A Generative Model of Power Law Distributions with Optimizing Agents with Constrained Information Access

Extended Abstract

Laszlo Gulyas

AITIA International Inc.
H-1039, Budapest, Czetz Janos utca 48-50.
Fax: +36 1 4538081, e-mail: lgulyas@aitia.ai

Short Title: A Generative Model of Power Laws

Abstract

US metropolitan areas show an intriguing empirical regularity. According to statistical surveys, the population of a city is inversely proportional to its rank (by population). This type of power law relationships are of special interest to complex systems studies, because they are surprisingly common in natural and man-made systems. Therefore, a number of general, abstract mechanisms and generative models have been proposed to explain the occurrence of power law distributions, independent of the particular domain they may occur in. These models make different assumptions of the system, its actors and their dynamics. In this paper a new generative model is presented, based on utility maximizing behavior of agents using limited local information. This model is capable of generating power law distributions given sufficient heterogeneity is present in the distribution of information among the agents. The model is related to the city formation model of Simon that has become more widely known in its variant by Albert and Barabási as the 'preferential attachment' model. The main difference between these models and the one presented in this paper is that the former models require global access to information (i.e., the arriving new agent or node has to assess the distribution of links/size of the whole pre-existing population), while our model operates with limited information and utility maximization subject to this set of information.

Keywords

Power Law, Scale-Free Distribution, Constrained Information Access, Limited Information, Heterogeneity, Generative Model

Introduction

US metropolitan areas show an intriguing empirical regularity. According to statistical surveys, the population of a city is inversely proportional to its rank (by population). If x_q denotes the size of the q^{th} largest city, then x_q equals $a \cdot r^{1}$, where a is the size of the largest city. This type of power law relationship is surprisingly common in many complex systems as testified by a vast and growing body of work in the literature of various disciplines. [3][4][5][6][7][8][9][10][11][14]

Power laws are of special interest to complex systems studies. A number of general, abstract mechanisms and generative models have been proposed to explain the occurrence of power law distributions, independent of the particular domain they may occur in. These models make different assumptions of the system, its actors and their dynamics. In this paper a new generative model is presented, based on utility maximizing behavior of agents using limited local information. This model is capable of generating power law distributions given sufficient heterogeneity is present in the distribution of information among the agents. The model is related to the city formation model of Simon [12][13] that has become more widely known in its variant by Albert and Barabási as the 'preferential attachment' model. [1][2] The main difference between these models and the one presented in this paper is that the former models require global access to information (i.e., the arriving new agent or node has to assess the distribution of links/size of the whole pre-existing population), while our model operates with limited information and utility maximization subject to this set of information.

The Model

Let $x_1^t,\ x_2^t,\ \ldots,\ x_L^t\ \in \mathbb{N}$ be positive integers for $t \geq 0$ and $L \in \mathbb{N}$, so that $x_l^0 = 1$, for all $l \in [1, L]$. At each time t, we take a random sample of x_l^t's without replacement. S^t will stand for the size of the sample, a random positive integer drawn uniformly from $[1,L]$: $S^t \in \mathbf{U}[1, L]$, for all $t \geq 0$. Z^t will stand for the sample itself:

$$(1) \qquad Z^t = \left\{ x_l^t \text{ such that } l \in \mathbf{U}[1, L] \right\} \text{ where } |Z^t| = S^t, \text{ for all } t \geq 0.$$

Let μ^t denote the largest element in the sample:

$$(2) \qquad \mu^t = \arg\max_{l \in Z^t} x_l^t,\ t \geq 0.$$

We will call μ^t as the *selected element* of the sample. If μ^t is not unique, we take the one with the lowest or highest index l, or simply a random 'maximum'. The dynamics of the system is the following:

$$(3) \qquad x_l^{t+1} = \begin{cases} x_l^t & ,\text{if } l \neq \mu^t \\ x_l^t + \gamma & ,\text{if } l = \mu^t \end{cases},\ \text{for all } t \geq 0, \text{ where } \gamma \in \mathbb{N}.$$

That is, the selected element of the sample is increased by a constant γ, while all the other elements remain the same.

We are interested in the properties of x_l^t, in the limit of $t \to \infty$. Especially, we are interested in the following 'frequency distribution': $FD^t(y) = \sum_{l \in [1,L]} \chi\left(x_l^t \geq y\right)$, where $\chi(.)$ is the membership function, and thus $FD^t(y)$ gives the number of x_l^t's that are at least the size of y at time t.

Numerical Results

Numerical simulations suggest that $\lim_{t \to \infty} FD^t(y) \sim y^{\alpha}$ with $\alpha \approx -2$, as shown by Figure 1. The figure shows the emerged power law distribution for L=100, 1000 and 10000 after $t=1000 \cdot L$ iterations on a log-log scale. The linear approximation of the L=1000 case is also included for convenience, together with its goodness-of-fit (R^2) value.

Figure 2 provides a summary of the sensitivity analysis. The sensitivity analysis consisted of creating 10 simulation runs with different pseudo random number sequences for each parameter combination (L=100, 1000, 10000, 100000). The resulting frequency distribution of each individual run was saved after t=1000000 iterations and linear regression was applied to their log-log scale transformations. The figure summarizes the goodness-of-fit (R^2) values obtained. It is clear that the results are independent of the particular pseudo random sequence used. On the other hand, the fit is quite naturally dependent on the number of iterations (*t*). The apparent drop in goodness-of-fit at L≥10000 is due to the insufficient number of iterations.

