Ecology, 81(4), 2000, pp. 1139–1142

DESCRIBING INEQUALITY IN PLANT SIZE OR FECUNDITY

Christian Damgaard[1] and Jacob Weiner[2]

[1]*Department of Terrestrial Ecology, National Environmental Research Institute, Vejlsøvej 25, DK-8600 Silkeborg, Denmark*
[2]*Department of Ecology, Royal Veterinary and Agricultural University, Rolighedsvej 21, DK-1958 Frederiksberg, Denmark*

Abstract. Lorenz curves have been used to describe inequality in plant size and fecundity, where the total inequality is summarized by the Gini coefficient. Here we propose a second and complementary statistic, the Lorenz asymmetry coefficient, which characterizes an important aspect of the shape of a Lorenz curve. The statistic tells us which size classes contribute most to the population's total inequality. This may be useful when interpreting the ecological significance of plant size or reproductive inequality.

Key words: asymmetry; gamma distribution; Gini coefficient; inequality; lognormal distribution; Lorenz curve; plant, fecundity, growth, size; size distributions.

Introduction

Variation in plant size and fecundity has traditionally been described and analyzed using the statistical moments of the size distribution, or statistics derived from the moments such as standard deviation and skewness. In recent years, the focus has shifted toward an emphasis on inequality in size and fecundity, after it was argued that ''size variability'' or ''size hierarchy,'' as the terms are used by ecologists, are often synonymous with the concept of size inequality or concentration (Weiner and Solbrig 1984). Similarly, inequality in fecundity reflects the degree of increase in the representation of the genes of the more fecund individuals in the next generation (e.g., Heywood 1986). Several measures of inequality from economics literature (Sen 1973) have been used to analyze variation in size and fecundity within plant populations. One approach to inequality that has been applied to plant populations is the Lorenz curve (Lorenz 1905, Gini 1912, Sen 1973, Dagum 1980, Weiner and Solbrig 1984).

In the Lorenz curve, individuals are ranked by size (or fecundity; we refer below only to size), and the cumulative proportion of plants (x-axis) is plotted against the corresponding cumulative proportion of their total size on the y-axis (Fig. 1). If we have a sample of n ordered plants, such that x_i' is the size of plant i, and $x_1' \leq x_2' \leq \ldots \leq x_n'$, then the sample Lorenz curve is the polygon joining the points $(h/n,\ L_h/L_n)$, where $h = 0, 1, \ldots, n$, $L_0 = 0$, and $L_h = \Sigma_{i=1}^{h} x_i'$ (Kotz et al. 1983). Alternatively, the Lorenz curve can be expressed as follows:

$$L(y) = \int_0^y x\ dF(x)/\mu \qquad (1)$$

where $F(y)$ is the cumulative distribution function of ordered plants, and μ is the average plant size (Dagum 1980, Kotz et al. 1983).

If all individuals are the same size, the Lorenz curve is a straight diagonal line, called the line of equality (Fig. 1). If there is any inequality in size, then the Lorenz curve is below the line of equality. The total amount of inequality can be summarized by the Gini coefficient (or Gini ratio), which is the ratio between the area enclosed by the line of equality and the Lorenz curve, and the total triangular area under the line of equality (Fig. 1). The Gini coefficient is most easily calculated from unordered plant size data as the ''relative mean difference,'' i.e., the mean of the difference between every possible pair of individuals, divided by the mean size (Sen 1973):

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2\mu}. \qquad (2)$$

Alternatively, if the data is ordered by increasing plant size (Dixon et al. 1987),

$$G = \frac{\sum_{i=1}^{n} (2i - n - 1)x_i'}{n^2\mu}. \qquad (3)$$

The Gini coefficient ranges from a minimum value of zero, when all individuals are equal, to a theoretical maximum of one in an infinite population in which every individual except one has a size of zero. It has been shown that sample Gini coefficients calculated by Eq. 2 or 3 should be multiplied with $n/(n - 1)$ to be-

Manuscript received 17 November 1998; revised 23 March 1999; accepted 12 April 1999; final version received 12 May 1999.

