Springer**Plus**

Open Access

CrossMark

A methodology to determine the maximum value of weighted Gini–Simpson index

José Pinto Casquilho[1,2]*

*Correspondence:
josecasquilho@gmail.com
[2] Centro de Ecologia
Aplicada "Prof Baeta Neves",
InBio, Instituto Superior de
Agronomia, Universidade
de Lisboa, Tapada da Ajuda,
1349-017 Lisbon, Portugal
Full list of author information
is available at the end of the
article

Abstract

Weighted Gini–Simpson index is an analytical tool that promises to be widely used concerning biological and economics applications, relative to the assessment of diversity measured by compositional proportions of a system defined with a finite number of elementary states characterized by positive weights. In this paper, a current literature review on the theme is presented and the mathematical properties of the index are outlined, focusing on the location of the maximizer (maximum point) and evaluation of the maximum value, with emphasis in the role of the Lagrange multiplier critical value—closely related with the harmonic mean of the weights—which is shown to be a barrier concerning the feasibility of the solution. Sequential procedures are presented, either backward or forward, which are used to obtain the correct values of the maximum point coordinates, thus allowing for the computation of the right maximum value of the index. Also, new theoretical results are provided, such as the calculus of limits and partial derivatives related to the critical solution, used to assess of the effectiveness of the algorithms herein proposed and discussed.

Keywords: Weighted Gini–Simpson index, Critical solution, Feasibility of solution, Sequential procedures, Maximum point, Maximum value, Harmonic mean

Background

Weighted Gini–Simpson index may be considered a recent analytical tool since explicit mentions don't seem available before the end of last century, and few empirical applications are yet available, though its number seems to be increasing fast. The problem that will be addressed in this paper concerns the location of maximum point and evaluation of the maximum value of the index, which are not trivial issues, as standard formulas referred to in the main literature just are straightforwardly applicable when a full set of inequalities are simultaneously verified. Otherwise, one has to proceed using algorithms as will be outlined in this paper. First, as background, the literature and main characteristics of Simpson and Gini–Simpson indices and of the correspondent weighted versions will be reviewed, followed by results and discussion of the methodology here proposed. Last, some conclusions are drawn.

Springer Open

Simpson and Gini–Simpson indices

Considering a simplex of dimension $m - 1$ defined as $\Delta^{m-1} = \left\{ p_i \geq 0, i = 1, \ldots m; \sum_{i=1}^{m} p_i = 1 \right\}$, where the numbers p_i denote relative extension measures, usually probabilities or proportions,[1] Simpson's index, originally mentioned as a measure of the concentration of a classification (Simpson 1949) is evaluated with the formula $C = \sum_{i=1}^{m} p_i^2$ and its symmetric form $D = 1 - C$ is usually named[2] Gini–Simpson index (e.g., Rao 1982), and used as a measure of biological or phylogenetic diversity until today (e.g., Tryjanowski et al. 2015; Zaller et al. 2015; Brocchieri 2015), since we can rewrite the correspondent formula as $D = \sum_{i=1}^{m} p_i(1 - p_i)$ and interpret it associated to the probability that any two random individuals in a population are assigned to different populations or genetic clusters (e.g., Chybicki et al. 2014). In biological studies, more than seven decades ago, the term $p_i(1 - p_i)$ was already mentioned as the contribution to the sampling variance due to any one species being sometimes observed and sometimes not (Fisher et al. 1943), later stated as the probability of interspecific encounters (Hurlbert 1971), or the probability of drawing two individuals of different type from a given collection (Gregorius and Gillet 2008). Good (1953), in a paper inspired by Alan Turing, defined parametric measures of heterogeneity of populations of s species, evaluated as $c_{m,n} = \sum_{i=1}^{s} p_i^m \left(- \log p_i \right)^n$ with $m, n = 0,1,2,\ldots$ Using this formalism it follows that the case $c_{2,0}$ is Simpson's concentration index C, while $c_{1,1}$ is Shannon (1948) statistical entropy.

The attribution to Corrado Gini of the earliest formulation of index D, more than a century ago, was related with the themes of variability devoted to the measurement of quantitative phenomena, and mutability, this one concerned with the measurement of qualitative phenomena. It is mentioned that Gini presented about 13 versions of the index (Ceriani and Verme 2012) and measuring variability is considered to be at the core of his procedure (de Finetti 1931). Sen (2005) says that Gini index opened the avenue for research in diversity analysis of qualitative categorical data models.

