Title

A maximum entropy framework for nonexponential distributions

Permalink
https://escholarship.org/uc/item/2983m4r8

Journal
Proceedings of the National Academy of Sciences of the United States of America, 110(51)

ISSN
0027-8424

Authors
Peterson, Jack
Dixit, Purushottam D
Dill, Ken A

Publication Date
2013-12-17

DOI
10.1073/pnas.1320578110

Peer reviewed

A maximum entropy framework for nonexponential distributions

Jack Peterson[a,b], Purushottam D. Dixit[c], and Ken A. Dill[b,1]

[a]Department of Mathematics, Oregon State University, Corvallis, OR 97331; [b]Laufer Center for Physical and Quantitative Biology, Departments of Physics and Chemistry, State University of New York, Stony Brook, NY 11794; and [c]Department of Systems Biology, Columbia University, New York, NY 10032

Contributed by Ken A. Dill, November 7, 2013 (sent for review June 26, 2013)

Probability distributions having power-law tails are observed in a broad range of social, economic, and biological systems. We describe here a potentially useful common framework. We derive distribution functions $\{p_k\}$ for situations in which a "joiner particle" k pays some form of price to enter a community of size $k-1$, where costs are subject to economies of scale. Maximizing the Boltzmann–Gibbs–Shannon entropy subject to this energy-like constraint predicts a distribution having a power-law tail; it reduces to the Boltzmann distribution in the absence of economies of scale. We show that the predicted function gives excellent fits to 13 different distribution functions, ranging from friendship links in social networks, to protein–protein interactions, to the severity of terrorist attacks. This approach may give useful insights into when to expect power-law distributions in the natural and social sciences.

heavy tail | fat tail | statistical mechanics | thermostatistics | social physics

Probability distributions are often observed to have power-law tails, particularly in social, economic, and biological systems. Examples include distributions of fluctuations in financial markets (1), the populations of cities (2), the distribution of Web site links (3), and others (4, 5). Such distributions have generated much popular interest (6, 7) because of their association with rare but consequential events, such as stock market bubbles and crashes.

If sufficient data are available, finding the mathematical shape of a distribution function can be as simple as curve-fitting, with a follow-up determination of the significance of the mathematical form used to fit it. However, it is often interesting to know if the shape of a given distribution function can be explained by an underlying generative principle. Principles underlying power-law distributions have been sought in various types of models. For example, the power-law distributions of node connectivities in social networks have been derived from dynamical network evolution models (8–17). A large and popular class of such models is based on the preferential attachment rule (18–27), wherein it is assumed that new nodes attach preferentially to the largest of the existing nodes. Explanations for power laws are also given by Ising models in critical phenomena (28–34), network models with thresholded "fitness" values (35), and random-energy models of hydrophobic contacts in protein interaction networks (36).

However, such approaches are often based on particular mechanisms or processes; they often predict particular power-law exponents, for example. Our interest here is in finding a broader vantage point, as well as a common language, for describing a range of distributions, from power law to exponential. For deriving exponential distributions, a well-known general principle is the method of maximum entropy (Max Ent) in statistical physics (37, 38). In such problems, you want to choose the best possible distribution from all candidate distributions that are consistent with certain set of constrained moments, such as the average energy. For this type of problem, which is highly underdetermined, a principle is needed for selecting a "best" mathematical function from among alternative model distribution functions. To find the mathematical form of the distribution function p_k over states $k = 1, 2, 3, \ldots$, the Max Ent principle asserts that you should maximize the Boltzmann–Gibbs–Shannon (BGS) entropy functional $S[\{p_k\}] = -\sum_k p_k \log\, p_k$ subject to constraints, such as the known value of the average energy $\langle E \rangle$. This procedure gives the exponential (Boltzmann) distribution, $p_k \propto e^{-\beta E_k}$, where β is the Lagrange multiplier that enforces the constraint. This variational principle has been the subject of various historical justifications. It is now commonly understood as the approach that chooses the least-biased model that is consistent with the known constraint(s) (39).

Is there an equally compelling principle that would select fat-tailed distributions, given limited information? There is a large literature that explores this. Inferring nonexponential distributions can be done by maximizing a different mathematical form of entropy, rather than the BGS form. Examples of these nontraditional entropies include those of Tsallis (40), Renyi (41), and others (42, 43). For example, the Tsallis entropy is defined as $\frac{K}{1-q}\left(\sum_k p_k^q - 1\right)$, where K is a constant and q is a parameter for the problem at hand. Such methods otherwise follow the same strategy as above: maximizing the chosen form of entropy subject to an extensive energy constraint gives nonexponential distributions. The Tsallis entropy has been applied widely (44–53).

