Scaling Behavior of Chinese City Size Distribution

Xiaowu Zhu[1,2], Aimin Xiong[3], Liangsheng Li[1], Maoxin Liu[1], and Xiaosong Chen[1]

[1] Institute of Theoretical Physics, Chinese Academy of Sciences, P.O. Box 2735, Beijing 100080, P.R. of China
[2] Business School, China University of Political Science And Law, No.27 FuXue Road, Beijing 102249, P.R. of China
[3] Department of Systems Science, School of Management, Beijing Normal University, Beijing 100875, P.R. of China
{zxw,xiong,liliangsheng,mxliu,chenxs}@itp.ac.cn

Abstract. We have investigated the population distribution of Chinese cities from 1997 to 2006. The rank-size distributions of Chinese cities deviate from the Pareto distribution. For city size distribution of each year we can find a population threshold P_c that characterizes the boundary of the deviation. The cities with population more than P_c follow the Pareto distribution, while the smaller cities deviate from the Pareto distribution. Using P_c for every year, the rank-size distribution from 1997 to 2006 can be written into a scaling form $R(P,T) = C(T)P^{-\alpha(T)}f(P/P_c(T))$, where the Pareto exponent $\alpha(T)$ is not equal to the value of Zipf's law and evolutes with time. According this scaling form, the data of the city size distributions of Chinese cities from 1997 to 2006 can collapses to a single curve, which is the scaling function of the city size distribution.

Keywords: city size distribution, Pareto distribution, scaling, Zipf's law.

1 Introduction

Over the last 30 years, China has experienced rapid urbanization. From 1978 to 2006 the Chinese national population increased from 962.59 million in 1978 to 1.314 billion in 2006, a 37 percent increase. The non-agricultural population in urban areas had a 144 percent increase, which is much greater than the national population growth rate. As a result, China has become more urbanized, with an urban population share that has increased from 17 percent in 1978 to 44 percent in 2006 [1]. The urban population share of China now is still much lower than the 70 percent of developed nations, but the Chinese urbanization speed is almost the three times of the world average. With the large number of cities and very active dynamics of urban organization, China's urban system and urbanization provide an important area for research.

The most economic activities of human being are happened in cities. One of the most striking regularities in the location of economic activity is the size

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 868–875, 2009.

distribution of cities in a country [2]. The empirical study of city size distribution has engaged scientists and economists since the beginning of last century [3]. In 1913 the German geographer Felix Auerbach found an interesting empirical regularity that the product of the population size of a city and its rank in the city size distribution appears to be roughly constant for a country. Since Auerbach proposed this basic proposition, it has been widely accepted by scholars in a variety of disciplines and refined by others. In 1941 Zipf [4] provided an empirical analysis which suggested that the city size distribution can be represented by a Pareto distribution with an exponent equal to 1.

Recently, the city size distribution has attracted renewed attention. The researches of Rosen&Resnick [5] in 1980 and others recently [2] have shown, using empirical analysis of the data from many countries, that the Zipf's law is not always tenable. The city size distribution as a Pareto distribution has an exponent seldom equal to one. Even some doubts were raised regarding the validity of the Pareto distribution. In 1990 Husing [6] argued that not only Zipf's law, but also the Pareto distribution as well were not supported by the empirical data. In 1998 Laherrere and Sornette [7] suggested the use of a stretched exponential distribution. Later it was thought that the Zipf's law is spurious in explaining city size distribution [8]. In 2007 Benguigui and Blumenfeld-Lieberthal [9] went beyond the power law and proposed a new approach to analyze city size distributions. More recently, it was suggested that cities are complex systems that mainly grow from the bottom up [10]. Like a physical system in nature, the size et al of cities follow scaling laws. As a manifestation of the underlying dynamics and structure, scaling is well known in physics. It has been instrumental in understanding problems across the entire spectrum of science [11]. Typically scaling laws reflect generic features of the systems which consist of a large number of interacting particles [12]. They are universal and independent of the microscopic details of systems. The scaling hypothesis has been well verified by a wealth of experimental data on diverse systems [13]. One prediction of the scaling hypothesis is the scaling law which relates the various critical point exponents characterizing the singular behavior of functions such as thermodynamic functions. Another prediction of the scaling hypothesis is a sort of data collapse, where diverse data collapse onto a single curve called a scaling function under appropriate axis normalization.

In this paper, we analyze the city size distributions of Chinese cities from 1997 to 2006. The large number of cities in China make us be able to check if the sizes of Chinese cities follow the Pareto distribution. The rapid changes of Chinese cities from 1997 to 2006 make us be able to investigate the scaling behavior of the city size distributions. The rest of our paper is organized as follows. In Section 2, we describe the date of Chinese cities. In section 3, we analyze the data of Chinese cities and propose a scaling for the city size distributions of China. With the scaling form the city size distributions from 1997 to 2006 can collapse onto a single curve, which is the scaling function of city size distributions. In Section 4 we give some conclusions.