Weighted Simpson and Gini–Simpson indices

It is not easy finding references relative to the use of weighted Simpson's concentration index $C_w = \sum_{i=1}^{m} w_i p_i^2$ which seems to have been first explicitly stated and used as an inverse measure for antigenic diversity of a virus population (Nowak 1994); it was also used recently as a price-weighted biodiversity index of catch in freshwater fisheries in Malawi (Kasulo and Perrings 2006).

Sharma et al. (1978) discussed a non-additive information measure they named "generalized useful information of degree α" relative to a utility information scheme with m positive real numbers w_i also defined in the simplex Δ^{m-1}, denoted parametrically[3] as $I^\alpha(W) = \sum_{i=1}^{m} w_i p_i \left(p_i^{\alpha-1} - 1 \right) / \left(2^{1-\alpha} - 1 \right)$ with $\alpha \neq 1$, from which we can retrieve weighted Gini–Simpson index evaluating the semi-value of $I^\alpha(W)$ with $\alpha = 2$, obtaining $\sum_{i=1}^{m} w_i p_i(1 - p_i)$. The weights $\{w_i\}_{i=1,\ldots,m}$ allow for taking into account different features

---

[1] In biodiversity studies, proportions of populations in a community are commonly designated as relative abundances of species.

[2] Also referred to as Simpson's index of diversity (e.g., Crist et al 2003; Niane et al. 2014).

[3] Aggarwal and Picard (1978) say that Emptoz had outlined an equivalent formula in 1976.

related to ecological or economic values of species or other components of a system characterized by the proportions $\{p_i\}_{i=1,\ldots,m}$ including the sampling effort, the phylogenetic distances or conservation values, to name a few possible applications.

Weighted Gini–Simpson (WGS) index seems to have been formerly conceived and studied as an analytical tool addressing the diagnosis of landscape mosaics composition (Casquilho 1999) where the maximum point of the index and its maximum value were discussed using Lagrange multipliers method. It was also stated as an approach used to assess inequality measures under the scope of utility theory (Sen 1999). Guiasu and Guiasu (2003) outlined conditional and weighted measures of ecological diversity presenting the formulas for the maximum value of WGS index and the optimal proportions, results which were further retrieved and generalized for triads of species (Guiasu and Guiasu 2010). Also, Casquilho (2009) discussed an issue relative to habitats valuation with complex numbers, conceiving weighted Gini–Simpson index as a sum of variances of interdependent Bernoulli variables indexed by positive characteristic values, either ecological or economic.

Several theoretical developments were presented in the following, from which stand out, concerning ecological or related biological fields: the application of weighted Gini–Simpson to assess ecomosaics compositional scenarios (Casquilho 2011); the application to biodiversity partitioning and measuring of diversity with respect to the pairs of species (Guiasu and Guiasu 2012); Ricotta et al. (2012), discussing Rao's quadratic index under the scope of functional rarefaction, claim that their method is suitable to be extended to any concave diversity measure including WGS index; also, weighted Gini–Simpson index was said to be closely related to a unified framework based on Hill numbers concerning specific, phylogenetic, functional and other diversity measures (Chiu and Chao 2012; Chao et al. 2014); Guiasu and Guiasu (2014) proceeded with developments concerning the use of the index as a biodiversity assessment tool for interdependent species; Pavoine and Izsák (2014) formulated a new parametric index of diversity related to Rao's quadratic entropy and discuss connections relative to other indices including WGS index; last, WGS index was combined with expected utility generating a non-expected utility device (Casquilho 2015). Other empirical studies or applications using WGS index will be mentioned in the discussion of results.

Stating the problem

The problem addressed and discussed in this paper is that, in general, the maximum point coordinates of WGS index must be computed with a sequential procedure, because the formulas available in the most relevant literature concerning the issue (e.g., Guiasu and Guiasu 2003, 2010) are valid only within limited ranges of values of the set of predefined nonnegative weights $\{w_i\}_{i=1,\ldots,m}$. If this remark is not scrutinized, the blind use of those formulas may inflate the proportions of the heaviest weighted components and lead to an erroneous "maximum value" evaluation, which can have pernicious consequences in subsequent normalization procedures or other inferences on the subject. Though the problem was previously mentioned (Casquilho 1999, 2009, 2011), it was not fully systematized and analytically focused, and one of the procedures proposed in this paper is new.

