



Article

# Measuring the Complexity of Continuous Distributions

Guillermo Santamaría-Bonfil <sup>1,2,\*</sup>, Nelson Fernández <sup>3,4,\*</sup> and Carlos Gershenson <sup>1,2,5,6,7,\*</sup>

<sup>1</sup> Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico

<sup>2</sup> Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico

<sup>3</sup> Laboratorio de Hidroinformática, Universidad de Pamplona, 543050 Pamplona, Colombia

<sup>4</sup> Grupo de Investigación en Ecología y Biogeografía, Universidad de Pamplona, 543050 Pamplona, Colombia

<sup>5</sup> SENSEable City Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>6</sup> MoBS Lab, Northeastern University, Boston, MA 02115, USA

<sup>7</sup> ITMO University, 199034 St. Petersburg, Russia

\* Correspondence: guillermo.santamaria@iimas.unam.mx (G.S.-B.); nfernandez@unipamplona.edu.co (N.F.); cgg@unam.mx (C.G.); Tel.: +52-55-56-22-38-99 (ext. 44027) (C.G.)

Academic Editors: Hermann Haken and Juval Portugali

Received: 30 October 2015; Accepted: 16 February 2016; Published: 26 February 2016

**Abstract:** We extend previously proposed measures of complexity, emergence, and self-organization to continuous distributions using differential entropy. Given that the measures were based on Shannon's information, the novel continuous complexity measures describe how a system's predictability changes in terms of the probability distribution parameters. This allows us to calculate the complexity of phenomena for which distributions are known. We find that a broad range of common parameters found in Gaussian and scale-free distributions present high complexity values. We also explore the relationship between our measure of complexity and information adaptation.

**Keywords:** complexity; emergence; self-organization; information; differential entropy; probability distributions

## 1. Introduction

Complexity is pervasive. However, there is no agreed definition of complexity. Perhaps complexity is so general that it resists definition [1]. Still, it is useful to have formal measures of complexity to study and compare different phenomena [2]. We have proposed measures of emergence, self-organization, and complexity [3,4] based on information theory [5]. Shannon information can be seen as a measure of novelty, so we use it as a measure of emergence, which is correlated with chaotic dynamics. Self-organization can be seen as a measure of order [6], which can be estimated with the inverse of Shannon's information and is correlated with regularity. Complexity can be seen as a balance between order and chaos [7,8], between emergence and self-organization [4,9].

We have studied the complexity of different phenomena for different purposes [10–14]. Instead of searching for more data and measuring its complexity, we decided to explore different distributions with our measures. This would allow us to study broad classes of dynamical systems in a general way, obtaining a deeper understanding of the nature of complexity, emergence, and self-organization. Nevertheless, our previously proposed measures use discrete Shannon information. The later statement has two implications: on the one hand, in the continuous domain, entropy is a proxy of

the average uncertainty for a probability distribution with a given parameter set, rather than a proxy of the system's average uncertainty; on the other, even when any distribution can be discretized, this always comes with caveats [15]. For these reasons, we base ourselves on differential entropy [15,16] to propose measures for continuous distributions.

The next section provides background concepts related to information and entropies. Next, discrete measures of emergence, self-organization, and complexity are reviewed [4]. Section 4 presents continuous versions of these measures, based on differential entropy. The probability density functions used in the experiments are described in Section 5. Section 6 presents results, which are discussed and related to information adaptation [17] in Section 7.

## 2. Information Theory

Let us have a set of possible events whose probabilities of occurrence are  $p_1, p_2, \dots, p_n \in P(X)$ . Can we measure the uncertainty described by the probability distribution  $P(X)$ ? To solve this endeavor in the context of telecommunications, Shannon proposed a measure of entropy [5], which corresponds to Boltzmann–Gibbs entropy in thermodynamics. This measure, as originally proposed by Shannon, possesses a dual meaning of both uncertainty and information, even when the latter term was later discouraged by Shannon himself [18]. Moreover, we encourage the concept of entropy as the average uncertainty given the property of asymptotic equipartition (described later in this section). From an information-theoretic perspective, entropy measures the average number of binary questions required to determine the value of  $p_i$ . In cybernetics, it is related to variety [19], a measure of the number of distinct states a system can be in.

In general, entropy is discussed regarding a discrete probability distribution. Shannon extended this concept to the continuous domain with differential entropy. However, some of the properties of its discrete counterpart are not maintained. This has relevant implications for extending to the continuous domain the measures proposed in [3,4]. Before delving into these differences, first we introduce the discrete entropy, the asymptotic equipartition property (AEP), and the properties of discrete entropy. Next, differential entropy is described, along with its relation to discrete entropy.

### 2.1. Discrete Entropy

Let  $X$  be a discrete random variable, with a probability mass function  $p(x) = \Pr\{X = x\}, x \in X$ . The entropy  $H(X)$  of a discrete random variable  $X$  is then defined by

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x). \quad (1)$$

The logarithm base provides the entropy's unit. For instance, base two measures entropy as bits, base ten as nats. If the base of the logarithm is  $\beta$ , we denote the entropy as  $H_\beta(X)$ . Unless otherwise stated, we will consider all logarithms to be of base two. Note that entropy does not depend on the value of  $X$ , but on the probabilities of the possible values  $X$  can take. Furthermore, Equation (1) can be understood as the expected value of the information of the distribution.

### 2.2. Asymptotic Equipartition Property for Discrete Random Variables

In probability, the large numbers law states that, for a sequence of  $n$  independent and identically distributed (i.i.d.) elements of a sample  $X$ , the average value of the sample  $\frac{1}{n} \sum_{i=1}^n X_i$  approximates the expected value  $\mathbb{E}(X)$ . In this sense, the Asymptotic Equipartition Property (AEP) establishes that  $H(X)$  can be approximated by

$$H(X) = \frac{1}{n} \log_2 \frac{1}{p(X_1, \dots, X_n)},$$

under the conditions that  $n \rightarrow \infty$ , and  $x_i \in X$  is i.i.d.

Therefore, discrete entropy can be written also as

$$H(X) = \mathbb{E} \left[ \log \frac{1}{p(X)} \right], \quad (2)$$

where  $\mathbb{E}$  is the expected value of  $P(X)$ . Consequently, Equation (2) describes the expected or average uncertainty of probability distribution  $P(X)$ .

A final note about entropy is that, in general, any process that makes the probability distribution more uniform increases its entropy [15].

### 2.3. Properties of Discrete Entropy

The following are properties of the discrete entropy function. Proofs and details can be found in textbooks [15].

1. Entropy is always non-negative,  $H(X) \geq 0$ .
2.  $H_\beta(X) = (\log_\beta a) H_a(X)$ .
3.  $H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n H(X_i)$ , with equality iff  $X_i$  are i.i.d.
4.  $H(X) \leq \log |X|$ , with equality iff  $X$  is distributed uniformly over  $X$ .
5.  $H(X)$  is concave.

### 2.4. Differential Entropy

Entropy was first formulated for discrete random variables, and was then generalized to continuous random variables in which case it is called *differential entropy* [20]. It has been related to the shortest description length, and thus, is similar to the entropy of a discrete random variable [21]. The differential entropy  $H(X)$  of a continuous random variable  $X$  with a density  $f(x)$  is defined as

$$H(f) = H(X) = - \int_S f(x) \log_2 f(x) dx, \quad (3)$$

where  $S$  is the support set of the random variable. It is well-known that this integral exists if and only if the density function of the random variables is Riemann-integrable [15,16]. The Riemann integral is fundamental in modern calculus. Loosely speaking, is the approximation of the area under any continuous curve given by the summation of ever smaller sub-intervals (*i.e.*, approximations), and implies a well-defined concept of limit [21].  $H(f)$  can also be used to denote differential entropy, and in the rest of the article, we shall employ this notation.

