems are uniquely characterized by their HT exponents, i.e., by their asymptotic properties in the limit $W \to \infty$ [16]. In this sense, we can also speak of the universality class (c, d).

As way of illustration, we are going to derive the HT exponents of S_{BGS}, T_q and R_q.

(E1) For the BGS entropy, $g(x) = -x \ln x$ (see Equation (11)), so

$$\frac{g\left(zx\right)}{g\left(x\right)} = \frac{zx \ln(zx)}{x \ln x} = \frac{z \ln z + z \ln x}{\ln x} \to z$$

as $x \to 0+$. Therefore, $c = 1$. Furthermore,

$$\frac{g\left(x^{1+a}\right)}{x^{ac} g\left(x\right)} = \frac{x^{1+a} \ln x^{1+a}}{x^{a+1} \ln x} = \frac{(1+a) \ln x}{\ln x} = 1 + a$$

for all $x > 0$, so $d = 1$.

(E2) For the Tsallis entropy, see Equation (12),

$$g(x) = \begin{cases} \frac{1}{1-q} x^q + \mathcal{O}(x) & \text{if } 0 < q < 1, \\ -\frac{1}{1-q} x + \mathcal{O}(x) & \text{if } q > 1. \end{cases}$$

It follows readily that $(c, d) = (q, 0)$ if $0 < q < 1$, and $(c, d) = (1, 0)$ if $q > 1$. Hence, although $\lim_{q\to 1} T_q = S_{BGS}$, there is no parallel convergence concerning the HT exponents.

(E3) For the Rényi entropy, $g(x) = x^q$ and $G(u) = \frac{1}{1-q} e^u$ (see Equation (15)), so

$$\frac{G(\frac{\lambda}{x} g\left(\frac{x}{\lambda}\right))}{G(\frac{1}{x} g\left(x\right))} = \frac{\ln\left(\frac{\lambda}{x}(\frac{x}{\lambda})^q\right)}{\ln\left(\frac{1}{x} x^q\right)} = \frac{\ln x^{q-1} - \ln \lambda^{q-1}}{\ln x^{q-1}} \to 1$$

as $x \to 0+$ (both for $0 \leq q \leq 1$ and $q \geq 1$). Therefore, $c = 1$. Furthermore,

$$\frac{G\left(\frac{1}{x^{1+a}} g\left(x^{1+a}\right)\right)}{G\left(\frac{1}{x} g\left(x\right)\right)} = \frac{\ln\left(\frac{1}{x^{1+a}} x^{q(1+a)}\right)}{\ln\left(\frac{1}{x} x^{q}\right)} = \frac{\ln x^{(q-1)(1+a)}}{\ln x^{q-1}} = 1 + a$$

for all $x > 0$, so that $d = 1$. In sum, $(c, d) = (1, 1)$ for all q.

As for the generalized entropies H_1, H_2, and H_3 considered in Section 3.3, we show in Appendix B that their HT exponents are $(1, 1)$, $(0, 0)$, and $(1, 1)$, respectively. Thus, H_1 and H_3 belong to the same universality class as S_{BGS}, while the HT exponents of H_2 and R_q (both of the same the type in Equation (14)) are different. Moreover, the interested reader will find in Table 1 of [16] the HT exponents of the generalized entropies listed in Appendix A.

An interesting issue that arises at this point is the inverse question: Given $c \in (0, 1]$ and $d \in \mathbb{R}$, is there an admissible system such that its HT exponents are precisely (c, d)? The answer is yes, at least under some restrictions on the values of c and d. Following [16], we show in Appendix C that, if

$$\begin{array}{ll} d > -1 & \text{for } 0 < c \leq \frac{1}{2}, \\ d \geq 1 - \frac{1}{c} & \text{for } \frac{1}{2} < c \leq 1, \end{array} \tag{31}$$

then the "generalized (c, d)-entropy"

$$S_{c,d}(p_1, ..., p_W) = eA \sum_{i=1}^{W} \Gamma(d + 1, 1 - c \ln p_i), \tag{32}$$

has HT exponents (c, d). Here, $A > 0$ and Γ is the incomplete Gamma function (Section 6.5 of [59]), that is,

$$\Gamma(r, s) = \int_s^\infty t^{r-1} e^{-t} dt \quad (r > 0). \tag{33}$$

Several application cases where generalized (c, d)-entropies are relevant have been discussed by Hanel and Thurner in [40] (super-diffusion, spin systems, binary processes, and self-organized critical systems) and [60] (aging random walks, i.e., random walks whose transition rates between states are path- and time-dependent).

