2008

On estimating the exponent of power-law frequency distributions

Ethan P. White
Utah State University

B. J. Enquist

J. L. Green

Follow this and additional works at: http://digitalcommons.usu.edu/biology_facpub

Part of the Biology Commons

Recommended Citation

1   ON ESTIMATING THE EXPONENT OF POWER-LAW FREQUENCY DISTRIBUTIONS

2   ETHAN P. WHITE,[1,2,4] BRIAN J. ENQUIST,[2] AND JESSICA L. GREEN[3]

3

4   [1]*Department of Biology and the Ecology Center, Utah State University, Logan, UT, 84322*

5   [2]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721*

6   [3]*Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR 97403*

7   [4] E-mail: epwhite@biology.usu.edu

8

9   Abstract: Power-law frequency distributions characterize a wide array of natural phenomena. In

10   ecology, biology, and many physical and social sciences, the exponents of these power-laws are

11   estimated to draw inference about the processes underlying the phenomenon, to test theoretical

12   models, and to scale up from local observations to global patterns. Therefore, it is essential that

13   these exponents be estimated accurately. Unfortunately, the binning-based methods traditionally

14   utilized in ecology and other disciplines perform quite poorly. Here we discuss more

15   sophisticated methods for fitting these exponents based on cumulative distribution functions and

16   maximum likelihood estimation. We illustrate their superior performance at estimating known

17   exponents and provide details on how and when ecologists should use them. Our results confirm

18   that maximum likelihood estimation out-performs other methods in both accuracy and precision.

19   Because of the use of biased statistical methods for estimating the exponent, the conclusions of

20   several recently published papers should be revisited.

21

1   INTRODUCTION

2   Power-laws have a long history in ecology and other disciplines (Bak 1996, Brown et al.

3   2002, Newman 2005). Power-law relationships appear in a wide variety of physical, social, and

4   biological systems and are often cited as evidence for fundamental processes that underlie the

5   dynamics structuring these systems (Bak 1996, Brown et al. 2002, Newman 2005). There are two

6   major classes of power-laws commonly reported in the ecological literature. The first are bi-

7   variate relationships between two variables. Examples of this type of relationship include the

8   species-area relationship and body-size allometries. Standard approaches to analyzing this type

9   of data are generally reasonable and discussions of statistical issues related to this kind of data

10   are presented elsewhere (e.g., Warton et al. 2006). The second type of power-law, and the focus

11   of this paper, is the frequency distribution, where the frequency of some event (e.g., the number

12   of individuals) is related to the size, or magnitude, of that event (e.g., the size of the individual).

13   Frequency distributions of a wide variety of ecological phenomena tend to be, at least

14   approximately, power-law distributed. These phenomena include distributions of species body

15   sizes (Morse et al. 1985), individual body sizes (Enquist and Niklas 2001), colony sizes (Jovani

16   and Tella 2007), abundance among species (Pueyo 2006), trends in abundance of species through

17   time (Keitt and Stanley 1998), step lengths in animal search patterns (i.e., Levy flights; Reynolds

18   et al. 2007), fire magnitude (Turcotte et al. 2002), island size (White and Brown 2005), lake size

19   (Wetzel 1991), flood magnitude (Malamud and Turcotte 2006), landslide magnitude (Guzzetti et

20   al. 2002), vegetation patch size (Kefi et al. 2007), and fluctuations in metabolic rate (Labra et al.

21   2007). Frequency distributions are usually displayed as simple histograms of the quantity of

22   interest. If a distribution is well characterized by a power-law then the frequency of an event

23   (e.g., the number of individuals with mass between 10 and 20 grams), f, is related to the size of

1   that event, x, by a function of the form,

2     $$f(x) = cx^{\lambda},\qquad\qquad\qquad (1)$$

3   where c and λ are constants, and λ is called the exponent and is typically negative (i.e., $\lambda < 0$).

4   Because $f(x)$ is a probability density function (PDF) the value of c is a simple function of λ and

5   the minimum and maximum values of x (Table 1). The specific form of the PDF depends on

6   whether the data are continuous or discrete, on the presence of minimum and maximum values,

7   and on whether λ is < -1 or > -1. The different forms are often given distinct names for clarity

8   (see Table 1).

9   There is substantial interest in using the parameters of these power-laws distributions to

10   make inferences about the processes underlying the distributions, to test mechanistic models, and

11   to estimate and predict patterns and processes operating beyond the scope of the observed data.

12   For example, power-law species abundance distributions with $\lambda \approx -1$ are considered to represent

13   evidence for the primary role of stochastic birth-death processes, combined with species input, in

14   community assembly (Pueyo 2006, Zillio and Condit 2007); quantitative models of tree size

15   distributions make specific predictions (e.g., $\lambda = -2$; Enquist and Niklas 2001) that can be used to

16   test these models (Coomes et al. 2003, Muller-Landau et al. 2006); and power-law frequency

17   distributions of individual size have been used to scale up from individual observations to

18   estimate ecosystem level processes (Enquist et al. 2003, Kerkhoff and Enquist 2006).

