

# Diversity analysis

## Theory

In general, diversity is a measure quantifying number of different states in a system. In case of ecological communities these states are usually species, but could be also genera, families, OTU's or functional types. Many important ecological theories predict number of species in a community (e.g. island biogeography, species-area relationship, productivity-diversity relationship, metacommunity dynamics etc.). Diversity is considered an “emergent property” of a community<sup>1)</sup>, acting at the community level, but not at the level of individual species. Diversity is also an important measure used in conservation management, as an indicator of “well-being” of ecological systems.

Diversity has two components: **species richness** (number of species in a community), and **evenness** (or shape of species abundance distribution (SAD) – the fact that some species are common and other are rare).

To understand why differences in abundances between species matter for the diversity, let's take a walk through two forests (example adapted from Gotelli & Chao 2013). Both communities have the same species richness of 20 different tree species. Note that here, richness refers to the number of species in the whole community, and we are surveying the community by sampling a limited number of individuals (it is unlikely that we would be able to survey the whole community, i.e. all individuals in it). The first forest (community A) is perfectly even, i.e. each species is represented by the same number of individuals. The second forest (community B) is highly uneven, i.e. one or a few) species are dominant and the others are rare (in this case, the first species represents 81% of all individuals in the community, and each of 19 other species represents 1% each). We take two walks through each forest, and within each walk, we inspect 20 trees.