5. Asymptotic Relation between the HT Exponent c and the Diffusion Scaling Exponent

In contrast to "non-interacting" systems, where both the additivity and extensivity of the BGS entropy S_{BGS} hold, in the case of general interacting statistical systems these properties can no longer be simultaneously satisfied, requiring a more general concept of entropy [16,40]. Following [16] (Section 4), a possible generalization of S_{BGS} for admissible systems is defined via the two asymptotic scaling relations in Equations (29) and (30), i.e., the HT exponents c and d, respectively. These asymptotic exponents can be interpreted as a measure of deviation from the "non-interacting" case regarding the stationary behavior.

5.1. The Non-Stationary Regime

In this section, we describe a relation between the exponent c and a similar macroscopic measure that characterizes the system in the non-stationary regime, thus providing a meaningful interpretation of the exponent. The non-stationary behavior of a system can possibly be described by the Fokker–Planck (FP) equation governing the time evolution of a probability density function $p = p(x, t)$. In this continuous limit, the generalized entropy F_g is assumed to be written as $F_g[p(s)] = \int g(p(s))\mathrm{d}s$, where g is asymptotically characterized by Equation (29) and $s = s(x)$ is a time-independent scalar function of the space coordinate x (for example, a potential) [61,62].

Going beyond the scope of the simplest FP equation, we consider systems for which the correlation among their (sub-)units can be taken into account by replacing the diffusive term $\partial_x^2 p$ with an effective term $\partial_x^2 \Phi[p]$, where $\Phi[p]$ is a pre-defined functional of the probability density. $\Phi[p]$ can be either derived directly from the microscopical transition rules or it may be defined based on macroscopic assumptions. The resulting FP equation can be written as

$$\partial_t p(x,t) = D\beta\partial_x \left(p(x,t)\partial_x u(x)\right) + D\partial_x^2 \Phi[p(x,t)], \tag{34}$$

where D, β are constants and $u(x)$ is a time-independent external potential.

For simplicity, hereafter we exclusively focus on one dimensional FP equations. In the special case of $\Phi[p] = p$ and no external forces, Equation (34) reduces to the well-known linear diffusion equation

$$\partial_t p(x,t) = D\partial_x^2 p(x,t). \tag{35}$$

The above equation is invariant under the space-time scaling transformation

$$p(x,t) = \tau^{-\gamma} p\left(\frac{x}{\tau^\gamma}, \frac{t}{\tau}\right) \tag{36}$$

with $\gamma = \frac{1}{2}$ [63,64]. This scaling property opens up the possibility of a phenomenological and macroscopic characterization of anomalous diffusion processes [15,44] as well, which correspond to more complicated non-stationary processes described by FP equations in the form of Equation (34) with a non-trivial value of γ. With the help of the transformation in Equation (36), we can also classify correlated statistical systems according to the rate of the spread of their probability density functions over time in the asymptotic limit and, thus, quantitatively describe their behavior in the non-stationary regime.

5.2. Relation between the Stationary and Non-Stationary Regime

To reasonably and consistently relate the generalized entropies to the formalism of FP equations—corresponding to the stationary and non-stationary regime, respectively—the functional $\Phi[p]$ has to be chosen such that the stationary solution of the general FP equation becomes equivalent to the Maximum Entropy (MaxEnt) probability distribution calculated with the generalized entropies. These MaxEnt distributions can be obtained analogously to the results by Hanel and Thurner in [16,40], where they used standard constrained optimization to find the most general form of MaxEnt distributions, which turned out to be $p(\epsilon) = \mathcal{E}_{c,d,r}(-\epsilon)$ with

$$\mathcal{E}_{c,d,r}(x) \propto \exp\left[-\frac{d}{1-c} W_k\left(B(1-\frac{x}{r})^{1/d}\right)\right]. \tag{37}$$

Here, B, r are constants depending only on the c, d parameters and W_k is the kth branch of the Lambert-W function (specifically, branch $k = 0$ for $d \geq 0$ and branch $k = 1$ for $d < 0$). The consistency criterion imposed above accords with the fact that many physical systems tend to converge towards maximum entropy configuration over time, however, it specifies the limits of our assumptions.

