Stochastic model of Zipf's law and the universality of the power-law exponent

Ken Yamamoto

Department of Physics, Faculty of Science and Engineering, Chuo University, Kasuga 1-chome, Bunkyo, Tokyo, Japan

(Received 4 September 2013; revised manuscript received 21 November 2013; published 8 April 2014)

We propose a stochastic model of Zipf's law, namely a power-law relation between rank and size, and clarify as to why a specific value of its power-law exponent is quite universal. We focus on the successive total of a multiplicative stochastic process. By employing properties of a well-known stochastic process, we concisely show that the successive total follows a stationary power-law distribution, which is directly related to Zipf's law. The formula of the power-law exponent is also derived. Finally, we conclude that the universality of the rank-size exponent is brought about by symmetry between an increase and a decrease in the random growth rate.

DOI: 10.1103/PhysRevE.89.042115 PACS number(s): 02.50.−r, 89.75.Da, 05.40.−a

Power-law distributions are common in almost all fields of physics and other natural and social sciences [1,2]. Their origins, universality, and applications are addressed in statistical physics in terms of critical phenomena [3], chaos and fractals [4], and self-organized criticality [5]. The emergence of a power-law distribution implies that the average value does not represent typical behavior of the system, and very large events occur more frequently than they do in "classical" distributions such as Gaussian and exponential distributions. Because a power-law distribution is an outstanding feature, special nomenclature is often used in individual fields: Zipf's law in linguistics [6] (this is the main topic of this paper), Pareto's law in economics [7], the Gutenberg-Richter law in seismology [8], and the scale-free property in complex network analysis [9].

Zipf's law is originally a statistical relation in linguistics [6]. In many types of text, the number of times the rth most frequent word appears is inversely proportional to its rank r. A similar power-law relation $s \propto r^{-\alpha}$ between the size s and its rank r has been also observed very commonly in systems other than linguistic ones [10–12], and the rank-size rule is generally called Zipf's law. A remarkable point is that the power-law exponent α approaches almost unity in many systems: $\alpha = 1.08$ in word frequency of Shakespeare's plays [13], $\alpha = 0.83$ in populations of cities in Japan [14], and $\alpha = 0.96$ in popularity of opening moves in chess [15].

From a theoretical viewpoint, the emergence of power-law behavior of rank-size distribution has been explained by several mechanisms, such as positive feedback [16], random typewriting [17], coherent noise [18], intermittency [19], asymmetric mobility [20], and algorithmic entropy [21]. However, it remains unclear as to why the exponent $\alpha = 1$ is so universal in Zipf's law. Very few works have considered the particularity of the exponent $\alpha = 1$ [19–21].

The main purpose of this paper is to explain why $\alpha = 1$ is a special value in Zipf's rank-size distribution. We focus on a quantity given by a successive total of the value at each time. For instance, the total number of citations of a scientific paper is given by the sum of citations at each time, and the total number of occurrences of a word is the sum of the occurrence number over a number of chapters in the text. We give a simple stochastic process which models a successive-total quantity and prove that it produces a power-law distribution, for which the exponent is also analytically derived. We show that Zipf's law of $\alpha = 1$ occurs naturally from the model when it possesses a type of symmetry.

The model we study is given by the following discrete-time stochastic process for two variables x_t and S_t:

$$x_{t+1} = \mu_t x_t, \tag{1a}$$

$$S_{t+1} = S_t + x_t, \tag{1b}$$

with the initial conditions $x_0 = 1$ and $S_0 = 0$. μ_t is a positive random variable that represents the growth rate of x_t, and we assume for simplicity that $\mu_0, \mu_1, \ldots$ are distributed independently and identically. The variable x_t represents the value at time t, and S_t is the sum of x_t's up to time $t-1$. Indeed, we can apply the second equation iteratively to obtain $S_t = x_0 + x_1 + \cdots + x_{t-1}$. We will show Zipf's law for S_t. Figure 1 illustrates the model; x_t varies with t according to Eq. (1a), and S_t is approximately given by the area between the x_t curve and the t axis up to $t-1$.

