

Weighted Gini-Simpson Quadratic Index of Biodiversity for Interdependent Species

Radu Cornel Guiasu¹, Silviu Guiasu²

¹Environmental and Health Studies Program, Department of Multidisciplinary Studies, Glendon College, York University, Toronto, Canada

²Department of Mathematics and Statistics, York University, Toronto, Canada
Email: rguiasu@glendon.yorku.ca, guiasus@pascal.math.yorku.ca

Received 31 December 2013; revised 31 January 2014; accepted 7 February 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The weighted Gini-Simpson quadratic index is the simplest measure of biodiversity which takes into account the relative abundance of species and some weights assigned to the species. These weights could be assigned based on factors such as the phylogenetic distance between species, or their relative conservation values, or even the species richness or vulnerability of the habitats where these species live. In the vast majority of cases where the biodiversity is measured the species are supposed to be independent, which means that the relative proportion of a pair of species is the product of the relative proportions of the component species making up the respective pair. In the first section of the paper, the main versions of the weighted Gini-Simpson index of biodiversity for the pairs and triads of independent species are presented. In the second section of the paper, the weighted Gini-Simpson quadratic index is calculated for the general case when the species are interdependent. In this instance, the weights reflect the conservation values of the species and the distribution pattern variability of the subsets of species in the respective habitat induced by the inter-dependence between species. The third section contains a numerical example.

Keywords

Weighted Gini-Simpson Index, Generalized Rao Index, Interdependent Species, Watanabe's Entropic Measure of Interdependence, Distribution Pattern Variability of the Subsets of Species

1. Introduction

In a certain habitat, let n be the number of species; p_i the relative abundance of species i ; $P = (p_1, \dots, p_n)$ the distribution of the relative abundance of species. We have:

$$p_i \geq 0, (i = 1, \dots, n); \sum_i p_i = 1.$$

The simplest measure of biodiversity is the Gini-Simpson quadratic index (Gini [1], Simpson [2]):

$$GS = GS(P) = \sum_i p_i(1 - p_i) = 1 - \sum_i p_i^2;$$

Recently, Jost [3] [4], and Jost *et al.* [5] gave some examples showing that the Gini-Simpson index does not behave well when the number of species n is very large. Guiasu and Guiasu [6] showed, however, that the Rich-Gini-Simpson index:

$$RGS(P) = n \sum_i p_i(1 - p_i),$$

which depends explicitly on the number of species (species richness), preserves all the properties of the classic Gini-Simpson index GS but, unlike GS , behaves very well when n is large.

Another measure of diversity is the weighted Gini-Simpson index introduced in Guiasu and Guiasu [7]:

$$GS_w(P) = \sum_i w_i p_i(1 - p_i),$$

where $W = (w_1, \dots, w_n)$ are some nonnegative weights assigned to the species, such as the conservation values of the respective species, for instance. Obviously, GS_w becomes RGS if $w_i = n$ for each $i = 1, \dots, n$.

Let us assume that in a certain region there are n species and m sites. In what follows, the subscripts i and j refer to species ($i, j = 1, \dots, n$) and the subscripts k and r refer to sites, ($k, r = 1, \dots, m$).

Let $P_k = (p_{1,k}, \dots, p_{n,k})$ be the vector whose components are the relative abundances of the individual species within site k , such that:

$$p_{i,k} \geq 0, (i = 1, \dots, n); \sum_i p_{i,k} = 1,$$

for each $k = 1, \dots, m$. A measure of diversity μ may be used in the standard additive partitioning of diversity induced by individual species if it is a concave function, which means that it satisfies the inequality:

$$\mu\left(\sum_k \lambda_k P_k\right) \geq \sum_k \lambda_k \mu(P_k) \tag{1}$$

for arbitrary parameters:

$$\lambda_k \geq 0, (k = 1, \dots, m), \sum_k \lambda_k = 1. \tag{2}$$

For each site k , we calculate the diversity $\mu(P_k)$ among its species, called the local diversity of the respective site. In such a case, as pointed out by Lande [8], the right-hand side of (1) is the α -diversity, denoted by α , measuring the average local diversity, or the average within-diversity. The left-hand side of (1) is the γ -diversity, denoted by γ , measuring the diversity among the species belonging to the averaged union of the habitats, called the regional or total diversity. As shown in [6] [7] [9], the measures of biodiversity GS , RGS , and GS_w , the last one when its weights do not depend on the relative proportions of species p_1, \dots, p_n , are concave functions and satisfy the inequality (1), for arbitrary parameters (2). Unlike the α -diversity and the γ -diversity, there is no consensus about how to interpret and calculate the β -diversity. According to Whittaker [10] [11] who introduced the terminology, β -diversity is the ratio between γ -diversity and α -diversity, namely, $\beta = \gamma/\alpha$. This is the multiplicative partitioning of diversity. According to MacArthur [12], MacArthur and Wilson [13], and Lande [8], β -diversity is the difference between γ -diversity and α -diversity, namely, $\beta = \gamma - \alpha$. This is the additive partitioning of diversity. In both cases, the β -diversity measures the variation of the diversity among habitats, or the average between-diversity in the additive case.