Consider systems described by Equation (34) in the absence of external force, i.e.,

$$\partial_t p(x,t) = D\partial_x^2 \Phi\left[p(x,t)\right]. \tag{38}$$

By assuming that the corresponding stationary solutions can be identified with the MaxEnt distributions in Equation (37), it can be shown that the functional form of the effective density $\Phi[p]$ must be expressed as

$$\Phi[p] \propto \int_0^p q\, \partial_q^2 g\left(q\right) \mathrm{d}q, \tag{39}$$

where we neglected additive and multiplicative constant factors for the sake of simplicity. Similar implicit equations have already been investigated in [61,62,65]. Once the asymptotic phase space volume scaling relation in Equation (29) holds, it can also be shown that the generalized FP in Equation (38) (with Φ as in Equation (39)) obeys the diffusion scaling property in Equation (36) with a non-trivial value of γ in the $p \to 0$ asymptotic limit [66] (assuming additionally the existence of the solution of Equation (38), at least from an appropriate initial condition). A simple algebraic relation between the diffusion scaling exponent γ and the phase space volume scaling exponent c can be established [66], which can be written as

$$\gamma = \frac{1}{1+c}. \tag{40}$$

Therefore, this relation between c and γ defines families of FP equations which show asymptotic invariance under the scaling relation in Equation (36).

6. Conclusions

This review concentrates on the concept of generalized entropy (Section 2), which is relevant in the study of real thermodynamical systems and, more generally, in the theory of complex systems. Possibly the first example of a generalized entropy was introduced by Rényi (Section 3.2), who was interested in the most general information measure which is additive in the sense of Equation (5), with the random variables X and Y being independent. Another very popular generalized entropy was introduced by Tsallis as a generalization of the Boltzmann–Gibbs entropy (Section 3.1) to describe the properties of physical systems with long range forces and complex dynamics in equilibrium. Some more exotic generalized entropies are considered in Section 3.3, while other examples that have been published in the last two decades are gathered in Appendix A. Our approach was to a great extent formal, with special emphasis in Sections 2 and 3 on axiomatic formulations and mathematical properties. For expository reasons, applications are mentioned and the original references given as our description of the main generalized entropies progressed, rather than addressing them jointly in a separate section.

An alternative approach to generalized entropies other than the axiomatic one (Section 2) consists in characterizing their asymptotic behavior in the thermodynamic limit $W \to \infty$. Hanel and Thurner showed that two scaling exponents (c, d) suffice for admissible generalized entropies, i.e., those entropies of the form in Equation (10) with g continuous, concave and $g(0) = 0$ (Section 4); it holds $c \in (0, 1]$ and $d \in \mathbb{R}$. As a result, the admissible systems fall in equivalence classes labeled by the exponents (c, d) of the corresponding entropies. Conversely, to each (c, d), there is a generalized entropy with those Hanel–Thurner exponents (see Equation (32)), at least for the most interesting value ranges.

It is also remarkable that, at asymptotically large times and volumes, there is a 1-to-1 relation between the equivalence class of generalized entropies with a given $c \in (0, 1]$ and the equivalence class of Fokker–Planck equations in which the invariance in Equation (36) holds with $\gamma = \frac{1}{1+c} \in \left[\frac{1}{2}, 1\right)$ (Section 5). This means that the equivalence classes of admissible systems can generally be mapped into anomalous diffusion processes and vice versa, thus conveying the same information about the system in the asymptotic limit (i.e., when $p(x, t) \to 0$) [66]. A schematic visualization of this relation is provided in Figure 5. Moreover, the above result can actually be understood as a possible generalization of the Tsallis–Bukman relation [44].