We can regard the model (1) as simplified dynamics of word frequency. Let us focus on a certain word, and count its frequency in each chapter in a book. We assume that x_t denotes the number of times the target word appears in the tth chapter. In the simplest description, the positive feedback, the effect that the use of a word in a chapter increases its use in the next chapter, gives a proportional relation between x_t and x_{t+1} as in Eq. (1a). Then, the total count of the word over a number of chapters is expressed by S_t appearing in Eq. (1b).

Zipf's law is a statistical relation for the set of observed values, while our model (1) contains one key variable S_t. We compare Eq. (1) to Zipf's rank-size distribution as follows. We arrange a large number of statistical copies, each of which evolves independently according to Eq. (1). At a certain time t, we record the value of S_t of each sample, and make the rank-size distribution of them; the rank can be defined within the collection of S_t values. We can also make the cumulative distribution $P(S_t > s)$ from these samples. The cumulative distribution is the inverse function of the rank-size distribution [2], so the power-law form of the rank-size distribution $s \propto r^{-\alpha}$ implies the power-law cumulative distribution $P(S_t > s) \propto s^{-\beta}$ whose exponent satisfies $\beta = 1/\alpha$. Note that $\alpha = 1$ is clearly equivalent to $\beta = 1$.

Our model (1) is a type of a multiplicative stochastic process, which means that the evolution of x_t is given by a multiplication of the random growth rate μ_t. In particular, Eq. (1a) is called the Gibrat process [22], and it is well

 ©2014 American Physical Society

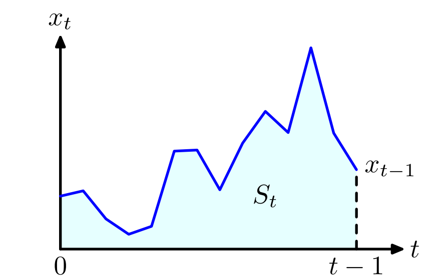

FIG. 1. (Color online) Illustration of our model. The x_t curve is given by the Gibrat process (1a), and the sum S_t is the area between the x_t curve and the t axis.

known that x_t follows a log-normal distribution for sufficiently large t [23]. A log-normal distribution is replaced by a power-law distribution when some additional conditions are considered along with the Gibrat process. For example, the introduction of additive noise [24], a reset event [25], and random stopping [26] have been reported to produce power-law distributions. We show below that the successive total (1b) is another mechanism with which a power-law distribution can be derived from the Gibrat process.

We present a numerical result first. In the calculation, we set μ_t as a uniform random variable on the interval [0.5,1.5] and computed cumulative distributions of S_t at $t = 10^3$ and 10^4 from 10^5 independent samples for each. As shown in Fig. 2, each of the two cumulative distributions has a clear power-law tail with an exponent $\beta = 1$: $P(S_t > s) \sim s^{-1}$ in large s. The two distributions almost perfectly overlap each other, and thus we expect that S_t reaches a stationary distribution after a long time. The existence of a stationary distribution and calculation of the power-law exponent are discussed later. According to the discussion below, the evaluation of the exponent $\beta = 1$ is justified by the fact $E(\mu_t) = 1$.

For the sake of the analysis of the model (1), we employ the stochastic process Z_t given by

$$Z_0 = 0, \quad Z_{t+1} = \tilde{\mu}_t Z_t + 1, \tag{2}$$

where $\tilde{\mu}_0, \tilde{\mu}_1, \dots$ are independently and identically distributed. In general, a multiplicative stochastic process with a random additive term is known as the Kesten process [27]; Eq. (2), which has a constant additive term, is a special case of it. We review some useful properties of the Kesten process as follows [24]. Z_t has a stationary power-law tail $P(Z_t > z) \sim z^{-\beta}$ in large t whose exponent β is characterized by

$$E\left(\tilde{\mu}_t^{\beta}\right) = 1. \tag{3}$$

Here $E(\cdot)$ denotes the average. The trivial solution $\beta = 0$ always exists, and a unique positive solution $\beta > 0$ exists if $E(\ln \tilde{\mu}_t) < 0$ and $\lim_{\beta' \to \infty} E(\tilde{\mu}_t^{\beta'}) = \infty$.

