q-exponential distribution in urban agglomeration

L. C. Malacarne and R. S. Mendes
Departamento de Física, Universidade Estadual de Maringá, Avenida Colombo 5790, 87020-900, Maringá-PR, Brazil

E. K. Lenzi
Centro Brasileiro de Pesquisas Físicas, R. Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro-RJ, Brazil

(Received 10 January 2001; published 21 December 2001)

Usually, the studies of distributions of city populations have been reduced to power laws. In such analyses, a common practice is to consider cities with more than one hundred thousand inhabitants. Here, we argue that the distribution of cities for all ranges of populations can be well described by using a q-exponential distribution. This function, which reproduces the Zipf-Mandelbrot law, is related to the generalized nonextensive statistical mechanics and satisfies an anomalous decay equation.

DOI: 10.1103/PhysRevE.65.017106 PACS number(s): 89.90.+n, 89.65.−s, 05.20.−y

In several areas in nature, besides the complexities, it is possible to identify macroscopic regularities that can be well described by simple laws. For example, frequency of words in a long text [1], forest fires [2], distribution of species lifetimes for North American breeding bird populations [3], scientific citations [4,5], World Wide Web surfing [6], ecology [7], solar flares [8], football goal distribution [9], economic index [10], and epidemics in isolated populations [11] among others.

In particular, recently, interest in the study of city population distribution has increased. Such interest is related to the analysis of data and to models that present asymptotic power-law behavior [12–16]. However, in such analyses, only cities with more than one hundred thousand inhabitants have been considered. This power-law behavior may be identified in terms of the distribution

$$N(x)dx \propto x^{-\alpha}dx, \tag{1}$$

which gives the number of cities with x and $x+dx$ inhabitants, where α is a positive constant. Another way to express the same relation is in terms of the relative number (rank or cumulative distribution) of cities with a population larger than a certain value x,

$$r(x)=\int_x^\infty N(y)dy \propto x^{1-\alpha}. \tag{2}$$

By expressing the population $x(n)$ of the cities in descending order [$x(1)$ being the city with the highest population, $x(2)$ the city with the second-highest population, and so on], it follows from Eq. (2) that

$$x(n) \propto n^{1/(1-\alpha)}. \tag{3}$$

The plot of $x(n)$ on a double logarithmic scale is called a "Zipf plot" [1] and leads to a straight line with slope $1/(1-\alpha)$. Note that the Zipf plot [from Eq. (3)] and the cumulative plot [from Eq. (2)] are equivalent, except when regarding the weight related to the rare (largest) elements.

The Zipf plot for cities with more than one hundred thousand inhabitants [17] for some countries and Europe is illustrated in Fig. 1(a). These graphics enable us to visualize how good the power law is at describing the population distribution for large cities. In the inset plot of Fig. 1(a), we show the cumulative plot for the same cities in Europe. However, the fraction of cities with more than a hundred thousand inhabitants is small. For instance, these cities represent about 15%