## Community A – 20 species, abundances even

Sample #1 – 20 individuals, 15 species observed, 5 unseen



Sample #2 – 20 individuals, 13 species observed, 7 unseen



## Community B – 20 species, abundances highly uneven

Sample #1 – 20 individuals, 3 species observed, 17 unseen



Sample #2 – 20 individuals, 4 species observed, 16 unseen

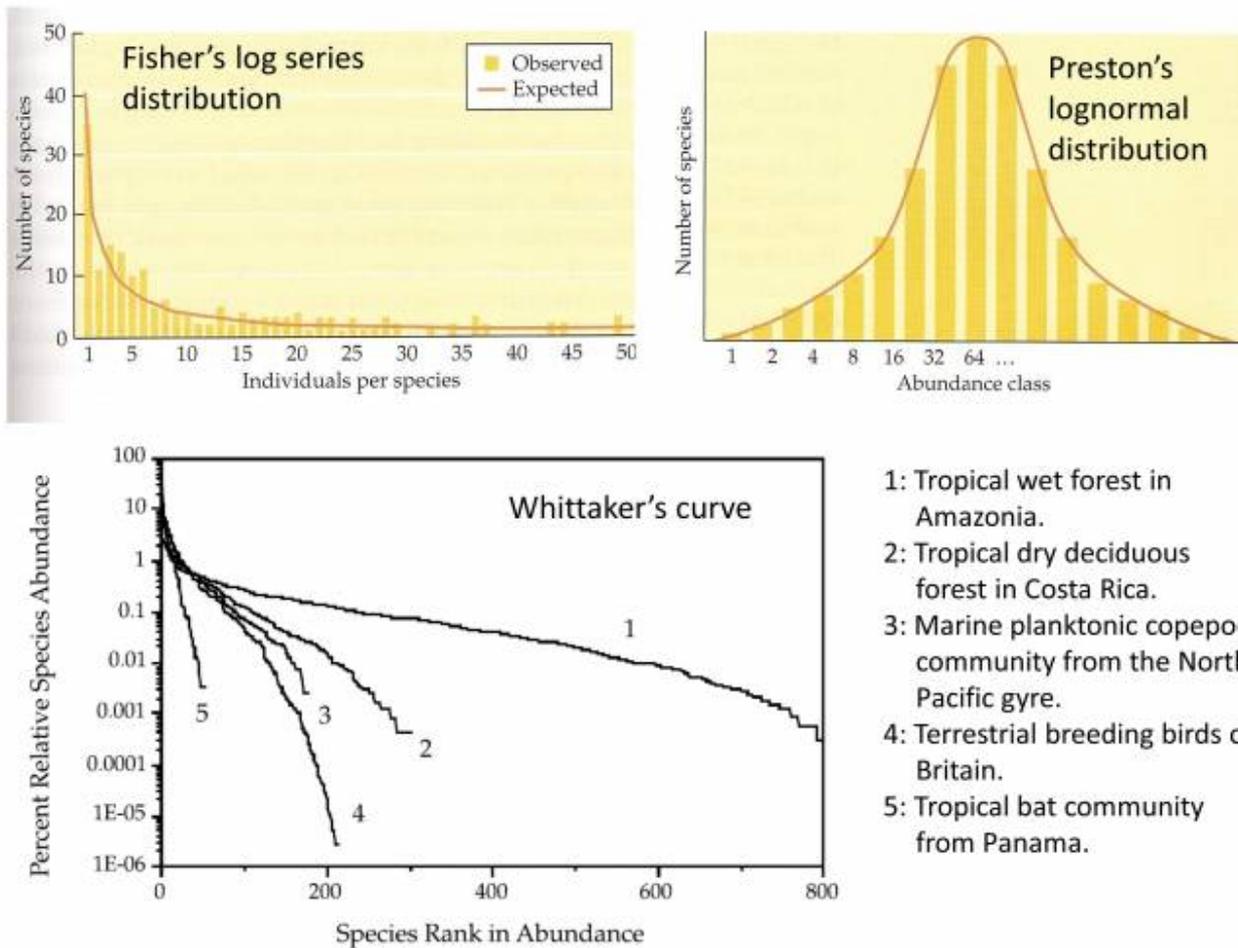


Adapted from Gotelli & Chao (2013)

As you can see from the figure above, in the even community A we have a high probability that each new individual will be a new species, while in the highly uneven community B we keep meeting still the same species, while rarely observing the other. In the forest A, during the first walk we observed 15 species, which means that 5 species remain undetected, and during the second walk, we observed 13 species (7 undetected). In forest B, the first walk brings 3 species (17 undetected), and the second walk brings 4 species (16 undetected). The feeling of diversity differs - community A feels much more diverse since we keep meeting different tree species. Functionally, community A and community B are also rather different - in community A, interacting individuals are mostly of different species, so interspecific competition prevails, while in community B interacting individuals are mostly of the same species, so intraspecific competition is more common.

The shape of species abundance distribution (SAD) has been for a rather long time considered as an important indicator of underlying community assembly processes. R.A. Fisher was perhaps the first to plot the SAD plot in which x-axis represents number of individuals per species and y-axis represents number of species, and fitted the data by log series, which was that time considered empirically as the most common shape of species abundance distribution and predicted that singletons (species with only one detected individuals) are always the most common. Fisher’s alpha, one of the oldest diversity measures considering both species richness and evenness, is derived from this empirical relationship. Preston modified this diagram and showed that if community is sampled more completely and the x-axis is transformed into octaves (numbers of species are binned into groups of 1, 2, 4, 8, 16, 32 etc. species), the resulting shape of SAD resembles symmetric bell shape of Gaussian distribution (more intensive sampling will make singletons less common or completely absent, since it’s a matter of time to find another one or more individuals for each species original represented by singletons). Robert H. Whittaker introduced “rank abundance curve”, called also diversity-dominance or Whittaker’s curve, where the x-axis represents species ranked according to their relative abundance (from commonest at the left to rarest at the right), and the y-axis represents

relative species abundances (often log-transformed). The shape (the steepness and the length of the tail) indicates the relative proportion of dominant and rare species in a community.

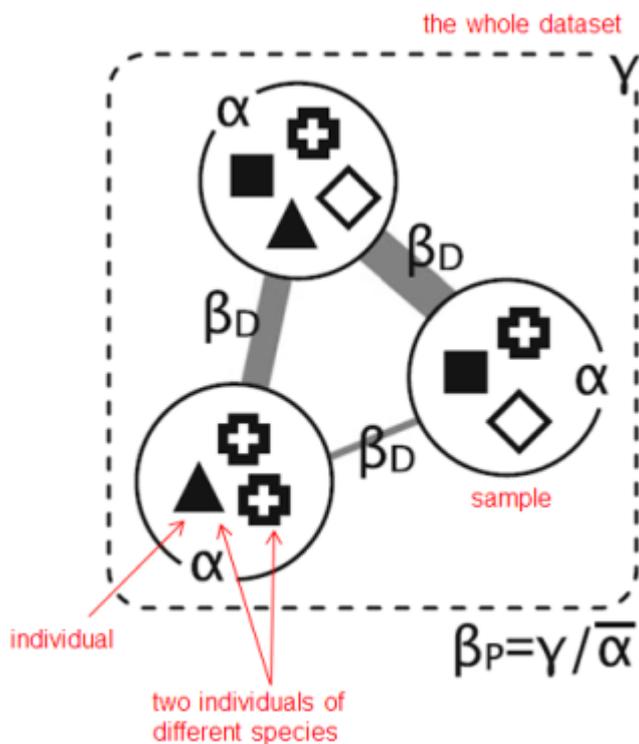


The Fisher's and Preston's species abundance distribution plot comes from Mittelbach (2012) and is based on the famous dataset of Rothamsted moths. The figure with Whittaker's curve comes from Hubbel (2001) and individual curves are based on data from communities listed beside (1 to 5).

