COENOSES 13(1):33-38, 1998

THE LORENZ CURVE: A GRAPHICAL REPRESENTATION OF EVENNESS

David Nijssen[1], Ronald Rousseau[2] and Piet Van Hecke[1*]

1 Department of Biology, UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium. Email: vanhecke@uia.ua.ac.be
2 Informatie- en Bibliotheekwetenschap, UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium; and, Department of Industrial Sciences and Technology, KHBO, Zeedijk 101, B-8400, Oostende, Belgium. Email: ronald.rousseau@kh.khbo.be
* To whom all correspondence should be addressed.

Keywords and phrases: Gini coefficient; Coefficient of variation; Shannon-Wiener index; Simpson's index; Modified Simpson index; Generalized Lorenz order; Partially ordered sets; Basic axioms and principles for evenness; Removing ambiguity.

Abstract: Problems with the notion of evenness, such as ambiguity, proliferation of indices, choice of indices, etc. can be overcome by a more fundamental, mathematical approach. We show that the Lorenz curve is an adequate representation of evenness. The corresponding Lorenz order induces a partial order in the set of equivalent abundance vectors. Classical indices such as the Gini index, the coefficient of variation, the Shannon-Wiener index and Simpson's index can be adapted in such a way as to confirm this partial order.

1. Introduction

Many authors such as Hurlbert (1971), Pielou (1975) and Peet (1975) have criticized the concept of diversity because of ambiguous meanings and interpretations attached to it. The most likely reason for this confusion is that diversity is a concept with many aspects. Indices such as Shannon-Wiener's or Simpson's which are commonly used to measure diversity combine, in an unstandardized, intuitive way, the two main subconcepts: evenness and number of species (Magurran 1991). Consequently, a thorough investigation of the mathematical background of functions used to measure evenness (or diversity) seems to be in order (Rousseau & Van Hecke, 1998). On the other hand, we consider *species richness* as a simple concept. It is synonymous with the number of different species present in a community. In this paper we will try to clarify the existing ambiguity concerning the notion of evenness by using partial orders (see below). We assume a complete community, studied at a fixed point in time. Dynamical aspects, sampling properties, and consequences on estimates for the proposed measures fall outside the scope of this article.

2. Evenness for a fixed number of species

2.1 Basic Ordering Mathematics

Suppose we have a set of species abundance vectors. From a diversity point of view such a vector is characterized by its length: the number of different species, denoted as S, and by the relative abundance of each species. Evenness, sometimes also called equitability (Magurran, 1991), can best be described by a partial order relation (as we will show). Recall that a relation R, defined on a set U, is a partial order if R is reflexive, transitive and antisymmetric (Roberts 1979, p.15). The class of all subsets of a fixed set U, considered with the inclusion relation is the prototype of a partially ordered set. Within a partially ordered set, elements may not be comparable. If, however, for every x and y in U we have xRy or yRx, the order is said to be complete or total. The inclusion does not yield a total order, but the set of real numbers, $\Re$, with the natural ordering $\leq$, is a totally ordered set.

2.2 Evenness

In the social sciences and in economics, questions involving evenness frequently arise. The econometric and linguistic terminology used by Dalton (1920), Muller (1964), Sen (1973), Allison (1978), Holmes (1985) and others will be converted in this paper to biological terminology. As such, sources, types and social classes will be named species; items, tokens and income will go by the general name of abundance. Magurran (1991) describes evenness roughly as 'how equally abundant species are' (p. 7). More precisely, evenness is a measure of the relative apportionment of abundances among the species present. This is certainly not the same as normalizing a so-called diversity measure (Pielou, 1975, p.15)!

A community is represented by the abundances of the constituent species. For example, (1, 4, 3, 10) represents four species with abundances of 1, 4, 3 and 10. These abundances can be expressed in biomass, individuals or any other quantity.