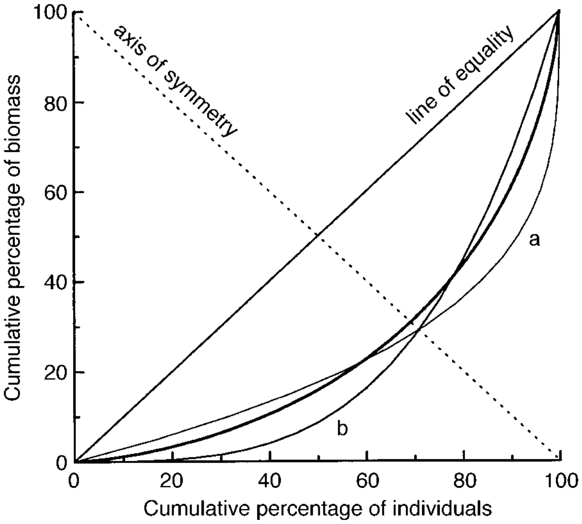

FIG. 1. Three Lorenz curves: a symmetric case (bold line), and two asymmetric cases (a and b).

come unbiased estimates (Glasser 1962). The Gini coefficient has been used as a measure of inequality in size and fecundity in plant populations in numerous studies (e.g., Weiner 1985, Geber 1989, Knox et al. 1989, Preston 1998).

Like any summary statistic, the Gini coefficient does not contain all the information in the Lorenz curve, and it has been pointed out that different Lorenz curves can have the same Gini coefficient (Weiner and Solbrig 1984, Shumway and Koide 1995). In case a of Fig. 1, most of the inequality within the population is due to the very few largest individuals, which contain a very large percentage of the population's biomass. In case b, the same overall degree of inequality is due primarily to the relatively large number of very small individuals, which are contributing very little to the population's total biomass. A simplified version of this can be demonstrated by example populations consisting of only two sizes. Populations a and b in Fig. 2 have the same mean (10), the same Gini coefficient (0.444), but different Lorenz curves (Fig. 3). This difference can be quantified by measuring the asymmetry of the Lorenz curve around the other diagonal (axis of symmetry in Fig. 1), specifically the location of the point at which the Lorenz curve has a slope equal to 1, the same slope as the line of equality (Fig. 1; Kotz et al. 1983, Shumway and Koide 1995). This asymmetry of the Lorenz curve has previously been quantified by Shumway and Koide (1995) to look at the effects of mycorrhizae on plant population structure. They fitted a sample Lorenz curve with a polynomial equation and estimated the point on the Lorenz curve where it was parallel to the line of equality by differentiating the fitted equation.

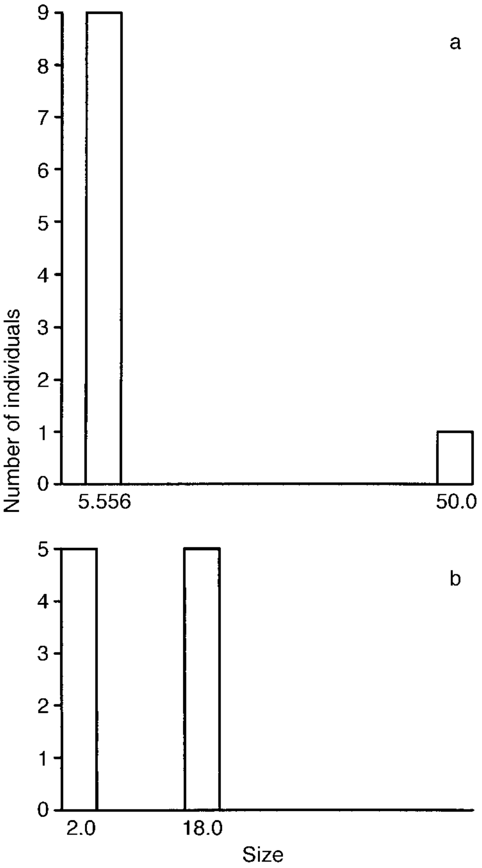

FIG. 2. Two simple populations with the same mean (10), Gini coefficient (0.444), and sample size (10), but with different Lorenz asymmetry coefficients: (a) 1.46; (b) 0.74.