19   One concern when interpreting the exponents of these distributions is that there are a wide

20   variety of different approaches currently being used to estimate the exponents (Sims et al. 2007,

21   White et al. 2007). These include techniques based on: 1) binning (e.g., Enquist and Niklas 2001,

22   Meehan 2006, Kefi et al. 2007); 2) the cumulative distribution function (e.g., Rinaldo et al.

23   2002); and 3) maximum likelihood estimation (e.g., Muller-Landau et al. 2006, Zillio and Condit

1   2007, Edwards et al. 2007). There has been little discussion in the ecological literature of how

2   the choice of methodology influences the parameter estimates, and methods other than binning

3   are rarely used. If different methods produce different results this could have important

4   consequences for the conclusions drawn about the ecology of the system (Edwards et al. 2007,

5   Sims et al. 2007).

6   Here we: 1) describe the different approaches used to quantify the exponents of power-law

7   frequency distributions; 2) show that some of these approaches give biased estimates; 3)

8   illustrate the superior performance of some approaches using Monte Carlo methods; 4) make

9   recommendations for best estimating parameters of power-law distributed data; and 5) show that

10   some of the conclusions of recent studies are effected by the use of biased statistical techniques.

11   METHODS FOR ESTIMATING THE EXPONENT

12   *Linear Binning*

13   Perhaps the most intuitive way to quantify an empirical frequency distribution is to bin the

14   observed data using bins of constant linear width. This generates the familiar histogram.

15   Specifically, linear binning entails choosing a bin i of constant width ($w = x_{i+1} - x_i$), counting the

16   number of observations in each bin (i.e., with values of x between x_i and $x_i + w$), and plotting this

17   count against the value of x at the center of the bin ($x_i/2 + x_{i+1}/2$). If the counts are divided by the

18   sum of all the counts, this plot is an estimate of the probability density function, $f(x)$. The

19   traditional approach to estimating the power-law exponent is to fit a linear regression to log

20   transformed values of $f(x)$ and x, with the slope of the line giving an estimate of the exponent, λ.

21   Bins with 0 observations are excluded (because log(0) is undefined) and sometimes bins with

22   low counts are also excluded (e.g., Enquist and Niklas 2001). While in practice the choice of bin

23   width is normally arbitrary, this choice represents a tradeoff between the number of bins

1   analyzed (i.e., the resolution of the frequency distribution) and the accuracy with which each

2   value of $f(x)$ is estimated (fewer observations/bin provide a poorer density estimate; Pickering et

3   al. 1995).

4   *Logarithmic Binning*

5   *Simple logarithmic binning* – This approach is similar to linear binning, except that instead of

6   the bins having constant linear width, they have constant logarithmic width, $b = \log(x_{i+1})\text{-}\log(x_i)$.

7   The estimate of λ is obtained by log-transforming the values of x and following the procedure

8   described above. Since the x data are transformed to begin with, it is not necessary to transform

9   the bin centers again prior to fitting the regression. For power-law like distributions, an

10   advantage of logarithmic binning is the reduction of the number of zero and low count bins at

11   larger values x because the linear width of a bin increases linearly with x; i.e. $w_i = x_i(e^b - 1)$.

12   However, this means that the number of observations within each bin is determined not only by

13   x, but also by the linear width of the bin. Therefore, the slope of the regression will give an

14   estimate of $\lambda+1$, not λ (Appendix A, Han and Straskraba 1998, Bonnet et al. 2001, Sims et al.

15   2007).

16   *Normalized logarithmic binning* – The problem of increasing linear width of logarthimic bins

17   can be dealt with by normalizing the number of observations in each bin by the linear width of

18   the bin, w. This converts the counts into densities (number of observations per unit of x) (Bonnet

19   et al. 2001, Christensen and Moloney 2005). The linear width of a logarithmic bin can be

20   calculated as $x_i(e^b - 1)$ (Appendix A). This normalization approach is typically used in the

21   characterization of aquatic size-spectra and power-law distributions in physics (Kerr and Dickie

22   2001, Christensen and Moloney 2005). It removes the artifact from traditional logarithmic

23   binning while maintaining the advantage of using larger bins where there are fewer values of x.

1   An alternative approach is to use simple logarithmic binning and subtract one from the estimated

2   exponent (Han and Straskraba 1998, Bonnet et al. 2001).

