Parameter dependence of correlation between the Shannon index and members of parametric diversity index family

János Izsák *

Department of Zoology, Berzsenyi College, Károlyi Gáspár tér 4, Szombathely 9701, Hungary

Abstract

We present detailed data on correlation of the Shannon diversity index with members of Hill's and Hurlbert's parametric index family. The correlation reaches its maximum in both cases at a definite parameter value. These specified data on correlation are exploitable when applying the above diversity indices. Namely, in case of their parallel use with the Shannon index a strong correlation between the indices should be avoided. The investigations were carried out on sets of moth and fly collections.

Moreover, we hypothesized that a smaller distance between so-called sensitivity profiles of diversity indices predicts a larger correlation between the indices concerned. Based on investigations of the sensitivity profile distance between the Shannon index and members of parametric diversity index family as a function of index parameter, we point out that on the whole the above hypothesis is true. This makes possible to estimate the correlation of two diversity indices omitting troublesome index calculations on a large set of abundance lists.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Correlation; Diversity indices; Hill's indices; Hurlbert's indices; Index sensitivity; Shannon index

1. Introduction

Diversity indices are classic scalar ecological indicators. The application of these indices is common in ecological analysis. A frequent finding is that species diversity indicates the status of the ecosystem or community (Fernández-Aláez et al., 2002; Diserud and Aagaard, 2002; Park et al., 2003; Salas et al., 2005) and, in a wider sense, the quality of the living environment (Gaston and Spicer, 2004). Moreover, a high species diversity contributes to the stability of the ecosystem (Sankaran and McNaughton, 1999; Naeem and Baker, 2005; Kiessling, 2005; Moore, 2005). In addition to hundreds of articles, several books discuss diversity aspects in ecology and the methodology of diversity measures (Pielou, 1975; Grassle et al., 1979; Patrick, 1983; Magurran, 1988, 2004).

The parallel use of some diversity indices is a general praxis. However, the well-known correlation between most diversity indices is considerable. As Ricklefs (1990) writes: ''. . . the results of most studies

* Tel.: +36 94 504 300; fax: +36 94 504 404.
  E-mail address: ijanos@bdtf.hu.

are relatively insensitive to which index of diversity is applied …''. If one interprets this correlation as the sign of an abstract diversity property of the community, so this correlation is reassuring. However, if the correlation expresses the fact that diversity indices inform almost uniformly on the level of diversity, then from a practical point of view the introduction of more and more new diversity indices seems to be questionable.

In her popular book Magurran (1988) summarized data on diversity index correlation. In the recent past Bryja and Kula (2000) analyzed Spearman's rank correlations of numerous diversity indices on bug collections. However, on the whole the investigations of diversity index correlation are sporadic and phenomenological. Our investigations are aimed to make a first step to a more systematic and formalized treatment of this topic. The correlation of diversity indices can be reduced to the property that the usually applied diversity indices measure mainly the dominance structure of the communities. Realizing this one-sidedness, some authors suggested the use of diversity indices more sensitive to changes in the range of the middle and small species abundances. Such indices are, for example, some members of Hill's and Hurlbert's index family.

Hill's index family can be associated with the well-known Shannon diversity index or multinomial entropy defined by the formula:

$$H'(p_1, \ldots, p_s) = -\sum_{i=1}^{s} p_i \ln p_i,$$

where s is the number of the occurring species in the sample and p_i is the theoretical probability of occurrence of species i; $p_1 + p_2 + \cdots + p_s = 1$ (Pielou, 1975). The quantity $H'(p_1, \ldots, p_s)$ expresses the information gain when one is informed on the distribution $(p_1, \ldots, p_s)$ after the a priori hypothesis on the uniform distribution $(1/s, \ldots, 1/s)$ in the sample. A generalization of index H' is Rényi's generalized entropy index or a-order entropy family with the formula

$$R_a = \frac{\ln \sum_{i=1}^{s} p_i^a}{1-a}, \quad a \neq 1, \ R_1 = \lim_1 R_a = H'$$

(Rényi, 1961). Hill's diversity index family (Hill, 1973), which is well-known in the literature of eco-logical diversity, is defined by the formula

$$N_a = \left(\sum_{i=1}^{s} p_i^a \right)^{1/(1-a)},$$

$$a \neq 1, N_1 = \lim_1 N_a = \exp(H').$$

Thus, N_a equals to $\exp R_a$.

Clearly, the N_a indices also have an information theoretical meaning, as they are directly related to Rényi's indices. A basic property of Hill's indices is that by decreasing the (positive) parameter a the index stresses more and more the distribution conditions in the range of the small species abundances. An extreme illustration can be that $N_0 = \lim_1 N_a = s$ for any abundance distribution. That is, upon the emergence of a new species in the population, represented if only by a single individual, the value of N_0 increases from s to $s + 1$, regardless of the size of other abundances. On the other hand, it is easy to show (see Appendix A) that $N_\infty = \lim_\infty N_a = 1/p_{i,\max}$, which is the reciprocal of the Berger–Parker concentration index. Apparently, N_∞ is sensitive only to changes in the maximum species probability. The application of N_a indices, mainly N_1, N_2 is frequent in the statistical ecology, although the application of N_2 emerges often in another context. Namely, $N_2 = (\sum p_i^2)^{1/(1-2)} = 1/\sum p_i^2$, which is the well-known reciprocal Simpson index. The sporadic application of Hill's indices with further parameter values can be reduced to insufficiency of data about the relation between the statistical behavior of N_a and frequently used diversity indices, such as H' index or Fisher's alpha index, etc.

The diversity measures $s(m)$, $m = 2$, 3, … (Hurlbert, 1971) were introduced originally as so-called *rarefaction indices* (Simberloff, 1979). Let us choose randomly m individuals from the population with species probabilities p_1, p_2, …, p_s ($p_1 + p_2 + \cdots + p_s = 1$). Let S_m be the probability variable relating to the number of represented species in the sample. It is easy to show that the expectation of the number of species in the sample is $ES_m = s(m)$.

Particularly, $s(2)$ is the mean of the number of species represented in a sample consisting of two randomly chosen individuals. The quantity $s(2)$ bears a relation also to the probability that two randomly chosen individuals belong to different species. Namely, this probability is $1 - \sum p_i^2$ (Gini-Simpson