Section: Diversity analysis

Diversity indices

Theory R functions Examples

This section will overview commonly used indices measuring diversity of ecological community (species richness, Shannon index, Simpson index). We will also introduce the measures of evenness, concept of effective number of species and general framework of Hill numbers.

As already mentioned in the general section, diversity consists of two components – species richness, i.e. the number of species in the community, and evenness, i.e. the fact that some species in community are common and other are rare. All diversity measures suffer from the same problem – they depend on the sampling effort, i.e. energy spent by researcher to discover all species present in a community. Rare species, i.e. those represented by few individuals, small biomass or low cover, are those most easily undetected, requiring disproportionately larger sampling effort than common species.

Diversity indices reviewed below differ from each other in the weight they put on either species richness or evenness.

Species richness

(denoted as *S* here) is the most intuitive and natural index of diversity, and I bet that it is used the most frequently in studies dealing with diversity. However, it is also the most sensitive to the difference in sampling effort, since it weights all species equally independent from their relative abundances, i.e. rare species count equally to common species although they are more likely to be undetected.

Shannon index

(or Shannon entropy[2], Shannon-Wiener or (incorrectly) Shannon-Wiever; denoted as *H*, *H'* or H_{Sh}) considers both species richness and evenness. The index is derived from information theory and represents the uncertainty with which we can predict of which species will be one randomly selected individual in the community. If community contains only one species, the uncertainty is zero, since we are sure that randomly chosen individual will belong to that one only species. The more species the community contains, the more uncertainty increases; in a diverse community, we are unlikely to guess of which species will be the randomly chosen individual. However, if community has many species, but only one (or few) prevails (many individuals of one or few species), uncertainty will not be so high, since we have high probability that randomly selected individual will be the most abundant species. This is why Shannon index increases with richness and evenness, and it puts more weight on the

Shannon index

$$H_{max} = \log S \;[1]$$
$$H = - \sum_{i=1} p_i \log p_i$$

where
S = species richness,
p_i = relative abundance of species *i*,
log = usually natural logarithm (i.e. log_e or *ln*)

richness than on evenness.

In real ecological data, values of H are usually between 1.5-3.5 (the units are *bits* of information); note that absolute value of the H depends on the base of the logarithm used for the calculation (usually $\log_e$, where e = 2.718). The maximum value of H index (H_{max}) for community of given richness occurs at situation that it is perfectly even (all species have the same relative proportion).

Simpson index

(also Simpson concentration index, denoted as D, H_s or λ) is also considering both richness and evenness, but compared to Shannon it is more influenced by evenness than richness. It represents the probability that two randomly selected individuals will be of the same species. Since this probability decreases with increasing species richness, the Simpson index also decreases with richness, which is not too intuitive. For that reason, more meaningful is to use **Gini-Simpson** index, which is simply 1-Simpson index, and which with increasing richness increases (it is identical with Hurlbert's *probability of interspecific encounter*, PIE).

The values of D are in the range between 0 and 1 and the units is a probability. When the species richness of community exceeds 10, the values of Simpson index are mostly influenced by evenness.

Simpson index

$$D = \sum_{i=1}^{S} p_i^2 \qquad D_{max} = \frac{1}{S} \ ^{3)}$$

Gini-Simpson index

$$GS = 1\text{-}D$$

where
S = species richness,
p_i = relative abundance of species i,

Comparison of species richness, Shannon index and Simpson index

In case of perfectly even communities, the Shannon and Gini-Simpson index increases non-linearly with number of species in the community; Gini-Simpson index increases faster. This relationship also illustrates that Gini-Simpson index changes very fast in low species richness values (0.5 for S = 2, 0.67 for S = 3, 0.75 for S = 4, ... 0.9 for S = 10), and with richness over 10 it changes much slower (0.95 for S = 20 and 0.99 for S = 100).

