The principle of least effort and Zipf distribution

Yueying Zhu\textsuperscript{1,2}, Benwei Zhang\textsuperscript{2}, Qiuping A. Wang\textsuperscript{3,4}, Wei Li\textsuperscript{2,5}, Xu Cai\textsuperscript{2}

\textsuperscript{1} Research Centre of nonlinear Science, Wuhan Textile University, Wuhan, Hubei, 430073, China
\textsuperscript{2} Key Laboratory of Quark & Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan, 430079, China
\textsuperscript{3} IMMM, UMR CNRS 6283, Le Mans Université, 72085 Le Mans, France
\textsuperscript{4} School of Hautes Etudes d’Ingenieur, Yncrea, 59014 Lille, France
\textsuperscript{5} Max-Planck Institute for Mathematics in the Sciences, Inselst. 22, 04103 Leipzig, Germany.

E-mail: zhuyy@mails.ccnu.edu.cn

Abstract. "Each individual will adopt a course of action that will involve the expenditure of the probably least average of his work." This statement was named "the principle of least effort". The principle of least effort is often known as a "deterministic description of human behavior". In this paper, we present a brief introduction of this principle. Applications of the principle in different fields are also summarized. As the principle of least effort is proposed by Zipf, it is also called Zipf’s law. We then discuss the correlation between three widely considered distributions: Zipf distribution, Pareto distribution and probability distribution. With empirical investigations, it is often stated that, most social behaviors are controlled by the pure Zipf’s law that corresponds to the Zipf distribution of exponent -1. We summarily present the discovery of Zipf’s law in different social behaviors. Some empirical studies are also given as examples, verifying that, in most countries, the distribution of city size by population follows Zipf’s law, and the exponent of Zipf distribution of individual income is about -0.5, the same as Zipf predicted in theory.

1. Introduction

The principle of least effort was first discovered in 1894 by a French philosopher: Guillaume Ferrero. He discussed this principle in his article entitled "L’inertie mentale et la loi du moindre effort" [1]. However, until 1949, the principle was proposed by George Kingsley Zipf, an American professor of philology at Harvard University, in his book "Human Behavior and the Principle of Least Effort" [2]. Zipf theorised that the distribution of word use was due to tendency to communicate efficiently with least effort. Hence, the principle of least effort is also known as Zipf’s Law. Based on the principle of least effort, it is human nature to want the greatest outcome at the least amount of work. And Zipf showed that useful behaviors were performed frequently. Frequent behaviors became quicker and easier to perform over time. With this phenomenon, it is presented that people often chose their entire behavior along the direction of minimizing the effort. Basically, Zipf’s law describes people’s social behavior in space.

Zipf studied the least effort of individual behavior and of collective behavior, separately. Regarding individual behavior, he statistically analyzed, for example, the speech, words and their meaning, the verbalizations of children. For the collective behavior, he mainly focused on the studies of economy of human social behavior, such as the economy of geography, the distribution...
of economic power and social status, the distribution of prestige symbols and vogues. In both of these two parts, Zipf began with the empiric aspect, presenting a large number of observations from a truly wide range of living phenomena. And then, he gave the underlying theoretic analysis, attempting to rationalize different kinds of empiric laws in terms of a single uniform principle. Therefore, the principle of least effort is a theory which could be well understood from both empirical and theoretical views.

Actually, the topic of effort first aroused the interest of some experimentalists. In 1930, J. A. Gengerelli published some results of a series of experiments performed with blinded and normal white rats. These rats were used to help determine the nature of the path that animals would eventually select from an indefinite number of possible paths leading to food [3]. Experimental results suggested that in practically all cases that the path finally chosen by animals (both normal and blinded ones) was the path of "least effort," namely, the path of minimal distance. Two years later, L. S. Tsai stated that "Among several alternatives of behavior leading to equivalent satisfaction of some potent organic need, the animal, within the limits of its discriminative ability, tends finally to select that which involves the least expenditure of energy" [4]. And also, in 1937, R. H. Waters declared that "Thus, Theseus, after slaying the minotaur, found his way out of the labyrinth and to his loved one by following the string which he had carried with him into the labyrinth. Perhaps, this was not the most direct route in terms of distance, time, or effort, but it was the only sure way he had of escaping. Likewise our rats found that by sticking to the outside pathways they more readily achieved the goal" [5].

Since the principle of least effort was proposed by Zipf, many scholars are attracted to study it and to apply it to different fields. In the next section, we will discuss applications of the principle of least effort in different fields.

