

RESEARCH ARTICLE

Dynamical analogues of rank distributions

Carlos Velarde¹, Alberto Robledo²*

1 Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico City, Mexico, **2** Instituto de Física y Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

☞ These authors contributed equally to this work.

* robledo@fisica.unam.mx



Abstract

We present an equivalence between stochastic and deterministic variable approaches to represent ranked data and find the expressions obtained to be suggestive of statistical-mechanical meanings. We first reproduce size-rank distributions $N(k)$ from real data sets by straightforward considerations based on the assumed knowledge of the background probability distribution $P(N)$ that generates samples of random variable values similar to real data. The choice of different functional expressions for $P(N)$: power law, exponential, Gaussian, etc., leads to different classes of distributions $N(k)$ for which we find examples in nature. Then we show that all of these types of functions can be alternatively obtained from deterministic dynamical systems. These correspond to one-dimensional nonlinear iterated maps near a tangent bifurcation whose trajectories are proved to be precise analogues of the $N(k)$. We provide explicit expressions for the maps and their trajectories and find they operate under conditions of vanishing or small Lyapunov exponent, therefore at or near a transition to or out of chaos. We give explicit examples ranging from exponential to logarithmic behavior, including Zipf's law. Adoption of the nonlinear map as the formalism central character is a useful viewpoint, as variation of its few parameters, that modify its tangency property, translate into the different classes for $N(k)$.

OPEN ACCESS

Citation: Velarde C, Robledo A (2019) Dynamical analogues of rank distributions. PLoS ONE 14(2): e0211226. <https://doi.org/10.1371/journal.pone.0211226>

Editor: Miguel A. F. Sanjuán, Universidad Rey Juan Carlos, SPAIN

Received: August 20, 2018

Accepted: January 8, 2019

Published: February 4, 2019

Copyright: © 2019 Velarde, Robledo. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data underlying the results presented in the study are available in the citations given in the bibliography in the manuscript. These are Refs. [10], [11], [12], [13] and [23].

Funding: This work was supported by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), Universidad Nacional Autónoma de México (UNAM) <http://dgapa.unam.mx/index.php/impulso-a-la-investigacion/papiit>.

Competing interests: The authors have declared that no competing interest exist.

Introduction

There exist countless sets of data detailing magnitudes or sizes of a vast variety of measurable properties from many different fields: astrophysical, geophysical, ecological, biological, technological, financial, urban, social, etc. The magnitude data sets can be ranked and after examination the resultant distributions can be classified into different families or groups according to the functional expression that fits them best, power law, exponential, logarithmic, inverse error function, etc. Fig 1 exemplifies four such types of distributions. Over the years, [1–7] interest has been mostly placed on ranked data that exhibits power-law behavior even if this is not observed over the entire collection of records since this suggests a possible relationship with the famous empirical Zipf's law [8, 9]. Less attention has been placed to the comprehensive study of a wider range of types of rank distributions with the purpose of uncovering the

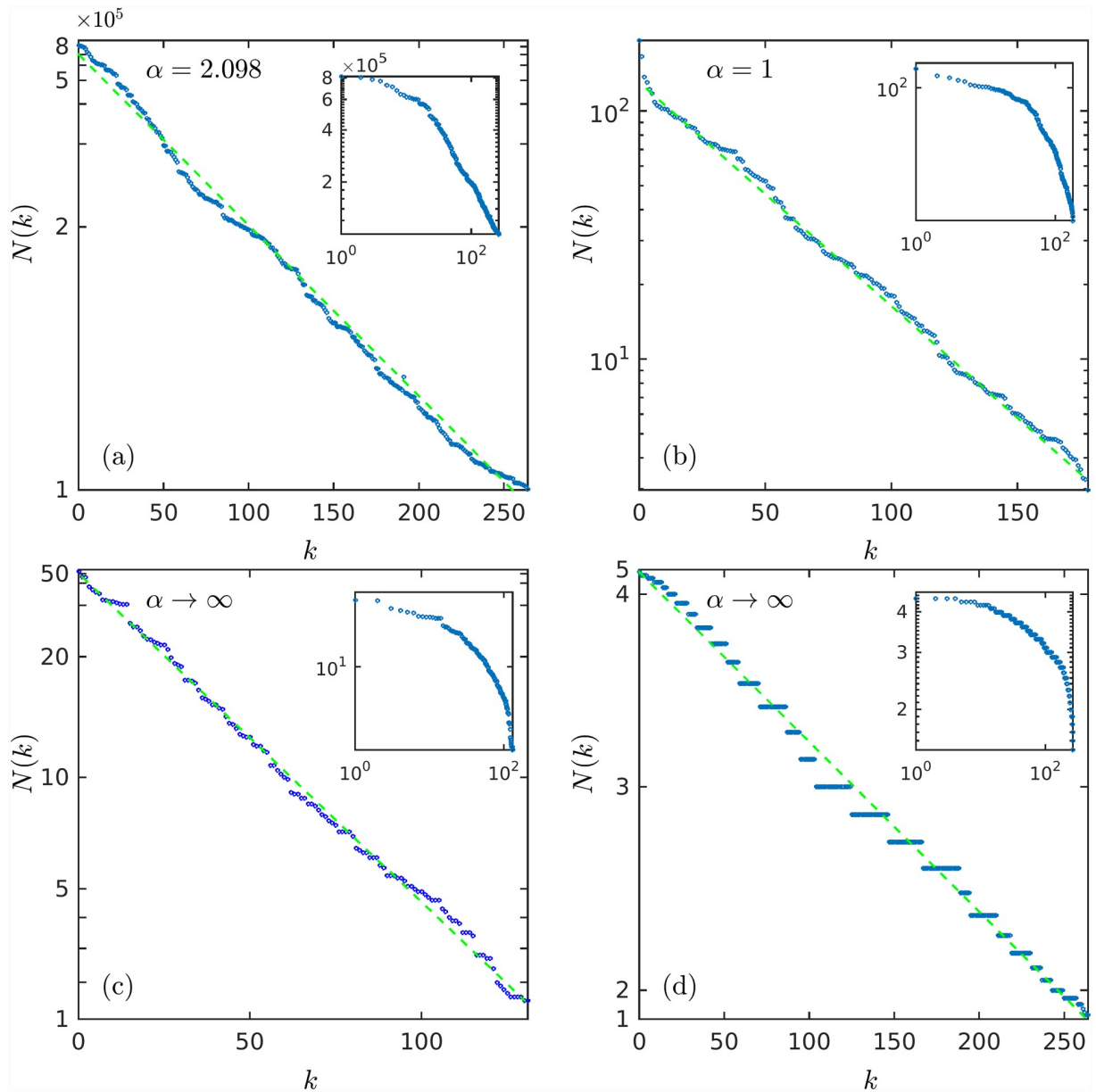


Fig 1. Ranked size data. Data $N(k)$ plotted in scales such that they appear fitted by decreasing straight lines. (a) USA city populations [10], ordinates in $\ln_e(N(k)/N_{\max})$ scale. (b) Infant mortality per country [11], ordinates in logarithmic scale. (c) Firearms owned per 100 capita and per country [12], ordinates in $-\exp(-N(k)/N_0)$ scale ($N_0 = 11.77$). (d) Los Angeles household sizes [13], ordinates in $\operatorname{erf}((N(k) - \mu)/(\sqrt{2}\sigma))$ scale ($\mu = 2.97, \sigma = 0.644$). The insets show the same data in logarithmic scales. See text for explanations, definitions and notation.

