Contents

Introduction

This script reports the code to reproduce the analyses and figures presented in the second lecture of the Conservation Biology class at Stanford University (Bio 144, 2020).

Software preparation

Install `R` software (https://www.r-project.org) if you do not have it yet Then, install and load the following packages

```
install.packages('vegan')
install.packages('bipartite')
```

```
library(vegan)
library(bipartite)
```

Data import

Download data at the following link:

https://github.com/losapio/bio144_c2/blob/master/maindata.csv

Move them to your working folder, set it up as your working directory (e.g., `mydir`)

```
setwd("/mydir")
```

and import the .csv file into `R`

```r
maindata <- read.csv("maindata.csv", sep=",", head=T)
```

We familiarize with the dataset and see how it looks like

```r
str(maindata)
```

```
## 'data.frame':    330 obs. of  3 variables:
##  $ Transect: int  10 3 4 6 6 8 8 7 6 5 ...
##  $ Plant   : int  16 7 2 7 8 7 10 6 26 10 ...
##  $ Species : Factor w/ 80 levels "Acalypha diversifolia",..: 75 1 1 1 1 1 1 2 3 4 ...
```

```r
head(maindata)
```

```
##   Transect Plant               Species
## 1       10    16            Sapranthus
## 2        3     7 Acalypha diversifolia
## 3        4     2 Acalypha diversifolia
## 4        6     7 Acalypha diversifolia
## 5        6     8 Acalypha diversifolia
## 6        8     7 Acalypha diversifolia
```

It is a `dataframe` composed of 330 observations (rows) and three variables (columns): (1) Transect is the transect ID, from one to ten; (2) Plant is the the number of individuals of each plant species in each transect; (3) Species is the taxonomic name of plants.

Data preparation

To produce the Species Accumulation Curve (SAC), we will make use of the existing function `specaccum` in `vegan` package. Take first a look at what this function does and requires.

```r
?specaccum
```

This function needs data organized into a matrix with species in columns (s), transects in rows (t), and species abundance (N) as entries

$$\begin{bmatrix} N_{1,1} & \cdots & N_{1,s} \\ \vdots & \ddots & \vdots \\ N_{t,1} & \cdots & N_{t,s} \end{bmatrix} \tag{1}$$

We then convert our dataframe into such a matrix.

```r
maindata$dummy <- 1
maindata$Transect <- as.character(maindata$Transect)
data_matrix <- frame2webs(maindata[,c(1,3,4)],
              c('Transect','Species','dummy'))[[1]]
# head(data_matrix)
```

Species Accumulation Curve

We now draw our SAC taking the whole dataset. We do so by running 10,000 permutations (all possible permutations of 10 transects are $n = 10!/(10-10)! = 3628800$, which is computationally intense) and extrapolating richness values by means of bootstrap.

