Notes and Comments

Testing the Power Law Model for Discrete Size Data

Andrew R. Solow,[1,*] Christopher J. Costello,[2,†] and Michael Ward[3,‡]

1. Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543;
2. Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106;
3. Department of Agricultural and Resource Economics, University of California, Berkeley, California 94720

Submitted May 20, 2002; Accepted March 20, 2003; Electronically published November 6, 2003

Keywords: goodness of fit, multinomial distribution, order restriction, parametric bootstrap.

Empirical regularities in biology, as in other fields, can be extremely interesting. In particular, such regularities may suggest the operation of fundamental laws. Unfortunately, apparent regularities sometimes cannot stand up under close scrutiny. An early example in biology is logistic growth. The visually impressive fit of the logistic curve to a wide variety of growth data was taken by some as evidence of a universal law of growth (Kingsland 1995). However, in an unrecognized classic of statistical ecology, Feller (1941) showed that the fit of the logistic curve was no better than that of other sigmoidal models, thereby depriving the logistic model of its universality. Incidentally, despite this, the logistic model has achieved cult status among a small but active group of scientists and engineers (e.g., Marchetti et al. 1996).

Recently, claims have been made about the ubiquity of so-called power laws underlying a variety of size distributions in biology. In the case of units of discrete size, let n_j be the observed number of units of size j, where $j = 1, 2, \ldots$. Roughly speaking, under a power law, n_j varies as $j^{-\gamma}$ for some $\gamma > 1$. Newman (1996) used the fossil record to argue that the frequency of extinction events of different sizes (measured by the number of extinguished taxa) follows a power law (see also Kauffman 1995). Using modern taxonomic data, Burlando (1990, 1993) concluded that the frequency of genera of different sizes (measured by the number of species) also follows a power law. This is actually a revival of an older observation by Willis (1922; see also Yule 1924). At least part of the interest in power laws stems from a potential connection to the notion of self-organized criticality (e.g., Bak 1996).

This note is concerned with the statistical problem of testing the null hypothesis that a collection of discrete size data follows a power law distribution. In a typical application, a power law model is fit by an ordinary least squares regression of $\log n_j$ (where here and below log denotes the natural logarithm) on $\log j$, omitting cases in which $n_j = 0$, and goodness of fit is assessed graphically or through summary statistics such as R^2. In some applications, goodness of fit is assessed more formally through a Kolmogorov-Smirnov or a χ^2 test. It is not uncommon to omit observations that are not consistent with a power law model. Whether inconvenient observations are omitted or not, these approaches are incorrect as formal goodness-of-fit tests. For example, the Kolmogorov-Smirnov test assumes that the power law parameter γ is specified independently of the data when, in fact, it is typically estimated from the data. The χ^2 test breaks down when the number of small expected counts is large, as is typical for power laws, while the Kolmogorov-Smirnov test also assumes that the underlying distribution is continuous, which is not the case when size is discrete. These are technical problems, and possible remedies are discussed in the monograph edited by D'Agostino and Stephens (1986).

In this note, we focus on a more fundamental problem: namely, the absence of an explicit alternative to the power law hypothesis. As in testing the logistic model, the absence of a specified alternative can make it difficult to reject the power law if only because it captures some gross feature of size data. One way to set up an alternative would be to embed the power law in a parametric family of models and test for departures from the power law within this family. Although this approach is appealing for its trac-

* E-mail: asolow@whoi.edu.

† E-mail: costello@bren.ucsb.edu.

‡ E-mail: mward@are.berkeley.edu.

Am. Nat. 2003. Vol. 162, pp. 685–689.

tability, there is no natural family in which to embed the power law. Moreover, the results of this approach may be sensitive to departures from the assumed parametric family. Here, we will test the power law model against the nonparametric alternative that frequency is nonincreasing in size. This choice was motivated by the impression that the apparent goodness of fit of regressions of $\log n_j$ on $\log j$ may be due to a general decline of n_j with j and not to the detailed shape of this decline. In the same way that it is undesirable to accept the logistic model simply because growth is sigmoidal, it would be undesirable to accept a power law simply because frequency declines with size. The idea of testing a parametric model against a nonparametric alternative originated with Azzalini et al. (1989) and is now a well-established practice (Hart 1997).

