

Universality of Zipf's law

Bernat Corominas-Murtra¹ and Ricard V. Solé^{1,2,3}¹*ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona (PRBB) Dr., Aiguader 80, 08003 Barcelona, Spain*²*Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA*³*Institut de Biologia Evolutiva, CSIC-UPF, Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain*

(Received 15 January 2010; revised manuscript received 21 April 2010; published 1 July 2010)

Zipf's law is the most common statistical distribution displaying scaling behavior. Cities, populations or firms are just examples of this seemingly universal law. Although many different models have been proposed, no general theoretical explanation has been shown to exist for its universality. Here, we show that Zipf's law is, in fact, an inevitable outcome of a very general class of stochastic systems. Borrowing concepts from Algorithmic Information Theory, our derivation is based on the properties of the symbolic sequence obtained through successive observations over a system with an unbounded number of possible states. Specifically, we assume that the complexity of the description of the system provided by the sequence of observations is the one expected for a system evolving to a stable state between order and disorder. This result is obtained from a small set of mild, physically relevant assumptions. The general nature of our derivation and its model-free basis would explain the ubiquity of such a law in real systems.

DOI: [10.1103/PhysRevE.82.011102](https://doi.org/10.1103/PhysRevE.82.011102)

PACS number(s): 05.40.-a, 05.65.+b

I. INTRODUCTION

Scaling laws are common in both natural and artificial systems [1]. Their ubiquity and universality is one of the fundamental issues in statistical physics [2–4]. One of the most prominent examples of power-law behavior is the so called *Zipf's law* [5–7]. It was popularized by the linguist G. K. Zipf, who observed that it accounts for the frequency of words within written texts [5,8]. But this law is extremely common [9], and has been found in the distribution of populations in city sizes [5,10–14], firm sizes in industrial countries [15], market fluctuations [16], money income [17,18], Internet file sizes [19], or family names [20]. For instance, if we rank all the cities in a country from the largest (in population size) to the smallest, Zipf's law states that the probability $p(s_i)$ that a given individual lives in the i th most populated city ($i=1, \dots, n$) falls off as

$$p(s_i) = \frac{1}{Z} i^{-\gamma}, \quad (1)$$

with the exponent, $\gamma \approx 1$, and being Z the normalization constant, i.e.,

$$Z = \left(\sum_{i \leq n} i^{-\gamma} \right). \quad (2)$$

Although systems exhibiting Zipf's-like statistics are clearly different in their constituent units, the nature of their interactions and intrinsic structure, most of them share a few essential commonalities. One is that they are stochastic, far from equilibrium systems changing in time, under mechanisms that prevent them to become homogeneous. Within the context of economic change, for example, wider varieties of goods and attraction for people are fueled by large developed areas. Increasing returns drive further growth and feedback between economy and city sizes [21–23]. Moreover, the presence of a scaling law seems fairly robust through time: in spite of widespread political and social changes, the sta-

tistical behavior of words in written texts, cities or firms has remained the same over decades or even centuries [5,7,15,23,24]. Such robustness is remarkable, given that it indicates a large insensitivity to multiple sources of external perturbation. In spite of their disparate nature, all seem to rapidly achieve the Zipf's law regime and remain there.

To account for the emergence and robustness of Zipf's law, several mechanisms have been proposed, including autocatalytic processes [25–27], extinction dynamics [28,29], intermittency [30,31], coherent noise [32], coagulation-fragmentation processes [33,34], self-organized criticality [35], communicative conflicts [36,37], random *typewriting* [38,46], multiplicative dynamics [39,40], or stochastic processes in systems with interacting units with complex internal structure [41]. The diverse character of such mechanisms sharing a common scaling exponent strongly points toward the hypothesis that some fundamental property (beyond a given specific dynamical mechanism) is at work. Such a universal trend asks for a generic explanation, which should avoid the use of a particular set of rules.

We address the problem from a very general, mechanism-free viewpoint; by studying the statistical properties of the sequence of successive observations over the system. More precisely, our observations can be understood as a sequence of symbols of a given alphabet (depending on the nature of the system) following some probability distribution. The elements of this alphabet can be coded in some way, for example, bits. From this conceptual starting point, we borrow concepts from algorithmic information theory (AIT) and propose a characterization of a wide family of stochastic systems, to which those systems displaying Zipf's law would belong. Such a characterization imposes special features on the behavior of the entropy, whose study leads us to conclude that, under generic mathematical assumptions, Zipf's law is the only solution.

The paper is organized as follows: In Sec. II, we briefly introduce the concept of stochastic object as defined within the context of AIT and how it helps to understand our prob-

lem. In Sec. III, we find the asymptotic solutions of the equations derived from the characterization provided in Sec. II. Section IV discusses the relevance of the obtained results.

II. ALGORITHMIC COMPLEXITY OF STOCHASTIC SYSTEMS

The cornerstone of our argument is an abstract characterization of the sequence of observations made on a given system in terms of AIT [42–45,47,48]—see also [49]. The key quantity of such theory is the so-called *Kolmogorov complexity*, which is a conceptual precursor of statistical entropy, and an indicator of the complexity (and predictability) of a dynamical system [50–52]. In a nutshell, let \mathbf{x} be a symbolic string generated by the successive observations of the system \mathcal{S} . Its Kolmogorov complexity, $K(\mathbf{x})$ is defined as the length $l(\pi^*)$ —in bits—of the shortest program π^* executed in a universal computer in order to reproduce \mathbf{x} . This measure has been often used in statistical physics [53–55] particularly in the context of symbolic dynamics [51]. In this context, K is known to be maximal for completely disordered systems, whereas it takes intermediate values when some asymmetry on the probabilities of appearance of symbols emerges.

Within the framework of statistical physics, a sequence of observations performed over a given system can be interpreted as a sequence of independent, identically distributed random variables, where the specific outcomes of the observations are obtained according to a given probability distribution. In mathematical terms, such a sequence of observations defines a *stochastic object*. By definition, the Kolmogorov Complexity of a stochastic object, described by a binary string $\mathbf{x}=x_1, \dots, x_m$ of length m , satisfies [56]:

$$\lim_{m \rightarrow \infty} \frac{K(\mathbf{x})}{m} = \mu \in (0, 1]. \tag{3}$$

In other words, the binary representation of a stochastic object is *linearly* compressible. The case where $\mu=1$ refers to a completely random object, and the string is called *incompressible*.