Species-based SAC

In the first case, SAC is based on adding transects (in `vegan` also called 'sites') in a random order and finding the mean SAC and its standard deviation from random permutations of the data (Gotelli & Colwell 2001).

```
sac <- specaccum(data_matrix,
                 method = 'random', permutations = 10000, gamma = 'boot')
sac
```

```
## Species Accumulation Curve
## Accumulation method: random, with 10000 permutations
## Call: specaccum(comm = data_matrix, method = "random", permutations = 10000,      gam
##
##
## Sites      1.00000  2.00000  3.00000  4.00000  5.00000  6.00000  7.0000  8.00000
## Richness 15.65030 26.81300 36.16710 44.18420 51.32240 57.78900 63.6986 69.25330
## sd         7.06106  8.79259  9.55097  9.73488  9.54198  9.14503  8.3568  7.14419
##
## Sites      9.00000 10
## Richness 74.69260 80
## sd         5.30735  0
```

```
plot(sac, ci.type='polygon', ci.col='green', las = 1,
     xlab = "Transect", ylab = "Species")
```
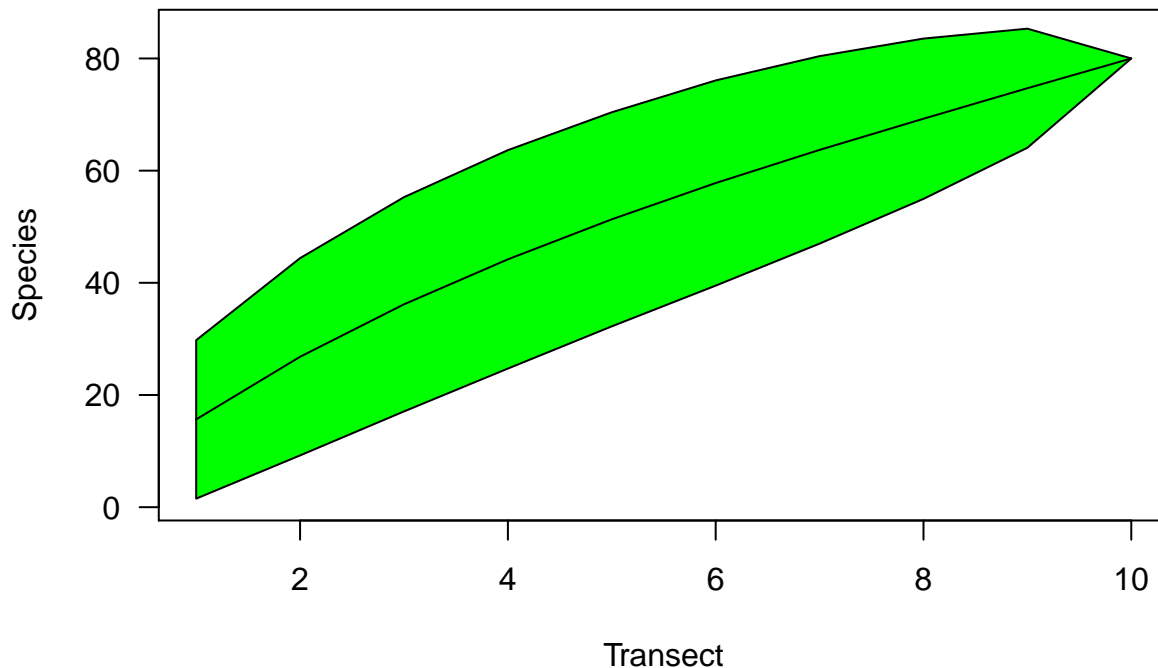
From the shape of the curve we can deduce our goodness of sampling. Since the curve is still far from tending to an asymptote we need more transects to have a better estimation of species richness. Indeed, there is a certain number of unseen, missing species.