Methods

To devise a formal test of the power law model, it is necessary to specify a statistical model of the data. We posit an otherwise undescribed stochastic process that, for a particular family, generates m independent genera, each containing a discrete number of species. Let p_j be the probability that a genus generated by this process contains j species ($j = 1, 2, \ldots$), with $0 \le p_j \le 1$ and $\sum_{j=1}^{\infty} p_j = 1$, and let the random variable N_j be the number of genera containing j species ($j = 1, 2, \ldots$). Under this model, the random variables $N_1, N_2, \ldots$ follow a multinomial distribution with m trials and probability vector $p = (p_1, p_2, \ldots)$. Here, we are treating the number m of genera as fixed. If this number is also viewed as a random variable, then the multinomial model holds upon conditioning on the observed number of genera. We note in passing that, conditional on m, the multinomial model holds if $N_1, N_2, \ldots$ are independent Poisson random variables with means $\mu_1, \mu_2, \ldots$, in which case $p_j = \mu_j / \sum_{j=1}^{\infty} \mu_j$. An unappealing feature of this statistical model is that it is not based on an underlying model of the evolutionary process. However, this is precisely in keeping with the empirical work in this area. The whole thrust of this work has been to identify power laws as a so-called emergent property—that is, without recourse to models of the underlying processes.

Let $E(N_j)$ be the expected value of N_j. Interest centers on testing the null hypothesis H_0: $E(N_j) = cj^{-\gamma}$ for unknown parameters $c > 0$ and $\gamma > 1$. As $E(N_j) = mp_j$, this can be restated as H_0: $p_j = j^{-\gamma}/\zeta(\gamma)$, where $\zeta(\gamma) = \sum_{j=1}^{\infty} j^{-\gamma}$ is the zeta function. As noted, we will test H_0 against the nonparametric alternative hypothesis H_1: $p_1 \ge p_2 \ge \ldots$ that frequency is nonincreasing with size. To do so, we will use the likelihood ratio (LR) statistic $\Lambda = \log L_1 - \log L_0$, where $\log L_0$ and $\log L_1$ are the values of the log likelihood maximized under H_0 and H_1, respectively. For the multinomial model outlined above, the log-likelihood function is given by $\log L = \sum_{j=1}^{\infty} N_j \log p_j$. It is straightforward to maximize $\log L$ numerically under H_0. The value $\hat{\gamma}$ that maximizes $\log L$ is the maximum likelihood estimate of the power law parameter γ. Maximum likelihood estimation of multinomial probabilities under the order restriction imposed by H_1 is described in Robertson et al. (1988). In the applications described in the following section, the log-likelihood function was maximized under H_1 by the pool adjacent violators algorithm. This algorithm is described in the appendix.

Once the observed value λ of the LR statistic Λ is found, its significance can be assessed through the following simulation procedure. A sample of m units is simulated from the multinomial distribution fitted under H_0 (i.e., with probabilities $p_j = j^{-\hat{\gamma}}/\varsigma(\hat{\gamma})$). The value of Λ is found for the simulated data in exactly the same way as for the original data. The procedure is repeated a large number of times, and the significance level is estimated by the proportion of the simulated values of Λ that exceed λ. This is an example of a parametric bootstrap (Efron and Tibshirani 1993).

To assess the performance of this parametric bootstrap test, we conducted a small simulation experiment. For each of several combinations of m and γ, we generated 1,000 samples from the power law model. For each sample, we applied the parametric bootstrap test at the nominal .05 significance level. Each bootstrap test was based on 200 bootstrap samples. For each combination of m and γ, the estimated probability of falsely rejecting H_0, given by the proportion of these 1,000 samples for which H_0 was rejected, is reported in table 1. The standard errors of these estimated probabilities are all around 0.007. The results suggest that the test based on the parametric bootstrap is

Table 1: Estimated true probability of false rejection of H_0 in testing at the nominal .05 significance level for selected values of m and γ

m and γ	Estimated probability
200:	
1.5	.046
1.75	.043
2.0	.054
500:	
1.5	.047
1.75	.046
2.0	.043
1,000:	
1.5	.043
1.75	.061
2.0	.043

slightly conservative (i.e., the true rate of false rejection of H_0 is slightly below the nominal significance level) for this range of parameter values.