There are several diversity indices differing by the degree in which they consider richness and evenness (species richness, Shannon entropy and Simpson concentration index in this order putting the weight on evenness from *no* in case of species richness to *high* in case of Simpson), and also several indices of evenness itself. Mark O. Hill (1973) showed that all three diversity indices can be summarized using so-called **Hill numbers** of order  $q$ , which represent effective numbers of species (increasing  $q$  puts less weight on rare species and more weight on abundant species). Hill numbers can be used to draw diversity profiles, which allow for an elegant comparison of diversity among communities considering both richness and evenness.

Since Earth is finite, each community has theoretically a countable number of species and their evenness. However, these theoretical numbers are usually not readily available, since we **estimate diversity of a community by sampling it**, and sampling is always incomplete. Diversity estimated from sampled data is dependent on sampling effort, and if diversity (alpha, beta, gamma) should be compared among different communities, the sampling effort should be standardized. This can be done using **rarefaction curves**, which allow for comparison of diversity based on the same number of individuals or the same number of samples. Eventually, we can go further and standardize our samples on the same community completeness (Chao & Jost 2012). Another way is to use some of the **diversity estimators** to estimate the number of unseen species which have not been detected by

sampling but are expected to be observed if the sampling effort increases. Two families of diversity estimators are available, the first based on abundance data (abundance of species in one sample quantified e.g. by number of individuals or amount of biomass) and the second on incidence data (frequency of species presences in the set of samples, i.e. in each sample only species incidence, i.e. presence-absence, is recorded).



**The schema of alpha, beta and gamma diversity** from [Jurasinsky et al. \(2009\)](#).  $\alpha$  – alpha diversity (number of species in the sample).  $\gamma$  – gamma diversity (number of species in the dataset, community or region).  $\beta_D$  – differentiation diversity, the similarity between pairs of samples (calculated e.g. by Jaccard similarity index); the wider the grey bar, the higher similarity (and lower dissimilarity).  $\beta_P$  – proportional diversity, the relationship between two scales of investigations considering species number on alpha and gamma level.

There are several concepts which aim to specify different flavours of diversity. Perhaps the oldest is Whittaker's concept of **alpha, beta and gamma diversity** ([Whittaker 1960](#)). Whittaker built upon Fisher's alpha and extended the concept of local species richness (alpha diversity) for regional species richness (gamma diversity) and change in species composition among samples (beta diversity). Among complementary approaches is the one introduced by [Jurasinsky et al. \(2009\)](#), identifying **inventory diversity** (alpha and gamma diversity, differing by scale at which it is applied), and **differentiation vs proportional diversity** (both being beta diversities – differentiation one calculated by dissimilarity indices or as a variation of species composition matrix, while proportional as the ratio between gamma and alpha diversity).

**Beta diversity** is a concept fundamentally different from alpha or gamma diversity, and itself represents a complex topic. Beta diversity can be seen as **species turnover** (directional exchange of species among pair of samples or along spatial, temporal or environmental gradient) or as **variation in species composition** (non-directional description of heterogeneity in species composition within the dataset)(e.g. [Anderson et al. 2011](#)). Alternatively (sensu [Jurasinsky et al. 2009](#)), beta diversity can be seen as either differential diversity (considering differences in species composition) or as proportional diversity (proportion of species on a regional and local level, gamma vs alpha diversity).

Emergent property is a property which the collection or complex system has, but which the individual members do not have. Often such properties “emerge” because of interactions between individual entities of the system. Emergent property of a flight is the ability to fly; individual parts of disassembled flight does not have this ability.

From:

<https://www.davidzeleny.net/anadat-r/> - **Analysis of community ecology data in R**

Permanent link:

[https://www.davidzeleny.net/anadat-r/doku.php/en:diversity\\_analysis](https://www.davidzeleny.net/anadat-r/doku.php/en:diversity_analysis)

Last update: **2019/01/26 20:28**