We can generalize the concept for non binary strings, whose elements belong to a given set $\Sigma=\{s_1, \dots, s_n\}$, being $|\Sigma|=n$. This is the case of a dice, for example, whose set of outcomes is $\Sigma_{dice}=\{1, 2, 3, 4, 5, 6\}$. If the behavior of the system is governed by the random variable $X(n)$, accordingly, the successive observations of our stochastic system define a sequence of independent, identically p_n -distributed random variables $X_1(n), \dots, X_m(n)$ taking values over the set Σ . The so-called *noiseless Coding* theorem [47,57,58], establishes that the minimum length, (in bits) of the string needed to code the event $s_i, l^*(s_i)$, satisfies

$$l^*(s_i) = -\log(p_n(s_i)) + \mathcal{O}(1). \tag{4}$$

Throughout the paper, $\log \equiv \log_2$, unless the contrary is indicated. The average minimum length will correspond to the minimum length of the code, which is, by definition, the Kolmogorov complexity. Thus, we obtain [47,59]:

$$\lim_{m \rightarrow \infty} \frac{K(X_1(n), \dots, X_m(n))}{m} = \sum_{i \leq n} p_n(s_i) l^*(s_i) = H(X(n)) + \mathcal{O}(1), \tag{5}$$

being $H(X(n))$ the Shannon or statistical entropy [57,58,47], namely:

$$H(X(n)) = - \sum_{i \leq n} p_n(s_i) \log p_n(s_i).$$

The complete random case is obtained when, $\forall s_i \in \Sigma p_n(s_i) = 1/n$ leading to $l^*(s_i) = \log n + \mathcal{O}(1)$. This indicates that we need $\approx \log n$ bits to code any element from Σ . Therefore, the length in bits of the sequence of m successive observations will be approximately $m \log n$. the average minimum length of the code will be lower than $\log n$. Using our previous result Eq. (3) for the binary case, it is not difficult to see that:

$$\lim_{m \rightarrow \infty} \frac{K(X_1(n), \dots, X_m(n))}{m \log n} = \mu; \quad \mu \in (0, 1]. \tag{6}$$

By defining $h(n)$ as the normalized entropy as:

$$h(n) \equiv \frac{H(X(n))}{\log n}, \tag{7}$$

and from Eq. (5), we observe that Eq. (6) can be rewritten as $h(n) \approx \mu; \mu \in (0, 1]$.

So far we have been concerned with the algorithmic characterization of stochastic systems for which the size of the configuration space is static. However, we must differentiate the properties of the systems we want to characterize from a standard stochastic object such as the ones obtained by tossing a dice or a coin. They both generate a bounded number of possible outcomes—namely, 6 and 2—with an associated probability, whereas those systems exhibiting power-laws lack an *a priori* constraint on the potential number of available outcomes. These systems are *open* concerning the size—or dimensionality—of the configuration space. Let $X(n)$ be a random variable taking values on Σ , where $|\Sigma|=n$ and with associated probability distribution p_n , where (without any loss of generality) an ordering

$$p_n(s_1) \geq p_n(s_2) \geq \dots \geq p_n(s_n) \tag{8}$$

is assumed. At a given time, the system satisfies Eq. (6), since it is a stochastic object with a given number of available states. However, we assume that the system changes (generally growing) maintaining its basic statistical properties stable [5,7,15,24]. Using Eq. (5), condition Eq. (6) is replaced by:

$$\lim_{n \rightarrow \infty} h(n) = \mu. \tag{9}$$

We can replace Eq. (9) alternatively by the following statement: For any $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that, for any $n' > n$:

$$|h(n') - \mu| < \epsilon. \tag{10}$$

The main objective of the paper is to find the expected distribution $p_n(s_i)$ consistent with Eq. (10). The case $\mu=0$ would correspond to systems where, although growing in

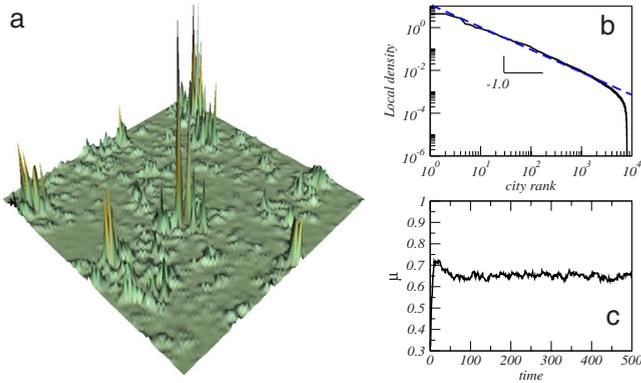


FIG. 1. (Color online) An example of the behavior of the normalized entropy for a multiplicative stochastic process exhibiting Zipf's law. Here, we use the model described in [23] using a 80×80 lattice where each node is described by a density of population $\rho(i, j)$. The rules of the model are very simple: i) At every time step, each node loses a fraction α of its contents, which is distributed among its four nearest neighbors. ii) At time $t+1$ the local population is multiplied, with probability p , by a factor p^{-1} . Furthermore, with probability $1-p$, the population of a node is set to zero. Additionally, at each step a random number η is added to every node. In this way, we avoid falling into an absorbing state $\rho=0$. Here we use $0 < \eta < 0.01$, $\alpha=1/4$, and $p=3/4$. This is an extremely simplified (and yet successful) model of urban population dynamics. A snapshot (for $t=500$) is shown in (a) where we can appreciate the wide range of local densities, following Zipf's law (b). If we plot the evolution of the normalized entropy μ over time (averaged over 10^2 replicas) we observe a convergence toward a stationary value $\mu \approx 0.65$.

size, its complexity (and thus, its statistical entropy) is bounded or grows sublinearly with $\log n$, a case studied in [51]. Here, we are interested in the intermediate case, where $\mu \in (0, 1)$. This characterization would depict systems with some balance among ordering and disordering forces, and thereby displaying a dissipation of statistical entropy proportional to the maximum entropy achievable for the system in equilibrium. Therefore, we will refer to the problem of finding solutions for Eq. (9) as the *entropy restriction problem*. A computational test for this result can be illustrated by the model results shown in Fig. 1. The picture shows a spatial snapshot of the local population densities of a model of urban growth displaying Zipf's law [23]. The normalized entropy evolves toward a stationary value $\mu \approx 0.65$ consistently with our discussion. This is true in spite that this model exhibits wide fluctuations due to its intermittent stochastic dynamics.