<https://doi.org/10.1371/journal.pone.0211226.g001>

broad-spectrum physics, if there is one, behind rank distributions. Here we consider this enterprise.

We start by recalling [3] a simple stochastic approach to reproduce size-rank data from an assumed parent distribution that governs the values of random variables that form finite samples. For every assumed form of the parent distribution one obtains a size-rank distribution. And this leads, when matched successfully with real data, to different possible families or universality classes. As an illustration we show (see Fig 1) four examples of data fitted by power law, exponential, logarithmic, and inverse error function expressions. The examples in Fig 1,

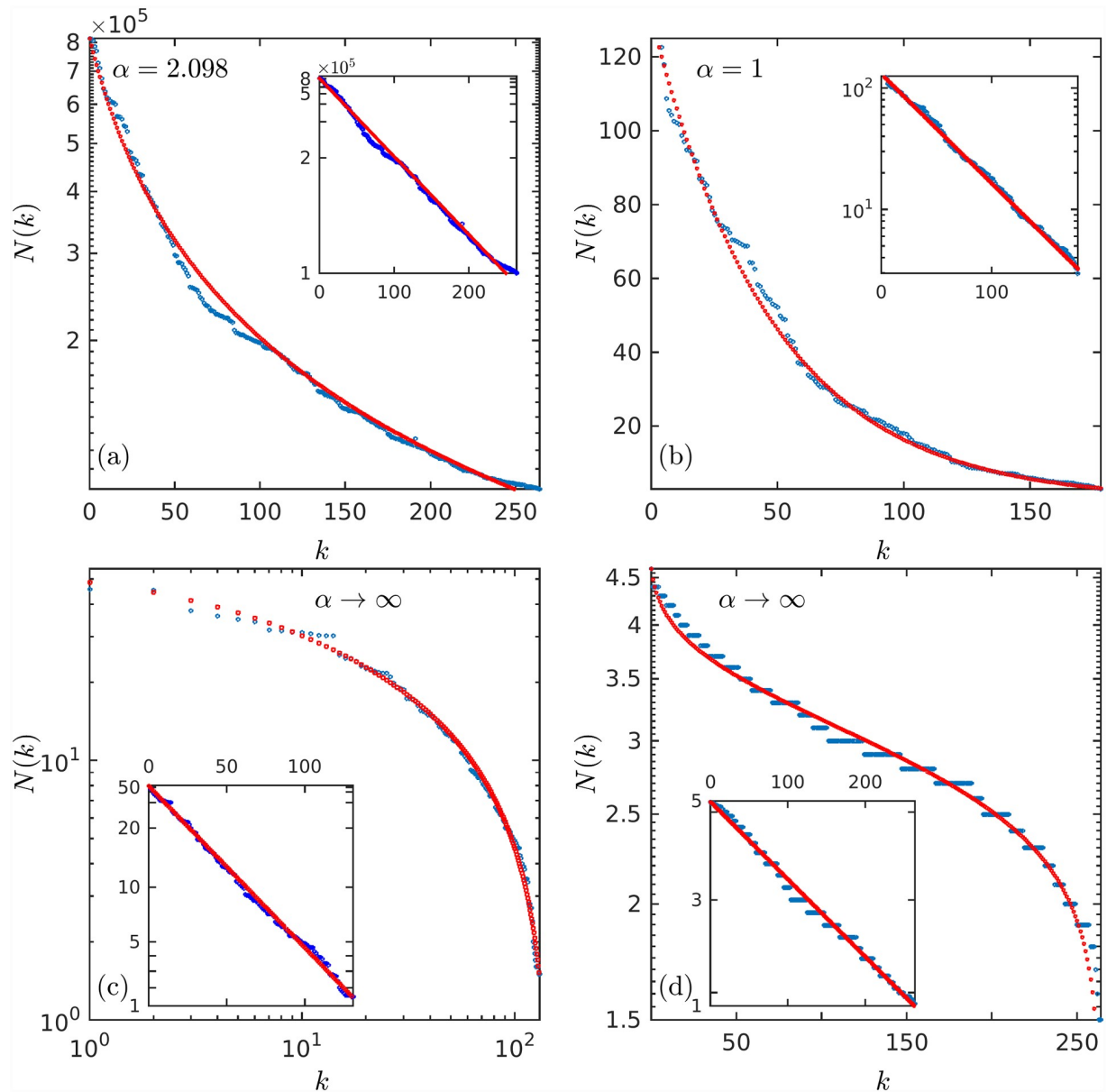


Fig 2. Data fitting. Another look at the fitting of data in Fig 1. (a) USA city populations, ordinates in ordinary logarithmic scale. (b) Infant mortality per country in ordinary linear scales. (c) Firearms owned per 100 capita and per country, in ordinary logarithmic scales, (d) Los Angeles household sizes, ordinates in ordinary logarithmic scale. The straight lines in the insets are obtained by using the same scales as those in Fig 1 or in the insets of Fig 3. See text for explanations, definitions and notation.

<https://doi.org/10.1371/journal.pone.0211226.g002>

and also further down in the text in Fig 2, correspond, respectively, to data on USA city populations [10], infant mortality [11], gun ownerships [12], and Los Angeles occupants of households [13]. The ordinates in Fig 1 and the ordinates in the insets in Fig 2 use scales such that the data approximate straight decreasing lines.

After this, we translate the stochastic approach into a deterministic one by deriving nonlinear one-dimensional iterated maps such that their trajectories match exactly the aforementioned classes of size-rank distributions. In Fig 3 we plot the maps that correspond to the same four examples in Figs 1 and 2 and where the insets show the trajectories plotted with the same

