Motivation

Neutral diversity

Estimation

Practical 1

Partitioning Diversity

Practical 2

Phylogenetic diversity

Practical 3

Functional Diversity

Conclusion

Diversity Metrics

Eric Marcon

October 8, 2018

Section 1

Motivation

The Questions

Which diversity metrics make more sense than mere indices?

Which of them may be less sensitive to metabarcoding issues?

Diversity of what?

Usually species (definition of E.O. Wilson).

May be any partitioned set.

Usually species in a clade (or a group), considered as a community: e.g. trees in a forest habitat.

Usually asymptotic diversity of a community that does not exist physically.

Course material

- Book: Marcon, E. (2017). Mesures de la Biodiversité. Kourou, France: UMR EcoFoG. https://hal-agroparistech.archives-ouvertes.fr/cel-01205813
- English version: https://github.com/EricMarcon/BDmeasurement
- R package entropart (Marcon and Hérault 2015b)

Motivation

Neutral diversity

Estimation

Practical 1

Partitioning Diversity

Practical 2

Phylogenetic diversity

Practical 3

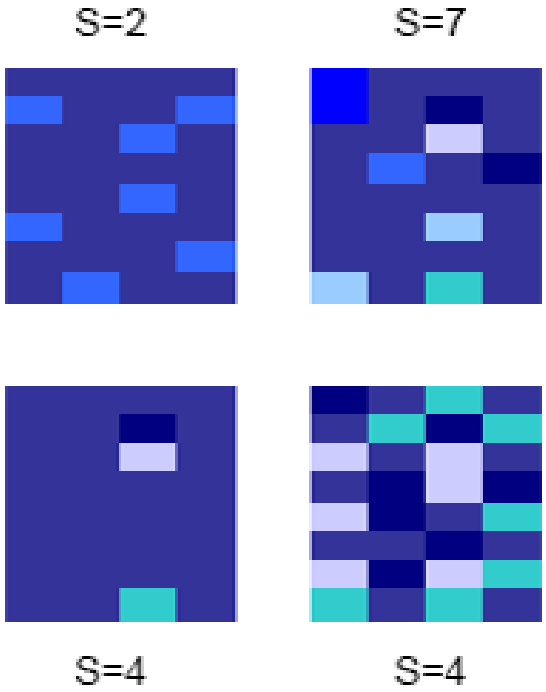
Functional Diversity

Conclusion

Section 2

Neutral diversity

Richness and Evenness



Which is the most diverse community?

Measuring Complexity

Define a character string:

- Length n
- Each letter has a probability

Example:

- 3 letters, {a, b, c}, probabilities (1/2, 1/3, 1/6)
- How many 60-character strings?
- The logarithm of the number of strings is n times entropy: 61

Shannon's entropy measures the *complexity* of the distribution of {a, b, c}, independently of n: 1.01

Entropy in Information Theory

An experiment with several outcomes:

- The probability to obtain r_s is p_s.

Information function: $I(p_s)$, between $I(0) = +\infty$ and $I(1) = 0$.

- Definition: rarity is $1/p_s$.
- The logarithm of rarity is Shannon's information function.

The expectation of the information carried by an individual is Shannon's entropy:

$$\sum_s p_s \ln \frac{1}{p_s}$$

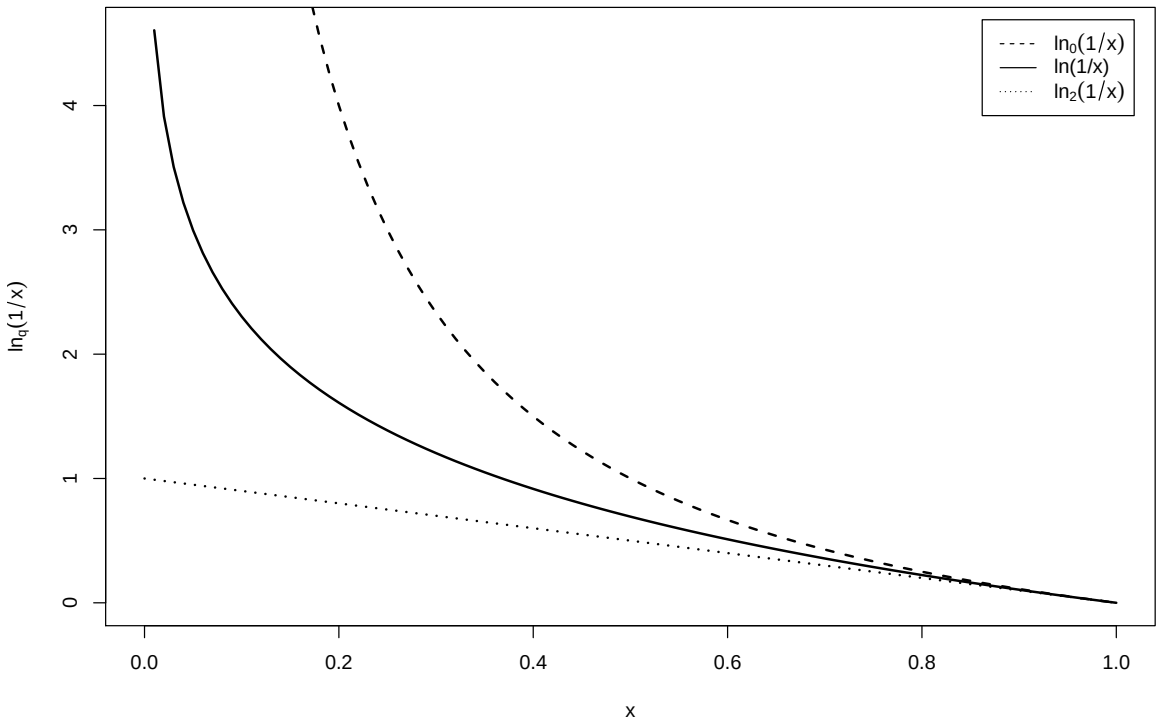
The average information is equivalent to complexity.