2. The principle of least effort in different fields

2.1. The field of information retrieval

The principle of least effort is exceptionally important in designing libraries and researching in the context of modern library. User’s desire to find information quickly and easily is often the primary consideration in a library design. In library literature, "least effort" is restated as Mooers’ law, an empirical observation of behavior provided by an American computer scientist Calvin Mooers in 1959 [6]. More commonly, Mooers’s law is considered to be a derivation of the principle of least effort. With Mooers’ law, it is stated that an information retrieval system will tend not to be used whenever it is more painful and troublesome for a user to have information than for him not to have it.

In 1987, T. Mann classified the principle of least effort as one of several principles controlling information seeking behavior [7]. He emphasized seven different research methods helping individuals get further into a subject more quickly, and with less wasted effort. Furthermore, E. G. Bierbaum also declared that "least effort" was one scholar’s suggestion for the uniform principle needed in research and practice, library and information science [8]. She stated "No other principle underlies as much of library and information science. Least effort explains the one-look-up reader, staff resistance to automation, the reliance of the scientist in colleagues rather than collections, and the rapid acceptance of CD-ROM compared to microfilm".

T. E. Chrzastowski, a professor of library administration, discussed in 1995 whether the workstations, aided by the principle of least effort, had changed the nature of how research was performed in academic libraries [9]. This discussion was proposed based on an investigation about library workstation popularity and the principle of least effort. With empirical data, she first analyzed the impacts of IBIS databases on the UIUC Chemistry library and then the journal use in the UIUC Chemistry library. Empirical results supported the principle of least effort and also showed that "least effort" method was often considered by many patrons as a suitable model for academic library research.
In 2004, the principle of least effort was further explored by Z. Liu and Z. Yang [10]. They provided a self-administered questionnaire to study the principal individual and environmental factors influencing a student’s decision process of selecting and using their information sources. Survey results say that reasons given by respondents for selecting and using primary information sources show their strong preference for fast and easy information retrieval. This phenomenon suggests that the principle of least effort also controls the respondent’s selection and use of information sources.

2.2. The field of human behavior

Generally speaking, interpretation of palaeoenvironmental changes is a key mission of researchers from archaeology, botany, geography and many other relevant disciplines. The principle of least effort also owns its specified scope of application in this field. Employing the principle, scholars A. Scholtz and M. L. Tusenius respectively provided a functional interpretation of charcoal data sets in 1986 [11, 12].

In 1992, C. M. Shackleton and F. Prins further discussed the applicability of the principle of least effort in explaining palaeoclimatic data [13]. They proposed a conceptual model to determine whether the principle of least effort was applicable to a given situation. In the conceptual model, four generalized areas used for inhabitant’s collection of fuelwood were described, occurring in chronological order. Area one regards the situation of plenty of fuelwood, both dead and live, with a considerable diversity of species available. Area two considers the situation of a medium abundance of dry wood, declining selectivity and maximum effort. Areas three and four take into account the conditions of low wood abundance, little selectivity, minimal effort and demand for wood far exceeding supply. Experimental results display that the principle of least effort is only applicable to the situation of area three. This conceptual model can not only help identify when the principle of least effort is appropriate for interpreting some data sets, but also be a valuable aid for understanding the past human behavior.

In 2002, R. F. Cancho and R. V. Sole illustrated a hypothesis of Zipf on the principle of least effort [14]. They aimed to provide new theoretical insights into the absence of intermediate stages between animal communication and language. Beginning with this idea and purpose, Cancho and Sole established a simple form of language game employing a mathematical model that involves a set of signals and objects. With the game, they studied the problem of compromise between speaker and hearer needs. Results strongly display that Zipf’s law is a hallmark of symbolic reference but not a meaningless feature.

2.3. The field of animal behavior

In animal ethology, an animal often wants to achieve the most energy with the lowest cost while foraging, so as to maximize the fitness. The optimal foraging theory is a model that helps predict the best strategy with which an animal can achieve the goal stated above. This theory is well known as the most essential theory helping predict how an animal behaves while collecting food. So, the optimal foraging theory can be considered as a derivative of the principle of least effort. The optimal foraging theory assumes that the most economically advantageous foraging pattern will be selected by a species through natural selection that has achieved optimal allocation of time and energy expenditures[15]. In 1984, A. Kacelnik found that starlings could maximize net energy gain per unit time [16]. J. R. Krebs and N. B. Davies also stated in 1989 that, with maximal energy efficiency, the bees were able to avoid expending too much energy per trip and to live long enough to maximize their lifetime productivity for their hive [17].

