Section: Diversity analysis

Rarefaction

Theory R functions **Examples**

Example 1: Comparing forest diversity of two plots standardized to number of individuals and sample completeness

We will use vegetation data from One-hectar plots in different forest types across Taiwan, containing the survey of woody species at seven localities sampled in different elevations in Taiwan (we already used it in Example 1 within the Diversity indices section). At each locality, a 1-ha plot has been established, and within the plot, 25 10×10-m subplots have been systematically sampled (in total, 25×10×10 m = 0.25-ha area was surveyed within each locality). We will use abundance data (`hp.abund`) containing the number of individuals for each species at each locality, and select two plots, one with high diversity but a low number of individuals (Pi-Lu) and the other with lower diversity but a much higher number of individuals (Yuan-Yang-Hu).

Our aim will be to compare the diversities of forest vegetation in these two plots when standardized to the same number of individuals or the same sample completeness (coverage).

First, upload the abundance dataset and subset data from the two localities, PL and YYH:

```
hp.abund <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/hp.abund.txt')
PY <- hp.abund[, c('PL', 'YYH')]  # we are selecting columns from the data.frame,
that's why the comma
```

Data are stored in `data.frame`, with species in rows and localities (1-ha plots) in columns; the cells are filled by the number of individuals for each species.

For all calculations in this exercise, we will use the library `iNEXT`, developed by the team of Prof. Anne Chao (Tsing-Hua University, Hsinchu, Taiwan). You may need to install the package from CRAN first if you haven't used it before:

```
# install.packages ('iNEXT')
library (iNEXT)
```

The package `iNEXT` offers function `DataInfo` for data overview:

```
DataInfo (x = PY, datatype = 'abundance')
```

```
  Assemblage    n S.obs     SC f1 f2 f3 f4 f5 f6 f7 f8 f9 f10
1         PL  394    29 0.9823  7  4  3  0  1  0  1  0  0   0
2        YYH 1551    34 0.9916 13  1  0  0  0  0  0  1  0   1
```

The meaning of the variables in the output of `DataInfo` for `datatype = "abundance"` is the following: `Assemblage` is the abbreviation of the locality; `n` is the number of individuals in each plot; `S.obs` is the observed number of species in each locality; SC is calculated sample completeness (coverage); `f1` is the number of singletons (species occurring only in one plot), `f2` is the number of doubletons (occurring only in 2 plots), etc.

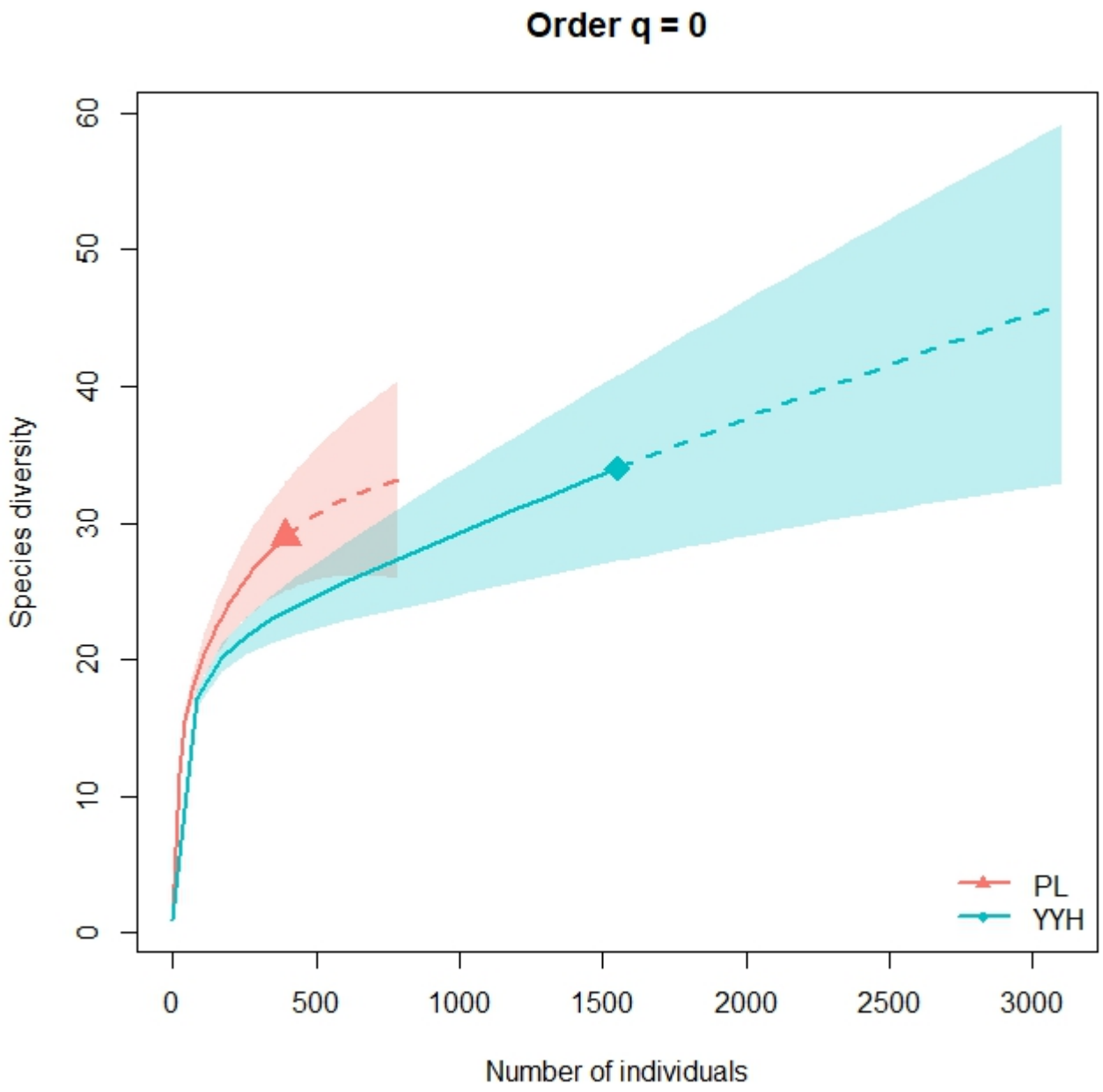
Let's first see the rarefaction curves of species richness for both communities. These can be calculated by the function `iNEXT`, with parameter `q = 0` (only species richness) and `datatype = 'abundance'` (abundance data):

```
PY_raref <- iNEXT (x = PY, q = 0, datatype = 'abundance')
PY_raref
```