In fact, the problem at stake has an old root, as Jaynes (1957) had already pointed out that the negative term $-\sum p_i^2$ has the difficulty arising from the fact that conditional maxima

cannot be found by a stationary property involving Lagrange multipliers, because the results do not, in general, satisfy the axiomatic condition $p_i \geq 0$. We will see that it is such a kind of problem which is at the core of the subject that will be discussed in the following.

Next, the main mathematical properties of weighted Gini–Simpson index will be reviewed, focusing on the critical solution and the meaning of the Lagrange multiplier value as a parameter controlling the feasibility of the solution. Then, sequential backward and forward procedures or algorithms are outlined, associated with simple numerical examples illustrating the performance of the method, and last, results will be discussed.

Review of the theoretical framework

Consider m interdependent Bernoulli variables B_i where v_i is a characteristic positive value, with associated probabilities $P[B_i = v_i] = p_i$ satisfying normalized measure space definition Δ^{m-1}, and $P[B_i = 0] = q_i$, with $q_i = 1 - p_i$; thus $\sum_{i=1}^{m} q_i = m - 1$ equates the dimension of the simplex with m vertices. Computing the variance of B_i we obtain $Var(B_i) = v_i^2 p_i(1 - p_i)$ from what follows the sum of variances $\sum_{i=1}^{m} Var(B_i) = \sum_{i=1}^{m} v_i^2 p_i(1 - p_i)$. Renaming v_i^2 as $v_i^2 = w_i$, the weighted Gini–Simpson index, measuring the variability of a system with such a characterization, is defined with the formula (1):

$$D_w = \sum_{i=1}^{m} w_i p_i (1 - p_i) \tag{1}$$

Index D_w is a continuous real function with domain in a compact set, the $m - 1$ simplex, what entails Bolzano–Weierstrass theorem to ensure that the index attains maximum and minimum values, as well as all the intermediate in its range. The inequality $D_w \geq 0$ is easily seen to be true as the index is conceived as a sum of nonnegative terms, thus one can conclude straightforwardly that the minimum value $D_w = 0$ is reached at every vertex of the simplex ($p_j = 1$ and $p_i = 0$ if $i \neq j$). Moreover, it is shown that index D_w is a concave function (e.g. Casquilho 1999; Guiasu and Guiasu 2010).

Lagrange multiplier method and feasibility of solution

Also, index D_w is a real differentiable function, hence one can use the auxiliary Lagrangian function denoted as $L = \sum_{i=1}^{m} w_i p_i (1 - p_i) - \alpha \left(\sum_{i=1}^{m} p_i - 1 \right)$ as an analytical tool for finding candidates to constrained extreme points of D_w (e.g. Bertsekas 1996) located in the hyperplane defined by the equation $\sum_{i=1}^{m} p_i = 1$. The calculus of generic partial derivative(s) in the variable(s) p_i entails the following result: $\partial L / \partial p_i = w_i - 2w_i p_i - \alpha$.

Searching the critical or stationary point of function L implies the system of equations:

$$p_i = (w_i - \alpha)/(2w_i) \quad \text{for } i = 1, \ldots, m. \tag{2}$$

As the weights are positive real numbers by hypothesis ($w_i > 0$) one can state the immediate conclusion that the optimal proportions should verify the conditions $p_i \geq 0 \Leftrightarrow w_i \geq \alpha$, from what follows that the value of the Lagrange multiplier is a barrier, or limit value, concerning the feasibility of the solution evaluated with this method.

Computing the associated closure condition $\sum_{i=1}^{m} p_i = 1$ using (2), one obtains the critical value of the Lagrange multiplier:

$$\alpha^* = (m-2) \left/ \left(\sum_{i=1}^{m} \frac{1}{w_i} \right) \right. . \tag{3}$$

From what follows that the critical point evaluated combining (2) and (3) is defined by the equations:

$$p_i^* = \frac{1}{2} + \left(1 - \frac{m}{2}\right) \left/ \left(w_i \sum_{j=1}^{m} \frac{1}{w_j} \right) \right. \quad \text{for } i = 1, \dots, m \tag{4}$$

Formula (5) presented next, relative to the presumed maximum value of the index[4] is the result of the evaluation of (1) replacing $\{p_i\}_{i=1,\dots,m}$ with the critical proportions defined in (4).

