Multidimensional indices and orders of diversity[*]

Karl Mosler

Universität zu Köln

Albertus-Magnus-Platz, 50923 Köln, Germany

E-mail: mosler@wiso.uni-koeln.de

November 2001

Keywords: Multivariate diversity and evenness indices, majorization, concentration order, Gini-Simpson index

Abstract: The paper presents several indices to describe multivariate diversity and evenness. Multivariate generalizations of the Gini-Simpson index and the Rosenbluth index are proposed to measure diversity. Further a multivariate Gini ratio to measure evenness is presented. These indices extend the usual univariate measures; they reflect not only the diversity of marginal distributions but also the dependence structure of abundance. The indices fulfill desirable measurement properties and are consistent with certain orderings of multivariate distributions. An order of concentration surfaces and a majorization order are also shortly surveyed.

1. Introduction

Consider a community of species living in a space that has several dimensions. The dimensions may relate to limited resources or to attributes of importance. For example, resource dimensions may be light and humidity for plants or prey and space for animals. Examples of importance dimensions are population density (or total number) and species biomass for plants as well as for animals. Our aim is to measure the α-diversity of the given community and to compare it with the diversities of other communities. The measures – indices and orders – shall be based not only on one dimension but on several dimensions simultaneously and they shall also reflect possible interactions between the dimensions.

There exists a classical literature on the univariate measurement of diversity; see e.g. the comprehensive and unified treatments by Pielou (1977), Solomon (1979), Rao (1982), Patil and Taillie (1982) and many recent more special studies. Tóthmérész(1995) surveys and discusses twelve univariate methods of diversity ordering. But there exist rather few approaches to multivariate measurement, e.g., Alatalo and Alatalo (1977) and Loehle and Wein (1994).

Section 2 treats the univariate measurement. Several popular diversity indices, the concentration curve and concentration order, two evenness indices, and the Lorenz order are surveyed. In Section 3 the multivariate Gini-Simpson index is introduced. Section 4 presents a multivariate majorization, called relative dilation, and a multivariate concentration order together with monotonicity properties of indices including a multivariate Rosenbluth diversity index. Section 5 discusses a multivariate Gini ratio and Section 6 concludes.

2. Univariate indices and orders

If just one resource or one importance value is relevant, a univariate measurement problem has to be solved. Many indices have been proposed and used in the literature to measure the diversity of a given community in a one-dimensional space. Let the community consist of S species and each of them use a share p_i of the limited resource (or bear relative importance p_i), $i = 1, 2, \ldots, S$.

2.1 Indices of Diversity

The following diversity indices are widely used in Ecology,

- *number of species S ,*

- *Shannon-Wiener index*
$$D_1 = -\sum_{i=1}^{S} p_i \log p_i \,,$$

- *Gini-Simpson index*
$$D_2 = -\log \sum_{i=1}^{S} p_i^2 \,,$$

- *Renyi's α-entropy*
$$D_\alpha = \frac{1}{1-\alpha} \log \sum_{i=1}^{S} p_i^\alpha \,, \quad \alpha > 0 \,,$$

where $\log y$ denotes the natural logarithm of y. The family D_α includes the Gini-Simpson index and, in the limit $\alpha \to 1$, also the Shannon-Wiener index. Renyi's D_α is frequently used in the following variant, which is mentioned as the

- *Hill index*

$$H_\alpha = \exp(D_\alpha) = \left(\sum_{i=1}^{S} p_i^\alpha \right)^{\frac{1}{1-\alpha}}, \quad \alpha > 0.$$

To measure concentration, which is the opposite of diversity, the *Rosenbluth concentration index* A is well known in Economics,

$$A = \frac{1}{S(1-R)},$$

where $R = \frac{1}{S} \sum_{i=1}^{S} (2i - S - 1) p_{(i)}$ and $p_{(i)}$ is the i-th share ordered from below, $p_{(1)} \le p_{(2)} \le \cdots \le p_{(S)}$. To measure diversity, we introduce the

- *Rosenbluth diversity index*

$$B = \frac{1-A}{1+A} = \frac{S(1-R)-1}{S(1-R)+1}.$$

All these diversity indices measure *which shares* of the resource are *used by how many* species.