3   *Fitting the Cumulative Distribution Function*

4   An alternative to binning methods is to work with the cumulative distribution function

5   (CDF):

6
$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(x)dx$$

7   The CDF describes the probability that a random variable, *X*, drawn from $f(x)$ is $\le x$. The CDF is

8   straightforward to construct for a set of observed data – and no binning is required. To construct

9   the CDF, first rank the *n* observed values (x_i) from smallest to largest ($i = 1…n$). The probability

10   that an observation is less than or equal to x_i (the CDF) is then estimated as i/n (this is the

11   Kaplan-Meier estimate, Evans et al. 2000). Analyzing the CDF avoids the subjective influence of

12   the choice of bin width and the problem of empty bins. Having determined the CDF for a power-

13   law distribution, the exponent, λ, of the PDF can be estimated using regression. The traditional

14   approach is to transform the equation for the CDF such that the slope of a linear equation is a

15   function of λ. The linearized equation differs among distributions (Appendix A). The slope of the

16   regression will be equal to $\lambda+1$, making it necessary to subtract 1 to obtain λ (Bonnet et al. 2001,

17   Rinaldo et al. 2002).

18   *Maximum Likelihood Estimation*

19   Maximum likelihood estimation (MLE) is one of the preferred approaches for estimating

20   frequency distribution parameters (e.g., Rice 1994). MLE determines the parameter values that

21   maximize the likelihood of the model (in this case, a power-law with an unknown exponent)

22   given the observed data. Specifically, MLE finds the value of λ that maximizes the product of the

23   probabilities of each observed value of *x* (i.e., the product of $f(x)$ evaluated at each data point; see

1   Rice (1994) for a good introduction to maximum likelihood methods). The specific solution for

2   the maximum likelihood estimate of λ and whether the solution is closed form or requires

3   numerical methods to solve depends on the minimum and maximum values of x and on the value

4   of λ (Table 1). Alternatively, the likelihood can be maximized directly using numerical methods

5   (Clauset et al. 2007, Zillio and Condit 2007). While MLE does not provide an opportunity for

6   visual inspection of the distribution to determine if the assumption of the power-law functional

7   form is reasonable, the validity of this assumption can be assessed using simple goodness of fit

8   tests such as the Chi-square on binned data (Clark et al. 1999, Clauset et al. 2007, Edwards et al.

9   2007), or by visually assessing the linearity of binned data, or the CDF (Benhamou 2007), under

10   the appropriate transformation.

11   <div align="center">COMPARING THE METHODS</div>

12   While uncorrected simple logarithmic binning clearly provides incorrect estimates of λ, the

13   alternative approaches discussed above all seem reasonable and intuitive. However, the different

14   approaches do not perform equally well, and some produce biased estimates of the exponent

15   (e.g., Pickering et al. 1995, Clark et al. 1999, Sims et al. 2007). We applied Monte Carlo

16   methods to illustrate the advantages and disadvantages of the various approaches and to explore

17   cases relevant to ecology that have not been previously addressed. Monte Carlo methods

18   generate data that are, by definition, power-law distributed with known exponents, making it

19   possible to compare the performance of the different techniques in estimating the value of λ.

20   We generated power-law distributed random numbers using the inverse transformation

21   method for the Pareto distribution (Ross 2006), and using the rejection method for the discrete

22   Pareto distribution (Devroye 1986). Each analysis consisted of the following: 1) generating

23   10,000 Monte Carlo datasets for each point in the analysis (e.g., for each sample size), 2)

1   estimating the exponent for each dataset using the methods described above, 3) evaluating the

2   distribution of exponents generated by each method, and, 4) comparing the performance of the

3   methods based on bias (i.e., accuracy) and on the variance in the estimate (i.e., precision). We

4   report on simulated distributions generated using $\lambda = -2$ and $a = 1$. The results for other

5   combinations of parameters are qualitatively similar. We also evaluated the influence of sample

6   size on the various estimation techniques, and for binning-based approaches we evaluated the

7   effect of bin width on the analysis.

8   GENERAL RULES

9   *Uncorrected simple logarithmic binning gives the wrong exponent* – Non-normalized

10   logarithmic binning does not estimate λ; it estimates $\lambda+1$ (Han and Straskraba 1998, Bonnet et al.

11   2001, Sims et al. 2007). Therefore if simple logarithmic binning is used, and an estimate of λ is

12   the desired result, then it is necessary to subtract one from the slope of the logarithmically binned

13   data. Not doing so will give the wrong value for the exponent.

14   *Binning-based approaches perform poorly* – Linear binning performs poorly by practically

15   any measure. In most cases it produces biased estimates of the exponent and its estimates are

16   highly variable (Figs. 1 and 2). In addition, the estimated exponent is highly dependent on the

17   choice of bin width, and this dependency varies as a function of sample size (Fig. 3). While

18   normalized logarithmic binning performs better than linear binning, its estimates are also

19   dependent on the choice of bin width and are more variable than alternate approaches. Our

20   results are based on recommended practices in binning analyses (following Pickering et al.