### 2.5. Asymptotic Equipartition Property of Continuous Random Variables

Given a set of i.i.d. random variables drawn from a continuous distribution with probability density  $f(x)$ , its differential entropy  $H(f)$  is given by

$$-\frac{1}{n} \log_2 (f(X_1, \dots, X_n)) \rightarrow \mathbb{E} [\log_2 (f(X))] = H(f), \quad (4)$$

such that  $n \rightarrow \infty$ . The convergence to expectation is a direct application of the weak law of large numbers.

### 2.6. Properties of Differential Entropy

1.  $H(f)$  depends on the coordinates. For different choices of coordinate systems for a given probability distribution  $P(X)$ , the corresponding differential entropies might be distinct.
2.  $H(f)$  is scale variant [15,22]. In this sense,  $H(af) = H(f) + \log_2 |a|$ , such that  $a \neq 0$ .
3.  $H(f)$  is translational invariant [15,16,22]. In this sense,  $H(f+c) = H(f)$ .

4.  $-\infty \leq H(f) \leq \infty$  [16]. The  $H(f)$  of a Dirac delta probability distribution, is considered the lowest  $H(f)$  bound, which corresponds to  $H(f) = -\infty$ .
5. Information measures such as relative entropy and mutual information are consistent, either in the discrete or continuous domain [22].

### 2.7. Differences between Discrete and Continuous Entropies

The derivation of Equation (3) comes from the assumption that its probability distribution is Riemann-integrable. If this is the case, then differential entropy can be defined just like discrete entropy. However, the notion of “average uncertainty” carried by the Equation (1) cannot be extended to its differential equivalent. Differential entropy is rather a function of the parameters of a distribution function, that describes *how uncertainty changes* as the parameters are modified [15].

To understand the differences between Equations (1) and (3), we will quantize a probability density function, and then calculate its discrete entropy [15,16].

First, consider the continuous random variable  $X$  with a probability density function  $f(x)$ . This function is then quantized by dividing its range into  $h$  bins of length  $\Delta$ . Then, in accordance to the Mean Value Theorem, within each  $h_i$  bin of size  $[i\Delta, (i+1)\Delta]$ , there exists a value  $x_i^*$  that satisfies

$$\int_{i\Delta}^{(i+1)\Delta} f(x) dx = f(x_i^*) \Delta. \quad (5)$$

Then, a quantized random variable  $X_i^\Delta$  is defined as

$$X_i^\Delta = x_i^*, \quad \text{if } i\Delta \leq X \leq (i+1)\Delta, \quad (6)$$

and, its probability is

$$p_i = \int_{i\Delta}^{(i+1)\Delta} f(x) dx = f(x_i^*) \Delta. \quad (7)$$

Consequently, the discrete entropy of the quantized variable  $X^\Delta$ , is formulated as

$$\begin{aligned} H(X^\Delta) &= - \sum_{-\infty}^{\infty} p_i \log_2 p_i \\ &= - \sum_{-\infty}^{\infty} (f(x_i^*) \Delta) \log_2 (f(x_i^*) \Delta) \\ &= - \sum \Delta f(x_i^*) \log_2 f(x_i^*) - \sum f(x_i^*) \Delta \log_2 \Delta \\ &= - \log_2 \Delta - \sum \Delta f(x_i^*) \log_2 f(x_i^*) \\ &\cong - \log_2 \Delta - \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx. \end{aligned} \quad (8)$$

To understand the final form of Equation (8), notice that as the size of each bin becomes infinitesimal,  $\Delta \rightarrow 0$ , the left-hand term of Equation (8) becomes  $\log_2(\Delta)$ . This is due to the fact that Equation (8) is a Riemann integral (as mentioned before)

$$\lim_{\Delta \rightarrow 0} \sum_{-\infty}^{\infty} f(x_i^*) \Delta = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Furthermore, as  $\Delta \rightarrow 0$ , the right-hand side of Equation (8) approximates the differential entropy of  $X$  such that

$$\lim_{\Delta \rightarrow 0} \sum_{-\infty}^{\infty} \Delta f(x_i^*) \log_2 f(x_i^*) = \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx.$$

Note that the left-hand side of Equation (8), explodes towards minus infinity such that

$$\lim_{\Delta \rightarrow 0} \log_2(\Delta) \approx -\infty.$$

Therefore, the difference between  $H(f)$  and  $H(X^\Delta)$  is  $H(f) - H(X^\Delta) = \log_2(\Delta)$ , which approaches to  $-\infty$  as the bin size becomes infinitesimal. Moreover, consistently with this is the fact that the differential entropy of a discrete value is  $-\infty$  [16].

Lastly, in accordance to [15], the average number of bits required to describe a continuous variable  $X$  with a  $n$ -bit accuracy (quantization) is  $H(X) + n \approx H(f)$  such that

$$H(X^\Delta)' = \lim_{\Delta \rightarrow 0} H(X^\Delta) + \log_2(\Delta) \rightarrow H(f). \quad (9)$$

### 3. Discrete Complexity Measures

Emergence  $E$ , self-organization  $S$ , and complexity  $C$  are close relatives of Shannon's entropy. These information-based measures inherit most of the properties of Shannon's discrete entropy [4], being the most valuable one that discrete entropy quantizes the average uncertainty of a probability distribution. In this sense, complexity  $C$  and its related measures ( $E$  and  $S$ ) are based on a quantization of the average information contained by a process described by its probability distribution.

#### 3.1. Emergence

Emergence has been used and debated for centuries [23]. Still, emergence can be understood [24]. The properties of a system are emergent if they are not present in their components, *i.e.*, global properties which are produced by local interactions are emergent. For example, the temperature of a gas can be said to be emergent [25], since the molecules do not possess such a property: it is a property of the collective. In a broad and informal way, emergence can be seen as differences in phenomena as they are observed at different scales [2,4].

Another form of entropy, rather related to the concept of information as uncertainty, is called emergence  $E$  [4]. Intuitively,  $E$  measures the ratio of uncertainty a process produces by new information that is a consequence of changes in (a) *dynamics* or (b) *scale* [4]. However, its formulation is more related to the thermodynamics entropy. Thus, it is defined as

$$E \equiv H(X) = -K \sum_{i=1}^N p_i \log_2 p_i, \quad (10)$$

where  $p_i = P(X = x)$  is the probability of the element  $i$ , and  $K$  is a normalizing constant.

Having the same equation to measure emergence, information, and entropy could be questioned. However, different phenomena such as gravity and electrostatic force are also described with the same equation. Still, this does not mean that gravitation and charge are the same. In the same fashion, there are differences between emergence, information, and entropy, which depend more on their usage and interpretation than on the equation describing them. Thus, it is justified to use the same expression to measure different phenomena [10].

#### 3.2. Multiple Scales

In thermodynamics, the Boltzmann constant  $K$ , is employed to normalize the entropy in accordance to the probability of each state. However, Shannon's entropy typical formulation [15–17] neglects the usage of  $K$  in Equation (10) (being its only constraint that  $K > 0$ , [4]). Nonetheless, for emergence as a measure of the average production of information for a given distribution,  $K$  plays

a fundamental role. In the cybernetic definition of *variety* [18],  $K$  is a function of the distinct states a system can be, *i.e.*, the system's alphabet size. Formally, it is defined as

$$K = \frac{1}{\log_2(b)}, \quad (11)$$

where  $b$  corresponds to the size of the alphabet of the sample or bins of a discrete probability distribution. Furthermore,  $K$  should guarantee that  $0 \leq E \leq 1$ ; therefore,  $b$  should be at least equal to the number of bins of the discrete probability distribution.

It is also worth noting that the denominator of Equation (11),  $\log_2(b)$ , is equivalent to the maximum entropy for a continuous distribution function, the uniform distribution. Consequently, emergence can be understood as the ratio between the entropy for given distribution  $P(X)$ , and the maximum entropy for the same alphabet size  $H(U)$  [26], is

$$E = \frac{H(P(X))}{H(U)}. \quad (12)$$

### 3.3. Self-Organization

Self-organisation,  $S$ , is the complement of  $E$ . In this sense, with more uncertainty less predictability is achieved, and *vice versa*. Thus, an entirely random process (e.g., uniform distribution) has the lowest organization, and a completely deterministic system one (Dirac delta distribution), has the highest. Furthermore, an extremely organized system yields no information with respect of *novelty*, while, on the other hand, the more chaotic a system is, the more information is yielded [4,26].

