Notes and Comments

BOOTSTRAPPING THE GINI COEFFICIENT OF INEQUALITY[1]

Philip M. Dixon,[2] Jacob Weiner,[3]
Thomas Mitchell-Olds,[4] and Robert Woodley[5]

Despite current interest in the causes and consequences of plant size hierarchies, there are different opinions about the best way to evaluate a size distribution. Weiner and Solbrig (1984) have argued that size hierarchy means size inequality, and that the Gini coefficient of inequality (Sen 1973) is more relevant than skewness or variance of plant size (e.g., Turner and Rabinowitz 1983) for most ecological questions. Weiner (1985) presents a formula to estimate the population Gini coefficient (G) from a sample and states that reasonable confidence intervals for the population Gini coefficient can be obtained by a bootstrapping procedure (Efron 1982). This note evaluates the accuracy of these bootstrap confidence intervals. We find that they are reasonably accurate when calculated from samples of 50 or more individuals, but that they are too narrow when calculated from smaller samples.

The bootstrap procedure uses the observed data to estimate the theoretical and usually unknown distribution from which the data came (Efron 1982, Meyer et al. 1986). Bootstrap samples of the same size as the original sample are repeatedly drawn by sampling with replacement from the observed data. The test statistic, e.g., the Gini coefficient, is calculated for each bootstrap sample. The distribution of G's obtained from bootstrap sampling can be used to estimate the standard deviation and set confidence limits on the observed statistic (Efron 1982). The bootstrap procedure does not require any knowledge of the distribution of the statistic in question, may have certain optimal properties (Efron 1981, but see Schenker 1985, Wu 1986), and can be used when the standard deviation or confidence intervals for the statistic are unknown or difficult to calculate analytically.

The accuracy of any method for computing confidence intervals can be evaluated by generating data from a known distribution with a known parameter. If many samples of data from the known distribution are generated and a confidence interval calculated from each, the number of confidence intervals that include the parameter can be determined. An accurate confidence interval includes the known parameter the stated percentage of the time; for example, a 95% confidence interval should include the true value in 95% of the random samples of data. Although in some situations bootstrap confidence intervals are relatively accurate (Efron 1982:79), in other situations they are too narrow (Schenker 1985, Meyer et al. 1986).

Methods

The accuracy of the bootstrap procedure can be evaluated by repeatedly generating random samples from a distribution with a known Gini coefficient. From each sample, the bootstrap procedure is used to calculate a confidence interval. Then, one can ask the question: how many of those confidence intervals include the known G for the population. The coverage probability is the observed percent of times the confidence interval includes the parameter.

We performed two simulation studies. In the first study, large simulated populations were generated from three statistical distributions chosen to be representative of observed plant size distributions: normal, lognormal, and bimodal (see Table 1). Normal random numbers were generated by transforming uniform random numbers from the VAX/VMS pseudo-random number generator RAN with an approximate inverse normal cumulative distribution function (Hastings 1955). The bimodal distribution was the sum of two normal distributions with means of 6.0 and 10.0 and variance of 1.0. Given those means, the probability of generating a negative value was less than 1 in 100 000. The size of each large population was arbitrarily set at 10 000.

The "true" Gini coefficient was calculated from each large population, then five hundred random samples of size N were taken from the population; bootstrap 95% confidence intervals were calculated from each sample as in Weiner (1985). In this study, we used sample sizes (N) of 20, 50, 100, and 500. Computations were done using a VAX 11/750 computer at the Smithsonian Environmental Research Center.

In the second simulation study, samples of sizes 20, 100, and 250 were generated from uniform, truncated-normal, and lognormal distributions. Since plant size must be positive, the data generated from the normal distribution were truncated by replacing any nonpositive values. The true value of the Gini coefficient can be calculated analytically when the size data follow one of these three statistical distributions (Appendix). Computations for the one-sample bootstrap were done on an IBM 3081D computer using IMSL (1982) random number generators and VS FORTRAN programs. Computations for the two-sample bootstrap were done in PASCAL on a microcomputer. A randomized shuffler was used to improve the uniform distribution from

RAN, the PASCAL pseudo-random number generator, and the Box-Müller method was used to generate normal deviates (Press et al. 1986).