Now, we derive the distribution of S_t in Eq. (1) by making a comparison with the above Z_t. Equation (1) is written by the matrix form as

$$\begin{pmatrix} x_{t+1} \\ S_{t+1} \end{pmatrix} = \begin{pmatrix} \mu_t & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ S_t \end{pmatrix}. \tag{4}$$

That is, the matrix

$$M_t := \begin{pmatrix} \mu_t & 0 \\ 1 & 1 \end{pmatrix}, \tag{5}$$

which involves a random variable μ_t, acts as the time-evolution operator. By using the initial conditions $x_0 = 0$ and $S_0 = 1$, x_t and S_t are formally given by

$$\begin{pmatrix} x_t \\ S_t \end{pmatrix} = M_{t-1} M_{t-2} \cdots M_0 \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{6}$$

Multiplying the vector $(0 \quad 1)$ from the left, one obtains

$$S_t = (0 \quad 1) M_{t-1} M_{t-2} \cdots M_0 \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{7}$$

At the same time, we find that the Kesten process (2) is expressed using M_t as

$$(Z_{t+1} \quad 1) = (Z_t \quad 1) \begin{pmatrix} \tilde{\mu}_t & 0 \\ 1 & 1 \end{pmatrix} = (Z_t \quad 1) \tilde{M}_t. \tag{8}$$

The second component of the vector gives the trivial relation "$1 = 1$." As with Eq. (7), Z_t is written as

$$Z_t = (0 \quad 1) \tilde{M}_0 \tilde{M}_1 \cdots \tilde{M}_{t-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{9}$$

For a given t, we consider a time-reversal transformation $\mu_\tau \mapsto \tilde{\mu}_{t-1-\tau}$ $(\tau = 0, 1, \dots, t-1)$, which causes $M_0 \mapsto \tilde{M}_{t-1}$, $M_1 \mapsto \tilde{M}_{t-2}$, and so on. Because of the independence of $\mu_0, \dots, \mu_{t-1}$, S_t is properly mapped to Z_t. Note that the independence is an essential factor in this argument. In fact, if μ_τ's are not independent of each other, e.g, μ_0 affects μ_1, μ_1 affects μ_2, and so on, then $\tilde{\mu}_0$ is affected by $\tilde{\mu}_1$, $\tilde{\mu}_1$ is affected by $\tilde{\mu}_2$, and so on; time reversal gives rise to "breaking of causality" between S_t in Eq. (7) and Z_t in (8) in this case. Therefore, S_t is statistically equivalent to Z_t for any t, if μ_t's and $\tilde{\mu}_t$'s are distributed independently and identically. Useful properties of the Kesten process Z_t are directly transferred to our S_t.

In conclusion, S_t in Eq. (1b) has a stationary power-law tail $P(S_t > s) \sim s^{-\beta}$, with an exponent β that is a unique positive solution of

$$E\left(\mu_t^{\beta}\right) = 1. \tag{10}$$

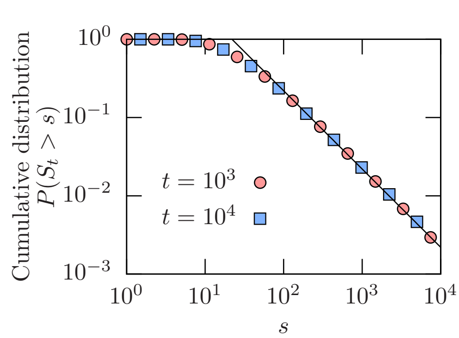

FIG. 2. (Color online) Cumulative distributions of S_t at $t = 10^3$ (circles) and 10^4 (squares), which are calculated from 10^5 independent numerical samples. The growth rate μ_t is a uniform random variable between 0.5 and 1.5. Circles and squares are alternately aligned in order to avoid overlapping. The straight line indicates a power law with exponent -1.