However, we adopt an alternative way to infer nonexponential distributions. To contrast our approach, we first switch from probabilities to their logarithms. Logarithms of probabilities can be parsed into energy-like and entropy-like components, as is standard in statistical physics. Said differently, a nonexponential distribution that is derived from a Max Ent principle requires that there be nonextensivity in either an energy-like or entropy-like term; that is, it is nonadditive over independent subsystems, not scaling linearly with system size. Tsallis and others have chosen to assign the nonextensivity to an entropy term, and retain extensivity in an energy term. Here, instead, we keep the canonical BGS form of entropy, and invoke a nonextensive energy-like term.

Significance

Many statistical distributions, particularly among social and biological systems, have "heavy tails," which are situations where rare events are not as improbable as would have been guessed from more traditional statistics. Heavy-tailed distributions are the basis for the phrase "the rich get richer." Here, we propose a basic principle underlying systems with heavy-tailed distributions. We show that it is the same principle (maximum entropy) used in statistical physics and statistics to estimate probabilistic models from relatively few constraints. The heavy-tail principle can be expressed in terms of shared costs and economies of scale. The probability distribution we derive is a mathematical digamma function, and we show that it accurately fits 13 real-world data sets.

Author contributions: J.P. and K.A.D. designed research; J.P. and P.D.D. performed research; J.P. and K.A.D. contributed new reagents/analytic tools; J.P. analyzed data; and J.P., P.D.D., and K.A.D. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dill@laufercenter.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1320578110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1320578110

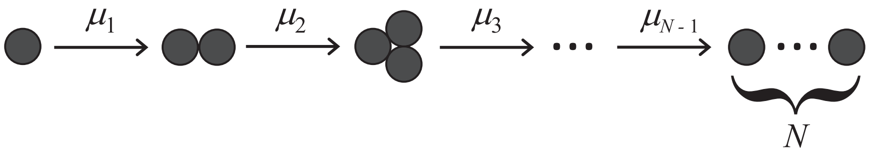

Fig. 1. The joining cost for a particle to join a size $k-1$ community is μ_k. This diagram can describe particles forming colloidal clusters, or social processes such as people joining cities, citations added to papers, or link creation in a social network.

In our view, only the latter approach is consistent with the principles elucidated by Shore and Johnson (37) (reviewed in ref. 39). Shore and Johnson (37) showed that the BGS form of entropy is uniquely the mathematical function that ensures satisfaction of the addition and multiplication rules of probability. Shore and Johnson (37) assert that any form of entropy other than BGS will impart a bias that is unwarranted by the data it aims to fit. We regard the Shore and Johnson (37) argument as a compelling first-principles basis for defining a proper variational principle for modeling distribution functions. Here, we describe a variational approach based on the BGS entropy function, and we seek an explanation for power-law distributions in the form of an energy-like function instead.

Theory

Assembly of Simple Colloidal Particles. We frame our discussion in terms of a joiner particle that enters a cluster or community of particles, as shown in Fig. 1. However, this is a natural way to describe the classical problem of the colloidal clustering of physical particles; it is readily shown (reviewed below) to give an exponential distribution of cluster sizes. However, this general description also pertains more broadly, such as when people populate cities, links are added to Web sites, or when papers accumulate citations. We want to compute the distribution, p_k, of populations of communities having size $k=1,2,\ldots,N$.

To begin, we express a cumulative cost of joining. For particles in colloids, this cost is expressed as a chemical potential, i.e., a free energy per particle. If μ_j represents the cost of adding particle j to a cluster of size $j-1$, the cumulative cost of assembling a whole cluster of k particles is the sum

$$w_k=\sum_{j=1}^{k-1}\mu_j. \qquad [1]$$

Max Ent asserts that we should choose the probability distribution that has the maximum entropy among all candidate distributions that are consistent with the mean value $\langle w\rangle$ of the total cost of assembly (54),

$$p_k=\frac{e^{-\lambda w_k}}{\sum_i e^{-\lambda w_i}}, \qquad [2]$$

where λ is a Lagrange multiplier that enforces the constraint.

In situations where the cost of joining does not depend on the size of the community a particle joins, then $\mu_k=\mu^\circ$, where μ° is a constant. The cumulative cost of assembling the cluster is then

$$w_k=(k-1)\mu^\circ. \qquad [3]$$

Substituting into Eq. **2** and absorbing the Lagrange multiplier λ into μ° yields the grand canonical exponential distribution, well known for problems such as this:

$$p_k=\frac{e^{-\mu^\circ k}}{\sum_i e^{-\mu^\circ i}}. \qquad [4]$$

In short, when the joining cost of a particle entry is independent of the size of the community it enters, the community size distribution is exponential.

