

**LS1102: Diversity Practicals.** Notes for working on Bird Data you collected and a sample Tree Data (from Sinharaja, Sri Lanka).

**Instructors:** Dr. Robert John Chandran, Department of Biological Sciences

**Date:** August 28, 2019

Your field exercise was to collect data on bird diversity in the IISER Kolkata campus. You were required to walk on trails and spot birds, count the ones you see and identify them. Your data should therefore consist of the long list of individuals you spotted and their respective identities. Where species identification was not possible you could assign an operational taxonomic unit (OTU1, OTU2, . . .) for the distinct individuals. Please note that the same OTU number should be given to all individuals that belong to the same unnamed species.

Using this method, you can estimate *alpha diversity*, (the number of species of a taxonomic group in a site). Estimating *beta diversity* would be harder as it requires comparison of species composition **among** sites. For example if you have sampled in 2 or more distinct locations, you could compute the geographical distances between each pairs of sites and also compute the corresponding *similarity* in species composition between the pairs. The rate at which the species composition changes (or turnover) as a function of geographical distance is a measure of beta diversity.

**Alpha Diversity** could be expressed as the following:

1. The total number of species (**S**) in a site.

(**Note:** This measure of alpha diversity does not consider the abundances of species. Ecologists believe that not only species numbers but their respective abundances are also important to consider in expressing diversity. Species abundances tend to be unequal, and the community typically consists of a few abundant species with increasing numbers of less-abundant species. The higher the inequality in species abundances, the lower the diversity of the community).

The following indices of diversity consider species abundances:

2. Simpson's Index:  $D = 1 - \sum_{i=1}^S p_i^2$ , where  $p_i$  is the proportional abundance of the  $i^{th}$  species.

3. Shannon's index:  $H = - \sum_{i=1}^S p_i \log(p_i)$

These are simple indices to compute. While the absolute value of these indices for a site is of limited use, these values are quite useful in comparing multiple sites.

For **Beta Diversity** we will compute the similarity in species composition between each pair of sites. Two simple indices can help:

4. Jaccard Index:  $J_{sim} = \frac{a}{a+b+c}$

5. Sorensen Index:  $S_{sim} = \frac{2a}{2a+b+c}$ ,

where  $a$  is the number of species shared between the two sites,  $b$  is the number of species found only in one of the sites and  $c$  is the number of species found only in the other site.

Geographical distances between pairs of sites are easy to compute from the  $(X, Y)$  locations, which can be the latitude and longitude of sites or simple  $(x, y)$  locations in a defined plane.

You may then plot *geographical distance* **versus**  $J_{sim}$  or  $S_{sim}$ , using all possible unique pairwise comparisons.

## Sampling Issues

Your assessment of diversity in a site is subject to sampling error. That is, due to chance, species may not be detected by you even when they are present, and this affects rare species more than common ones. So when you count a small number of individuals you not only miss out species, but also get biased estimates of species abundance. There is no way to make up for small samples other than sampling more, that is counting greater numbers of individuals randomly across your study site. One way to check sampling effort is to plot the **species-individual** relationship with your data. Sample a fixed number of individuals randomly *without replacement* and count the number of species in the sample. Increase the numbers of individuals sampled steadily and keep track of the number of unique species in each sample. The relationship tends to saturate (show asymptotic behaviour) at large sample sizes, depending on the diversity in the site. You may use this to assess the sample size required to estimate diversity in your study site.

## Species abundances

Species abundances are usually examined using the **species-abundance** distribution. This is easy to plot. Just plot a **histogram** of species abundance after *transforming* the species abundance values to the log2 scale. That is the same as counting how many species (Y-axis) fall in the following doubling abundance classes (X-axis): (0-1), (1-2), (3-4), (4-8), (8-16), (16-32), (32-64), (64-128), . . . and so on.

## Instructions for installing R

R is a free programming and statistical analyses environment. You may install this on Windows, Mac OS, or Linux platforms. You typically install R Base first, and then install additional packages depending on your needs. R Base is enough to do most of the basic stuff, while the packages serve specialised requirements.

After you have installed R Base, you may install R Studio, which is an elegant user-friendly front-end to work in R.

R can be installed by following instructions and downloads available at <https://cran.r-project.org/>

R Studio can be downloaded from <https://www.rstudio.com/>

You may learn the basics of R online from <https://www.datacamp.com/courses/free-introduction-to-r>

An introductory manual can be downloaded from <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

Start your tutorials after installing R and R Studio (optional)