As the classic Gini-Simpson index GS is a concave function of the relative abundance of species $P = (p_1, \dots, p_n)$, it may be used in the additive partitioning of diversity. Hill [14] showed that:

$$1/(1 - GS)$$

may be used in the multiplicative partitioning of diversity. Similarly, the weighted Gini-Simpson quadratic index GS_w is a concave function of the relative abundance of species $P = (p_1, \dots, p_n)$ if the weights

$W = (w_1, \dots, w_n)$ do not depend on the relative abundance $P = (p_1, \dots, p_n)$ and, therefore, may be used in the additive partitioning of diversity, where $\beta = \gamma - \alpha$. The corresponding multiplicative weighted Gini-Simpson quadratic index for individual species:

$$1/\left[\sum_i w_i p_i - GS_W(P)\right] = \left(\sum_i w_i p_i^2\right)^{-1}$$

may be used in the multiplicative partitioning of diversity where the β -diversity is the ratio between the γ -diversity and the α -diversity, i.e. $\beta = \gamma/\alpha$.

Let $D = [d_{ij}]$ be a square matrix whose entries are the distances between species, such as the phylogenetic distances, for instance. Thus, d_{ij} is the distance between the species i and j , and we have:

$$d_{ij} \geq 0, \quad d_{ii} = 0, \quad (i, j = 1, \dots, n).$$

Rao's index [15], also called quadratic entropy or dissimilarity measure, is:

$$R_D(P) = \sum_{ij} d_{ij} p_i p_j.$$

Viewed as a function of the relative abundance of individual species, R_D is a quadratic function. Unfortunately, if the distance matrix D is arbitrary, Rao's index R_D , taken as a diversity measure μ , does not satisfy the inequality (1) for arbitrary parameters (2) and, therefore, cannot be used in a standard additive partitioning of the diversity. There are some excellent papers dealing with diversity (Pavoine *et al.* [16], Ricotta [17], Ricotta and Szeidel [18], Hardy and Senterre [19], Villéger and Moullot [20], Hardy and Jost [21], Ricotta and Szeidel [22], Sherwin [23], De Bello *et al.* [24], and Tuomisto [25] [26]). Some of them discuss how to use Rao's index R_D in the additive partitioning of diversity. Some research focused on finding special kinds of distance matrices D for which the corresponding Rao's index R_D is a concave function, such as the matrix \sqrt{D} assumed to be Euclidean, for instance. Some other research focused on how to use Rao's index for getting a nonstandard additive partitioning of diversity, which means determining whether some special parameters (2) could be used in order to define analog α -, β -, and γ -diversities corresponding to these particular parameters. However, the weighted Gini-Simpson quadratic index of biodiversity, applied to the pairs of species, provides a concave replacement of Rao's index and, therefore, is appropriate for use in the additive partitioning of diversity for arbitrary distance matrix D .

Let w_{ij} be the weight assigned to the pair of species (i, j) ; π_{ij} is the joint probability of the ordered pair of species (i, j) . The joint probability of the pair of species $\{i, j\}$ is: $\pi_{ij} + \pi_{ji}$. We have:

$$\pi_{ij} \geq 0, \quad \sum_{ij} \pi_{ij} = 1; \quad w_{ij} \geq 0;$$

Denote by: $W = [w_{ij}]$ and $\pi = [\pi_{ij}]$. Let $c_i \geq 0$ be the value of species i and $c = (c_1, \dots, c_n)$. Again, let $D = [d_{ij}]$ be the matrix whose entry d_{ij} is the distance between the species i and j . The weighted Gini-Simpson quadratic index of biodiversity for pairs of species is:

$$GS_W(\pi) = \sum_{ij} w_{ij} \pi_{ij} (1 - \pi_{ij}).$$

As shown in Guiasu and Guiasu ([27] [28]), $GS_W(\pi)$ is a concave function of the joint distribution π , when the weights do not depend on the probability distribution π , and, therefore, may be used in the additive partitioning of diversity, where $\beta = \gamma - \alpha$. The corresponding multiplicative weighted Gini-Simpson quadratic index for pairs of species:

$$1/\left[\sum_{i,j} w_{ij} \pi_{ij} - GS_W(\pi)\right] = \left(\sum_{i,j} w_{ij} \pi_{ij}^2\right)^{-1};$$

may be used in the multiplicative partitioning of diversity where the β -diversity is the ratio between the γ -diversity and the α -diversity, i.e. $\beta = \gamma/\alpha$.

Let us assume now that the weights are:

$$w_{ij} = \frac{n(n-1)}{2} \frac{1}{2} (c_i + c_j) d_{ij},$$

where: $\frac{n(n-1)}{2}$ is the number of distinct pairs of species (the richness of the distinct pairs of species), and $\frac{1}{2}(c_i + c_j)$ is the average value of the pair of species (i, j) . If the species are considered to be independent from the point of view of their relative abundance, then:

$$\pi_{ij} = p_i p_j,$$

and the corresponding weighted Gini-Simpson quadratic index of biodiversity for the distinct pairs of species becomes:

$$GS_{n,c,D} = \frac{n(n-1)}{2} \sum_{i < j} (c_i + c_j) d_{ij} p_i p_j (1 - p_i p_j).$$

As mentioned in [27] [28], this measure is a concave replacement of the Rao's index. It depends on the distance between species, the richness of the pairs of distinct species, the value of the species, and the relative proportion of the species.