2.3 Construction of a Lorenz curve (Lorenz, 1905)

First, species are ranked according to their abundances (from low to high). Then cumulative proportions of the species as abscissae (X) are drawn against correspondingly

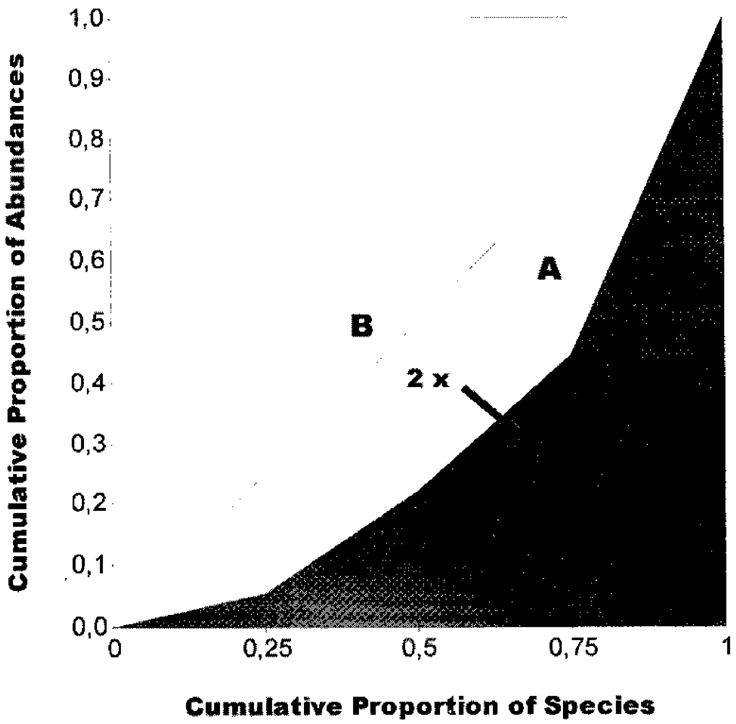

Figure 1. Lorenz curves: the Lorenz curve of A=(1,4,3,10) and the Lorenz curve of perfect evenness (B). Twice the area of the shaded region is equal to the Gini evenness index.

ranked cumulative proportions of species abundance as ordinates (Y). In our example, the vector (1,4,3,10) is first rewritten as (1/18, 3/18, 4/18, 10/18), then straight lines join the coordinates (0,0), (1/4,1/18), (2/4,4/18), (3/4,8/18) and (4/4,18/18). In Fig. 1, two Lorenz curves are drawn. Line B, the straight diagonal line, is the Lorenz curve of perfect evenness. This line is often taken as a reference line. If any change or shift occurs in this perfect distribution of abundances, the line will fall below the line of perfect evenness. Line A represents the Lorenz curve of (1,4,3,10).

If abundances are ranked from low to high, one obtains a convex curve, as shown in Fig. 1. Note that ranking abundances from high to low yields an equivalent representation as a concave Lorenz curve.

2.4 Ordering

Lorenz curves determine a partial order in the set of abundance vectors with the same length. Consider the set of all Lorenz curves drawn using a fixed number of species. Using the convex representation we say that curve L_1 is dominated by curve L_2 if for every x-value the corresponding y-value on L_1 is smaller than or equal to the corresponding value on L_2. Graphically, this means that the curve L_1 never lies strictly above L_2. This relation will be denoted as: L_1 -< L_2. The relation -< introduces a partial order in the set of Lorenz curves: a curve never lies above itself (L_1 -< L_1); if L_1 never lies strictly above L_2 (L_1 -< L_2), and L_2 never lies strictly above L_1 (L_2 -< L_1) then they clearly must coincide ($L_1 = L_2$), and finally, if L_1 -< L_2 and L_2 -< L_3 then L_1 -< L_3. The line of perfect equality is the largest one in this set, partially ordered by evenness. The ordering is only partial in the sense that two, or more, Lorenz curves that intersect are non-comparable. We will refer to this partial order as the Lorenz order.

2.5 Properties

Equivalent vectors have to be introduced to transfer this partial order to the set of all abundance vectors with the same length. Abundance vectors that differ only in the order of their components yield the same Lorenz curve, so they are said to be equivalent. This means that evenness is not a property of individual species, but of the community as a whole. We can also express this by saying that evenness, as a concept, is permutation invariant.