$$D_w^* = \frac{1}{4} \sum_{j=1}^{m} w_j - \left(1 - \frac{m}{2}\right)^2 \left/ \left(\sum_{j=1}^{m} \frac{1}{w_j} \right) \right. . \tag{5}$$

Results and discussion

From Eq. (3) one can conclude immediately that for $m \geq 3$ the inequality $\alpha^* > 0$ is verified and for $m = 2$ reduces to $\alpha^* = 0$ which implies the trivial result when the simplex is 1-dimensional: $p_1^* = p_2^* = 0.5$; also, the critical coordinates p_i^* evaluated with (2) verify intrinsically the condition $p_i^* \leq 1$, as the following equivalences show $p_i^* \leq 1 \Leftrightarrow w_i - \alpha^* \leq 2w_i \Leftrightarrow -\alpha^* \leq w_i$ which is true because $\alpha^* \geq 0$, the equality sign just holding for the 1-simplex; last, whether $w_i = \alpha^*$ one gets the value $p_i^* = 0$. Next, it will be proved that the critical solution defined by Eq. (4) may not be feasible and, subsequently, cannot be used straightforwardly to evaluate the maximum value of the index as defined by formula (5).

Analytical study of the critical point

Both formulas (4) and (5) are the right results whenever we have $w_i \geq \alpha^*$ for $i = 1,\dots,m$ meaning that m inequalities must be verified simultaneously. Whether there is at least one value that verifies $w_i < \alpha^*$ the evaluation of the optimal solution needs a revision in a sequential procedure, though that can only happen if $m \geq 4$. In fact, successive replacements and simplifications allow for obtaining the equivalent results:

$$w_i > \alpha^* \overset{i=1,\dots,m}{\Leftrightarrow} w_i > (m-3) \left/ \left(\sum_{j \neq i} \frac{1}{w_j} \right) \right. \tag{6}$$

Thus, for $m = 3$ one can see that the condition $w_i > 0$ is trivially verified and the critical proportions are properly defined as the maximizer coordinates:

$$p_i^* = \frac{1}{2} - 1 \left/ \left(2w_i \sum_{j=1}^{3} \frac{1}{w_j} \right) \right. \quad \text{for i = 1,2,3} . \tag{7}$$

---

[4] Here we use the notation of Guiasu and Guiasu (2003); an equivalent formula with a different notation may be found in Casquilho (1999:121,122).

A direct inspection of formula (4)—for which may be helpful to rewrite the denominator as $w_i \sum_{j=1}^{m} \frac{1}{w_j} = 1 + \sum_{j \neq i} \frac{w_i}{w_j}$—shows that the critical proportion p_i^* increases with the value of the corresponding w_i when all the other weights remain fixed, and, on the contrary, decreases with the increasing value(s) of other $w_j (j \neq i)$.

The calculus of limits on formulas (4) for $m \geq 3$ also clarifies the issue: $\lim_{w_i \to +\infty} p_i^* = \lim_{w_i \to +\infty} \left(\frac{1}{2} + \left(1 - \frac{m}{2} \right) / \left(1 + \sum_{j \neq i} \frac{w_i}{w_j} \right) \right) = \frac{1}{2}$, and from this result one can conclude that there is a supremum (least upper bound) for the optimal point coordinates: $p_i^* < 0.5$ in the context (and $p_i^* = 0.5$ if $m = 2$); also, $\lim_{w_i \to 0^+} p_i^* = 3/2 - m/2$ which is the same result that is obtained when w_i is fixed and all the other weights $w_k (k \neq i)$ tend simultaneously to positive infinite. For example, if $m = 5$ we get the result $\lim_{\{w_k\}_{k \neq i} \to +\infty} p_i^* = -1$. This negative value shows that when applying formulas (4) we can obtain non-feasible solutions, becoming negative without bound as the dimension of the simplex increases.