21   1995). Many alternative approaches to constructing bins and performing regressions on binned

22   data are conceivable, and it is possible that some of these may improve the performance of the

23   estimates. However, this highlights the fact that binning-based methods are sensitive to a variety

1   of decisions, and it appears that no amount of tweaking will be able to produce a consistent

2   binning-based method for estimating the exponent. In general, binning results in a loss of

3   information about the distributions of points within a bin and is thus expected to perform poorly

4   (Clauset et al. 2007, Edwards et al. 2007). Therefore, while binning is useful for visualizing the

5   frequency distribution, and normalized logarithmic binning performs well at this task, binning-

6   based approaches should be avoided for parameter estimation (Clauset et al. 2007).

7   *Maximum likelihood estimation performs best* – While fitting the CDF generally produces

8   good results, estimates of λ using the CDF approach are often biased at small sample sizes and

9   are consistently more variable than those using MLE (Fig. 2; Clark et al. 1999, Newman 2005).

10   This probably results because the logarithmic transformation used in fitting the CDF weights a

11   small number of points more heavily, and because the points in the CDF are not independent

12   thus violating regression assumptions (see Clauset et al. 2007 for other issues with regression

13   based approaches). While alternative approaches to fitting the CDF (e.g., non-linear regression)

14   could improve the performance of this estimator, MLE has been shown mathematically to be the

15   single best approach for estimating power-law exponents (i.e., it is the minimum variance

16   unbiased estimator, Johnson et al. 1994, Clark et al. 1999, Newman 2005). In addition, MLE

17   produces valid confidence intervals for the estimated exponent (Appendix A), which the other

18   methods do not (Clark et al. 1999, Newman 2005, Clauset et al. 2007).

19   COMPLICATIONS

20   *Minimum and maximum values* – Minimum and maximum attainable values of ecological

21   quantities can result either from natural limits on the quantity being measured (e.g., trees cannot

22   grow above some maximum size), or from methodological limits on the values that can be

23   observed (e.g., fires <1 ha are not recorded). In addition, the power-law form of the distribution

1   may not hold over the entire range of x, making it necessary to select a restricted range of x on

2   which to estimate the exponent. While binning-based approaches do not assume particular limits

3   on x (but see Pickering et al. 1995), CDF and MLE approaches assume the minimum and

4   maximum attainable values of x given in Table 1. In some cases these limits may be known, but

5   if not it may be necessary to estimate them (e.g., Kijko 2004, Clauset et al. 2007). Because

6   maximum likelihood estimation for the truncated Pareto requires numerical methods, it has been

7   suggested that in some cases with both a minimum and maximum value that the error introduced

8   by assuming that there is no maximum is small enough that it is reasonable to estimate the

9   exponent using the maximum likelihood estimate for the Pareto distribution. Clark et al. (1999)

10   suggest this approximation in cases where the maximum value is at least two orders of

11   magnitude greater than the minimum, i.e., $\max(x) > 100 \times \min(x)$.

12   *Deviations from the power-law* – Empirical data are rarely perfectly power-law distributed

13   over the entire range of x (Brown et al. 2002, Newman 2005). MLE and CDF approaches

14   respond to deviations differently because the traditional MLE analysis implicitly weights data on

15   a linear scale while the traditional CDF approach weights it on a logarithmic scale (McGill

16   2003). The CDF approach will therefore respond more strongly to deviations from the power-law

17   at large values of x (such as those observed in individual size distributions; e.g., Coomes et al.

18   2003) than the MLE approach, whereas MLE will respond more strongly to deviations at small

19   values of x (commonly observed in many power-law distributions; e.g., Newman 2005). It is

20   common to truncate data in the tails that exhibits deviations from the power-law before fitting

21   the exponent (e.g., Newman 2005). However, these deviations should also not be ignored, as

22   they may help identify important biological processes (e.g., Coomes et al. 2003). In some cases

23   deviations may suggest that the power-law is in fact not the appropriate model for the data. This

1 can be evaluated using goodness of fit tests on binned data (Clark et al. 1999, Clauset et al. 2007,

2 Edwards et al. 2007) or by using model selection techniques (e.g., Burnham and Anderson 2002)

3 to compare the power-law to alternative distributions (Muller-Landau et al. 2006, Clauset et al.

4 2007, Edwards et al. 2007).

5   *Discrete data* – Most of the MLE and CDF methods presented here assume that the data are

6 continuously distributed, as is often the case (e.g., body size). However, some ecological patterns

7 (e.g., species-abundance distributions) are comprised of discrete observations (e.g., it is

8 impossible to census 4.3 individuals). It is therefore necessary to use analogous discrete

9 distributions. In the case of the Pareto distribution a discrete analog exists in the form of the aptly

10 named discrete Pareto distribution (Johnson et al. 2005, Newman 2005) (Table 1; also called the

11 Zipf or Riemann-zeta distribution). In some cases continuous distributions can reasonably

12 approximate discrete data; but in the case of the Pareto, using the continuous maximum

13 likelihood estimate instead of that derived from the discrete distribution produces strongly biased

14 results and should be avoided (Appendix C, Clauset et al. 2007).