An evenness index, considered for the same community, but weighted once in pounds and once in kilos should render the same results (Dalton 1920, Bendel et al. 1989). This property is expressed by saying that evenness is scale invariant. As Lorenz curves are drawn using proportions, they are scale invariant. Consequently, abundance vectors that differ only by a proportionality factor are considered equivalent.

So bringing vectors such as (2,6,8), (1,3,4), and (3,1,4) into one equivalence class gives a one to one correspondence between abundance vectors and Lorenz curves. To make the presentation not unnecessarily complicated, we will always identify abundance vectors with their equivalence classes. Consequently, we say that vector X -< vector X' if, and only if, the Lorenz curve corresponding to X' dominates the Lorenz curve corresponding to X. Note that we have used the same notation (namely -<) for the two partial orders.

Dalton (1920) came up with another requirement: the transfer principle. Slightly adapted, this principle says that when the biomass of a less abundant species decreases in favor of the biomass of an already more abundant species, the evenness should decrease. It has been shown that the Lorenz order also meets the transfer principle (Rousseau 1992). Consequently, communities with the same number of species are (partially) ordered by the Lorenz order. This order on abundance vectors coincides with the majorization order as studied by mathematicians since the beginning of the century (e.g. Muirhead 1903). We will not go into details here, but refer the interested reader to the classical works of Hardy et al. (1952), Marshall and Olkin (1979), and Solomon (1979).

3. Evenness functions

In the previous section we have shown that evenness is best expressed as a partial order and that this structure can adequately be visualized by Lorenz curves. Historically, ecologists have tried to quantify evenness as a scalar. This procedure automatically maps the Lorenz order into a finer partial order. A mapping that realizes this transition is called an evenness function. Such a function F associates a non-negative number with each abundance vector. Moreover, it should respect the Lorenz order, -<, i.e. if X and X' are abundance vectors and X -< X' then $F(X) \le F(X')$. For a fixed number of species S, it can be shown that, e.g. the reciprocal of Simpson's index ($1/\lambda$), $1-\lambda$, the reciprocal of the coefficient of variation (1/V), Shannon-Wiener's diversity index (H') and G' = 1-G (where G denotes the Gini coefficient) are good evenness measures in the sense that they meet the three requirements: permutation invariance, scale invariance and the transfer principle (Rousseau 1992). The functions λ, V, H' and G' = 1-G are defined as follows: if $X = (x_i)_{i=1,\ldots,S}$ and $N = \Sigma x_i$ then:

$$\lambda(X) = \sum_{i=1}^{S} \left(\frac{x_i}{N}\right)^2$$

The coefficient of variation, denoted as V, is defined as the standard deviation, σ, divided by the mean, μ. Hence:

$$V = \frac{\sigma}{\mu} \text{ and hence } \frac{1}{V} = \frac{\mu}{\sigma}$$

The Shannon-Wiener diversity index or entropy index is defined as

$$H'(X) = -\sum_{i=1}^{S} \frac{x_i}{N} \ln \frac{x_i}{N}$$

and finally G' = 1 - G (referred further on as the Gini evenness index) is defined as:

$$G'(X) = \frac{2}{\mu S^2} \left(\sum_{i=1}^{S} (S+1-i)x_i \right) - \frac{1}{S}$$

where the x_i's are ranked from low to high and μ denotes the mean of the set $\{x_i\}$. It can be shown that G'(X) is equal to twice the area under the Lorenz curve (Fig. 1). Consequently, the value of the Gini evenness measure for the equality situation is equal to one.

Note that, e.g., the Brillouin index and McIntosh' index of diversity (Magurran 1991) do not meet the partial Lorenz order requirement, and therefore should not be used as evenness functions.

