

## DESCRIPTIVE MEASURES OF ECOLOGICAL DIVERSITY

**B. V. Frosini**

*Institute of Statistics, Catholic University of Milan, Italy*

**Keywords:** Diversity, heterogeneity, homogeneity, concentration, inequality, evenness, equitability.

### Contents

1. Diversity, richness, evenness
    - 1.1. Introduction to Essential Properties
    - 1.2. A Comprehensive List of Diversity Indices
    - 1.3 Richness
    - 1.4 Evenness and Normalization of Diversity Indices
  2. General properties of diversity indices
    - 2.1. Statistical Proposals of General Properties
    - 2.2. Diversity and Inequality
    - 2.3 Majorization and Lorenz Curve
  3. Special indices and families of indices
    - 3.1. Gini-Simpson Indices and Generalizations
      - 3.1.1. Derivation of the Gini-Simpson Index
      - 3.1.2. A family of Indices Depending on the Euclidean Distance
      - 3.1.3. Rao's Family based on Dissimilarity Coefficients
    - 3.2. Diversity Indices based on Distances between Distributions
    - 3.3 Shannon and Entropy Measures
    - 3.4. Diversity Indices Derived from Averages
- Glossary  
Bibliography  
Biographical Sketch

### Summary

Ecological diversity relates to the different forms of life which are present in a particular site; in a more precise sense, it concerns the different species of a particular genus which are present in an ecological community. The measures, or indices, of ecological diversity, are statistical summaries of the abundance vector, that is, the frequencies or proportions of each species in the community. As a concept, diversity relates both to the number of species (richness) and to their apportionment within the community (evenness or equitability); other things being equal, there is greater diversity when the number of species grows, and when all the species are fairly represented. According to the aims pursued with the employment of diversity indices, some formal properties have been recognized essential for such indices; in the introductory Section 1.1, and in greater detail in Section 2.1, such properties are exposed and commented on. In agreement with these properties, the literature about diversity measures has proposed a great deal of particular instances, which answer different purposes; many of them are worked out for some artificial examples in Section 1.2. The likeness of diversity measures with some indices

used by economists in the study of income inequality is stressed. A special insight is devoted to some families of indices, derived from geometric distances between statistical distributions, from proposals of entropy measures, and as applications of Rao's approach based on dissimilarity coefficients, established in the pairwise comparison between species.

## 1. Diversity, Richness, Evenness

### 1.1. Introduction to Essential Properties

Given a number of many-species communities of the same kind (e.g. water algae, or beetles, in particular environments - see Magurran, 1988, p. 12), it is of interest in ecological studies to work out statistical summaries of the observations. As concerns evaluations and comparisons of diversity, the basic reference is a table like Table 1, which associates observed frequencies of individuals to the five species  $s_1, \dots, s_5$  in four different environments  $EN_1, \dots, EN_4$ ; to facilitate making comparisons, relative frequencies, or proportions, are also provided.

Sites Species	$EN_1$	$EN_2$	$EN_3$	$EN_4$	$EN_1$	$EN_2$	$EN_3$	$EN_4$
$s_1$	7	5	6	9	0.7	0.5	0.6	0.45
$s_2$	3	4	2	6	0.3	0.4	0.2	0.30
$s_3$	0	1	2	2	0	0.1	0.2	0.10
$s_4$	0	0	0	2	0	0	0	0.10
$s_5$	0	0	0	1	0	0	0	0.05
Total	10	10	10	20	1	1	1	1

Table 1. Hypothetical distributions of individuals of five species in four sites

In most examples, the absolute *dimension* of a species in a community is simply given by the *number* of individuals belonging to the species; in some cases, however, the biomass (or, alternatively, dry weight) or the surface covered is a suitable measure (Pielou, 1975, p. 6). In the latter case, when the cardinality of the dimension is lost, the arbitrariness of the unit measure (weight, volume, surface) to be employed renders all tables practically usable wholly equivalent to a table like the one composed of the last four columns of Table 1.

When the number of species is as low as four or five, and the number of sites to be compared is quite limited, it is not generally advisable to proceed to further data reduction: in fact, any process of data reduction causes us to lose some information, and can be responsible for a fading away of important aspects, of the original data. However, we are faced with large numbers of species in many studies, sometimes as high as a hundred. Thus, for summary and comparison purposes, it is practically inevitable to have recourse to one or a few summary statistics.

