Diversity indices

1 Shannon index

Also known as the Shannon-Weaver index, Shannon-Wiener index, and Shannon-entropy. It is used to calculate the diversity of a community by taking into account species richness and abundance of those species. It is calculated as:

$$H' = -\sum_{i=1}^{s} p_i \ln p_i \tag{1}$$

where p is the proportion of individuals of a species i divided by the total number of individuals found in all species (proportional abundance). The Shannon index therefore takes the sum of all the proportional abundances of each species multiplied by the natural log of the proportional abundances of each species. When there is only one species in the dataset, $H' = 0$ and can increase to $\ln S$ if each species has the same abundance, where S is the number of species in the community.

The Shannon equitability index is the Shannon index divided by maximum diversity:

$$E_H = \frac{H}{\log S} \tag{2}$$

E_H is normalised between 0 and 1, where higher values denote greater abundance evenness. In a community of equally common species, $E_H = 1$.

The Shannon index varies in a non-linear fashion with diversity. For example, $H' = 6$ corresponds to a community of 403 equally common species, while $H' = 5.5$ corresponds to a community of 244 equally-common species. The figures below shows how the value of the Shannon index varies with the number of equally common species, and cohorts of species with different sequential abundance distances:
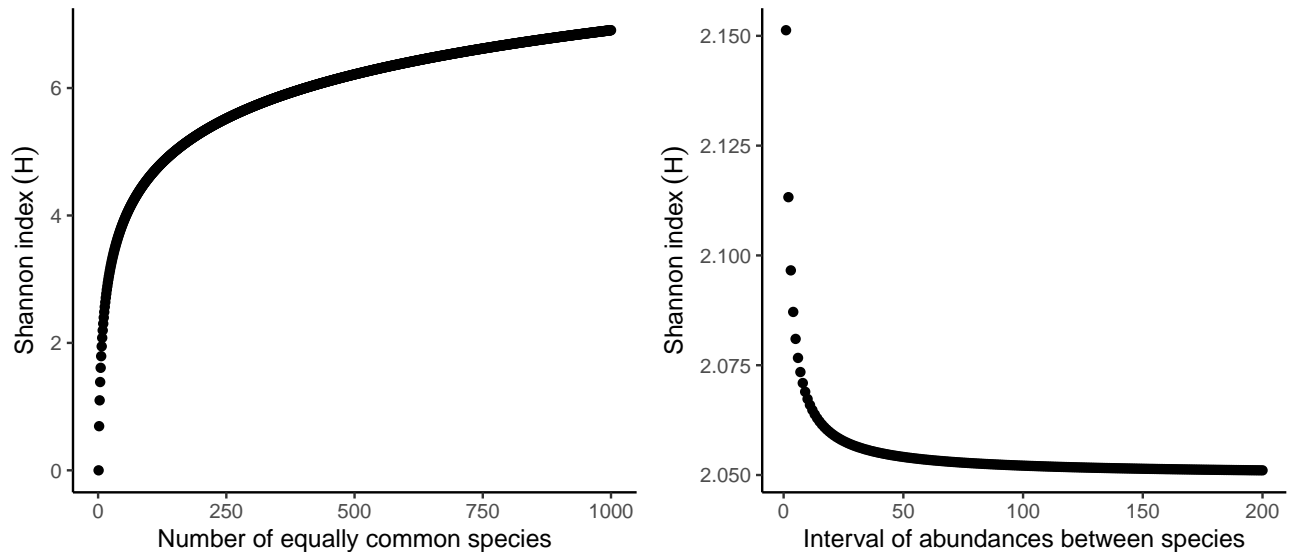
The Shannon index gives more weight per individual to additions of rare species (relative to the rarity of other species in the community) via the natural log term.