15   IMPLICATIONS FOR PUBLISHED RESULTS

16   One of the most important implications for published results is that studies that estimated

17 exponents using uncorrected simple logarithmic binning (e.g., Morse et al. 1988, Meehan 2006)

18 have reported the wrong exponent. This is particularly important in cases where the exponent is

19 used to test quantitative predictions. For example, an analysis in Meehan (2006) evaluates

20 whether observed individual size distribution exponents were consistent with those predicted,

21 using simple logarithmic binning. Meehan concluded that the empirical data matched the

22 predictions (Fig. 4a). However, since the reported exponents are equal to $\lambda + 1$, the analysis

23 suggests that the size distribution is substantially steeper than expected, thus refuting rather than

1   supporting the hypothesized mechanism (Fig. 4a; Appendix B).

2   Analyses based on linearly binned data should also be revisited due to the potential for

3   biased estimates and the strong influence of bin-width on the estimated exponent. In particular,

4   studies that have used linear binning to test the predictions of theoretical models or compare

5   exponents from different datasets (e.g., Enquist and Niklas 2001, Coomes et al. 2003, Niklas et

6   al. 2003, Kefi et al. 2007) may have reached incorrect conclusions. We reanalyzed the original

7   data from Enquist and Niklas (2001) and found that while the original linear binning analyses

8   suggested that observed diameter distribution exponents were near the theoretical prediction of -

9   2, MLE suggests that the observed exponents are actually closer, on average, to -2.5 (Fig. 4b;

10   Appendix B). Our reanalysis indicates that the size-frequency distributions in Gentry's plots are

11   not, in general, adequately represented by a power law with an exponent of -2, as originally

12   claimed by Enquist and Niklas (2001) (see Appendix B for an important caveat).

13   While normalized logarithmic binning performs better than linear binning, it can still

14   introduce biases of ~10% depending on the bin width. While many analyses based on normalized

15   logarithmic binning are probably reasonable, the recent suggestion that normalized logarithmic

16   binning is the best approach for fitting exponents (Sims et al. 2007) is unwarranted, and MLE

17   should be used whenever possible (Clark et al. 1999, Clauset et al. 2007).

18   Compared to binning-based approaches, results from fitting the CDF are probably

19   reasonable. In cases with low sample sizes, where small errors in the estimated exponent could

20   influence the conclusions of the study, or where minimum or maximum attainable values of x

21   have been ignored (see Pickering et al. 1995), it may be worth checking the results using MLE.

22   Regardless, MLE is the single best method for estimating exponents and should be used in future

23   studies.

1   CONCLUSIONS

2   The vast majority of ecological studies that estimate exponents for power-law like

3 distributions use approaches based on binning the empirical data (e.g., Morse et al. 1988, Enquist

4 and Niklas 2001, Coomes et al. 2003, Niklas et al. 2003, Meehan 2006, Jovani and Tella 2007,

5 Kefi et al. 2007, Reynolds et al. 2007, Sims et al. 2007). These binning based methods tend to

6 produce results that are biased, have high variance, and are contingent on the choice of bin

7 width. Instead of binning, maximum likelihood estimation should be used when fitting power-

8 law exponents to empirical data (Clark et al. 1999, Newman 2005, Edwards et al. 2007).

9   We have focused on power-laws because they, at least approximately, characterize a

10 number of distributions of interest to ecologists. The issues raised here, and the conclusions

11 discussed, should apply broadly to frequency distributions in general, and in particular to other

12 distributions with heavy tails. Paying careful attention to fitting methodologies and consultation

13 of statistical references (e.g., Johnson et al. 1994) should help improve the estimation of

14 distributional parameters.

1   LITERATURE CITED

2   Bak, P. 1996. How Nature Works: The Science of Self-Organized Criticality. Springer-Verlag,

3   New York.

4   Benhamou, S. 2007. How many animals really do the Levy walk? Ecology **88**:1962-1969.

5   Bonnet, E., O. Bour, N. E. Odling, P. Davy, I. Main, P. Cowie, and B. Berkowitz. 2001. Scaling

6   of fracture systems in geological media. Reviews of Geophysics **39**:347-383.

7   Brown, J. H., V. K. Gupta, B. L. Li, B. T. Milne, C. Restrepo, and G. B. West. 2002. The fractal

8   nature of nature: power laws, ecological complexity and biodiversity. Philosophical

9   Transactions of the Royal Society of London Series B-Biological Sciences **357**:619-626.

10   Christensen, K., and N. R. Moloney. 2005. Complexity and Criticality. Imperial College Press,

11   London.

12   Clark, R. M., S. J. D. Cox, and G. M. Laslett. 1999. Generalizations of power-law distributions

13   applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats.