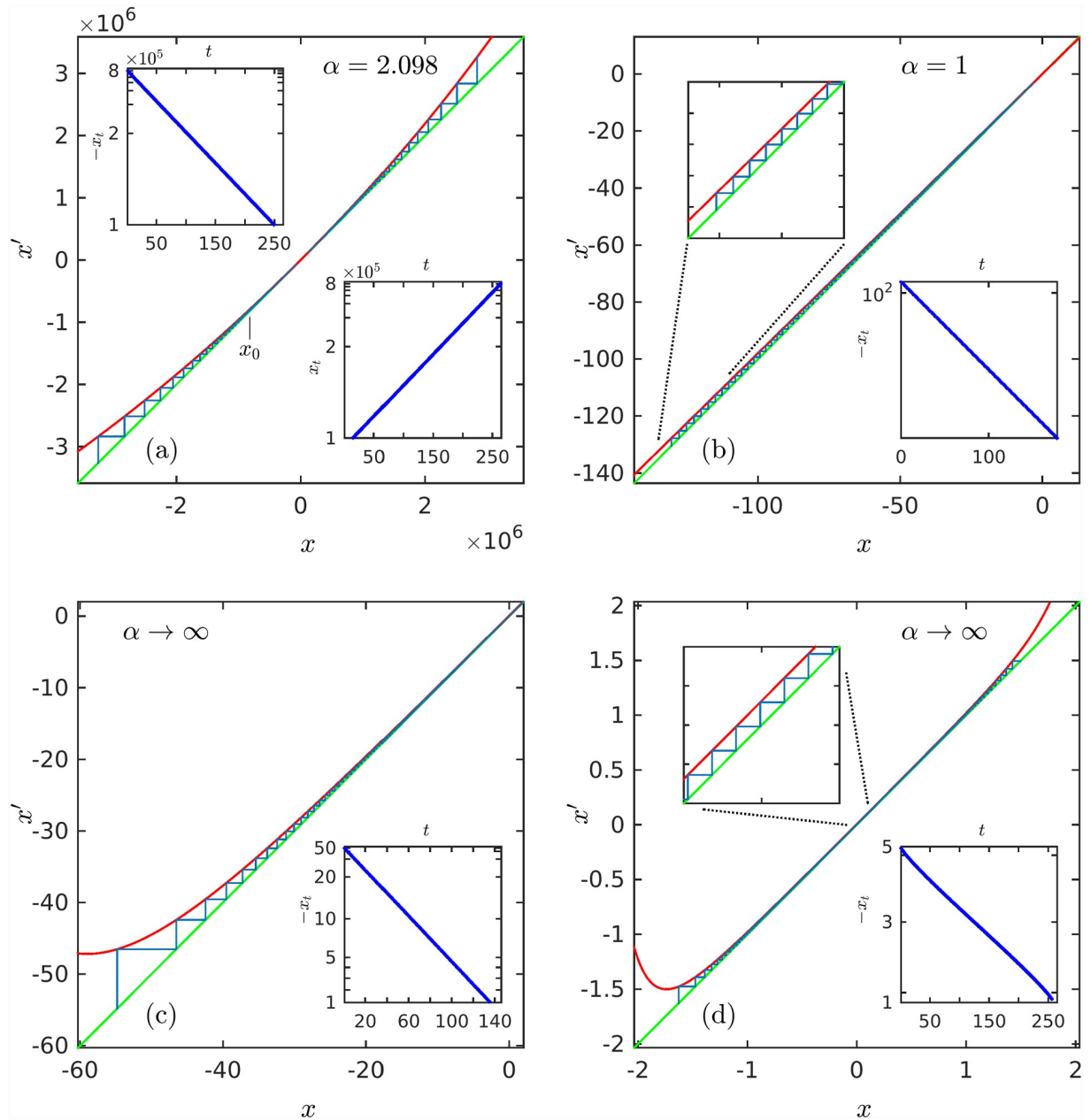


Fig 3. Nonlinear iterated maps and trajectories. Maps and trajectories that fit the data shown in Fig 1. The insets show the trajectories as (iterated) time series. (a) $x' = x \exp_2(ux^{\alpha-1})$, inset ordinates in $\mp \ln_2 [x_t/x_0]$ scale. (b) $x' = (1-a)x$, inset ordinates in logarithmic scale. (c) $x' = x + u \exp(-x/N_0)N_0$, inset ordinates in $-\exp(-x/N_0)$ scale. (d) $x' = x + u\sqrt{2\pi}\sigma \exp((x-\mu)/(\sqrt{2}\sigma))^2$, inset ordinates in $\text{erf}((-x, -\mu)/(\sqrt{2}\sigma))$ scale. See text for explanations, definitions and notation.

<https://doi.org/10.1371/journal.pone.0211226.g003>

scales as before such that they become straight decreasing lines. We find that these maps have as a common feature being close to or at tangency with the identity line and therefore their trajectories have zero or small Lyapunov exponent as in the transition in or out of chaos via the tangent bifurcation [14].

Consideration of the nonlinear map as the central object in the formalism offers a unifying viewpoint. As we shall see, variation of its nonlinearity takes the map from intersection (with the identity line) to tangency, then to a shift of the tangency position to infinity, or to a shape

with a central sector parallel and close to tangency. These circumstances translate, respectively, into different expressions for the rank distribution, first the mentioned exponential, then power law, next logarithmic, or inverse error function forms. We also present a novel use of the perturbed Renormalization Group (RG) fixed-point map of Hu and Rudnick [15] for the description of finite-size samples.

Finally, we indicate that the reciprocal of each value of the size-rank distribution (ranked magnitudes) can be used to define uniformly-distributed probabilities, one for each fixed value of the rank. And these in turn can be used to determine entropies. We discuss the occurrence of generalized entropy or information measure expressions in the limit when the Lyapunov exponent vanishes [16, 17]. These measures possess the extensivity property, where system (trajectory or sample) size is quantified by either final iteration time or maximum rank.

Originally [9], Zipf law referred to the number of occurrences of words in texts and since then it has been assigned also to the number of occurrences of other items and more freely to magnitudes or sizes of other entities. A large number of studies on the subject have been developed since and general considerations made, a taster is given in Refs [1–7]. Here we refer only to ranked magnitude or size data. For clarity of presentation a discussion on the distinction between the two kinds of rank distributions, frequency and magnitude, is not included here but can be found in [18].

A stochastic approach to size-rank distributions

The starting quantity in our formalism is the probability distribution $P(N)$ of the data N under consideration. This could have any form, we start by making use of the power law form [3, 18–20],

$$P(N) \sim N^{-\alpha}, \quad 1 \leq \alpha < \infty. \tag{1}$$

A set of data, or an ensemble of them, can be extracted from Eq (1) to be compared with real data sets. We shall be interested in size-rank distributions so that the values of N represent the magnitude or size of the items in the data set, and therefore N takes positive values within the interval $N_{\min} \leq N \leq N_{\max}$, with possibly $N_{\min} = 0$ and/or $N_{\max} \rightarrow \infty$. We use \mathcal{N} for the total number of data obtained from $P(N)$.

The \mathcal{N} items in the data set can be ordered or sorted out starting with the largest, N_{\max} , and proceeding with decreasing sizes down to N_{\min} . We label them with a new variable k , the rank, with $k = 0$ for N_{\max} and $k = k_{\max}$ for N_{\min} . We call the function $N(k)$ the size-rank distribution. The rank k can be an integer $k = 0, 1, 2, 3, \dots, k_{\max}$ (frequently the 1st value is $k = 1$) but it can be extended to be a real number. The main task is to determine $N(k)$ from $P(N)$.