Dependence of Shannon and Gini-Simpson index on number of species in perfectly even community
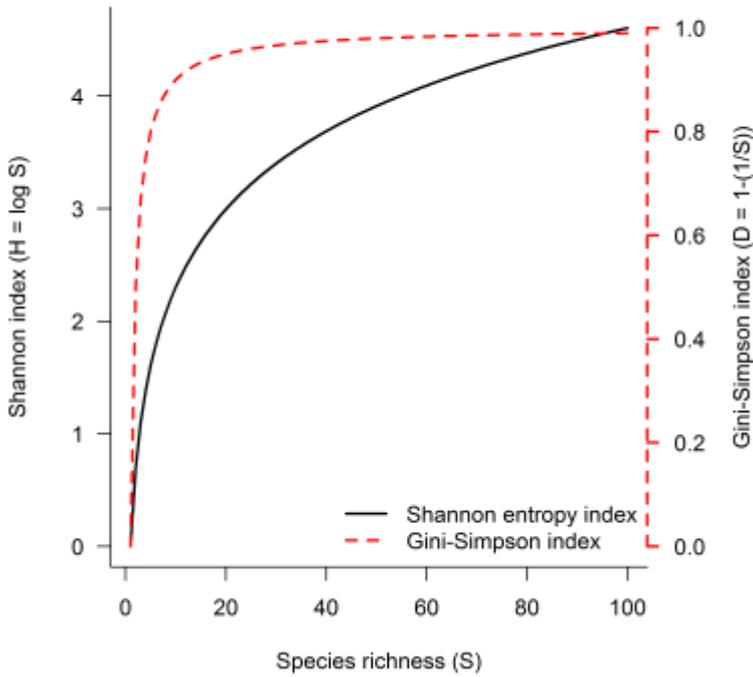

Dependence of the three diversity indices (richness, Shannon and Simpson) on the (un)evenness and diversity of the community is illustrated below.
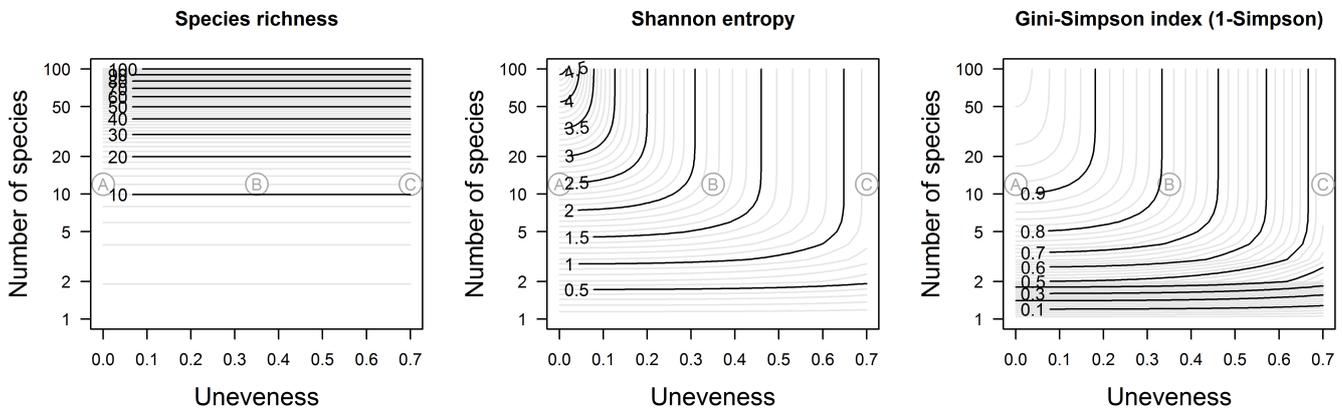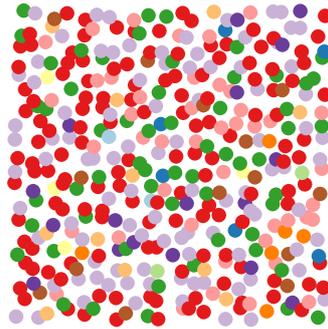