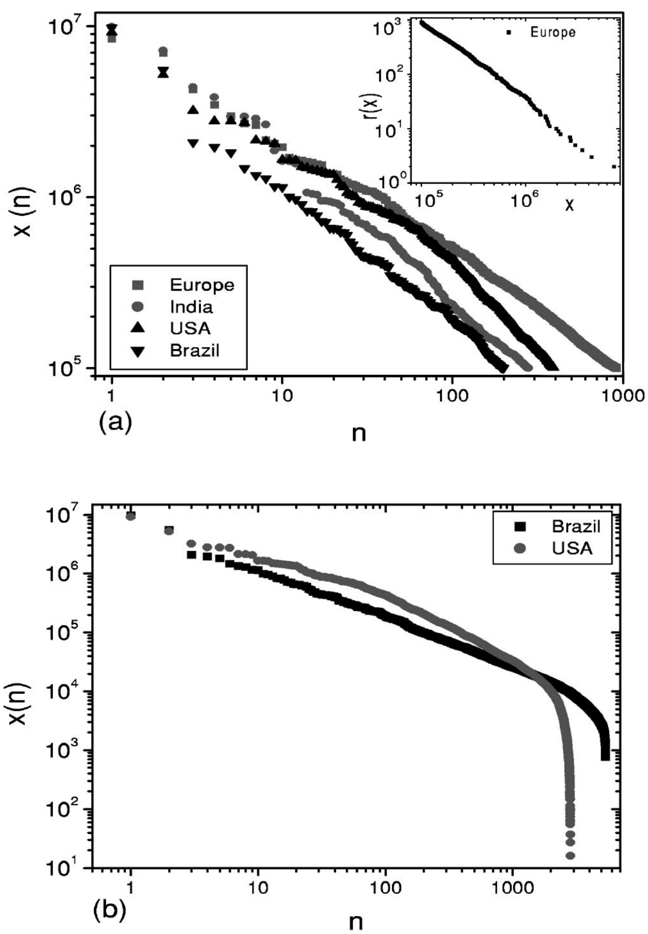

FIG. 1. (a) Zipf plot for cities with populations bigger than 100 000 and, in inset plot, the cumulative Zipf plot to the same cities in Europe. (b) Zipf plot for all cities in the U.S.A. and Brazil. In the above graphics, x is the population of the cities, n is the descending rank, and r is the cumulative rank.

of American cities and 4% of Brazilian cities. Furthermore, if we take into account all cities [18,19] in the country, and by using the Zipf plot in Fig. 1(b), we can identify a notorious deviation from the asymptotic power law when cities with small populations are considered. Thus, an analysis that considers all cities is an important task. In this direction, this paper is dedicated to an empirical analysis of this question.

An alternative approach to incorporate the deviation from power law is employed in Ref. [20] by considering the stretched distribution (Weibull distribution), $N(x) = N_0 x^{c-1}\exp(-\lambda x^c)$, to fit data of some complex systems. In particular, for city formation, they also show an adjustment to cities with populations bigger than a hundred thousand inhabitants, by using a kind of Zipf plot for x^c versus $\ln(n)$, where c is an adjustable parameter. However, the Weibull distribution leads to a poor adjustment for the complete set of data, i.e., this distribution gives us a satisfactory adjustment only for a restricted range of data. Furthermore, it is clear that the stretched function does not lead to an asymptotic straight line in a log-log plot, i.e., a power law.

On the other hand, the Zipf-Mandelbrot law [21] $N(x) = b/(c+x)^{\alpha}$ (b, c, and α all being positive constants), gives a curvature in a log-log plot, presents an asymptotic power-law behavior, and may be normalized for $\alpha>1$. In this way, the Zipf-Mandelbrot distribution is a natural generalization of an inverse power law. This distribution has been applied in many contexts; in particular, it was recently employed in the discussion of scientific citations [5] and football goal distribution [9]. Another important aspect of the Zipf-Mandelbrot distribution is that it arises naturally in the context of a generalized statistical mechanics proposed some years ago [22–25]. In this framework, the above distribution is usually rewritten as a q-exponential function

$$N(x) = N_0 \exp_{q'}(-ax) \equiv N_0[1-(1-q')ax]^{1/(1-q')}, \tag{4}$$

where $N_0 = bc^{-\alpha}$, $a = \alpha/c$, and $q' = 1 + 1/\alpha$ are positive parameters. Moreover, the above distribution has been largely used with $q'<1$ in other contexts [26]. In this case, Eq. (4) is defined equal to zero when $1-(1-q')ax<0$ in order to overcome imaginary values for $N(x)$. Thus, the distribution (4) is equivalent to the Zipf-Mandelbrot law only for $q'>1$ and gives an extension for such a law when $q'<1$ is employed. Note, also, that $\exp_{q'}(-x)$ reduces to the usual exponential function $\exp(-x)$, in the limit $q'\to 1$. In addition, Eq. (4) satisfies an anomalous decay equation

$$\frac{d}{dx}\left(\frac{N(x)}{N_0}\right) = -a\left(\frac{N(x)}{N_0}\right)^{q'}, \tag{5}$$

independently of the q' value. Since this equation reduces to the usual decay one in the limit $q'\to 1$, the parameter q' can be interpreted as a measure of how anomalous is the decay. These aspects put the Zipf-Mandelbrot law in a broader context, motivating us to employ the generalized Tsallis exponential Eq. (4), instead of the Zipf-Mandelbrot form to study the city population distribution.