Following [3] a well-known procedure to sort out random variable data [21], the next step is to introduce the (complementary) cumulative distribution of $P(N)$ as

$$\Pi(N, N_{\max}) = \int_N^{N_{\max}} P(N) dN, \tag{2}$$

so that normalization of $P(N)$ means $\Pi(N_{\min}, N_{\max}) = 1$. It is through the distribution $\Pi(N, N_{\max})$ that the set of data generated by $P(N)$ is ordered by size to deliver $N(k)$. As N is varied from N_{\max} all the way down to N_{\min} the distribution $\Pi(N, N_{\max})$ increases monotonically taking values from $\Pi(N_{\max}, N_{\max}) = 0$ to $\Pi(N_{\min}, N_{\max}) = 1$, and this distribution can be

identified with k if properly scaled, that is

$$k \equiv \mathcal{N}\Pi(N(k), N_{\max}). \tag{3}$$

The dependence of N on k can be obtained by solving

$$\frac{k}{\mathcal{N}} = \int_{N(k)}^{N_{\max}} P(N)dN, \tag{4}$$

for $N(k)$. We call $N(k)$ the size-rank distribution even though it is not a probability distribution [18], as $P(N)$ and $\Pi(N, N_{\max})$ are. If $P(N)$ requires normalization, it must be divided by the constant

$$C = \int_{N_{\min}}^{N_{\max}} P(N)dN, \tag{5}$$

and Eq (4) generalises to

$$\frac{k}{\mathcal{N}} = \frac{1}{C} \int_{N(k)}^{N_{\max}} P(N)dN.$$

Defining $\mathcal{N}' \equiv \mathcal{N}/C$ we have

$$\frac{k}{\mathcal{N}'} = \int_{N(k)}^{N_{\max}} P(N)dN. \tag{6}$$

Different classes of size-rank distributions

We look now at the specific expressions that come out of Eq (4) when $P(N)$ is given by Eq (1). We have

$$\begin{aligned} \Pi(N(k), N_{\max}) &= \frac{1}{C_\alpha} \int_{N(k)}^{N_{\max}} N^{-\alpha} dN \\ &= \frac{1}{C_\alpha(1-\alpha)} [N_{\max}^{1-\alpha} - N(k)^{1-\alpha}] \\ &= \frac{1}{C_\alpha} [\ln_\alpha N_{\max} - \ln_\alpha N(k)], \end{aligned} \tag{7}$$

where $\ln_q(x) \equiv (1-q)^{-1}[x^{1-q} - 1]$ with q a positive real number is the q -deformed logarithmic function, and $C_\alpha = \ln_\alpha N_{\max} - \ln_\alpha N_{\min}$. From Eqs (6) and (7) it follows

$$\ln_\alpha N(k) = \ln_\alpha N_{\max} - \mathcal{N}'^{-1}k. \tag{8}$$

The normalization constant C_α is finite if $N_{\min} > 0$ and $N_{\max} < \infty$. When $\alpha > 1$ C_α is finite when $N_{\max} \rightarrow \infty$ but $N_{\min} > 0$.

The size-rank distribution $N(k)$ is explicitly obtained from the above with use of the inverse of $\ln_q(x)$, the q -deformed exponential function $\exp_q(x) \equiv [1 + (1 - q)x]^{1/(1 - q)}$, this is

$$N(k) = N_{\max} \exp_z[-N_{\max}^{\alpha-1} \mathcal{N}^{\alpha-1} k], \tag{9}$$

so that when $\alpha = 1$ we have the ordinary exponential form

$$N(k) = N_{\max} \exp(-\mathcal{N}^{\alpha-1} k). \tag{10}$$

We can take the limit $\alpha \rightarrow \infty$ to signify, for instance, that

$$P(N) = N_0^{-1} \exp(-N/N_0), \tag{11}$$

and we choose $N_0 = 1$ to find

$$N(k) = -\ln[\exp(-N_{\max}) + \mathcal{N}^{\alpha-1} k]. \tag{12}$$

In the limit $N_{\max} \rightarrow \infty$ Eq (9) becomes the power law $N(k) \sim k^{1/(1-\alpha)}$ that when $\alpha = 2$ gives the simple hyperbolic form $N(k) \sim k^{-1}$, the classical Zipf law for ranked size data. In contrast, when $\alpha \rightarrow \infty$, in the limit $N_{\max} \rightarrow \infty$ the rank distribution becomes $N(k) = -\ln(k/\mathcal{N}^\alpha)$, $N(k)$ decays very fast as k increases since the argument in the logarithmic function lies in the interval $0 \leq k/\mathcal{N}^\alpha \leq 1$. This can be compared with the case $\alpha = 1$, but N_{\max} finite, when $N(k)$ decays exponentially as k increases.

Another, important, option for $P(N)$ for the limit when $\alpha \rightarrow \infty$ is the Gaussian distribution

$$P(N) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(N - \mu)^2}{2\sigma^2}\right), \tag{13}$$

for which we find

$$N(k) = \mu + \sqrt{2}\sigma \operatorname{erfc}^{-1}\left[\operatorname{erfc}\left(\frac{N_{\max} - \mu}{\sqrt{2}\sigma}\right) + 2k\mathcal{N}^{\alpha-1}\right]. \tag{14}$$

The data in the panels of Fig 1 are plotted using scales such that $N(k)$ given by Eqs (9), (10), (12) and (14) appear as decreasing straight lines.

Nonlinear dynamical analogues of size-rank distributions

The stochastic procedure described above to obtain size-rank functions can be transformed into a deterministic one where trajectories $x_t, t = 0, 1, \dots$ of a one-dimensional nonlinear iterated map $x' = f(x)$, reproduce the functions $N(k), k = 0, 1, \dots$. The general expression for this map is

$$f(x) = x + u/P(-x), u > 0, \tag{15}$$

where $P(-x) \geq 0$ is, as above, the parent distribution and $u = \mathcal{N}^{\alpha-1}$. To obtain Eq (15) we differentiate Eq (6) to give

$$-\frac{1}{\mathcal{N}^\alpha} \frac{dk(N)}{dN} = P(N), \tag{16}$$

or, in terms of dynamical variables,

$$u \frac{dt}{dx_t} = P(-x_t). \tag{17}$$

Use of $dx_t/dt \simeq x_{t+1} - x_t$ recovers Eq (15).

The size-rank distributions $N(k)$ obtained from the trajectories x_t of any map $f(x)$ in our procedure are monotonically decreasing functions of its rank k . This is so if all trajectories x_t generated by $f(x)$ are monotonically increasing functions of the iterated time t , as $N = -x$ and $k = t$. This is so because the condition $f(x) \geq x$ is always satisfied by Eq (15). Notice that $f(x)$ (or $P(N)$) may not be monotonic functions themselves.