In this study, bootstrap confidence intervals were computed using the percentile method (Efron 1982), the bias-corrected method (Efron 1982), an unequal-tailed percentile method (*personal communication* from J. Weiner in Zimmerman and Weis 1985), and the bias-corrected (alpha) method (Efron and Tibshirani 1986). The percentile method sets a 95% confidence interval such that 2.5% of the bootstrap G values are below the lower bound and 2.5% are above the upper bound. The unequal-tailed method arbitrarily adjusts the confidence interval so that 3% of the bootstrap values are below the lower bound and 2% are above the upper bound. The bias-corrected method and the bias-corrected (alpha) method attempt to correct for biased and narrow confidence intervals by using properties of the bootstrap distribution.

The sample Gini coefficient (Weiner and Solbrig 1984) was calculated using the formula (Glasser 1962):

$$G = \frac{1}{2\bar{X}n(n-1)} \sum_{i=1}^{n} (2i - n - 1)X_i,$$

where X_i are the sizes sorted from smallest to largest, $X_1 \le X_2 \le \ldots X_n$. When an efficient sort routine is used, this formula is computationally faster than the usual formula given in Weiner (1985).

Results and Discussion

The results of the two studies are similar. The bootstrap confidence intervals (CI's) are too narrow in almost all cases. In the first study, bootstrap CI's are unacceptably inaccurate for small samples ($N < 50$) for all three distributions, and for all except the largest sample size ($N = 500$) for the lognormal distribution (Table 1). Bootstrapping performs adequately for larger samples ($N \ge 50$) from the normal and bimodal populations.

The second study confirms these results. The empirical coverage of the bootstrap confidence interval decreases at larger Gini values and depends on the distribution of plant sizes (Table 1). For similar Gini values, the coverage of the 95% CI is better for plant sizes following uniform or truncated-normal distributions than for plant sizes that follow the more skewed lognormal distribution. The coverage of the 95% CI's from the other bootstrap methods is not substantially different, and results for 90% and 99% CI's are similar to those for 95% CI's. Using a bootstrap confidence interval from a small sample to test a hypothesis would misstate the type I error rate (alpha level); one would reject a null hypothesis more often than one should.

The empirical coverage of the bootstrap confidence

TABLE 1. Estimated coverage probabilities (in percent) for nominal 95% bootstrap confidence intervals.*

Single-sample confidence interval—study I

Distribution of the data:	Population Gini value	Sample size (N)			
		20	50	100	500
Normal (6, 1)	0.11	86.0	92.6	92.4	95.4
Bimodal (see Methods)	0.16	91.8	93.0	95.4	95.0
Lognormal (0, 1)	0.52	81.0	84.2	89.0	93.8

Single-sample confidence interval—study II

Distribution of the data:	Population Gini value	Sample size (N)		
		20	100	250
Uniform (1, 1.86)	0.10	87.6		
Lognormal (0, 0.21)	0.10	89.2		
Uniform (1, 4.00)	0.20	92.4		
Lognormal (0, 0.40)	0.20	89.7		
Uniform (1, 19.0)	0.30	91.8	95.4	94.3
Lognormal (0, 0.54)	0.30	85.1	92.5	95.2
Trunc. normal (0, 1)	0.41	92.5	92.7	93.9
Lognormal (0, 1.0)	0.52	78.2	87.9	
Lognormal (0, 1.5)	0.68	75.3		
Lognormal (0, 2.0)	0.84	63.4		

* Estimates in study I are based on 500 intervals. Estimates in study II are based on 845 intervals (for $N = 20$ and $N = 100$) or 212 intervals ($N = 250$). Each interval was based on 500 bootstrap samples. The data distributions are explained in the Methods and Appendix.