2.2 Concentration order

The distribution of resources among species is nicely described by a graphical device, the concentration curve (which is also popular in Economics): Calculate the

- *Concentration Rates*

$$C_i = p_{(S)} + \cdots + p_{(S-i+1)}, \quad i = 1, 2, \ldots, S,$$

and draw the points

$$(0,0), (C_1, 1), \ldots, (C_{S-1}, S-1), (1, S)$$

in two-space and connect them by a piecewise linear curve. This curve is the *concentration curve*. For each i, the i-th concentration rate C_i equals the total share

3

of the resource held by the i "largest" species; its negative, $-C_i$, serves as another index of diversity. The area G below the concentration curve is connected with the Rosenbluth diversity index B, there holds $B = (2G-1)/(2G+1)$.

When two communities P and Q are compared, the first having species number S and shares $p_1, \ldots p_S$ and the second having species number T and shares $q_1, \ldots q_T$, both concentration curves may be drawn. If the concentration curve of the second lies *below* that of the first, we say that Q majorizes P in *concentration order*. This is tantamount saying that all concentration rates are larger with Q than with P; the second community is more concentrated, that is *less diverse,* than the first. Obviously, the concentration order implies that $S \geq T$. Actually, it implies much more, namely:

Proposition 1: If of two communities one is less diverse than the other in the concentration order then each of the above mentioned diversity indices has a smaller value at the first community than at the second.

The proposition is known. For A it follows from the fact that B is an increasing function of the area below the concentration curve. For D_α and, hence, for H_α it follows from standard results on majorization (Marshall and Olkin 1979, Tong 1983, Mosler 1994).

As the concentration order has an obvious interpretation in terms of diversity and in view of the Proposition, it seems rather natural to postulate for *any univariate index of diversity* that it *should decrease with the concentration order.*

2.3 Indices of Evenness

The measurement of *evenness* abstracts from the number of species present in the community but focusses on their relative importance. It quantifies the slope of the importance value sequence and measures its flatness (Whittaker 1972, pp 217 ff).

The question is here *which shares* of the resource are *used by which parts* of the species. Univariate indices that measure *non-evenness* are

- *Gini's ratio*

$$R = \frac{1}{S}\sum_{i=1}^{S}(2i - S - 1)p_{(i)}\,,$$

- *Coefficient of variation*

$$v = \sqrt{S\sum_{i=1}^{S}p_i^2 - 1}\,.$$

4

The coefficient of variation equals the standard deviation of the p_i divided by their mean. Indices of *evenness* are obtained by going either to the negative or the reciprocal of R and v.

Both the Gini ratio R and the coefficient of variation v are closely related with diversity indices, R with the Rosenbluth diversity index B, and v with the Gini-Simpson index D_2,

$$D_2 = -\log \frac{v^2 + 1}{S}, \tag{1}$$

$$B = \frac{S(1-R) - 1}{S(1-R) + 1}. \tag{2}$$

Equation (1) will be employed in the sequel to define a proper multivariate extension of the index D_2.

The relevant partial order in the univariate measurement of evenness is the *Lorenz order*, which is closely related to the concentration order. The piecewise linear connection of the points

$$\left(0,0\right), \left(\frac{1}{S}, 1 - C_{S-1}\right), \left(\frac{2}{S}, 1 - C_{S-2}\right), \ldots, \left(\frac{S-1}{S}, 1 - C_1\right), \left(1,1\right)$$

in two-space is the *Lorenz curve* of P with shares $p_1, \ldots, p_S$. If the Lorenz curve of a community Q lies *below* that of a community P, then Q is said to *majorize P in Lorenz order*. Note that in case $S = T$ this is equivalent to majorizing in the concentration order.

A natural postulate for *any index of evenness* is that it *should decrease with the Lorenz order*. As the usual Gini ratio equals twice the area between the Lorenz curve and the diagonal of the unit square, the negative (or reciprocal) of the Gini ratio fulfills that postulate. The same does the negative (or reciprocal) of the coefficient of variation.