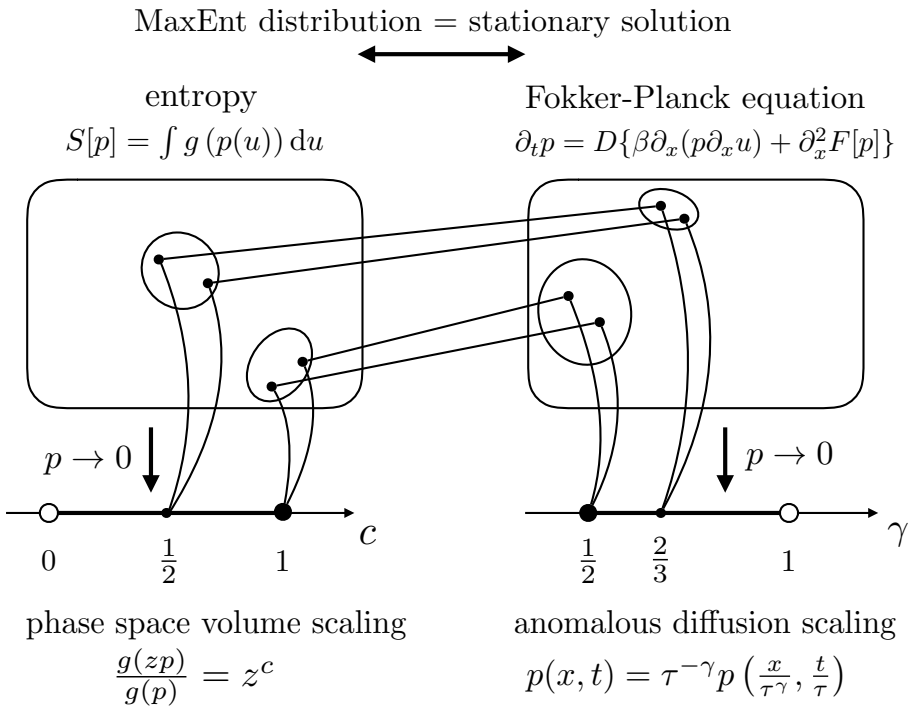

Figure 5. Visual summary of the main result presented in Section 5 schematically depicting the relation between the exponents γ and c. Source: [66].

Author Contributions: All the authors have contributed to conceptualization, methodology, validation, formal analysis, investigation, writing, review and editing, both of the initial draft and the final version.

Funding: J.M.A. was supported by the Spanish Ministry of Economy, Industry and Competitiveness, grant MTM2016-74921-P (AEI/FEDER, EU). S.G.B. was partially supported by the Hungarian National Research, Development and Innovation Office (grant no. K 128780) and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 740688.

Acknowledgments: We thank our referees for their helpful and constructive criticism. J.M.A. was supported by the Spanish Ministry of Economy, Industry and Competitiveness, grant MTM2016-74921-P (AEI/FEDER, EU). This research was also partially supported by the Hungarian National Research, Development and Innovation Office (grant no. K 128780) and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 740688.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

We list in this appendix further generalized entropies of the form in Equation (10) and the original references (notation as in Table 1 of [16]). $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function, in Equation (33).

1. $S_\eta(\{p_i\}) = \sum_i \Gamma(\frac{\eta+1}{\eta}, \ln p_i) - p_i\Gamma(\frac{\eta+1}{\eta})$ $(\eta > 0)$ [67].
2. $S_\kappa(\{p_i\}) = \sum_i \frac{p_i^{1-\kappa} - p_i^{1+\kappa}}{2\kappa}$ $(0 < \kappa < 1)$ [68].
3. $S_b(\{p_i\}) = \sum_i (1 - e^{bp_i}) + e^{-b} - 1$ $(b > 0)$ [69].
4. $S_E(\{p_i\}) = \sum_i p_i(1 - e^{(p_i-1)/p_i})$ [70].
5. $S_\beta(\{p_i\}) = \sum_i p_i^\beta \ln(1/p_i)$ $(0 < \beta \leq 1)$ [71].
6. $S_\gamma(\{p_i\}) = \sum_i p_i \ln^{1/\gamma}(1/p_i)$ ([15], page 60).

Appendix B

From Equations (18)–(20), it follows that the functions g of H_1, H_2, and H_3 are the following:

$$\begin{aligned} g_1(\varepsilon) &\equiv 1-\varepsilon^{\varepsilon} \approx -\varepsilon\ln\varepsilon, \\ g_2(\varepsilon) &\equiv \ln(2-\varepsilon^{\varepsilon}) \approx 1-\varepsilon^{\varepsilon} \approx -\varepsilon\ln\varepsilon, \end{aligned}$$

$$g_3(\varepsilon) \equiv \varepsilon + \ln\left(2-\varepsilon^{\varepsilon}\right) \approx \varepsilon - \varepsilon\ln\varepsilon \approx -\varepsilon\ln\varepsilon.$$

Since H_1 and H_3 are generalized entropies of the type in Equation (10), we conclude that both belong to the same class as S_{BGS} (see Equation (11)), hence $(c,d)=(1,1)$.