The metric of self-organization  $S$  was proposed to measure the organization a system has regarding its average uncertainty [4,27].  $S$  is also related to the cybernetic concept of *constraint*, which measures changes in due entropy restrictions on the state space of a system [8]. These constraints confine the system's behavior, increasing its predictability, and reducing the (novel) information it provides to an observer. Consequently, the more self-organized a system is, the less average uncertainty it has. Formally,  $S$  is defined as

$$S = 1 - E = 1 - \left( \frac{H(P(X))}{H(U)} \right), \quad (13)$$

such that  $0 \leq S \leq 1$ . It is worth noting that, the maximal  $S$  (*i.e.*,  $S = 1$ ) is only achievable when the entropy for a given probability density function (PDF) is such that  $H(P(X)) \rightarrow 0$ , which corresponds to the entropy of a Dirac delta (only in the discrete case).

### 3.4. Complexity

Complexity  $C$  can be described as a *balance* between order (stability), and chaos (scale or dynamical changes) [4]. More precisely, this function describes a system's behavior in terms of the average uncertainty produced by its probability distribution in relation the dynamics of a system. Thus, the complexity measure is defined as

$$C = 4 \cdot E \cdot S, \quad (14)$$

such that,  $0 \leq C \leq 1$ .

## 4. Continuous Complexity Measures

As mentioned before, discrete and differential entropies do not share the same properties. In fact, the property of discrete entropy as the average uncertainty in terms of probability, cannot be extended to its continuous counterpart. As a consequence, the proposed continuous information-based measures describe how the production of information changes respect to the probability distribution

parameters. In particular, this characteristic could be employed as a feature selection method, where the most relevant variables are those which have a high emergence (the most *informative*).

The proposed measures are differential emergence ( $E_D$ ), differential self-organization ( $S_D$ ), and differential complexity ( $C_D$ ). However, given that the interpretation and formulation (in terms of emergence) of discrete and continuous  $S$  (Equation (13)) and  $C$  (Equation (14)) are the same, we only provide details on  $E_D$ . The difference between  $S_D$ ,  $C_D$  and  $S$ ,  $C$  is that the former are defined on  $E_D$ , while the latter are on  $E$ . Furthermore, we make an emphasis in the definition of the normalizing constant  $K$ , which plays a significant role in constraining  $E_D \in [0, 1]$ , and consequently,  $S_D$  and  $C_D$  as well.

#### 4.1. Differential Emergence

As for its discrete form, the emergence for continuous random variables is defined as

$$E_D = -K \int_v^{\zeta} f(X) \log_2 f(X), \quad (15)$$

where,  $[v, \zeta]$  is the domain, and  $K$  stands for a normalizing constant related to the distribution's alphabet size. It is worth noting that this formulation is highly related to the view of emergence as the ratio of information production of a probability distribution respect the maximum differential entropy for the same range. However, since  $E_D$  can be negative (*i.e.*, entropy of a single discrete value), we choose  $E_D$  such that

$$E'_D = \begin{cases} E_D & E_D > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

$E'_D$  is rather a more convenient function than  $E_D$ , as  $0 \leq E'_D \leq 1$ . This statement is justified in the fact that the differential entropy of a discrete value is  $-\infty$  [15]. In practice, differential entropy becomes negative only when the probability distribution is extremely narrow, *i.e.*, there is a high probability for few states. In the context of information changes due to parameter manipulation, an  $E_D < 0$  means that the probability distribution is becoming a Dirac delta distribution. For notation convenience, from now on we will employ  $E_D$  and  $E'_D$  interchangeably.

#### 4.2. Multiple Scales

The  $K$  constant expresses the relation between uncertainty of a given  $P(X)$  Defined by  $H(X)$ , respect to the entropy of a maximum entropy over the same domain [26]. In this setup, as the uncertainty grows,  $E'_D$  becomes closer to unity.

To constrain the value of  $H(X) = [0, 1]$  in the discrete emergence case, it was enough to establish the distribution's alphabet size,  $b$  of Equation (10), such that  $b \geq \# \text{bins}$  [4]. However, for any PDF, the number of elements between a pair of points  $a$  and  $b$ , such that  $a \neq b$ , is infinite. Moreover, as the size of each bin becomes infinitesimal,  $\Delta \rightarrow 0$ , the entropy for each bin becomes  $-\infty$  [15]. In addition, it has been stated that  $b$  value should be equal to the cardinality of  $X$  [26]; however, this applies only to discrete emergence. Therefore, rather than a generalization, we propose an heuristic for the selection of a proper  $K$  in the case of differential emergence. Moreover, we differentiate between  $b$  for  $H(f)$ , and  $b'$  for  $H(X^\Delta)$ .

As in the discrete case,  $K$  is defined as Equation (11). In order to determine the proper alphabet size  $b$ , we propose the next algorithm:

1. If we know *a priori* the true  $P(X)$ , we calculate  $H(f)$ , and  $b = |P(X)|$  is the cardinality within the interval of Equation (15). In this sense, a large value will denote the cardinality of a “ghost” sample [16]. (It is *ghost*, in the concrete sense that it does not exist. Its only purpose is to provide a bound for the maximum entropy accordingly to some large alphabet size.)



- If we do not know the true  $P(X)$ , or we are interested rather in  $H(X^\Delta)'$  where a sample of finite size is involved, we calculate  $b'$  as

$$b' = \sum_i ind(x_i), \quad (17)$$

such that, the non-negative function  $ind(\cdot)$  is defined as

$$ind(x_i) = \begin{cases} 1 & \text{iff } P(x_i) > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (18)$$

For instance, in the quantized version of the standard normal distribution ( $N(0, 1)$ ), only values within  $\pm 3\sigma$  satisfy this constraint despite the domain of Equation (15). In particular, if we employ  $b = |X|$  rather than  $b'$ , we compress the  $E_D$  value as it will be shown in the next section. On the other hand, for a uniform distribution or a Power-Law (such that  $0 < x_{min} < x$ ), the whole range of points satisfies this constraint.

## 5. Probability Density Functions

In communication and information theory, uniform (U) and normal, also known as Gaussian (G) distributions play a significant role. Both are referent to maximum entropy: on the one hand, U has the maximum entropy within a continuous domain; on the other hand, G has the maximum entropy for distributions with a fixed mean ( $\mu$ ), and a finite support set for a fixed standard deviation ( $\sigma$ ) [15,16]. Moreover, as mentioned earlier,  $H(f)$  is useful when comparing the entropies of two distributions over some reference space [15,16,28]. Consequently, U, but mainly G, are heavily used in the context of telecommunications for signal processing [16]. Nevertheless, many natural and man-made phenomena can be approximated with Power-Law (PL) distributions. These types of distributions typically present complex patterns that are difficult to predict, making them a relevant research topic [29]. Furthermore, Power-Laws have been related to the presence of multifractal structures in certain types of processes [28]. Moreover, Power-Laws are tightly related to self-organization and criticality theory, and have been studied under information frameworks before (e.g., Tsallis's, and Renyi's maximum entropy principle) [29,30].

Therefore, in this work we focus our attention to these three PDFs. First, we provide a short description of each PDF, then, we summarize its formulation, and the corresponding  $H(f)$  in Table 1.

**Table 1.** Studied PDFs with their corresponding analytical differential entropies.

Distribution	PDF	Differential Entropy
Uniform	$p(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$	$H(p(x)) = \log_2(b)$
Normal	$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$H(p(x)) = \frac{1}{2} \log_2(2\pi e\sigma^2)$
Power-law	$p(x) = \left(\frac{\alpha-1}{x_{min}}\right) \left(\frac{x}{x_{min}}\right)^{-\alpha}$	$H(p(x)) = \log_2(x_{min}) - \log_2(\alpha-1) + \left(\frac{\alpha}{\alpha-1}\right)$

### 5.1. Uniform Distribution

The simplest PDF, as its name states, establishes that for each possible value of  $X$ , the probability is constant over the whole support set (defined by the range between  $a$  and  $b$ ), and 0 elsewhere. This PDF has no parameters besides the starting and ending points of the support set. Furthermore, this distribution appears frequently in signal processing as white noise, and it has the maximum entropy for continuous random variables [16].