4. Evenness and a varying number of species

Until now, we have kept the number of species (S) fixed. Of course, in real situations, one wants to compare the evenness of communities with different numbers of species. To deal with this nothing has to be changed. Indeed, any abundance vector can be represented by a Lorenz curve, and any two Lorenz curves can be compared as we did for the case of a fixed number of species. An abundance vector X over S species is then considered to describe a more even situation than a vector X' over S' species if the Lorenz curve of the first never lies under the Lorenz curve of the second. We will refer to this partial order on all abundance vectors as the generalized Lorenz order. We will denote it as -<< (as it is, strictly speaking, a different order relation). Using Lorenz curves and the corresponding generalized Lorenz order leads to two important advantages. They are formulated as properties.

1) The inheritance property. This means that the restriction of the generalized Lorenz order to the case of fixed S, recovers the (usual) Lorenz order.

2) The replication property.

In the literature we find the following requirement (Dalton 1920; Hill 1973; Taillie 1979), called the replication axiom (or property): the evenness of a community equals the evenness of any replication of that community. In other words, the evenness of (1,3,4,10) is the same as the evenness of (1,3,4,10,1,3,4,10), or of (1,3,4,10,1,3,4,10,1,3,4,10,1, 3,4,10) etc... According to permutation invariance, this is the same as the evenness of (1,1,3,3,4,4,10,10), or of (1,1,1,1,3,3,3,3,4,4,4,4,10,10,10,10). Replication clearly has no influence on the Lorenz curve.

Functions that respect the Lorenz order with fixed S and yield the same value for replicated vectors respect -<< , i.e. the generalized Lorenz order. From the four evenness functions mentioned in the previous section only the Gini evenness measure G', and the reciprocal of the coefficient of variation, 1/V, respect the generalized Lorenz order (we refer the reader to Appendix A for a mathematical proof of this). However, using $1/(S\lambda)$ instead of $1/\lambda$, and 1/(ln(S) - H') yields measures that do confirm the generalized Lorenz order (see Appendix A). The function $1/(S\lambda)$ will be referred to as the modified Simpson index. The measure 1/V has one disadvantage with respect to the other ones, namely, it has no upper bound. This makes comparisons between evenness values somewhat more difficult. Yet, composing it with a strictly increasing function with range [0,1] yields an evenness measure which respects the generalized Lorenz order and which, moreover, takes values between zero and one (for a proof, see Rousseau, 1992, p.24-25). The most natural function to perform this transformation is the arctan function. Consequently, the resulting function

$$X \to \frac{2}{\pi} \arctan\left(\frac{1}{V(X)}\right)$$

yields a good evenness measure.

Generalizing this observation, we suggest normalizing evenness measures so that they take values between zero and one. Once a function that respects the generalized Lorenz order is given, such a normalization can always be performed.

An example of the use of some of these indices is presented in Appendix B.

5. Conclusions

Based on the previous investigations, we agree with Taillie (1979) that the Lorenz curve is the best possible representation of the notion of 'evenness'. This graphical representation induces a partial order in the set of abundance vectors or more precisely, the equivalence classes of abundance vectors. If one wants to associate a number expressing the evenness of an abundance vector then the Gini evenness index is probably the simplest acceptable function. We think, however, that for modeling purposes, one should not depend too much on this or other indices. The generalized Lorenz order is probably the best we can achieve concerning a consensus of a mathematical expression for evenness. Striving for a total order representing evenness is a futile endeavor. However, diversity as a concept combines the notions evenness and number of species. How this can be done will be studied in another article (Rousseau et al., 1998).

Acknowledgments. We thank Prof. J. Izsák (Szombathely, Hungary) for helpful comments on an earlier draft. We also thank the referees for pertinent and useful comments.

References

Allison, P.D. 1978. Measures of inequality. American Sociological Review 43: 865-880.

Bendel, R.B., S.S. Higgins, 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: 394-400.

Daget, J. 1976. Les modèles mathématiques en écologie. Masson, Paris.

Dalton, H. 1920. The measurement of the inequality of incomes. The Economic Journal 30: 348-361.

Hardy, G., J.E. Littlewood, and G. Pólya . 1952. Inequalities (sec.ed.). Cambridge University Press, Cambridge, England.

Hill, M.O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54: 427-432.