Also, the calculus of partial derivatives in Eq. (4) shows that the value p_i^* increases with the corresponding weight w_i and decreases when any other weight w_k increases. In fact, one can see that, for $m \geq 3$, the following inequalities hold:

$$\frac{\partial p_i^*}{\partial w_i} = \left(\frac{m}{2} - 1 \right) \left(\sum_{j \neq i} \frac{1}{w_j} \right) \left(w_i \sum_{j=1}^{m} \frac{1}{w_j} \right)^{-2} > 0 \quad \text{and}$$

$$\frac{\partial p_i^*}{\partial w_k} = \left(1 - \frac{m}{2} \right) \left(w_i \sum_{j=1}^{m} \frac{1}{w_j} \right)^{-2} \left(\frac{w_i}{w_k^2} \right) < 0.$$

So, what happens if there is any $w_i < \alpha^*$? Then, the corresponding critical value p_i^*, although lying in the hyperplane defined by the equation $\sum_{i=1}^{m} p_i^* = 1$ is not located in the simplex, and the solution is not feasible. In the general case with $m \geq 4$ there will be $m' \geq 3$ non-null optimal proportions as stated by Eq. (7) and $m - m'$ null coordinates. In some well balanced sets of weights it may happen that $m' = m$ but it is a particular case, not the general one.

Sequential forward procedure

Next, it is outlined a sequential forward procedure to reach the maximum point (maximizer) of the index. First, one sorts the weights in a decreasing order: $w_{(1)} \geq w_{(2)} \geq \cdots \geq w_{(m)}$. From Eq. (7), combined with the evaluation of limits and partial derivatives, it is known that it is guaranteed that the three highest weighted components will be in the optimal solution, with strictly positive proportions; the fourth highest weighted component is the first candidate to have a null value; then, one computes $\alpha_4^* = 2 / \left(\sum_{i=1}^{4} 1/w_{(i)} \right)$ and if $w_{(4)} \leq \alpha_4^*$ stop; hence $p_{(4)}^* = \cdots = p_{(m)}^* = 0$ and the number of vertices is set $m' = 3$; otherwise one has $w_{(4)} > \alpha_4^*$ and proceeds computing $\alpha_5^* = 3 / \left(\sum_{i=1}^{5} 1/w_{(i)} \right)$; whether $w_{(5)} \leq \alpha_5^*$ stop, and reset the values $p_{(5)}^* = \cdots = p_{(m)}^* = 0$ with $m' = 4$; otherwise, $w_{(5)} > \alpha_5^*$ and one proceeds until obtaining $w_{(k)} \leq \alpha_k^*$, then stop, setting $p_{(k)}^* = \cdots = p_{(m)}^* = 0$; hence $m' = k - 1$; in any case, the maximizer is located in a m'-face of the original $m - 1$ simplex.

Now, formulas (4) and (5) may be used replacing m by m' and calculating the optimal proportions and the maximum value of the index D_w with the corresponding set of weights—all the remaining optimal proportions being null and the respective weights discarded from the evaluation.

Exemplifying with a relatively small dimension $m = 5$, which enables the lower bound of a pseudo-optimal coordinate to be -1, as was shown in the calculus of the limits in the previous section. If the values of the weights are $w_{(1)} = 5$, $w_{(2)} = 4$, $w_{(3)} = 3$, $w_{(4)} = 2$ and $w_{(5)} = 1$, then the value of the Lagrange multiplier computed with (3) and all weights ($m = 5$) gives the result $\alpha^* = 1.3139$ implying that $w_{(5)} < \alpha^*$; using the sequential procedure, one calculates $\alpha_4^* = 1.5585$ and as $w_{(4)} > \alpha_4^*$ hence $p_{(5)}^* = 0$ and formulas (4) and (5) may be applied with $m' = 4$, discarding $w_{(5)} = 1$ from the calculations, giving the results of the optimal proportions: $p_{(1)}^* = 0.3441$, $p_{(2)}^* = 0.3052$, $p_{(3)}^* = 0.2403$, $p_{(4)}^* = 0.1104$ and $p_{(5)}^* = 0$. The maximum of the index in this case evaluates to $D_w^* = 2.7208$.

Changing the weights to be: $w_{(1)} = 50$, $w_{(2)} = 40$, $w_{(3)} = 30$, $w_{(4)} = 2$ and $w_{(5)} = 1$, and using the sequential forward procedure one computes $\alpha_4^* = 3.4582$ and verify that $w_{(4)} < \alpha_4^*$ hence sets $p_{(4)}^* = p_{(5)}^* = 0$ and $m' = 3$, thus discarding $w_{(4)}$ and $w_{(5)}$, proceeding to evaluate the non-null coordinates with Eq. (7), so obtaining the results: $p_{(1)}^* = 0.3724$, $p_{(2)}^* = 0.3404$ and $p_{(3)}^* = 0.2872$. In this case, the maximum value is $D_w^* = 26.808$ and, in this example, whether formula (5) was used blindly with all the original weights ($m = 5$) one would obtain the wrong pseudo-maximum value of 29.324 is misvalued about 10 % relative to the true value. When the dimension of the simplex increases, and the weights are disparate, this type of error could get worse in a kind of curse of dimensionality.