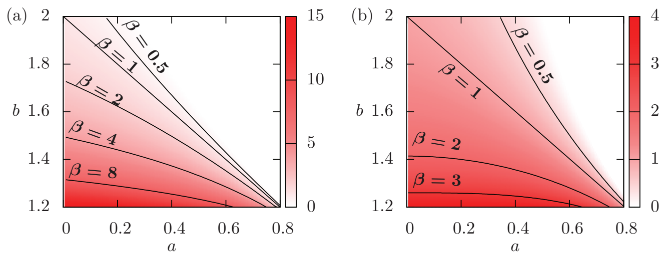

FIG. 3. (Color online) Numerical calculations of the exponent β, expressed by gradation of shading. (a) μ_t is a uniform random variable on the interval $[a,b]$ such that the probability density is $f(\xi) = 1/(b-a)$ for $a \leqslant \xi \leqslant b$. The contours of $\beta = 0.5$, 1, 2, 4, and 8 are shown. (b) μ_t can take either a or b with probability 1/2, such that $f(\xi) = [\delta(\xi - a) + \delta(\xi - b)]/2$. The contours of $\beta = 0.5$, 1, 2, and 3 are shown. White areas in both panels indicate that the distribution of S_t has no stationary power-law tail.

If the random variable μ_t has a probability density function f, this equation can be rewritten as

$$\int_0^\infty \xi^\beta f(\xi) d\xi = 1. \tag{11}$$

A positive solution β uniquely exists if $E(\ln \mu_t) < 0$ and $\lim_{\beta' \to \infty} E(\mu_t^{\beta'}) = \infty$.

In order to grasp behavior of the exponent β against the growth rate μ_t (or its probability density f), we take two types of simple f, which has two parameters a and b. The numerical results are shown in Fig. 3. The first case (a) is $f(\xi) = 1/(b-a)$ in $a \leqslant \xi \leqslant b$ and $f(\xi) = 0$ otherwise; that is, μ_t is a uniform random variable on the interval $[a,b]$. The value of β is shown to be a function of a and b. The second case (b) is $f(\xi) = [\delta(\xi - a) + \delta(\xi - b)]/2$, where $\delta(\cdot)$ is the Dirac δ function; that is, either $\mu_t = a$ or $\mu_t = b$ occurs with probability 1/2 each. The two f give the same average $E(\mu_t) = (a+b)/2$, but values of β in Figs. 3(a) and 3(b) greatly differ from each other. Qualitatively, we find that β tends to become large if a and b are small in both cases. We also observe that S_t has no stationary power-law tail if a and b are large, corresponding to $\beta = 0$ in the figure. Large a and b increase S_t rapidly, so the stationary distribution of S_t vanishes. From these observations, β tends to become large when $E(\mu_t)$ is small, and the power law vanishes when $E(\mu_t)$ is large.

We explain that our model leads Zipf's law of $\alpha = \beta = 1$ for a very wide class of μ_t. The *only* additional assumption is $E(\mu_t) = 1$, which means the symmetry between an increase and a decrease in x_t. In fact, x_t, on average, does not increase or decrease when $E(\mu_t) = 1$:

$$E(x_{t+1}) = E(\mu_t x_t) = E(\mu_t)E(x_t) = E(x_t). \tag{12}$$

With this condition, we immediately find that $\beta = 1$ is the positive solution of $E(\mu_t^\beta) = 1$. Therefore, Zipf's law of $\alpha = \beta = 1$ is immediately obtained. This result provides a piece of theoretical evidence for why the exponent $\beta = 1$ (or equivalently $\alpha = 1$) is special and universal. That is, the successive total S_t of the random variable x_t, whose dynamics are described by the Gibrat process, gives Zipf's law of $\alpha = 1$, if the growth rate μ_t exhibits the symmetry between an increase and a decrease as described above. We note again that the exponent $\alpha = 1$ is obtained irrespective of a choice of probability distribution of μ_t, provided that $E(\mu_t) = 1$ holds.