This exact analogy between the expressions for the rank distributions and those for the dynamics in a one-dimensional map was recognized before for the special case of the power law parent distribution in Eq (1) [18–20], where the map is at a tangent bifurcation [14]. In this case $f(x)$ is written locally as

$$x' = f(x) = x + u|x|^z + \dots, \quad x \leq 0, \quad z > 1, \tag{18}$$

and trajectories initiated at $x_0 \leq 0$ are obtained from

$$x_{\tau+1} = x_\tau + u|x_\tau|^z, \quad \tau = 0, 1, \dots \tag{19}$$

These trajectories move monotonically towards the point of tangency at $x = 0$. If we make the replacement, valid for large time τ , of the difference $x_{\tau+1} - x_\tau$ by $dx_\tau/d\tau$ in Eq (19) (written as $u|x_\tau|^z = x_{\tau+1} - x_\tau$) we obtain the differential form $ud\tau = |x_\tau|^{-z} dx_\tau$, and integration of both sides of it yields

$$\begin{aligned} ut &= \int_{x_0}^{x_t} \frac{dx_\tau}{|x_\tau|^z} = \int_{-x_t}^{-x_0} \frac{dx_\tau}{|x_\tau|^z} \\ &= \frac{1}{1-z} [|x_0|^{1-z} - |x_t|^{1-z}], \end{aligned} \tag{20}$$

or

$$\ln_z |x_t| = \ln_z |x_0| - ut. \tag{21}$$

The iteration number or time t dependence of all trajectories is obtained by solving the above for x_t , i.e.

$$x_t = x_0 \exp_z [x_0^{z-1} ut], \tag{22}$$

where $x^{z-1} \equiv \text{sign}(x) |x|^{z-1}$. The equivalence of the trajectory positions x_t with the size-rank distribution $N(k)$ is made clear by comparison of Eqs (21) and (22) with Eqs (8) and (9), respectively, together with the identifications $t = k$, $u = \mathcal{N}^{z-1}$, $x_t = -N(k)$, $x_0 = -N_{\max}$ and $z = \alpha$. When $z = 1$ Eq (22) becomes an ordinary exponential,

$$x_t = x_0 \exp(-at), \tag{23}$$

that matches, with use of the previous identifications, Eq (10). As we see below, a linear map that intersects the identity line,

$$x' = f(x) = (1 - a)x, \tag{24}$$

reproduces the trajectories in Eq (23).

In establishing the equivalence between trajectories x_t and rank functions $N(k)$ when $1 < \alpha = z < \infty$ we used starting positions at the left of the point of tangency $x_0 < 0$, but we could have obtained the same monotonically decreasing $N(k)$ from starting positions at the right of the tangency point $x_0 > 0$ with the use of the identifications $k = T - t$, $x_T = N_{\max}$, $x_t = N(k)$, where T is the last iteration time.

We call attention to the fact [22] that the map trajectories given by Eq (22) are precisely those that are generated by the renormalization group (RG) fixed-point map [14, 15]

$$x' = f(x) = x \exp_z(ux^{z-1}), \tag{25}$$

where this map was obtained [14, 15] as the solution of the functional composition RG fixed point equation

$$f(f(x)) = \gamma^{-1}f(\gamma x), \tag{26}$$

with $\gamma = 2^{z-1}$ and where the expansion of $f(x)$ starts as the required local form for tangency $f(x) = x + u|x|^z + \dots$, $x \leq 0$, $z > 1$.

In the limit $z \rightarrow \infty$ the counterpart of Eq (22) is

$$x_t = \ln [\exp(x_0) + ut], \tag{27}$$

as this expression transforms into Eq (12) for $N(k)$ under the same equivalences $t = k$, $u = N'^{-1}$, $x_t = -N(k)$, $x_0 = -N_{\max}$. Differentiation of Eq (27) gives

$$\frac{dx_t}{dt} = u \exp(-x_t), \tag{28}$$

so that use of $x_{t+1} - x_t \simeq dx_t/dt$, $t \gg 1$, yields the map

$$x' = x + u \exp(-x). \tag{29}$$

We note that the tangency of this map with the identity line is located at $x \rightarrow \infty$ and not at $x = 0$ as it is the case of the map in Eq (25), the effect of taking $z = \alpha \rightarrow \infty$ has shifted the tangency position. To obtain the trajectories x_t in Eq (27) from the map in Eq (29) it is necessary to perform a coordinate transformation to bring the point of tangency back to $x = 0$.

We can follow the same procedure for the above cases to determine the map that corresponds to the Gaussian distribution choice for $P(N)$ and, consequently, the inverse complementary error function for $N(k)$. This is

$$x' = x + u\sqrt{2\pi}\sigma \exp\left(\frac{(x - \mu)^2}{2\sigma^2}\right). \tag{30}$$

This map never touches the identity line but it comes increasingly closer and parallel to this line as $u \rightarrow 0$. The trajectories generated by the map in Eq (30) are given by

$$x_t = -\mu - \sqrt{2}\sigma \operatorname{erfc}^{-1}\left[\operatorname{erfc}\left(\frac{-x_0 - \mu}{\sqrt{2}\sigma}\right) + 2ut\right]. \tag{31}$$

In Fig 3 we show the maps and their trajectories that correspond to the four cases described here, those in Eqs (25), (24), (29) and (30). The insets in the panels of Fig 3 are plotted using scales such that the trajectories x_t given by Eqs (22), (23), (27) and (31) appear as decreasing straight lines. An exception is the bottom inset of panel (a) that corresponds to an initial position placed at the right of the point of tangency. In Fig 2 we reproduce the data in the panels of

Fig 1 using the same scales as before so that the straight lines correspond to the maps trajectory positions $-x_t = N(k), t = k$.

A closer look at rank distributions for finite-size samples

As mentioned, for finite data sets the distribution $P(N)$ cannot be a pure power law as in Eq (1). We use the analogy with the map at tangency to deal with this circumstance without the need to find the explicit form for $P(N)$. When the tangent RG fixed-point map in Eq (25) is perturbed by a small amount ϵ one obtains [15]

$$f(x, \epsilon) = x \exp_z[ux^{z-1}] + \frac{\epsilon}{z-1} x^{-(z-1)} (1 - (z-1)ux^{z-1} - \exp_z[ux^{z-1}]^z) + O(\epsilon^2). \tag{32}$$

According to the sign of ϵ the map in Eq (25) is shifted off tangency letting trajectories progress across a narrow channel (see Fig 4(a)), or producing a bisection of the map with the identity line (see Fig 4(b)). The effect on the rank distribution $N(k)$ can be observed by use of the identifications given above between nonlinear map and rank distribution variables to find the appropriate results for $\mathcal{N} < \infty$. In doing this the dependence of x_t on the shift parameter ϵ is translated into $N(k)$, and one immediate way of assessing its relevance to the description of finite $\mathcal{N} < \infty$ samples is through direct comparison.

When $\epsilon > 0$ the map intersects the identity line $x' = x$ twice, say at $x_l < 0$ and $x_r > 0$. The map in the neighborhood of x_l has the form $f(x) = x - ax + \dots$, where the variable x has been redefined by a shift such that $x = 0$ is the old x_l , and trajectories close to x_l evolve as $x_{t+1} = x_t - ax_t + \dots$. The differential version of the latter is $adt = -x_t^{-1} dx_t$, so that, integration of both sides of it yields $-\ln(x_t/x_0) = at$, or