Its PDF and its corresponding  $H(f)$  are shown in first row of Table 1. It is worth noting that, as the cardinality of the domain of  $U$  grows, its differential entropy increases as well.

### 5.2. Normal Distribution

The normal or Gaussian distribution is one of the most important probability distribution families [31]. It is fundamental in the central limit theorem [16], time series forecasting models such as classical autoregressive models [32], modelling economic instruments [33], encryption, modelling electronic noise [16], error analysis and statistical hypothesis testing.

Its PDF is characterized by a symmetric, bell-shaped function whose parameters are: location (*i.e.*, mean  $\mu$ ), and dispersion (*i.e.*, standard deviation  $\sigma^2$ ). The standard normal distribution is the simplest and most used case of this family, its parameters are  $N(\mu = 0, \sigma^2 = 1)$ . A continuous random variable  $x \in X$  is said to belong to a Gaussian distribution,  $X \sim N(\mu, \sigma^2)$ , if its PDF  $p(x)$  is given by the one described in the second row of Table 1. As is shown in the table, the differential entropy of  $G$  only depends on the standard deviation. Furthermore, it is well known that its differential entropy is monotonically increasing concave in relation to  $\sigma$  [31]. This is consistent with the aforementioned fact that  $H(f)$  is translation-invariant. Thus, as  $\sigma$  grows, so does the value of  $H(G)$ , while as  $\sigma \rightarrow 0$  such that  $0 < \sigma < 1$ , it becomes a Dirac delta with  $H(f) \approx 0$ .

### 5.3. Power-Law Distribution

Power-Law distributions are commonly employed to describe multiple phenomena (e.g., turbulence, DNA sequences, city populations, linguistics, cosmic rays, moon craters, biological networks, data storage in organisms, chaotic open systems, and so on) across numerous scientific disciplines [28–30,34–37]. These type of processes are known for being scale invariant, being the typically scales ( $\alpha$ , see below) in nature between one and 3.5 [30]. In addition, the closeness of this type of PDF to chaotic systems and fractals is such that, some fractal dimensions are called entropy dimensions (e.g., box-counting dimension, and Renyi entropy) [36].

Power-Law distributions can be described by continuous and discrete distributions. Furthermore, Power-laws in comparison with Normal distribution, generate events of large orders of magnitude more often, and are not well represented by a simple mean. A Power-Law density distribution is defined as

$$p(x)dx = P(x \leq X \leq x + dx) = Cx^{-\alpha}dx, \quad (19)$$

such that,  $C$  is a normalization factor,  $\alpha$  is the scale exponent, and  $X | x > x_{min} > 0$  is the observed continuous random variable. This PDF diverges as  $x \rightarrow 0$ , and do not hold for all  $x \geq 0$  [37]. Thus,  $x_{min}$  corresponds to lower bound of a Power-Law. Consequently, in Table 1, we provide the PDF of a Power-Law as proposed by [35], and its corresponding  $H(f)$  as proposed by [38].

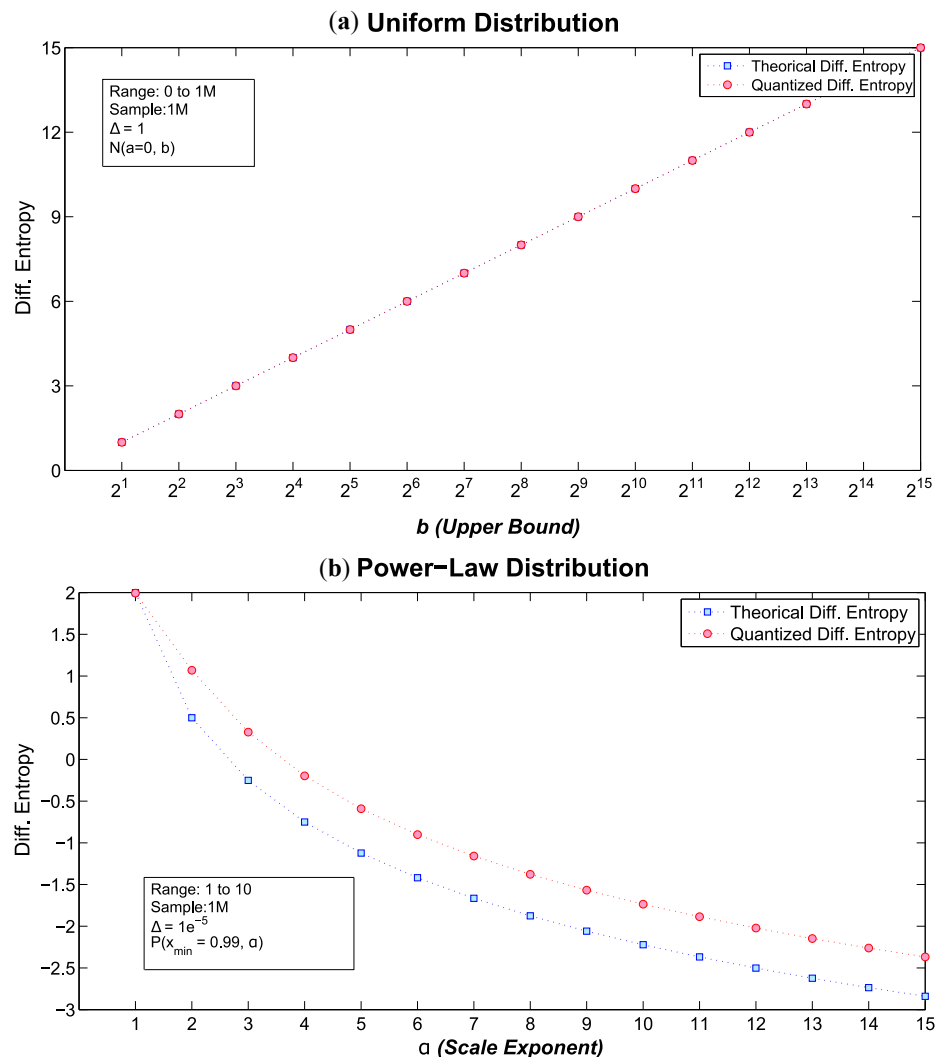
The aforementioned PDFs, and their corresponding  $H(f)$  are shown in Table 1. Further details about the derivation of  $H(f)$  for  $U$ , and  $G$  can be found in [15,16]. For additional details on the differential entropy of the Power-Law, we refer the reader to [28,38].

## 6. Results

In this section, comparisons of theoretical *vs.* quantized differential entropy for the PDFs considered are shown. Next, we provide differential complexity results ( $E'_D$ ,  $S_D$ , and  $C_D$ ) for the mentioned PDFs. Furthermore, in the case of Power-Laws, we also provide and discuss the corresponding complexity measures results for real world phenomena, already described in [39]. In addition, it is worth noting that, since quantized  $H(f)$  of the Power-Law yielded poor results, the Power-Law's analytical  $H(f)$  form was used.

### 6.1. Theoretical vs. Quantized Differential Entropies

Numerical results of theoretical and quantized differential entropies are shown in Figures 1 and 2. Analytical  $H(f)$  results are displayed in blue, whereas the quantized  $H(X^\Delta)'$  ones are shown in red. For each PDF, a sample of one million (*i.e.*,  $1 \times 10^6 \equiv 1\text{M}$ ) points were employed for calculations. The bin size  $\Delta$  required by  $H(X^\Delta)'$ , is obtained as the ratio  $\Delta = \frac{\text{Range}}{|\text{Sample}|}$ . However, the value of  $\Delta$  has considerable influence in the resulting quantized differential entropy.



**Figure 1.** Theoretical and quantized differential entropies for (a) Uniform distribution and (b) Power-Law distributions.