We have derived $\beta = 1$ from the assumption $E(\mu_t) = 1$, and the converse is also true; that is, $\beta = 1$ immediately implies $E(\mu_t) = 1$, via Eq. (10). Hence the symmetry $E(\mu_t) = 1$ is *equivalent* to Zipf's law of $\alpha = 1$.

We can estimate the exponent β even when the symmetry $E(\mu_t) = 1$ breaks slightly. We consider the case $E(\mu_t) = 1 - \epsilon$, where ϵ is a small parameter that can be positive or negative. It is reasonable to assume $\beta = 1 + \eta$, where η is a small correction term. Expanding Eq. (10) with respect to η as

$$\begin{aligned} 1 &= E(\mu_t^{1+\eta}) = E[\mu_t + \eta \mu_t \ln \mu_t + O(\eta^2)] \\ &= 1 - \epsilon + \eta E(\mu_t \ln \mu_t) + O(\eta^2), \end{aligned} \tag{13}$$

and neglecting $O(\eta^2)$ terms, we obtain

$$\eta = \frac{\epsilon}{E(\mu_t \ln \mu_t)}. \tag{14}$$

This is the correction formula of the exponent $\beta = 1 + \eta$. Thus $\beta \approx 1$ if ϵ is sufficiently small.

We comment on the importance of the multiplicative nature in Eq. (1a). For comparison, we look into the following "additive" model, similar to the multiplicative model (1):

$$x_{t+1} = x_t + \mu_t, \quad S_{t+1} = S_t + x_t. \tag{15}$$

This model is related to self-organized criticality [28], where S_t represents the size of an avalanche. If μ_t is symmetric about zero and has a finite variance, the variable x_t is essentially the same as a one-dimensional random walk, and S_t is the area under the walk. The cumulative distribution of S_t in this case also has a power-law tail, but the exponent is known to be $\beta = 1/3$ [29], which implies $\alpha = 3$. Thus Zipf's law of $\alpha = 1$ cannot be obtained from this additive model.

Our model (1) mathematically resembles the Kesten process (2). However, we stress here that their realizations in physics are different; Eq. (1) models a successive-total quantity to produce Zipf's law, but Eq. (2) cannot be directly connected with a successive total. Our focus in this paper is Zipf's law of successive-total variable; at the same time, we do not think that the mechanism in this paper provides a complete answer for the universality of Zipf's law. It is clear that not all quantities showing Zipf's law can be expressed as a successive total. We hope that this paper stimulates theoretical studies of Zipf's law on the basis of more general framework.

Let us give insight to the meaning of $E(\mu_t) = 1$ in the case of word frequency. As described above, x_t in Eq. (1) can be regarded as the number of appearance of a certain word in the tth chapter. If the size of each chapter is approximately constant, the increase in use of some words causes the decrease of some other words. Due to this constraint, the average growth rate is expected to become around unity. Therefore, we consider that the word count statistics can be approximately treated by the model proposed in this study.