interval depends mainly on the number of observations in the raw data, not on the number of bootstrap repetitions. Because the results for 100, 500, and 1000 bootstrap repetitions are similar, only the data for 500 repetitions are presented. With small samples (20 individuals), bootstrap 95% CI's include the parameter 63 to 92% of the time (Table 1). With 100 individuals, 95% CI's include the parameter 88–95% of the time, and they are relatively accurate when $N = 250$ plants (Table 1). As the number of observations increases, the frequency distribution of bootstrap values approaches the frequency distribution of the original data (Nash 1981), so it is not surprising that bootstrap CI's are more accurate with larger data sets. What is noticeable is their poor small-sample performance.

Comparison of the location of the upper and lower bounds to the confidence intervals suggests why the bootstrap confidence intervals are poorly behaved. The true lower and upper bounds on G for a size distribution can be estimated by Monte Carlo methods by calculating G from many independent samples of data. Generally, the lower bound of the bootstrap CI was close to the true lower bound, but the upper bound was too low by as much as 0.10 units. Using an unequal-tailed CI (from the 3rd to the 98th percentiles, instead of the 2.5th to 97.5th percentiles) brings the

TABLE 2. Estimated coverage probabilities (in percent) for nominal 95% bootstrap confidence intervals for the difference between two populations.*

Gini value for popula-tion 1	Gini value for population 2			
	0.10	0.20	0.52	0.84
0.10	93.9	91.5	77.8	60.3
0.20		95.2	84.9	61.8
0.52			93.4	86.9
0.84				94.8

* Estimates are based on 212 confidence intervals, 500 bootstrap samples for each interval, and a sample size of 20 from each population. All populations are lognormal.

upper bound closer to the true value, but the upper bound is still too low.

The upper bound is too low partly because the sample estimator of G is biased. The bias, the difference between the true G for the population and the average of many sample estimates of G, is largest when the sample size is small and the size distribution is very unequal (e.g., lognormal). This bias cannot be corrected unless one can estimate the true distribution of plant sizes, because the amount and direction of the bias depend on the distribution. The estimated Gini value is larger than the true value for uniformly distributed plant sizes and smaller than the true value for normally and lognormally distributed plant sizes. The "unbiased" sample estimator of the Gini coefficient used by Weiner and Solbrig (1984) reduces but does not eliminate the bias, although the bias is practically zero when $N > 100$ or when $G < 0.2$.

A confidence interval around a sample Gini value can be used to test a null hypothesis of the form: the population G equals some constant. These confidence intervals have been used (e.g., Weiner 1985, 1986) to test if two populations are different by examining either the overlap between two confidence intervals, or whether each confidence interval includes the other sample's G value. The proper test of whether two populations are similar is to see if a confidence interval around the difference between two G values includes 0. Such a confidence interval can be derived by bootstrapping. From the two samples of raw data, repeatedly draw a pair of bootstrap samples and calculate the difference between their Gini coefficients. The bootstrap distribution of differences can be used to set the bounds on a CI for the difference between the population G values. These confidence intervals are quite accurate if the two populations have similar Gini coefficients, but are too narrow if the two samples are small and have different Gini coefficients (Table 2). We applied this method to Weiner's (1986) results on the differences in inequality in *Ipomoea tricolor* populations grown in different competitive regimes. When the

difference between two Gini coefficients is bootstrapped with the method suggested here, the significance level of his results increases, although the present study indicates that his sample sizes are too small to produce accurate single-sample bootstrap confidence intervals.

We should not be surprised that bootstrapping yields "overconfident" single-sample confidence intervals, since the bootstrap samples, which are repeatedly drawn from one sample of raw data, are treated as if they were truly independent samples taken from some population. Reasonably good confidence intervals can be obtained by bootstrapping when sample sizes are > 100. When the distribution includes a few very large individuals (leading to large G's), the bootstrap frequency distribution is a poor approximation to the true, but unknown, frequency distribution, and the confidence intervals generated are unacceptably narrow unless the sample size is very large.