Individual-based SAC

In the second case, we will draw a SAC according to a rarefaction method that is based on sampling individuals (i.e., single plants) rather than transects (Hurlbert 1971).

```
sac <- specaccum(data_matrix,
                 method = 'rarefaction', permutations = 10000, gamma = 'boot')
sac
```

```
## Species Accumulation Curve
## Accumulation method: rarefaction
## Call: specaccum(comm = data_matrix, method = "rarefaction", permutations = 10000,
##
##
## Sites            1.0000  2.0000  3.0000    4.0000    5.0000    6.0000    7.000   8.0000
## Individuals     33.0000 66.0000 99.0000 132.0000 165.0000 198.0000 231.000 264.0000
## Richness        17.8053 29.4306 38.6249  46.3796  53.1888  59.3384  65.004  70.2952
## sd               2.3845  3.0002  3.2845   3.3941   3.3753   3.2400   2.981   2.5673
##
## Sites            9.0000  10
## Individuals    297.0000 330
## Richness        75.2801  80
## sd               1.9051   0
```

```
plot(sac, ci.type='polygon', ci.col='green', xvar = "individuals",
     las = 1, xlab = "Individuals", ylab = "Species")
```
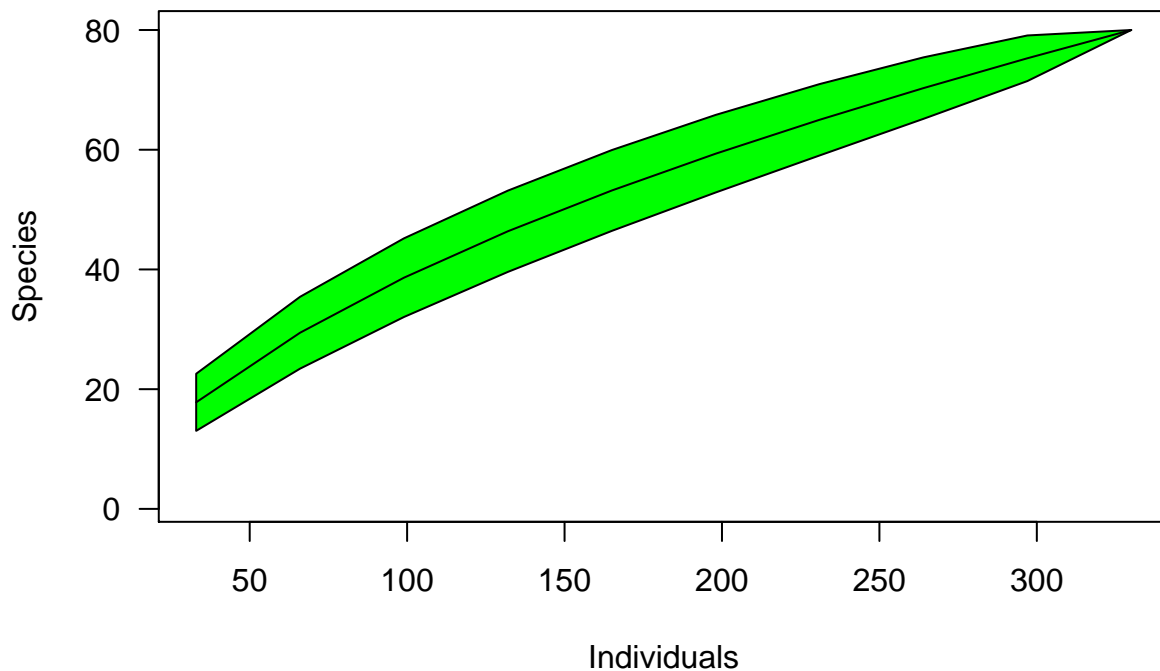

We obtain a similar curve to the previous one. We can see how the number of observed species increases by increasing the number of transects but it does not saturate, indicating that species richness should continue to increase with sampling more than 330 plants.

SAC-based Predictions

Unseen species

We estimate the number of missing, unseen species by extrapolating species richness using the function `specpool`.

```
?specpool
```

```
specpool(data_matrix)[,7:8]
```

```
##          boot   boot.se
## All 99.17644 10.48784
```

We estimate the 'real community' to be composed of 99 ± 10 species, which corresponds to an increase c 25% of species with respect to the original pool (i.e., 80 species). This means that we did not observe c 20 species with our 10 transects.

Notice that if we sample less than 10 transects, the estimated species number is reduced too.

```
poolaccum(data_matrix)$means[,c(2,6)]
```

```
##          S Bootstrap
```

```
## [1,] 36.00  44.48630
## [2,] 43.62  54.05977
## [3,] 50.72  62.84470
## [4,] 57.73  71.50459
## [5,] 63.49  78.60651
## [6,] 69.25  85.74540
## [7,] 74.45  92.22855
## [8,] 80.00  99.17644
```

With 3 transects we would have estimated 46 species (out of 37 observed), with 4 57, and so on, until 99 estimated species out of 80 observed over 10 transects.

