Section: Diversity analysis

Diversity indices

Theory R functions **Examples**

Example 1: Diversity of two forest communities in Taiwan, sampled within 1-ha plots

In this example, we will use individual-based data from two forest communities, sampled within 1-ha non-permanent plots. The whole dataset, consisting of seven such communities, can be seen here. We will select two contrasting localities: Pi-Lu (PL), a high-elevation forest dominated by tsuga, and Yuan-Yang-Hu (YYH, Yuan-Yang Lake), a cloud forest vegetation dominated by large Taiwan yellow cypress. Let's first upload the data (a list) and subset the two localities to independent vectors:

```
hp.abund <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/hp.abund.txt')
PL <- hp.abund$PL  # Pilu - fewer individuals
YYH <- hp.abund$YYH  # Yuan-Yang-Hu - many individuals, diversity comparable to PL
```

Each vector contains the number of woody individuals for given species (their abundance). For some of the follow-up formulas, we will need the relative abundances p_{ij}, so let's prepare them:

```
p_PL <- PL/sum (PL)
p_YYH <- YYH/sum (YYH)
```

Let's first calculate the basic indices of diversity by hand: species richness, Shannon entropy index and Simpson's concentration index (and related Ginni-Simpson index):

```
# Richness
S_PL <- sum (p_PL > 0) # 29
S_YYH <- sum (p_YYH > 0) # 34

# Shannon
H_PL <- -sum (p_PL*log (p_PL), na.rm = TRUE) # 2.76
H_YYH <- -sum (p_YYH*log (p_YYH), na.rm = TRUE) # 2.45

# Simpson
D_PL <- sum (p_PL^2) # 0.083
D_YYH <- sum (p_YYH^2) #0.138

# and Ginni-Simpson
GS_PL <- 1 - D_PL  # 0.917
GS_YYH <- 1 - D_YYH  # 0.862
```

As you can see, Pi-Lu has lower species richness than Yuan-Yang-Hu (29 vs 34 species), but when the other diversity indices are considered, YYH actually has lower diversity than PL. This becomes more obvious when we translate the Shannon and Simpson entropies into diversities, using the concept of the effective number of species (check theory, section Hill numbers, for details how the recalculation works).

```
# Shannon diversity (effective number of species based on Shannon's entropy index)
H_ENS_PL <- exp (H_PL)  # 15.7
H_ENS_YYH <- exp (H_YYH) # 11.7
```

```
# Simpson diversity (effective number of species based on Simpson's concentration index)
D_ENS_PL <- 1/D_PL  # 12.0
D_ENS_YYH <- 1/D_YYH  # 7.26
```

There are alternative ways how to calculate the diversity indices, as you can see in R functions tab of this section. Worth pointing out is the use of the package called `vegetarian` (I guess a joke name, related to the fact that the most popular package for ordination analysis is called `vegan`). The package needs to be installed manually from CRAN archive since the current version is not available on CRAN.