The A, B and C labels within the figures note three example communities, each with 12 species (species richness = 12). Community A is perfectly even, B is moderately uneven, and C is highly uneven. See the species abundance distribution barplots below.
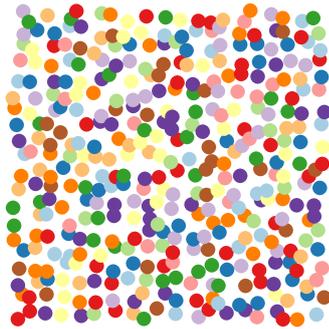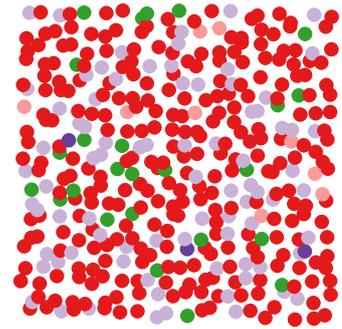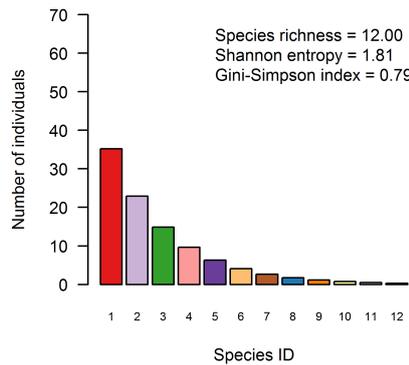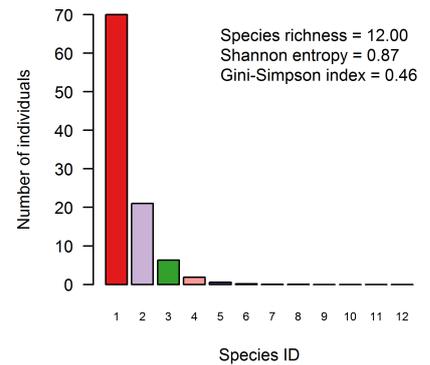
Evenness

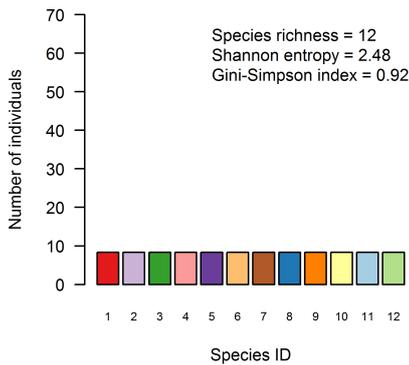is a synthetic measure describing pattern of relative species abundances in a community. There are many ways how evenness can be calculated, here I mention just two common ones, one derived from Shannon and the other from Simpson index.

Shannon's evenness

(called also Pielou's *J*) is calculated as a ratio of Shannon index calculated from real community (with S species and p_1, p_{2i}, p_3....p_i relative species abundances), and maximum Shannon index for the community with the same richness (i.e. with S species all having $p_1 = p_2 = p_i = 1/S$). The value is 1 in case that all species have the same relative abundances.

Shannon's evenness

$$J = \frac{H}{H_{max}} = \frac{H}{\log S}$$

Simpson's evenness

(called also *equitability*) is calculated from Simpson's effective number of species divided by observed number of species. Effective number of species (ENS) is the number of equally abundant species which would need to be in a community so as it has the same Simpson's index as the one really calculated (more about the concept of *effective number of species* below). In case of Simpson's D, effective number of

Simpson's evenness

$$equitability = \frac{\frac{1}{D}}{S} = \frac{1}{DS}$$

species is simply 1/D.

Effective numbers of species (ENS)