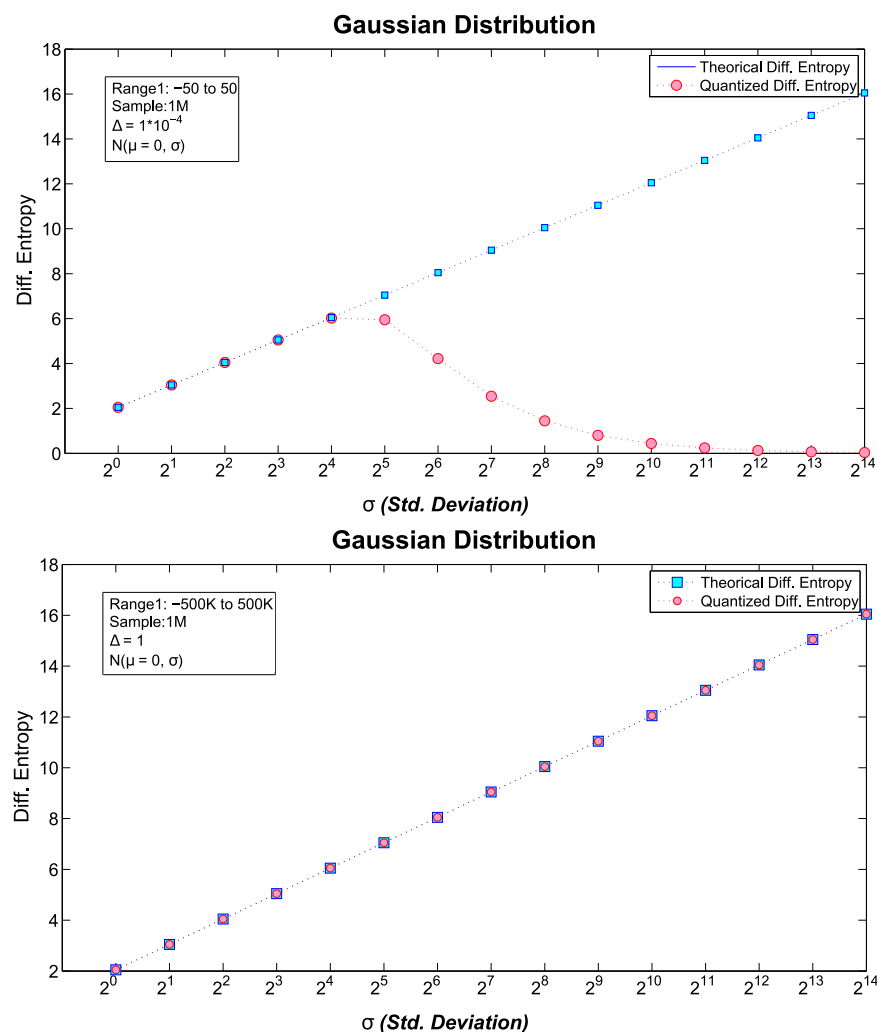
#### 6.1.1. Uniform Distribution

The results for U were expected. We tested several values of the cardinality of  $P(X)$ , such that  $b = 2^i \mid i = 1, \dots, 15$ . Using the analytical  $H(f)$  formula of Table 1, the quantized  $H(X^\Delta)'$ , and  $\Delta = 1$ , we achieved exactly the same differential entropy values. Results for U are shown in the left side of Figure 1. As was mentioned earlier, as the cardinality of the distribution grows, so does the differential entropy of U.

### 6.1.2. Normal Distribution

Results for the Gaussian distribution were less trivial. As in the U case, we calculate both  $H(f)$  and  $H(X^\Delta)'$ , for a fixed  $\mu = 0$ , and modified the standard deviation parameter such that,  $\sigma = 2^i \mid i = 0, 1, \dots, 14$ . Notice that the first tested distribution is the *standard normal distribution*.

In Figure 2, results obtained for the n-bit quantized differential entropy, and for the analytical form of Table 1 are shown. Moreover, we displayed two cases of the normal distribution: the left side of Figure 2 shows results for  $P(X)$  with range  $[-50, 50]$  and a bin size,  $\Delta = \frac{100}{1M} = 1 \times 10^{-4}$ , whereas, the right side provides results for a  $P(X)$  with range  $[-500e3, 500e3]$  and  $\Delta = 1$ . It is worth noting that, in the former case, the quantized differential entropy shows a discrepancy with  $H(f)$  after only  $\sigma = 2^4 = 16$ , which quickly increases with growing  $\sigma$ . On the other hand, for the latter case, there is an almost perfect match between the analytical and quantized differential entropies; however, the same mismatch will be observed if the standard deviation parameter is allowed to grow unboundedly ( $\sigma \rightarrow \infty$ ). Nonetheless, this is a consequence of how  $H(X^\Delta)'$  is computed. As mentioned earlier, as  $\Delta \rightarrow 0$ , the value of each quantized  $X^\Delta$  grows towards  $-\infty$ . Therefore, in the G case, it seems convenient employing a Probability Mass Function (PMF) rather than a PDF. Consequently, the experimental setup of right side image of Figure 2 is employed for the calculation of the continuous complexity measures of G.



**Figure 2.** Two comparisons of theoretical *vs.* quantized differential entropy for the Gaussian distribution.

### 6.1.3. Power-Law Distribution

Results for the Power-Law distribution are shown in the right side of Figure 1. In both U and G, a PMF instead of a PDF was used to avoid cumbersome results (as depicted in the corresponding images). However, for the Power-Law distribution, the use of a PDF is rather convenient. As discussed in the next section and highlighted by [35],  $x_{min}$  has a considerable impact on the value of  $H(f)$ . For Figure 1, the range employed was  $[1, 50]$ , with a bin size of  $\Delta = 1 \times 10^{-5}$ , a  $x_{min} = 0.99$ , and modified the scale exponent parameter such that,  $\alpha = i \mid i = 1, \dots, 15$ . For this particular setup, we can observe that as  $\alpha$  increases,  $H(f)$  and  $H(X^\Delta)$  decreases its value towards  $-\infty$ . This effect is consequence of increasing the scale of the Power-Law such that, the slope of the function in a log-log space, approaches to zero. In this sense, with larger  $\alpha$ s, the  $P(X)$  becomes closer to a Dirac delta distribution, thus,  $H(f) \rightarrow -\infty$ . However, as will be discussed later, for larger  $\alpha$ s, larger  $x_{min}$  values are required, in order for  $H(f)$  to display positive values.

### 6.2. Differential Complexity: $E_D$ , $S_D$ , and $C_D$

U results are trivial:  $E_D = 1$ , and  $S_D, C_D = 0$ . For each upper bound of U,  $E'_D = \frac{H(U)}{H(U)} = 1$ , which is exactly the same as its discrete counterpart. Thus, U results are not considered in the following analysis.

Continuous complexity results for G and PL are shown in Figures 3 and 4, respectively. In the following, we provide details of these measures.

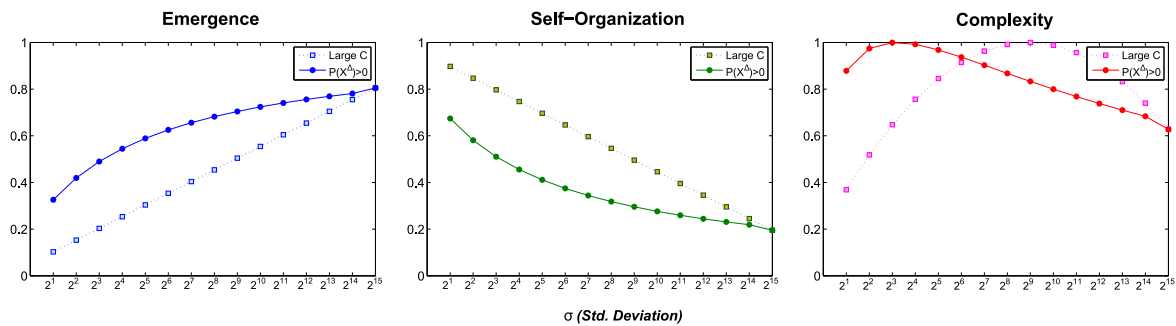


Figure 3. Complexity of the Gaussian distribution.