Summary statistics of general applicability are of the non-parametric kind; this article is specifically devoted to such kind of diversity indices or measures. This means that it is not required, for the validity content of an index or summary measure - that it is a parameter - or a function of the parameters - of the theoretical distribution or the stochastic process from which the sample observations (presumably) resulted. Of course, when the observations *for all the experiences* to be compared show a good fit with one and the same model distribution, there are good reasons to stick to some parameters of such distribution in order to judge and compare diversity between communities; but, also in these cases, a non-parametric measure could be profitably associated to the parametric one (for the main models used in ecological studies, see Pielou (1975) and Magurran (1988)).

Luckily enough, diversity indices or measures are among the best defined within descriptive statistical indices; postponing to Section 2 more formal definitions, in this introduction, it is sufficient to display the essential features of these indices. As there are indices depending on the absolute frequencies, and indices depending on relative frequencies (or biomass proportions), such essential requisites will be presented using both references. Given an absolute abundance vector  $n = (n_1, \dots, n_s)$ , with  $n_i$  = number of individuals (in the community) belonging to species  $i$  ( $N = \sum n_i$ ), and  $p = (p_1, \dots, p_s)$  the relative abundance vector ( $\sum p_i = 1$ ), the essential requisites of a diversity index  $I(p)$  or  $I^w(n)$  are as follows:

- (1)  $I(p) = I(p_1, \dots, p_s)$  and  $I^w(n) = I^w(n_1, \dots, n_s)$  are non-negative symmetric functions; i.e., they are constant over permutations of the elements of the vector  $p$  or  $n$ ;
- (2a)  $I(p)$ , as a function of the vector  $p$ ,  $p_i \geq 0$ ,  $\sum p_i = 1$  is a minimum when all except one of the proportions  $p_i$  are zero, the remaining being one;
- (2b)  $I^w(n)$ , as a function of the vector  $n$ ,  $n_i$  integer  $\geq 0$ ,  $\sum n_i = N$ , is a minimum when all except one of the frequencies  $n_i$  are zero, the remaining being  $N$ ;

In these situations the community is said to be *perfectly homogeneous*: only one species is present;

- (3a)  $I(p)$  is a maximum when all the proportions coincide, i.e. when  $p_i = 1/s \forall i$ ;
- (3b)  $I^w(n)$  is a maximum when all the frequencies coincide, i.e.  $n_i = N/s$ , *if this is possible*, that is, if  $N/s$  is an integer; otherwise,  $I^w(n)$  is a maximum when the vector  $n$  is the admissible vector nearest to  $(N/s, \dots, N/s)$ . Practically, if  $m < N/s < m+1$ , i.e.  $N/s$  is included between the integers  $m$  and  $m+1$ , the number  $h$  of species with frequency  $m$  in the maximizing vector is obtained by the equation  $mh + (m+1)(s-h) = N$ , namely  $h = s(m+1) - N$ ; the remaining  $s-h$  species have the frequency  $m+1$ . For example, if  $n = (5,4,1)$ ,  $N/s = 10/3$  is included between 3 and 4, so  $h = 3 \times 4 - 10 = 2$ ; the maximizing vector will be  $(4,3,3)$ , or a permutation of it.

The above maximization and minimization are implied by the following *coherence*

*property*, which will be given a more precise characterization in Section 2:

(4a)  $I(p)$  must increase (or, at least, not decrease) when some larger proportions are redistributed, so as to raise some smaller ones, aiming to approach the equalizing vector  $(1/s, \dots, 1/s)$ ;

(4b)  $I^w(n)$  must increase (or at least not decrease) when some larger frequencies are redistributed so as to raise some smaller ones, aiming to approach the equalizing vector, which is  $(N/s, \dots, N/s)$  whenever possible.

In the statistical literature, the indices satisfying the above property are called *heterogeneity indices* (see e.g. Leti, 1965); if the complementary property is used so that a function  $K(p)$  decreases when  $p$  approaches the equalizing vector,  $K(p)$  is called a *homogeneity index* (and also, a *concentration index*). A decreasing function of a homogeneity index is a heterogeneity index, and vice versa.

When the communities to be compared share the same number of species represented, only the different apportionment of the total number of individuals (or biomass) among the species is of interest for a judgment of diversity; in this case, the coherence property provides an *evenness* - or *equitability* - index, which orders the communities according to their *distance* from the evenness vector  $p = (1/s, \dots, 1/s)$ .

But a judgment on diversity depends as well - and first of all - on the number of species represented, when this number varies among communities: the larger the number of species, the higher the diversity of the community, other things being equal in some sense. One possibility to take into account of both components of diversity is to dispose of two kinds of indices: an index related to *evenness* (between species), and an index related to *richness* (of all the species represented). This point will be resumed at Sections 1.3 and 1.4)

A diversity index, instead, must be sensitive to both factors, thus must also be sensitive to the different number of species in two or more communities. In order to avoid having the two effects (evenness and richness) overlap, some kind of *ceteris paribus* condition must be imposed; the simpler one is as follows:

(5) Calling  $I_s(p)$  a diversity index applied to the case of  $s$  species represented,  $I_s(1/s, \dots, 1/s)$  must be an increasing function of  $s$ ; in other words, in cases of perfect evenness, the judgment about diversity is highest for the largest number of species.