Communal Assemblies and Economies of Scale. Now, we develop a general model of communal assembly based on economies of scale. Consider a situation where the joining cost for a particle depends on the size of the community it joins. In particular, consider situations in which the costs are lower for joining a larger community. Said differently, the cost-minus-benefit function μ_k is now allowed to be subject to economies of scale, which, as we note below, can also be interpreted instead as a form of discount in which the community pays down some of the joining costs for the joiner particle.

To see the idea of economy-of-scale cost function, imagine building a network of telephones. In this case, a community of size 1 is a single unconnected phone. A community of size 2 is two connected phones, etc. Consider the first phone: The cost of creating the first phone is high because it requires initial investment in the phone assembly plant. And the benefit is low, because there is no value in having a single phone. Now, for the second phone, the cost-minus-benefit is lower. The cost of producing the second phone is lower than the first because the production plant already exists, and the benefit is higher because two connected phones are more useful than one unconnected phone. For the third phone, the cost-minus-benefit is even lower than for the second because the production cost is even lower (economy of scale) and because the benefits increase with the number of phones in the network.

To illustrate, suppose the cost-minus-benefit for the first phone is 150, for the second phone is 80, and for the third phone is 50. To express these cost relationships, we define an intrinsic cost for the first phone (joiner particle), 150 in this example. We define the difference in cost-minus-benefit between the first and second phones as the discount provided by the first phone when the second phone joins the community of two phones. In this example, the first phone provides a discount of 70 when the second phone joins. Similarly, the total discount provided by the two-phone community is 100 when the third phone joins the community.

In this language, the existing community is paying down some fraction of the joining costs for the next particle. Mathematically, this communal cost-minus-benefit function can be expressed as

$$\mu_k=\mu^\circ-\frac{k\mu_k}{k_0}. \qquad [5]$$

The quantity μ_k on the left side of Eq. **5** is the total cost-minus-benefit when a particle joins a k-mer community. The joining cost has two components, expressed on the right side: each joining event has an intrinsic cost μ° that must be paid, and each joining event involves some discount that is provided by the community. Because there are k members of the existing community, the quantity μ_k/k_0 is the discount given to a joiner by each existing community particle, where k_0 is a problem-specific parameter that characterizes how much of the joining cost burden is shouldered by each member of the community. In the phone example, we assumed $k_0=1$. The value of $k_0=1$ represents fully equal cost-sharing between joiner and community member: each communal particle gives the joining particle a discount equal to what the joiner itself pays. The opposite extreme