The cumulative distribution for $1<q'<1.5$ is

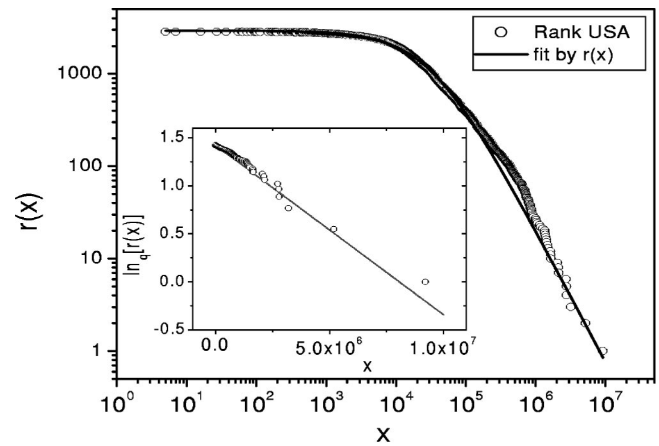

FIG. 2. Fit of cumulative distribution for all cities in the U.S.A. The parameters are $q=1.7$, $r_0=2919.4$, and $a=0.000\,08$. The coefficient of determination in nonlinear fit is $R^2=0.99$. Inset plot: generalized monolog plot for American cities.

$$r(x) = r_0\left[1-\frac{(1-q)}{q}ax\right]^{1/(1-q)}, \tag{6}$$

where $r_0 = N_0 q/a$, and $q=(2-q')^{-1}$. Usually, to compare this cumulative distribution with that obtained from data, a log-log plot is employed. Here, we introduce another possible way to analyze data by using a generalized monolog plot based on the generalized logarithm function, $\ln_q(x) \equiv (x^{1-q}-1)/(1-q)$. This generalized function arises naturally in the framework of Tsallis statistics [22,23,25] and reduces to the usual logarithm $\ln(x)$ for $q\to 1$. It is easy to verify that the plot of $\ln_q[r(x)]$ versus x leads to a straight line. So, if the data are well described by the distribution (4), we are able to obtain the q value that gives the best linear fit in the generalized monolog plot, independently of other parameters.

Here, we used this generalized monolog plot analysis and we found that $q\approx 1.7$ gives a good adjustment to all American and Brazilian cities. Inset plots of Figs. 2 and 3 show this adjustment for American and Brazilian cities, respectively.

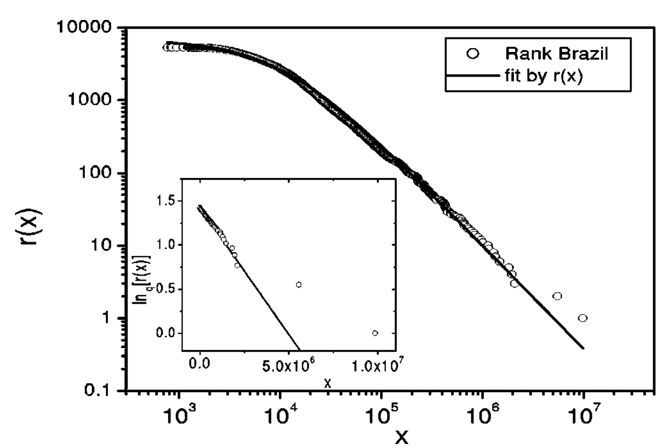

FIG. 3. Fit of cumulative distribution for all cities in Brazil. The parameters are $q=1.7$, $r_0=6968.6$, and $a=0.000\,24$. The coefficient of determination in nonlinear fit is $R^2=0.99$. Inset plot: generalized monolog plot for Brazilian cities.

Note that in Fig. (3), the two biggest cities are above the straight line formed by all other cities. This fact is known as the ''king'' effect [20,27], and occurs because a few cities in some of the countries, by a specific cause (economic, political, etc.), play an irregular competition to attract people and do not follow the same rules that most of the other cities do. These cities that dominate a region or country, which are highly centralized, are also referred to as the ''primate cities'' effect [28]. Of course, this effect can also be observed if you restrict it to cities with more than one hundred thousand inhabitants. For example, if we consider countries such as England and France, the ''king'' effect is related to London and Paris [20].

By fixing $q=1.7$, we obtain the other parameters from a nonlinear fit for the cumulative distribution. This fit is shown in Fig. 2 for American cities and in Fig. 3 for Brazilian ones.

In order to analyze the agreement between the data and the obtained distribution, beyond what has been visualized in Figs. 2 and 3, we calculate the total population $p=\int_{x_{min}}^{\infty} xN(x)dx$ and the average population by cities by $\langle x\rangle=\int_{x_{min}}^{\infty} xN(x)dx/\int_{x_{min}}^{\infty} N(x)dx$ [29]. Comparing p and $\langle x\rangle$ with experimental value, we obtain the deviation $\Delta p \equiv[(p_{data}-p_{model})/p_{data}]100\%=3.9\%$ for U.S.A. cities. Now, considering cities with less than one hundred thousand inhabitants, we have $\Delta p_{<}=4.6\%$, which is better than the one obtained in Ref. [20] using the stretched exponential distribution. For the U.S.A. average population, we obtain $\Delta\langle x\rangle=6.3\%$. In the Brazilian case, we obtain $\Delta p=7.0\%$ and $\Delta\langle x\rangle=9.0\%$. It is interesting to remark that the deviations $\Delta\langle x\rangle$ and Δp could be smaller if the ''king'' effect is not present.