III. EMERGENCE OF ZIPF'S LAW IN STOCHASTIC SYSTEMS

As pointed out in [36], the main difficulty we face in this kind of equations is that we are not dealing with an extremal problem, since our value of entropy is previously fixed and it is neither minimum nor maximum, in Jaynes' sense [60]. Thus, classical variational methods, which have been widely used with great success in statistical mechanics [60–63], do

not apply to our problem—although recently it has been shown that variational approaches using Fisher information and physically relevant constraints lead to Power laws whose exponent can be close to 1 [64]. We also must take into account that the particular properties of Zipf's law create an additional difficulty if the studied systems display, *a priori*, an unbounded number of possible states. Specifically, we refer to the nonexistence of finite moments and normalization constant in the thermodynamical limit. However, as we shall see, these apparently undesirable properties will be the key to our derivation.

A. Properties of the entropies of a power law

Let us briefly summarize the properties of the entropies of power-law distributed systems, which will be used to derive the main results of this work (For details, see Appendix). Such properties are intimately linked with the behavior of the *Riemann Zeta function*, $\zeta(\gamma)$ [65]:

$$\zeta(\gamma) = \sum_{k=1}^{\infty} \frac{1}{k^\gamma}. \quad (11)$$

In the real line, this function is defined in the interval $\gamma \in (1, \infty)$, displaying a singularity for $\gamma \rightarrow 1^+$.

Now, let us suppose that the system contains n states and the probability to find the i th most likely states decay as a power law, i.e., $p_n(s_i) \propto i^{-\gamma}$. For the sake of simplicity, we will refer to its associated entropy as $H(n, \gamma)$ and to its normalized counterpart as $h(n, \gamma)$, i.e.,:

$$h(n, \gamma) = \frac{1}{\log n} \left(\frac{\gamma}{Z} \sum_{i=1}^n \frac{\log i}{i^\gamma} + \log Z \right). \quad (12)$$

The most basic properties concern the global behavior of $H(n, \gamma)$. It is straightforward to check that $H(n, \gamma)$ is (i) a monotonous increasing function on n , and (ii) a monotonous decreasing function on γ . Moreover, the normalized entropy of Zipf's law of a system with n states converges to $1/2$ [66], i.e.,

$$\lim_{n \rightarrow \infty} h(n, 1) = \frac{1}{2}. \quad (13)$$

We also note that the entropy of a power law with exponent higher than one is bounded i.e., if $\gamma > 1$ is the exponent of our power law, there exists a finite constant $\phi(\gamma)$ such that:

$$\lim_{n \rightarrow \infty} H(n, \gamma) < \phi(\gamma). \quad (14)$$

A key consequence of this result is that, if our (unknown) probability distribution is dominated [67] from some k by some power law with exponent $\gamma > 1 + \delta$ (for any $\delta > 0$), our entropy will be bounded.

Furthermore, it can be shown that the normalized entropy of a power-law distribution in a system with n different states, with exponent $\gamma < 1$, converges to 1, i.e.,

$$\lim_{n \rightarrow \infty} h(n, \gamma) = 1. \quad (15)$$

Consistently, we can conclude that, if an (unknown) probability distribution is not dominated from any m by a power

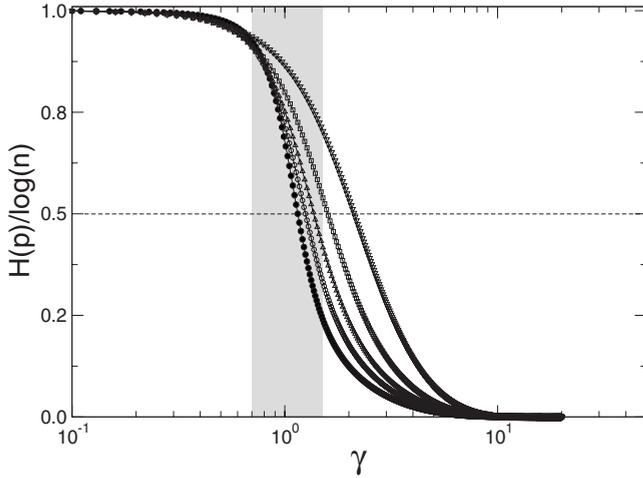


FIG. 2. Normalized entropies of five power-law distributed systems of different size as functions of the exponent. The curves display 5 different sizes. $n=500\,000$ black circles, $n=10\,000$ white circles, $n=10\,000$ up triangles, $n=1000$ squares and $n=100$ down triangles, respectively. The most interesting feature of the numerical computations is the sharp decay of the normalized entropy when the values of the exponent are close to 1, which implies that a wide range of normalized entropies are obtained by tuning the exponent of the power-law distribution around unity. Furthermore, we observe that the decay is sharper as the size of the system grows, concentrating an increasing range of relative entropies near the exponent 1 (gray area).

law with exponent lower than $1 - \delta$ (for any $\delta > 0$), the normalized entropy of our system will converge to 1.

Using these properties, in the following sections we proceed to derive Zipf's using two complementary approaches, namely, (1) proposing a power law as the asymptotic solution of Eq. (9)—Sec. III B—and (2) Assuming that the entropy behaves in a scale—invariant way—Sec. III C.