Generalized Entropy

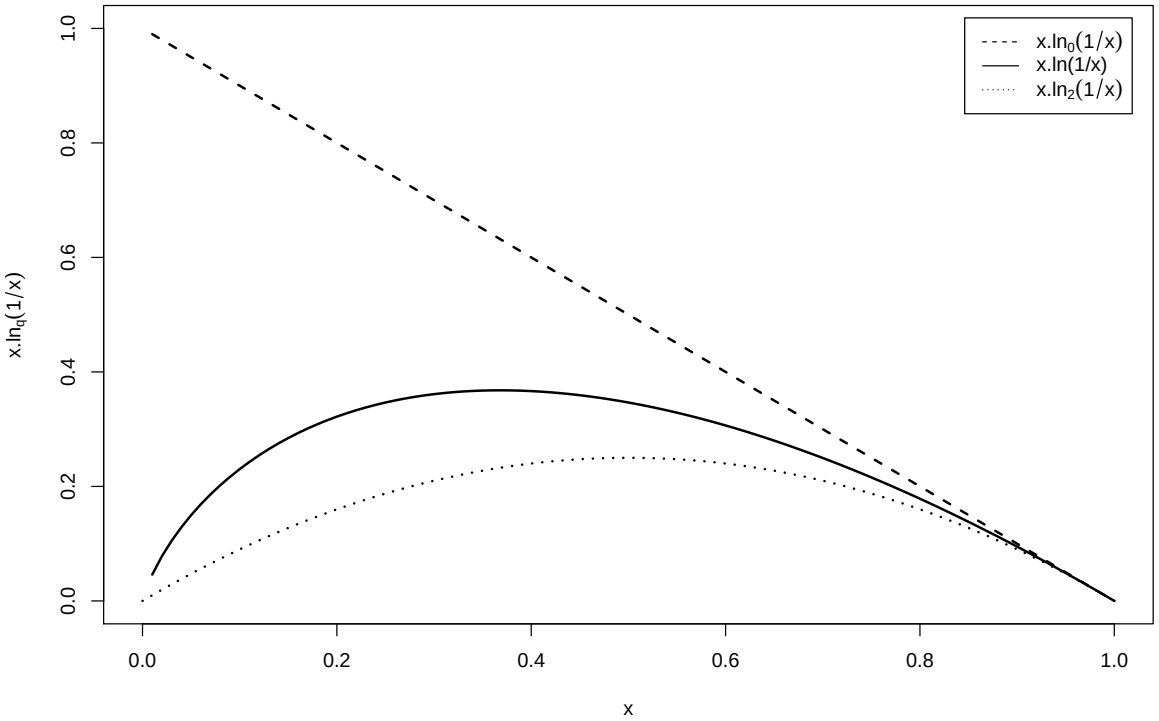
Other entropies: Rényi, Shorrocks... Tsallis (1988)

Parametric, to focus on rare or abundant species.

Deformed logarithm: $\ln_q x = \frac{x^{1-q}-1}{1-q}$



Tsallis (HCDT) Entropy 1/2



Contribution of a species to entropy of order $q = 0, 1, 2$.

Tsallis (HCDT) Entropy 2/2

Tsallis entropy is the average (deformed, of order q) logarithm of rarity (Tsallis 1994)

The order q stresses small or high probabilities.

- Entropy of order 0: the number of species -1.
- Entropy of order 1: Shannon.
- Entropy of order 2: Simpson.

Hill Numbers

The number of equiprobable outcomes that have the same entropy as the observed system (Hill 1973): *effective number of species*.

They are the (deformed, of order q) exponential of entropy (Marcon et al. 2014).

$$e_q^x = [1 + (1 - q)x]^{1/(1-q)}.$$

Diversity is noted ${}^qD(\mathbf{p_s})$.

