Short communication

A simple formula for calculating the lower limit of Shannon's diversity index

Jean-Nicolas Beisel *, Jean-Claude Moreteau

Centre de Recherches Ecologiques, Université de Metz, 1 Rue des Récollets, BP 4116, 57040 Metz Cedex 01, France

Accepted 19 December 1996

Abstract

The purpose of this note is to demonstrate that the exact value of the lower limit of Shannon's diversity index (called $H_{min}$) can be calculated with a very simple formula when data are in the form of counts: $H_{min} = \ln(Q) - \frac{(Q - S + 1)\ln(Q - S + 1)}{Q}$ where $S =$ species richness, $Q =$ total number of individuals, $S$ and $Q$ are integers and $Q \geq S$. $H_{min}$ is generally neglected in ecological studies although it can be very different from zero. $H_{min}$ is more and more important when the ratio $S/Q$ approaches unity and, then, the interpretation of Shannon's diversity or of an evenness index should take it into account. For this reason, we recommend the use of Hurlbert's index as an evenness measure rather than Pielou's index when $H_{min} > 0$. Calculations realized on three examples show that results can be appreciably different. Depending on the cases, evenness differences between two communities can be increased or decreased if we use Hurlbert's instead of Pielou's index. Moreover, one example allowed us to show that an evenness calculated with Pielou's index, even if it is large ($= 0.67$), can be in fact the minimal diversity which is mathematically possible. In all of these cases, the ecological interpretation of data can be highly simplified when using Hurlbert's index, owing to the fact that $H_{min}$ is integrated into the calculation of this index. © 1997 Elsevier Science B.V.

Keywords: Shannon index; Minimal diversity; Eveness measures

1. Introduction

Shannon's measure (Shannon and Weaver, 1963) has probably been the most widely used diversity index in ecology (Ludwig and Reynolds, 1988). The information theory from which it derived, and its use, widely exceeds the different fields of biology.

Shannon's index is estimated using the equation (Pect, 1974):
where \( S \) = species richness, \( q_i \) = the number of individuals in the \( i \)th species, and
\[
Q = \sum_{i=1}^{S} q_i
\]
i.e., the total number of individuals.

If the data are in the form of counts, for a given species number \( S \) and a sample size \( Q \), its minimum \( \left(H'_{\text{min}}\right) \) and maximum \( \left(H'_{\text{max}}\right) \) possible values can be calculated with the equations (Boyle et al., 1990):
\[
H'_{\text{min}} = \ln(Q!) - \ln[(Q - (S - 1))!] \quad (2)
\]
and
\[
H'_{\text{max}} = \ln(S) \quad (3)
\]
where \( S \) and \( Q \) are integer, \( Q \geq S \) and \( Q! = \text{total number of individuals factorial} \).

These upper and lower limits of the Shannon diversity index are helpful for the interpretation of data and are useful for calculating an evenness index (i.e., a relative diversity index) which assesses how individuals of the community are distributed among present species (see Hurlbert, 1971). Two formulations of an evenness index (that will be labeled \( J \) and \( J' \) in this note), which converge for large samples, were initially frequently used (Hurlbert, 1971; Peet, 1974) (1) \( J = H'/H'_{\text{max}} \) (Pielou, 1966) and (2) \( J' = (H' - H'_{\text{min}})/(H'_{\text{max}} - H'_{\text{min}}) \) or a derived measure like redundancy \((= 1 - J')\). The difference between these two formulations is that \( J \) does not take into account the minimum possible diversity value \( \left(H'_{\text{min}}\right) \) for both the given species number \( S \) and the sample size \( Q \).

In practice, the calculation of \( H'_{\text{min}} \) is easy when \( Q \) is very small (< 170). When \( Q \) is larger, we can only estimate \( H'_{\text{min}} \) due to the difficulty in calculating \( \ln(Q!) \). The reason being that the representation of numbers in a computer is limited (for a 64 byte computer, it is impossible to codify a figure superior to \( 1.8 \times 10^{308} \), this very figure being smaller than 171!). This makes it necessary to use an approximate method of calculation, such as Stirling’s, for calculating \( \ln(Q!) \). Many authors consider \( H'_{\text{min}} \) as been equal to zero, probably to avoid this difficulty. Therefore, the evenness measure systematically used is Pielou’s index \( J \).