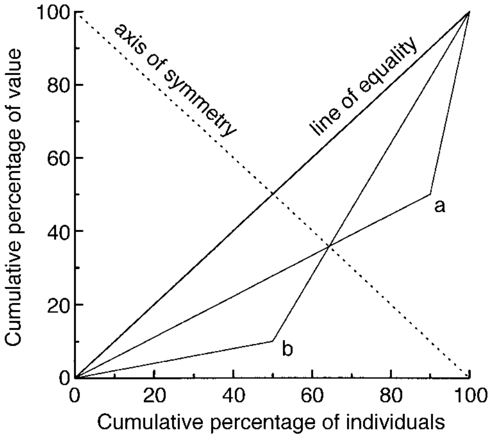

FIG. 3. Lorenz curves for the example populations in Fig. 2.

TABLE 1. Analysis of variance of Lorenz asymmetry coefficient (S) on plant density (high vs. low) and mycorrhizae (present vs. absent) for capsule production by *Abutilon theophrasti* (data from Shumway and Koide [1995]).

Source	SS	df	F	P
Density	0.0287	1	2.26	0.15
Mycorrhizae	0.0826	1	6.48	0.02
Residual	0.2166	17		
Total	0.3280	19		

Here we propose an improved method to quantify the asymmetry in which the data are used directly.

RESULTS

We define the Lorenz asymmetry coefficient as $S = F(\hat{\mu}) + L(\hat{\mu})$, where the functions F and L are as in Eq. 1 (Kotz et al. 1983). To interpret the graphical meaning of the statistic we note that the continuous Lorenz curve as defined by Eq. 1 only is parallel with the line of equality in the point $(F(\mu), L(\mu))$ (Dagum 1980). Then a Lorenz curve is called symmetric if the curve is parallel with the line of equality at the axis of symmetry, and since the axis of symmetry can be expressed by $F(y) + L(y) = 1$, we have that a Lorenz curve is symmetric if and only if $S = F(\hat{\mu}) + L(\hat{\mu}) = 1$ (Kotz et al. 1983). If $S > 1$, then the point where the Lorenz curve is parallel with the line of equality is above the axis of symmetry. Correspondingly, if $S < 1$, then the point where the Lorenz curve is parallel with the line of equality is below the axis of symmetry.

Since the sample Lorenz curve is a polygon, we can calculate the sample statistic S from the ordered plant size data using the following equations:

$$\delta = \frac{\hat{\mu} - x'_m}{x'_{m+1} - x'_m} \tag{4}$$

$$F(\hat{\mu}) = \frac{m + \delta}{n} \tag{5}$$

$$L(\hat{\mu}) = \frac{L_m + \delta x'_{m+1}}{L_n} \tag{6}$$

where m is the number of plants with a plant size less than μ. (If one or more of the data points are equal to $\hat{\mu}$, see Appendix A.) For the two example data sets a and b in Fig. 3, $S = 1.46$ and $S = 0.74$, respectively. Note, that distribution b has an asymmetric Lorenz curve ($S < 1$) although the distribution itself is not skewed.

As a more concrete example of its usefulness, we calculated S for data from Shumway and Koide (1995) on the effect of mycorrhizae and plant density on the number of capsules produced by *Abutilon theophrasti* individuals. Shumway and Koide found that mycorrhizae increased inequality in the number of capsules produced. When S was calculated for the Lorenz curves describing inequality in capsule number, we found that plants without mycorrhizae had significantly more asymmetric Lorenz curves (high plant density, S with mycorrhizae = 0.89, S without mycorrhizae = 0.71; low plant density, S with mycorrhizae = 0.92, S without mycorrhizae = 0.84; Table 1). Thus, the increase in reproductive inequality when mycorrhizae were present was primarily due to an increase in the contribution of plants with highest capsule production.

