**Patterns in community data**

The **occurrence** of species (i.e., their distribution) as well as their **abundance** are two interrelated factors that, ultimately, define biodiversity. These properties often respond in tandem to generate patterns that have been documented in a wide variety of regions and taxa.

For example, **most species are rare; few are common**. This is true both respect to occurrence (i.e., most species only occur at a few locations; only a few are more widespread in distribution) and abundance (i.e., for most communities, only a few species will comprise the majority of the individuals).

In addition, not all species present in a given region (part of the regional species pool) will be present in any given community, for a variety of reasons (resource limitation, dispersal limitation, biotic interactions, and/or chance). Thus, **most communities will contain only a subset of what species occur within that region**.

Furthermore, common species tend to be widespread whereas rare species tend to have smaller distributions (in other words, there tends to be a **positive relationship between occurrence and abundance**, with widespread species occurring in higher densities compared to species restricted in their geographic distribution).

Finally, **communities that have high species richness also tend to have high overall abundance**.

There are also some patterns in community data that are a function not of nature (as above) but of sampling. For example, **species richness is positively associated with the number of samples collected** (because of the increased likelihood of encountering a rare species with more sampling).

We will explore these patterns today. If your data show patterns that are contrary to the above “standards,” you need to determine why. It could be something as trivial as an error in data transcription, or it could be due to low sample size, or it could be something real and odd about your data. And all of those are things you need to know!

**Exercise:**

Open a new RStudio session and set your working directory with the following libraries:

*BiodiversityR* – this is a package we haven’t used before, so be sure to install it first

*labdsv*

*MASS*

*MVA*

*optpart*

*stats*

*vegan*

Today we will start off once again with bryceveg.R:

veg <- read.table("C://your/file/path/bryceveg.R",header=TRUE)

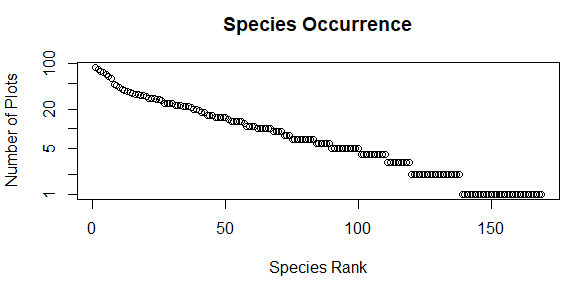
**Graphically depicting patterns of abundance and occurrence:**

The function abuocc() from *labdsv* is extremely useful in examining patterns of occurrence and abundance. It yields four plots of species abundance and occurrence:

abuocc(veg)

*Most species are rare; few are common - occurrence:*

The first graph, titled Species Occurrence, shows the # of plots (Y axis) vs. species ranked from most frequently encountered to those found at only 1 site (X-axis).

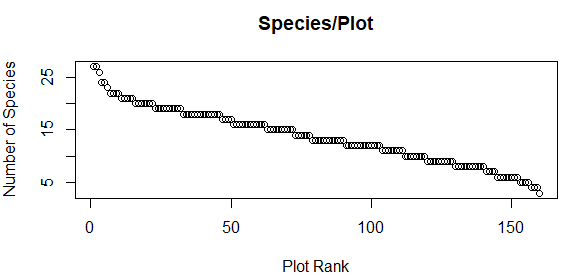


This graph indicates that most species occurred at only 20 or fewer plots.

This is one of two ways of depicting that most species are rare, few are common: it’s based on occurrence. This pattern is also true with respect to abundance; we will come back to that shortly (after we go through all four of the output graphs of the abuocc() function).

*Most communities contain only a subset of the species of that region:*

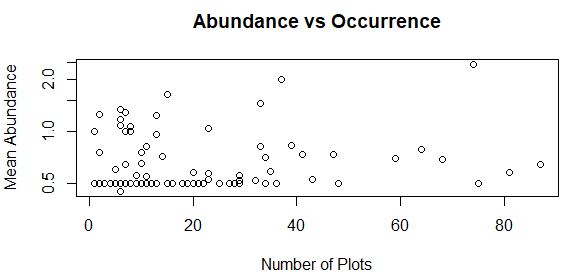
The next graph, representing species per plot, has the # of species (Y axis) vs. plots ranked from most to least speciose (X-axis).