3. A multivariate Gini-Simpson index

Now, we turn to the problem of measuring diversity when there is more than one dimension of resources or importance variables. Assume there are S species and J resources and let p_{ij} denote the share of resource j that is used by species i, $\sum_{i=1}^{S} p_{ij} = 1$ for all j. Assume that every species uses at least one resource, that is, $\min_j p_{ij} > 0$ for every i. The shares of each species are collected in a vector $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iJ})$, $i = 1, \ldots, S$.

We propose a multivariate Gini-Simpson index that is based on a multivariate standardized variance as follows. Define, for $j, k = 1, \ldots, J$,

$$\sigma_{jk}^* = S \sum_{i=1}^{S} p_{ij} p_{ik} - 1 \,,$$

which corresponds to the covariance between the resources j and k, and let $\mathbf{\Sigma}^* = [\sigma_{jk}^*]$ be the respective $J \times J$ matrix. More precisely, σ_{jk}^* is S^2 times the covariance between the vectors of shares which belong to resources j and k. The J-variate *Gini-Simpson index* is defined by (1) with $v = \sqrt{\det \mathbf{\Sigma}^*}$,

$$D_2 = - \log \frac{\det \mathbf{\Sigma}^* + 1}{S} \,,$$

and the corresponding *Hill index* is

$$H_2 = e^{D_2} = \frac{S}{\det \mathbf{\Sigma}^* + 1} \,.$$

Several properties are desirable postulates with any index of diversity or evenness:

- *Anonymity*: The index does not change if the species are renamed.

- *Scale invariance*: The index does not change if the scale of measurement is changed in one or several attributes.

- *Continuity*: The index is continuous in all its arguments.

Proposition 2: The multivariate Gini-Simpson index D_2 and the multivariate Hill index H_2 fulfill the postulates of anonymity, scale invariance, and continuity.

The proposition is obvious. The ranges of the indices are

$$\left] \log \frac{S}{J(S-1)+1}, \ \log S \right[\quad \text{for} \quad D_2 \,,$$

$$\left] \frac{S}{J(S-1)+1}, \ S \right[\quad \text{for} \quad H_2 \,,$$

where $]\cdot, \cdot[$ denotes the open interval. Further, any index of diversity or evenness is supposed to increase with so-called *regressive transfers*. In the next section we introduce a multivariate version of such transfers and state that D_2 and H_2 increase with them.

4. Multivariate orders of diversity and evenness

We first define a majorization order, relative dilation, for comparisons of evenness. This order is also relevant for diversity comparisons when the number of species does not change. After, we define a concentration order for comparisons of diversity when the numbers of species are different.

4.1 Multivariate dilations

Consider two communities P and Q which are characterized by their sequences of share vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_S$ and $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_T$, respectively. Note that the number of species, S, in P may differ from the number of species, T, in Q. Community Q is said to be a *relative dilation* of P if there exist numbers $\beta_{ik} \geq 0$ with

$$\boldsymbol{p}_i = \frac{T}{S} \sum_{k=1}^{T} \beta_{ik} \boldsymbol{q}_k \,, \quad \sum_{k=1}^{T} \beta_{ik} = 1 \,, \quad \sum_{i=1}^{S} \beta_{ik} = \frac{S}{T} \,. \tag{3}$$

For example let

$$\boldsymbol{p}_1 = (0.6, 0.48)\,, \quad \boldsymbol{p}_2 = (0.4, 0.52)\,,$$
$$\boldsymbol{q}_1 = (0.3, 0.2)\,, \quad \boldsymbol{q}_2 = (0.5, 0.2)\,, \quad \boldsymbol{q}_3 = (0.1, 0.4)\,, \quad \boldsymbol{q}_4 = (0.1, 0.2)\,.$$

Then with

$$\beta_{11} = 0.0\,, \quad \beta_{12} = 0.5\,, \quad \beta_{13} = 0.2\,, \quad \beta_{14} = 0.3\,,$$
$$\beta_{21} = 0.5\,, \quad \beta_{22} = 0.0\,, \quad \beta_{23} = 0.3\,, \quad \beta_{24} = 0.2\,,$$

we see that Q is a relative dilation of P.