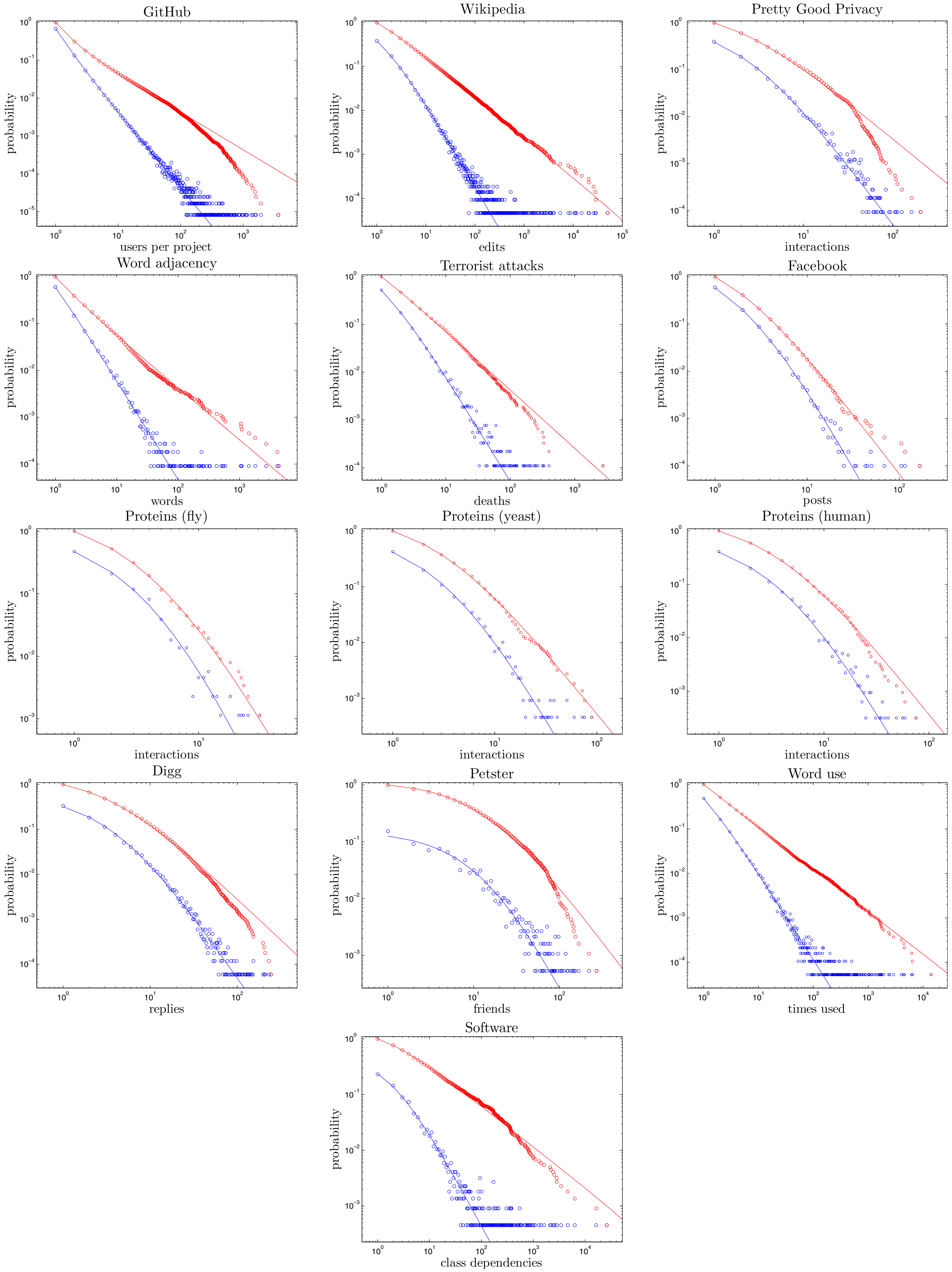

Fig. 2. Eq. **7** gives good fits ($P > 0.05$; see *SI Text* for details) to 13 empirical distributions, with the values of μ° and k_0 given in Table 1. Points are empirical data, and lines represent best-fit distributions. The probability p_k of exactly k is shown in blue, and the probability of at least k (the complementary cumulative distribution, $\sum_{j=k}^{\infty} p_j$) is shown in red. Descriptions and references for these datasets can be found in *SI Text*.

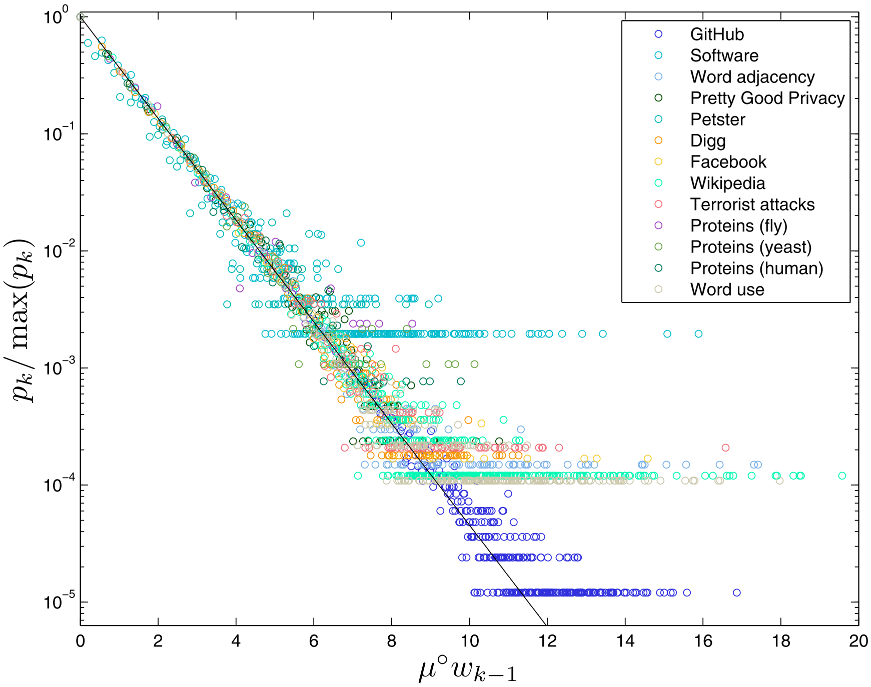

Fig. 3. Eq. **7** fitted to the 13 datasets in Table 1, plotted against the total cost to assemble a size k community, $\mu^\circ w_{k-1}$. Values of μ° and k_0 are shown in Table 1. The y axis has been rescaled by dividing by the maximum p_k, so that all curves begin at $p_k/\max(p_k) = 1$. All data sets are fit by the log $y = -x$ line. See Fig. 2 for fits to individual datasets.

limit is represented by $k_0 \to \infty$; in this case, the community gives no discount at all to the joining particle.