There are two main functions in the `vegetarian` package, function H, which calculates the original diversity indices (species richness, Shannon's entropy, Simpson's concentration index or Ginni-Simpson alternative), and function d, which calculates the effective number of species for the same indices. Here is how it works (for example, using the PL dataset with the original abundance values):

```
library (vegetarian)
H (PL, q = 0)  # 29
H (PL, q = 1)  # 2.75
H (PL, q = 2)  # 0.083
H (PL, q = 2, gini = TRUE)  # 0.880

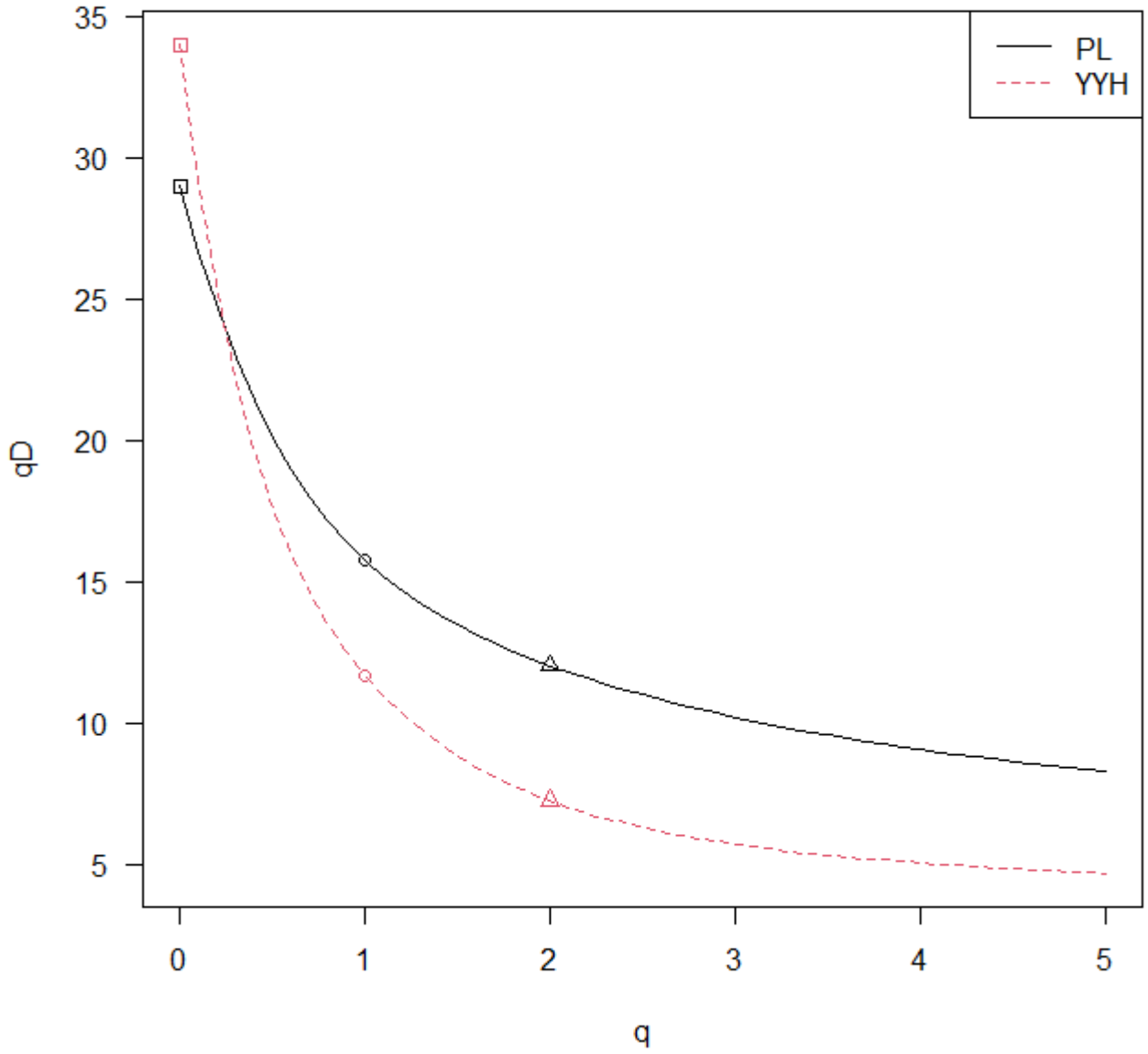
d (PL, q = 0)  # 29
d (PL, q = 1)  # 15.7
d (PL, q = 2)  # 12.0
```

The function d in the `vegetarian` package has the benefit that it can calculate an effective number of species for any value of q, following the Hill number equation. We can use this property to construct the "diversity profile" for the two communities by calculating diversity values for the q parameter in the sequence from zero to, e.g. five (in the following example, I define q by function `seq (from = 0, to = 5, by = 0.1)`:

```
dp_PL <- sapply (seq (0, 5, by = .1), FUN = function (q) d (PL, q = q))
dp_YYH <- sapply (seq (0, 5, by = .1), FUN = function (q) d (YYH, q = q))
```

The curve with diversity profiles can be plotted using the function `matplot` (I suggest you study `?matplot` to grasp the basics of this less used function):

```
matplot (y = data.frame (PL = dp_PL, YYH = dp_YYH), x =  seq (0, 5, by = .1), type = 'l', xlab = 'q', ylab = 'qD', las = 1)
points (x = c(0,1,2), y = c(d(PL, q = 0), d(PL, q = 1), d (PL, q = 2)), pch = 0:2, col = 1)
points (x = c(0,1,2), y = c(d(YYH, q = 0), d(YYH, q = 1), d (YYH, q = 2)), pch = 0:2, col = 2)
legend ('topright', col = 1:2, lty = 1:2, legend = c('PL', 'YYH'))
```



On the diversity profile above, I highlighted the diversity values for the three main indices, species richness (square), Shannon diversity (circle) and Simpson diversity (triangle). You can see that there is a flip in the leading position between the communities; YYH is more diverse when the richness of all species is considered, but when the focus is shifted on common or only dominant species (i.e. Shannon or Simpson diversity), PL becomes more diverse. I bet that the main reason is the difference in species abundance distribution, i.e. the fact that PL has more even abundances than YYH. Let's calculate evenness for both communities to verify this.

Evenness for Shannon's entropy is calculated as the ration between real observed Shannon entropy value for the community, divided by the maximum possible Shannon entropy value for community with the same number of equally abundant species:

```
H_PL_max <- log (sum (PL > 0))
H_YYH_max <- log (sum (YYH > 0))

Pi_PL <- H_PL/H_PL_max  # 0.819
Pi_YYH <- H_YYH/H_YYH_max  # 0.697
```

Evenness for Simspon's concentration index is calculated as the ratio between the effective number of species

for a given Simpson's index, divided by the number of all species in the community (I simply use species richness in the formula below):

```
Eq_PL <- D_ENS_PL / S_PL  # 0.415
Eq_YYH <- D_ENS_YYH / S_YYH  # 0.213
```

In both cases, the evenness for PL is higher than for YYH, even if PL has lower species richness (lower richness may be caused by higher elevation of PL compared to YYH).

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

Permanent link:
https://anadat-r.davidzeleny.net/doku.php/en:div-ind_examples

Last update: **2026/05/14 07:15**