However, a long term assumption regarding livestock trails in grazing procedure said that livestock often established their pathways of least resistance between frequented portions of their pastures [18, 19]. This assumption was proposed by considering only a component of MacArthur and Pianka’s optimum foraging theory, stating that animals were expected to minimize energy
expenditures while maximizing energy gains in their daily endeavors. It had never been verified because of the difficulty in quantification of foraging activities and of optimal energy in practice. Until 2000, with the help of Geographic Information Systems (GIS) and Global Positioning Systems (GPS), Ganskopp et al. built a method to quantify the extent and characteristics of cattle trails in three large pastures, and indirectly verified three hypotheses [20]. They mapped cattle trails in three 800+ ha pastures containing global positioning units. Then they used GIS to quantify characteristics of both trails and the landscape, and also to plot least-effort pathways connecting water sources and distant points on selected trails in each pasture. Furthermore, the assumption that ”cattle develop least-effort routes of travel in rugged terrain by comparing the characteristics of cattle trails and least-effort pathways” was also testified.

3. Zipf distribution

3.1. The relationship between three kinds of distribution

In the field of empirical investigation, three widely considered distributions include the probability distribution, Pareto distribution, and Zipf distribution. Many man made and naturally occurring phenomena, such as city sizes, incomes, word frequencies, and earthquake magnitudes, are distributed according to a power-law distribution [21]. This power-law behavior is often presented by the probability distribution that describes the probability of occurrence of an event during the whole system. Zipf distribution focuses on the correlation between the frequency of occurrence of an event and the rank of the underlying event. Pareto distribution is a cumulative distribution, illustrating the probability that a person has an income no less than a given number. Though these three distributions are established in different aspects, they refer to the same thing. In this section, we will discuss the correlation between them. Specifically, we mainly focus on the derivation of both Pareto and probability distributions from a Zipf one.

We take the distribution of individual income as example. Zipf’s law states that the size of income for individual people is inversely proportional to the rank of this income in decreasing order. Mathematically, Zipf’s law is formulated as

\[ I_r \propto C_1 * r^{-1/2}, \]  

where \( I_r \) is the income of rank \( r \) in decreasing order, and \( C_1 \) the size of income with rank 1. For convenience, we label the exponent in Zipf distribution by a general variable \( \alpha \), that is

\[ I_r = C_1 * r^{-\alpha}. \]  

The \( r \)th income owns size \( C_1 * r^{-\alpha} \), meaning that there are \( r \) kinds of income with size no less than \( C_1 * r^{-\alpha} \). With this description, the probability that an income is no less than \( C_1 * r^{-\alpha} \) is proportional to \( r \), expressed as

\[ P(X \geq C_1 * r^{-\alpha}) = C_2 * r. \]

Setting \( x = C_1 * r^{-\alpha} \), we can get \( r = \left[ \frac{x}{C_1} \right]^{-\frac{1}{\alpha}} \). Pareto distribution regarding the same event is then represented as

\[ P(X \geq x) = C_2 * \left[ \frac{x}{C_1} \right]^{-\frac{1}{\alpha}}. \]

With Pareto distribution, we can also directly get the cumulative probability distribution, that is

\[ P_r(X < x) = 1 - P(X \geq x) = 1 - C_2 * \left[ \frac{x}{C_1} \right]^{-\frac{1}{\alpha}}. \]
Taking the first derivative of the above equation with respect to \( x \), we then have the probability distribution:

\[
P_r(X = x) = \frac{dP_r(X < x)}{dx} = \frac{C_2}{\alpha} * C_1^{\frac{1}{\alpha}} * x^{-(\frac{1}{\alpha}+1)}.
\]

According to the above discussion, it is noticed that a Zipf frequency-rank distribution of exponent \( \alpha \) corresponds to a Pareto cumulative distribution of exponent \( \frac{1}{\alpha} \) and a power-law probability distribution of exponent \( (\frac{1}{\alpha}+1) \). X. Gabaix and Y. M. Ioannides illustrated in 2004, by Monte Carlo simulations, that it was considered as a successful application of Zipf’s law when the exponent of Pareto distribution was between 0.8 and 1.2 [22].