H_2 is a generalized entropy of the type in Equation (14) with $G(u)=e^u$. Therefore,

$$\frac{G(\frac{\lambda}{\varepsilon}g_2\left(\frac{\varepsilon}{\lambda}\right))}{G(\frac{1}{\varepsilon}g_2\left(\varepsilon\right))} \approx \frac{\exp\left(-\frac{\lambda}{\varepsilon}\frac{\varepsilon}{\lambda}\ln\frac{\varepsilon}{\lambda}\right)}{\exp\left(-\ln\varepsilon\right)} = \frac{\lambda/\varepsilon}{1/\varepsilon} = \lambda.$$

Comparison with Equation (27) shows that $c=0$.

Moreover,

$$\frac{G\left(\frac{1}{\varepsilon^{1+a}}g_2\left(\varepsilon^{1+a}\right)\right)}{\varepsilon^{a(c-1)}G\left(\frac{1}{\varepsilon}g_2\left(\varepsilon\right)\right)} \approx \frac{\exp\left(-(1+a)\ln\varepsilon\right)}{\varepsilon^{-a}\exp\left(-\ln\varepsilon\right)} = \frac{\varepsilon^{-(1+a)}}{\varepsilon^{-(1+a)}} = 1.$$

Comparison with Equation (28) shows that $d=0$.

Appendix C

1. First, note from Equation (32) that

$$g_{c,d}(x) = eA\Gamma(d+1, 1-c\ln x), \tag{A1}$$

where the incomplete Gamma function $\Gamma(d+1,1-c\ln x)$ exists for $d>-1$ and all $x\in(0,1]$ (see Equation (33)), with $g_{c,d}(0)=\lim_{x\to 0+} eA\Gamma(d+1,1-c\ln x)=0$.

Among Conditions (C1)–(C3) on $g_{c,d}$ (Section 2), for the entropy $S_{c,d}$ in Equation (32) to be admissible, only concavity (Condition (C2)) needs to be checked. Since

$$\begin{aligned} \frac{d^2}{dx^2}g_{c,d}(x) &= eA\frac{d^2}{dx^2}\Gamma(d+1,1-c\ln x) \\ &= eA\left(\frac{c}{x}\right)^2 e^{-1+c\ln x}(1-c\ln x)^{d-1}\times \\ &\quad \times\left(1-\tfrac{1}{c}+(1-c)\ln x-d\right), \end{aligned}$$

it holds $g''_{c,d}(x)\le 0$ if and only if $d\ge 1-\frac{1}{c}+(1-c)\ln x$, where $-\infty<(1-c)\ln x\le 0$ for each $c\in(0,1]$ and $x\in[0,1]$. Therefore, $g''_{c,d}(x)\le 0$ for all $x\in[0,1]$ if and only if $d\ge 1-\frac{1}{c}$, where $-\infty<1-\frac{1}{c}\le 0$. On the other hand, $d>-1$ for the integral $\Gamma(d+1,1-c\ln x)$ to exist. Both restrictions together lead then to the condition in Equation (31) on d for $S_{c,d}$ to be a generalized entropy.

2. Use the asymptotic approximation 6.5.32 of [59]

$$\begin{aligned} \Gamma(d+1,1-c\ln\varepsilon) &= (1-c\ln\varepsilon)^d e^{c\ln\varepsilon-1}+\mathcal{O}(\tfrac{1}{\ln\varepsilon}) \\ &\approx e^{-1}\varepsilon^c(1-c\ln\varepsilon)^d \end{aligned} \tag{A2}$$

($d>-1$, $0<c\le 1$) to obtain the leading approximation of $g_{c,d}(x)$ in an infinitesimal neighborhood of 0:

$$g_{c,d}(\varepsilon) \approx A\varepsilon^c(1-c\ln\varepsilon)^d \approx Ac^d\varepsilon^c(\ln\tfrac{1}{\varepsilon})^d. \tag{A3}$$

Using Equation (A3), the following can be derived:

$$\frac{g_{c,d}(z\varepsilon)}{g_{c,d}(\varepsilon)} \approx \frac{z^c\varepsilon^c(\ln\frac{1}{z\varepsilon})^d}{\varepsilon^c(\ln\frac{1}{\varepsilon})^d} = z^c\frac{(\ln\frac{1}{z}+\ln\frac{1}{\varepsilon})^d}{(\ln\frac{1}{\varepsilon})^d} \approx z^c \tag{A4}$$

(see Equation (29)) and

$$\frac{g_{c,d}(\varepsilon^{1+a})}{\varepsilon^{ac}g_{c,d}(\varepsilon)} = \frac{\varepsilon^{c(1+a)}(1+a)^d(\ln\frac{1}{\varepsilon})^d}{\varepsilon^{ac}\varepsilon^c(\ln\frac{1}{\varepsilon})^d} = (1+a)^d \tag{A5}$$

(see Equation (30)).