Another, subtler condition about evenness comparability relates to mixtures of distributions. Let us consider  $h$  communities which are replicas of one another in the sense that they have the same absolute abundance vector, but no species in common (Hill, 1973, p. 429; Taillie, 1979, p. 55). Then the mixture of the  $h$  distributions *could* be judged of equal evenness as each component, while the richness has obviously increased; as a consequence, a diversity index is bound to increase as well. For example, if two communities have the same abundance vector  $(5,4,1)$  for different species, the mixture gives rise to the vector  $(5,5,4,4,1,1)$ . This proposal is interesting, but rather questionable; a finer examination of this point is postponed to Sections 2.2 and 2.3.

-  
-  
-

TO ACCESS ALL THE 29 PAGES OF THIS CHAPTER,  
Visit: <http://www.eolss.net/Eolss-sampleAllChapter.aspx>

### Bibliography

- Baczkowski A.J., Joanes D.N. & Shamia G.M. (1998). Range of validity of  $\alpha$  and  $\beta$  for a generalized diversity index  $H(\alpha, \beta)$  due to Good. *Mathematical Biosciences* 148, 115-128. [Entropy-based diversity measures]
- Bickel P.J. & Lehmann E.L. (1975). Descriptive statistics for non-parametric models. I. Introduction. *Annals of Statistics* 3, 1038-1044. [Descriptive measures and general properties]
- Brillouin L. (1962). *Science and Information Theory*. Second edition. New York: Academic Press. [Entropy-based diversity measures]
- Burgio G. (1969). Sulla misura dell'eterogeneità di un collettivo statistico. *Statistica* 29, 5-24. [Families of diversity indices]
- Dalton H. (1920). The measurement of the inequality of incomes. *Economic J.*30, 348-361. [Concentration (or inequality) indices]
- Daroczi Z. (1970). Generalized information functions. *Information and Control* 16, 35-51. [Entropy-based diversity measures]
- De Simoni S. (1979). Entropie non additive e misura dell'eterogeneità. *Quaderni dell'Istituto di Statistica dell'Università di Bologna*, n. 3, 341-387. [Entropy-based diversity measures]
- Ferreri C. (1980). Hypoentropy and related heterogeneity, divergency and information measures. *Statistica* 40, 155-168. [Entropy-based diversity measures. Families of diversity indices]
- Forcina A. & Giovagnoli A. (1982). Homogeneity indices and Schur-convex functions. *Statistica*, 42, 529-542. [Heterogeneity (diversity) indices, general properties]
- Frosini B.V. (1976a). Indici di variabilità applicati a fenomeni qualitativi. *Statistica* 36, 237-259. [A comparison of the different indices. Distance based diversity indices. Families of diversity indices. Heterogeneity (diversity) indices, general properties]
- Frosini B.V. (1976b). Gli indici di eterogeneità dal punto di vista geometrico. *Statistica* 36, 513-530. [A comparison of the different indices. Distance based diversity indices. Families of diversity indices. Heterogeneity (diversity) indices, general properties]
- Frosini B.V. (1981). Heterogeneity indices and distances between distributions. *Metron* 34, 95-108. [Distance based diversity indices. Heterogeneity (diversity) indices, general properties]
- Frosini B.V. (1987). Types and properties of inequality measures. In "M. Zenga ed., *La distribuzione personale del reddito*", Milano: Vita e Pensiero, 81-94. [Concentration (or inequality) indices. Heterogeneity (diversity) indices, general properties]
- Gini C. (1912). Variabilità e mutabilità. *Studi economico-giuridici della Facoltà di Giurisprudenza dell'Università di Cagliari*, Anno terzo, Parte seconda. [Concentration (or inequality) indices. Gini diversity index]
- Gini C. (1918). Di una estensione del concetto di scostamento medio e di alcune applicazioni alla misura

della variabilità dei caratteri qualitativi. *Atti del Regio Istituto Veneto di Scienze, Lettere ed Arti* 78, Parte seconda, 397-461. [Gini diversity index]

Good I.J. (1982). Discussion of the article by G.B. Patil & C. Taillie "Diversity as a concept and its measurement", *J. of the Amer. Statist. Assoc.* 77, 561-563. [A general paper on ecological diversity]