Sequential backward procedure

Whether the forward procedure previously discussed helps checking the consistency of the problem stated by the feasibility condition expressed in inequalities (6), one can see that a sequential backward procedure is more effective, applying directly Eq. (4) and nothing else. Adopting the same ordering $w_{(1)} \geq w_{(2)} \geq \cdots \geq w_{(m-1)} \geq w_{(m)}$ begin computing $p_{(m)}^*$ and if $p_{(m)}^* > 0$ then Eq. (4) are proper, all the coordinates can be calculated directly and also Eq. (5) applies straightforwardly with no problem; otherwise, if $p_{(m)}^* < 0$ then set $p_{(m)}^* = 0$ withdraw $w_{(m)}$ from further calculations and compute $p_{(m-1)}^*$ with Eq. (4) modified with $m' = m - 1$ and the corresponding set of weights; proceed with the same reasoning, recurring, until one finds an order (k) such that $p_{(k)}^* > 0$, then stop; set all null coordinates $p_{(k+1)}^* = \cdots = p_{(m)}^* = 0$, and Eq. (4) apply with dimension reset as $m = k$ evaluated with the corresponding weights $\{w_{(i)}\}_{i=1,\ldots,k}$.

Retrieving the numerical examples from the previous section one has again $w_{(1)} = 5$, $w_{(2)} = 4$, $w_{(3)} = 3$, $w_{(4)} = 2$ and $w_{(5)} = 1$ then computing $p_{(5)}^*$ with formula (4) and $m = 5$ one gets $p_{(5)}^* = -0.15693 < 0$; so, one sets $p_{(5)}^* = 0$, discards $w_{(5)} = 1$ and evaluates $p_{(4)}^*$ with $m' = 4$ and $\{w_{(i)}\}_{i=1,\ldots,4}$; next result is $p_{(4)}^* = 0.11039 \cong 0.1104 > 0$, so stop; all the other coordinates can be calculated now with Eq. (4) and $m' = 4$.

With the other example relative to the set of weights $w_{(1)} = 50$, $w_{(2)} = 40$, $w_{(3)} = 30$, $w_{(4)} = 2$ and $w_{(5)} = 1$ one has the following sequence: in the first step computes $p_{(5)}^* = -0.45037 < 0$ so one discards $w_{(5)} = 1$, sets $m' = 4$ and proceeds evaluating $p_{(4)}^*$

with the corresponding weights $\{w_{(i)}\}_{i=1,\cdots,4}$ obtaining $p^*_{(4)} = -0.36455 < 0$; then, one sets $p^*_{(4)} = 0$ with $m' = 3$ and proceeds evaluating $p^*_{(3)}, p^*_{(2)}$ and $p^*_{(1)}$ with the set $\{w_{(i)}\}_{i=1,\dots,3}$ and Eq. (7), obtaining, for example, the value $p^*_{(3)} = 0.28723 \cong 0.2872$; and the maximum value of D_w now can be evaluated using Eq. (5) with $m' = 3$.

Discussion

The index D_w is a continuous real function defined in a compact domain and its range is $0 \leq D_w \leq D^*_w$, the minimum value $D_w = 0$ occurring in every vertex of the simplex Δ^{m-1}. The maximum value of the index denoted D^*_w has to be evaluated verifying the feasibility condition expressed by inequalities (6) before applying straightforwardly Eq. (5). In general, except for very specific and balanced sets of weights, the maximum point of D_w will not occur in the interior of the simplex but in a *k*-face with $3 \leq k < m$, as was shown by the theoretical results followed by sequential procedures and illustrated with numeric examples, leading to some null optimal coordinates.[5] Obviously, the maximum value of WGS index could also be computed as $D^*_w = \sum_{i=1}^{k} w_i p^*_i (1 - p^*_i)$ thus avoiding Eq. (5), but that still implies checking the feasibility condition as the summing procedure just applies for positive proportions.