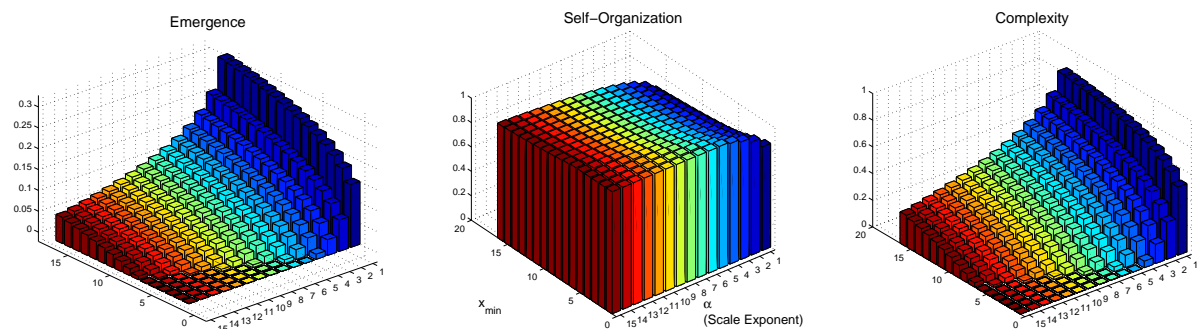


Figure 4. Complexity measures for the Power-Law. Lower values of the scale exponent  $\alpha$  are displayed in dark blue, colors turns into reddish for larger scale exponents.

#### 6.2.1. Normal Distribution

It was stated in Section 4 that the size of the alphabet is given by the function  $ind(P(X))$ . This rule establishes a valid cardinality such that  $P(X) > 0$ , thus, only those states with a positive probability are considered. For  $P(X^\Delta)$ , such operation can be performed. Nevertheless, when the

analytical  $H(f)$  is used, the proper cardinality of the set is unavailable. Therefore, in the Gaussian distribution case, we tested two criteria for selecting the value of  $b$ :

1.  $\sum_{x_i} \text{ind}(\cdot)$  is employed for  $H(X^\Delta)'$ .
2. A constant with a large value ( $C = 1 \times 10^6$ ) is used for the analytical formula of  $H(f)$ .

In Figure 3, solid dots are used when  $K$  is equal to the cardinality of  $P(X) > 0$ , whereas solid squares are used for an arbitrary large constant. Moreover, for the quantized case of  $P(G)$ , Table 2 shows the cardinality for each sigma,  $b'_i$ , and its corresponding  $K_i$ . As can be observed, for a large normalizing constant  $K$ , a logarithmic relation is displayed for  $E_D$  and  $S_D$ . In addition, the maximum  $C_D$  is achieved for  $\sigma = 2^8 = 256$ , which is where  $E_D = S_D$ . However, for  $H(X^\Delta)'$  the the maximum  $C_D$  is found around  $\sigma = 2^{1,2,3} = 2, 4, 8$ , such that  $C_D \leq \epsilon \mid \epsilon \rightarrow 0$ . A word of advice must be made here. The required cardinality to normalize the continuous complexity measures such that  $0 \leq E_D, S_D, C_D \leq 1$ , must have a lower bound. This bound should be related to the scale of the  $P(X)$  [40], and the quantization size  $\Delta$ . In our case, when a large cardinality  $|U| = 1 \times 10^6$ , and  $\Delta = 1$  are used, the normalizing constant flattens  $E_D$  results respect those obtained by  $b'$ ; moreover, the large constant increases  $S_D$ , and takes greater standard deviations for achieving the maximum  $C_D$ . However, these complexity results are rather artificial in the sense that, if we arbitrarily let  $|U| \rightarrow \infty$  then trivially we will obtain  $E_D = 0$ ,  $S_D = 1$ , and  $C_D = 0$ . Moreover, it has been stated that the cardinality of  $P(X)$  should be employed as a proper size of  $b$  [26]. Therefore, when  $H(X^\Delta)'$  is employed, the cardinality of  $P(X) > 0$  must be used. On the contrary, when  $H(f)$  is employed, a coarse search for increasing alphabet sizes could be used so that the maximal  $H(f)$  satisfies  $\frac{H(f)}{H(U)} \leq 1$ .

**Table 2.** Alphabet size  $b'$ , and its corresponding normalizing  $K$  constant for the normal distribution  $G$ .

$\sigma$	$b' = \sum \text{ind}(\text{Pr}(X) > 0)$	$H(U) = \log_2(b')$	$K = \frac{1}{H(U)}$
$2^0 = 1$	78	6.28	0.16
$2^1 = 2$	154	7.26	0.14
$2^2 = 4$	308	8.27	0.12
$2^3 = 8$	616	9.27	0.11
$2^4 = 16$	1232	10.27	0.10
$2^5 = 32$	2464	11.27	0.09
$2^6 = 64$	4924	12.27	0.08
$2^7 = 128$	9844	13.27	0.075
$2^8 = 256$	19,680	14.26	0.0701
$2^9 = 512$	39,340	15.26	0.0655
$2^{10} = 1024$	78,644	16.26	0.0615
$2^{11} = 2048$	157,212	17.26	0.058
$2^{12} = 4096$	314,278	18.26	0.055
$2^{13} = 8192$	628,258	19.26	0.0520
$2^{14} = 16,384$	1,000,000	19.93	0.050

### 6.2.2. Power-Law Distribution

In this case,  $H(f)$  rather than  $H(X^\Delta)'$  is used for computational convenience. Although the cardinality of  $P(X) > 0$  is not available, by simply substituting  $p(x_i) > 0 \mid x = \{1, \dots, 1 \times 10^6\}$  we can see that the condition is fulfilled by the whole set. Therefore, the large  $C$  criterium, earlier detailed, is used. Still, given that a numerical Power-Law distribution is given by two parameters, a lower bound  $x_{min}$  and the scale exponent  $\alpha$ , we depict our results in 3D in Figure 4. From left to right,  $E_D$ ,  $S_D$ , and  $C_D$  for the Power-Law distribution are shown, respectively. In the three images, the same coding is used:  $x$ -axis displays the scale exponent ( $\alpha$ ) values,  $y$ -axis shows  $x_{min}$  values, and  $z$ -axis depicts the continuous measure values; lower values of  $\alpha$  are displayed in dark blue, turning into reddish colors for larger exponents.

As it can be appreciated in Figure 4, for small  $x_{min}$  (e.g.,  $x_{min} = 1$ ) values, low emergence is produced despite the scale exponent. Moreover, maximal self-organization (i.e.,  $S_D = 1$ ) is quickly achieved (i.e.,  $\alpha = 4$ ), providing a PL with at most fair complexity values. However, if we let  $x_{min}$  take larger numbers,  $E_D$  grows, achieving the maximal complexity (i.e.,  $C_D \approx 0.8$ ) of this experimental setup at  $x_{min} = 15$ ,  $\alpha = 1$ . This behavior is also observed for other scale exponent values, where emergence of new information is produced as the  $x_{min}$  value grows. Furthermore, it has been stated that for  $P(X)$ , which displays a Power-Law behavior, it is required that  $\forall x_i \in P(X) \mid x_i > x_{min}$  [37]. Thus, for every  $\alpha$  there should be an  $x_{min}$  such that  $E_D > 0$ . Moreover, for larger scale exponents, larger  $x_{min}$  values are required for the distribution shows emergence of new information at all.