14   Geophysical Journal International **136**:357-372.

15   Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2007. Power-law distributions in empirical

16   data. arXiv:0706.1062v1 [physics.data-an].

17   Coomes, D. A., R. P. Duncan, R. B. Allen, and J. Truscott. 2003. Disturbances prevent stem

18   size-density distributions in natural forests from following scaling relationships. Ecology

19   Letters **6**:980-989.

20   Devroye, L. 1986. Non-Uniform Random Variate Generation. Springer-Verlag, New York.

21   Edwards, A. M., R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S.

22   V. Buldyrev, M. G. E. da Luz, E. P. Raposo, H. E. Stanley, and G. M. Viswanathan. 2007.

23   Revisiting Levy flight search patterns of wandering albatrosses, bumblebees, and deer. Nature

449:1044-1048.

Enquist, B., E. Economo, T. Huxman, A. Allen, D. Ignace, and J. Gillooly. 2003. Scaling metabolism from organisms to ecosystems. Nature **423**:639-642.

Enquist, B. J., and K. J. Niklas. 2001. Invariant scaling relations across tree-dominated communities. Nature **410**:655-660.

Evans, M., N. Hastings, and B. Peacock. 2000. Statistical Distributions, third edition. John Wiley & Sons, Inc., New York.

Guzzetti, F., B. D. Malamud, D. L. Turcotte, and P. Reichenbach. 2002. Power-law correlations of landslide areas in central Italy. Earth and Planetary Science Letters **195**:169-183.

Han, B. P., and M. Straskraba. 1998. Size dependence of biomass spectra and population density - I. The effects of size scales and size intervals. Journal of Theoretical Biology **191**:259-265.

Johnson, N. L., A. W. Kemp, and S. Kotz. 2005. Univariate Discrete Distributions, 3rd edition. Wiley-Interscience, New York.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. Continuous Univariate Distributions: Volume 1, 2nd edition. Wiley-Interscience, New York.

Jovani, R., and J. L. Tella. 2007. Fractal bird nest distribution produces scale-free colony sizes. Proceedings of the Royal Society of London Series B-Biological Sciences **274**:2465-2469.

Kefi, S., M. Rietkerk, C. L. Alados, Y. Pueyo, V. P. Papanastasis, A. ElAich, and P. C. de Ruiter. 2007. Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. Nature **449**:213-217.

Keitt, T. H., and H. E. Stanley. 1998. Dynamics of North American breeding bird populations. Nature **393**:257-260.

Kerkhoff, A. J., and B. J. Enquist. 2006. Ecosystem allometry: the scaling of nutrient stocks and

1   primary productivity across plant communities. Ecology Letters **9**:419-427.

2   Kerr, S. R., and L. M. Dickie. 2001. Biomass Spectrum. Columbia University Press, New York.

3   Kijko, A. 2004. Esimtation of the maximum earthquak magnitudes, mmax. Pure and Applied

4   Geophysics **161**:1655-1681.

5   Labra, F. A., P. A. Marquet, and F. Bozinovic. 2007. Scaling metabolic rate fluctuations.

6   Proceedings of the National Academy of Sciences of the United States of America **104**:10900-

7   10903.

8   Malamud, B. D., and D. L. Turcotte. 2006. The applicability of power-law frequency statistics to

9   floods. Journal of Hydrology **322**:168-180.

10 McGill, B. 2003. Strong and weak tests of macroecological theory. Oikos **102**:679-685.

11 Meehan, T. D. 2006. Energy use and animal abundance in litter and soil communities. Ecology

12   **87**:1650-1658.

13 Morse, D. R., J. H. Lawton, M. M. Dodson, and M. H. Williamson. 1985. Fractal dimension of

14   vegetation and the distribution of arthropod body lengths. Nature **314**:731-733.

15 Morse, D. R., N. E. Stork, and J. H. Lawton. 1988. Species number, species abundance and body

16   length relationships of arboreal beetles in bornean lowland rain-forest trees. Ecological

17   Entomology **13**:25-37.

18 Muller-Landau, H. C., R. S. Condit, K. E. Harms, C. O. Marks, S. C. Thomas, S.

19   Bunyavejchewin, G. Chuyong, L. Co, S. Davies, R. Foster, et al. 2006. Comparing tropical

20   forest tree size distributions with the predictions of metabolic ecology and equilibrium models.

21   Ecology Letters **9**:589-602.

22 Newman, M. E. J. 2005. Power laws, Pareto distributions and Zipf's law. Contemporary Physics

23   **46**:323-351.

1   Niklas, K. J., J. J. Midgley, and R. H. Rand. 2003. Tree size frequency distributions, plant

2   density, age and community disturbance. Ecology Letters **6**:405-411.