The null hypothesis considered here is that the power law holds over all sizes. In some situations, interest centers on testing the null hypothesis that the power law holds conditional on size exceeding a specified value j_0. The same method described here can be applied in that case, provided j_0 is selected independently of the data used in the test. However, these methods are not valid if the selection of j_0 is based on the data. In particular, if j_0 is even informally selected from the data to favor the fit of a power law, then the significance level of any test that fails to account for this selection will be distorted in favor of the power law. In principle, it may be possible to account for the selection of j_0 in the assessment of significance, but whether and how this could be done would depend on the details of the selection procedure.

Results

We applied the methods of the previous section to four data sets extracted from the article of Yule (1924). The data represent the frequencies of genera of different sizes for snakes, lizards, and two Coleopterans (*Chrysomelidae* and *Cerambycinae*). The standard plots of $\log n_j$ against

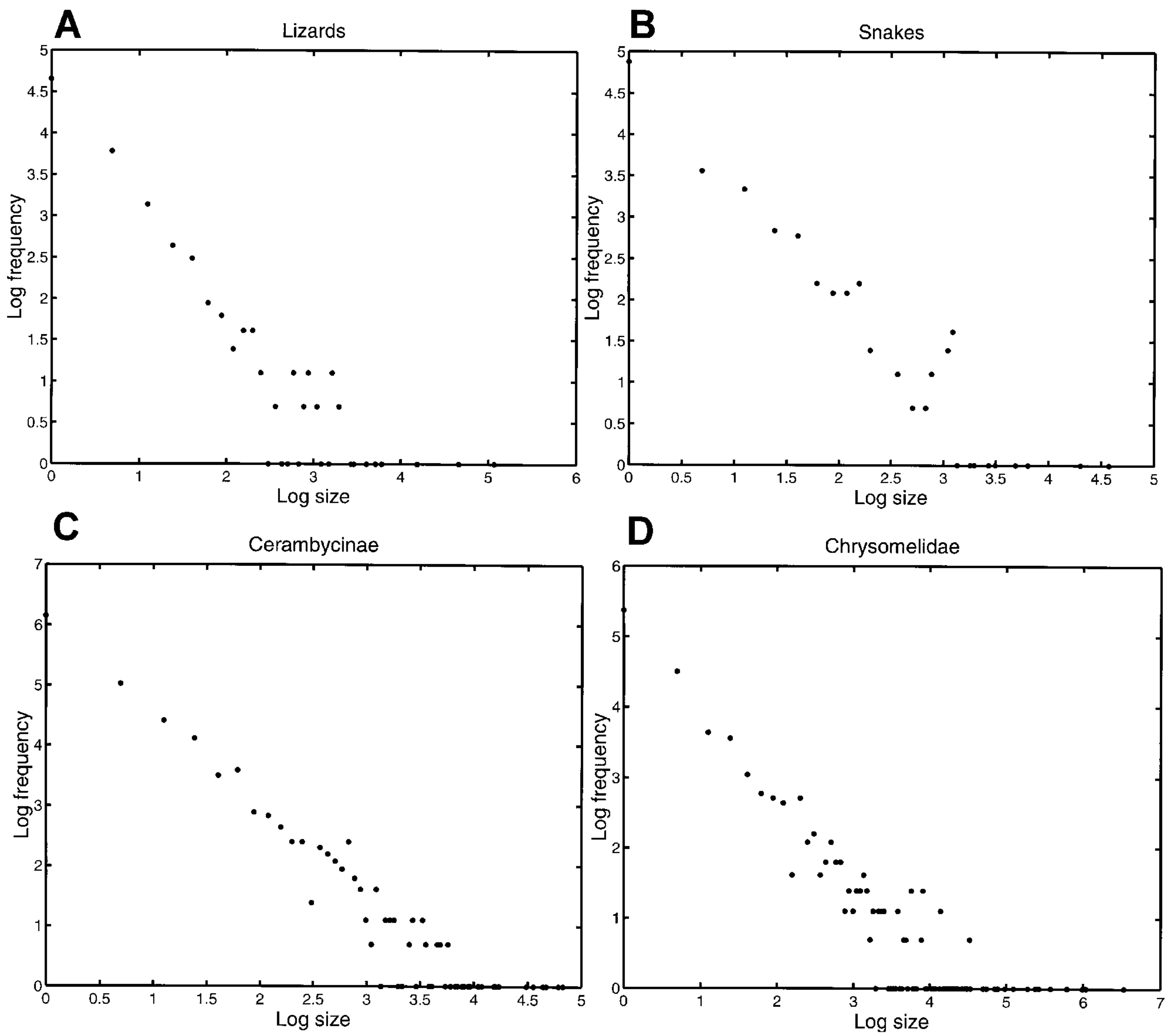

Figure 1: Plot of $\log n_j$ against $\log j$ with values for which $n_j > 0$ for (*A*) lizards, (*B*) snakes, (*C*) *Cerambycinae*, and (*D*) *Chrysomelidae*

Table 2: Values of m, $\hat{\gamma}$, λ, and estimated significance level for four data sets extracted from Yule (1924)

Group	m	$\hat{\gamma}$	λ	Estimated significance level
Lizards	259	1.67	16.68	.092
Snakes	293	1.71	26.40	.000
Cerambycinae	1,024	1.71	39.35	.000
Chrysomelidae	627	1.48	30.09	.006

Note: In each case, the significance level was estimated from 1,000 bootstrap samples.

$\log j$ for $n_j > 0$ are shown in figure 1. The behavior is typical for taxonomic data of this kind. The results are summarized in table 2. In each case, the significance level was estimated from 1,000 bootstrap samples. By conventional standards of significance, the power law model can be decisively rejected for all groups except lizards. Even in the case of lizards, in which the power law model is known to provide an impressive overall fit (Hill 1970), the result is marginally significant.

Discussion