3. From Equation (A3), we obtain

$$g_{1,1}(\varepsilon) \approx -A\varepsilon\ln\varepsilon \tag{A6}$$

(see Example (E1)) and,

$$g_{c,0}(\varepsilon) \approx A\varepsilon^c \tag{A7}$$

(see Example (E2)). Set $A = (1-c+cd)^{-1}$ [16] in Equations (A6) and (A7) to reproduce the g functions of S_{BGS} ($A = 1$) and T_c, $0 < c < 1$ ($A = \frac{1}{1-c}$), respectively.

References

1. Clausius, R. *The Mechanical Theory of Heat*; McMillan and Co.: London, UK, 1865.
2. Boltzmann, L. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitz. Ber. Akad. Wiss. Wien (II)* **1872**, *66*, 275–370.
3. Boltzmann, L. Über die Beziehung eines allgemeinen mechanischen Satzes zum zweiten Hauptsatz der Wärmetheorie. *Sitz. Ber. Akad. Wiss. Wien (II)* **1877**, *75*, 67–73.
4. Gibbs, J.W. *Elementary Principles in Statistical Mechanics—Developed with Especial References to the Rational Foundation of Thermodynamics*; C. Scribner's Sons: New York, NY, USA, 1902.
5. Dewar, R. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in nonequilibrium stationary state. *J. Phys. A Math. Gen.* **2003**, *36*, 631–641. [CrossRef]
6. Martyushev, L.M. Entropy and entropy production: old misconceptions and new breakthroughs. *Entropy* **2013**, *15*, 1152–1170. [CrossRef]
7. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
8. Wissner-Gross, A.D.; Freer, C.E. Causal entropic forces. *Phys. Rev. Lett.* **2013**, *110*, 168702. [CrossRef] [PubMed]
9. Mann, R.P.; Garnett, R. The entropic basis of collective behaviour. *J. R. Soc. Interface* **2015**, *12*, 20150037. [CrossRef] [PubMed]
10. Kolmogorov, A.N. A new metric invariant of transitive dynamical systems and Lebesgue space endomorphisms. *Dokl. Acad. Sci. USSR* **1958**, *119*, 861–864.
11. Rényi, A. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*; Neyman, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
12. Tsallis, C. Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
13. Amigó, J.M.; Keller, K.; Unakafova, V. On entropy, entropy-like quantities, and applications. *Disc. Cont. Dyn. Syst. B* **2015**, *20*, 3301–3343. [CrossRef]
14. Csiszár, I. Axiomatic characterization of information measures. *Entropy* **2008**, *10*, 261–273. [CrossRef]
15. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 2009.
16. Hanel, R.; Thurner, S. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *EPL* **2011**, *93*, 20006. [CrossRef]
17. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer: New York, NY, USA, 2010.