Figure 1 presents a simple empirical investigation regarding both Zipf and Pareto distributions of American city size by population in 2013. It is illustrated that the exponent of Pareto distribution is about -1.366, quite close to the reciprocal of the exponent of Zipf distribution -0.823. This statistical phenomenon directly provides the empirical evidence for the correlation between exponents of Zipf and Pareto distributions.

3.2. Alternative expressions of Zipf’s law

Zipf’s law can be generalized by an approximation on the relationship between rank and frequency

\[
r^\alpha * f = C,
\]

where \( r \) is the rank of a word-type in decreasing order of frequency, \( f \) the frequency of occurrence of the corresponding word in a given text, and \( C \) a constant. \( C \), depending upon the underlying text, is often about one-tenth of the size of the text (the total number of running words). When \( \alpha = 1 \), Eq. (9) is well known as the pure form of Zipf’s law:

\[
f = C * r^{-1}.
\]

Pure Zipf’s law states that the size of the \( r \)th largest occurrence of an event is inversely proportional to its rank.

In 1954, B. Mandelbrot proposed a further refinement of Zipf’s law:

\[
f = C * (r + \rho)^{-B},
\]

Figure 1. Zipf and Pareto distributions of city size by population in United States for 2013.
in which \( r \) is the rank of a word, \( f \) the frequency, \( C, \rho \) and \( B \) the constants dependent upon the underlying text[23]. This expression forms the basis of statistical LNRE (large number of rare events) models and provides a better fit to low rank but high frequency words. H. P. Edmundson developed a new 3-parameter expression in 1972 considering the relationship between frequency and rank:

\[
f(r; c, b, a) = c (r + a)^{-b}, \quad c > 0, \quad b > 0, \quad a \geq 0, \quad (12)
\]

where \( a, b \) and \( c \) are constants[24].

Except for the distribution of words by frequency in a given text, the behavior of Zipf’s law is also exhibited in many other aspects: the distribution of city size by population (as presented above), the distribution of individual income, the distribution of scientist by the number of published papers, etc.

### 3.3. Zipf’s law in different aspects

#### 3.3.1. The distribution of city size

The distribution of city size by population has been a longstanding topic of interest since the last century. In 1913, F. Auerbach, a German physicist, launched an initial interest in the distribution of city size [25]. With empirical data, he found there was a universal relationship connecting the city’s population and city’s rank in United States and five European countries. This relationship can be denoted by,

\[
P_i \cdot r_i = A; \quad (13)
\]

where \( A \) is a constant, \( P_i \) the average population of cities in size-class \( i \), and \( r_i \) the rank of class \( i \) in the order of size decreasing.

A. J. Lotka, a US scholar, found empirically in 1925 a better fit function for the first 100 largest cities in United States in 1920 [26]. The function is expressed as

\[
P_i \cdot r_i^{0.93} = 5,000,000. \quad (14)
\]

In 1940, Zipf analyzed the rank-frequency distribution of the first 100 largest Metropolitan Districts in United States in 1940, according to the Sixteenth Census. He discovered that the slope of the underlying distribution was -0.98 (see Fig. 9-2 in Ref. [2]). Zipf’s law in the distribution of city size by population suggests that the city with the largest population in any country is generally twice as large as the next-biggest, and three times as large as the third biggest, and so on. Incredibly, Zipf’s law in city size distribution has always held true for most countries in the world.

Furthermore, K. T. Rosen and M. Resnick classically studied in 1980 the rank-size distribution of cities of 44 countries encompassing both developing and developed nations, using 1970 census data [27]. With their results, the first 50 largest cities in most countries can be predicted by a Pareto distribution of the form:

\[
R = AS^{-a}, \quad (15)
\]

where \( R \) is the number of cities with population \( S \) or more, \( A \) a constant, \( S \) the population of city, and \( a \) the Pareto exponent. For 44 countries under their investigation, exponent \( a \) ranges from 0.809 to 1.963 and has mean value 1.136. And for 32 of them, the Pareto exponent is greater than one. This discovery suggests that the population in most countries is more evenly distributed than that could be predicted by the pure Zipf’s law which corresponds to a Pareto distribution of exponent 1. Similar results were also provided by K. T. Soo in 2005 who assessed the empirical validity of Zipf’s law in the distribution of city size using new data of 73 countries and two estimation methods [28]. In 2007, he also performed a test of Zipf’s law in the rank-size distribution of cities, based on five years population censuses from Malaysian (1957, 1970, 1980, 1991 and 2000) [29]. Meanwhile, the factors that possibly influenced the growth of a city in
Malaysia were also explored. Soo found that Zipf’s law held for the size-rank distribution of cities in Malaysia in 1957. But since then, cities were more unequal in size than that would be predicted by Zipf’s law. Furthermore, he also stated that a city growth was negatively related to the city size. This statement is against Gibrat’s law, a rule defined by R. Gibrat, stating that the growth rate of a city is independent of the city size [30, 31].

