

Species richness for microbial ecologists

Amy Willis

6 August 2016

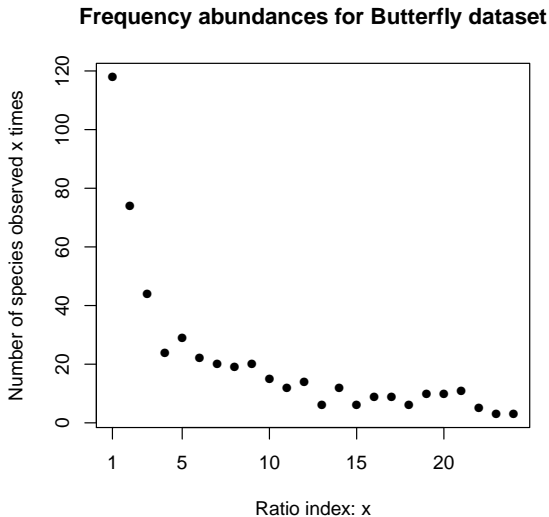
Goal: Estimate the total number of different taxa



First appearance: Fisher *et al.* (*J Anim Ecol* 1943)

Species richness for microbial ecologists

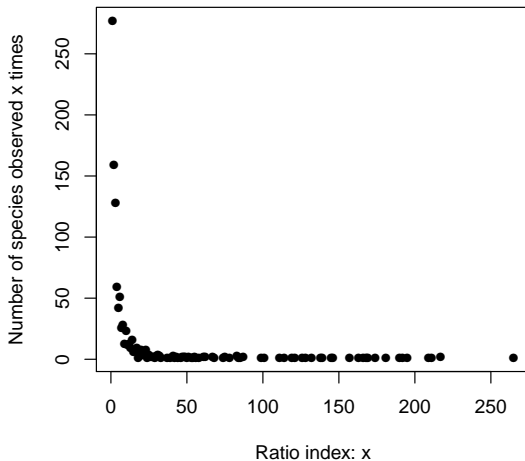
Data structure: Traditional



Fisher *et al.* (*J Anim Ecol* 1943)

Data structure: Modern

Frequency abundances for Apples dataset



Walsh *et al.* (*Front Microbiol* 2014)

Richness

Species richness: The total number of different taxa/species/OTUs in the *population*

- Will inherit from the definition your data uses (eg. 97%)
- Key is *population*

Richness: sample richness

A common, but bad, estimator of the (total, or population) richness is the sample richness

sample richness = number of unique species in your sample

Don't use this! Your sample is not the target of interest!

Richness: Chao1

Another common estimator is the Chao1 “estimator” (or its friends)

$$\text{Chao1} = \text{number of unique species in sample} + \frac{(\# \text{ singletons})^2}{2 \times \# \text{ doubletons}}$$

$$= c + \frac{f_1^2}{2f_2}$$

$$\text{s.e.}(\text{Chao1}) = \sqrt{f_2 \left(\frac{1}{2} \left(\frac{f_1}{f_2} \right)^2 + \left(\frac{f_1}{f_2} \right)^3 + \frac{1}{4} \left(\frac{f_1}{f_2} \right)^4 \right)}$$

We advise against this – too restrictive.

Richness: CatchAll

CatchAll is software for richness estimation

- Fits lots of statistical models and selects best one
- Good for microbial studies in “low-diversity” case
 - ▶ Homogeneous systems: gut, vaginal...
- Very stable
 - ▶ Too stable? Underreports errors

Richness: Objective Bayes

The newest distribution of richness software is *objective_bayes_**

- Massive improvements in statistical honesty compared to CatchAll
- The same class of models, but better estimation and error quantification
- Just released by Kathryn Barger
 - ▶ Low-to-medium-to-medium-high diversity cases?
- Easy to use; available in R

Richness: breakaway

Other good option for richness estimation is *breakaway*

- Fits different class of statistical models and selects best one
- Good in higher diversity cases
 - ▶ Non-homogeneous systems: Soil, water...
- A different set of assumptions
 - ▶ Pros: More flexible; nests many other cases
 - ▶ Cons: large standard errors
- Easy to use; available in R

Species richness for microbial ecologists

Amy Willis

6 August 2016