2 Data of Chinese Cities

The definition of city in China is not straightforward. According to administrative level, Chinese cities are classified into three different groups: county-level cities (XianJiShi), prefecture-level cities (DiJiShi) and province-level cities (ShengJiShi). The small settlements with townships or lower administrative levels are not called as city. The administrative criteria distinguishing cities are the scale of urban population, the economic and political importance of an urban agglomeration [14]. The counties or the county-level cities are administrated by prefecture-level cities. The economic situations play the principal roles in the dynamics of city population [15]. The dynamic movement of population in China has resulted in that many counties had been reclassified as cities. The total number of county-level cities increased sharply from 223 in 1980 to 656 in 2006. The prefecture-level cities increase from 78 in 1978 to 283 in 2006. Even though prefecture-level and province-level cities share only about 30 percent of the national population, but they contribute the main part of economics in the whole nation.

When we look up the Chinese Urban Statistic Year Book, only the data of prefecture-level and province-level cities are provided year by year from 1997 to 2006. There are three categories of population in the Chinese Urban Statistic Year Book: the population of whole city including all rural and urban population, the urban population who live in the urban region, the non-agriculture population of the urban population who do non-agricultural work. In this article, we study the city size distribution of the urban population in prefecture-level and province-level cities from 1997 to 2006. We discuss the exponent of Pareto distribution and the scaling behavior of the city size distributions. The urban population can characterize really the size of a city. Other scholars have taken the data of Chinese cities for discussions similarly [14,16].

3 Data Analysis

3.1 City Size Distribution Function

Auerbach(1913), Singer(1936) and Zipf(1949) demonstrated that the city size distributions could be represented by a Pareto distribution

$$R = CP^{-\alpha}, \tag{1}$$

where P is the population of a city and R is its rank which is ordered from the largest to the smallest. α is the exponent of Pareto distribution and C is a constant.

In Fig.1a, the rank-size distribution of Chinese cities in 2006 is shown. There is a threshold of population P_c in the city size distribution. The cities with population $P > P_c$ follow the Pareto distribution, but the cities with population $P < P_c$ deviate from the Pareto distribution. The rank-size distributions of Chinese cities in other years are similar to the year 2006. We can conclude that

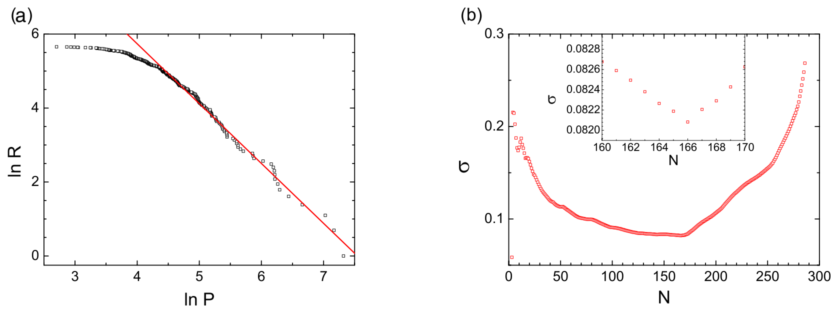

Fig. 1. (a) The rank-size distribution of 2006. (b) The mean square deviation versus number of sample N.

the city size distributions of Chinese cities do not follow the Pareto distribution as a whole. The small cities have obvious deviation. To describe the city size distributions of most Chinese cities in different years, a modified form of the usual Pareto distribution (1) is introduced as

$$R(P,T) = C(T)P^{-\alpha(T)}f(P,T) \tag{2}$$

Where T denotes the year. The function $f(P,T)$ of population and year characterizes the modification to the Pareto distribution. There is a Pareto distribution when $f(P,T) = 1$ and there is deviation when $f(P,T) \neq 1$.

We can rewrite Eq.(2) in a double-logarithmic representation as

$$\ln R(P,T) = -\alpha(T)\ln P + M(P,T) + \ln C(T)\ , \tag{3}$$

where $M(P,T)$ is defined as

$$M(P,T) \equiv \ln f(P,T) = \ln R(P,T) + \alpha(T)\ln P - \ln C(T). \tag{4}$$

The function $M(P,T)$ characterizes precisely the deviation of city size distribution $R(P,T)$ from the Pareto distribution.