18. Hernández, S. Introducing Graph Entropy. Available online: http://entropicai.blogspot.com/search/label/Graph%20entropy (accessed on 22 October 2018).
19. Salicrú, M.; Menéndez, M.L.; Morales, D.; Pardo, L. Asymptotic distribution of (h, ϕ)-entropies. *Commun. Stat. Theory Meth.* **1993**, *22*, 2015–2031. [CrossRef]
20. Bosyk, G.M.; Zozor, S.; Holik, F.; Portesi, M.; Lamberti, P.W. A family of generalized quantum entropies: Definition and properties. *Quantum Inf. Process.* **2016**, *15*, 3393–3420. [CrossRef]
21. Von Neumann, J. Thermodynamik quantenmechanischer Gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen* **1927**, *1927*, 273–291. (In German)
22. Hein, C.A. Entropy in Operational Statistics and Quantum Logic. *Found. Phys.* **1979**, *9*, 751–786. [CrossRef]
23. Short, A.J.; Wehner, S. Entropy in general physical theories. *New J. Phys.* **2010**, *12*, 033023. [CrossRef]
24. Holik, F.; Bosyk, G.M.; Bellomo, G. Quantum information as a non-Kolmogovian generalization of Shannon's theory. *Entropy* **2015**, *17*, 7349–7373. [CrossRef]
25. Portesi, M.; Holik, F.; Lamberti, P.W.; Bosyk, G.M.; Bellomo, G.; Zozor, S. Generalized entropie in quantum and classical statistical theories. *Eur. Phys. J. Spec. Top.* **2018**, *227*, 335–344. [CrossRef]
26. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley and Sons: Hoboken, NJ, USA, 2006.
27. Enciso, A.; Tempesta, P. Uniqueness and characterization theorems for generalized entropies. *J. Stat. Mech.* **2017**, 123101. [CrossRef]
28. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover Publications: New York, NY, USA, 1957.
29. Ash, R.B. *Information Theory*; Dover Publications: New York, NY, USA, 1990.
30. MacKay, D.J. *Information Theory, Inference, and Earning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
31. Bandt, C. A new kind of permutation entropy used to classify sleep stages from invisible EEG microstructure. *Entropy* **2017**, *19*, 197. [CrossRef]
32. Havrda, J.; Charvát, F. Quantification method of classification processes. Concept of structural α-entropy. *Kybernetika* **1967**, *3*, 30–35.
33. Abe, S. Stability of Tsallis entropy and instabilities of Renyi and normalized Tsallis entropies. *Phys. Rev. E* **2002**, *66*, 046134. [CrossRef] [PubMed]
34. Tsallis, C.; Brigatti, E. Nonextensive statistical mechanics: A brief introduction. *Contin. Mech. Thermodyn.* **2004**, *16*, 223–235. [CrossRef]
35. Abe, S. Tsallis entropy: How unique? *Contin. Mech. Thermodyn.* **2004**, *16*, 237–244. [CrossRef]
36. Dos Santos, R.J.V. Generalization of Shannon's theorem for Tsallis entropy. *J. Math. Phys.* **1997**, *38*, 4104–4107. [CrossRef]
37. Suyari, H. Generalization of Shannon–Khinchin axioms to nonextensive systems and the uniqueness theorem for the nonextensive entropy. *IEEE Trans. Inf. Theory* **2004**, *50*, 1783–1787. [CrossRef]
38. Furuichi, S. On uniqueness theorems for Tsallis entropy and Tsallis relative entropy. *IEEE Trans. Inf. Theory* **2005,** *51*, 3638–3645. [CrossRef]
39. Jäckle, S.; Keller, K. Tsallis entropy and generalized Shannon additivity. *Axioms* **2016**, *6*, 14. [CrossRef]
40. Hanel, R.; Thurner, S. When do generalized entropies apply? How phase space volume determines entropy. *Europhys. Lett.* **2011**, *96*, 50003. [CrossRef]
41. Plastino, A.R.; Plastino, A. Stellar polytropes and Tsallis' entropy. *Phys. Lett. A* **1993**, *174*, 384–286. [CrossRef]
42. Alemany, P.A.; Zanette, D.H. Fractal random walks from a variational formalism for Tsallis entropies. *Phys. Rev. E* **1994**, *49*, R956–R958. [CrossRef]
43. Plastino, A.R.; Plastino, A. Non-extensive statistical mechanics and generalized Fokker–Planck equation. *Physica A* **1995**, *222*, 347–354. [CrossRef]
44. Tsallis, C.; Bukman, D.J. Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis. *Phys. Rev. E* **1996**, *54*, R2197. [CrossRef]
45. Capurro, A.; Diambra, L.; Lorenzo, D.; Macadar, O.; Martin, M.T.; Mostaccio, C.; Plastino, A.; Rofman, E.; Torres, M.E.; Velluti, J. Tsallis entropy and cortical dynamics: The analysis of EEG signals. *Physica A* **1998**, *257*, 149–155. [CrossRef]
46. Maszczyk, T.; Duch, W. Comparison of Shannon, Renyi and Tsallis entropy used in decision trees. In Proceedings of the International Conference on Artificial Intelligence and Soft Computing, Zakopane, Poland, 22–26 June 2008; Springer: Berlin, Germany, 2008; pp. 643–651.