These statistical considerations are not an argument to abandon the use of the Gini coefficient in favor of other measures of size inequality. As always, parameters should be chosen to represent best the phenomenon of interest. Accurate confidence intervals for variance and skewness are also difficult to determine from small samples (Schenker 1985). To set a confidence interval around a measure of size inequality accurately, or at least conservatively, using the bootstrap requires a larger experiment than often used.

Literature Cited

Efron, B. 1981. Nonparametric standard errors and confidence intervals. Canadian Journal of Statistics 9:139–158.
———. 1982. The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.
Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1:54–77.
Glasser, G. J. 1962. Variance formulas for the mean difference and coefficient of concentration. Journal of the American Statistical Association 57:648–654.
Hastings, C. 1955. Approximations for digital computers. Princeton University Press, Princeton, New Jersey, USA.
IMSL. 1982. IMSL library reference manual. Edition 9. IMSL, Houston, Texas, USA.
Meyer, J. S., C. G. Ingersoll, L. L. McDonald, and M. S. Boyce. 1986. Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. Ecology 67:1156–1166.
Nash, S. W. 1981. Discussion of Efron paper. Canadian Journal of Statistics 9:163–164.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1986. Numerical recipes. Cambridge University Press, New York, New York, USA.

Schenker. N. 1985. Qualms about bootstrap confidence intervals. Journal of the American Statistical Association 80: 360–361.

Sen, A. 1973. On economic inequality. Clarendon, Oxford, England.

Turner, M. D., and D. Rabinowitz. 1983. Factors affecting frequency distributions of plant mass: the absence of dominance and suppression in competing monocultures of Festuca paradoxa. Ecology 64:469–475.

Weiner, J. 1985. Size hierarchies in experimental populations of annual plants. Ecology 66:743–752.

———. 1986. How competition for light and nutrients affects size variability in Ipomoea tricolor populations. Ecology 67:1425–1427.

Weiner, J., and O. Solbrig. 1984. The meaning and measurement of size hierarchies in plant populations. Oecologia (Berlin) 61:334–336.

Wu, C. J. F. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14:1261–1295.

Zimmerman, J. K., and I. M. Weis. 1985. Factors affecting survivorship, growth, and fruit production in a beach population of Xanthium strumarium. Canadian Journal of Botany 62:2122–2127.

[2] Cornell Plantations, Cornell University, Ithaca, New York 14853 USA.
[3] Department of Biology, Swarthmore College, Swarthmore, Pennsylvania 19081 USA.
[4] Department of Botany, University of Washington, Seattle, Washington 98105 USA.
[5] Smithsonian Environmental Research Center, Edgewater, Maryland 21037 USA.

APPENDIX

Analytical expressions for the population Gini coefficient and parameterization of distributions.*

Distribution	Explanation	Population Gini coefficient
Truncated-normal (0, 1)	Generate a normal (0, 1) deviate, reject if not larger than 0.	$4\Phi(0) + \sqrt{2} - 3 = \sqrt{2} - 1$
Lognormal (μ, σ)	Generate x, a normal (μ, σ) deviate, transform to $\exp(x)$	$2\Phi(\sigma/\sqrt{2}) - 1$
Uniform (a, b)	a = minimum value, b = maximum value.	$(b - a)/3(b + a)$

* $\Phi(x)$ is the normal cumulative distribution function. Note that the population Gini coefficient for the lognormal distribution depends only on the standard deviation.

ERRATA

In an article by Philip Dixon et al. ("Bootstrapping the Gini coefficient of inequality," Ecology 68(5):1548–1551) there was an error in the equation on page 1549. The correct equation should be:

$$G = \frac{1}{\bar{X}n(n-1)} \sum_{i=1}^{n} (2i - n - 1) X_i.$$

The numerical results are correct.