You can see that most plots had 15 or fewer species.

*There is usually a positive relationship between occurrence and abundance:*

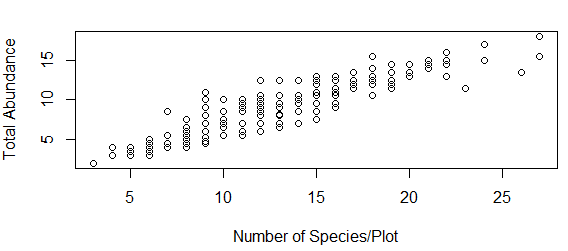
The third graph, Abundance vs. Occurrence, plots mean abundance (Y axis) vs. number of plots (X axis). When it asks if you want to identify individual species, I suggest you answer N for No (can take a loooooooooong time otherwise).



If you were to fit a regression line through the points, you’d see that it would have a positive slope, indicating that the most widespread species (occurred in the greatest number of plots) were the most abundant on average. Even without such a line, you can see that the rarest species (lowest abundance) were found in the fewest plots.

*Communities with many species also tend to have high overall abundance:*

The final graph is total abundance (Y axis) vs. the # of species per plot (X axis). When it asks if you want to identify individual plots, again I suggest answering N.



This graph illustrates that plots that had the most species also had the most individuals.

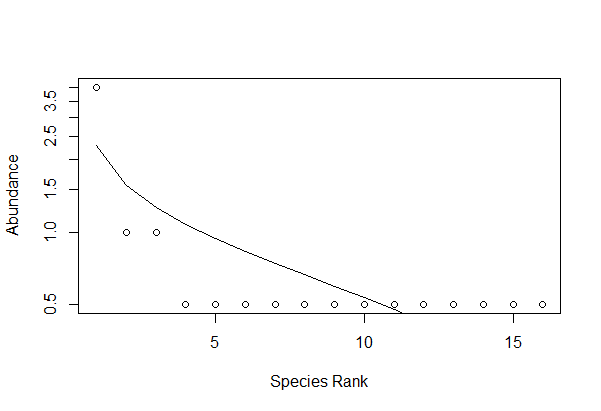
*Most species are rare; few are common - abundance:*

When you plot the species on the X-axis ranked from those with the most individuals to those with the fewest (down to species represented by singletons) and the abundance of each species on the Y-axis, the resulting plot is called a **rank-abundance curve** or **dominance-diversity plot** (or Whittaker plot or species-abundance distribution). From it, you can see that most of the species in a community are represented by relatively few individuals. Different communities can be compared by comparing the shape (slope) of their rank-abundance curves. There are over 2 dozen shapes that species-abundance distributions can take, each resulting from a different process. We will cover more on this topic in an upcoming week.

For an example of this with the Bryce Canyon vegetation data, we’ll use a function from *vegan* to fit some of these data with the most common shape (slope) found, the log-normal. We’ll cover this in more detail in a few weeks; for now, this is just to illustrate the point that in most communities, species have unequal (and usually quite skewed) distributions:

veg.ra <- rad.lognormal(veg[50,])

plot(veg.ra, ylab="Abundance", xlab="Species Rank")



All of the above graphs illustrate patterns very commonly seen in community data that are due to inherent factors about species themselves. Now let’s explore a pattern that emerges as a result of sampling.

*Species richness as a function of sampling effort:*

For this example, we will work with one of the same datasets from the previous week’s assignment, grassland.community.csv. Read it in as an object named comm, with header = TRUE and row.names = 1.

Each cell contains the percent cover of a species in a sample as a proxy of counts of individuals (just as we have dealt with previously in bryceveg.R).