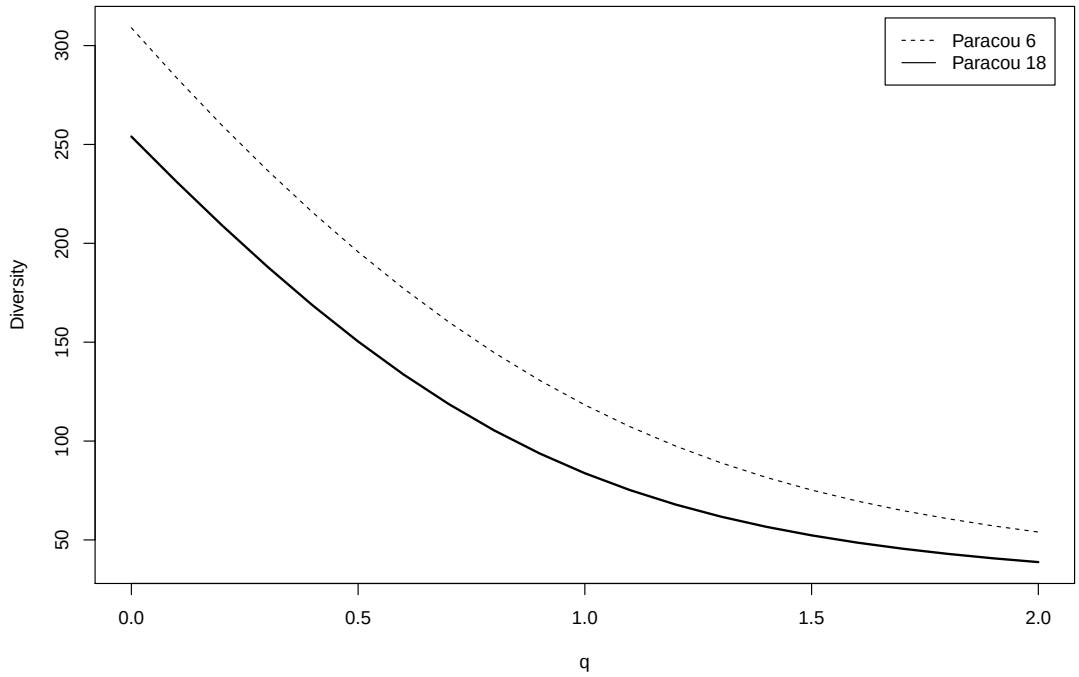
Eric Marcon

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Diversity Profiles

A diversity profile is ${}^qD \sim q$.

Compare two communities:



Summary

Entropy is the average log of rarity:

$$ {}^{q}H(\mathbf{p_s}) = \sum_s p_s \ln_q (1/p_s) $$

Diversity is its exponential:

$$ {}^{q}D(\mathbf{p_s}) = e_q^{{}^{q}H(\mathbf{p_s})} $$

Section 3

Estimation

Estimation bias

Rare species are difficult to sample

$\rightarrow$ diversity is generally underestimated.

Sampling effort is measured by n, the sample size.

The estimation bias decreases with n and q.

Simpson diversity is almost unbiased if $n/(n-1) \approx 1$.

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Estimation of richness

High bias but easiest correction methods.

Jackknife 1 estimator: just add the number of singletons S_1.

Correct if sample completeness $> 3/4$, i.e. singletons are less than 1 species out of 4.

Example: 225 species including 19 singletons $\rightarrow$ 244 species.

Estimation of diversity

Eric Marcon

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Many estimators of entropy, most based on sample coverage.
See section 4.6 of the book.

Entropy is estimated, then transformed to diversity.

Sample coverage is the probability of an individual of the community to belong to a sample species.

Far more than sample completeness.

Estimated by

$$\hat{C} = 1 - \frac{S_1}{n}.$$

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Practical estimation

All estimators available in *entropart::Diversity*.

Rule of thumb:

- Chao-Wang-Jost estimator is the best if singletons are less than 1 species out of 4.
- Else, the unveiled estimator chooses the appropriate Jackknife estimator for richness and is less biased but has higher variance.

Section 4

Practical 1

Data

Use Paracou forest tree inventory (2016).

```
library("EcoFoG")
```

```
## Loading required package: magrittr
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
Paracou2df("Plot=6 AND CensusYear=2016") %>% # Année 2016
  # Arbres vivants seulement
  filter(CodeAlive == TRUE) %>%
  # Filtrage des colonnes utiles
  select(Plot, SubPlot:Yfield, -Project, -Protocole, Family:Species, Cir
  # Création d'une colonne contenant "Genre espèce".
  unite(col = spName, Genus, Species, remove = FALSE) -> Paracou
```

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Richness

Number of species in plot 6.

```
library("entropart")
AbdP6 %>% Richness(Correction="None")
```

```
## None
##  763
```

Number of singletons in plot 6.

```
sum(AbdP6 == 1)
```

```
## [1] 337
```

Estimation.

```
AbdP6 %>% Richness(Correction="Jackknife")
```

```
## Jackknife 3
##        1372
```

