Effect of pooling species on Good’s generalized diversity index

A. J. Baczkowski, D.N. Joanes, G.M. Shamia

Department of Statistics, University of Leeds
Leeds LS2 9JT, UK

Internal Report STAT 01/01
February 2001
Effect of pooling species on Good’s generalized diversity index

A. J. Baczkowski, D.N. Joanes, G.M. Shamia

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK
Telephone 0113-2335156, Fax: 0113-233-5102, Email: sta6ajb@leeds.ac.uk

Internal Report STAT 01/01
February 2001

Abstract

Gadagkar (1989) presented numerical evidence which showed that Hill’s (1973) diversity index did not possess a natural pooling property. We extend these results to a generalized diversity index $H(\alpha, \beta)$ due to Good (1953). Some simple transformations of $H(\alpha, \beta)$ are examined.

Keywords: Diversity indices, pooling communities.

Running title: Effect of pooling species.
1 Introduction

Consider a population of \( s \) species with proportion of the \( i \)th largest species in the population being denoted by \( p_i \). An index of diversity attempts to simultaneously measure both the population richness or size, and the differences between the species proportions.

Hill (1973) advocated the use of diversity indices \( N_k \),

\[
N_k = \left\{ \sum_{i=1}^{s} p_i^k \right\}^{1/(1-k)}.
\]

In particular,

\[
N_1 = \exp \left\{ \sum_{i=1}^{s} p_i (\log p_i) \right\},
\]

and

\[
N_2 = 1/\sum_{i=1}^{s} p_i^2.
\]

Routledge (1979) suggested use of \( N_2 \) as the “best” measure of diversity. However, Gadagkar (1989) presented evidence which showed that \( N_2 \) did not possess a desirable pooling property which was possessed by \( N_1 \) and two other diversity indices, that due to Simpson (1949),

\[
\lambda = 1 - \sum_{i=1}^{s} p_i^2,
\]

and Shannon’s (1948) index,

\[
H = -\sum_{i=1}^{s} p_i (\log p_i).
\]

Good (1953) proposed a general class of diversity index, given by,

\[
H(\alpha, \beta) = \sum_{i=1}^{s} p_i^\alpha (\log p_i)^\beta,
\]

and defined for non-negative integer values of \( \alpha \) and \( \beta \). This includes both Shannon’s and Simpson’s indices as special cases. Baczkowski et al. (1997) numerically examined the region for real values of \( \alpha \) and \( \beta \) for which the generalized index \( H(\alpha, \beta) \) satisfies the two requirements of Pielou (1975) for a good diversity index, namely,
• (P1) for given \( s \), the index should be a maximum when the \( p_i \) are equal,

• (P2) if the \( p_i \) are equal the index should be an increasing function of \( s \).

Clearly not all \( \alpha \) and \( \beta \) are suitable. For example, \( H(0,0) = s \), which has property (P2) but does not depend upon the \{\( p_i \}\}. Also \( H(1,0) = 1 \) for all \( s \) and \{\( p_i \}\}. The results of Baczkowski et al. (1997, 1998) showed that \( H(\alpha, \beta) \) satisfied the two properties (P1) and (P2) for \( (\alpha, \beta) \) within a closed region in the quadrant \( 0 < \alpha \leq 1 \) and \( \beta \geq 0 \). The moments of \( H(\alpha, \beta) \) are given in Baczkowski et al. (2000) and their use in some simple hypothesis tests are examined.

This paper considers whether the generalized diversity index \( H(\alpha, \beta) \) satisfies the pooling property of Gadagkar. This is done numerically in Section 2. The effect of simple transformations of \( H(\alpha, \beta) \) are examined in a similar way in Section 3.

2 The behaviour of \( H(\alpha, \beta) \) when two communities are pooled

Gadagkar (1989) considered the behaviour of Hill’s index \( N_2 \) for the case of two communities of equal size each having \( s = 2 \) species and for which the relative abundances are \( p_1 = (p_1, 1 - p_1) \) and \( p_2 = (p_2, 1 - p_2) \) respectively. Let \( N_{21} \) and \( N_{22} \) be the values of Hill’s diversity indices for community 1 and community 2 respectively. For the pooled community the relative abundances of the two species are given by

\[
p = \left( \frac{1}{2}(p_1 + p_2), 1 - \frac{1}{2}(p_1 + p_2) \right) \,.
\]

Let \( N_{2p} \) denote the diversity index for this pooled community. It is desirable that the pooled diversity index \( N_{2p} \) is greater than or equal to the mean of \( N_{21} \) and \( N_{22} \); we refer to this as the pooling criterion (P3). This is so because when two communities are pooled the evenness of the species abundances increases. Gadagkar (1989) showed that \( N_2 \) gives anomalous behaviour when the proportion of one of the species is close to unity.
Consider now the diversity index $H(\alpha, \beta)$.

To investigate the behaviour of $H(\alpha, \beta)$ numerically for different values of $\alpha$ and $\beta$, consider initially the case of two communities having $s$ species with relative abundances $p_1 = (p_{11}, p_{12}, \ldots, p_{1s})$ and $p_2 = (p_{21}, p_{22}, \ldots, p_{2s})$ respectively. For different values of $p_{1i}$ and $p_{2j}$ we can numerically evaluate the generalized diversity index $H(\alpha, \beta)$ for each community separately and for the case where the two communities are pooled together.