The stringency with which the goodness of a fitted model should be assessed depends to a degree on the claims that are being made about the model. The claim that a model is correct, as opposed merely to providing a useful approximation, should be subjected to particularly close scrutiny. Such claims have been made about the power law model for size-frequency data without adequate scrutiny. The purpose of this note has been to develop and apply a procedure for testing the goodness of fit of the power law model. The key to this procedure is embedding the power law model in the nonparametric family of models in which frequency is nonincreasing in size. This was done to ensure that the power law model is not accepted simply because it exhibits this behavior. After all, claims about this model go well beyond this particular feature.

Turning to the results of the previous section, it is obvious that rejecting the power law model for three of four data sets does not constitute a wholesale rejection of the model. However, a cursory review of the literature suggests that these data sets are fairly typical, and we suspect that similar results would be found for many, if not most, such data sets. Failure of the power law model can be explained in a number of ways. For example, the frequencies of genera of different sizes may simply not follow a power law. Alternatively, errors of incompleteness or misclassification could cause the power law model to be rejected even if it held for the complete, error-free data. Choosing between these and possibly other alternatives is beyond the scope of this note. However, the standing claim that observed size distributions of genera follow power laws does not appear to be tenable.

This note has focused on the case in which size is discrete. Power laws have also been invoked for continuous size data. Let the continuous random variables $X_1, X_2, \ldots, X_m$ denote the sizes of m objects. Under a continuous version of the power law, these follow the Pareto distribution with probability density function $f(x) = (\gamma - 1)a^{\gamma-1}x^{-\gamma}$ for $x \geq a > 0$ and $\gamma > 1$. For this model, $\log f(x)$ is a declining linear function of $\log x$. It is possible to test the null hypothesis that size follows a Pareto distribution against the alternative that $f(x)$ is nonincreasing. In this case, the log-likelihood function is simply $\sum_{j=1}^{m} \log f(x_j)$, where x_j is the observed value of X_j. While maximum likelihood estimation of f is not possible under the alternative, Ramsay (1998) described an approach to estimating a smooth, nonincreasing probability density function. This estimate can be used to form a pseudolikelihood ratio statistic whose significance can be assessed through a parametric bootstrap by simulating from the fitted Pareto distribution.