Additional sampling

Therefore, how many species can we observe by increasing the number of transects? How many new transects are needed to sample the estimated 99 species?

```r
sac <- specaccum(data_matrix, method = 'exact')
sac.fit <- fitspecaccum(sac, 'lomolino')
sac.pred <- predict(sac.fit, 11:20)
names(sac.pred) <- paste('transect',11:20)
sac.pred
```

```
## transect 11 transect 12 transect 13 transect 14 transect 15 transect 16
##    84.26879    88.56517    92.62555    96.47302   100.12729   103.60535
## transect 17 transect 18 transect 19 transect 20
##   106.92201   110.09020   113.12137   116.02565
```

```r
plot(c(sac.fit$richness,sac.pred), type = 'l', lwd = 2, las = 1,
     col = 'green', xlab = 'Transect', ylab = 'Species')
abline(h = 100, lty = 2, col = 'blue')
```
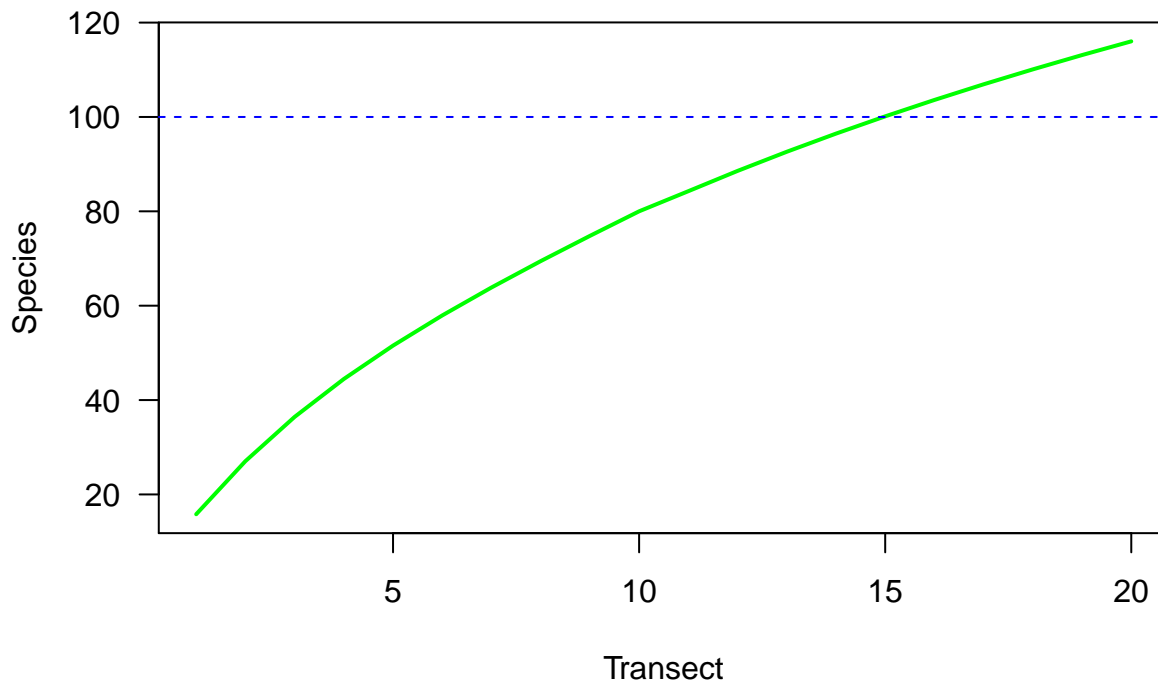
By sampling one more transect we observe 84 species, 89 with two more transects, 93 with three, 96 with four, and 100 with five. Thus, with fifteen total transects we may have a quite good sample of our forest.

Still, we notice that the curve does not saturate yet. Indeed, the derivate of the SAC at 20 transects is still much higher than zero.

```r
specslope(sac.fit, 20)
```

```
## [1] 2.843932
```