The idea of communal sharing of cost-minus-benefit is applicable to various domains; it can express that one person is more likely to join a well-populated group on a social networking site because the many existing links to it make it is easier to find (i.e., lower cost) and because its bigger hub offers the newcomer more relationships to other people (i.e., greater benefit). Or, it can express that people prefer larger cities to smaller ones because of the greater benefits that accrue to the joiner in terms of jobs, services, and entertainment. (In our terminology, a larger community pays down more of the cost-minus-benefit for the next immigrant to join.) We use the terms "economy of scale" (EOS) or "communal" to refer to any system that can be described by a cost function, such as Eq. **5**, in which the community can be regarded as sharing in the joining costs, although other functional forms might also be of value for expressing EOS.

Rearranging Eq. **5** gives $\mu_k = \mu^\circ k_0/(k+k_0)$. The total cost-minus-benefit, w_k, of assembling a community of size k is

$$w_k = \mu^\circ k_0 \sum_{j=1}^{k-1} \frac{1}{j+k_0} = \mu^\circ k_0 \psi(k+k_0) - C, \qquad \textbf{[6]}$$

where $\psi(k) = -\gamma + \sum_{j=1}^{k-1} j^{-1}$ is the digamma function ($\gamma = 0.5772\ldots$ is Euler's constant), and the constant term $C = \mu^\circ k_0 \psi(k_0) + \mu^\circ$ will be absorbed into the normalization.

From this cost-minus-benefit expression (Eq. **6**), for a given k_0, we can now uniquely determine the probability distribution by maximizing the entropy. Substituting Eq. **6** into Eq. **2** yields

$$p_k = \frac{e^{-\mu^\circ k_0 \psi(k+k_0)}}{\sum_i e^{-\mu^\circ k_0 \psi(i+k_0)}}. \qquad \textbf{[7]}$$

Eq. **7** describes a broad class of distributions. These distributions have a power-law tail for large k, with exponent $\mu^\circ k_0$, and a cross-over at $k = k_0$ from exponential to power law. To see this, expand $\psi(k+k_0)$ asymptotically and drop terms of order $1/k^2$; this yields $w_k \sim \mu^\circ k_0 \ln\left(k+k_0-\frac{1}{2}\right)$, so Eq. **7** obeys a power law $p_k \sim \left(k+k_0-\frac{1}{2}\right)^{-\mu^\circ k_0}$ for large k, and p_k becomes a simple exponential in the limit of $k_0 \to \infty$ (zero cost-sharing). One quantitative measure of a distribution's position along the continuum from exponential to power law is the value of its scaling exponent, $\mu^\circ k_0$. A small exponent indicates that the system has extensive social sharing, thus power-law behavior. As the exponent becomes large, the distribution approaches an exponential function. Eq. **7** has a power-law scaling only when the cost of joining a community has a linear dependence on the community size. The linear dependence arises because the joiner particle interacts identically with all other particles in the community.

What is the role of detailed balance in our modeling? Fig. 1 shows no reverse arrows from k to $k-1$. The principle of Max Ent can be regarded as a general way to infer distribution functions from limited information, irrespective of whether there is an underlying a kinetic model. So, it poses no problem that

Table 1. Fitting parameters and statistics

Data set	μ°	k_0	$\langle k \rangle$	N	$\mu^\circ k_0$	P
GitHub	9(1)	0.21(2)	3.642	120,866	2(2)	0.78
Wikipedia	1.5(1)	1.3(1)	25.418	21,607	1.9(1)	0.79
Pretty Good Privacy	1(1)	2.6(2)	4.558	10,680	2.6(3)	0.16
Word adjacency	3.6(4)	0.6(1)	5.243	11,018	2.1(3)	0.09
Terrorist attacks	2.1(2)	1(1)	4.346	9,101	2.2(3)	0.38
Facebook wall	1.6(1)	2.3(3)	2.128	10,082	3.6(5)	0.99
Proteins (fly)	0.9(2)	5(2)	2.527	878	5(2)	0.89
Proteins (yeast)	0.9(1)	4(1)	3.404	2,170	3(1)	0.48
Proteins (human)	0.8(1)	4(1)	3.391	3,165	4(1)	0.52
Digg	0.68(3)	4.2(3)	5.202	16,844	2.8(2)	0.05
Petster	0.21(3)	15(3)	13.492	1,858	3(1)	0.08
Word use	2.3(1)	0.8(1)	11.137	18,855	1.9(2)	0.56
Software	0.8(1)	2.1(3)	62.82	2,208	1.7(3)	0.69

some of our distributions, such as scientific citations, are not taken from reversible processes.