As shown in [9], the weighted Gini-Simpson quadratic index may be applied to the study of the biodiversity of the triads of species as well. Let $D = [d_{ij}]$ be the matrix whose generic entry d_{ij} is the distance between species i and j . The area of the triangle formed by the distinct species i , j , and k , is given by the Heron's formula:

$$A_{ijk} = [s(s - d_{ij})(s - d_{jk})(s - d_{ki})]^{1/2},$$

where s is the semi-perimeter of this triangle:

$$s = \frac{1}{2} [d_{ij} + d_{jk} + d_{ki}].$$

If the species are independent from the point of view of their relative abundance, taking $A = [A_{ijk}]$ and the richness of the triads of species as weights, the corresponding weighted Gini-Simpson quadratic index for triads of independent species is:

$$GS_{n,D} = \frac{n(n-1)(n-2)}{6} \sum_{i \neq j \neq k} A_{ijk} p_i p_j p_k (1 - p_i p_j p_k),$$

where $\frac{n(n-1)(n-2)}{6}$ is the number of distinct triads of species. If the values of the species are also taken into account, the weighted Gini-Simpson quadratic index for triads of species is:

$$GS_{n,c,D} = \sum_{i \neq j \neq k} w_{ijk} p_i p_j p_k (1 - p_i p_j p_k),$$

where the weights are:

$$w_{ijk} = \frac{n(n-1)(n-2)}{6} \frac{c_i + c_j + c_k}{3} A_{ijk}, \quad (i \neq j \neq k).$$

2. Interdependent Species

In the current literature on biodiversity the species from a habitat are considered to be independent from the point of view of their relative abundance and the separate abundance of each species is taken into account. Thus, if p_i is the relative proportion of species i , then the relative proportion of the ordered pair of species (i, j) is $p_i p_j$. In real life, however, the species can be interdependent and the biodiversity in the ecosystem is influenced by this interdependence. For instance, pairs of species can interact in various types of inter-specific ecological relationships, including mutualism, where both species benefit from the interaction, commensalism, where one species benefits and the other one is not significantly affected, and parasitism, where the parasite

benefits and the host species is harmed. While interactions between pairs of species are easier to study, and have generally been the focus of most inter-specific studies, the interdependence among several species within a particular ecosystem can often be much more complex and crucial for determining the biodiversity found at a certain location. As an example, in sage-scrub habitat fragments in southern California, the presence of coyotes (the top predators in these habitats) keeps the numbers of mesopredators, such as domestic cats and raccoons, down. Since domestic cats are effective predators of local scrub-breeding birds, these bird species actually benefit from the presence of coyotes. Therefore, when the coyote populations decline, local scrub-breeding birds are more likely to become extinct, and bird species diversity decreases (Crooks and Soulé [29]). This is an example of multiple interactions involving several species, which shows how these inter-specific relationships can affect the biodiversity of the ecosystem.

Let $\{x_1, \dots, x_n\}$ be the distinct species that may be found in a given large habitat, like some species of plants on a large island, for instance. Making observations at several different locations of this island, denote by $N_{i_1 \dots i_k}$ the number of locations where only the species $\{x_{i_1}, \dots, x_{i_k}\}$ have been found, coexisting together. If N is the total number of locations where observations have been made, the relative proportion of the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$ is:

$$\theta_{i_1 \dots i_k} = N_{i_1 \dots i_k} / N.$$

We have:

$$\theta_0 + \sum_i \theta_i + \sum_{i \neq j} \theta_{ij} + \sum_{i \neq j \neq k} \theta_{ijk} + \dots + \theta_{1 \dots n} = 1,$$

where θ_0 is the relative proportion of finding none of the species $\{x_1, \dots, x_n\}$ at the locations on the island where observations have been made.

Let $w_{i_1 \dots i_k}$ be some nonnegative weights assigned to the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$. The weighted Gini-Simpson quadratic index for interdependent species is:

$$GS_W = w_0 \theta_0 (1 - \theta_0) + \sum_i w_i \theta_i (1 - \theta_i) + \sum_{i \neq j} w_{ij} \theta_{ij} (1 - \theta_{ij}) + \sum_{i \neq j \neq k} w_{ijk} \theta_{ijk} (1 - \theta_{ijk}) + \dots + w_{1 \dots n} \theta_{1 \dots n} (1 - \theta_{1 \dots n}). \tag{3}$$

The main problem is to investigate what kind of weights could adequately reflect the species biodiversity on the island, along with the abundance of these species. A reasonable answer is offered by the use of Shannon's [30] entropy, as a measure of uncertainty, and Watanabe's [31] entropic measure of interdependence.