Relative dilation may be interpreted in a dynamic context as follows: The community Q changes to community P while each species i in P takes part $\frac{T}{S}\beta_{ik}$ of the resources from each species k in Q. As a simple special case, let $S = T$ and assume that all species do not change resources besides two, say species $i = 1$ and 2. This means that $\beta_{ik} = 0$ if $i, k \notin \{1, 2\}$. In this case species 1 obtains some part β_{12} of the other species' resource vector and retains part β_{11} of its own, and species 2 viceversa. The special case is called a *simple regressive transfer* from Q to P. Any relative dilation with $S = T$ can be split into a reverse series of simple regressive transfers, each from one single species to another. A general relative dilation with $S \neq T$ differs from the previous by a rescaling, which multiplies the Q-shares with the factor $\frac{T}{S}$; for details, see Koshevoy and Mosler (1997).

Relative dilation induces a partial order between communities. When there is only one resource, $J = 1$, this order is equivalent to the usual Lorenz order of $T \cdot \boldsymbol{q}_1, \ldots, T \cdot \boldsymbol{q}_T$ over $S \cdot \boldsymbol{p}_1, \ldots, S \cdot \boldsymbol{p}_S$. Multivariate relative dilation implies relative dilation in every single dimension.

Increasingness with respect to a simple regressive transfer is a postulate for any index of evenness as well as diversity:

- *Regressive Transfer*: The index increases if two species exchange parts of their resources.

Proposition 3: The multivariate Gini-Simpson index decreases with relative dilations. In particular, it satisfies the postulate of regressive transfer.

The proposition is proved by standard majorization arguments (e.g. Mosler 1994).

4.2 Multivariate concentration order

To introduce multivariate versions of the concentration curve and the concentration order we first define a *concentration function* $c(\boldsymbol{z})$ with argument $\boldsymbol{z}$ in the J-variate cube,

$$c(\boldsymbol{z}) = \min \sum_{i=1}^{S} \theta_i \,, \quad \boldsymbol{z} \in [0,1]^J \,,$$

where the minimum is taken over all $\theta_1, \ldots \theta_S$ with

$$\sum_{i=1}^{S} \theta_i p_{ij} \geq z_j \quad \text{for all } j \,.$$

The graph of $c(\boldsymbol{z})$ is a set in $(J+1)$-space. If $J = 2$ it is named the *concentration surface*, in larger dimensions the *concentration hypersurface*.

Let us explain the concentration hypersurface. For a given vector of resource shares, $\boldsymbol{z} \in [0,1]^J$, determine the minimal size of an aggregate species which uses these shares or more. The aggregate species consists of portions θ_i of each species i, and the sum of the portions is the size of the aggregate species. Its minimum size defines the value of the J-variate concentration function at $\boldsymbol{z}$. Note that in case $J = 1$ the concentration hypersurface is just the usual concentration curve; see Figure 1 below.

To compare two communities with respect to their diversity, their concentration functions may be employed. If one concentration function is smaller than the other at every argument $\boldsymbol{z}$, the concentration hypersurfaces are ordered. We call this the multivariate *concentration order*.

It can be proved that relative dilation with equal species number implies concentration order: If $S = T$ and Q is a relative dilation of P then the concentration function of Q is smaller than that of P. Moreover, it can be shown that the index D_2 is decreases with the concentration order:

8

Proposition 4: If $S \geq T$ and the concentration hypersurface of Q is below that of P then the Gini-Simpson index D_2 of Q is smaller than that of P.