The Shannon index is based on the weighted geometric mean of proportional abundances, and it equals the

logarithm of "true diversity" (^1D), hence the non-linear variation with abundance evenness and number of species:

$$H' = \ln {}^1D \tag{3}$$

Therefore, to calculate the order 1 modified numbers-equivalent "true diversity":

$$^1D = e^{(\sum_{i=1}^{S} p_i \ln p_i)} = e^{H'} \tag{4}$$

Where p_i is the proportion of taxa i, S is the number of species. "True" diversity quantifies how many "effective species" the dataset represents given the mean proportional abundance of the observed species. In a community of equally common species $^1D = S$. So a community of 16 equally common species $^1D = 16$ and for a community of 8 equally common species $^1D = 8$. This idea of a "true diversity" can be formalised as:

The "true diversity" is the inverse of the mean of the proportional abundances
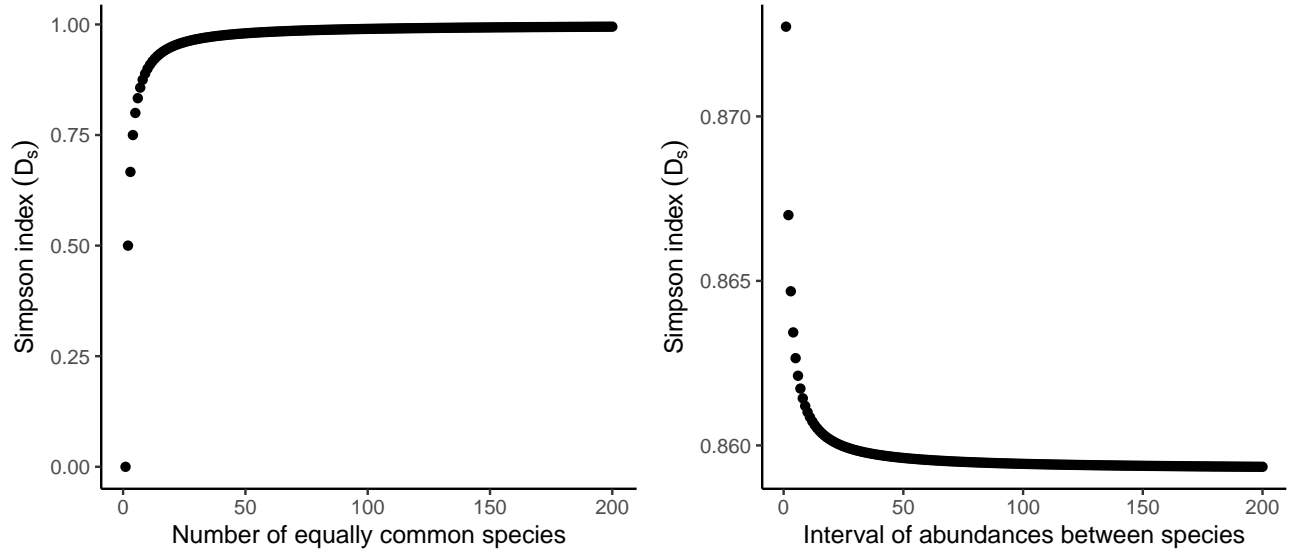
2 Simpson index

$$D = \sum_{i=1}^{S} p_i^2 = \sum_{i=1}^{S} (n/N)^2 = \frac{\sum_{i=1}^{s} n(n-1)}{N(N-1)} \tag{5}$$

Where p_i is the proportional abundance of species i, n is the number of individuals of a particular species, N is the total number of individuals across species, and S is the number of species. Simpson's index therefore takes the sum of all squared proportional abundances of each species. Simpson's index is actually a similarity index, higher values indicate lower diversity and range between 0 and 1. $D = 1$ indicates that the probability of two individuals randomly selected from the dataset are the same. To make the index more intuitive, it is often expressed as:

$$D_s = 1 - D \tag{6}$$

Where D_s is the Gini-Simpson index, or the Simpson's diversity index. This represents the probability that two individuals taken at random from the dataset are of different species. This is also known as the Probability of Interspecific Encounter (PIE). It also ranges between 0 and 1, where 1 represents infinite diversity, i.e. it is guaranteed that every individual sampled will be of a different species.