3   Page, R. 1968. Aftershocks and microaftershocks of the great Alaska earthquake of 1964.

4   Bulletin of the Seismological Society of America **58**:1131-1168.

5   Pickering, G., J. M. Bull, and D. J. Sanderson. 1995. Sampling power-law distributions.

6   Tectonophysics **248**:1-20.

7   Pueyo, S. 2006. Diversity: between neutrality and structure. Oikos **112**:392-405.

8   Rice, J. A. 1994. Mathematical Statistics and Data Analysis. Duxbury Press, Pacific Grove, CA.

9   Rinaldo, A., A. Maritan, K. K. Cavender-Bares, and S. W. Chisholm. 2002. Cross-scale

10   ecological dynamics and microbial size spectra in marine ecosystems. Proceedings of the

11   Royal Society of London Series B-Biological Sciences **269**:2051-2059.

12 Reynolds, A. M., A. D. Smith, R. Menzel, U. Greggers, D. R. Reynolds, and J. R. Riley. 2007.

13   Displaced honey bees perform optimal scale-free search flights. Ecology **88**:1955-1961.

14 Ross, S. 2006. A First Course in Probability, Seventh edition. Pearson Prentice Hall, Upper

15   Saddle River, New Jersey.

16 Sims, D. W., D. Righton, and J. W. Pitchford. 2007. Minimizing errors in identifiying Levy

17   flight behavior of organisms. Journal of Animal Ecology **76**:222-229.

18 Tscharntke, T., M. E. Hochberg, T. A. Rand, V. H. Resh, and J. Krauss. 2007. Author sequence

19   and credit for contributions in multiauthored publications. PLOS Biology **5**:e18.

20 Turcotte, D. L., B. D. Malamud, F. Guzzetti, and P. Reichenbach. 2002. Self-organization, the

21   cascade model, and natural hazards. Proceedings of the National Academy of Sciences of the

22   United States of America **99**:2530-2537.

23 Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby. 2006. Bivariate line-fitting methods

1   for allometry. Biological Reviews **81**:259-291.

2   Wetzel, R. G. 1991. Land-water interfaces: metabolic and limnological regulators.

3   Verhandlungen Internationale Vereinigung für Theoretische und Angewandte Limnologie

4   **24**:6-24.

5   White, E. P., and J. H. Brown. 2005. The template: patterns and processes of spatial variation.

6   Pages 31-47 *in* G. M. Lovett, C. G. Jones, M. G. Turner, and K. C. Weathers, editors.

7   Ecosystem Function in Heterogeneous Landscapes. Springer, New York.

8   White, E. P., S. K. M. Ernest, A. J. Kerkhoff, and B. J. Enquist. 2007. Relationships between

9   body size and abundance in ecology. Trends in Ecology and Evolution.

10   Zillio, T., and R. Condit. 2007. The impact of neutrality, niche differentiation and species input

11   on diversity and abundance distributions. Oikos **116**:931-940.

Table 1. Descriptions of different power-law frequency distributions, including the name of the distribution, the range of data and parameter values over which it applies, its probability density function (or probability mass function), $f(x)$, its cumulative distribution function, $F(x)$, and the maximum likelihood estimate (MLE) for λ based on the PDF[1]. The minimum value of x for which a distribution is valid is given by a, which is defined to be greater than 0. The maximum value of x for which a distribution is valid is given by b, which is defined to be less than infinity.

Distribution	$f(x)$	$F(x)$	MLE for λ
(1) Pareto Range $a \leq x < \infty$ Parameters $\lambda < -1,\, a > 0$	$-(\lambda+1)a^{-(\lambda+1)}x^{\lambda}$	$1-a^{-(\lambda+1)}x^{\lambda+1}$	$\hat{\lambda} = -1 - \left[\dfrac{1}{n}\sum_{i=1}^{n} \log\left(\dfrac{x_i}{a}\right) \right]^{-1}$
(2) Truncated Pareto[2] Range $a \leq x \leq b$ Parameters $\lambda \neq -1,\, a \geq 0,\, b \geq 0$	$(\lambda+1)\left(b^{\lambda+1}-a^{\lambda+1}\right)^{-1}x^{\lambda}$	$\dfrac{x^{\lambda+1}-a^{\lambda+1}}{b^{\lambda+1}-a^{\lambda+1}}$	$\overline{\ln x} = \dfrac{-1}{\left(\hat{\lambda}+1\right)} + \dfrac{b^{\hat{\lambda}+1}\ln b - a^{\hat{\lambda}+1}\ln a}{b^{\hat{\lambda}+1}-a^{\hat{\lambda}+1}}$
(3) Discrete Pareto[2,3] Range $x = a, a+1, a+2, \dots \infty$ Parameters $\lambda < -1,\, a \geq 1$	$\dfrac{x^{\lambda}}{\zeta(-\lambda,a)}$	$\dfrac{\sum_{j=a}^{x} j^{\lambda}}{\zeta(-\lambda,a)}$	$\overline{\ln x} = \dfrac{-\zeta'\left(-\hat{\lambda},a\right)}{\zeta\left(-\hat{\lambda},a\right)}$