We estimate that the curve will start saturating (i.e., increasing rate smaller than one) with 48 additional transects.

```r
specslope(sac.fit, 21:60)
```

```
##  [1] 2.7302940 2.6243435 2.5253289 2.4325955 2.3455701 2.2637486 2.1866853
##  [8] 2.1139849 2.0452949 1.9803004 1.9187187 1.8602954 1.8048010 1.7520277
## [15] 1.7017868 1.6539068 1.6082312 1.5646170 1.5229334 1.4830602 1.4448870
## [22] 1.4083120 1.3732415 1.3395887 1.3072735 1.2762214 1.2463634 1.2176355
## [29] 1.1899778 1.1633348 1.1376546 1.1128886 1.0889916 1.0659211 1.0436373
## [36] 1.0221030 1.0012831 0.9811446 0.9616566 0.9427899
```

```r
sac.pred <- predict(sac.fit, 11:58)
plot(c(sac.fit$richness,sac.pred), type = 'l', lwd = 2, las = 1,
     col = 'red', xlab = 'Transect', ylab = 'Species')
```
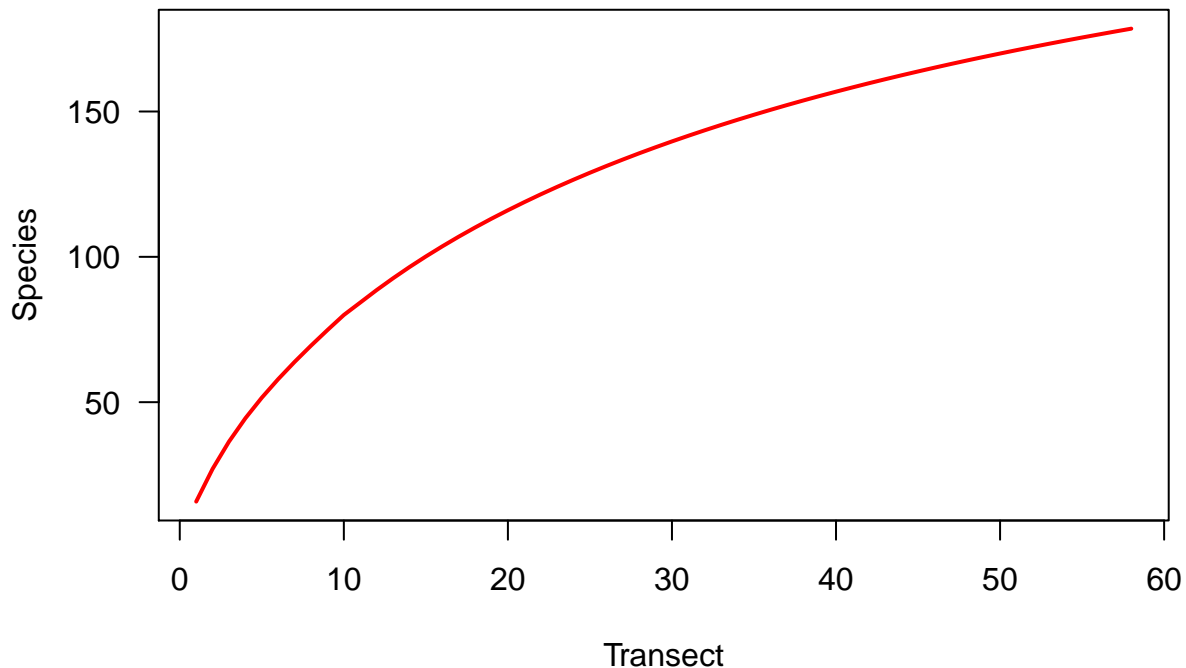
```
sac.pred[48]
```

```
## [1] 178.4879
```

We notice how the number of newly discovered species decreases with further increasing the number of transects. With a total of 58 transects we estimate to observe 178 species.

Rank–Abundance Curve

We finally examine the distribution of species abundances by means of rank–abundance curvers (RAD), also called dominance–diversity curves or Whittaker plots (Whittaker 1965, Wilson 1991). These plots show the logarithmic species abundances as a function of species rank order, that is in decreasing order from the most abundant to the least abundant species. RAD represents an important diagnostic tool for inspecting patterns of abundance distributions and diversity in communities.

We make use of the function `radfit`, which fits the following models to our abundance data: brokenstick (i.e., null model, Hurlbert 1971; Pielou 1975), niche preemption (Koleff et al. 2003), log-normal (McCune 1987), Zipf (O'Hara 2005), Zipf-Mandelbrot (Petchey & Gaston 2002). For more details, consult the help page of the function (type in R `?radfit`), see Oksanen 2019 and references therein.