$$x_t = x_0 \exp(-at), \quad x_0 < 0. \tag{33}$$

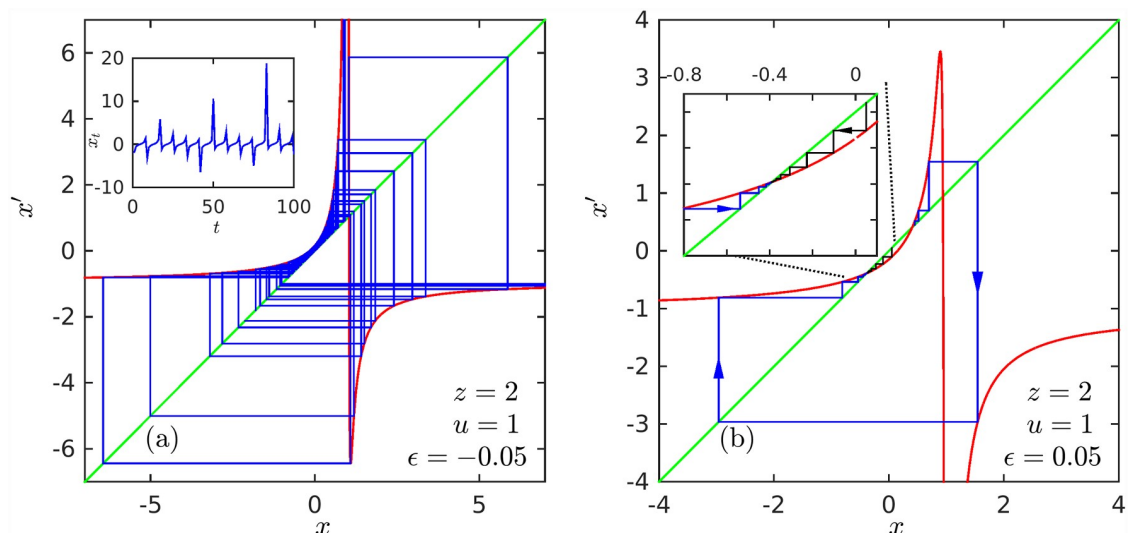


Fig 4. Perturbed RG fixed-point maps. (a) $\epsilon < 0$. A chaotic trajectory is shown performing passages through the map narrow channel close to the identity line mediated by reinjections that make use of the bottom branch of the map that occurs after the map singularity at 0. (b) $\epsilon > 0$. The intersections of the map with the identity line, attractor and repeller positions, generate trajectories that are guided towards the attractor and/or driven away from the repeller.

<https://doi.org/10.1371/journal.pone.0211226.g004>

This is the map equivalent to the size-rank distribution $N(k)$ given in Eq (10) for the case when $\alpha = 1$ (where again $t = k$, $a = \mathcal{N}^{\prime-1}$, $x_t = -N(k)$ and $x_0 = -N_{\max}$). Therefore the effect of a positive shift $\varepsilon > 0$ is to transform the power law $P(N) \sim N^{-\alpha}$, $\alpha > 1$, by the hyperbolic form $P(N) \sim N^{-1}$ as N increases. When $\varepsilon > 0$ is very small the map is very close to tangency, the intersection $x_t \sim 0$, and a crossover should be observed in $N(k)$ from power law $k^{-\alpha}$ to exponential decay.

In establishing the equivalence between the trajectory x_t and the rank function $N(k)$ we used starting positions at the left of the intersection $x_t < 0$, but we could have obtained the same monotonically decreasing $N(k)$ from starting positions at the right of the intersection $x_t > 0$, after a shift such that $x = 0$ is the old x_t , and with the use of the identifications $k = T - t$, $x_T = N_{\max}$, $x_t = N(k)$, where T is the last iteration time.

When $\varepsilon < 0$ the map does not intersect the identity line $x' = x$ but there is a narrow channel through which trajectories started at $x_0 < 0$ slip into positions $x_t > 0$. Once a trajectory traverses the channel it increases its value x_t rapidly, and this behavior is reflected in $N(k)$ as a sharp bend towards $N = 0$ reaching this value at a finite rank, $k_{\max} < \infty$. In this way the outcome of a negative shift $\varepsilon < 0$ is a finite size effect. The transformation of x_t into $N(k)$ follows the earlier identifications $t = k$, $u = \mathcal{N}^{\prime-1}$, and $z = \alpha$, but now $x_0 - x^* = -N_{\max}$ and $x_t - x^* = -N(k)$, where the translation x^* ensures that all $N(k) \geq 0$. In particular, for a map with parameters (z, u) and given value of ε , a trajectory started at x_0 and recorded until a certain maximum value x_T at time T before escaping from the narrow channel we have $x^* = x_T$ and $T = k_{\max}$. We show in Fig 5 (see inset) that this indeed is observed in real data. We used data in this particular case of numbers of tallest buildings [23].

Notice that for the map at tangency, $\varepsilon = 0$, any trajectory with $x_0 < 0$ evolves towards $x = 0$ asymptotically as a power law $t^{1/(1-z)}$, $z > 1$, reaching $x = 0$ at $t \rightarrow \infty$. This implies $k_{\max} = \mathcal{N} \rightarrow \infty$, and signifies infinite size or a ‘thermodynamic limit’ as suggested in the following Section. The shift off tangency $\varepsilon < 0$ translates into $k_{\max} = \mathcal{N} < \infty$ for the rank distribution.

Chaotic dynamics arises via the intermittency property of trajectories when $\varepsilon = 0$ turns $\varepsilon \lesssim 0$ (inset of Fig 4a). The passages through the narrow channel are the so-called laminar episodes (monotonic curves) separated by re-injections (that break monotonicity). One such laminar episode was used to fit the data in Fig 5. The duration D of the laminar episodes depends on the value of ε (D diverges as ε vanishes as a known power law [14]). In the analogy D equals the maximum value of the rank k_{\max} , the size of the data sample. This connection provides a means to treat finite size without the need to determine the appropriate parent distribution $P(N)$ that is required to replace the pure power law (strictly compatible only with infinite size) $P(N)$ in Eq (1). A sequence of laminar episodes (as shown in the inset of Fig 4a) provides an ensemble of finite-size data samples.

A statistical-mechanical structure for size-rank distributions

As pointed out before [18–20] the size-rank expressions such as those in Eqs (8) and (9) are reminiscent of those in a statistical-mechanical and thermodynamic structure. First consider that for a fixed value of k the probability for each unit of N that makes up $N(k)$ to occur is $p_i^{(k)}$, $i = 1, \dots, N(k)$. If this probability is uniform as in a microcanonical ensemble, we have $p_i^{(k)} = p^{(k)} \equiv 1/N(k)$ for all $i = 1, \dots, N(k)$, and this for each $k = 0, 1, \dots, k_{\max}$. Then, if we consider the case $\alpha = 1$ and define $S(k) \equiv \ln N(k)$ and $S_{\max} \equiv \ln N_{\max}$, Eq (8) reads

