Annals of Physics 355 (2015) 322–337

Contents lists available at ScienceDirect

Annals of Physics

journal homepage: www.elsevier.com/locate/aop

Power-law and exponential rank distributions: A panoramic Gibbsian perspective

Iddo Eliazar

School of Chemistry, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Article history:
Received 11 July 2014
Accepted 11 February 2015
Available online 19 February 2015

Keywords:
Zipf's law
Power-law rank distributions
Exponential rank distributions
Size and log-size distributions
Entropy

ABSTRACT

Rank distributions are collections of positive sizes ordered either increasingly or decreasingly. Many decreasing rank distributions, formed by the collective collaboration of human actions, follow an inverse power-law relation between ranks and sizes. This remarkable empirical fact is termed Zipf's law, and one of its quintessential manifestations is the demography of human settlements — which exhibits a harmonic relation between ranks and sizes. In this paper we present a comprehensive statistical-physics analysis of rank distributions, establish that power-law and exponential rank distributions stand out as optimal in various entropy-based senses, and unveil the special role of the harmonic relation between ranks and sizes. Our results extend the contemporary entropy-maximization view of Zipf's law to a broader, panoramic, Gibbsian perspective of increasing and decreasing power-law and exponential rank distributions — of which Zipf's law is one out of four pillars.

1. Introduction

Zipf's law is perhaps the most striking empirical law emanating from collaborative human endeavors [1,2]. Observed across numerous fields of science, Zipf's law turns out to manifest, on a system-level, the collective actions of humans [3–7]. An illuminating example of Zipf's law is the formation of human settlements [8–10], which is described as follows.

Consider a country comprised of n settlements labeled $r = 1, 2, \ldots, n$, and ranked in a decreasing order. Namely, the largest city is labeled $r = 1$, the second largest city is labeled $r = 2$, and the smallest

E-mail address: eliazar@post.tau.ac.il.

http://dx.doi.org/10.1016/j.aop.2015.02.016

settlement is labeled $r = n$. Also, set S_r to denote the number of citizens living in the rth settlement. Clearly, the distribution of the country's population among its cities is the amalgamated outcome of highly complex processes — demographic, social, economic, environmental, etc. Nonetheless, the settlement sizes S_r turn out to be governed by the remarkable quantitative formula $S_r = b/r$ ($r = 1, 2, \ldots, n$), where b is a positive amplitude. Namely, the sizes of human settlements are inversely proportionate to their ranks [8–10]. This formula is no less than mindboggling — as it manifests a universal system-level structure of human societies: the citizens of countries miraculously self-organize so as to produce harmonic demographics.

Zipf's law is the generic term for an inverse power-law relation between sizes and ranks:

$$S_r = \frac{b}{r^{\beta}} \tag{1}$$

($r = 1, 2, \ldots, n$), where b is a positive amplitude, and where β is a positive exponent. In general, the sizes S_r may represent any collection of positive-valued quantities ranked from largest to smallest. The *harmonic Zipf law*, with exponent $\beta = 1$, is the quintessential example of Eq. (1). Astonishingly, Zipf's law holds for a plethora of quantities generated by collective human actions. Examples include word frequencies in large texts [11–13], publications of scientists [14–16], sizes of firms and of firm bankruptcies [17–19], and in-degrees and out-degrees of social-networks' nodes [20–22]. We note that although the power-law connection of Eq. (1) is named after Zipf, it was first discovered in demography by Auerbach [8], in linguistics by Estoup [11], and in scientific productivity by Lotka [14].

The statistical inference of Zipf's law is usually carried out via a log–log plot, of log-sizes $\ln(S_r)$ vs. log-ranks $\ln(r)$—which depicts a straight line with slope $-\beta$ and with intercept $\ln(b)$, i.e.

$$\ln(S_r) = \ln(b) - \beta \ln(r) \tag{2}$$

($r = 1, 2, \ldots, n$). The affine log–log plot of Eq. (2) is the graphical hallmark of Zipf's law. Upon the first encounter with real-world data following Zipf's law, people usually express the following response: the immediate reaction is an amazed "*wow!* ", and after a short while comes the bewildered "*why?*". And indeed, why do so many collaborative human endeavors result in quantities that are governed by Zipf's law? This intriguing question is a matter of vigorous scientific exploration, and familiar explanations include growth processes [10,18], preferential attachment [20–22], self-organized criticality [23,24], and entropy maximization [25–31]. In this paper we extend the entropy-maximization approach, and present a comprehensive *Gibbsian study* of *rank distributions*—collections of positive-valued quantities that are ordered either *decreasingly* (as in the case of Zipf's law) or *increasingly*.

We begin with a general transformation of rank distributions to their corresponding *size-distributions* and *log-size distributions* (Section 2), followed up by the specific examples of *power-law* and *exponential* rank distributions (Section 3). An entropy-based *optimization analysis* of the corresponding log-size distributions then establishes the power-law and exponential rank distributions as optimal in various senses (Sections 4 *and* 5). Moreover, an intimate relation between these optimal rank distributions and *diffusion processes* is unveiled (Section 6), followed up by a discussion of the results (Section 7). We conclude with addressing *composite rank distributions* (Section 8), and the *Poisson-process modeling* of size and log-size distributions (Section 9).

The results established in this paper extend the contemporary entropy-maximization view of Zipf's law to a broader, panoramic, *Gibbsian* perspective of increasing and decreasing power-law and exponential rank distributions — of which Zipf's law is one out of four pillars.

2. Setting

Underlying Zipf's law is a sequence of positive sizes ordered from largest to smallest: $S_1 \geq S_2 \geq \cdots \geq S_n$. Alternatively, the underlying sequence of positive sizes can be ordered from smallest to largest: $S_1 \leq S_2 \leq \cdots \leq S_n$. The first ordering yields a *decreasing rank distribution*, and the second ordering yields an *increasing rank distribution*. Extending Zipf's law we consider rank distributions that are governed by the general rank-size relation

$$S_r = \phi(r) \tag{3}$$