In 2012, M. Cristelli et al. also found that Zipf’s law held approximately for the distribution of city size in each European country (France, Italy, Germany, etc). While for the aggregated data in European union, Zipf’s law completely failed. They declared that “In fact, historically, the geographic level for Europe, at which an integrated evolution is observed, is the national state, while in the US, the whole confederation, not each independent state, has collectively and organically evolved towards a distribution of cities that follows Zipf’s law. From this perspective, the US is an organic, integrated economic federation, while the EU has not yet become so, and shows little convergence to such an economic unit” [32]. This would seem to support the idea that Zipf’s law is a response to economic conditions, since it works only if you compare cities that are connected economically in a country.

With theoretical analysis and empirical studies, Y. Chen found that Zipf’s law was closely related to the hierarchical scaling law [33]. Beginning with a general form of Zipf’s law of exponent $q$, the author firstly defined a self-similar hierarchy of cities based on $q$-sequence. And then he deduced theoretically the hierarchical scaling law of cities, and observed that the exponent of rank-size scaling law was the reciprocal of that of size-number scaling law. Empirical data from both America and China showed the existence of Zipf’s law in the rank-size distribution of cities with exponent -0.738 and -0.889, respectively. The exponent of the number-size distribution was -1.364 for America and -1.193 for China.

In 2013, S. Li and D. Sui studied the rank-size distribution of China’s urban system based on empirical data in the period of 24 years: from 1984 to 2008 [34]. They displayed that the upper tails of rank-size distributions of Chinese cities followed a power-law function of slope slightly less than -1. This characteristic suggested that Chinese cities were more evenly distributed than that predicted by Zipf’s law. A similar trend was also found in many other countries [35].

J. Luckstead and S. Devadoss provided in 2014 a comparison and examination of size distributions of Chinese and India cities from 1950 to 2010, using log-normal, Pareto, and general Pareto distributions. Results provide that large number of cities from China and India have similar trends: the rank-size distribution of cities is log-normal in the early periods but power-law in 2010. Furthermore, in both 2000 and 2010, the distribution of city size in India is controlled by pure Zipf’s law, different from the situation in China [36].

An empirical investigation regarding the distribution of city size is also presented here, see Fig. 2(a – e). We mainly focus on the first 295 largest cities in United States in both 2010 and 2013, major cities in Germany of more than 100,000 inhabitants in 2010, major cities in France of more than 75,000 inhabitants in 1999, major cities in China of more than 155,540 inhabitants in 2010, and major cities in Japan of more than 202,283 inhabitants in 2010. Results reveal that Zipf distributions of the size (denoted by population) of cities under investigation all behave obviously power-law distribution:

$$ P \sim r^{-\alpha}, \quad (16) $$

$P$ labels the population of a city and $r$ the underlying city’s rank. For United States, Germany, France and Japan, the exponent $\alpha$ of the distribution of city size by population values in sequence -0.823/-0.827, -0.809, -0.996 and -0.846. These values are quite close to -1, suggesting the distribution of city size by population in countries mentioned above is approximately controlled by the pure Zipf’s law. In China, however, the exponent $\alpha$ is -0.590, diverging from the value -1 indicating the pure Zipf’s law. This phenomenon sates again that the distribution of population in China is more homogeneous than that predicted by pure Zipf law. This might be because China owns the largest population, compared to other countries. More similar distributions
from other countries could be found in Ref. [2].

3.3.2. The distribution of firm size  As early as 1949, Zipf illustrated the rank-frequency (the number of wage earners) distribution of manufactures in United States. With data collected in 1939 and manufactures ranked in the order of their decreasing number of wage earners, he showed the slope of the underlying distribution was -2/3 (Fig. 9-8 in Ref. [2]). Zipf also studied the rank-frequency distribution of corporation assets in United States, based on data dating from 1931 to 1936. Results exhibit that corporation assets act as a power-law function of corporation’s rank with exponent close to -1 [2].