$$S(k) = S_{\max} - \mathcal{N}^{\prime-1}k, \tag{34}$$

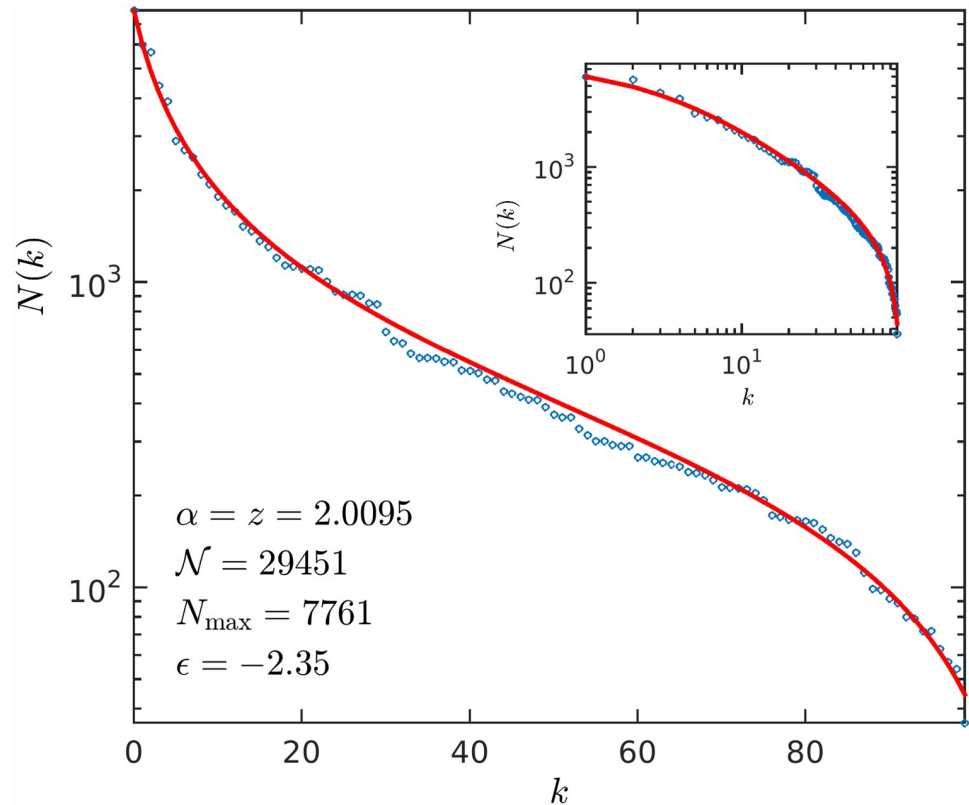


Fig 5. Rank-order statistics for cities with largest numbers of tallest buildings. Data (empty circles) from Reference [23]. Fitting obtained from a trajectory (smooth curve) of the map in Eq (32) with the identifications provided in the text.

<https://doi.org/10.1371/journal.pone.0211226.g005>

where if $S_{\max}(\mathcal{N}'^{-1})$ is thought of as an entropy for the system with fixed \mathcal{N}' then $S(k)$ would be a Massieu potential [24] after the variable \mathcal{N}'^{-1} has been replaced by the (conjugate) variable k as in a Legendre transformation. Eq (10),

$$N(k) = N_{\max} \exp(-\mathcal{N}'^{-1}k), \tag{35}$$

can be taken to be the outcome of the steepest-descent property arising for large \mathcal{N}' on

$$N(k) = \int N_{\max}(\mathcal{N}') \exp(-\mathcal{N}'^{-1}k) d\mathcal{N}'^{-1}. \tag{36}$$

That is, we consider that there are fluctuations in the data sample size \mathcal{N}' , and consequently also in the degeneration $N_{\max}(\mathcal{N}')$, but these become increasingly irrelevant for large \mathcal{N}' as the limit $k_{\max} = \mathcal{N} \rightarrow \infty$, and $N_{\min} \rightarrow 0$, and a dominant configuration is established in analogy with the thermodynamic limit.

The case $\alpha > 1$, where now $S(k) \equiv \ln_{\alpha} N(k)$ and $S_{\max} \equiv \ln_{\alpha} N_{\max}$, together with the use of \exp_{α} instead of the ordinary \exp in Eqs (35) and (36) demands careful analysis of the limit $k_{\max} = \mathcal{N} \rightarrow \infty$. It has been argued recently [16, 17] that this type of generalized expression for the entropy appears in situations where a drastic reduction of accessibility to configuration space is forced upon a system, such that the final set of reachable configurations has a vanishing measure with respect to the initial set. Attractors in dissipative nonlinear systems provide a mechanism for such severe reduction of accessible states, and the case of the tangent

bifurcation, as other transitions to chaos in low-dimensional iterated maps, have proved to be suitable model systems to analyze this circumstance [16, 17]. Interestingly, the RG fixed-point map Eq (25) provides a simple analytical expression that illustrates both this phase space (or position interval) contraction (from a set of real numbers to only a finite set) and a connection with the size-rank distribution $N(k)$ in Eq (9).

The above entropy expressions form part of a set of extensive entropy (or information) measures that comprise the four types of distributions $N(k)$ we have described (Eqs (9), (10), (12) and (14) and shown in Fig 2). These measures are given by their corresponding unnormalized cumulative distributions $\mathcal{N}\Pi(N(k), N_{\max})$:

$$\mathcal{N}\Pi(N(k), N_{\max}) = k_{\max}[\ln N_{\max} - \ln N(k)], \quad \alpha = 1, \tag{37}$$

$$\mathcal{N}\Pi(N(k), N_{\max}) = k_{\max}[\ln_{\alpha} N_{\max} - \ln_{\alpha} N(k)], \quad 1 < \alpha < \infty, \tag{38}$$

$$\mathcal{N}\Pi(N(k), N_{\max}) = k_{\max}[\exp(N(k)) - \exp(N_{\max})], \quad \alpha \rightarrow \infty, \tag{39}$$

and

$$2\mathcal{N}\Pi(N(k), N_{\max}) = k_{\max} \left[\operatorname{erfc} \left(\frac{N(k) - \mu}{\sqrt{2}\sigma} \right) - \operatorname{erfc} \left(\frac{N_{\max} - \mu}{\sqrt{2}\sigma} \right) \right], \quad \alpha \rightarrow \infty. \tag{40}$$

Since the size of the system is given by the rank k , $0 < k < k_{\max}$ that generates the data set $N(k)$, $k = 0, 1, \dots$, then by virtue of Eq (3) the above measures are extensive. In the nonlinear map language the measures are areas under the maps $f(x)$ that grow linearly with iteration time (c.f. Eq (20)).

Summary and discussion

We described different classes of size-rank distributions $N(k)$ that originate each from a different parent distribution $P(N)$ for data samples of the size random variable N . We considered power law, exponential, and Gaussian parent distributions to obtain analytic expressions for $N(k)$. Then we derived for each of these cases expressions of one-dimensional nonlinear iterated maps such that their trajectories x_t are exact functional analogues of the size-rank distributions $N(k)$, via $t = k$, $x_t = -N(k)$, etc. Significantly, all the equivalent maps appear at or close to tangency with the identity line, as can be clearly seen in Fig 3, so that the Lyapunov exponent vanishes or is close to zero, and dynamical evolution takes place at or is close to a transition in or out of chaos via the tangent bifurcation. The Lyapunov exponent λ vanishes for the map in Eq (25) and is proportional to $\epsilon \simeq 0$ for the perturbed map in Eq (32) [14, 22]. Straightforward calculations show that λ vanishes for the maps in Eq (29) (for all u) and Eq (30) (for $u \simeq 0$). Seen via the dynamical analogues there is a unifying thread that links these classes of rank functions: When the nonlinearity is $z = 1$ the map intersects the identity line and we have an exponentially-decaying $N(k)$, Eq (10). When $1 < z < \infty$ the map is tangent to the line and $N(k)$ decays as a power law, Eq (9). When $z \rightarrow \infty$ we observed that the point of tangency shifts to infinity and $N(k)$ acquires a logarithmic form, Eq (12). Also, for the same $z \rightarrow \infty$ the map can show a shape with a central sector parallel and close to tangency in which case $N(k)$ exhibits an inverse error function form, Eq (14). The latter class represents the common situation in which the size variable N is controlled by the Central Limit Theorem. The description of rank functions in the alternative language of nonlinear dynamics opens the possibility of gains with the use of a different narrative. All one-dimensional nonlinear maps are dissipative so that they possess attractors and repellers and their known dynamical properties [14], some of

which we have used here, may find further application. For all classes of size-rank distributions described we made comparison with real data and found agreement.