As mentioned before, $\theta_{i_1 \dots i_k}$ is the relative abundance of the separate subset of species $\{x_{i_1}, \dots, x_{i_k}\}$ alone, or the probability of finding only the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$, and no other species, at a location on the island where observations have been made. Let $\pi_{i_1 \dots i_k}$ be the probability distribution that describes the presence, denoted by the symbol 1, or absence, denoted by the symbol 0, of the species belonging to the subset $\{x_{i_1}, \dots, x_{i_k}\}$. The Shannon entropy $H_{i_1 \dots i_k}$ corresponding to the probability distribution $\pi_{i_1 \dots i_k}$ measures how much variability is induced on the respective island by the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$ at the locations where observations have been made. The subset of species $\{x_{i_1}, \dots, x_{i_k}\}$ contributes more to the diversity of the respective habitat if its variability, as measured by the corresponding entropy $H_{i_1 \dots i_k}$, is larger. We have:

$$0 \leq H_{i_1 \dots i_k} \leq \ln 2^k = (0.693147181)k.$$

The minimum value corresponds to the case when the same configuration present/absent of the species of the subset $\{x_{i_1}, \dots, x_{i_k}\}$ is found at every location where observations were made. This happens, for instance, when all the species of the subset $\{x_{i_1}, \dots, x_{i_k}\}$ were found at every location. In such a case, there is no variability as far as the distribution of the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$ on the respective island is concerned. The maximum value corresponds to the case when all the configurations indicating presence/absence of the species from the subset $\{x_{i_1}, \dots, x_{i_k}\}$ are encountered at the different locations where observations were made, with equal likelihood, thereby inducing a maximum variability on the island.

For the subset $\{x_i\}$, the components of the probability distribution π_i , which describes its distribution pattern variability, are $\pi_i(1)$ and $\pi_i(0)$ giving the probability that the species x_i is present or absent at the locations where observations were made. The corresponding entropy is:

$$H_i = -\pi_i(1)\ln \pi_i(1) - \pi_i(0)\ln \pi_i(0).$$

We have: $0 \leq H_i \leq \ln 2 = 0.693147181$, where the minimum is obtained when $\pi_i(1) = 1$, which means that the species x_i is present, with certainty, at every location, or when $\pi_i(0) = 1$, which means that the species x_i is absent, with certainty, at every location. In both such cases, the species x_i induces no distribution variability on the island. The maximum value of H_i corresponds to the case when $\pi_i(1) = \pi_i(0) = 1/2$, which means that the species x_i is present at half of the locations and absent at the other half, inducing a maximum variability on the island. This has nothing to do with the abundance of the species x_i , which obviously contributes to the diversity on the island; it has to do with the way in which the species x_i is distributed within the respective island. Abundance and distribution variability are both components of diversity.

For the subset $\{x_i, x_j\}$, the components of the probability distribution π_{ij} , which describes its distribution pattern variability, are $\pi_{ij}(11)$, $\pi_{ij}(10)$, $\pi_{ij}(01)$ and $\pi_{ij}(00)$, giving the probability that the species x_i and x_j are present or absent at the locations where observations were made. Thus, $\pi_{ij}(11)$, for instance, is the probability that both species are present, or proportion of locations where both x_i and x_j are present. The corresponding entropy is:

$$H_{ij} = -\pi_{ij}(11)\ln \pi_{ij}(11) - \pi_{ij}(10)\ln \pi_{ij}(10) - \pi_{ij}(01)\ln \pi_{ij}(01) - \pi_{ij}(00)\ln \pi_{ij}(00).$$

We have: $0 \leq H_{ij} \leq \ln 2^2 = 1.386294361$, where the minimum is obtained when $\pi_{ij}(11) = 1$, or when $\pi_{ij}(10) = 1$, or when $\pi_{ij}(01) = 1$, or when $\pi_{ij}(00) = 1$, which means that only one of the patterns 11, 10, 01, or 00 is observed, with certainty, at every location. This indicates that the subset of species $\{x_i, x_j\}$ induces no distribution pattern variability on the respective island. The maximum value of H_{ij} corresponds to the case when $\pi_{ij}(11) = \pi_{ij}(10) = \pi_{ij}(01) = \pi_{ij}(00) = 1/4$. This indicates that the four different patterns of presence/absence of the two species are uniformly distributed at the locations on the island where observations were made, inducing a maximum distribution pattern variability on the island.

If the species of the subset $\{x_i, \dots, x_k\}$ were independent, as far as the individual presence/absence is concerned, then the entropy $H_{i_1 \dots i_k}$ would be a suitable weight reflecting the distribution pattern variability induced by this subset of species. But, more often than not, the species are not independent as far as the individual presence/absence is concerned. The interdependence between species will diminish variability. To give a simple example, if two species, x_i and x_j , are independent, as far as the individual presence/absence is concerned, then there are four possibilities to consider at each location: (a) x_1 present and x_2 present; (b) x_1 present and x_2 absent; (c) x_1 absent and x_2 present; (d) x_1 absent and x_2 absent. If, however, the two species are interdependent and, for instance, can be present only together, then there are only two possibilities to deal with, namely: (a) x_1 present and x_2 present; (b) x_1 absent and x_2 absent. Obviously, we have less variability due to the interdependence between the two species.

The amount of interdependence between the species of the subset $\{x_i, \dots, x_k\}$, as introduced by Watanabe [31], is given by the sum of the entropies of the species of the subset minus the entropy of the entire subset of species, namely:

$$I_{i_1 \dots i_k} = H_{i_1} + \dots + H_{i_k} - H_{i_1 \dots i_k}.$$

We have: $I_{i_1 \dots i_k} \geq 0$, and $I_{i_1 \dots i_k} = 0$ if and only if the species of the subset $\{x_i, \dots, x_k\}$ are independent. Details may be found in Guiasu [32].