Results

Eq. **7** and Fig. 2 show the central results of this paper. Consider three types of plots. On the one hand, exponential functions can be seen in data by plotting log p_k vs. k. Or, power-law functions are seen by plotting log p_k vs. log k. Here, we find that plotting log p_k vs. a digamma function provides a universal fit to several disparate experimental data sets over their full distributions (Fig. 3). Fig. 2 shows fits of Eqs. **7**–**13** datasets, using μ° and k_0 as fitting parameters that are determined by a maximum-likelihood procedure (see *SI Text* for dataset and goodness-of-fit test details). The μ° and k_0 characterize the intrinsic cost of joining any cluster, and the communal contribution to sharing that cost, respectively.

Rare events are less rare under fat-tailed distributions than under exponential distributions. For dynamical systems, the risk of such events can be quantified by the coefficient of variation (CV), defined as the ratio of the SD σ_k to the mean $\langle k\rangle$. For equilibrium/steady-state systems, the CV quantifies the spread of a probability distribution, and is determined by the power-law exponent, $\mu^\circ k_0$. Systems with small scaling exponents ($\mu^\circ k_0 \leq 3$) experience an unbounded, power-law growth of their CV as the system size N becomes large, $\sigma_k/\langle k\rangle \sim N^{\beta}$. This growth is particularly rapid in systems with $1.8 < \mu^\circ k_0 < 2.2$, because the average community size $\langle k\rangle$ diverges at $\mu^\circ k_0 = 2$. For these systems, $\beta = 1/2$ is observed. Several of our datasets fall into this high-risk category, such as the number of deaths due to terrorist attacks (Table 1).

Discussion

We have expressed a range of probability distributions in terms of a generalized energy-like cost function. In particular, we have considered types of costs that can be subject to economies of scale, which we have also called "community discounts." We maximize the BGS entropy, subject to such cost-minus-benefit functions. This procedure predicts probability distributions that are exponential functions of a digamma function. Such a distribution function has a power-law tail, but reduces to a Boltzmann distribution in the absence of EOS. This function gives good fits to distributions ranging from scientific citations and patents, to protein-protein interactions, to friendship networks, and to Web links and terrorist networks—over their full distributions, not just in their tails.

Framed in this way, each new joiner particle must pay an intrinsic buy-in cost to join a community, but that cost may be reduced by a communal discount (an economy of scale). Here, we discuss a few points. First, both exponential and power-law distributions are ubiquitous. How can we rationalize this? One perspective is given by switching viewpoint from probabilities to their logarithms, which are commonly expressed in a language of dimensionless cost functions, such as energy$/RT$. There are many forms of energy (e.g., gravitational, magnetic, electrostatic, springs, and interatomic interactions). The ubiquity of the exponential distribution can be seen in terms of the diversity and interchangeability of energies.

A broad swath of physics problems can be expressed in terms of the different types of energy and their ability to combine, add, or exchange with each other in various ways. Here, we indicate that nonexponential distributions, too, can be expressed in a language of costs, particularly those that are shared and are subject to economies of scale. Second, where do we expect exponentials vs. power laws? What sets Eq. **5** apart from typical energy functions in physical systems is that EOS costs are both independent of distance and long-ranged (the joiner particle interacts with all particles in given community). Consequently, when the system size becomes large, due to the absence of a correlation length-scale, the energy of the system does not increase linearly with system size, giving rise to a nonextensive energy function. This view is consistent with the appearance of power laws in critical phenomena, where interactions are effectively long-ranged.

Third, interestingly, the concept of cost-minus-benefit in Eq. **5** can be further generalized, also leading to either Gaussian or stretched-exponential distributions. A Gaussian distribution results when the cost-minus-benefit function grows linearly with cluster size, $\mu_k \sim k$; this would arise if the joiner particle were to pay a tax to each member of a community, and this leads to a total cost of $w_k \sim k^2$ (Eq. **1**). These would be "hostile" communities, leading to mostly very small communities and few large ones, because a Gaussian function drops off even faster with k than an exponential does. An example would be a Coulombic particle of charge q joining a community of k other such charged particles, as in the Born model of ion hydration (55). A stretched-exponential distribution can arise if the joiner particle instead pays a tax to only a subset of the community. For example, in a charged sphere with strong shielding, if only the particles at the sphere's surface interact with the joiner particle, then $\mu_k \sim k^{2/3}$ and $w_k \sim k^{5/3}$, leading to a stretched-exponential distribution. In these situations, EOS can affect the community-size distribution not only through cost-sharing but also through the topology of interactions.