From an ecological perspective, \( H'_{\text{min}} \) becomes all the more important when \( Q \) approaches the \( S \) value, that is to say, when the ratio \( S/Q \) approaches unity. In fact, if \( H'_{\text{min}} \) cannot be considered equal to 0, the evenness calculated using Pielou’s index does not vary between 0 and 1, as commonly read, but between \( J'_{\text{min}} = H'_{\text{min}}/H'_{\text{max}} \) and 1. Biological interpretation of data, then, should take this into account.

The specific purpose of this note is to demonstrate that the exact value of \( H'_{\text{min}} \) can be calculated with a very simple formula valid when \( S \) and \( Q \) are integers and when \( Q \geq S \) (= data in the form of counts). To our knowledge, it is not very well known and has never been used in biology. For example, in rare ecological studies where an evenness measure as redundancy was used, the authors estimate \( H'_{\text{min}} \) both with the Eq. (2) and the Stirling’s approximation. The utility of \( H'_{\text{min}} \) is discussed with the help of three examples of work where the authors used both Shannon and Pielou indices without considering \( H'_{\text{min}} \).

2. Demonstration

Shannon’s diversity can be rewritten in a formula easily computable:

\[
H' = -\frac{1}{Q} \sum_{i=1}^{S} q_i \ln\left(\frac{q_i}{Q}\right) \quad (1)
\]
\[
\Rightarrow H' = -\frac{1}{Q} \sum_{i=1}^{S} q_i [\ln(q_i) - \ln(Q)] \quad (2)
\]
\[
\Rightarrow H' = -\frac{1}{Q} \sum_{i=1}^{S} [q_i \ln(q_i)] + \frac{\ln(Q)}{Q} \sum_{i=1}^{S} [q_i] \quad (3)
\]
with
\[
\sum_{i=1}^{S} [q_i] = Q
\]
\[
\Rightarrow H' = \ln(Q) - \frac{1}{Q} \sum_{i=1}^{S} [q_i \ln(q_i)] \quad (4)
\]

What is minimal diversity? The lower limit of a diversity measure corresponds to a community for which (Kvalseth, 1991):
Table 1
Examples of calculation results of (1) the lower limit of Shannon index ($H^{\text{min}}$), (2) Hurlbert's evenness ($J'$) and (3), the difference in percent between Pielou's evenness ($J$) and Hurlbert's evenness. All the other informations are those presented by their authors in references cited. $S$, richness; $Q$, abundance; $H'$, Shannon diversity.

<table>
<thead>
<tr>
<th>References</th>
<th>Location</th>
<th>$S$</th>
<th>$Q$</th>
<th>$H'$</th>
<th>$J$</th>
<th>$H^{\text{min}}$</th>
<th>$J'$</th>
<th>$J-J'$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Battegazzore et al., 1992 (Table 3)$^a$</td>
<td>Station 1</td>
<td>26</td>
<td>139</td>
<td>2.12</td>
<td>0.65</td>
<td>1.05</td>
<td>0.48</td>
<td>26.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>1382</td>
<td>1.33</td>
<td>0.42</td>
<td>0.14</td>
<td>0.39</td>
<td>7.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>233</td>
<td>1.03</td>
<td>0.38</td>
<td>0.39</td>
<td>0.28</td>
<td>26.32</td>
</tr>
<tr>
<td>Example 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sousa and da Gama, 1994 (Table 2)$^b$</td>
<td>Eucalyptus (Humus)</td>
<td>18</td>
<td>263</td>
<td>2.38</td>
<td>0.57</td>
<td>0.61</td>
<td>0.5</td>
<td>12.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>27</td>
<td>2.04</td>
<td>0.79</td>
<td>1.12</td>
<td>0.63</td>
<td>20.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>106</td>
<td>2.47</td>
<td>0.69</td>
<td>0.84</td>
<td>0.59</td>
<td>14.49</td>
</tr>
<tr>
<td>Example 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waringer, 1996 (Table 4, site 1)$^c$</td>
<td>Ephemeroptera only</td>
<td>3</td>
<td>8</td>
<td>0.74</td>
<td>0.67</td>
<td>0.74</td>
<td>0.01</td>
<td>98.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>72</td>
<td>0.7</td>
<td>0.64</td>
<td>0.15</td>
<td>0.58</td>
<td>9.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>46</td>
<td>1.3</td>
<td>0.63</td>
<td>0.72</td>
<td>0.43</td>
<td>31.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>126</td>
<td>1.79</td>
<td>0.68</td>
<td>0.6</td>
<td>0.58</td>
<td>14.71</td>
</tr>
</tbody>
</table>

$^a$Macroinvertebrates (River Po, Italy).
$^b$Collembolan in coppices (Castelo Branco, Portugal).
$^c$Aquatic insect larvae (Austria).