Holmes, D.I. 1985. The analysis of literary style - a review. Journal of the Royal Statistical Society A 148: 328-341.

Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577-586.

Lorenz, M.O. 1905. Methods of measuring concentration of wealth. Journal of the American Statistical Association 9: 209-219.

Magurran, A.E. 1991. Ecological diversity and its measurement. Chapman & Hall, London, England.

Marshall, A.W., and I. Olkin. 1979. Inequalities: theory of majorization and its applications. Academic Press, New York, NY, USA.

Muirhead, R.F. 1903. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. Proceedings of the Edinburgh Mathematical Society 21: 144-157.

Muller, C. 1964. Calcul des probabilités et calcul d'un vocabulaire. Travaux de Linguistique et de Litérature 2: 235-244.

Peet, R.K. 1975. Relative diversity indices. Ecology 56: 496-498.

Pielou, E.C. 1975. Ecological diversity. Wiley, New York.

Roberts, F. 1979. Measurement theory. Addison-Wesley,Reading, MA,USA.

Rousseau, R. 1992. Concentration and diversity in informetric research. Ph.D. thesis, Universitaire Instelling Antwerpen.

Rousseau, R. and P. Van Hecke 1998. Measuring biodiversity. Acta Biotheoretica (to appear).

Rousseau, R., P. Van Hecke, D. Nijssen and J. Bogaert. 1998. The relationship between diversity profiles, evenness and species richness, based on partial ordering. Environmental and Ecological Statistics (to appear).

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

Solomon, D.L. 1979. A comparative approach to species diversity. In: J.F. Grassle, G.P. Patil, W. Smith and C. Taillie, (eds.). Ecological Diversity in Theory and Practice. pp. 29- 35. International Co-operative Publishing House, Fairland, MD, USA.

Taillie,C. 1979. Species equitability: a comparative approach. In: J.F. Grassle, G.P. Patil, W. Smith and C. Taillie, (eds.). Ecological Diversity in Theory and Practice. pp. 51- 65. International Co-operative Publishing House, Fairland, MD, USA.

Appendix A

If $X = (x_1, x_2,...,x_S)$ is an abundance vector and c is a natural number, then we denote by cX the replication of X obtained by taking every x_i, $i = 1,..., S$, c times. A measure F meets the replication requirement if $F(X) = F(cX)$. We will check this requirement for $F = 1/\lambda$, $F = 1-\lambda$ $F = 1/S\lambda$, $F = 1/V$, $F = H'$, $F = 1/(\ln(S) - H')$ and $F = G' = 1-G$.

a) The reciprocal of Simpson's index ($1/\lambda$) *and* $1-\lambda$

$$\lambda(cX) = \sum_{i=1}^{S} c\left(\frac{a_i}{c}\right)^2 = \frac{\lambda(X)}{c}, \text{where } a_i = \frac{x_i}{\sum_{j=1}^{S} x_j} = \frac{x_i}{N}$$

Consequently, neither $1/\lambda$ nor $1-\lambda$ meet the replication requirement. Yet:

$$\frac{1}{cS\lambda(cX)} = \frac{c}{cS\lambda(X)} = \frac{1}{S\lambda(X)},$$

so that this function does meet the replication requirement. It will be referred to as the modified Simpson index.

b) The reciprocal of the coefficient of variation

It is clear that $\mu(X) = \mu(cX)$. Consequently,

$$V^2(cX) = \frac{\left(\frac{1}{cS}\sum_{i=1}^{S} cx_i^2 - \mu(X)^2\right)}{\frac{\sum_{i=1}^{S} cx_i}{cS}} = V^2(X)$$

Hence, V(cX) = V(X), and so 1/V(cX) = 1/V(X).

c) The Shannon-Wiener index H'

$$H'(cX) = -\sum_{i=1}^{S} c\left(\frac{x_i}{cN}\right)\ln\left(\frac{x_i}{cN}\right) = H'(X) + \ln(c)$$

This shows that H' does not satisfy the replication axiom. Yet,

$$\frac{1}{\ln(cS) - H'(cX)} = \frac{1}{\ln(c) + \ln(S) - H'(X) - \ln(c)} = \frac{1}{\ln(S) - H'(X)}$$

showing that this adapted Shannon-Wiener index satisfies the replication axiom.