B. Power-law ansatz: Convergence of exponents to $\gamma=1$

In this section we make use of the power-law ansatz as a solution of our problem, i.e., we assume that the solution is a power law with an arbitrary exponent, i.e., $p_n(s_i) \propto i^{-\gamma}$. The objective of this section is to demonstrate that, being $h(n, \gamma)$ as defined in Eq. (12), then the following limit holds:

$$\lim_{n \rightarrow \infty} h(n, \gamma) = \Theta(\gamma), \quad (16)$$

being $\Theta(\gamma)$ the step function, i.e., $\Theta(\gamma)=1$ if $\gamma < 1$ and $\Theta(\gamma)=0$ if $\gamma > 1$. It implies that, for large values of n , the whole range of normalized entropies between 0 and 1 is obtained from exponents γ arbitrarily close to $\gamma=1$ —see Fig. 2.

Let us rewrite the convergence assumptions provided in Eqs. (9) and (10) assuming that our probability distribution is a power law: For any $\epsilon > 0$ we can find an n such that, for any $n' > n$ we have an exponent, $\gamma(n')$ such that,

$$|h(n', \gamma(n')) - \mu| < \epsilon, \quad (17)$$

i.e., the sequence of normalized entropies \mathcal{H} , associated to system's growth, namely,

$$\mathcal{H} = h(1, \gamma(1)), h(2, \gamma(2)), \dots, h(k, \gamma(k)), \dots, \quad (18)$$

converges to μ . Below we split the problem in two different scenarios.

1. First case: $\mu < \frac{1}{2}$

We begin by exploring the following scenario:

$$\lim_{n \rightarrow \infty} h(n, \gamma(n)) = \mu \in \left(0, \frac{1}{2}\right). \quad (19)$$

From Eq. (13) we can ensure that, for large values of n , $\gamma(n) > 1$. Since we assumed that the sequence \mathcal{H} converges to μ , we can state that, for a given $\epsilon > 0$, there is an arbitrary n_1 such that:

$$\mu - \epsilon < h(n_1, \gamma(n_1)) < \mu + \epsilon. \quad (20)$$

We know, from the properties of the entropies of power-law distributed systems, that $H(n_1, \gamma(n_1)) < \phi(\gamma(n_1))$, where $\phi(\gamma(n_1))$ is some positive, finite constant [see Eq. (14) and Appendix]. Then, since $\log x$ is an unbounded, increasing function of x , we can find $n_2 > n_1$ such that

$$\phi(\gamma(n_1)) < (\mu + \epsilon) \log n_2. \quad (21)$$

Thus, since $h(n, \gamma)$ is a decreasing function on γ , we need to find $\gamma(n_2) < \gamma(n_1)$ such that

$$\mu - \epsilon' < h(n_2, \gamma(n_2)) < \mu + \epsilon', \quad (22)$$

with $\epsilon' \leq \epsilon$, in order to satisfy the entropy restriction. Furthermore, since $H(n, 1) = \frac{1}{2} \log n + \mathcal{O}(\log(\log n))$, we conclude that $1 < \gamma(n_2) < \gamma(n_1)$. Let us expand this process recursively, thus generating an infinite decreasing sequence of exponents,

$$\{\gamma(n_k)\}_{k=1}^{\infty} = \gamma(n_1), \dots, \gamma(n_i), \dots, \quad (23)$$

such that, for any $\gamma(n_i) \in \{\gamma(n_k)\}_{k=1}^{\infty}$, $\gamma(n_i) > 1$. We notice that, for any $\alpha > 0$, we can find a n_k such that, if $n_j > n_k$,

$$|\gamma(n_j) - 1| < \alpha, \quad (24)$$

since, for every $\gamma(n_k)$, we always find a $n_j > n_k$ such that

$$\phi(\gamma(n_k)) < (\mu + \epsilon) \log n_j. \quad (25)$$

2. Second case: $\mu > \frac{1}{2}$

Let us now consider the following entropy restriction problem:

$$\lim_{n \rightarrow \infty} h(n, \gamma(n)) = \mu \in \left(\frac{1}{2}, 1\right). \quad (26)$$

From Eq. (13), we can ensure that, for any n , $\gamma(n) < 1$. Furthermore, from Eq. (15), we again find a problem close to the one solved above, since for n_1 large enough and $\gamma < 1$, we have:

$$H(n_1 + 1, \gamma) - H(n_1, \gamma) > \mu(\log(n_1 + 1) - \log n_1). \quad (27)$$

Now, since we assumed that the sequence \mathcal{H} converges, we can state that given an arbitrary step n_1 ,

$$\mu - \epsilon < h(n_1, \gamma(n_1)) < \mu + \epsilon. \quad (28)$$

Since $H(n, \gamma)$ is a decreasing function on γ , we need to find $\gamma(n_2) > \gamma(n_1)$ such that:

$$\mu - \epsilon' < h(n_2, \gamma(n_2)) < \mu + \epsilon', \quad (29)$$

with $\epsilon' \leq \epsilon$, to satisfy the entropy restriction. However, from Eq. (13), we know that $1 > \gamma(n_2) > \gamma(n_1)$. Proceeding as above, we expand this process, thus generating an infinite increasing sequence of exponents $\{\gamma(n_k)\}_{k=1}^{\infty}$. By virtue of Eq. (13) and Eq. (15), and taking into account the decreasing behavior of h as a function of the exponent, we observe that, for any $\alpha > 0$, we can find a n_k such that, if $n_j > n_k$,

$$|\gamma(n_j) - 1| < \alpha. \quad (30)$$

In summary, under the power law ansatz, the only solution for Eq. (9), in the limit of large systems, is $\gamma=1$, i.e., Zipf's law.