$q_i = Q - (S - 1)$
for one species $i$.

$q_j = 1$
for all species $j \neq i$.

If we use the Eq. (4) to calculate $H^{\text{min}}$, then the $S - 1$ species $j$ with 1 individual do not contribute to the calculation of $H^{\text{min}}$ because $\ln(1) = 0$. Thus, minimal diversity value depends only on the species $i$ which has $(Q - S + 1)$ individuals. If we replace $q_i$ with $(Q - S + 1)$ in the Eq. (4), we obtain:

$$H^{\text{min}} = \ln(Q) - \frac{[(Q - S + 1) \ln(Q - S + 1)]}{Q}$$

In conclusion, for a community with $S$ species and $Q$ individuals, this simple formula Eq. (5) gives the exact minimal value ($H^{\text{min}}$) of Shannon's diversity index.

3. Discussion

To illustrate the useful purpose of $H^{\text{min}}$ computation, Table 1 presents three recent examples of work where Shannon's diversity and Pielou's evenness were used. The data are exactly those presented by the authors but in addition we add our calculations of $H^{\text{min}}$, Hurlbert's evenness and the difference in percentage between Pielou's and Hurlbert's indices.

It has long been recognized that structural indices are efficient in comparative frameworks (Magurran, 1988). For different communities of the same study, $H^{\text{min}}$ can be considered negligible because it is close to zero, or on the contrary, can be considered as being important (Table 1). These differences have particularly strong repercussions on differences between Pielou's and Hurlbert's indices. Because Hurlbert's evenness integrates $H^{\text{min}}$, it can help in differentiating communities which appear as having the same evenness value to that of Pielou's index (Stations 1 and 2 from the first example). On the other hand, the fact of integrating $H^{\text{min}}$ into Hurlbert's evenness can reduce evenness differences between two communities (Communities of soil and pitfall from the second example). We come to the same observation when realizing calculations on only one part of a community as shown in the third example of Table 1. This example illustrates the diversity of
calculations carried out on each of the three insect orders and then, on the orders all together. Pielou’s evenness did not allow us to clearly differentiate the four group structures even when Hurlbert’s index permitted it, $H_{\min}^\prime$, being important for 3 out of the 4 calculations realized. We observe that in the specific case of $S$ and $Q$ having close values (Example 3, Ephemeroptera only), the diversity calculated is the minimal diversity which is mathematically possible ($H^\prime = H_{\min}^\prime$). Consequently, the evenness value observed with Pielou’s index ($J = 0.67$), even if it is large, is in fact approximately the minimal evenness which is mathematically possible (i.e. $J_{\min} = H_{\min}^\prime/H_{\max}^\prime$). This demonstration is very tedious without knowing the value of $H_{\min}^\prime$, and the conclusion is obvious when using Hurlbert’s index ($J' = 0.01$).

Depending on the problematic studied, taking into consideration the lower limit of the Shannon’s diversity index can give differences in results. This could modify the biological interpretation of the structure of studied communities considerably. For this reason, when $H_{\min}^\prime$ is different from zero, we think that Hurlbert’s index (Hurlbert; 1971), or a derived formula like redundancy (Peet, 1974), is a better evenness measure than Pielou’s index. The reason is that Hurlbert’s measure integrates $H_{\min}^\prime$ and consequently always ranges from 0 to 1 (Pielou’s index ranges from $H_{\min}^\prime/H_{\max}^\prime$ to 1), which simplifies the interpretation of ecological data.

This very simple formula presented gives the exact value of the lower limit of Shannon’s diversity index. It allows Shannon index users to verify if $H_{\min}^\prime$ is negligible or not, so that they can integrate this result into their interpretation of data. Potentially, the minimal diversity value could also be incorporated into a biodiversity monitoring model as the DIMO (Qinghong, 1995), based on the Shannon index for analyzing and presenting state and change in plant species diversity at community level.

Acknowledgements

We wish to thank the anonymous referees for their comments on an earlier version of the manuscript. Many thanks to Mrs Antoine and Mrs Chaumeil for the linguistic corrections of the manuscript.

References