The degree of distribution pattern variability induced by the subset of species $\{x_i, \dots, x_k\}$ on the respective island is obtained by subtracting the interdependence between the species of the subset $\{x_i, \dots, x_k\}$ from the entropy of this subset of species, namely:

$$V_{i_1 \dots i_k} = H_{i_1 \dots i_k} - I_{i_1 \dots i_k}.$$

Summarizing, diversity depends on the abundance of species, the number of species (richness), the values of species, and the distribution pattern variability of the species. The classic Gini-Simpson index measures diversity taking into account only the abundance of species. The weighted Gini-Simpson quadratic index takes into account all four factors just mentioned.

Let $c = (c_1, \dots, c_n)$ be the positive numbers reflecting the value of the species $\{x_1, \dots, x_n\}$, respectively, such as the conservation values of species, for instance. Normally, $c_i \geq 1, (i = 1, \dots, n)$. The weight assigned to the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$ is:

$$w_{i_1 \dots i_k} = (c_{i_1} + \dots + c_{i_k}) V_{i_1 \dots i_k}. \tag{4}$$

Obviously, $w_0 = 1$ and $w_i = c_i H_i$, because:

$$I_i = H_i - H_i = 0; \quad V_i = H_i - I_i = H_i.$$

If we do not take the value of the species into account or have no such information available, then we assume that $c_1 = \dots = c_n = 1$, in which case $c_{i_1} + \dots + c_{i_k} = k$, which is the richness of the subset of species $\{x_{i_1}, \dots, x_{i_k}\}$, and the weights (4) become:

$$w_{i_1 \dots i_k} = k V_{i_1 \dots i_k}.$$

The main conclusion is that for interdependent species, the diversity is measured by the formula (3) where the weights are given by (4).

3. Numerical Example

On an island, observations are made at several locations, looking for four species of plants, denoted by $\{x_1, x_2, x_3, x_4\}$. Seven different patterns have been found, denoted by $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$, as shown in the table below, with the relative frequencies:

$$q_1 = 0.325, \quad q_2 = 0.050, \quad q_3 = 0.325, \quad q_4 = 0.050, \\ q_5 = 0.025, \quad q_6 = 0.025, \quad q_7 = 0.200,$$

respectively.

| | y_1 | y_2 | y_3 | y_4 | y_5 | y_6 | y_7 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| x_1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| x_2 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| x_3 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| x_4 | 0 | 1 | 0 | 1 | 0 | 1 | 1 |

The entries in this table are: 1 meaning “presence” and 0 meaning “absence” of the species from the rows in the patterns from the columns. Thus, pattern y_4 , for instance, shows that the species x_1, x_3 , and x_4 have been found together at a certain location, but the species x_2 was absent. Of course, the same pattern may be found at different locations on the island, but no pattern different from those listed in the table was found at any location on the island where observations were made. The number q_j is the relative frequency of the pattern y_j , or the percentage of the locations where the pattern y_j was found. Let us assume that the conservation values of the species are: $c_1 = 5, c_2 = 2, c_3 = 3, c_4 = 1$. The relative abundance of the subsets of species, taken separately, are:

$$\theta_1 = \theta_2 = \theta_3 = 0, \quad \theta_4 = q_7 = 0.200, \\ \theta_{12} = q_1 = 0.325, \quad \theta_{13} = q_3 = 0.325, \quad \theta_{14} = 0, \\ \theta_{23} = q_5 = 0.025, \quad \theta_{24} = 0, \quad \theta_{34} = 0, \\ \theta_{123} = 0, \quad \theta_{124} = q_2 = 0.050, \quad \theta_{134} = q_4 = 0.050, \\ \theta_{234} = q_6 = 0.025, \quad \theta_{1234} = 0.$$

For the subset of species $\{x_1\}$:

$$\pi_1(1) = q_1 + q_2 + q_3 + q_4 = 0.750, \quad \pi_1(0) = 0.250; \quad H_1 = 0.562335, \quad I_1 = 0, \quad V_1 = 0.562335;$$

For the subset of species $\{x_2\}$:

$$\pi_2(1) = q_1 + q_2 + q_5 + q_6 = 0.425, \quad \pi_2(0) = 0.575; \quad H_2 = 0.681855, \quad I_2 = 0, \quad V_2 = 0.681855;$$

For the subset of species $\{x_3\}$:

$$\pi_3(1) = q_3 + q_4 + q_5 + q_6 = 0.425, \quad \pi_3(0) = 0.575; \quad H_3 = 0.681855, \quad I_3 = 0, \quad V_3 = 0.681855;$$

For the subset of species $\{x_4\}$:

$$\pi_4(1) = q_2 + q_4 + q_6 + q_7 = 0.325, \quad \pi_4(0) = 0.675; \quad H_4 = 0.630581, \quad I_4 = 0, \quad V_4 = 0.630581;$$