C. Scale invariance Condition

The above power-law ansatz is purely mathematical, and can be replaced by a more physically realistic assumption. This leads us to the second strategy to solve our problem, which is based on the assumption that the mechanisms responsible for the growth and stabilization of the system do not depend on the size of the configuration space, and, thus, a partial observation of the system will satisfy also condition Eq. (9). We will refer to this assumption as the *scale invariance condition*, and it is formulated as follows. Let $\Sigma^{(k)} \subseteq \Sigma$ be the set of the first k elements of Σ , observing a labeling consistent with the ordering of probabilities provided in Eq. (8)—roughly speaking, the k most probable elements of Σ . The random variable which accounts for the observations of such k elements is notated $X(k \leq n)$. We observe that, if $X(n)$ follows the probability distribution p_n , the random variable $X(k \leq n)$ obeys the following probability distribution, to be notated p_n^k :

$$p_n^k(i) \equiv \text{P}(s_i | i \leq k) = \left(\sum_{j \leq k} p_n(s_j) \right)^{-1} p_n(s_i). \quad (31)$$

Thus, if $H(X(k \leq n))$ is the entropy of $X(k \leq n)$, its normalized counterpart is defined as $h(k \leq n)$:

$$h(k \leq n) \equiv \frac{H(X(k \leq n))}{\log k}. \quad (32)$$

We remark that these derivations are valid at the limit of large systems, thereby considering that, at every step, n is arbitrarily greater than k . Furthermore, let us define ϵ' as:

$$\epsilon' \equiv |h(k \leq n) - \mu| + \delta, \quad (33)$$

being δ arbitrarily small. Then, the scale invariance assumption for the entropy states that, for any $n \geq k' \geq k$,

$$|h(k' \leq n) - \mu| < \epsilon'. \quad (34)$$

In summary, condition Eq. (34), is grounded on the assumption that the entropy restriction works at all levels of observation. Thus, the partial probability distributions of states we

obtain must reflect the effect of the entropy restriction, introducing a scale invariance of the normalized entropy of the partial samples of the system.

As we saw in the above sections, the decay of this tail is strongly constrained by the entropy restriction, since only special cases avoid the normalized entropy to fall to 0 or 1. To study in detail how it constrains the tail of the distribution we will work with the coefficients $f_n(k, k+1)$, defined as:

$$f_n(i, i+1) = \frac{p_n(s_i)}{p_n(s_{i+1})},$$

instead of the raw probability distribution, to avoid multiplying factors due to normalization. Now we observe that, for a given, very large n , our probability distributions p_n^k must be able, as k increases, to unboundedly increase the entropy of the whole system to reach the global value $H(X(n))$, which lies in the interval $((\mu - \epsilon) \log n, (\mu + \epsilon) \log n)$. Furthermore, scale invariance condition depicted in Eq. (34) forces that, as k increases, contributions to the entropy never go neither to 0 nor to $\log(k+1/\log k)$, but lie within this interval. In other words, the sum defined by the entropies must diverge as k increases over a system where n is arbitrary large, whereas the sequence of its normalized versions must converge to μ . The above derivations concerning the convergence properties of the entropy—see also Appendix—clearly state that those properties hold if p_n satisfies, on one hand, for large i 's,

$$f_n(i, i+1) < \left(\frac{i+1}{i} \right)^{(1-\delta)}, \quad (35)$$

to avoid that $h(n) \rightarrow 1$. On the other hand, if we want to avoid that $h(n) \rightarrow 0$, the following inequality must hold:

$$f_n(i, i+1) > \left(\frac{i+1}{i} \right)^{(1+\delta)}. \quad (36)$$

Therefore, the solution of our problem lies in the range defined by:

$$\left(\frac{i+1}{i} \right)^{(1-\delta)} > f_n(i, i+1) > \left(\frac{i+1}{i} \right)^{(1+\delta)}. \quad (37)$$

From the study of the entropies of a power law performed in the previous section, we know that δ can be arbitrarily small if the size of the system is large enough. Thus,

$$f_n(i, i+1) = \frac{p_n(s_i)}{p_n(s_{i+1})} \approx \frac{i+1}{i} \quad (38)$$

which leads us to Zipf's law as the unique asymptotic solution:

$$p_n(s_i) \propto i^{-1}. \quad (39)$$

IV. DISCUSSION

Complex, far from equilibrium systems involve a tension between amplifying mechanisms and negative feedbacks able to buffer the impact of fluctuations. In this paper we have considered the consequences of such tension in terms of one of its most well known outcomes: the presence of an inverse scaling law connecting the size of observed events

and its rank. The commonality of Zipf’s law in both natural and man-made systems has been a puzzle that attracted for years the attention of scientists, sociologists and economists alike. The fact that such a plethora of apparently unrelated systems display the same statistical pattern points toward some fundamental, unifying principle.

In this paper, we treat complex systems as stochastic systems describable in terms of algorithmic complexity and thus statistical entropy. A general result from the algorithmic complexity theory is that Eq. (3) holds for stochastic systems. Taking this general result as the starting point, we define a characterization of a wide class of complex systems, which grasps the open nature of many complex systems, summarized in Eq. (9). The main achievement of this equation is that it encodes the concepts of growing and, even most important, the stabilization of complexity properties in an intermediate point between order and disorder, a feature observed in many systems displaying Zipf’s-like statistics. From this equation we derived Zipf’s law as the natural outcome of systems belonging to this class of stochastic systems.

Our development avoids the classical procedures based on maximization (minimization) of some functional in order to find the most probable configuration of states, since in far from equilibrium the ensemble formalism, together with Jaynes’ maximum entropy principle [60] can fail due to the open, nonreversible behavior of the systems considered here. Thus, we do not introduce moment constraints, as it is usual in equilibrium statistical mechanics [63], but instead a constraint on the value achieved by the normalized entropy, no matter the scale we observe the system. Both a scaling ansatz and a more general scale invariance assumption lead to Zipf’s law as the unique solution for this problem. We observe that the finite size effects define an interval of exponents around 1, namely, $(1 - \delta, 1 + \delta)$, which could partly explain the variation observed in finite, natural systems. However, it is true that a system satisfying Eq. (9) does not necessarily exhibit Zipf’s law. Further work should explore in depth the physically relevant conditions leading the evolution of Zipf’s like systems to remove the mathematical assumptions made in this paper, thereby obtaining a complete description of them from a completely general, theoretical viewpoint.

ACKNOWLEDGMENTS

We thank D. Jou, S. Manrubia, J. Fortuny, and our colleagues at the Complex Systems Lab for their useful comments. We also acknowledge the helpful comments provided by the anonymous referees. This work has been founded by the McDonnell Foundation (B.C.M.) and by the Santa Fe Institute (R.S.).