As you can see, object `PY_raref` is a list with quite verbal output, containing data overview (`$DataInfo`), estimates of diversity based on number of individuals (the `$iNextEst$size_based`) or coverage (`$iNextEst$coverage_based`), and also asymptotic estimates of diversity (`$AsyEst`).

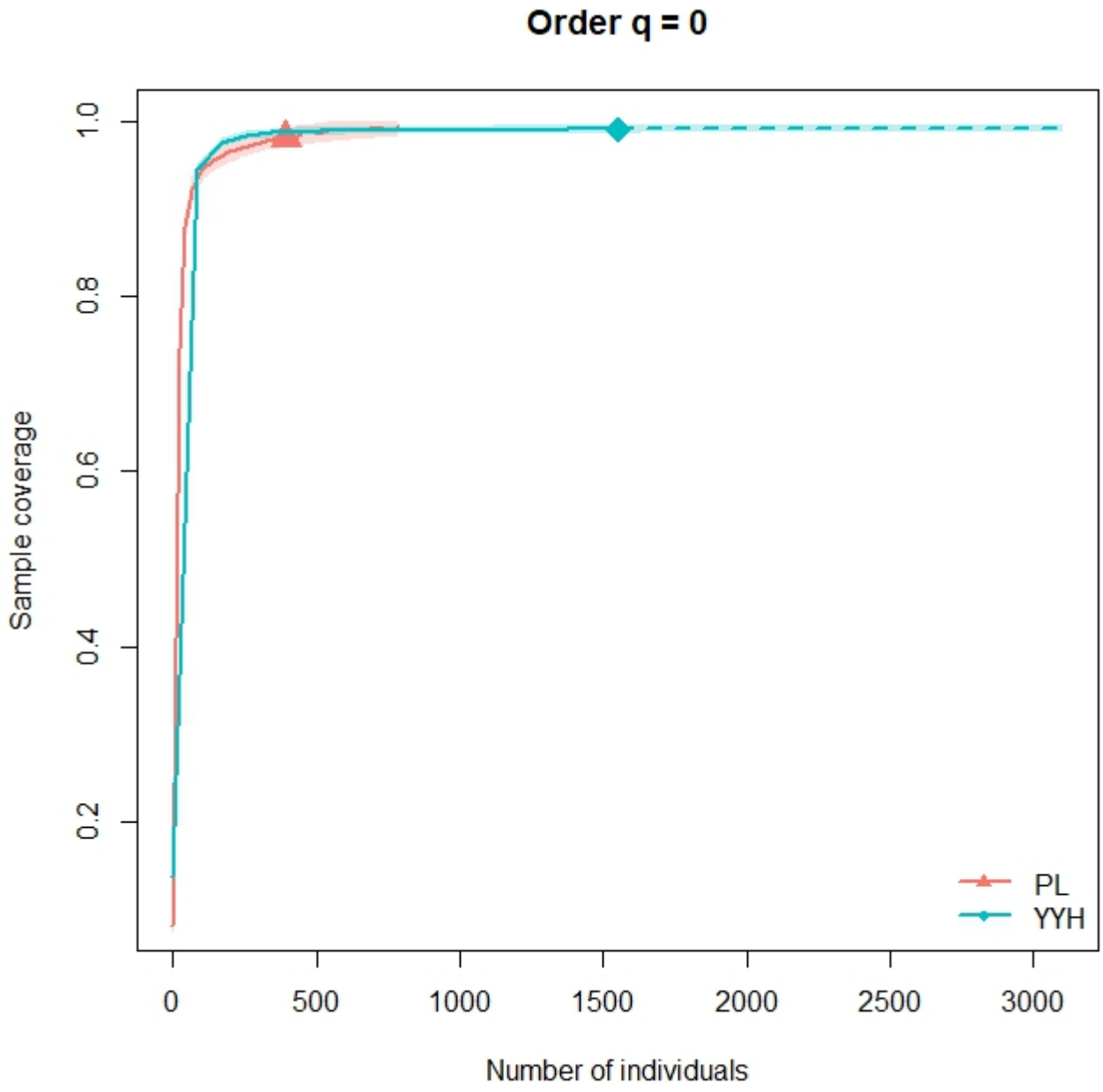
Let's plot rarefaction curves for both communities, using the function `plot.iNEXT`:

```
plot (PY_raref)
```