When the parent distribution is a power law, $P(N) \sim N^{-\alpha}$, $\alpha > 1$, the resulting map is that which corresponds to the fixed point of the RG approach where the basic transformation is functional composition, i.e. a deformed exponential map, c.f. Eq (25), which describes scaling at the tangent bifurcation. There, when the largest data $N_{\max} \rightarrow \infty$, $N(k)$ becomes a power law $N(k) \sim k^{1/(1-\alpha)}$, the classical Zipf law for $\alpha = 2$. This is the most common case since then the local map at tangency is an analytic function with nonzero curvature. We have also argued about the statistical-mechanical meaning of the expressions we use to reproduce ranked data.

Significantly, the RG fixed-point map also appears as a central element in the description of critical fluctuations [25, 26] and helps determine their temporary, intermittent, nature. Also, the divergence of the critical correlation length implies the vanishing of the Lyapunov exponent at the tangent bifurcation [25, 26].

Author Contributions

Conceptualization: Carlos Velarde, Alberto Robledo.

Formal analysis: Carlos Velarde, Alberto Robledo.

Funding acquisition: Alberto Robledo.

Investigation: Carlos Velarde, Alberto Robledo.

Methodology: Carlos Velarde, Alberto Robledo.

Writing – original draft: Carlos Velarde, Alberto Robledo.

Writing – review & editing: Carlos Velarde, Alberto Robledo.

References

1. To Honor G.K. Zipf Glottometrics 3,4,5 Ludenscheid: RAM-Verl., 2002 ISSN 1617-8351. Available from: <http://www.ram-verlag.eu/journals-e-journals/glottometrics/>.
2. Newman M.E.J. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 2005; 46(5):323–351. <https://doi.org/10.1080/00107510500052444>.
3. Pietronero L.Tosatti E.Tosatti V.Vespignani A. Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. Physica A. 2001; 293(1-2):297–304. [https://doi.org/10.1016/S0378-4371\(00\)00633-6](https://doi.org/10.1016/S0378-4371(00)00633-6).
4. Cristelli M.Batty M.Pietronero L. There is more than a power law in Zipf. Sci Rep 2012; 2(812):1–7. <https://doi.org/10.1038/srep0082>.
5. Finley B.J.Kilki K. Exploring empirical rank-frequency distributions longitudinally through a simple stochastic process. PLoS ONE 2014; 9(4):1–14. <https://doi.org/10.1371/journal.pone.0094920>.
6. Moreno-Sánchez I.Font-Clos F.Corrall A. Large scale analysis of Zipf's law in English texts. PLoS ONE 2016; 11(1):1–19. <https://doi.org/10.1371/journal.pone.0147073>.
7. Lestrade S. Unzipping Zipf's law. PLoS ONE 2018; 12(8):1–13. <https://doi.org/10.1371/journal.pone.0181987>.
8. van der Galien S.J.G. Zipf's law. Wikipedia. Available from: http://en.wikipedia.org/wiki/Zipfs_law (2003).
9. Zipf G.K. Human Behavior and the Principle of Least Effort. Cambridge: Addison-Wesley, 1949.
10. USA city populations. The Rank-Size Rule of City Populations (USACities2009.csv). Available from: <http://www.statisticalconsultants.co.nz/blog/the-rank-size-rule-of-city-populations.html>.
11. Infant Mortality. NationMaster Geographical Data. Available from: <http://www.nationmaster.com/country-info/stats/Health/Infant-mortality-rate>.
12. Estimated number of guns per 100 capita by country. Wikipedia. Available from: http://en.wikipedia.org/wiki/Number_of_guns_per_capita_by_country.

13. Household Size. Los Angeles Times. Available from: <http://projects.latimes.com/mapping-la/neighborhoods/household-size/neighborhood/list/>.
14. Schuster H.G. Deterministic Chaos. An Introduction. VCH Publishers, Weinheim; 1988.
15. Hu B.Rudnick J. Exact solutions to the Feigenbaum renormalization- group equations for intermittency. Phys Rev Lett, 1982; 48: 1645–1648. <https://doi.org/10.1103/PhysRevLett.48.1645>.
16. Yalcin G.C.Robledo A.Gell-Mann M. Incidence of q statistics in rank distributions. Proc Natl Acad Sci USA. 2014; 111(39):14082–14087. <https://doi.org/10.1073/pnas.1412093111>. PMID: 25189773
17. Yalcin G.C.Velarde C.Robledo A. Entropies for severely contracted configuration space. Heliyon, 2015; 1:e00045. Available from: <http://www.heliyon.com/article/e00045/> <https://doi.org/10.1016/j.heliyon.2015.e00045>. PMID: 27441229
18. Velarde C.Robledo A. Rank distributions: Frequency vs. Magnitude. PLoS ONE 2017; 12(10):1–13. <https://doi.org/10.1371/journal.pone.0186015>.
19. Altamirano C.Robledo A. Possible thermodynamic structure underlying the laws of Zipf and Benford. Eur Phys J B. 2011; 81(3):345–351. <https://doi.org/10.1140/epjb/e2011-10968-5>.
20. Robledo A. Laws of Zipf and Benford, intermittency, and critical fluctuations. Chinese Sci Bull. 2011; 56(34):3645–3648. <https://doi.org/10.1007/s11434-011-4827-y>.
21. Empirical cumulative distribution function. Wikipedia. Available from: https://en.wikipedia.org/wiki/Empirical_distribution_function.
22. Baldovin F.Robledo A. Sensitivity to initial conditions at bifurcations in one-dimensional nonlinear maps: Rigorous nonextensive solutions. Eur Phys Lett. 2002; 60:518–524. <http://stacks.iop.org/0295-5075/60/i=4/a=518>.
23. Skyline Ranking. Emporis. Available from: <http://www.emporis.com/statistics/skyline-ranking>.
24. Callen H.B. Thermodynamics and an introduction to thermostatistics. Wiley, Cambridge; 1993.
25. Robledo A. Unorthodox properties of critical clusters. Mol Phys. 2005; 103:3025–3030. <https://doi.org/10.1080/00268970500185989>.
26. Riquelme-Galván M.Robledo A. Dual characterization of critical fluctuations: Density functional theory & nonlinear dynamics close to a tangent bifurcation. Eur Phys J S T. 2017; 226:341–351. <https://doi.org/10.1140/epjst/e2016-60268-0>.