APPENDIX: ENTROPIC PROPERTIES OF POWER-LAW DISTRIBUTED SYSTEMS

Consider a system whose behavior is described by the random variable $X(n)$ taking values on the set $\Sigma = \{s_1, \dots, s_n\}$, $|\Sigma| = n$, according to the probability distribu-

tion $p_n(s_i)$. The labeling ‘ i ’ of the state is chosen in such a way that $p_n(s_1) \geq p_n(s_2) \geq \dots \geq p_n(s_i) \geq \dots \geq p_n(s_n)$.

The Shannon entropy of our system of n states, to be noted $H(X(n))$, is defined as [58]:

$$H(X(n)) = - \sum_{k \leq n} p_n(s_k) \log p_n(s_k). \tag{A1}$$

The normalized entropy of the system, to be written, $h(n)$, is defined as:

$$h(n) \equiv \frac{H(X(n))}{\log n}. \tag{A2}$$

We will work with power-law distributions, by which $p_n(s_i) = \frac{1}{Z} i^{-\gamma}$ where n is the number of available states, and Z the normalization constant, which depends on the size of the system, n . Let us rewrite the function $H(X(n))$ as a function of the exponent and the size of Σ , $H(n, \gamma)$. Consistently,

$$h(n, \gamma) \equiv \frac{H(n, \gamma)}{\log n}. \tag{A3}$$

This appendix is devoted to derive five properties of the entropy of power-law distributed systems.

(1) $H(n, \gamma)$ is a continuous, monotonous decreasing function with respect to γ in the range $(0, \infty)$.

Indeed, the dominant term of its derivative is:

$$\frac{\partial H(n, \gamma)}{\partial \gamma} \sim - \sum_{i \leq n} \frac{(\log i)^2}{i^\gamma} < 0. \tag{A4}$$

(2) The entropy of a power-law is a monotonous, increasing function on the size of the system [68].

We want to show that $H(n, \gamma)$ is a monotonous increasing function on n . In order to prove it, we must compute the difference $H(n+1, \gamma) - H(n, \gamma)$. For simplicity, let us define:

$$S_n \equiv \sum_{k \leq n} \frac{1}{k^\gamma}. \tag{A5}$$

Using the trivial inequality:

$$\log \left[S_n + \frac{1}{(n+1)^\gamma} \right] > \log(S_n), \tag{A6}$$

we can state that:

$$\begin{aligned} & H(n+1, \gamma) - H(n, \gamma) \\ &= \frac{\gamma}{S_n + \frac{1}{(n+1)^\gamma}} \sum_{k \leq n+1} \frac{\log k}{k^\gamma} + \log \left[S_n + \frac{1}{(n+1)^\gamma} \right] \\ &\quad - \frac{\gamma}{S_n} \sum_{k \leq n} \frac{\log k}{k^\gamma} + \log(S_n) \\ &> \gamma \sum_{k \leq n} \frac{\log k}{k^\gamma} \left[\frac{1}{S_n + \frac{1}{(n+1)^\gamma}} - \frac{1}{S_n} \right] + \gamma \frac{\log(n+1)}{S_n + \frac{1}{(n+1)^\gamma}} \\ &= \frac{\gamma}{S_n^2(n+1)^\gamma + S_n} \left[S_n(n+1)^\gamma \log(n+1) - \sum_{k \leq n} \frac{\log k}{k^\gamma} \right] \\ &> 0. \end{aligned}$$

Finally, it is easy to check that the following properties also hold:

$$\lim_{\gamma \rightarrow \infty} H(n, \gamma) = 0, \tag{A7}$$

$$\lim_{\gamma \rightarrow 0} H(n, \gamma) = \log n. \tag{A8}$$

(3) The normalized entropy of Zipf's law of a system with n states ($p_n(s_i) \propto i^{-1}$) converges to $1/2$:

We want to show that the sequence

$$\mathcal{H} = \{h(k, 1)\}_{k=1}^{\infty} = h(1, 1), h(2, 1), \dots, h(k, 1), \dots \tag{A9}$$

converges to $\frac{1}{2}$. Let us suppose that \mathcal{H} is a sequence satisfying the above requirements. Then, the entropy for a given n can be approached by [66]:

$$H(n, 1) = \frac{1}{2} \log n + \mathcal{O}(\log(\log n)). \tag{A10}$$

Thus, if $h(n, 1) = H(n, 1) / \log n$, let us define $\epsilon(n)$ like:

$$\epsilon(n) \equiv \left| h(n, 1) - \frac{1}{2} \right| = \left| \frac{\mathcal{O}(\log(\log n))}{\log n} \right|. \tag{A11}$$

Clearly, $\epsilon(n)$ is strictly decreasing on n , and, furthermore,

$$\lim_{n \rightarrow \infty} \epsilon(n) = 0. \tag{A12}$$

(4) The Entropy of a power law with exponent higher than 1 is bounded.

Here we demonstrate that the entropy of a power law with exponent higher than 1 is bounded [69]. Specifically, we assume there exists a pair of positive constants Z, δ , such that:

$$p_n(i) = \frac{1}{Z} i^{-(1+\delta)}. \tag{A13}$$

Then, the sequence of $\mathcal{H} = \{h(k, 1 + \delta)\}_{k=1}^{\infty}$ converges to 0. Indeed, let us first note that:

$$\lim_{n \rightarrow \infty} p_n(s_i) = \frac{1}{\zeta(1 + \delta)} i^{-(1+\delta)}, \tag{A14}$$

where

$$\zeta(1 + \delta) \equiv \sum_k \frac{1}{k^{1+\delta}} \tag{A15}$$

is the Riemann zeta function [65]. The function is defined by an infinite sum which converges, in the real line, if $\delta > 0$, i.e.,:

$$\sum_k \frac{1}{k^{1+\delta}} < \infty. \tag{A16}$$

otherwise, the sum diverges. Furthermore, it is also true that the above condition also holds for the following series:

$$\sum_k \frac{\log k}{k^{1+\delta}}. \tag{A17}$$

Indeed, note that, given an arbitrary $\delta > 0$ there exists a finite number i^* such that:

$$i^* \equiv \min \left\{ i: \left(\delta - \frac{\log(\log i)}{\log i} \right) > 0 \right\} \tag{A18}$$

and, if we define the following exponent, $\beta(i^*)$:

$$\beta(i^*) \equiv 1 + \delta - \frac{\log(\log i^*)}{\log i^*}, \tag{A19}$$

there exists a finite constant, $\Psi(\delta)$, defined as:

$$\Psi(\delta) \equiv \sum_{i < i^*} \left(\frac{\log i}{i^{1+\delta}} - \frac{1}{i^{\beta(i^*)}} \right) + \zeta(\beta(i^*)), \tag{A20}$$

such that:

$$\sum_k \frac{\log k}{k^{1+\delta}} < \Psi(\delta). \tag{A21}$$

With the above properties, it is clear that, if there exists a constant $\phi(1 + \delta) < \infty$ such that:

$$\lim_{n \rightarrow \infty} H(n, 1 + \delta) < \phi(1 + \delta), \tag{A22}$$

then, the entropy of a power law with exponent higher than 1 is bounded. As we shall see, it is straightforward by checking directly the behavior of $H(n, 1 + \delta)$:

$$\lim_{n \rightarrow \infty} H(n, 1 + \delta) = \frac{1 + \delta}{\zeta(1 + \delta)} \sum_{i=1}^{\infty} \frac{\log i}{i^{1+\delta}} + \log(\zeta(1 + \delta)).$$

Since $H(n, \gamma)$ is an increasing function on n , and

$$\frac{1 + \delta}{\zeta(1 + \delta)} \sum_{i=1}^{\infty} \frac{\log i}{i^{1+\delta}} + \log(\zeta(1 + \delta)) < \infty, \tag{A23}$$

we can define a constant $\phi(1 + \delta)$,

$$\phi(1 + \delta) \equiv \lim_{n \rightarrow \infty} H(n, 1 + \delta) + \epsilon \tag{A24}$$

(where ϵ is any positive, finite constant). Clearly,

$$H(n, 1 + \delta) < \phi(1 + \delta). \tag{A25}$$

Thus,

$$\lim_{n \rightarrow \infty} h(n, 1 + \delta) = \lim_{n \rightarrow \infty} \frac{H(n, 1 + \delta)}{\log n} \leq \lim_{n \rightarrow \infty} \frac{\phi(1 + \delta)}{\log n} = 0.$$

Consequence, if an unknown probability distribution is dominated from some k by some power-law with exponent higher than $1 + \delta$, our entropy will be bounded.

Consequence, if an unknown probability distribution is dominated from some k by some power-law with exponent higher than $1 + \delta$, our normalized entropy will tend to 0.

(5) The normalized entropy of a power-law distribution in a system with n different states, p_n with exponent lower than 1 converges to 1.

Let us suppose that we have the following probability distribution, with $0 < \delta < 1$:

$$p_n(s_i) = \frac{1}{Z} i^{-(1-\delta)}. \quad (\text{A26})$$

Note that [66]:

$$\sum_{k \leq n} \frac{1}{k^{1-\delta}} = \int_1^n \frac{1}{x^{1-\delta}} + \mathcal{O}(1) \approx \frac{n^\delta}{\delta}. \quad (\text{A27})$$

Applying directly the definition of entropy,

$$H(n, 1-\delta) = \frac{\delta(1-\delta)}{n^\delta} \sum_{k \leq n} \frac{\log k}{k^{1-\delta}} + \delta \log n - \log \delta. \quad (\text{A28})$$

If we compute the limit of $h(n, 1-\delta)$:

$$\begin{aligned} \lim_{n \rightarrow \infty} h(n, 1-\delta) &= \lim_{n \rightarrow \infty} \left[\frac{\delta(1-\delta)}{\log n \cdot n^\delta} \sum_{k \leq n} \frac{\log k}{k^{1-\delta}} + \delta \right] \\ &= \lim_{n \rightarrow \infty} \frac{1-\delta}{\log n} \left(\log n - \frac{1}{\delta} \right) + \delta = 1 - \delta + \delta = 1. \end{aligned}$$

Consequence, if our (unknown) probability distribution is not dominated from some k by a power law with exponent higher than $1-\delta$, our normalized entropy will converge to 1.