Herzel A. (1967). Considerazioni sulla variabilità. *Biblioteca del Metron* 3, 187-203. [Descriptive measures and, general properties. Heterogeneity (diversity) indices, general properties]

Herzel A. (1968). Alcuni problemi di massimi e minimi condizionati di interesse statistico. *Metron* 27, 69-96. [Descriptive measures and, general properties. Heterogeneity (diversity) indices, general properties]

Herzel A. (1977). Indici di eterogeneità o omogeneità, «coerenza» e convessità. *Statistica* 37, 3-12. [Descriptive measures and, general properties. Heterogeneity (diversity) indices, general properties]

Hill M.O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology* 54, 427-432. [Evenness. Families of diversity indices]

Hurlbert S.H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology* 52, 577-586. [A comparison of the different indices]

Leti G. (1965). Sull'entropia, su un indice del Gini e su altre misure dell'eterogeneità di un collettivo. *Metron* 24, 332-378. [Entropy-based diversity measures. Heterogeneity (diversity) indices, general properties]

Lloyd M. & Ghelardi R.J. (1964). A table for calculating the 'equitability' component of species diversity. *J. Anim. Ecology* 33, 217-55. [Models (for the distribution of species)]

Magurran A.E. (1988). *Ecological Diversity and Its Measurement*. London: Croom Helm. [A general book on ecological diversity. Models (for the distribution of species)]

Marshall A.W. & Olkin I. (1979). *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press. [Concentration (or inequality) indices]

Nayak T.K. (1985). On diversity measures based on entropy functions. *Commun. Statist. - Theor. Meth.* 14, 203-215. [Entropy-based diversity measures]

Patil G.P. & Taillie C. (1982). Diversity as a concept and its measurement. *J. Amer. Statist. Assoc.* 77, 548-567. [A general paper on ecological diversity and a comparison of the different indices. Concentration (or inequality) indices. Families of diversity indices. Heterogeneity (diversity) indices, general properties]

Peccati L. & Riva L. (1975). Alcune osservazioni concernenti gli indici di eterogeneità, in "Problemi di coerenza in statistica", Milano: Giuffrè, 41-58. [Heterogeneity (diversity) indices, general properties]

Peet R.K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics* 5, 285-307. [A general paper on ecological diversity. Entropy-based diversity measures]

Pielou E.C. (1975). *Ecological Diversity*. New York: Wiley. [A general book on ecological diversity. Entropy-based diversity measures. Models (for the distribution of species)]

Rao C.R. (1982a). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhya A* 44, 1-22. [Families of diversity indices]

Rao C.R. (1982b). Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biology* 21, 24-43. [Families of diversity indices]

Rao C.R. (1986). Rao's axiomatization of diversity measures. In "*Encyclopedia of Statistical Sciences*", New York: Wiley, Vol. 7, 614-617. [Families of diversity indices]

Rényi A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 547-561. [Entropy-based diversity measures. Families of diversity indices]

Sen A. (1973). *On Economic Inequality*. Oxford: Clarendon Press.

Shannon C.E. & Weaver W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press. [Entropy-based diversity measures]

Simpson E.H (1949). Measurement of diversity. *Nature* 163, 688. [Gini diversity index]

Solomon D.L. (1979). A comparative approach to species diversity. In “*Ecological Diversity in Theory and Practice*”, J.F. Grassle, G.P. Patil, W. Smith & C. Taillie eds, Fairland, Maryland: International Cooperative Publishing House, 29-35. [Evenness]

Taillie C. (1979). Species equitability: A comparative approach. In “*Ecological Diversity in Theory and Practice*”, J.F. Grassle, G.P. Patil, W. Smith & C. Taillie eds, Fairland, Maryland: International Cooperative Publishing House, 51-61. [Concentration (or inequality) indices. Evenness]

Theil H. (1967). *Economics and Information Theory*. Amsterdam: North Holland. [Concentration (or inequality) indices]

### **Biographical Sketch**

**Benito V. Frosini** is a professor of statistics at the Catholic University of Milan, Italy. His research interests have been mostly directed towards statistical methods. The main research topics have been the following: (1) within Descriptive statistics (about populations or random variables): Variability and concentration, Inequality measures and their decomposition, Heterogeneity (diversity) indices, Asymmetry indices; (2) within Estimation theory: Parametric estimation for the lognormal distribution, Cramér-Rao inequality in the multiparameter case, Confidence intervals vs. Bayesian intervals; (3) within Test theory: Power function of the Chi-square test, EDF goodness-of-fit tests, Conditional tests vs. global tests; Foundations: Likelihood principle, Coherency of statistical decisions, Risk assessment.