Lou Jost (2002) argued that to call Shannon and Simpson (or Ginni-Simpson, respectively) indices as *diversity* is misleading, since diversity should be measured in intuitive units of *species*, while each of the two indices have different units (Shannon *bits* and Simpson *probability*)[4]. This problem can be overcome by introducing concept of `effective number of species` (ENS, MacArthur 1965), i.e. number of species in equivalent community (i.e. the one which has the same value of diversity index as the community in question) composed of equally-abundant species. In cace of perfectly even community, ENS is equal to species richness; for unevenn communities, ENS is always smaller than S. Each of the indices above can be converted into effective number of species following a simple formulas.
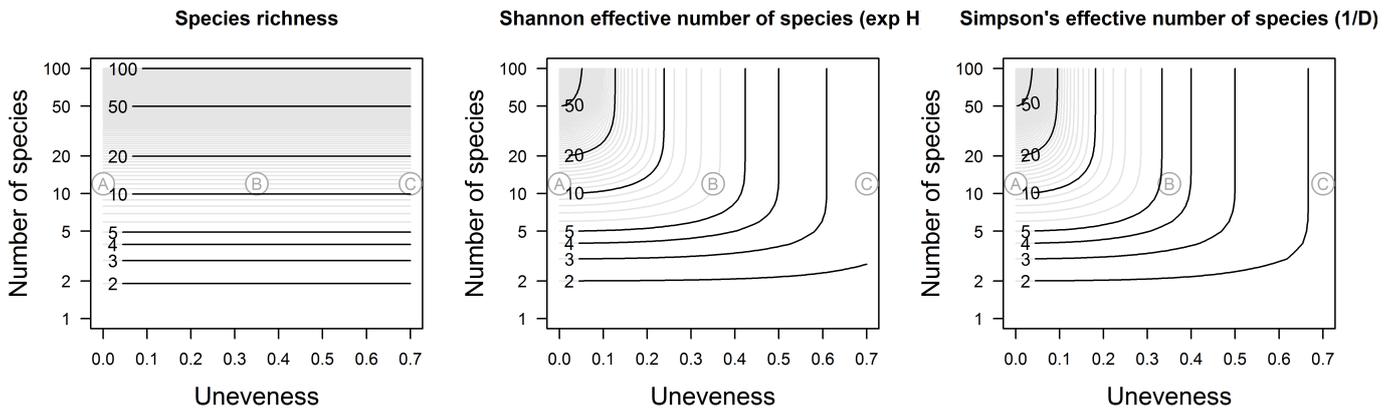
Effective number of species

- for species richness = S
- for Shannon index = e^H (exponential of Shannon entropy index)
- for Simpson index = 1/D (reciprocal of Simpson concentration index)

Hill numbers

Mark Hill (British scientist, known also for introducing Detrended correspondence analysis (DCA), Twinspan, and recallibrating Ellenberg species indicator values for Britain) realized that species richness, Shannon entropy and Simpson's concentration index are all members of the same family of diversity indices, later called as Hill numbers. Individual Hill numbers differ by the parameter *q*, which quantifies how much the measure discounts rare species when calculating diversity. Hill number for *q* = 0 is simply species richness, for *q* = 1[5] it is **Shannon diversity**, i.e. effective number of species derived from Shannon entropy, and for *q* = 2 it is **Simpson diversity**, i.e. ENS for Simpson concentration index. For *q* > 0, indices discount rare species, while for *q* < 0 the indices discount common species and focus on number of rare species (usually not meaningful).

Hill numbers

For q = 0, 1 and 2 (also noted as N_0, N_1 and N_2):

$$^qD = \left(\sum_{i=1}^{S} p_i^q\right)^{\frac{1}{1-q}}$$
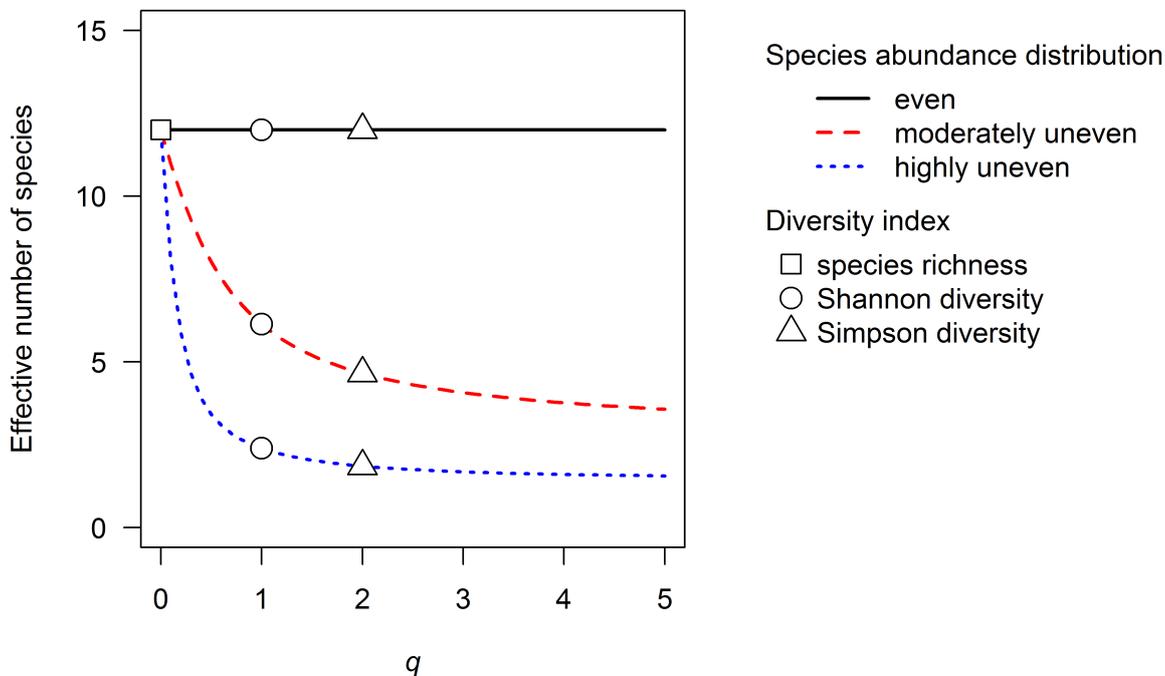
$^0D = S$ (species richness)
$^1D = e^H$ (exponential of Shannon entropy)
$^2D = \frac{1}{D}$ (reciprocal of Simpson index)