Sample coverage

```
Coverage(AbdP6)
```

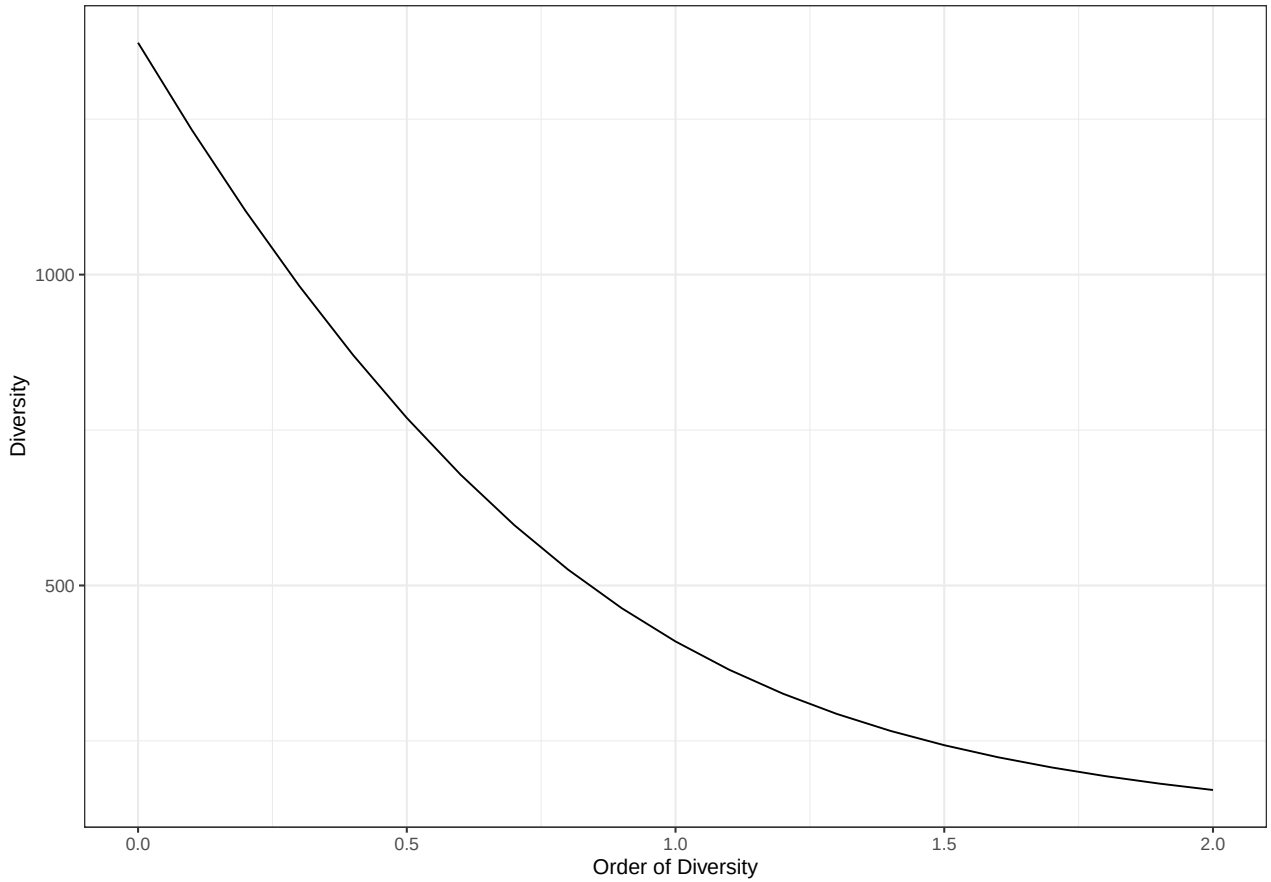
```
## ZhangHuang
##   0.904854
```

More than 1/3 of species are not observed but they contain less than 3% of the number of trees.

Diversity Profile

Unveiled-Jackknife estimator.

```
CommunityProfile(Diversity, AbdP6, Correction="UnveilJ") %>%
  autoplot
```



Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Section 5

Partitioning Diversity

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

α, β and γ diversity

Defined by Whittaker (1972):

- α: Average richness of a set of habitats, i.e. $\bar{S}$ species per habitat.
- γ: Richness of the assemblage, i.e. S species
- β: The ratio γ/α, i.e. a number of habitats.

Extensions to:

- All orders of diversity.
- Any embedded spatial scales.

The controversy

After Lande (1996): additive partitioning of diversity.

Resolution:

- β diversity = the ratio γ/α.
- β entropy = the difference $\gamma - \alpha$.

Exponentials multiply, logs sum...

Effective number of communities

α diversity at the community level is an effective number of species / community.

γ diversity at the assemblage ("meta-community") level is an effective number of species.

$\rightarrow \beta$ diversity is an effective number of communities.

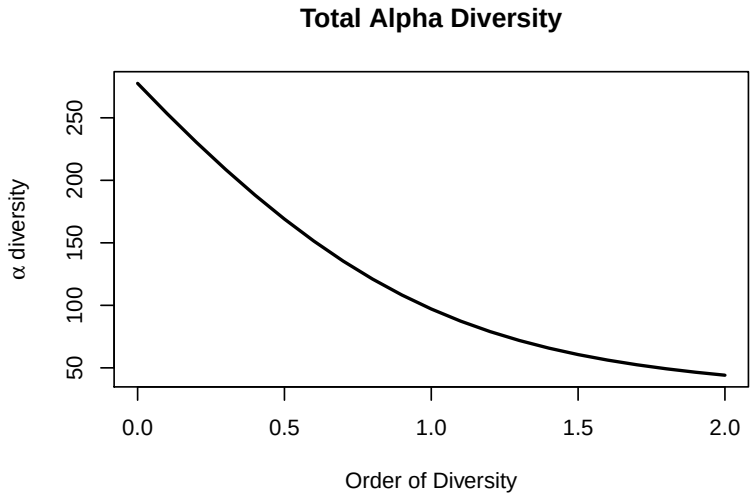
Effective number of communities

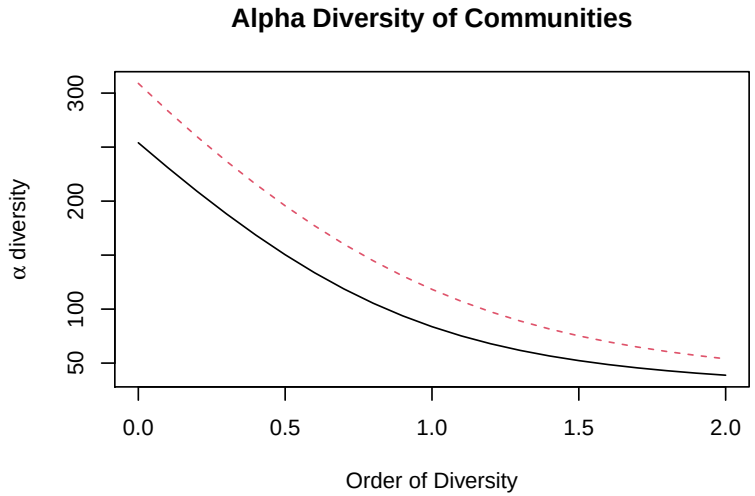

```
summary(DivPart(q = 1, MC = Paracou618.MC,
                Biased=FALSE, Correction="UnveilJ"))
```