4.3 A numerical example

Consider a community of $S = 8$ species with shares in two resource dimensions. Let the transpose of the 8×2 matrix of shares be given by

$$[p_{ij}]' = \begin{bmatrix} .0179 & .1519 & .0521 & .1715 & .1563 & .2348 & .1915 & .0238 \\ .0065 & .0877 & .0214 & .2809 & .0715 & .3155 & .1781 & .0384 \end{bmatrix}.$$

Figure 1 exhibits the concentration curve of the first diension alone, and Figure$\tilde{2}$ the concentration surface of both dimensions. The figures are reproduced from Koshevoy and Mosler (1999). Curves like these may also be mentioned as *diversity profiles*; the term was proposed by Patil and Taillie (1979).

4.4 A multivariate Rosenbluth diversity index

As in the univariate case the Rosenbluth diversity index is determined by the area below the concentration curve, in dimension $J \geq 2$, the same can be done with the concentration hypersurface. Let G denote the volume below the concentration hypersurface. A multivariate *Rosenbluth diversity index* is defined by

$$B = \frac{G - \frac{J}{J+1}}{G + \frac{J}{J+1}}.$$
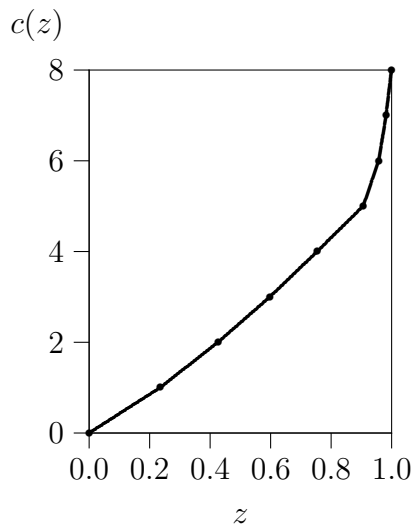
Concerning diversity orderings there holds:

FIGURE 1: The concentration curve for one dimension.

9

file=mconfig3.ps,height=11cm,width=11cm,angle=-90

FIGURE 2: The concentration hypersurface for two dimensions.

Proposition 5:
(i) The multivariate Rosenbluth diversity index B satisfies the postulates of anonymity, scale invariance, and continuity. There holds always $B \geq 0$. The maximum $B = \frac{S-1}{S+1}$ is achieved if and only if all share vectors are identical, that is $\boldsymbol{p}_i = (\frac{1}{S}, \ldots, \frac{1}{S})$ for all species i.
(ii) B decreases with the multivariate concentration order and, if $S = T$, also with the relative dilation order.

Part (i) and monotonicity in the concentration order are immediately seen from the definition. Many other properties of the multivariate concentration order are found in Koshevoy and Mosler (1999).

5. A multivariate Gini ratio

To measure evenness in more than one dimension, various extensions of the usual Gini ratio are possible. Instead of defining a multivariate Gini ratio through Equation (2) and the above multivariate Rosenbluth index, we introduce a notion which is easier to handle.

The multivariate *Gini ratio* is given by

$$R = \frac{1}{2S^2 J} \sum_{i=1}^{S} \sum_{k=1}^{S} ||\boldsymbol{p}_i - \boldsymbol{p}_k||,$$

with $||\boldsymbol{p}_i - \boldsymbol{p}_k|| = \left(\sum_{j=1}^{J} (p_{ij} - p_{kj})^2 \right)^{\frac{1}{2}}$ being the Euclidean distance. Note that R reduces to the usual Gini ratio when $J = 1$.

Proposition 6:
(i) The multivariate Gini ratio R satisfies the postulates of anonymity, scale invariance, and continuity. There holds always $R \geq 0$. The minimum $R = 0$ is achieved if and only if all share vectors are identical.
(ii) R has the following ceteris paribus *property: If we add a $(J+1)$-th dimension in which the species do not differ, the Gini ratio changes by a factor $\frac{J}{J+1}$ only.*
(iii) R is strictly increasing with relative dilation.

Thus, as an index of evenness, the negative (or the reciprocal) of the multivariate Gini ratio increases with the reverse of a relative dilation. Part (i) of the proposition is immediate from the definition; for (ii) and (iii) see Koshevoy and Mosler (1997).