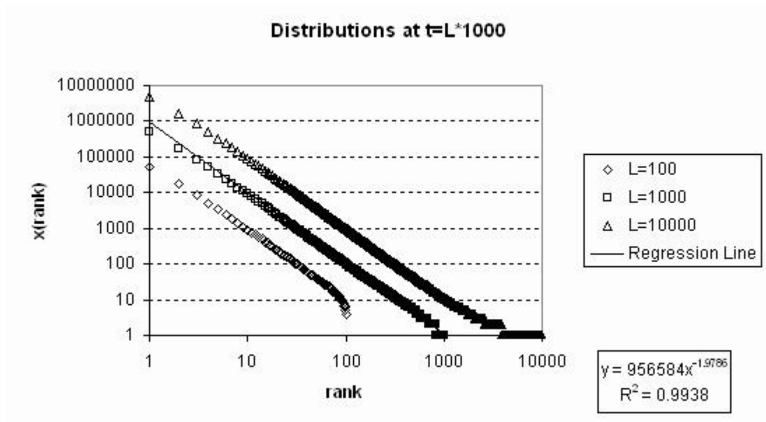

Figure 1: The emerging power law distributions for L=100, 1000 and 10000.

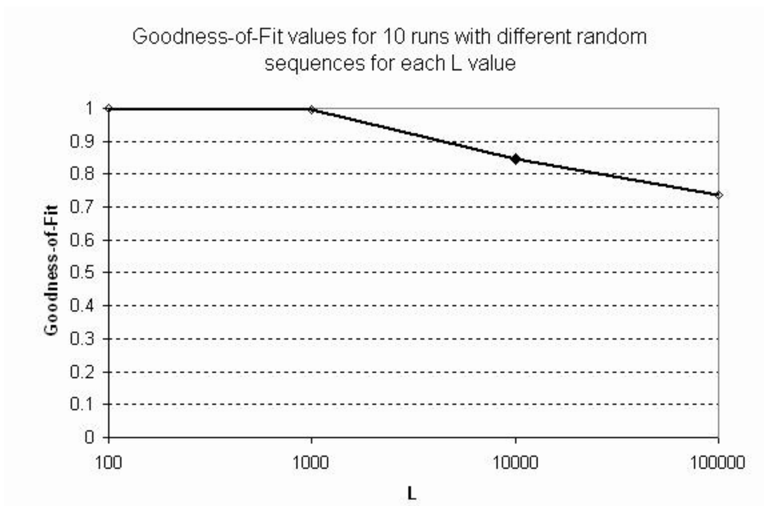

Figure 2: Sensitivity analysis of the results. Goodness-of-fit values for 10 runs with different pseudo random sequences for L=100, 1000, 10000 and 100000.

Conclusions

This paper presented a novel generative model of power law distributions. The driving force behind the process resulting in the scale-free distribution is the maximizing behavior of agents, acting on limited local information that is distributed unevenly among them. We believe that this entirely local approach has a great potential to explain the occurrence of power law distributions in social systems with bounded rational agents, where the assumption global information access is unfeasible.

Acknowledgements

The author gratefully acknowledges the partial support of the Informatics Cooperative Research and Education Center, Loránd Eötvös University, Budapest and the GVOP-3.2.2-2004-07-0005/3.0 grant of the Hungarian National Office for Research and Technology. He would also like to thank Prof. George Kampis at Loránd Eötvös University, Budapest and Prof. Peter Erdi at the Center for Complex Systems Studies, Physics Department, Kalamazoo College, MI for their help.

References

[1] Albert, R. and Barabási, A.-L.: "Statistical mechanics of complex networks", *Reviews of Modern Physics*, Vol. 74:47, 2002.

[2] Barabási, A.-L. and Albert, R.: "Emergence of scaling in random networks", *Science*, Vol. 286, pp. 509-512, 1999.

[3] Cederman, L.-E.: "Modeling the Size of Wars: From Billiard Balls to Sandpiles", *American Political Science Review*, Vol. 97, pp. 135-150, 2003.

[4] Champernowne, D.: "A model of income distribution", *Economic Journal*, Vol. 63, pp. 318-351, 1953.

[5] Faloutsos, M., Faloutsos, P. and Faloutsos, C.: "On power-law relationships of the Internet topology", In *Proceedings of ACM SIGCOMM 1999 Conference*, pp. 251-261. 1999.

[6] Gabaix, X.: "Zipf's law for cities: an explanation", *Quarterly Journal of Economics*, Vol. 114, pp. 739-767, 1999

[7] Gulyás, L.: "Generation of Robust Networks with Optimization under Budget Constraints", In *Proceedings of The 5th International Workshop on Emergent Synthesis (IWES'04)*, Budapest, Hungary, 2004.

[8] Mega, M. S., Allegrini, P., Grigolini, P., Latora, V., Palatella, L., Rapisarda, A. and Vinciguerra, S.: "Power-Law Time Distribution of Large Earthquakes", *Physical Review Letters*, Vol. 90:18, pp. 188501, 2003.

[9] Mitzenmacher, M.: "A brief history of generative models for power law and lognormal distributions", *Internet Mathematics*, Vol. 1:2, pp. 226-251, 2003.

[10] Neukum, G. and Ivanov, B. A.: "Crater size distributions and impact probabilities on Earth from lunar, terrestrial-planet, and asteroid cratering data", *Hazards due to comets and asteroids* (Ed. Gehrels, T.), University of Arizona Press, pp. 359-390. 1994.

[11] Pareto, V.: *Cours d'Economie Politique*, Droz, Geneva, Switzerland. 1896.

[12] Simon, H. A.: "On A Class of Skew Distribution Functions", *Biometrika*, Vol. 52, pp. 425-440, 1955.

[13] Simon, H. A. and Bonini, C. P.: "The Size Distribution of Business Firms", *American Economic Review*, Vol. 48, pp. 607-617, 1958.

[14] Willis, J. C.: *Age and Area: A Study in Geographic Distribution and Origin of Species*, Cambridge University Press, Cambridge, UK. 1922.