The curve contains both rarefied (solid) and extrapolated (dashed) parts, with a symbol showing the size and diversity of each locality, and each locality has a different colour. In this plot, the x-axis is “Number of individuals”. An alternative option is to use the rarefaction curve to plot sample coverage over the number of individuals:

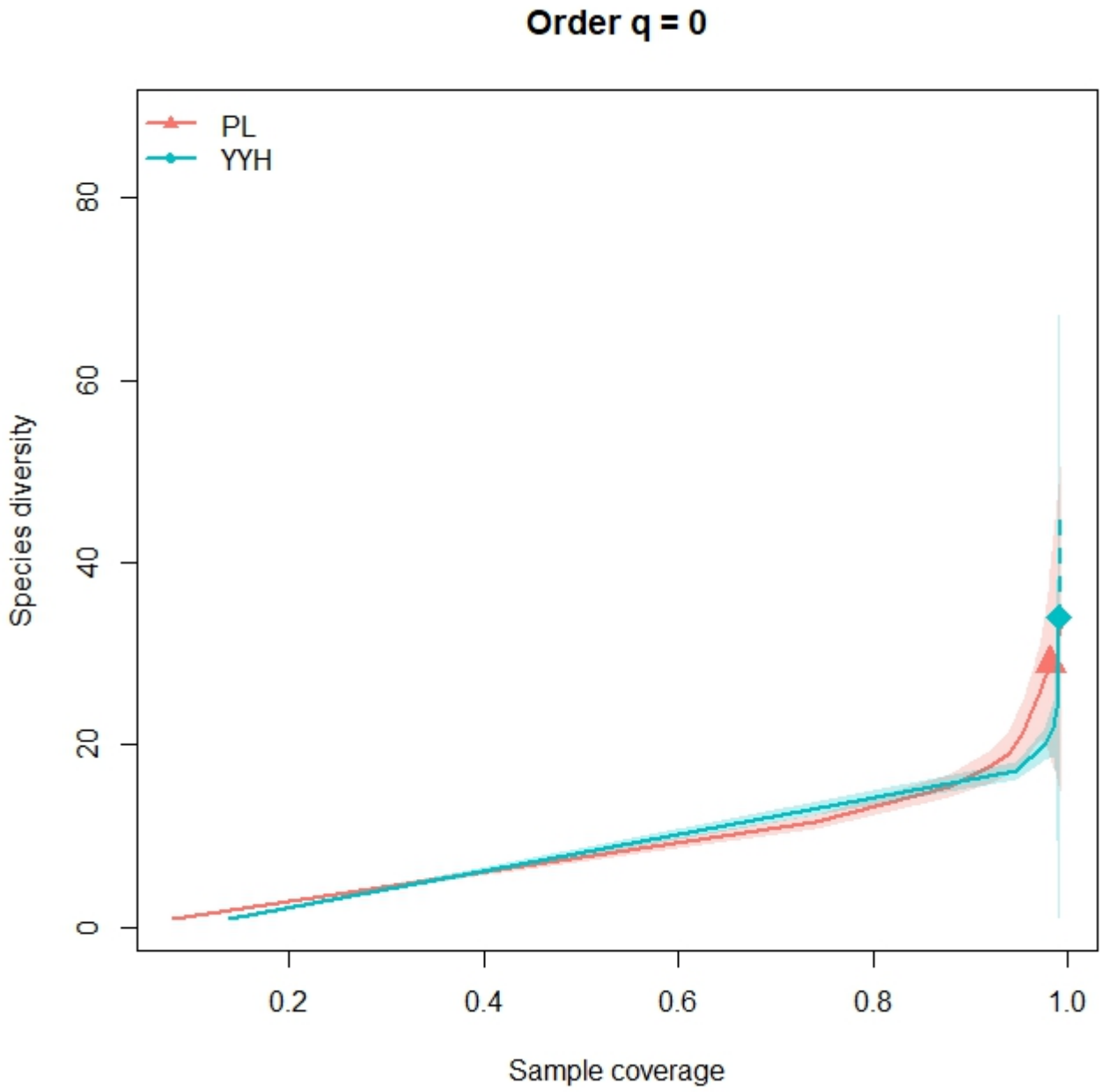
```
plot (PY_raref, type = 2)
```



In this figure, it is obvious that both localities have rather complete species surveys, with coverages close to one. The argument `type` can be either 1 (diversity vs number of individuals, default value), 2 (coverage vs number of individuals), or 3 (diversity vs coverage).

The last option is to draw rarefaction curves of diversity against sample coverage:

```
plot (PY_raref, type = 3)
```



Here you can see that if standardized to the same coverage, Pi-Lu (PL) is, in reality, species-richer than Yuan-Yang-Hu (YYH). The standardized coverage is the maximum non-extrapolated coverage for Pi-Lu (0.9823).

Now, let's calculate richness for each community, standardized in two different ways: first by the number of individuals, and second by sample coverage. In both cases, we choose the lowest value from available data, i.e., 394 individuals for standardization to the number of individuals, and 0.9823 for standardization to coverage. We will use the function `estimateD` for this calculation. The function has additional arguments, which allow us to specify the level of diversity we want to estimate. Argument `base` quantifies whether we want to calculate an estimate based on the number of individuals (`size`) or sample coverage (`coverage`). The argument `level` specifies the target numbers for each estimate (394 or 0.9823, respectively). If `level` is left blank, the default option is what the function judges as optimal (check `?estimateD` for help).