Because many multivariate methods are sensitive to the total abundance in a sample, we must convert these absolute abundance estimates to relative abundance estimates. First, check total abundance in each sample:

apply(comm, 1, sum)

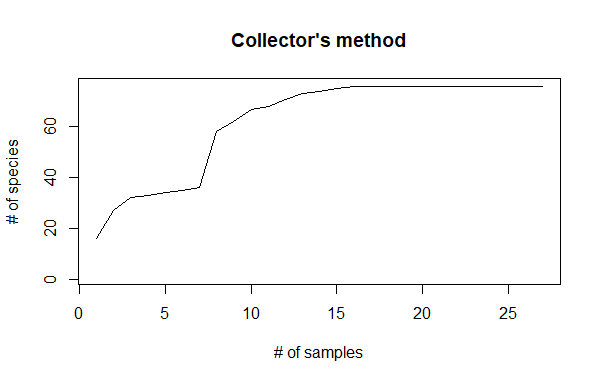
then convert to relative abundance by dividing each value by sample total abundance:

comm <- decostand(comm, method = "total")

Now you can determine whether sampling was sufficient to represent species richness. To do so, you can plot a **species accumulation curve**, of which there are several types. A species accumulation curve allows you to determine how much sampling effort is needed to represent richness, and whether your sampling has done so, by presence of an asymptote. These kinds of graphs are sometimes called “collector’s curves,” “sampling effort curves,” or “species discovery curves.”

The simplest type just plots the raw numbers of species as a function of the number of samples collected. This results in a “collector’s curve”:

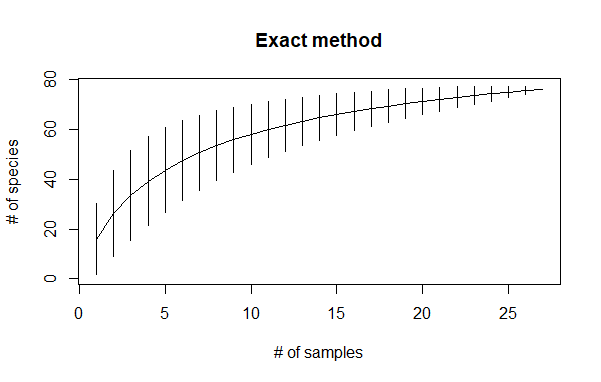
plot(specaccum(comm, method = "collector"), main = "Collector’s method", xlab = "# of samples", ylab = "# of species")



This graph indicates that only 16 samples are needed to adequate represent species richness.

There is also a method that produces a smoother curve with error bars (which are calculated by dividing the standard deviation by the square root of sample size [i.e., the number of measurements that make up the mean]):

plot(specaccum(comm, method = "exact"), main = "Exact method", xlab = "# of samples", ylab = "# of species")



For this curve, the error bars appear to plateau around 16-17 samples (consistent with the previous curve).

If you examine the data in comm, you’ll see that the sites have names like “mix-O-1” and “fes-K-8”. These alphanumeric codes encompass two habitat types (mixedgrass and fescue). Last time you used these data to examine species richness by habitat type. Now let’s see whether sampling was adequate for us to feel confident that those answers are accurate, or whether they are underestimates.

To do so, first load the site information about this dataset:

metadata <- read.csv("plot.metadata.csv", header = TRUE, row.names = 1)

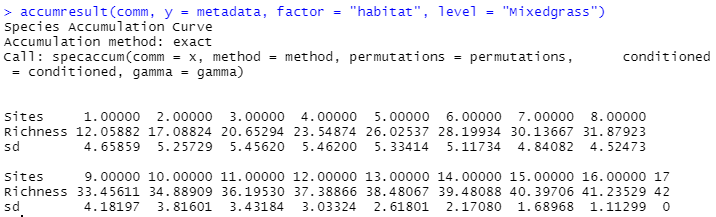
Examine it to refresh your memory about the exact names (and capitalizations; remember, things in R are case-sensitive) of the two habitat types.

Now we will use the accumresult() function from *BiodiversityR* as follows to obtain info for each habitat type separately:

accumresult(comm, y = metadata, factor = "habitat", level = "Mixedgrass")

accumresult(comm, y = metadata, factor = "habitat", level = "Fescue")

These commands will return the following (just the one for Mixedgrass shown):



These are data that would be used to make a plot; we’ll use the accumcomp() function from *BiodiversityR* to do so. *BiodiversityR* has some nice plotting options that make it a little more useful for some purposes. Its default method of calculating a species accumulation curve is the exact method, but it also has other options; use ?accumresult to learn more.