Confidence intervals for the estimates of S for a sample can be obtained with the bootstrapping procedure, as has previously been demonstrated for the Gini coefficient (Dixon et al. 1987). A *Mathematica* notebook (Wolfram 1996), which draws the sample Lorenz curves and calculates the estimates of G and S as well as the bootstrap confidence intervals from input data, is available on the Internet.[3]

Since the differences between Lorenz curves with different values of S may also be reflected in the skewness of the size distribution, we compared the behavior of S with the behavior of the standard measure of skewness ($m_3/m_2^{3/2}$, where $m_i = 1/n \sum_{j=1}^{n} (x_j - \mu)^i$) in two often used positively skewed distributions, the gamma distribution and the lognormal distribution (see Appendix B).

DISCUSSION

The Lorenz curve expresses a concept of inequality. This concept is fundamentally different from, and therefore represents an alternative to, the concept of variation expressed using distribution moments. If the Lorenz curve framework is appropriate for analysis of variation in a population, then the Lorenz asymmetry coefficient may be useful in describing the population's inequality.

In some cases, it will not matter which framework one uses for looking at inequality. For example, all measures of inequality, such as the Gini coefficient, the coefficient of variation, Theil's measure of entropy (van Andel et al. 1984), or the standard deviation of the log-transformed distribution, will tend to be correlated. If one is just asking if a specific factor or treatment will increase or decrease inequality, then the interpretation of the data will probably not be affected by which framework the researcher takes. But as we obtain more data, both empirical data and artificial data from simulation models, it may be useful to look not only at the overall degree of inequality within a population, but also at how this inequality is distributed. In this context, the statistic S may be useful in analyzing and interpreting inequality in plant size and/or reproductive output.

---

[3] URL: http://www.dmu.dk/TerrestrialEcology/index.htm

The Lorenz asymmetry coefficient in the Lorenz curve model of inequality is in some ways analogous to skewness in the traditional distribution-moment analysis. While skewness has been used as a measure of the size hierarchy within a population, Weiner and Solbrig (1984) argued that this was inappropriate because a distribution can be highly skewed while having very little variation and, similarly, a distribution can contain huge variation while being non- or even negatively skewed. The biological interpretation of skewness of a size or fecundity distribution is problematic, although the use of skewness does have its defenders (Bendel et al. 1989). We suggest that the Lorenz asymmetry coefficient is much easier to interpret biologically in the context of the Lorenz curve, which itself has a clear biological interpretation.

ACKNOWLEDGMENTS

We thank Durland Shumway for providing us with his data and comments, and Gösta Kjellsson and Søren Larsen for comments on a previous version of the manuscript. This research was funded, in part, by the Danish Environmental Research Program and was performed within the Centre for Ecological Risk Assessment of Transgenic Plants.

LITERATURE CITED

Bendel, R. B., S. S. Higgens, J. E. Teberg, and D. A. Pyke. 1989. Comparison of skewness coefficient, coefficient of variation, and Gini coefficient as inequality measures within populations. Oecologia **78**:398–400.

Dagum, C. 1980. The generation and distribution of income, the Lorenz curve and the Gini ratio. Économie Appliquée **33**:327–367.

Dixon, P. M., J. Weiner, T. Mitchell-Olds, and R. Woodley. 1987. Bootstrapping the Gini coefficient of inequality. Ecology **68**:1548–1551.

Geber, M. 1989. Interplay of morphology and development on size inequality: a polygonum greenhouse study. Ecological Monographs **59**:267–288.

Gini, C. 1912. Variabilitá e mutabilita. Pages 211–382 *in* E. Pizetti and T. Salvemini, editors. Memorie di metodologia statistica (1955). Libreria Eredi Virgilio Veschi, Rome, Italy.

Glasser, G. J. 1962. Variance formulas for the mean difference and coefficient of concentration. Journal of the American Statistical Association **57**:648–654.

Heywood, J. S. 1986. The effect of plant size variation on genetic drift in populations of annuals. American Naturalist **137**:851–861.

