Ecological Archives **E089-052-A1**

Ethan P. White, Brian J. Enquist, and Jessica L. Green. 2008. On estimating the exponents of power-law frequency distributions. Ecology 89:905-912.

Appendix A.Details of logarithmic binning, cumulative distribution function fitting, and maximum likelihood estimation.

Why simple logarithmic binning estimates λ+1

By definition a bin of constant logarithmic width means that the logarithm of the upper edge of a bin (x_{i+1}) is equal to the logarithm of the lower edge of that bin (x_i) plus the bin width (b). That is,

$$\log\left(x_{i+1}\right) = \log\left(x_i\right) + b$$
$$\Rightarrow x_{i+1} = e^{\left(\log\left(x_i\right)+b\right)} = e^{\log\left(x_i\right)}e^b = x_i e^b$$

Since the linear bin width of bin i, w_i, is defined as

$$w_i = x_{i+1} - x_i$$

the linear bin width is directly proportional to x_i because

$$w_i = x_i e^b - x_i = x_i\left(e^b - 1\right).$$

The number of observations in a bin (n) is equal to the density of observations in that bin times the width of that bin. Therefore if the probability density function, $f\left(x\right) \propto x^{\lambda}$ and the width of the bin, $w \propto x$, then

$$n \propto x^{\lambda} x = x^{\lambda+1} = x^{\lambda+1}$$

and regressing log(n) against log(x) yields a slope equal to λ+1, not λ. If n is divided by the linear width of the bin then,

$$\frac{n}{w} \propto \frac{x^{\lambda} x}{x}$$
$$\propto x^{\lambda}$$

and thus a regression of the normalized logarithmic bin counts against the logarithm of x will estimate λ.

Linearizations of the cumulative distribution functions

Distribution	CDF	Linearization of CDF
Pareto[1]	$1-a^{-(\lambda+1)}x^{\lambda+1}$	$\log\left(1-F(x)\right)=-(\lambda+1)\log(a)+(\lambda+1)\log(x)$
Truncated Pareto[2]	$\dfrac{x^{\lambda+1}-a^{\lambda+1}}{b^{\lambda+1}-a^{\lambda+1}}$	------
Discrete Pareto[2]	$\dfrac{\sum_{j=a}^{x} j^{\lambda}}{\zeta(-\lambda,a)}$	------
Power Function[1]	$(x/b)^{\lambda+1}$	$\log\left(F(x)\right)=-(\lambda+1)\log(b)+(\lambda+1)\log(x)$

[1] Typically the linearization of the CDF is then fit using simple linear regression. This ignores the fact that the intercept is also a function of lambda, or alternatively that the CDF at the minimum attainable value of x must be equal to 0. This could be dealt with using non-linear regression, but this is not done in the ecological literature.
[2]There is no obvious way to isolate λ as a simple component of the slope for these distributions.

<u>Confidence intervals for maximum likelihood estimates</u>

Solutions for the standard error (SE) of the estimated value of λ[1]. Estimates of the SE can also be obtained using the bootstrap or jackknife techniques (Newman 2005).

Distribution	SE of MLE
Pareto	$\dfrac{\left(-\hat{\lambda}-1\right)}{\sqrt{n}}$
Truncated Pareto	$\dfrac{1}{\sqrt{n}}\left(\dfrac{1}{\hat{\lambda}^2}-\dfrac{(b/a)^{-\lambda}\left[\ln(b/a)\right]^2}{\left[1-(b/a)^{-\hat{\lambda}}\right]^2}\right)^{-\frac{1}{2}}$
Discrete Pareto[2]	$\dfrac{1}{\sqrt{n\left[\dfrac{\zeta''(-\hat{\lambda},a)}{\zeta(-\hat{\lambda},a)}-\left(\dfrac{\zeta'(-\hat{\lambda},a)}{\zeta(-\hat{\lambda},a)}\right)^2\right]}}$
Power Function	$\dfrac{\hat{\lambda}+1}{\sqrt{n}}$

[1]Source: Pareto (Clauset et al. 2007); Truncated Pareto (Aban et al. 2006); Discrete Pareto (Clauset et al. 2007); Power Function (JLG). Solutions for SE are in the limit of

large *n*. It is possible to correct the SE for small *n* and solutions for this correction are available for the Pareto distribution (Johnson et al. 1994, Newman 2005, Clauset et al. 2007).

[2]The estimates of the SE for the Pareto distribution can be used as an approximation for the Discrete Pareto for reasonably large *n* and *a* (Clauset et al. 2007).

Aban, I. B., M. M. Meerschaert, and A. K. Panorska. 2006. Parameter estimation for the truncated pareto distribution. Journal Of The American Statistical Association **101**:270-277.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2007. Power-law distributions in empirical data. arXiv:0706.1062v1 [physics.data-an].

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

Newman, M. E. J. 2005. Power laws, Pareto distributions and Zipf's law. Contemporary Physics **46**:323-351.