For two communities we consider the three cases $s = 2, 3, 4$ and values of the relative abundances $p_{1i}$ and $p_{2j}$ given by $p_{1i} = 0.05(0.05)0.95$ and $p_{2j} = 0.05(0.05)p_{1i}$, such that $\sum p_{1i} = 1$ and $\sum p_{2j} = 1$.

For each choice of $(\alpha, \beta)$ and for each combination of $p_{1i}$ and $p_{2j}$ we determine whether property (P3) is satisfied. Only if (P3) holds for all values of $p_{1i}$ and $p_{2j}$ is that choice of $(\alpha, \beta)$ regarded as acceptable.

The set of acceptable $(\alpha, \beta)$ for the index $H(\alpha, \beta)$ for $s = 2, 3, 4$ are shown shaded in Figures (1), (2) and (3) respectively. As $s$ increases the region satisfying property (P3) becomes larger. The shaded region is similar to the region $(\alpha, \beta)$ shown to be valid in Baczkowski et al. (1997, 1998), and satisfying the two properties (P1) and (P2). The upper bound for this valid region is shown in bold in the three figures.

**INSERT FIGURES (1), (2) AND (3).**

### 3 Pooling property for transformations of $H(\alpha, \beta)$

Simpson’s index, $H(2, 0)$, is known to vary inversely with evenness of the relative abundances. To overcome this Greenberg (1956) and Berger and Parker (1970) proposed using $1 - H(2, 0)$ as a diversity index, whereas Williams (1964) suggested $1/H(2, 0)$ as a diversity index, and $-\log H(2, 0)$ has been proposed by Pielou (1977). Though $H(2, 0)$ does not satisfy the properties (P1) and (P2), its “inverse” (in some sense) does. It is possible to similarly consider whether the “inverse” of $H(\alpha, \beta)$
satisfies properties (P1) and (P2) for given \((\alpha, \beta)\).

In Baczkowski et al. (1997, 1998) we show that the “inverse” of \(H(\alpha, \beta)\) satisfies properties (P1) and (P2) for \(\alpha > 1\) and \(\beta < 0\), but excluding the case \(\alpha = 1\) and \(\beta = 0\).

Gadagkar (1989) examined the pooling property (P3) with particular reference to Hill’s index \(N_2\). This is equivalent to \(1/H(2, 0)\) and motivates us to considering the effect of pooling on the more general index \(N(\alpha, \beta) = 1/H(\alpha, \beta)\).

This index was studied in the same way as described for \(H(\alpha, \beta)\) in Section 2. The results for \(N(\alpha, \beta)\) for \(s = 2, 3, 4\) are shown in Figures (4), (5) and (6) respectively.

INSERT FIGURES (4), (5) AND (6).

By comparing the three figures it can be seen that not all \((\alpha, \beta)\) values for \(N(\alpha, \beta)\) satisfy property (P3). Gadagkar showed that \(N_2\) was unsuitable in the case \(s = 2\). Our results confirm this and extend his conclusions to the cases \(s = 3\) and \(s = 4\). As \(s\) increases the acceptable region for \((\alpha, \beta)\) becomes smaller. The results suggest that for large \(s\), \(N(\alpha, \beta)\) would not satisfy property (P3) for any \((\alpha, \beta)\) in the given quadrant.

An alternative “inverse” transformation is to use the complement of \(H(\alpha, \beta)\), \(L(\alpha, \beta) = 1 - H(\alpha, \beta)\). Repeating the above procedure for \(L(\alpha, \beta)\) showed that for all values \(\alpha > 1\) and \(\beta < 0\), and for all the values of \(s\) examined, the pooling property (P3) was satisfied.

4 Discussion

We have applied the same procedures described in Sections 2 and 3 to the case when three communities are pooled and obtained similar results.

Our results show that in order to satisfy the natural pooling criterion (P3) one can use the index \(H(\alpha, \beta)\) in its acceptable region with \(0 < \alpha < 1\), or one can use
the complement $L(\alpha, \beta)$ for cases $\alpha > 1$ and $\beta < 0$. In conclusion, for two or more communities, we recommend the use of $L(\alpha, \beta)$ as a transformation of $H(\alpha, \beta)$ in the quadrant $\alpha > 1$ and $\beta < 0$ rather than the reciprocal $N(\alpha, \beta)$.

References


Figure captions

Figure 1. The (shaded) region for $H(\alpha, \beta)$ satisfying property (P3) when pooling two communities each with $s = 2$ species. Bold lines show upper bound for region satisfying properties (P1) and (P2).

Figure 2. The (shaded) region for $H(\alpha, \beta)$ satisfying property (P3) when pooling two communities each with $s = 3$ species. Bold lines show upper bound for region satisfying properties (P1) and (P2).

Figure 3. The (shaded) region for $H(\alpha, \beta)$ satisfying property (P3) when pooling two communities each with $s = 4$ species. Bold lines show upper bound for region satisfying properties (P1) and (P2).

Figure 4. The (shaded) region for $N(\alpha, \beta)$ satisfying property (P3) when pooling two communities each with $s = 2$ species.

Figure 5. The (shaded) region for $N(\alpha, \beta)$ satisfying property (P3) when pooling two communities each with $s = 3$ species.

Figure 6. The (shaded) region for $N(\alpha, \beta)$ satisfying property (P3) when pooling two communities each with $s = 4$ species.