For the subset of species $\{x_1, x_2\}$:

$$\pi_{12}(11) = q_1 + q_2 = 0.375, \quad \pi_{12}(10) = q_3 + q_4 = 0.375,$$

$$\pi_{12}(01) = q_5 + q_6 = 0.050, \quad \pi_{12}(00) = q_7 = 0.200;$$

$$H_{12} = 1.20730, \quad I_{12} = 0.0368900, \quad V_{12} = 1.17041;$$

For the subset of species $\{x_1, x_3\}$:

$$\pi_{13}(11) = q_3 + q_4 = 0.375, \quad \pi_{13}(10) = q_1 + q_2 = 0.375,$$

$$\pi_{13}(01) = q_5 + q_6 = 0.050, \quad \pi_{13}(00) = q_7 = 0.200;$$

$$H_{13} = 1.20730, \quad I_{13} = 0.0368900, \quad V_{13} = 1.17041;$$

For the subset of species $\{x_1, x_4\}$:

$$\pi_{14}(11) = q_2 + q_4 = 0.100, \quad \pi_{14}(10) = q_1 + q_3 = 0.650,$$

$$\pi_{14}(01) = q_6 + q_7 = 0.225, \quad \pi_{14}(00) = q_5 = 0.025;$$

$$H_{14} = 0.938112, \quad I_{14} = 0.254804, \quad V_{14} = 0.683308;$$

For the subset of species $\{x_2, x_3\}$:

$$\pi_{23}(11) = q_5 + q_6 = 0.050, \quad \pi_{23}(10) = q_1 + q_2 = 0.375,$$

$$\pi_{23}(01) = q_3 + q_4 = 0.375, \quad \pi_{23}(00) = q_7 = 0.200;$$

$$H_{23} = 1.20730, \quad I_{23} = 0.156410, \quad V_{23} = 1.05089;$$

For the subset of species $\{x_2, x_4\}$:

$$\pi_{24}(11) = q_2 + q_6 = 0.075, \quad \pi_{24}(10) = q_1 + q_5 = 0.350,$$

$$\pi_{24}(01) = q_4 + q_7 = 0.250, \quad \pi_{24}(00) = q_3 = 0.325;$$

$$H_{24} = 1.27356, \quad I_{24} = 0.0388760, \quad V_{24} = 1.23468;$$

For the subset of species $\{x_3, x_4\}$:

$$\pi_{34}(11) = q_4 + q_6 = 0.075, \quad \pi_{34}(10) = q_3 + q_5 = 0.350,$$

$$\pi_{34}(01) = q_2 + q_7 = 0.250, \quad \pi_{34}(00) = q_1 = 0.325;$$

$$H_{34} = 1.27356, \quad I_{34} = 0.0388760, \quad V_{34} = 1.23468;$$

For the subset of species $\{x_1, x_2, x_3\}$:

$$\pi_{123}(111) = 0, \quad \pi_{123}(110) = q_1 + q_2 = 0.375,$$

$$\pi_{123}(101) = q_3 + q_4 = 0.375, \quad \pi_{123}(011) = q_5 + q_6 = 0.050,$$

$$\pi_{123}(100) = 0, \quad \pi_{123}(010) = 0,$$

$$\pi_{123}(001) = 0, \quad \pi_{123}(000) = q_7 = 0.200.$$

$$H_{123} = 1.20730, \quad I_{123} = 0.718745, \quad V_{123} = 0.488555;$$

For the subset of species $\{x_1, x_2, x_4\}$:

$$\pi_{124}(111) = q_2 = 0.050, \quad \pi_{124}(110) = q_1 = 0.325,$$

$$\pi_{124}(101) = q_4 = 0.050, \quad \pi_{124}(011) = q_6 = 0.025,$$

$$\pi_{124}(100) = q_3 = 0.325, \quad \pi_{124}(010) = q_5 = 0.025,$$

$$\pi_{124}(001) = q_7 = 0.200, \quad \pi_{124}(000) = 0.$$

$$H_{124} = 1.53646, \quad I_{124} = 0.338311, \quad V_{124} = 1.19815;$$

For the subset of species $\{x_1, x_3, x_4\}$:

$$\pi_{134}(111) = q_4 = 0.050, \quad \pi_{134}(110) = q_3 = 0.325,$$

$$\pi_{134}(101) = q_2 = 0.050, \quad \pi_{134}(011) = q_6 = 0.025,$$

$$\pi_{134}(100) = q_1 = 0.325, \quad \pi_{134}(010) = q_5 = 0.025,$$

$$\pi_{134}(001) = q_7 = 0.200, \quad \pi_{134}(000) = 0.$$

$$H_{134} = 1.53646, \quad I_{134} = 0.338311, \quad V_{134} = 1.19815;$$

For the subset of species $\{x_2, x_3, x_4\}$:

$$\pi_{234}(111) = q_6 = 0.025, \quad \pi_{234}(110) = q_5 = 0.025,$$

$$\pi_{234}(101) = q_2 = 0.050, \quad \pi_{234}(011) = q_4 = 0.050,$$

$$\pi_{234}(100) = q_1 = 0.325, \quad \pi_{234}(010) = q_3 = 0.325,$$

$$\pi_{234}(001) = q_7 = 0.200, \quad \pi_{234}(000) = 0.$$

$$H_{234} = 1.53646, \quad I_{234} = 0.457831, \quad V_{234} = 1.07863;$$

For the subset of species $\{x_1, x_2, x_3, x_4\}$:

$$\pi_{1234}(1100) = q_1 = 0.325, \quad \pi_{1234}(1101) = q_2 = 0.050,$$

$$\pi_{1234}(1010) = q_3 = 0.325, \quad \pi_{1234}(1011) = q_4 = 0.050,$$

$$\pi_{1234}(0110) = q_5 = 0.025, \quad \pi_{1234}(0111) = q_6 = 0.025,$$

$$\pi_{1234}(0001) = q_7 = 0.200,$$

(all the other patterns have a probability equal to 0)

$$H_{1234} = 1.53646, \quad I_{1234} = 1.02017, \quad V_{1234} = 0.516294.$$

Applying Formula (4), the weights are:

$$\begin{aligned} w_1 &= 2.81168, & w_2 &= 1.36371, & w_3 &= 2.04557, \\ w_4 &= 0.63058, & w_{12} &= 8.19287, & w_{13} &= 9.36328, \\ w_{14} &= 4.09985, & w_{23} &= 5.25445, & w_{24} &= 3.70404, \\ w_{34} &= 4.93872, & w_{123} &= 4.88555, & w_{124} &= 9.58520, \\ w_{134} &= 10.78335, & w_{234} &= 6.47178, & w_{1234} &= 5.67923. \end{aligned}$$

Applying Formula (3) with these weights, we obtain the amount of plant biodiversity on the island (at least as far as the four species, in this simplified example, are concerned), as measured by the weighted Gini-Simpson quadratic index: $GS_w = 5.20575$.

If the weights are ignored and we apply Formula (3) with all the weights equal to 1, we obtain the diversity on the island induced only by the abundance of the subsets of species, as measured by the classic Gini-Simpson index: $GS = 0.74252$.

4. Conclusions

The classic Gini-Simpson index is the oldest and simplest measure of biodiversity. It takes into account only the relative abundance of the species from a habitat. As pointed out recently, this index of biodiversity does not behave well when the number of species is very large. The weighted Gini-Simpson index preserves the qualities of the classic index but behaves well when the number of species is very large. Moreover, the weights may be chosen in such a way that not only the relative abundance of species but also the number of species (richness), the distance between species, and the value of species are taken into account when the diversity in the habitat is measured. The weighted Gini-Simpson index continues to be a concave function of the relative abundance of species, allowing it to be used in the additive and multiplicative partitioning of diversity.

In the current applications of the weighted Gini-Simpson index the species are considered to be independent from the point of view of their individual relative abundance, which means that the joint relative proportion of a set of species is the product of the relative proportions of the species of this set. The first section of this paper summarizes the main variants of the weighted Gini-Simpson index recently introduced for dealing with such independent species. There are cases, however, when the species are not independent from the point of view of the relative abundance. The second section of this paper deals with interdependent species and the weighted Gini-Simpson index is adapted for dealing with such a more complex case. The new concept of distribution pattern variability of species is introduced, essentially based on Shannon's entropy and Watanabe's entropic measure of global interdependence. The new weighted Gini-Simpson index of biodiversity for interdependent species measures the amount of biodiversity induced by the abundance of species, the species richness, the value of species, and the distribution pattern variability of the subsets of interdependent species in the respective habitat.

The third section contains a numerical example which shows how the general formulas from the second section have to be applied.

References

- [1] Gini, C. (1912) Variabilità e mutabilità. In: Pizetti, E. and Salvemini, T., Eds., *Rome: Libreria Eredi Virgilio Veschi*, Memorie di metodologica statistica.
- [2] Simpson, E.H. (1949) Measurement of Diversity. *Nature*, **163**, 688. <http://dx.doi.org/10.1038/163688a0>
- [3] Jost, L. (2007) Partitioning Diversity into Independent Alpha and Beta Components. *Ecology*, **88**, 2427-2439. <http://dx.doi.org/10.1890/06-1736.1>