Lastly, we discuss application of our model to real phenomena. A possible example is Twitter. It was reported that user influence based on the number of retweets follows Zipf's law [30], whose β is slightly less than unity. Meanwhile, modeling of information diffusion on Twitter has been proposed [31]. Information flow on the Twitter network forms nested shells according to the network distance from the source user. The total retweets $n_{\rm RT}$ is given by

$$n_{\rm RT} = \sum_{g=1}^{\infty} n_g, \tag{16}$$

where n_g is the number of retweets by users in gth shell. In turn, n_g approximately grows by the Gibrat process

$$n_g \simeq b_g \bar{k}_{g-1} n_{g-1}, \tag{17}$$

where $\bar{k}_{g-1}$ is the average number of followers of the retweeters in the $(g-1)$st shell, and b_g is a random number called the retweet rate. These equations for $n_{\rm RT}$ and n_g are strikingly similar to our model (1); $n_{\rm RT}$ is the successive total of n_g's. Although b_g's are not distributed independently and identically to be exact, our result probably contributes to the study of Twitter. We expect that our model connects Zipf's law of Twitter and statistical properties of b_g and $\bar{k}_g$.

The present work was supported by a Grant-in-Aid for Young Scientists (B) (25870743) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. The author is grateful to a referee, who has provided constructive comments.

[1] M. Buchanan, *Ubiquity: Why Catastrophes Happen* (Three Rivers Press, New York, 2000).
[2] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).
[3] D. Sornette, *Critical Phenomena in Natural Sciences*, 2nd ed. (Springer, Berlin, 2006).
[4] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science* (Springer, New York, 1992).
[5] P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Copernicus, New York, 1999).
[6] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA, 1949).
[7] V. Pareto, *Le Cours d'Économie Politique* (Macmillan, Paris, 1897).
[8] B. Gutenberg and C. F. Richter, *Seismicity of the Earth and Associated Phenomena*, 2nd ed. (Princeton University Press, Princeton, NJ, 1954).
[9] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[10] W. J. Reed and B. D. Hughes, Phys. Rev. E **66**, 067103 (2002).
[11] C. Furusawa and K. Kaneko, Phys. Rev. Lett. **90**, 088102 (2003).
[12] H. Yamada and K. Iguchi, Physica A **387**, 1628 (2008).
[13] M. A. Montemurro, Physica A **300**, 567 (2001).
[14] Y. Sasaki, H. Kuninaka, N. Kobayashi, and M. Matsushita, J. Phys. Soc. Jpn. **76**, 074801 (2007).
[15] B. Blasius and R. Tönjes, Phys. Rev. Lett. **103**, 218701 (2009).
[16] H. A. Simon, Biometrika **42**, 425 (1955).
[17] W. Li, IEEE Trans. Inf. Theory **38**, 1842 (1992).
[18] M. E. J. Newman and K. Sneppen, Phys. Rev. E **54**, 6226 (1996).
[19] D. H. Zanette and S. C. Manrubia, Phys. Rev. Lett. **79**, 523 (1997).
[20] K. Kawamura and N. Hatano, J. Phys. Soc. Jpn. **71**, 1211 (2002).
[21] B. Corominas-Murtra and R. V. Solé, Phys. Rev. E **82**, 011102 (2010).
[22] R. Gibrat, *Les Inégalités Économiques* (Librairie du Recueil Sirey, Paris, 1931).
[23] *Lognormal Distributions: Theory and Applications*, edited by E. L. Crow and K. Shimizu (Marcel Dekker, New York, 1988).
[24] H. Takayasu, A.-H. Sato, and M. Takayasu, Phys. Rev. Lett. **79**, 966 (1997).
[25] S. C. Manrubia and D. H. Zanette, Phys. Rev. E **59**, 4945 (1999).
[26] K. Yamamoto and Y. Yamazaki, Phys. Rev. E **85**, 011145 (2012).
[27] H. Kesten, Acta Math. **131**, 207 (1973).
[28] D. Dhar and R. Ramaswamy, Phys. Rev. Lett. **63**, 1659 (1989).
[29] M. J. Kearney and S. N. Majumdar, J. Phys. A **38**, 4097 (2005).
[30] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, in *Proceedings of the Fourth International Conference on Weblogs and Social Media* (The AAAI Press, Menlo Park, California, 2010), pp. 10–17.
[31] T. Kawamoto, Physica A **392**, 3470 (2013).