3.2 Scaling Behavior of Cities Size Distribution

To calculate the function $M(P,T)$, we need to determine the exponent $\alpha(T)$ from the rank-size distribution of the cities with population $P > P_c$ at first. We estimate P_c by using OLS method. We sort the cities with the order from larger to smaller city. With suitable number of cities that $N > 10$, we can fit the rank-size distribution by a Pareto distribution. Of course, the real data have deviations from a Pareto distribution. Using the OLS method we can calculate the mean square deviation σ between the data and the Pareto distribution with an exponent α which fits the data at best. In Fig. 1b the mean square deviation σ is shown as a function of N. σ has a minimum of σ at the number of cities

Table 1. Threshold of population and exponent of Pareto distribution from 1997 to 2006

year	N	N_c	P_c	α	$\ln C$
1997	223	99	91.15	1.79148	12.60805
1998	229	125	68.33	1.62408	11.80417
1999	236	130	67.6	1.65552	11.96679
2000	262	144	67.14	1.68128	12.1731
2001	266	152	70	1.66078	12.16288
2002	278	154	74.53	1.64056	12.170393
2003	284	162	72.15	1.62331	12.1327
2004	285	166	74.53	1.63023	12.20675
2005	286	167	75.03	1.5887	12.04478
2006	286	166	77.36	1.61828	12.21067

$N = N_c$. When $N > N_c$, σ increases with the number of cities. So we define the population of N_c-th city as the threshold of population P_c. For the city size distribution of 2006, we find that $N_c = 166$ (see the sub-figure in Fig. 1b). In the same way, we can obtain N_c, P_c, α and $\ln C$ of each year from 1997 to 2005. All results are given in Table 1.

With the parameters $\alpha(T)$ and $C(T)$ given in the Table 1, we can calculate the deviation function $M(P,T)$ by using Eq.4 for each year. For $P < P_c$ of each year, $M(P,T)$ is approximately equal to zero. $M(P,T)$ at $P < P_c$ from 1997 to 2006 is shown in Fig.2.

In physics, there is scaling when a system is near its critical point. As a function of temperature and external field, the correlation length and thermodynamic quantities can be written in a scaling form which is the product of a power law term and an one-variable scaling function. The variable of the scaling function is the ratio of a physical quantity to its characteristic quantity. Inspired by the scaling in physics, we introduce a scaling form for the evolution of city size distribution in China as

$$R(P,T) = C(T)P^{-\alpha(T)} f_s\left(P/P_c(T)\right). \tag{5}$$

Here we define the ratio of population to the threshold P_c as the scaling variable. Correspondingly we can define the scaling form of the deviation function $M(P,T)$ as

$$M_s(P/P_c) = \ln f_s(P/P_c). \tag{6}$$

The validity of our scaling form Eq.(5) for the evolution of city size distribution should be tested by the data of cities. In Fig.2a, it has been shown that the deviation function $M(P,T)$ of Chinese cities depends really on population P and time T. If $M(P,T)$ has a scaling form, the different curves in Fig.2a could be collapsed onto a single curve after using the scaling variable $P/P_c(T)$. So the two-variable function $M(P,T)$ becomes a single-variable function $M_s\left(P/P_c(T)\right)$. In Fig.2b, it is shown that the data of Chinese cities from 1997 to 2006 collapse

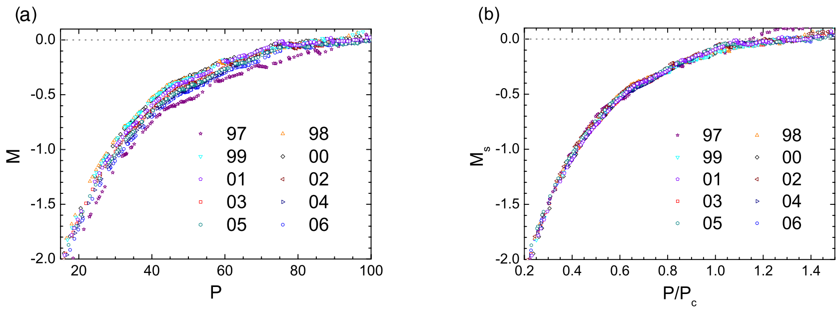

Fig. 2. The deviation function $M(P,T)$ at $P < P_c$ from 1997 to 2006: (a)with population P as variable,(b) with the ratio $P/P_c(T)$ as variable

quite well onto a single curve. We can conclude that the evolution of city size distributions in China from 1997 to 2006 satisfies the scaling form Eq.(5). In this case, the evolution of city size distribution can be described by the evolution of the exponent $\alpha(T)$ and the population threshold $P_c(T)$. The scaling functions in the critical phenomena are universal. Here we could suggest that the scaling functions $f_s(x)$ and $M_s(x)$ here have also universality which depends on the macroscopic properties of a country, but not the microscopic details.