(4) Power Function[4]			
Range $0 \le x \le b$	$(\lambda+1)b^{-(\lambda+1)}x^{\lambda}$	$(x/b)^{\lambda+1}$	$\hat{\lambda} = \left[\log(b) - \frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right]^{-1} - 1$
Parameters $\lambda > -1, b > 0$			

[1]Sources: Pareto (Johnson et al. 1994); Truncated Pareto (Page 1968); Discrete Pareto (Clauset et al. 2007); Power Function (Evans et al. 2000). There is an error in the MLE solution given by Evans et al. (2000) that has been corrected. Note that MLEs are only guaranteed to be minimum variance unbiased estimators in the limit of large n. If n is small, corrections to the MLE are available (Johnson et al. 1994, Clark et al. 1999, Clauset et al. 2007). All solutions assume that a and b are known.

[2]The MLE equations for these distributions cannot be solved analytically for $\hat{\lambda}$, so they must be solved using numerical methods such as bracketing and bisection.

[3] $\zeta(\lambda,a) = \sum_{k=0}^{\infty}(k+a)^{-\lambda}$ is the generalized zeta function and $\zeta'(\lambda,a)$ is its derivative with respect to λ.

[4]The Power Function distribution is often ignored in discussions of power-law distributions because it rarely occurs in natural systems (Newman 2005, Clauset et al. 2007). We include it here for completeness and because it has been suggested that in some groups individual size distributions based on mass may be approximately power-law distributed with $\lambda > -1$ (e.g., Enquist and Niklas 2001).

Fig. 1. Example of Monte Carlo results for the different methods of fitting the power-law exponent. (a) A single Monte Carlo sample from a Pareto distribution plotted as one minus the cumulative distribution. Data are plotted as gray circles along with the fits to the data using the four different methods: linear binning (red; Linear), normalized logarithmic binning (blue; Nlog), cumulative distribution function fitting (black; CDF), and maximum likelihood estimation (green; MLE). (b) Kernel density estimates of the distribution of exponents from 10,000 Monte Carlo runs. Line colors are the same as for (a) and the value of λ used to generate the data is indicated by the dashed line. Parameter values were n = 500, $\lambda = -2$, $1 \leq x < \infty$, linear bin width = 3, logarithmic bin width = 0.3, and the binning analyses utilized a minimum value of x and excluded the last bin and bins containing ≤ 1 individual. Exclusion of the last bin is not necessary, but improves the performance of binning based approaches and is thus conservative in the context of our conclusions. The single sample for (a) was chosen to illustrate the general results shown in (b). Binning methods generate biased estimates of the exponent and result in more variable estimates than MLE and CDF based approaches.

Fig. 2. Effect of sample size on the mean estimated exponent (a) and the variance of that exponent (b), for the four estimation methods: linear binning (red), normalized logarithmic binning (blue), cumulative distribution function fitting (black solid), and maximum likelihood estimation (green). Values for each sample size were generated using 10,000 Monte Carlo runs from the Pareto distribution with parameter values: $\lambda = -2$ (black dashed), $1 \leq x < \infty$, linear bin width = 7.5, logarithmic bin width = 0.75. Other binning methods as in Figure 1. Linear binning fails to converge to the correct estimate. While the other methods all appear to converge at large sample sizes, maximum likelihood estimation always yields the lowest variance in the estimated exponent.

Fig. 3. Effect of bin width on the estimated exponents for linear (a) and normalized logarithmic (b) binning for three different sample sizes: $n = 200$ (solid black line), $n = 500$ (dashed gray line) and $n = 1000$ (dotted black line); based on 1000 Monte Carlo runs from the Pareto distribution per point. Parameter values were $\lambda = -2$ (dashed black line) and $1 \le x < \infty$. Error bars are ± 2 SEs. Other binning methods as in Figure 1. Changing bin width changes the estimated exponent for all sample sizes.

Fig. 4. Reanalysis of individual size distribution data from (a) Meehan (2006) and (b) Enquist and Niklas (2001) using less biased methods. Plots are probability densities of the estimated exponents using the studies original methodology (dashed line; simple logarithmic binning in Meehan, linear binning in Enquist and Niklas), and using less biased methods (sold line; normalized logarithmic binning for Meehan, MLE for Enquist and Niklas). Both studies purported to support a theoretically derived exponent (dotted line). However, when the data are reanalyzed using a more accurate estimator of the exponent it becomes clear that the observed data deviate significantly from the theoretical prediction.