-
- [1] H. E. Stanley, L. A. N. Amaral, P. Gopikrishnan, P. Ch. Ivanov, T. H. Keitt, and V. Plerou, *Physica A* **281**, 60 (2000).
- [2] H. E. Stanley, *Rev. Mod. Phys.* **71**, S358 (1999).
- [3] M. E. J. Newman, *Contemp. Phys.* **46**, 323 (2005).
- [4] R. V. Solé and B. Goodwin, *Signs of Life: How Complexity Pervades Biology* (Basic Books, New York, 2001).
- [5] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA, 1949).
- [6] F. Auerbach, *Petermanns Geographische Mitteilungen* **59**, 74 (1913).
- [7] X. Gabaix, *Q. J. Econ.* **114**, 739 (1999).
- [8] R. Ferrer i Cancho and R. V. Solé, *Adv. Complex Syst.* **5**, 1 (2002).
- [9] For a very complete collection of the works concerning the topic we refer the reader to: <http://www.nslj-genetics.org/wli/zipf/>
- [10] H. A. Simon, *Biometrika* **42**, 425 (1955).
- [11] H. A. Makse, S. Havlin, and H. E. Stanley, *Nature (London)* **377**, 608 (1995).
- [12] P. Krugman, *J. Jpn. Int. Econ.* **10**, 399 (1996).
- [13] A. Blank and S. Solomon, *Physica A* **287**, 279 (2000).
- [14] E. H. Decker, A. J. Kerkhoff, and M. E. Moses, *PLoS ONE* **2**, e934 (2007).
- [15] R. L. Axtell, *Science* **293**, 1818 (2001).
- [16] X. Gabaix, P. Gopikrishnan, V. Plerou, and E. H. Stanley, *Nature (London)* **423**, 267 (2003).
- [17] V. Pareto, *Cours d'Economie Politique* (Droz, Geneva, 1896).
- [18] K. Okuyama, M. Takayasu, and H. Takayasu, *Physica A* **269**, 125 (1999).
- [19] W. J. Reed and B. D. Hughes, *Phys. Rev. E* **66**, 067103 (2002).
- [20] D. H. Zanette and S. C. Manrubia, *Physica A* **295**, 1 (2001).
- [21] P. Krugman, *The Self-Organizing Economy* (Blackwell, Oxford, 1996).
- [22] B. W. Arthur, *Increasing Returns and Path Dependence in Economy* (Michigan University Press, Michigan, 1996).
- [23] S. C. Manrubia, D. H. Zanette, and R. V. Solé, *Fractals* **7**, 1 (1999).
- [24] G. Rozman, in *East Asian Urbanization in the Nineteenth Century: Comparisons with Europe*, edited by V. der Woude et al. (Clarendon, Oxford, 1990), pp. 61–63.
- [25] S. Solomon and M. Levy, *Int. J. Mod. Phys. C* **7**, 745 (1996).
- [26] O. Malcai, O. Biham, and S. Solomon, *Phys. Rev. E* **60**, 1299 (1999).
- [27] Z. F. Huang and S. Solomon, *Eur. Phys. J. B* **20**, 601 (2001).
- [28] R. V. Solé and S. C. Manrubia, *Phys. Rev. E* **54**, R42 (1996).
- [29] M. E. J. Newman and R. G. Palmer, *Modelling Extinction* (Oxford University Press, New York, 2003).
- [30] D. H. Zanette and S. C. Manrubia, *Phys. Rev. Lett.* **79**, 523 (1997).
- [31] S. C. Manrubia and D. H. Zanette, *Phys. Rev. E* **58**, 295 (1998).
- [32] M. E. J. Newman and K. Sneppen, *Phys. Rev. E* **54**, 6226 (1996).
- [33] W. H. White, *J. Colloid Interface Sci.* **87**, 204 (1982).
- [34] F. Family and P. Meakin, *Phys. Rev. A* **40**, 3836 (1989).
- [35] P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987).
- [36] P. Harremoës and F. Topsøe, *Entropy* **3**, 191 (2001).
- [37] R. Ferrer i Cancho and R. V. Solé, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 788 (2003).
- [38] W. Li, *IEEE Trans. Inf. Theory* **38**, 1842 (1992).
- [39] E. W. Montroll and M. F. Shlesinger, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3380 (1982).
- [40] K. Kawamura and N. Hatano, *J. Phys. Soc. Jpn.* **71**, 1211 (2002).
- [41] Luis A. Nunes Amaral, S. V. Buldyrev, S. Havlin, M. A. Salinger, and H. E. Stanley, *Phys. Rev. Lett.* **80**, 1385 (1998).
- [42] R. Solomonoff, *Inform and Control* **7**, 1 (1964).
- [43] A. Kolmogorov, *Probl. Inf. Transm.* **1**, 1 (1965).
- [44] G. J. Chaitin, *J. ACM* **13**, 547 (1966).
- [45] L. Ming and P. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications* (Springer, New York, 1997).
- [46] R. Perline, *Phys. Rev. E* **54**, 220 (1996).
- [47] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- [48] C. Adami, *Introduction to Artificial Life* (Springer, New York, 1999).
- [49] G. J. Chaitin, *J. ACM* **22**, 329 (1975).
- [50] J. S. Nicolis, *Rep. Prog. Phys.* **49**, 1109 (1986).

- [51] W. Ebeling and G. Nocolis, *EPL* **14**, 191 (1991).
- [52] W. Ebeling, *Physica A* **194**, 563 (1993).
- [53] F. Kaspar and H. G. Schuster, *Phys. Rev. A* **36**, 842 (1987).
- [54] W. H. Zurek, *Nature (London)* **341**, 119 (1989).
- [55] T. G. Dewey, *Phys. Rev. E* **54**, R39 (1996).
- [56] P. D. Grünwald and P. M. Vitányi, in *Handbook of the Philosophy of Science, Volume 8: Philosophy of Information*, edited by P. Adriaans and J. van Benthem (Elsevier, New York, 2008), pp. 289–325.
- [57] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [58] R. B. Ash, *Information Theory* (Dover, New York, 1990).
- [59] P. D. Grünwald and P. M. B. Vitányi, *J. Log. Lang. Inf.* **12**, 497 (2003).
- [60] E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- [61] H. Haken, *Synergetics: An Introduction. Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology*, Springer Series in Synergetics (Springer, New York, 1978).
- [62] J. N. Kapur, *Maximum-entropy Models in Science and Engineering* (Wiley Eastern Limited, New Delhi, 1989).
- [63] R. K. Pathria, *Statistical Mechanics*, 2nd ed. (Butterworth-Heinemann, 1996).
- [64] A. Hernando, D. Puigdomènech, D. Villuendas, C. Vesperinas, and A. Plastino, *Phys. Lett. A* **374**, 18 (2009).
- [65] *Handbook of Mathematical Functions*, NBS, Applied Mathematics Series, edited by M. Abramowitz and I. Stegun (U.S. Government Printing office, Washington, D.C., 1965), Vol. 55.
- [66] D. S. Jones, *Elementary Information Theory* (Oxford University Press, Oxford, 1979).
- [67] A probability distribution is dominated from some k by a power law with exponent $1 + \delta$ if $(\exists m): (\forall i > m) \left(\frac{p(i+1)}{p(i)} < \left(\frac{i}{i+1} \right)^{1+\delta} \right)$.
- [68] The reader could object that this section is unnecessary, since the axiomatic derivation of the uncertainty function (which we call entropy) assumes that the entropy increases with the size of the system. However, the explicit statement of this axiom corresponds to the special case of uniform probabilities [58]. Specifically, the axiom states that, if we have two systems A, B such that A contains n states a_1, \dots, a_n and B contains $n+1$ states, b_1, \dots, b_{n+1} , then, if $(\forall i \leq n) p(a_i) = 1/n$ and $(\forall i \leq n+1) p(b_i) = 1/(n+1)$ $H(A) < H(B)$. Thus, if we are not dealing with this special case, we need to explicitly demonstrate that it holds for our purposes.
- [69] This derivation is equivalent to the one found in [36], Theorem 8.2. In this theorem, the authors demonstrate that every infinite distribution with infinite entropy is *hyperbolic*, which implies that the distribution is not dominated by a power law with an exponent higher than 1.