Dependence of species richness, Shannon diversity (effective number of species based on Shannon entropy index) and Simpson's diversity (effective number of species based on Simpson's index) on (un)evenness and diversity is illustrated below.

Species richness   **Shannon effective number of species (exp H**   **Simpson's effective number of species (1/D)**

Diversity profiles

It is possible to draw the effective number of species as a function of coefficient q - increasing q decreases the impact of rare species on the measure of diversity. The value for $q = 0$ equals to species richness (in the diagram displayed by squares), for $q = 1$ equals to Shannon diversity (circles) and for $q = 2$ Simpson diversity (triangles). The shape of the diversity profile considers the differences in evenness between the three communities; the more is the community species abundance uneven, the faster the curve declines with increasing coefficient q. The future will see what exactly can this form of diversity visualization bring new.

Diversity profiles for three communities

Species abundance distribution
— even
- - - moderately uneven
······ highly uneven

Diversity index
□ species richness
○ Shannon diversity
△ Simpson diversity

Summary of values for diversity measures discussed in this chapter

	Community A (perfectly even)	Community B (moderately uneven)	Community C (highly uneven)

	Community A (perfectly even)	Community B (moderately uneven)	Community C (highly uneven)
Species richness	12	12	12
Shannon entropy	2.48	1.81	0.87
Simpson index	0.92	0.79	0.46
Shannon evenness	1	0.73	0.35
Simpson evenness	1	0.38	0.15
Shannon diversity (^1D, N1)	12	6.14	2.39
Simpson diversity (^2D, N2)	12	4.66	1.86

[1]

Why log (S)? $H = -\sum p_i \log p_i$. In case of S equally abundant species, each $p_i = 1/S$. Then, $H = -\sum 1/S \log 1/S = - S * 1/S * \log (1/S) = - \log (1/S) = - \log S^{-1} = - (- \log S) = \log S$.

[2]

Entropy of the system represents the uncertainty, expected measure of surprise.

[3]

Why 1/S? $D = \sum p_i^2$. In case of S equally abundant species, each $p_i = 1/S$. Then $D = \sum 1/S^2 = S*1/S^2 = S/S^2 = 1/S$.

[4]

Jost (2002) argues: "The radius of a sphere is an index of its volume but is not itself the volume, and using the radius in place of the volume in engineering equations will give dangerously misleading results. This is what biologists have done with diversity indices. The most common diversity measure, the Shannon-Wiener index, is an entropy, giving the uncertainty in the outcome of a sampling process…. Entropies are reasonable indices of diversity, but this is no reason to claim that entropy *is* diversity."

[5]

In fact the Hill's formula is not defined for $q = 1$, but it can be shown that when q approaches 1 from below or above, the index gets equal to exponential Shannon.