Knox, R. G., R. K. Peet, and N. L. Christiensen. 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. Ecology **70**:1153–1166.

Kotz, S., N. L. Johnson, and C. B. Read. 1983. Encyclopedia of statistical science. John Wiley and Sons, New York, New York, USA.

Lorenz, M. O. 1905. Methods for measuring the concentration of wealth. American Statistics Association **9**:209–219.

Preston, K. A. 1998. Architectural constraints on flower number in a photoperiodic annual. Oikos **81**:279–288.

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

Shumway, D. L., and R. T. Koide. 1995. Size and reproductive inequality in mycorrhizal and nonmycorrhizal populations of *Abutilon theophrasti.* Journal of Ecology **83**:613–620.

van Andel, J. H., J. M. Nelissen, E. Wattell, T. A. V. Valen, and A. T. Wassenaar. 1984. Theils inequality index applied to quantify population variation in plants with regard to dry matter allocation. Acta Botanica Neerlandica **33**:161–176.

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

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

Wolfram, S. 1996. The Mathematica book. Cambridge University Press, Cambridge, UK.

APPENDIX A

Eqs. 4–6 assume that none of the plant sizes are equal to μ; strictly speaking we assume that plant size are continuously distributed, so that $P(x_i = \mu) \approx 0$. Otherwise, if one or more of $x_i = \hat{\mu}$, then a section of the Lorenz curve is parallel to the diagonal, and S has to be defined as an interval instead of a number. The interval can be defined as follows:

$$\left[\frac{m}{n} + \frac{L_m}{L_n}, \frac{m+a}{n} + \frac{L_{m+a}}{L_n}\right] \tag{A.1}$$

where a is the number of data values that are equal to $\hat{\mu}$.

APPENDIX B

For two commonly used positively skewed distributions, the gamma distribution and the lognormal distribution, the relationship between F and L in Eq. 1 is particularly simple (Dagum 1980). If a sample of plant sizes can be assumed to come from a gamma distribution with two parameters, so that the cumulative plant distribution is

$$F(y) = G(y; \alpha, \beta) = \int_0^y \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx \tag{B.1}$$

($\alpha > 0$; $\beta > 0$), then the Lorenz curve $(F(y), L(y)) = (G(y; \alpha, \beta), G(y; \alpha + 1, \beta)$ (Dagum 1980). For gamma-distributed plant sizes,

$$S = F(\mu) + L(\mu) = \frac{\int_0^\infty t^{\alpha-1} e^{-t}\, dt - \int_\alpha^\infty t^{\alpha-1} e^{-t}\, dt}{\Gamma(\alpha)} + \frac{\int_0^\infty t^{\alpha} e^{-t}\, dt - \int_\alpha^\infty t^{\alpha} e^{-t}\, dt}{\Gamma(1+\alpha)}. \tag{B.2}$$

Thus, S is independent of β. This is comparable to the skewness of the gamma distribution ($= 2/\sqrt{\alpha}$), which also is independent of the parameter β.

Second, if a sample of plant sizes can be assumed to come from a lognormal distribution, so that the cumulative plant distribution is

$$F(y) = LN(y; \alpha, \sigma) = \int_0^y \frac{1}{\sqrt{2\pi}\sigma x} \exp\{-[\log(x) - \alpha]^2/2\sigma^2\}\, dx \tag{B.3}$$

($\sigma > 0$), then the Lorenz curve $(F(y), L(y)) = (LN(y; \alpha, \sigma), LN(y; \alpha + \sigma^2, \sigma)$ (Dagum 1980). Interestingly, for lognormal distributed plant sizes $S = F(\mu) + L(\mu) = 1$. Thus, the Lorenz curve is always symmetric when the data are lognormal distributed. This should be compared to the skewness of the lognormal distribution ($= \sqrt{-1 + e^{\sigma^2}}(2 + e^{\sigma^2})$), which also is independent of α, but depends on the parameter σ. The observation that $S = 1$ for lognormal distributed data, may be used as a basis for deciding whether a lognormal distribution is appropriate.