Finally, we reiterate a matter of principle. On the one hand, nonexponential distributions could be derived by using a nonextensive entropy-like quantity, such as those of Tsallis, combined with an extensive energy-like quantity. Here, instead, our derivation is based on using the BGS entropy combined with a nonextensive energy-like quantity. We favor the latter because it is consistent with the foundational premises of Shore and Johnson (37). In short, in the absence of energies or costs, the BGS entropy alone predicts a uniform distribution; any other alternative would introduce bias and structure into p_k that is not warranted by the data. Models based on nonextensive entropies intrinsically prefer larger clusters, but without any basis to justify them. The present treatment invokes the same nature of randomness as when physical particles populate energy levels. The present work provides a cost-like language for expressing various different types of probability distribution functions.

ACKNOWLEDGMENTS. We thank A. de Graff, H. Ge, D. Farrell, K. Ghosh, S. Maslov, and C. Shalizi for helpful discussions, and K. Sneppen, M. S. Shell, and H. Qian for comments on our manuscript. Support for this work was provided by a US Department of Defense National Defense Science and Engineering Graduate Fellowship (to J.P.), the National Science Foundation and Laufer Center (J.P. and K.A.D.), and Department of Energy Grant PM-031 from the Office of Biological Research (to P.D.D.).

1. Mantegna R, Stanley H (1995) Scaling behaviour in the dynamics of an economic index. *Nature* 376:46–49.
2. Zipf GK (1949) *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA).
3. Broder A, et al. (2000) Graph structure in the web. *Comput Netw* 33(1-6):309–320.
4. Newman M (2005) Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46(5):323–351.
5. Clauset A, Shalizi C, Newman M (2009) Power-law distributions in empirical data. *SIAM Rev* 51(4):661–703.
6. Taleb NN (2007) *The Black Swan: The Impact of the Highly Improbable* (Random House, New York).
7. Bremmer I, Keats P (2009) *The Fat Tail: The Power of Political Knowledge for Strategic Investing* (Oxford Univ Press, London).
8. Vázquez A, Flammini A, Maritan A, Vespignani A (2003) Modeling of protein interaction networks. *Complexus* 1(1):38–44.
9. Berg J, Lässig M, Wagner A (2004) Structure and evolution of protein interaction networks: A statistical model for link dynamics and gene duplications. *BMC Evol Biol* 4(1):51–63.
10. Maslov S, Krishna S, Pang TY, Sneppen K (2009) Toolbox model of evolution of prokaryotic metabolic networks and their regulation. *Proc Natl Acad Sci USA* 106(24): 9743–9748.
11. Pang TY, Maslov S (2011) A toolbox model of evolution of metabolic pathways on networks of arbitrary topology. *PLOS Comput Biol* 7(5):e1001137.
12. Leskovec J, Chakrabarti D, Kleinberg J, Faloutsos C, Ghahramani Z (2010) Kronecker graphs: An approach to modeling networks. *J Mach Learn Res* 11:985–1042.
13. Karagiannis T, Le Boudec J-Y, Vojnovic M (2010) Power law and exponential decay of intercontact times between mobile devices. *IEEE Trans Mobile Comput* 9(10):1377–1390.