In this Brief Report, we show that the population of a country (U.S.A. and Brazil), distributed in its cities, is well described by a q-exponential with $q=1.7$. Thus, this fact indicates a possible connection among the previous results, Tsallis statistics, and anomalous decay. Furthermore, when one deals with a distribution that may be adjusted by a q-exponential, the generalized monolog plot introduced here gives a practical way to determine the q value, independent of other parameters of the distribution.

We thank CNPq (Brazilian Agency) for partial financial support.

[1] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949).
[2] B.D. Malamud, G. Morein, and D.L. Turcotte, Science **281**, 1840 (1998).
[3] T.H. Keit and H.E. Stanley, Nature (London) **393**, 257 (1998).
[4] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[5] C. Tsallis and M.P. Albuquerque, Eur. Phys. J. B **13**, 777 (2000).
[6] B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, and R.M. Lukose, Science **280**, 95 (1998).
[7] J.R. Banavar, J.L. Green, J. Harte, and A. Maritan, Phys. Rev. Lett. **83**, 4212 (1999).
[8] G. Boffetta, V. Carbone, P. Giuliani, P. Veltri, and A. Vulpiani, Phys. Rev. Lett. **83**, 4662 (1999).
[9] L.C. Malacarne and R.S. Mendes, Physica A **286**, 391 (2000).
[10] R.N. Mantegna and H.E. Stanley, Nature (London) **376**, 46 (1995).
[11] C.J. Rhodes and M. Anderson, Nature (London) **381**, 600 (1996).
[12] A.H. Makse, S. Havlin, and H.E. Stanley, Nature (London) **377**, 608 (1995).
[13] H. A. Makse, J.S. Andrade Jr., M. Batty, S. Havlin, and H.E. Stanley, Phys. Rev. E **58**, 7054 (1998).
[14] D.H. Zanette and S.C. Manrubia, Phys. Rev. Lett. **79**, 523 (1997).
[15] M. Marsili and Y.-C. Zhang, Phys. Rev. Lett. **80**, 2741 (1998).
[16] G. Malescio, N.V. Dokholyan, S.V. Buldyrev, and H.E. Stanley, e-print cond-mat/0005178.
[17] For Europe and India the data were obained in http://www.un.org/Depts/unsd/demog/index.html. For the U.S.A. and Brazil, we used the data of Refs. [18] and [19], respectively.
[18] County Population Estimates for July 1, 1998. Source: Population Estimates Program, Population Division, U.S. Bureau of the Census, Washington, DC 20233. Contact: Statistical Information Staff, Population Division, U.S. Bureau of the Census (301-457-2422). Internet Release Date: March 12, 1999.
[19] We consider the population of the 5336 Brazilian cities obtained from the census of 1991. http://200.255.94.114/ibge/ftp/ftp.php?dir=/Censos
[20] J. Laherrere and D. Sornette, Eur. Phys. J. B **2**, 525 (1998).
[21] B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1977).
[22] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[23] E.M.F. Curado and C. Tsallis, J. Phys. A **24**, L69 (1991); **24**, 3187(E) (1991); **25**, 1019(E) (1992).
[24] S. Denisov, Phys. Lett. A **235**, 447 (1997).
[25] C. Tsallis, R.S. Mendes, and A.R. Plastino, Physica A **261**, 534 (1998).
[26] *Nonextensive Statistical Mechanics and Its Applications*, Lectures Notes in Physics, Vol. 560, edited by S. Abe and Y. Okamoto (Springer, New York, 2001).
[27] J. Laherrère, C. R. Acad. Sci., Ser. IIa: Sci. Terre Planetes **322**, 535 (1996).
[28] *Geographic Perspectives on Urban Systems*, edited by B. J. Berry and F. Horton (Prentice Hall, Englewood Cliffs, NJ, 1970).
[29] In the calculation of total population and average population, we use the limit of integration, $\int_{x_{min}}^{x_{max}}dx$, where x_{min} is the population of the smallest city of the country to be analyzed and x_{max} the largest one. Usually, we can replace $x_{max}\rightarrow\infty$ since it does not substantially affect the result.