4 Conclusions

Nowadays cities play more and more important role in the world. It is estimated that sometime in 2007 more people are living in cities than outside them, for the first time in history. Cities are undeniably centers of politics, economics and innovation. At the same time, they are also centers for the production of waste, pollution and heat. It is of essential interests for mankind to find out the regularities of organization and evolution of cities. Among them, the regularity of city size distribution is a relatively simple and important one. Even this problem has been investigated for nearly one century and many progresses have been obtained, there are still many fundamental questions which are open.

In this paper, we investigate the rank-size distribution of Chinese cities from 1997 to 2006. In these rank-size distributions cities can be divided into two parts. The cities with population more than a threshold P_c follow the Pareto distribution. The exponent of the Pareto distribution α varies with time from the smallest value 1.58 in 2005 to the largest value 1.79 in 1997. They are obviously larger than $\alpha = 1$ of the Zipf's law. The cities with population less than the threshold P_c deviate from the Pareto distribution. So the rank-size distributions of Chinese cities cannot be described by a Pareto distribution as a whole.

In physics, thermodynamic quantities can follow scaling laws when the correlation length of a system becomes very large. In the societies today, the correlations between cities and people are also very large due to modern technologies in communication and transportation. So it is reasonable for us to suggest scaling law for city size distributions. With respect to the evolution of rank-size distributions of Chinese cities from 1997 to 2006, we introduce a scaling form in Eq. 5, which is the product of a Pareto distribution and a scaling function with the scaling variable defined by the ratio of population to the threshold P_c. With this scaling form we can describe the evolution of city size distribution by the evolution of Pareto exponent $\alpha(T)$, population threshold $P_c(T)$ and a constant $C(T)$. The data of Chinese cities from 1997 to 2007 have been shown to collapse onto a single curve and have confirmed the scaling form.

We plan to use the scaling form to investigate the rank-size distribution of other countries in the future. We will check if there is also such scaling form for cities of other countries and investigate the relationship between the scaling functions of different countries if they exist. Taking cities as complex systems that grow from bottom up, many ideas and methods in physics, especially statistical physics, can be learned for the researches of phenomena in societies.

Acknowledgement

The project was supported by the National Natural Science Foundation of China under Grant 10325418.

References

1. National Bureau of Statistics: China Urban Statistical Yearbook. China Statistics Press, Beijing (1998-2007)
2. Soo, K.T.: Zipf's Law for cities: a cross-country investigation. Regional Science and Urban Economics 35(3), 239–263 (2005)
3. Dobkins, L.H., Ioannides, Y.M.: Dynamic Evolution of the U.S. City Size Distribution. Discussion Papers Series, Department of Economics, Tufts University (1999)
4. Zipf, G.K.: National Unity and Disunity. The Principia Press, Bloomington Indiana (1941)
5. Rosen, K.T., Resnick, M.: The Size Distribution of Cities: An Examination of the Pareto Law and Primacy. Journal of Urban Economics 8(2), 165–186 (1980)
6. Husing, Y.: A note on functional forms and the urban size distribution. Journal of Urban Economics 27, 70–73 (1990)
7. Laherrere, J., Sornette, D.: Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. The European Physical Journal B 2, 525–539 (1998)
8. Gan, L., Li, D., Song, S.: Is the Zipf law spurious in explaining city-size distributions? Economics Letters 92(2), 256–262 (2006)
9. Benguigui, L., Blumenfeld-Lieberthal, E.: Beyond the power law - a new approach to analyze city size distributions. Computers, Environment and Urban Systems 31, 648–666 (2007)
10. Batty, M.: The Size, Scale, and Shape of Cities. Science 319, 769–771 (2008)
11. Bettencourt, L.M.A., Lobo, J., Helbing, D., Kuhnert, C., West, G.B.: Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Sciences of the United States of America 104(17), 7301–7306 (2007)

12. Isalgue, A., Coch, H., Serra, R.: Scaling laws and the modern city. Physica A: Statistical Mechanics and its Applications 382(2), 643–649 (2007)
13. Privman, V., Hohencerg, P.C., Aharony, A.: Universal Critical-Point Amplitude Relations. In: Domb, C., Lebowitz, J.L. (eds.) Phase Transitions and Critical Phenomena, vol. 14, p. 1. Academic Press, New York (1991)
14. Song, S., Zhang, K.H.: Urbanisation and City Size Distribution in China. Urban Studies 39(12), 2317–2327 (2002)
15. Eeckhout, J.: Gibrat's Law for (All) Cities. The American Economic Review 94(5), 1429–1451 (2004)
16. Anderson, G., Ge, Y.: The size distribution of Chinese cities. Regional Science and Urban Economics 35(6), 756–776 (2005)