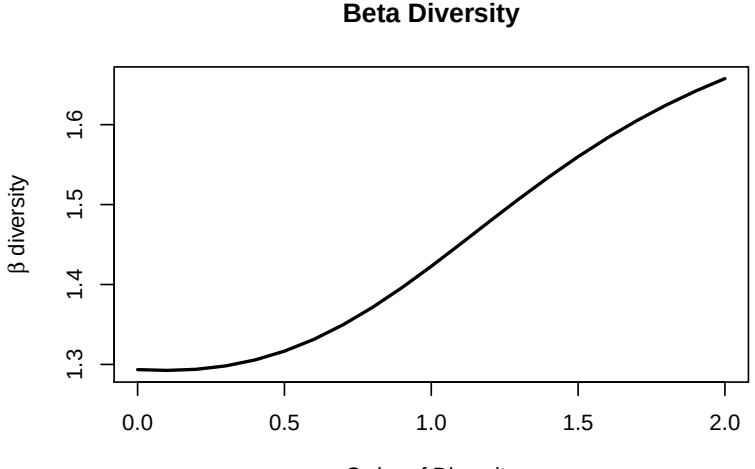
```
## HCDT diversity partitioning of order 1
## of metaCommunity Paracou618.MC
##  with correction: UnveilJ
## Alpha diversity of communities:
##     P006     P018
##  83.7268 118.2713
## Total alpha diversity of the communities:
## [1] 97.06467
## Beta diversity of the communities:
##  UnveilJ
## 1.422843
## Gamma diversity of the metacommunity:
##  UnveilJ
## 138.1078
```

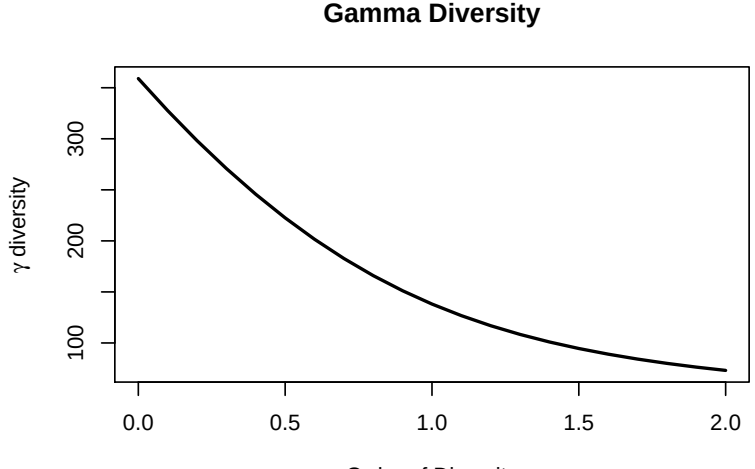
Diversity profile


```
plot(DivProfile(MC = Paracou618.MC,
                Biased=FALSE, Correction="UnveilJ"))
```









Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Relative Entropy

Departure of the observed distribution from the expected distribution (Kullback and Leibler 1951's divergence).

Generalization to order q (Marcon et al. 2014).

If the expected distribution is the average distribution (that of the assemblage) then the relative entropy is the difference between the entropy of the assemblage (γ) and that of each community (α): it is β entropy.

Useless but important.

Differentiation vs Proportional Diversity

This definition of β diversity measures the average departure of a community from the meta-community.

$\rightarrow$ Proportional Diversity.

Other measures exist to define how different two communities are from each other.

$\rightarrow$ Differentiation Diversity.

Motivation

Neutral diversity

Estimation

Practical 1

Partitioning Diversity

Practical 2

Phylogenetic diversity

Practical 3

Functional Diversity

Conclusion

Section 6

Practical 2

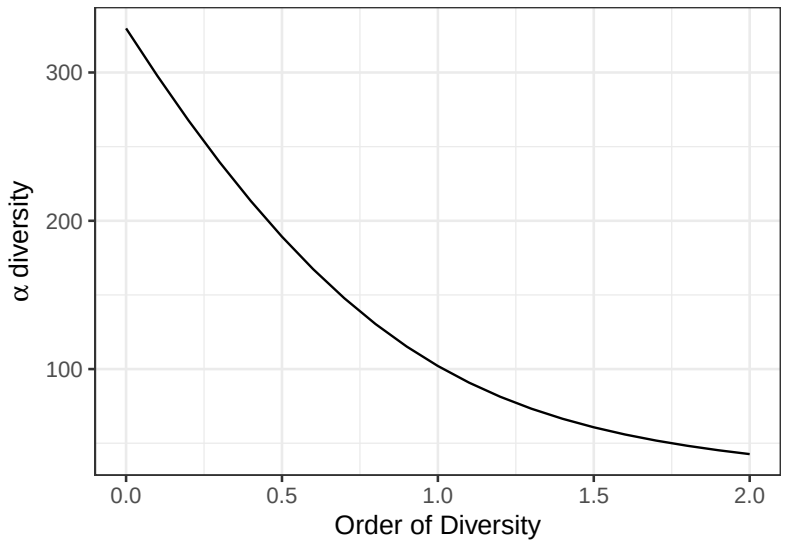
Data

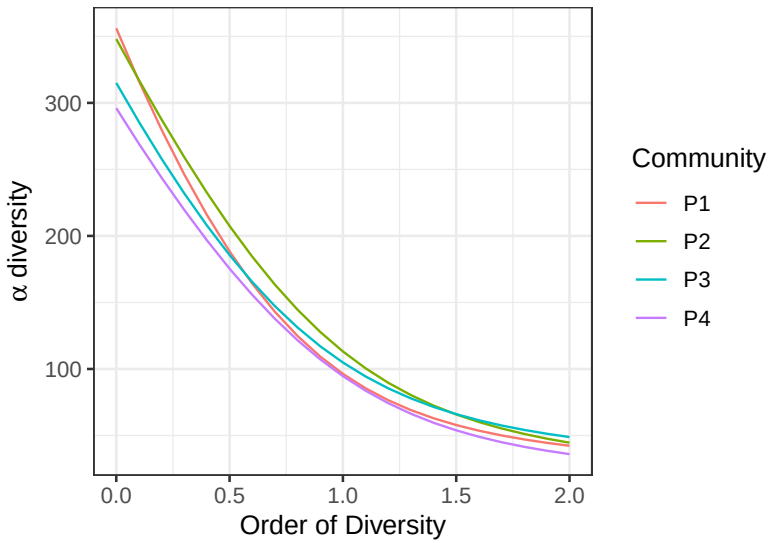
Formatting the data as a dataframe, species in lines, subplots in columns.