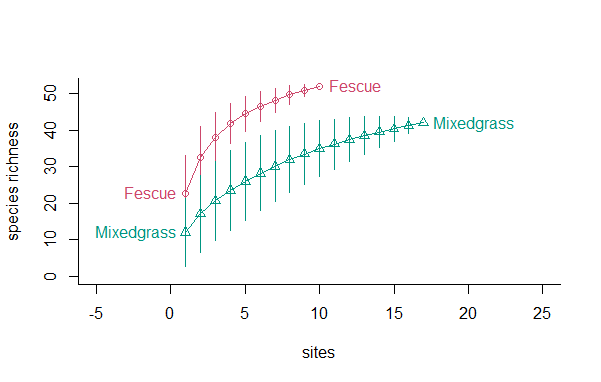
accumcomp(comm, y = metadata, factor = "habitat",

xlim = c(-5,25), plotit = TRUE, rainbow = TRUE,

legend = FALSE)

The rainbow = T makes the graph in color (use rainbow = F for black and white). The legend = F omits a legend; if you want one, use legend = T and then click on the graph where you want the legend.

I had to play around a bit with the scale of the X-axis in the commands above to make the graph labels show in their entirety. You should also remember to stretch the plot window to try to see the curves better!



From this graph, you can see that Fescue had more species even though it was sampled at fewer sites; however, its curve doesn’t show a strong asymptote, indicating that there are likely even more species present in fescue-dominated area, and perhaps surveying more sites would increase the species richness value for that habitat type. The curve for Mixedgrass appears to be levelling off, indicating that the species richness value for it is likely correct and wouldn’t change much with more sampling.

Use Zhou et al. 2013 to illustrate some real-world example of pattern of occurrence and abundance (Fig 4)

Zhou, Y., et al. 2013. Biogeography of the ecosystems of a healthy human body. Genome Biology 14:R1. <http://genomebiology.com/content/14/1/R1>.

**Assignment:** due 0800 Monday, 1 March

Start a fresh RStudio session. Remember to set your working directory to your course folder and use the same package libraries as we used today.

Use the Ground\_beetles\_abundance.csv data; read it in as an object named GBA, with row.names = 1. Examine the data: you can see that this is a site x species matrix. There were three site types (habitat types: E = edge, G = grassland, and W = wood), with six replicate sites of each. (Compare Ground\_beetles\_abundance.csv with GBbiol.csv; they are both the same site x species data but are expressed in different ways: GBbiol.csv is a presence/absence site x species matrix whereas Ground\_beetles\_abundance.csv is an abundance site x species matrix. It is typically MUCH easier to use Excel to create alternative forms of a dataset than to try to wrestle with an existing dataset in R.)

Now read in GBsite.csv as an object named gb.site, with row.names = 1. This site x environment .csv is similar to Ground\_beetles\_habitat.csv, but again, compare the two .csv files and note the differences: GBsite.csv is much pared down, with information only on the 18 sites, their habitat type, and the maximum height of vegetation at each. It has rows of sites (which it calls Samples) rather than rows of species as in Ground\_beetles\_habitat.csv.

Last week you addressed the most basic questions about any ecological community: what is the overall species richness, and which habitat type had the most species? But without an assessment of species accumulation, the values of richness that you obtain may be underestimates if your sampling hasn’t been adequate. Using skills you gained today as well as over the past few weeks, answer the following questions:

Redo analyses with GBA rather than GBbiol.csv (as was done in Sp21)

**Q1. Are rare species common and common ones rare in ground beetles?**

**Q2. Is per-site richness only a subset of overall richness?**

**Q3. Is there a positive relationship between occurrence and abundance?**

**Q4. Do sites with high richness also have high abundance?**

**Q5. Was sampling adequate for each of the three habitat types (in other words, were six replicates enough to represent species richness)?**

Make an RMarkdown Word file of your work and turn that in. Be sure to include your answers to the questions asked!