It has often been observed that the upper tail of the size distribution of firms resembles a Pareto distribution. However, people does not know how to explain this kind of distribution by economic theory. In 1958, H. A. Simon and C. P. Bonini developed a stochastic model of the growth process of firms on the basis of two assumptions [37]. Simulation results provided by the model were quite consistent with empirical data. With their assumptions, Yule distribution would be the steady-state distribution of the same process. Regarding the distribution of firm size, Yule distribution in the upper tail can be approximated by the Pareto distribution: $f(s) \propto s^{-(\rho+1)}$ as $s \to \infty$. Simon and Bonini showed that parameter $\rho$ here was represented by

$$\rho = \frac{1}{1 - g/G},$$

in which, $G$ is the net growth during some specified periods of assets of all firms in an industry, and $g$ the part of net growth attributable to new firms (firms that have reached the minimum size during the period).

In 1995, with data of American firms from CompuStat, M. H. R. Stanely et al demonstrated using Zipf plot that the upper tail of the distribution of firm size was too thin, relative to the log normal, rather than too fat [38]. They reported that the distribution of American firm size could be approximated as a log-normal distribution. This discovery, however, could not explain the general property of firms as the empirical data they used were unrepresentative of overall population of American firms. Regarding this limitation, R. L. Axtell studied in 2001 the distribution of American firm size, with the entire population considered, based on the data combined from Census/SBA and CompuStat [39]. With "non-employee" firms (of size 0) neglected, he found the distribution of firm size by the number of employees followed Zipf’s law of slope -1.008. And the slope would be -0.963 if those firms of size 0 were included. Based on a large number of empirical results, researchers also stated that "the Zipf distribution of firm size is robust across varying definition of 'size', and so too it is quantitatively invariant over time" [40].

In view of the fact that most of researches only focus on the economy in developed countries, J. Zhang et al analyzed in 2009 the data of top 500 Chinese firms from 2002 to 2007 [41]. China is commonly regarded as the biggest developing country in the world. Her rapid growth in economy always attracts the interest of many researchers from different fields. With empirical investigation, Zhang et al revealed that the revenues and ranks of Chinese firms exactly obeyed the pure Zipf’s law, with exponent being -1 for each year under consideration. Furthermore, they offered an explanation of this characteristic using a simple economic model, namely AK model. The model is usually applied to describe how the revenue of a firm depends upon its capital and technology level by neglecting the impact of human resource [42]. AK model, however, can satisfactorily represent the formation process of firm size only in China. This is because human resource is more abundant and much cheaper for Chinese firms, compared to other aspects of consumption [43].
3.4. The distribution of individual income

The distribution of individual income has always been a central concern in economic theory and policy. We have discussed above the rank-frequency distribution of firm size which is analogous to the distribution of group income. Zipf indicated theoretically that the slope of the rank-frequency distribution of firm size was -1, in good agreement with empirical results.

However, for the rank-frequency distribution of individual income, the slope should equal -0.5 except the case of the settlement of liability claims against automobile insurance companies whose slope would often rise to 1. This value was predicted theoretically by Zipf in 1949 [44]. Based on individual income data samples from seven different countries including the United Kingdom, France, Finland, etc, he also illustrated some empirical results of the rank-frequency distribution of individual income. All these results exhibit a very good verification of the theoretical prediction of slope -0.5, independent of the country and the year that data samples come from.

But differently, H. T. Davis claimed as early as 1941 that the slope of Pareto distribution of individual income was -1.5, according to a very large amount of empirical data [45]. We have discussed before the reciprocal relationship between the exponent of Pareto cumulative distribution and that of Zipf rank-frequency distribution. Based on Davis’s results, the slope of rank-frequency distribution of individual income in double logarithmic coordinates was -2/3. Davis also argued that the slope would tend to fall in good times and to increase in bad times.

A simple empirical investigation is also given as an example here, regarding the Pareto distribution of individual income in United States for 2010. The underlying result is presented in Fig. 2(f), indicating the exponent of Pareto distribution is close to -2. This suggests the slope of Zipf rank-frequency distribution in double logarithmic coordinates is about -0.5, the same as Zipf predicted in theory.