14. Shou C, et al. (2011) Measuring the evolutionary rewiring of biological networks. *PLOS Comput Biol* 7(1):e1001050.
15. Fortuna MA, Bonachela JA, Levin SA (2011) Evolution of a modular software network. *Proc Natl Acad Sci USA* 108(50):19985–19989.
16. Peterson GJ, Pressé S, Peterson KS, Dill KA (2012) Simulated evolution of protein-protein interaction networks with realistic topology. *PLoS ONE* 7(6):e39052.
17. Pang TY, Maslov S (2013) Universal distribution of component frequencies in biological and technological systems. *Proc Nat Acad Sci* 110(15):6235–6239.
18. Simon H (1955) On a class of skew distribution functions. *Biometrika* 42:425–440.
19. de Solla Price D (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inf Sci* 27(5):292–306.
20. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.
21. Vázquez A (2003) Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys Rev E Stat Nonlin Soft Matter Phys* 67(5 Pt 2):056104.
22. Yook S-H, Jeong H, Barabási A-L (2002) Modeling the Internet's large-scale topology. *Proc Natl Acad Sci USA* 99(21):13382–13386.
23. Capocci A, et al. (2006) Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(3 Pt 2): 036116.
24. Newman ME (2001) Clustering and preferential attachment in growing networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 64(2 Pt 2):025102.
25. Jeong H, Néda Z, Barabási A-L (2003) Measuring preferential attachment in evolving networks. *Europhys Lett* 61(4):567.
26. Poncela J, Gómez-Gardeñes J, Floría LM, Sánchez A, Moreno Y (2008) Complex cooperative networks from evolutionary preferential attachment. *PLoS ONE* 3(6):e2449.
27. Peterson GJ, Pressé S, Dill KA (2010) Nonuniversal power law scaling in the probability distribution of scientific citations. *Proc Natl Acad Sci USA* 107(37):16023–16027.
28. Fisher M (1974) The renormalization group in the theory of critical behavior. *Rev Mod Phys* 46(4):597–616.
29. Yeomans J (1992) *Statistical Mechanics of Phase Transitions* (Oxford Univ Press, New York).
30. Stanley H (1999) Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev Mod Phys* 71(2):S358–S366.
31. Gefen Y, Mandelbrot BB, Aharony A (1980) Critical phenomena on fractal lattices. *Phys Rev Lett* 45(11):855–858.
32. Fisher DS (1986) Scaling and critical slowing down in random-field Ising systems. *Phys Rev Lett* 56(5):416–419.
33. Suzuki M, Kubo R (1968) Dynamics of the Ising model near the critical point. *J Phys Soc Jpn* 24:51–60.
34. Glauber RJ (1963) Time-dependent statistics of the Ising model. *J Math Phys* 4(2): 294–304.
35. Caldarelli G, Capocci A, De Los Rios P, Muñoz MA (2002) Scale-free networks from varying vertex intrinsic fitness. *Phys Rev Lett* 89(25):258702–258705.
36. Deeds EJ, Ashenberg O, Shakhnovich EI (2006) A simple physical model for scaling in protein-protein interaction networks. *Proc Natl Acad Sci USA* 103(2):311–316.
37. Shore J, Johnson R (1980) Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans Inf Theory* 26(1):26–37.
38. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106(4): 620–630.
39. Pressé S, Ghosh K, Lee J, Dill KA (2013) The principles of Maximum Entropy and Maximum Caliber in statistical physics. *Rev Mod Phys* 85(3):1115–1141.
40. Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. *J Stat Phys* 52(1-2): 479–487.
41. Rènyi A (1961) On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, ed Neyman J (Univ of California Press, Berkeley, CA), pp 547–561.
42. Aczél J, Daróczy Z (1975) *On Measures of Information and Their Characterizations* (Academic, New York), Vol 40.
43. Amari S-i (1985) *Differential-Geometrical Methods in Statistic* (Springer, Berlin).
44. Lutz E (2003) Anomalous diffusion and Tsallis statistics in an optical lattice. *Phys Rev A* 67(5):051402.
45. Douglas P, Bergamini S, Renzoni F (2006) Tunable Tsallis distributions in dissipative optical lattices. *Phys Rev Lett* 96(11):110601.
46. Burlaga L, Vinas A (2005) Triangle for the entropic index q of non-extensive statistical mechanics observed by Voyager 1 in the distant heliosphere. *Phys A Stat Mech Appl* 356(2):375–384.
47. Pickup RM, Cywinski R, Pappas C, Farago B, Fouquet P (2009) Generalized spin-glass relaxation. *Phys Rev Lett* 102(9):097202.
48. DeVoe RG (2009) Power-law distributions for a trapped ion interacting with a classical buffer gas. *Phys Rev Lett* 102(6):063001.
49. Plastino A, Plastino A (1995) Non-extensive statistical mechanics and generalized Fokker-Planck equation. *Phys A Stat Mech Appl* 222(1):347–354.
50. Tsallis C, Bukman DJ (1996) Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 54(3):R2197–R2200.
51. Caruso F, Tsallis C (2008) Nonadditive entropy reconciles the area law in quantum systems with classical thermodynamics. *Phys Rev E Stat Nonlin Soft Matter Phys* 78(2 Pt 1):021102.
52. Abe S (2000) Axioms and uniqueness theorem for Tsallis entropy. *Phys Lett A* 271(1): 74–79.
53. Gell-Mann M, Tsallis C (2004) *Nonextensive Entropy: Interdisciplinary Applications: Interdisciplinary Applications* (Oxford Univ Press, New York).
54. Dill K, Bromberg S (2010) *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience* (Garland Science, New York), 2nd Ed.
55. Born M (1920) [Volumes and heats of hydration of ions]. *Z Phys* 1:45–48. German.