The question remains in which respect an index of evenness like R differs from an index of diversity like D_2 or B. Consider the following postulates:

- *Replication invariance*: The index remains unchanged if the community is replicated to $k \geq 2$ similar communities, that is, if each species i, which holds a vector $(p_{i1}, \ldots, p_{iJ})$ of resource shares, is substituted by k different species each holding a vector $(\frac{p_{i1}}{k}, \ldots, \frac{p_{iJ}}{k})$.

- *Replication monotonicity*: The index increases strictly if the community is replicated to $k \geq 2$ similar communities.

- *Cardinal replication monotonicity*: The index increases by a factor k if the community is replicated to $k \geq 2$ similar communities.

Proposition 7:
(i) R satisfies the postulate of replication invariance.
(ii) D_2, H_2 and B satisfy the postulate of replication monotonicity.
(iii) H_2 satisfies also the postulate of cardinal replication monotonicity, but D_2 and B do not.

The proposition is easily verified from the definitions. Thus, the principal difference between diversity and evenness is that a diversity index increases under replication, while an evenness index remains constant.

6. Concluding remarks

To compare the diversity and evenness of communities, we have presented several indices and orderings and have discussed their relations. Also we have introduced a new diversity index, the multivariate Gini-Simpson index, and investigated some of its properties.

Throughout the paper, diversity and evenness have been compared at given points in time or space. All measures presented here are local in the sense that no resource gradients and no allocation of species in some niche hyperspace are considered. They are also static in the sense that no evolution of diversity over time is modelled.

Several variants of the above multivariate indices make sense. Other Gini-Simpson indices are obtained by plugging other notions of multivariate variance into Equation (1). In particular, the trace of the matrix $\boldsymbol{\Sigma}^*$ may be used in place of the determinant; then a multivariate Gini-Simpson index, $\widetilde{D}_2$, is obtained which neglects possible interactions between the dimensions. Note that a properly transformed index, $\exp(\widetilde{D}_2)$, can be additively decomposed into the diversities of single dimensions.

Also there are different ways to extend the Gini ratio to several dimensions. As the univariate Gini ratio equals twice the area between the Lorenz curve in two-space, a J-variate Gini ratio has been defined as the volume of a certain convex body in $(J+1)$-space (Mosler 1994, Koshevoy and Mosler 1997); this Gini ratio has similar properties as the above one but needs heavier computations. Also, distances other than the Euclidean may be employed in the multivariate Gini ratio, e.g. with different weights in the dimensions.

Multivariate entropy indices have been proposed and investigated by Maasoumi (1986).

References

Alatalo, R. and R. Alatalo. 1977. Components of diversity: Multivariate analysis with interaction. Ecology 58: 900–906.

Koshevoy, G. and K. Mosler. 1997. Multivariate Gini indices. Journal of Multivariate Analysis 60: 252–276.

Koshevoy, G. and K. Mosler. 1999. Measuring multidimensional concentration: A geometric approach. Allgemeines Statistisches Archiv 83: 173–189.

Loehle, C. and G. Wein. 1994. Landscape habitat diversity: A multiscale information theory approach. Ecological Modelling 73: 311–329.

Maasoumi, E. 1986. The measurement and decomposition of multi-dimensional inequality. Econometrica 54: 991–997.

Marshall, A.W. and I. Olkin. 1979. Inequalities: Theory of Majorization and its Applications. Academic Press, New York.

Mosler, K. 1994. Majorization in economic disparity measures. Linear Algebra and Its Applications 199: 91–114.

Patil, G.P. and C. Taillie. 1982. Diversity as a concept and its measurement. JASA 77: 548–561.

Pielou, E.C. 1977. Mathematical Ecology. Wiley, New York.

Rao, C.R. 1982. Diversity and dissimilarity coefficients: A unified approach. Theoretical Population Biology 21: 24–43.

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

Tong, Y.L. 1983. Some distribution properties of the sample species-diversity indices and their applications. Biometrics 39: 999–1008.

Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. Journal of Vegetation Science 6: 283–290.

Whittaker, R.H. 1972. Evolution and measurement of species diversity. Taxon 21: 213–251.