First, estimates of diversity for a fixed number of individuals:

```
PY_ind <- estimateD (x = PY, q = 0, datatype = 'abundance', base = 'size', level =
394)
PY_ind
```

```
  Assemblage   m      Method Order.q        SC       qD   qD.LCL   qD.UCL
1         PL 394    Observed       0 0.9822850 29.00000 25.12918 32.87082
2        YYH 394 Rarefaction       0 0.9886819 23.51271 21.78464 25.24078
```

Output shows the estimated data for a given number of individuals (`m`), which part of the rarefaction curve was used for that (`Method`), and the estimated diversity (`qD`). It also reports sample coverage for a given number of individuals (`SC`) and confidence interval for diversity estimates (lower one, `qD.LCL`, and upper one, `qD.UCL`). Estimated richness is 29.0 species for Pi-Lu and 23.5 species for Yuan-Yang-Hu.

Second, estimates under fixed coverage:

```
PY_compl <- estimateD (x = PY, q = 0, datatype = 'abundance', base = 'coverage',
level = 0.9823)
PY_compl
```

```
  Assemblage        m        Method Order.q     SC       qD   qD.LCL   qD.UCL
1         PL 394.2914 Extrapolation       0 0.9823 29.00517 17.77778 40.23255
2        YYH 224.3293   Rarefaction       0 0.9823 21.18177 18.73626 23.62728
```

Here, the numbers of individuals (`m`) are different (Yan-Yang-Hu has only 224 individuals), but the sample coverage (`SC`) is fixed. Richness for Pi-Lu is 29.0, while for Yuan-Yang-Hu is 21.2 species.

In the original dataset, Pi-Lu has 29 species and 394 individuals, while Yuan-Yang-Hu has 34 species and 1551 individuals; Yuan-Yang-Hu is richer, but also has many more individuals, which is perhaps why it appears to have more species. After standardization to the same number of individuals or the same sample coverage, Pi-Lu became richer than Yuan-Yang-Hu.

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

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

Last update: **2026/05/24 17:16**