47. Gajowniczek, K.; Karpio, K.; Łukasiewicz, P.; Orłowski, A.; Zabkowski, T. Q-Entropy approach to selecting high income households. *Acta Phys. Pol. A* **2015**, *127*, 38–44. [CrossRef]
48. Gajowniczek, K.; Orłowski, A.; Zabkowski, T. Simulation study on the application of the generalized entropy concept in artificial neural networks. *Entropy* **2018**, *20*, 249. [CrossRef]
49. Lesche, B. Instabilities of Renyi entropies. *J. Stat. Phys.* **1982**, *27*, 419–422. [CrossRef]
50. Mariz, A.M. On the irreversible nature of the Tsallis and Renyi entropies. *Phys. Lett. A* **1992**, *165*, 409–411. [CrossRef]
51. Aczél, J.; Daróczy, Z. Charakterisierung der Entropien positiver Ordnung und der Shannonschen Entropie. *Acta Math. Acad. Sci. Hung.* **1963**, *14*, 95–121. (In German) [CrossRef]
52. Jizba, P.; Arimitsu, T. The world according to Rényi: Thermodynamics of multifractal systems. *Ann. Phys.* **2004**, *312*, 17–59. [CrossRef]
53. Rényi, A. On the foundations of information theory. *Rev. Inst. Int. Stat.* **1965**, *33*, 1–4. [CrossRef]
54. Campbell, L.L. A coding theorem and Rényi's entropy. *Inf. Control* **1965**, *8*, 423–429. [CrossRef]
55. Csiszár, I. Generalized cutoff rates and Rényi information measures. *IEEE Trans. Inf. Theory* **1995**, *41*, 26–34. [CrossRef]
56. Bennett, C.; Brassard, G.; Crépeau, C.; Maurer, U. Generalized privacy amplification. *IEEE Trans. Inf. Theory* **1995**, *41*, 1915–1923. [CrossRef]
57. Kannathal, N.; Choo, M.L.; Acharya, U.R.; Sadasivan, P.K. Entropies for detection of epilepsy in EEG. *Comput. Meth. Prog. Biomed.* **2005**, *80*, 187–194. [CrossRef] [PubMed]
58. Contreras-Reyes, J.E.; Cortés, D.D. Bounds on Rényi and Shannon Entropies for Finite Mixtures of Multivariate Skew-Normal Distributions: Application to Swordfish (*Xiphias gladius Linnaeus*). *Entropy* **2016**, *11*, 382. [CrossRef]
59. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Tables*; Dover Publications: New York, NY, USA, 1972.
60. Hanel, R.; Thurner, S. Generalized (c,d)-entropy and aging random walks. *Entropy* **2013**, *15*, 5324–5337. [CrossRef]
61. Chavanis, P.H. Nonlinear mean field Fokker–Planck equations. Application to the chemotaxis of biological populations. *Eur. Phys. J. B* **2008**, *62*, 179–208. [CrossRef]
62. Martinez, S.; Plastino, A.R.; Plastino, A. Nonlinear Fokker–Planck equations and generalized entropies. *Physica A* **1998**, *259*, 183–192. [CrossRef]
63. Bouchaud, J.P.; Georges, A. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* **1990**, *195*, 127–293. [CrossRef]
64. Dubkov, A.A.; Spagnolo, B.; Uchaikin, V.V. Lévy flight superdiffusion: An introduction. *Int. J. Bifurcat. Chaos* **2008**, *18*, 2649–2672. [CrossRef]
65. Schwämmle, V.; Curado, E.M.F.; Nobre, F.D. A general nonlinear Fokker–Planck equation and its associated entropy. *EPJ B* **2007**, *58*, 159–165. [CrossRef]
66. Czégel, D.; Balogh, S.G.; Pollner, P.; Palla, G. Phase space volume scaling of generalized entropies and anomalous diffusion scaling governed by corresponding nonlinear Fokker–Planck equations. *Sci. Rep.* **2018**, *8*, 1883. [CrossRef] [PubMed]
67. Anteneodo, C.; Plastino, A.R. Maximum entropy approach to stretched exponential probability distributions. *J. Phys. A Math. Gen.* **1999**, *32*, 1089–1098. [CrossRef]
68. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **2002**, *66*, 056125. [CrossRef] [PubMed]
69. Curado, E.M.; Nobre, F.D. On the stability of analytic entropic forms. *Physica A* **2004**, *335*, 94–106. [CrossRef]
70. Tsekouras, G.A.; Tsallis, C. Generalized entropy arising from a distribution of q indices. *Phys. Rev. E* **2005**, *71*, 046144. [CrossRef] [PubMed]
71. Shafee, F. Lambert function and a new non-extensive form of entropy. *IMA J. Appl. Math.* **2007**, *72*, 785–800. [CrossRef]