4. Conclusions

We first present a brief introduction of the principle of least effort. Applications of this principle in the field of information retrieval, human behavior and animal behavior are also summarized. Furthermore, we also discuss the correlation between three widely used distributions: the Zipf distribution, Pareto distribution and probability distribution. Results suggest that both Pareto and probability distributions can be derived from Zipf distribution. A Zipf distribution of exponent $\alpha$ is analogous to the Pareto distribution of exponent $\frac{1}{\alpha}$, and the probability distribution of exponent $(1 + \frac{1}{\alpha})$. Zipf distribution considers the correlation between the frequency of occurrence of an event and the rank of the underlying event. With empirical investigations, it is widely stated that Zipf distribution acts as power-law decay function in most social behaviors. And the corresponding exponent is always close to -1, namely the pure Zipf’s law. Regarding individual income, however, the exponent of Zipf distribution is close to -0.5, predicted in theory and also verified by empirical investigation by Zipf. We summary discuss the discovery of Zipf’s law in different aspects. Some empirical investigations are also provided as examples in this paper. Results verify again that, in different countries, the distribution of city size by population is controlled by Zipf’s law of slope close to -1. The slope of Zipf distribution of individual income is about -0.5, the same as Zipf predicted in theory.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China under Grant Nos. 11435004 and 11521064, and by the Programme of Introducing Talents of Discipline to Universities under Grant No. B08033.

References

[1] Ferrero G 1894 Revue Philosophique de la France et de l’Étranger 37 169C182
Figure 2. The Zipf distribution of city size by population (a – e) and Pareto distribution of individual income (f). a) Results for American cities. Black and red symbols indicate the data of April 1, 2010 enumerated by the 2010 United States Census and of July 1, 2013 estimated by the United States Census Bureau, respectively. b) Results for German cities of more than 100,000 inhabitants in 2010. c) Results for major cities in France of more than 75,000 inhabitants in 1999. d) Results for major cities in China of more than 155,540 inhabitants in 2010. e) Results for major cities in Japan of more than 202,283 inhabitants in 2010. f) The Pareto distribution of individual income in United States in 2010 (line+black squares). Line+red circles has an ideal slope of -2.
[2] Zipf G K 1949 Human behavior and the principle of least effort: An introduction to human ecology (Cambridge: Addison-Wes-ley)

[3] Gengeregli J A 1930 J. Comp. Psychol. 11 193-236

[4] Tsai T S 1932 The laws of Minimum Effort and Maximum satisfaction in Animal Behavior Monograph of the National Institute of Psychology

[5] Waters R H 1937 J. Gen. Psychol. 16 3-20

[6] Mooers C N 1959 Bull. Am. Soci. Inf. Sci. & Tech. 23 22-23

[7] Mann T 1987 A guide to library research methods (Oxford: Oxford University Press)

[8] Bierbaum E G 1990 American Libraries 21 18-19

[9] Chrzastowski T E 1995 J. Am. Soc. Inf. Sci. 46 638-641

[10] Scholtz A 1986 Palynological and palaeobotanical studies in the southern Cape M. A. thesis (Stellenbosch: University of Stellenbosch)

[11] Shackleton C M and Prinsh F 1992 J. Archaeolog. Sci. 19 631-637

[12] Adamic L A 2000 Zipf, Power-laws, and Pareto-A ranking tutorial http://ginger. hpl. hp.com/shl/papers/rankin-g/ranking.html.

[13] Cristelli M, Batty M and Pietronero L 2012 Scientific Reports 2 812

[14] Chen Y 2012 Physica A 391 3285-99

[15] Li S and Sui D 2013 GeoJournal 78 615-25

[16] Carroll G R 1982 Progress in Human geography 6 1-43

[17] Luckstead J and Devadoss S 2014 Econ. Lett. 124 290-95

[18] Simon H A and Bonini C P 1958 Am. Econ. Rev. 48 607-17

[19] Stanely M H R et al. 1995 Econ. Lett. 49 453-57

[20] Simon H A and Bonini C P 1958 Am. Econ. Rev. 48 607-17

[21] Axtell R L 2001 Science 293 1818-20

[22] Ijiri Y and Simon H A 1935 J. Roy. Stat. Soc. 140 219-23

[23] Zhang J, Chen Q and Wang Y 2009 Physica A 388 2020-24

[24] Romer P 1986 J. Politi. Econ. 94 1002-37

[25] Su Y and Xu X 2002 Econ. Res. J. 11 3-11

[26] Zipf G K 1949 Human behavior and the principle of least effort (Addison-Wesley Press) 496-500

[27] Davis H T 1941 The Theory of Econometrics (Bloomington: The Principia press)