Before fitting the RAD, we pool species abundance data over the ten transects. This way, we obtain abundance data for each species at the forest site.

```
rad_data <- colSums(data_matrix)
rad_data
```

```
##        Acalypha diversifolia       Aegephila costaricensis
```

```
##                              6                             1
##      Allophylus campostachys        Amphitecna tuxtlensis
##                              1                             1
##                  Apocynaceae 1                  Apocynacee 2
##                              1                             1
##                  Apoxynacaea 2          Arrabidea verrucosa
##                              1                             1
##          Astrocaryum mexicanum             Bactris mexicana
##                            116                             6
##              Bernoullia flammea         Brosimum alicastrum
##                              1                            10
##              Calatola laevigata              Capparis baduca
##                              1                             1
##           Chamaedorea alternans         Chamaedorea concolor
##                             17                             2
## Chamaedorea ernesti-augustii        Chamaedorea oblongata
##                              1                            10
##        Chamaedorea pinnatifrons              Clarisa biflora
##                              9                             1
##                  Cojoba arborea                Combretum sp.
##                              1                             2
##                Cordia alliodora             Cordia megalantha
##                              1                             1
##                    Costus scaber            Croton Schediarius
##                              1                             2
##              Croton schiedianus        Cymbopetalum baillonii
##                              1                             2
##            Dendropanax arboreus             Dialium guianense
##                              4                             2
##                Diospiros digyna              Eugenia mexicana
##                              2                             1
##            Faramea occidentalis                   Ficus aurea
##                              2                             1
##                Flacocurtace ?       Fornsteronia viridescens
##                              1                             1
##                   Guarea bijuga                 Guarea glabra
##                              2                             5
##                 Hampea nutricia                   Inga glabra
##                              1                             1
##          Macherium floribundum                    Makania ?
##                              1                             1
## Mortoniodendron guatemalense         Myriocarpa longipes
##                              1                             3
##               Nectandra ambigens        Nectandra dendrodaphne
##                              5                             1
```