Table 1. Index values for the Cocody Bay data (Daget, 1976)

Month	Modified Simpson	Gini evenness	Reciprocal Simpson
January	0.176	0.234	6.135
February	0.171	0.224	4.796
March	0.116	0.181	3.356
April	0.097	0.173	2.625
May	0.064	0.107	1.848
June	0.055	0.076	1.307
July	0.069	0.109	1.451
August	0.260	0.255	5.988
September	0.089	0.130	2.227
October	0.179	0.220	3.759
November	0.242	0.292	4.587
December	0.073	0.114	1.748

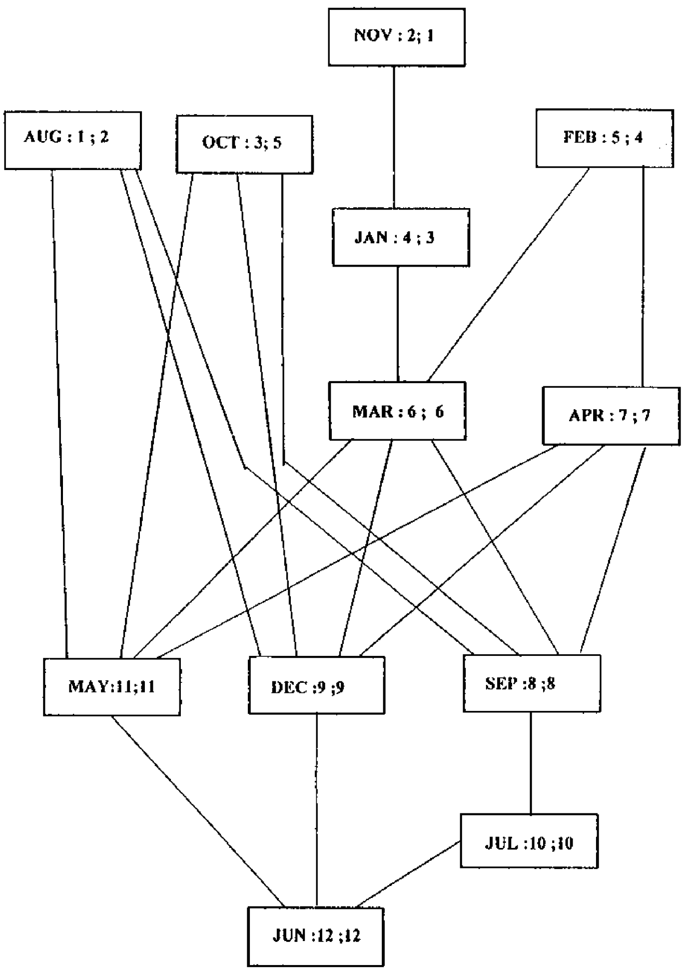

Figure 2. Evenness order for the Cocody Bay data. Each rectangle gives the month, the rank order according to the modified Simpson index and the rank order according to the Gini evenness index. It can be verified that both indices respect this partial order.

d) The Gini evenness index G'

As the Lorenz curve of X coincides with the Lorenz curve of cX, the areas under these Lorenz curves also coincide, hence G'(cX) = G'(X).

Appendix B

As an example we have analyzed the monthly data of fish sampled in Cocody Bay (Daget, 1976). Table 1 shows the monthly values of the modified Simpson index, the Gini evenness index and the reciprocal of the Simpson index. Fig.2 displays the (partial) evenness order, derived from Lorenz curves. This figure should be read as follows: Month A is more even than Month B, if Month A is situated higher than Month B, and if, moreover, they are connected. So, February is more even than April, while February and January are not comparable. It can easily be verified that both the modified Simpson index and the Gini evenness index confirm these orderings. The reciprocal Simpson index does not as shown by the values for January and November.