### 6.3. Real World Phenomena and Their Complexity

Data of phenomena that follows a Power-Law is provided in Table 3. These Power-Laws have been studied by [35,37,39], and the Power-Law parameters were published by [39]. The phenomena in the table mentioned above compromises data from:

1. Numbers of occurrences of words in the novel Moby Dick by Hermann Melville.
2. Numbers of citations to scientific papers published in 1981, from the time of publication until June 1997.
3. Numbers of hits on websites by users of America Online Internet services during a single day.
4. Number of received calls to A.T.&T. U.S. long-distance telephone services on a single day.
5. Earthquake magnitudes occurred in California between 1910 and 1992.
6. Distribution of the diameter of moon craters.
7. Peak gamma-ray intensity of solar flares between 1980 and 1989.
8. War intensity between 1816–1980, where intensity is a formula related to the number of deaths and warring nations populations.
9. Frequency of family names accordance with U.S. 1990 census.
10. Population per city in the U.S. in agreement with U.S. 2000 census.

More details about these Power-Laws can be found in [35,37,39].

For each phenomenon, the corresponding differential entropy and complexity measures are shown in Table 3. Furthermore, we also provide Table 4 which is a color coding for complexity measures proposed in [4]. Five colors are employed to simplify the different value ranges of  $E_D$ ,  $S_D$ , and  $C_D$  results. According to the nomenclature suggested in [4], results for these sets show that, very high complexity  $0.8 \leq C_D \leq 1$  is obtained by the number of citations set (i.e., 2), and intensity of solar flares (i.e., 7). High complexity,  $0.6 \leq C_D < 0.8$  is obtained for received telephone calls (i.e., 4), intensity of wars (i.e., 8), and frequency of family names (i.e., 9). Fair complexity  $0.4 \leq C_D < 0.6$  is displayed by earthquakes magnitude (i.e., 5), and population of U.S. cities (i.e., 10). Low complexity,  $0.2 \leq C_D < 0.4$  is obtained for frequency of used words in Moby Dick (i.e., 1) and web hits (i.e., 3), whereas, moon craters (i.e., 6) have very low complexity  $0 \leq C_D < 0.2$ . In fact, earthquakes, and web hits, have been found not to follow a Power-Law [35]. Furthermore, if such sets were to follow a Power-Law, a greater value of  $x_{min}$  would be required as can be observed in Figure 4. In fact, the former case is found for the frequency of words used in Moby Dick. In [39], parameters of Table 3 are proposed. However, in [35], another set of parameters are estimated (i.e.,  $x_{min} = 7$ ,  $\alpha = 1.95$ ). For the more recent estimated set of parameters, a high complexity is achieved (i.e.,  $C_D = 0.74$ ), which is more consistent with literature about Zipf's law [39]. Lastly, in the case of moon craters, the  $x_{min} = 0.01$  is rather a poor choice according to Figure 4. For the chosen scale exponent, it would require at least a  $x_{min} \approx 1$ , for the Power-Law to produce any information at all. It should be noted that  $x_{min}$  can be adjusted to change the values of all measures. In addition, it is worth mentioning that if we were to normalize and discretize a Power-Law distribution to calculate its discrete entropy (as in [4]), all Power-Law distributions present a very high complexity, independently of  $x_{min}$  and  $\alpha$ ,



precisely because these are normalized. Still, this is not useful for comparing different Power-Law distributions.

**Table 3.** Power-Law parameters and information-based measures of real world phenomena.

	Phenomenon	$x_{min}$	$\alpha$ (Scale Exponent)	$H(f)$	$E_D'$	$S_D$	$C_D$
1	Frequency of use of words	1	2.2	1.57	0.078	0.92	0.29
2	Number of citations to papers	100	3.04	7.1	0.36	0.64	0.91
3	Number of hits on web sites	1	2.4	1.23	0.06	0.94	0.23
4	Telephone calls received	10	2.22	4.85	0.24	0.76	0.74
5	Magnitude of earthquakes	3.8	3.04	2.38	0.12	0.88	0.42
6	Diameter of moon craters	0.01	3.14	−6.27	0	1	0
7	Intensity of solar flares	200	1.83	10.11	0.51	0.49	0.99
8	Intensity of wars	3	1.80	4.15	0.21	0.79	0.66
9	Frequency of family names	10000	1.94	15.44	0.78	0.22	0.7
10	Population of U.S. cities	40000	2.30	16.67	0.83	0.17	0.55

**Table 4.** Color coding for  $E_D$ ,  $S_D$ , and  $C_D$  results.

Category	Very High	High	Fair	Low	Very Low
Range	[0.8, 1]	[0.6, 0.8)	[0.4, 0.6)	[0.2, 0.4)	[0, 0.2)
Color	Blue	Green	Yellow	Orange	Red

## 7. Discussion

In this paper, we extended complexity measures based on Shannon’s entropy to the continuous domain. Shannon’s continuous entropy cannot measure the average predictability of a system as its discrete counterpart. Rather, it measures the average uncertainty of a system given a configuration of a probability distribution function. Therefore, continuous Emergence, Self-organization, and Complexity describe the expected predictability of a system given it follows a probability distribution with a specific parameters set. It is common in many disciplines to describe real world systems using a particular probability distribution function. Therefore, the proposed measures can be useful to describe the production of information and novelty in such cases, or how information and uncertainty would change in the system if parameters were perturbed. Certainly, the *interpretation* of the measures is not given, as this will depend on the use we make of the measures for specific purposes.

From exploring the parameter space of the uniform, normal, and scale-free distributions, we can corroborate that high complexity values require a form of *balance* between extreme cases. On the one hand, uniform distributions, by definition, are homogeneous and thus all states are equiprobable, yielding the highest emergence. This is also the case of normal distributions with a very large standard deviation and for Power-Law distributions with an exponent close to zero. On the other hand, highly biased distributions (very small standard deviation in G or very large exponent in PL) yield a high self-organization, as few states accumulate most of the probability. Complexity is found between these two extremes. From the values of  $\sigma$  and  $\alpha$ , this coincides with a broad range of phenomena. This does not tell us something new: complexity is common. The relevant aspect is that this provides a common framework to study of the processes that lead phenomena to have a high complexity [41]. It should be noted that this also depends on the time scales at which change occurs [42].

Many real-world phenomena are modelled under some probability distribution assumption. These impositions are rather a consequence of analytical convenience than intrinsic to the phenomenon under study. Nevertheless, these usually provide a rough but useful approximation for explanatory purposes: ranging from models of wind speed (Weibull distribution) to economic models (Logistic distribution), PDFs “effectively” describe phenomena data. Thus, in future work, it



would be very useful to characterize PDFs like the Laplace, Logistic, and Weibull distribution to shed light in terms of complexity, emergence, and self-organization of systems, which can be modelled by these PDFs.

On the other hand, it is interesting to relate our results with information adaptation [17]. In a variety of systems, adaptation takes place by *inflating* or *deflating* information, so that the “right” balance is achieved. Certainly, given that it is possible to derive upper and lower bounds for the differential entropy of a PDF (e.g., [20]), it should be also possible to define analytical bounds for the complexity measures for the given PDF. However, for practical purposes, complexity measures are constrained by the selected range of the PDF parameters. Thus, the precise balance change from system to system and from context to context. Still, the capability of information adaptation has to be correlated with complexity, as the measure also reflects a balance between emergence (inflated information) and self-organization (deflated information).

As future work, it will be interesting to study the relationship between complexity and semantic information. There seems to be a connection with complexity as well, as we have proposed a measure of *autopoiesis* as the ratio of the complexity of a system over the complexity of its environment [4,43]. These efforts should be valuable in the study of the relationship between information and meaning, in particular in cognitive systems.

Another future line of research lies in the relationship between the proposed measures and complex networks [44–47], exploring questions such as: how does the topology of a network affect its dynamics? How much can we predict the dynamics of a network based on its topology? What is the relationship between topological complexity and dynamic complexity? How controllable are networks [48] depending on their complexity?

**Acknowledgments:** Guillermo Santamaría-Bonfil was supported by the Universidad Nacional Autónoma de México (UNAM) under grant CJIC/CTIC/0706/2014. Carlos Gershenson was supported by projects 212802 and 221341 from the Consejo Nacional de Ciencia y Tecnología, and membership 47907 of the Sistema Nacional de Investigadores of México.