Simpson's diversity index is not sensitive to rare species. The addition of a few species with very low abundances will not affect the index much, this is due to the proportional abundances being squared in the equation. Simpson's index therefore gives more weight to common or dominant species. Simpson's index saturates quickly if more equally common species are added. The figures below shows how the value of the Simpson's index varies with the number of equally common species, and cohorts of species with different sequential abundance distances:

Simpson's index is also sometimes expressed as:

$$^{2}D = 1/D \tag{7}$$

Where ^{2}D is the Simpson's reciprocal index, or the Simpson's dominance index. ^{2}D is the "true diversity" of order 2. The lowest value is 1, with only one species, and increases to infinity.

3 Beta diversity

Beta diversity measures the difference between communities and can be expressed as:

$$\beta = \gamma/\alpha \tag{8}$$

Where β is beta diversity, α is alpha diversity, i.e. the mean of the diversities of individual community, and γ is gamma diversity, i.e. the pooled diversity of all communities. α and γ can be calculated using many of the diversity indices already mentioned, including Simpson's index, the Shannon index, Simpson's reciprocal index, or the modified true numbers equivalent Shannon index. Larger values of β indicate greater differences in diversity between communities.

4 Proportional species turnover

$$B_{Mt-1} = \frac{\gamma - \alpha}{\gamma} \tag{9}$$

Where α is the mean alpha diversity across communities and γ is the total diversity pooled across all communities. The α and γ components of B_{Mt-1} can be calculated using many different diversity indices, including Shannon and Simpson's diversity indices. B_{Mt-1} gives the proportion of species which are not contained in the average unit, when comparing multiple units.

5 Jaccard and Sørensen similarity coefficient

For measuring the difference in diversity between communities.

$$S_J = a/(a + b + c) \tag{10}$$

Where a is the number of species shared between two communities, b is the number of species unique to community 1, and c is the number of species unique to community 2. S_J is frequently multiplied by 100 to give a "percentage difference" and can be transformed to a measure of difference $D_J = 1 - S_J$.

The Sørensen coefficient is very similar but gives greater weight to species common between communities than those unique to a single community, by multiplying shared species by 2:

$$S_S = 2a/(2a + b + c) \tag{11}$$

Again, it can be multiplied by 100 to give a "percentage difference" and expressed as a difference metric as $D_S = 1 - S_S$.

6 Problems with diversity indices

Abundance based diversity indices provide useful results, effectively combining species richness and abundance into a single value, but they are not themselves "true diversities". Consider the Shannon index, we would intuitively expect that a community containing 16 equally common species should be considered twice as diverse as a community with 8 equally common species. However, the Shannon index would give values of 2.77 and 2.08, respectively. You may choose to quantify diversity using a "true diversity" index instead.

There are many diversity indices and each is useful in different types of study, so care should be taken to choose the right indices. There is value also in choosing an index which is widely used and so can be compared more easily with other literature values. The Simpson's index and Shannon index are two such widely used choices. It is common to report both Simpson's and Shannon's indices in a report, when comparing multiple sites.

7 "Order" of diversity indices

The term "order" refers to an unexpected unity underlying all standard diversity indices, given their shared form:

$$^q\gamma = \sum_{i=1}^{S} p_i^q \tag{12}$$

Where γ is the diversity index value of order q. Every diversity measure has a numbers equivalent, which increases as expected as more equally common species are added, which is symbolised as ^qD, where q is the "order". The numbers equivalent of any standard diversity index can be given as:

$$^qD = (\sum_{i=1}^{S} p_i^q)^{1/(1-q)} = (^q\gamma)^{1/(1-q)} \tag{13}$$

Species richness has an order 0, the Shannon index is of order 1 because the proportional abundances p_i aren't raised to the power of anything, Simpson's index is of order 2 because all the proportional abundances are squared. Orders higher than 1 are disproportionately sensitive to additions of common species. Expressions with $q = 1$ will equally favour common or rare species

8 Further reading

Jost, L. (2006), 'Entropy and diversity', *Oikos* **113**(2), 363–375.

Jost, L. (2007), 'Partitioning diversity into independent alpha and beta components', *Ecology* **88**(10), 2427–2439.

Jost, L. (2019), 'Effective number of species', http://www.loujost.com/. Accessed: 2019-09-03.

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.