9

```
##         Nectandra salicifolia        Ocotea dendrodaphne
##                          1                           1
##               Ouratea sp.                Piper sanctum
##                          1                           1
##          Pithecoctenium sp               Platimisium sp
##                          1                           1
##    Pleuratodendron lindenii          Posoqueria latifolia
##                          3                           1
##           Poulsenia armata          Pouteria campechiana
##                          6                           1
##          Pouteria durlandii           Pouteria reticulata
##                          1                           2
##             Pouteria sapota    Pseudolmedia oxyphyllaria
##                          2                          22
##        Psychotria chiapensis         Psychotria faxlucens
##                          8                           5
##            Psychotria flava         Psychotria limonensis
##                          4                           1
##       Psychotria papantlensis               Psychotria sp
##                          4                           1
##             Psychotria sp.      Psychotria veracruzensis
##                          1                           1
##          Pterocarpus rohrii            Quararibea funebris
##                          1                           3
##       Quararibea guatemalteca                   Randia sp
##                          1                           1
##             Rheedia edulis       Salacia megistophylla
##                          5                           5
##                 Sapranthus  Stemadenia donnell-smitthii
##                          1                           1
##          Tetracera volubilis          Trichilia martiana
##                          1                           2
##            Trophis mexicana             Trophis racemosa
##                          4                           2
```

```r
plot(radfit(rad_data), las = 1)
```
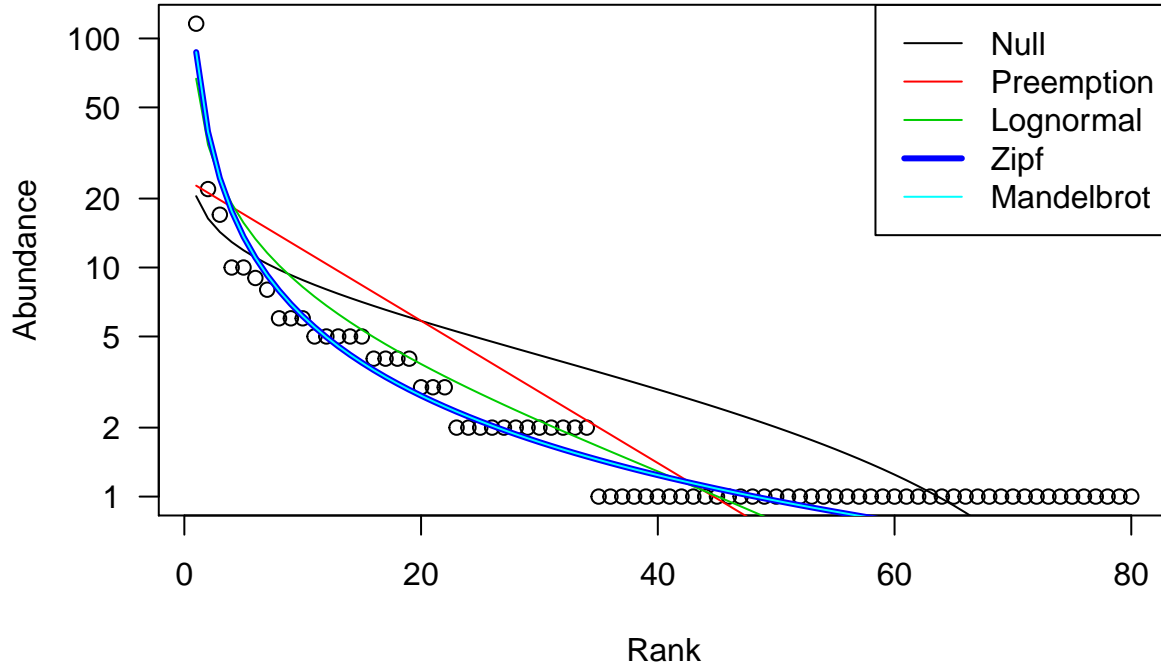
We can see that the null model fails to represent our data, indicating the distribution of individuals across species is not random. Indeed, we observe that few species have more individuals than expected by chance (very abundant, dominant species with low rank), while many species have less individuals than expected by chance (the majority of species with rank from 4–60). Furthermore, we can also see that the long tail of rare species with just one individuals is more pronounced that expected by chance.

The model that best represents our data is a Zipf/Zipf-Mandelbrot model, which follows a power-law distribution. This type of power-law distribution is a pattern widespread across ecological communities and many other systems, including e.g. the size distribution of river networks, the traffic volume across airports, and the occurrence of actors/actress in Hollywood movies.

References

Gotelli NJ & Colwell RK (2001). Quantifying biodiversity: procedures and pitfalls in measurement and comparison of species richness. Ecol. Lett. 4, 379–391.

Hurlbert SH (1971). The nonconcept of species diversity: a critique and alternative parameters. Ecology, 52, 577–586.

Koleff P, Gaston KJ & Lennon JJ (2003). Measuring beta diversity for presence-absence data. Journal of Animal Ecology, 72, 367–382.

McCune B (1987). Improving community ordination with the Beals smoothing function. Ecoscience, 1, 82–86.

O'Hara RB (2005). "Species richness estimators: how many species can dance on the head of a pin." Journal of Animal Ecology, 74, 375–386.

Oksanen J (2019). Vegan: ecological diversity. https://cran.r-project.org/web/packages/vegan/vignettes/diversity-vegan.pdf

Pielou EC (1975). Ecological Diversity. Wiley & Sons.

Petchey OL & Gaston KJ (2002). "Functional diver- sity (FD), species richness and community com- position." Ecology Letters, 5, 402–411.

Preston FW (1948). The commonness and rarity of species. Ecology 29, 254–283.

Whittaker RH (1965). Dominance and diversity in plant communities. Science 147, 250–260.

Wilson JB (1991). Methods for fitting dominance/diversity curves. Journal of Vegetation Science 2, 35–46.