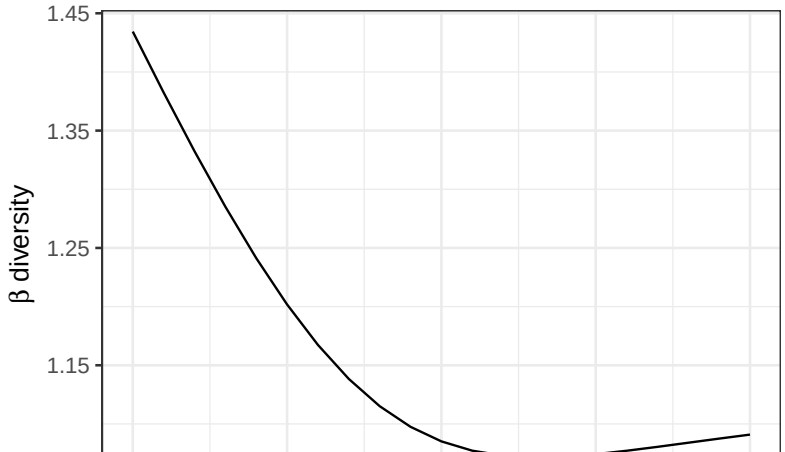
```
AbundancesP6 %>%
  spread(key = SubPlot, value = Abundance, fill=0) %>%
  as.data.frame -> df
# Name rows and columns
rownames(df) <- df$spName
df <- df[, -1]
colnames(df) <- paste("P", colnames(df), sep="")
# Create a MetaCommunity object
ParacouMC <- MetaCommunity(df, Weights = colSums(df))
```

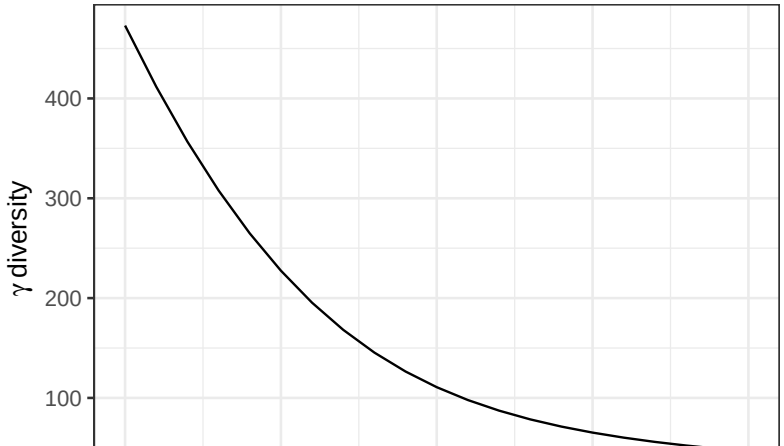
Diversity Profile


```
dp <- DivProfile(, ParacouMC,
                 Biased = FALSE, Correction="UnveilJ")
autoplot(dp)
```