**Author Contributions:** Guillermo Santamaría-Bonfil, Carlos Gershenson conceived and designed the experiments; Guillermo Santamaría-Bonfil performed the experiments; Guillermo Santamaría-Bonfil, Nelson Fernández and Carlos Gershenson wrote the paper. All authors have read and approved the final manuscript.

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

## References

1. Gershenson, C. (Ed.) *Complexity: 5 Questions*; Automatic Peess/VIP: Copenhagen, Denmark, 2008.
2. Prokopenko, M.; Boschetti, F.; Ryan, A. An Information-Theoretic Primer on Complexity, Self-Organisation and Emergence. *Complexity* **2009**, *15*, 11–28.
3. Gershenson, C.; Fernández, N. Complexity and Information: Measuring Emergence, Self-organization, and Homeostasis at Multiple Scales. *Complexity* **2012**, *18*, 29–44.
4. Fernández, N.; Maldonado, C.; Gershenson, C. Information Measures of Complexity, Emergence, Self-organization, Homeostasis, and Autopoiesis. In *Guided Self-Organization: Inception*; Prokopenko, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2014; Volume 9, pp. 19–51.
5. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
6. Gershenson, C.; Heylighen, F. When Can We Call a System Self-Organizing? In *Advances in Artificial Life*; Banzhaf, W., Ziegler, J., Christaller, T., Dittrich, P., Kim, J.T., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 606–614.
7. Langton, C.G. Computation at the Edge of Chaos: Phase Transitions and Emergent Computation. *Physica D* **1990**, *42*, 12–37.
8. Kauffman, S.A. *The Origins of Order*; Oxford University Press: Oxford, UK, 1993.
9. Lopez-Ruiz, R.; Mancini, H.L.; Calbet, X. A statistical measure of complexity. *Phys. Lett. A* **1995**, *209*, 321–326.
10. Zubillaga, D.; Cruz, G.; Aguilar, L.D.; Zapotécatl, J.; Fernández, N.; Aguilar, J.; Rosenblueth, D.A.; Gershenson, C. Measuring the Complexity of Self-organizing Traffic Lights. *Entropy* **2014**, *16*, 2384–2407.

11. Amoretti, M.; Gershenson, C. Measuring the complexity of adaptive peer-to-peer systems. *Peer-to-Peer Netw. Appl.* **2015**, 1–16, doi:10.1007/s12083-015-0385-4.
12. Febres, G.; Jaffe, K.; Gershenson, C. Complexity measurement of natural and artificial languages. *Complexity* **2015**, 20, 25–48.
13. Santamaría-Bonfil, G.; Reyes-Ballesteros, A.; Gershenson, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renew. Energy* **2016**, 85, 790–809.
14. Fernández, N.; Villate, C.; Terán, O.; Aguilar, J.; Gershenson, C. Complexity of Lakes in a Latitudinal Gradient. *Ecol. Complex.* **2016**, submitted.
15. Cover, T.; Thomas, J. *Elements of Information Theory*; John Wiley Sons: Hoboken, NJ, USA, 2005; pp. 1–748.
16. Michalowicz, J.; Nichols, J.; Bucholtz, F. *Handbook of Differential Entropy*; CRC Press: Boca Raton, FL, USA, 2013; pp. 19–43.
17. Haken, H.; Portugali, J. *Information Adaptation: The Interplay Between Shannon Information and Semantic Information in Cognition*; Springer: Berlin/Heideberg, Germany, 2015.
18. Heylighen, F.; Joslyn, C. Cybernetics and Second-Order Cybernetics. In *Encyclopedia of Physical Science & Technology*, 3rd ed.; Meyers, R.A., Ed.; Academic Press: New York, NY, USA, 2003; Volume 4, pp. 155–169.
19. Ashby, W.R. *An Introduction to Cybernetics*; Chapman & Hall: London, UK, 1956.
20. Michalowicz, J.V.; Nichols, J.M.; Bucholtz, F. Calculation of differential entropy for a mixed Gaussian distribution. *Entropy* **2008**, 10, 200–206.
21. Calmet, J.; Calmet, X. Differential Entropy on Statistical Spaces. **2005**, arXiv:cond-mat/0505397.
22. Yeung, R. *Information Theory and Network Coding*, 1st ed.; Springer: Berlin/Heideberg, Germany, 2008; pp. 229–256.
23. Bedau, M.A.; Humphreys, P. (Eds.) *Emergence: Contemporary Readings in Philosophy and Science*; MIT Press: Cambridge, MA, USA, 2008.
24. Anderson, P.W. More is Different. *Science* **1972**, 177, 393–396.
25. Shalizi, C.R. Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata. Ph.D. thesis, University of Wisconsin, Madison, WI, USA, 2001.
26. Singh, V. *Entropy Theory and its Application in Environmental and Water Engineering*; John Wiley Sons: Chichester, UK, 2013; pp. 1–136.
27. Gershenson, C. The Implications of Interactions for Science and Philosophy. *Found. Sci.* **2012**, 18, 781–790.
28. Sharma, K.; Sharma, S. Power Law and Tsallis Entropy: Network Traffic and Applications. In *Chaos, Nonlinearity, Complexity*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 178, pp. 162–178.
29. Dover, Y. A short account of a connection of Power-Laws to the information entropy. *Physica A* **2004**, 334, 591–599.
30. Bashkirov, A.; Vityazev, A. Information entropy and Power-Law distributions for chaotic systems. *Physica A* **2000**, 277, 136–145.
31. Ahsanullah, M.; Kibria, B.; Shakil, M. Normal and Student's t-Distributions and Their Applications; In *Atlantis Studies in Probability and Statistics*; Atlantis Press: Paris, France, 2014.
32. Box, G.; Jenkins, G.; Reinsel, G. *Time Series Analysis: Forecasting and Control*, 4th ed.; John Wiley Sons: Hoboken, NJ, USA, 2008.
33. Mitzenmacher, M. A Brief History of Generative Models for Power Law and Lognormal Distributions. **2009**, arXiv:arXiv:cond-mat/0402594v3.
34. Mitzenmacher, M. A brief history of generative models for Power-Law and lognormal distributions. *Internet Math.* **2001**, 1, 226–251.
35. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **2009**, 51, 661–703.
36. Frigg, R.; Werndl, C. Entropy: A Guide for the Perplexed. In *Probabilities in Physics*; Beisbart, C.; Hartmann, S., Eds.; Oxford University Press: Oxford, UK, 2011; pp. 115–142.
37. Virkar, Y.; Clauset, A. Power-law distributions in binned empirical data. *Ann. Appl. Stat.* **2014**, 8, 89–119.
38. Yapage, N. Some Information measures of Power-law Distributions Some Information measures of Power-law Distributions. In Proceedings of the 1st Ruhuna International Science and Technology Conference, Matara, Sri Lanka, 22–23 January 2014.
39. Newman, M. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, 46, 323–351.

40. Landsberg, P. Self-Organization, Entropy and Order. In *On Self-Organization*; Mishra, R.K., Maaß, D., Zwielerlein, E., Eds.; Springer: Berlin/Heidelberg, Germany, 1994; Volume 61, pp. 157–184.
41. Gershenson, C.; Lenaerts, T. Evolution of Complexity. *Artif. Life* **2008**, *14*, 241–243.
42. Cocho, G.; Flores, J.; Gershenson, C.; Pineda, C.; Sánchez, S. Rank Diversity of Languages: Generic Behavior in Computational Linguistics. *PLoS ONE* **2015**, *10*, doi:10.1371/journal.pone.0121898.
43. Gershenson, C. Requisite Variety, Autopoiesis, and Self-organization. *Kybernetes* **2015**, *44*, 866–873.
44. Newman, M.E.J. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256.
45. Newman, M.; Barabási, A.L.; Watts, D.J. (Eds.) *The Structure and Dynamics of Networks*; Princeton University Press: Princeton, NJ, USA, 2006.
46. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308.
47. Gershenson, C.; Prokopenko, M. Complex Networks. *Artif. Life* **2011**, *17*, 259–261.
48. Motter, A.E. Networkcontrolology. *Chaos* **2015**, *25*, 097621.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).