Acknowledgments

We acknowledge with gratitude the extremely helpful comments of anonymous reviewers.

APPENDIX

This appendix briefly describes the pool adjacent violators algorithm used to find the maximum likelihood estimates $\hat{p}_1, \hat{p}_2, \ldots$ of the multinomial probabilities $p_1, p_2, \ldots$ under the order restriction $p_1 \geq p_2 \geq \ldots$. Details are given in Robertson et al. (1988). To begin with, note that $\hat{p}_j = 0$ for all $j > j_{\max}$, where $j_{\max}$ is the size of the largest genus in the sample of m genera. The algorithm proceeds in the following way: Let $\tilde{p}_j = n_j/m$ be the sample frequency of genera of size j for $j = 1, 2, \ldots, j_{\max}$. If $\tilde{p}_1 \geq \tilde{p}_2 \geq \ldots \geq \tilde{p}_{j_{\max}}$, then the sample frequencies satisfy the order restriction $\hat{p}_j = \tilde{p}_j$ for $j = 1, 2, \ldots, j_{\max}$, and the algorithm terminates. Otherwise, $\tilde{p}_j < \tilde{p}_{j+1}$ for at least one value of j. In that case, both of these values are replaced by their average. The algorithm continues in this way until the order restriction is met.

It can be shown that the algorithm works regardless of the order in which violations of the order restriction are corrected. For programming purposes, it is convenient to begin at $j = 1$ and to increment j until a violation is found and corrected by averaging. Suppose that the first violation is found at j and $j + 1$. Its correction can give rise to a violation at $j - 1$ and j. To account for this possibility, j can be decremented, and the order restriction can be checked. If it is satisfied, then the algorithm returns to incrementing j. If the order restriction is violated, then the violation is corrected and the algorithm continues to decrement j, correcting any new violations by averaging, until there is no violation of the order restriction. At that point, the algorithm resumes incrementing j.

Literature Cited

Azzalini, A., A. W. Bowman, and W. Hardle. 1989. On the use of nonparametric regression for model checking. Biometrika 76:1–11.

Bak, P. 1996. How nature works. Springer, New York.

Burlando, B. 1990. The fractal dimension of taxonomic systems. Journal of Theoretical Biology 146:99–114.

———. 1993. The fractal geometry of evolution. Journal of Theoretical Biology 163:161–172.

D'Agostino, R. B., and M. A. Stephens. 1986. Goodness-of-fit techniques. Dekker, New York.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, London.

Feller, W. 1941. On the logistic law of growth and its empirical verifications in biology. Acta Biotheoretica 5: 51–66.

Hart, J. D. 1997. Nonparametric smoothing and lack-of-fit tests. Springer, New York.

Hill, B. M. 1970. Zipf's law and prior distributions for the composition of a population. Journal of the American Statistical Association 65:1220–1232.

Kauffman, S. A. 1995. At home in the universe. Oxford University Press, Oxford.

Kingsland, S. E. 1995. Modeling nature: episodes in the history of population biology. University of Chicago Press, Chicago.

Marchetti, C., P. S. Meyer, and J. H. Ausubel. 1996. Human population dynamics revisited with the logistic model: how much can be modeled and predicted? Technological Forecasting and Social Change 52:1–30.

Newman, M. E. J. 1996. Self-organized criticality, evolution and the fossil extinction record. Proceedings of the Royal Society of London B, Biological Sciences 263: 1605–1610.

Ramsay, J. O. 1998. Estimating smooth monotone functions. Journal of the Royal Statistical Society B 60:365–375.

Robertson, T., F. T. Wright, and R. L. Dykstra. 1988. Order restricted statistical inference. Wiley, New York.

Willis, J. C. 1922. Age and area. Cambridge University Press, Cambridge.

Yule, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis. Philosophical Transactions of the Royal Society of London B, Biological Sciences 213:21–87.

Associate Editors: William J. Boecklen
Jonathan B. Losos