Motivation

Neutral diversity

Estimation

Practical 1

Partitioning Diversity

Practical 2

Phylogenetic diversity

Practical 3

Functional Diversity

Conclusion

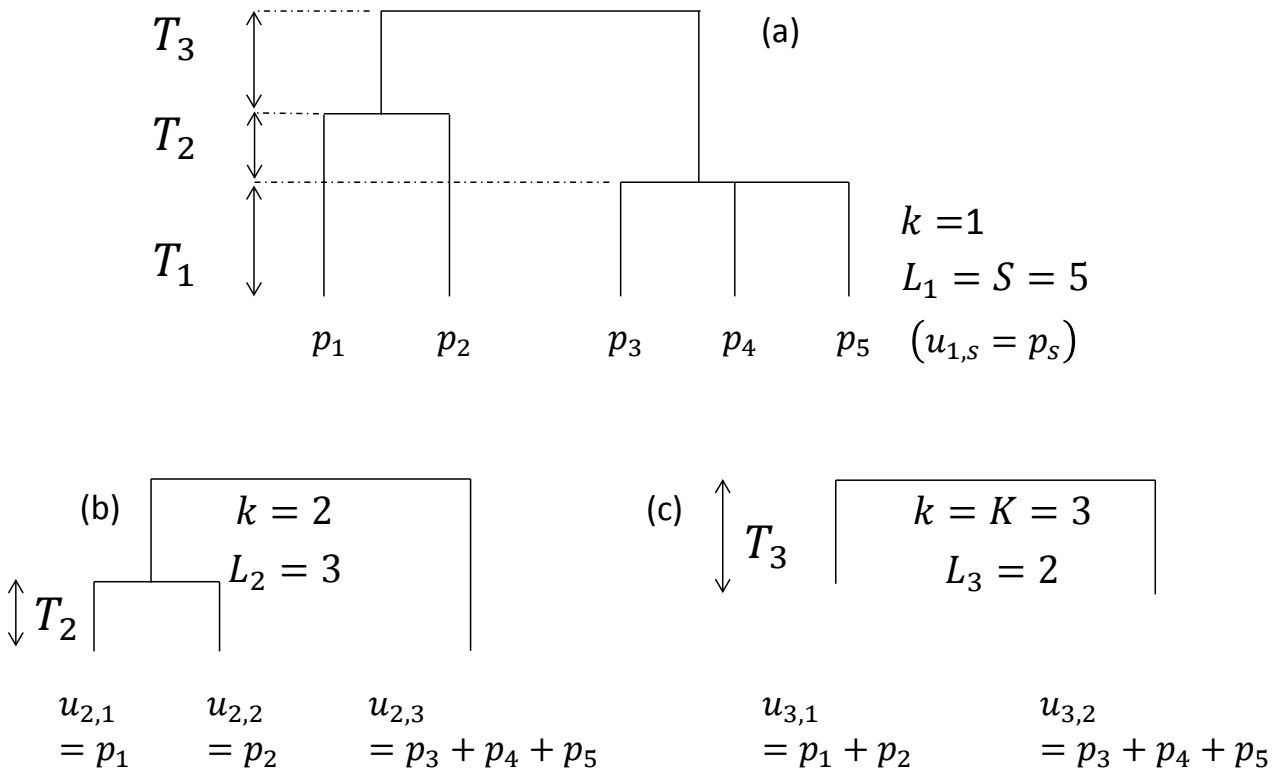
Section 7

Phylogenetic diversity

Eric Marcon

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Ultrametric phylogeny



Phylogenetic Entropy and Diversity

Entropy sums along the tree.

Diversity is its exponential (Marcon and Hérault 2015a). It is the number of equiprobable species in a star phylogeny of height 1.

$\rightarrow$ Estimate entropy along the tree, average it, transform the phyloentropy into phylodiversity.

Phylogenetic Diversity

Eric Marcon

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Phylogenetic diversity of order 0 is called PD. It reduces to richness in a star tree of height 1.

Phylogenetic entropy of order 2 is Rao's quadratic entropy. It reduces to Simpson's entropy in a star tree of height 1.

Motivation

Neutral diversity

Estimation

Practical 1

Partitioning Diversity

Practical 2

Phylogenetic diversity

Practical 3

Functional Diversity

Conclusion

Section 8

Practical 3

Data

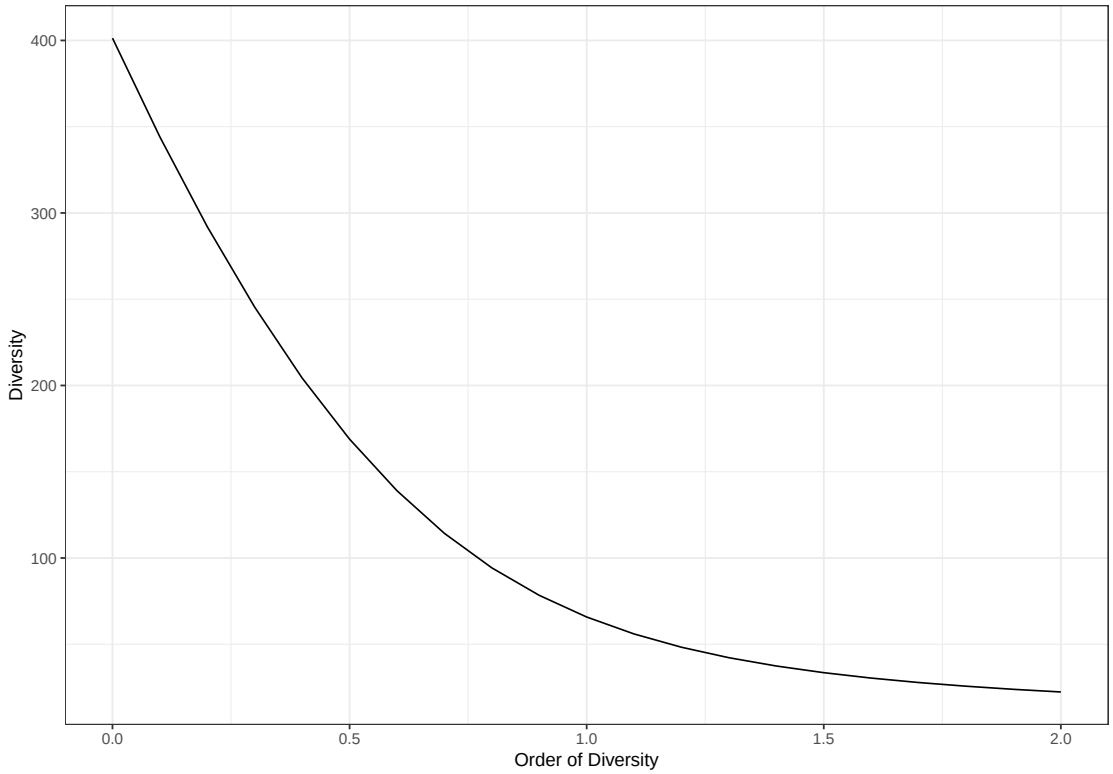
Make a taxonomic tree.

```
library("ape")
library("magrittr")
Paracou %>%
  filter(Plot == 6) %>%
  select(Family:Species) %>%
  unite(col=spName, Genus, Species, remove=FALSE) %>%
  mutate_if(is.character, as.factor)  %>%
  {as.phylo(~Family/Genus/spName, data=., collapse=FALSE)} %>%
  compute.brlen(method=1) %>%
  collapse.singles %>%
  multi2di %T>%
  plot(show.tip.label = FALSE) -> p6Phylo
```

Diversity Profile

Same estimator as that of neutral diversity.

```
dp <- CommunityProfile(function(Abd, q, CheckArguments)
  PhyloDiversity(Abd, q, Correction="UnveilJ", Tree=p6Phylo)$Total,
  AbdP6)
autoplot(dp)
```



Section 9

Functional Diversity

Functional space

Distance between species are not ultrametric.

Metrics based on the distance matrix.