The optimal proportions p^*_i are insensitive to a change in unities: the positive linear transformation $u_i = c w_i$, $c > 0$ implies that the critical solution remains the same and the new value of the Lagrange multiplier is also linearly transformed to be $\alpha^{**} = c\alpha^*$ entailing that the feasibility condition (6) remains unchanged. The optimal value of the Lagrange multiplier α^* defined in (3) is closed related to the harmonic mean of the weights. How can we justify that α^* has numerator $m - 2$ instead of m? It seems that the most appealing interpretation is that when discussing result (3) we deduced that for $m = 2$ the value α^* vanishes and the maximum point is fixed: $(p^*_1, p^*_2) = (0.5, 0.5)$, independent of the weights. So, $m - 2$ is the number of relevant weights that affect the subsequent calculation of the maximum point coordinates and maximum value of the index.

The problem addressed in this paper is particularly important when the evaluation of D^*_w aims to be used in further normalization assessments with range $0 \leq D_w/D^*_w \leq 1$ and an erroneous computation of the maximum value can induce wrong conclusions when comparing different compositional systems. There are several empirical studies that use the maximum value of WGS index as a reference for further normalization assessments: besides Guiasu and Guiasu (2010) numeric examples such as the one relative to 10 species in two habitats with data retrieved from Jost et al. (2010), also Subburayalu and Sydnor (2012) used formula (5) when assessing street tree diversity in four Ohio communities and, probably, the feasibility condition here discussed was not checked. Weighted Gini–Simpson goes on being mentioned (e.g., Niane et al. 2014) and the problem handled in the present article seems to become relevant for the next future.

Conclusions

In this paper it was summarized an issue that seems to be relevant in the field of diversity measures: the proper evaluation of the maximum point and maximum value of weighted Gini–Simpson index. The main literature on the subject does not refer to the

---

[5] As it can happen with other biodiversity indices (e.g., Pavoine and Izsák 2014).

feasibility condition here discussed, what can involve consequent wrong results in applications. New theoretical results concerning the analytical study of the critical solution are provided, such as the calculus of limits and partial derivatives, as well as are sketched forward and backward procedures conceived to solve the issue at stake, also illustrated with numeric examples.

Author details

^{1} Universidade Nacional Timor Lorosa'e, Rua Formosa, Díli, Timor-Leste. ^{2} Centro de Ecologia Aplicada "Prof Baeta Neves", InBio, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal.

References

Aggarwal NL, Picard C-F (1978) Functional equations and information measures with preference. Kybernetika 14:174–181

Bertsekas DP (1996) Constrained optimization and Lagrange multiplier methods. Athena Scientific, Belmont

Brocchieri L (2015) Phylogenetic diversity and the evolution of molecular sequences. J Phylogenet Evol Biol 3:1. doi:10.4172/2329-9002.1000e109

Casquilho JAP (1999) Ecomosaico: índices para o diagnóstico de proporções de composição. Dissertation (doctoral thesis), Universidade Técnica de Lisboa. doi:10.13140/RG.2.1.4211.5608

Casquilho JP (2009) Complex number valuation of habitats and information index of landscape mosaic. Silva Lus 17(2):171–180

Casquilho JP (2011) Ecomosaic composition and expected utility indices. Silva Lus 19(1):55–65

Casquilho JP (2015) Combining expected utility and weighted Gini–Simpson index into a non-expected utility device. Theor Econ Lett 5(2):185–195. doi:10.4236/tel.2015.52023

Ceriani L, Verme P (2012) The origins of Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. J Econ Inequal 10(3):421–443. doi:10.1007/s10888-011-9188-x

Chao A, Chiu C-H, Jost L (2014) Unifying species diversity, phylogenetic diversity, functional diversity and related similarity and differentiation measures through Hill numbers. Annu Rev Ecol Evol Syst 45:297–324. doi:10.1146/annurev-ecolsys-120213-091540

Chiu C-H, Chao A (2012) Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers. PLoS ONE 9(7):e100014. doi:10.1371/journal.pone.0100014

Chybicki IJ, Waldon-Rudzionek B, Meyza K (2014) Population at the edge: increased divergence but not inbreeding towards northern range limit in *Acer campestre*. Tree Genet Genomes 10:1739–1753. doi:10.1007/s11295-014-0793-2

Crist TO, Veech JA, Gering JC, Summerville KS (2003) Partitioning species diversity across landscapes and regions: a hierarchical analisys of α, β, and γ diversity. Am Nat 162(6):734–743. doi:10.1086/378901

De Finetti B (1931) Sui metodi proposti per il calcolo della differenza media. Metron 9(1):47–52

Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. J Anim Ecol 12(1):42–58

Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40(3–4):237–264

Gregorius H-R, Gillet EM (2008) Generalized Simpson-diversity. Ecol Model 211(1–2):90–96. doi:10.1016/j.ecolmodel.2007.08.026

Guiasu RC, Guiasu S (2003) Conditional and weighted measures of ecological diversity. Int J Uncertain Fuzziness Knowl Based Syst 11:283–300. doi:10.1142/S0218488503002089

Guiasu RC, Guiasu S (2010) New measures for comparing the species diversity found in two or more habitats. Int J Uncertain Fuzziness Knowl Based Syst 18(6):691–720. doi:10.1142/S0218488510006763

Guiasu RC, Guiasu S (2012) The weighted Gini–Simpson index: revitalizing an old index of biodiversity. Int J Ecol. doi:10.1155/2012/478728

Guiasu RC, Guiasu S (2014) Weighted Gini–Simpson quadratic index of biodiversity for interdependent species. Nat Sci 6(7):455–466. doi:10.4236/ns.2014.67044

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52(4):577–586. doi:10.2307/1934145

Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106(4):620–630. doi:10.1103/PhysRev.106.620

Jost L, DeVries P, Walla T, Greeney H, Chao A, Ricotta C (2010) Partitioning diversity for conservation analysis. Divers Distrib 16:65–76. doi:10.1111/j.1472-4642.2009.00626.x

Kasulo V, Perrings C (2006) Fishing down the value chain: biodiversity and access regimes in freshwater fisheries—the case of Malawi. Ecol Econ 59:106–114. doi:10.1016/j.ecolecon.2005.09.029

Niane AA, Singh M, Struik PC (2014) Bayesian estimation of shrubs diversity in rangelands under two management systems in northern Syria. Open J Ecol 4:163–173. doi:10.4236/oje.2014.44017

Nowak MA (1994) The evolutionary dynamics of HIV infections. In: Joseph A, Mignot F, Murat F, Prum B, Rentschler R (eds) First European congress of mathematics Paris, July 1992. Progress in mathematics, Birkhäuser Basel, vol 120, pp 311–326. doi:10.1007/978-3-0348-9112-7_13

Pavoine S, Izsák J (2014) New biodiversity measure that includes consistent interspecific and intraspecific components. Methods Ecol Evol 5(2):165–172. doi:10.1111/2041-210X.12142

Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. Theor Popul Biol 21(1):24–43. doi:10.1016/0040-5809(82)90004-1

Ricotta C, Pavoine S, Bacaro G, Acosta ATR (2012) Functional rarefaction for species abundance data. Methods Ecol Evol 3(3):519–525. doi:10.1111/j.2041-210X.2011.00178.x

Sen PK (1999) Utility-oriented Simpson-type indexes and inequality measures. Calcutta Stat Assoc Bull 49:1–21

Sen PK (2005) Gini diversity index, Hamming distance, and curse of dimensionality. METRON Int J Stat 63(3):329–349

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Sharma BD, Mitter J, Mohan M (1978) On measures of "useful" information. Inf Control 39(3):323–336. doi:10.1016/S0019-9958(78)90671-X

Simpson EH (1949) Measurement of diversity. Nature 163:688. doi:10.1038/163688a0

Subburayalu S, Sydnor TD (2012) Assessing street tree diversity in four Ohio communities using the weighted Simpson index. Landsc Urban Plan 106(1):44–50. doi:10.1016/j.landurbplan.2012.02.004

Tryjanowski P, Sparks TH, Biadu W, Brauze T, Hetmański T, Martyka R, Skórka P, Indykiewicz P, Myczko L, Kunysz P, Kawa P, Czyż S, Czechowski P, Polakowski M, Zduniak P, Jerzak L, Janiszewski T, Goławski A, Duduś L, Nowakowski JJ, Wuczyński A, Wysocki D (2015) Winter bird assemblages in rural and urban environments: a national survey. PLoS ONE 10(6):e0130299. doi:10.1371/journal.pone.0130299

Zaller JG, Kerschbaumer G, Rizzoli R, Tiefenbacher A, Gruber E, Schedl H (2015) Monitoring arthropods in protected grasslands: comparing pitfall trapping, quadrat sampling and video monitoring. Web Ecol 15:15–23. doi:10.5194/we-15-15-2015