- [4] Jost, L. (2009) Mismeasuring Biological Diversity: Response to Hoffmann and Hoffmann. *Ecological Economics*, **68**, 925-928. <http://dx.doi.org/10.1016/j.ecolecon.2008.10.015>
- [5] Jost, L., DeVries, P., Walla, T., Greeney, H., Chao, A. and Ricotta, C. (2010) Partitioning Diversity for Conservation Analyses. *Diversity and Distributions*, **16**, 65-76. <http://dx.doi.org/10.1111/j.1472-4642.2009.00626.x>
- [6] Guiasu, R.C. and Guiasu, S. (2010) The Rich-Gini-Simpson Quadratic Index of Biodiversity. *Natural Science*, **2**, 1130-1137. <http://dx.doi.org/10.4236/ns.2010.210140>
- [7] Guiasu, R.C. and Guiasu, S. (2003) Conditional and Weighted Measures of Ecological Diversity. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **11**, 283-300.
- [8] Lande, R. (1996) Statistics and Partitioning of Species Diversity and Similarity among Multiple Communities. *Oikos*, **76**, 5-13. <http://dx.doi.org/10.2307/3545743>
- [9] Guiasu, R.C. and Guiasu, S. (2010) New Measures for Comparing the Species Diversity Found in Two or More Habitats. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **18**, 691-720.
- [10] Whittaker, R.H. (1972) Evolution and Measurement of Species Diversity. *Taxon*, **21**, 213-251. <http://dx.doi.org/10.2307/1218190>
- [11] Whittaker, R.H. (1977) Evolution of Species Diversity in Land Communities. In: Hecht, M.K. and Steere, B.W.N.C., Eds., *Evolutionary Biology*, Plenum Press, New York, 1-67. http://dx.doi.org/10.1007/978-1-4615-6953-4_1
- [12] MacArthur, R.H. (1965) Patterns of Species Diversity. *Biological Review*, **40**, 510-533. <http://dx.doi.org/10.1111/j.1469-185X.1965.tb00815.x>
- [13] MacArthur, R.H. and Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton.
- [14] Hill, M. (1973) Diversity and Evenness. A Unifying Notation and Its Consequences. *Ecology*, **88**, 2427-2439.
- [15] Rao, C.R. (1982) Diversity and Dissimilarity Coefficients: A Unified Approach. *Theoretical Population Biology*, **21**, 24-43. [http://dx.doi.org/10.1016/0040-5809\(82\)90004-1](http://dx.doi.org/10.1016/0040-5809(82)90004-1)
- [16] Pavoine, S., Ollier, S. and Pontier, D. (2005) Measuring Diversity from Dissimilarities with Rao's Quadratic Entropy: Are Any Dissimilarities Suitable? *Theoretical Population Biology*, **67**, 231-239. <http://dx.doi.org/10.1016/j.tpb.2005.01.004>
- [17] Ricotta, C. (2005) Additive Partitioning of Rao's Quadratic Diversity: A Hierarchical Approach. *Ecological Modelling*, **183**, 365-371. <http://dx.doi.org/10.1016/j.ecolmodel.2004.08.020>
- [18] Ricotta, C. and Szeidel, L. (2006) Towards a Unifying Approach to Diversity Measures: Bridging the Gap between the Shannon Entropy and Rao's Quadratic Index. *Theoretical Population Biology*, **70**, 237-243. <http://dx.doi.org/10.1016/j.tpb.2006.06.003>
- [19] Hardy, O.J. and Senterre, B. (2007) Characterizing the Phylogenetic Structure of Communities by an Additive Partitioning of Phylogenetic Diversity. *Journal of Ecology*, **95**, 493-506. <http://dx.doi.org/10.1111/j.1365-2745.2007.01222.x>
- [20] Villéger, S. and Mouillot, D. (2008) Additive Partitioning of Diversity Including Species Differences: A Comment on Hardy and Senterre (2007). *Journal of Ecology*, **96**, 845-848. <http://dx.doi.org/10.1111/j.1365-2745.2007.01351.x>
- [21] Hardy, O.J. and Jost, L. (2008) Interpreting Measures of Community Phylogenetic Structuring. *Journal of Ecology*, **96**, 849-852. <http://dx.doi.org/10.1111/j.1365-2745.2008.01423.x>
- [22] Ricotta, C. and Szeidel, L. (2009) Diversity Partitioning of Rao's Quadratic Entropy. *Theoretical Population Biology*, **76**, 299-302. <http://dx.doi.org/10.1016/j.tpb.2009.10.001>
- [23] Sherwin, W.B. (2010) Entropy and Information Approaches to Genetic Diversity and Its Expression: Genomic Geography. *Entropy*, **12**, 1765-1798. <http://dx.doi.org/10.3390/e12071765>
- [24] De Bello, F., Lavergne, S., Meynard, C.N., Lepš, J. and Thuiller, W. (2010) The Partitioning of Diversity: Showing Theseus a Way out of the Labyrinth. *Journal of Vegetation Science*, **21**, 992-1000. <http://dx.doi.org/10.1111/j.1654-1103.2010.01195.x>
- [25] Tuomisto, H. (2010) A Diversity of Beta Diversities: Straightening up a Concept Gone Awry. Part 1. Defining Beta Diversity as a Function of Alpha and Gamma Diversity. *Ecography*, **33**, 2-22. <http://dx.doi.org/10.1111/j.1600-0587.2009.05880.x>
- [26] Tuomisto, H. (2010) A Diversity of Beta Diversities: Straightening up a Concept Gone Awry. Part 2. Quantifying Beta Diversity and Related Phenomena. *Ecography*, **33**, 23-45. <http://dx.doi.org/10.1111/j.1600-0587.2009.06148.x>
- [27] Guiasu, R.C. and Guiasu, S. (2011) The Weighted Quadratic Index of Biodiversity for Pairs of Species: A Generalization of Rao's Index. *Natural Science*, **3**, 795-801. <http://dx.doi.org/10.4236/ns.2011.39104>
- [28] Guiasu, R.C. and Guiasu, S. (2012) The Weighted Gini-Simpson Index: Revitalizing an Old Index of Biodiversity. *International Journal of Ecology*, **2012**, 10 p.

- [29] Crooks, K.R. and Soulé, M.E. (1999) Mesopredator Release and Avifaunal Extinctions in a Fragmented System. *Nature*, **400**, 563-566. <http://dx.doi.org/10.1038/23028>
- [30] Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379-423, 623-656. <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [31] Watanabe, S. (1969) *Knowing and Guessing*. Wiley, New York.
- [32] Guiasu, S. (1977) *Information Theory with Applications*. McGraw-Hill, New York.