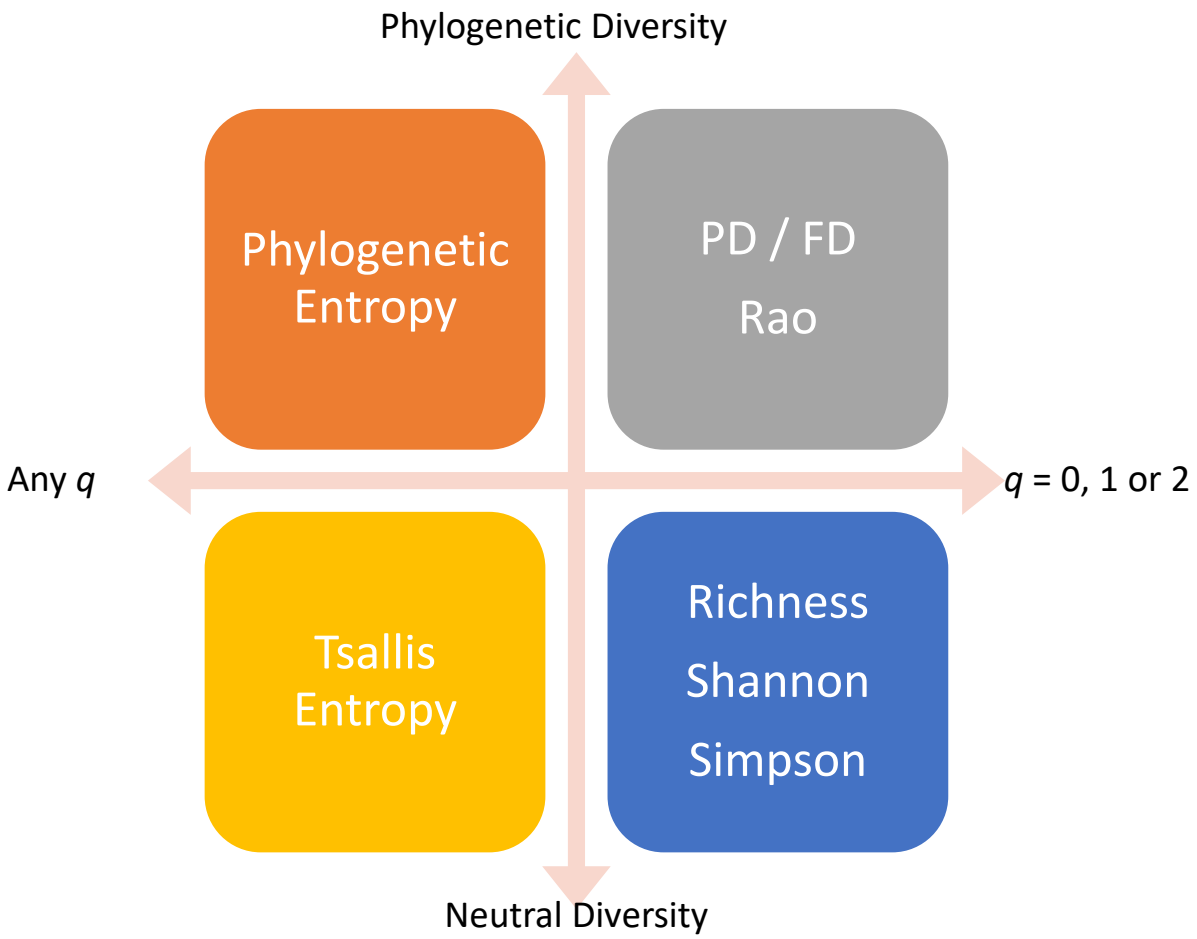
No time here: see part 3 of the book.

Section 10

Conclusion

Entropy to unify diversity measures

Diversity Metrics
Eric Marcon
Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion



Metabarcoding

Motivation
Neutral diversity
Estimation
Practical 1
Partitioning Diversity
Practical 2
Phylogenetic diversity
Practical 3
Functional Diversity
Conclusion

Satisfactory results when supervised clustering is possible (but no estimation bias correction available)

Risky results with unsupervised clustering. Seems to work quite well around Shannon's diversity.

References

Hill, M. O. 1973. "Diversity and Evenness: A Unifying Notation and Its Consequences." *Ecology* 54 (2): 427–32. https://doi.org/10.2307/1934352.

Kullback, S., and R. A. Leibler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22 (1): 79–86.

Lande, Russell. 1996. "Statistics and partitioning of species diversity, and similarity among multiple communities." *Oikos* 76 (1): 5–13.

Marcon, Eric, and Bruno Hérault. 2015a. "Decomposing Phylodiversity." *Methods in Ecology and Evolution* 6 (3): 333–39. https://doi.org/10.1111/2041-210X.12323.

———. 2015b. "entropart, an R Package to Measure and Partition Diversity." *Journal of Statistical Software* 67 (8): 1–26. https://doi.org/10.18637/jss.v067.i08.

Marcon, Eric, Ivan Scotti, Bruno Hérault, Vivien Rossi, and Gabriel Lang. 2014. "Generalization of the Partitioning of Shannon Diversity." *Plos One* 9 (3): e90289. https://doi.org/10.1371/journal.pone.0090289.

Tsallis, Constantino. 1988. "Possible generalization of Boltzmann-Gibbs statistics." *Journal of Statistical Physics* 52 (1): 479–87. https://doi.org/10.1007/BF01016429.

———. 1994. "What are the numbers that experiments provide?" *Química Nova* 17 (6): 468–71. http://quimicanova.sbq.org.br/detalhe%7B/_%7Dartigo.asp?id=5517.

Whittaker, R. H. 1972. "Evolution and Measurement of Species Diversity." *Taxon* 21 (2/3): 213–51. https://doi.org/10.2307/1218190.