Shannon diversity background reading

Introduction

In this lab we will be analyzing the data from the compost arthropod lab to determine values for species **richness**, **evenness**, and **diversity**. We will compare two metrics of soil invertebrate biodiversity:

We will be using the Shannon's index (H') to calculate species **diversity**. This is one of the most common measures of species diversity in ecology, though there are others. Shannon's index takes into account both **richness** (the number of species) and **evenness**, or how evenly individuals are distributed among species. A large value of H' denotes high biodiversity. Shannon's index is advantageous over simply counting the total number of different species, because a simple count is affected by sampling effort (plot size and total number of individuals sampled): the larger the sample, the more rare species you find. H' is superior because it is calculated from proportions (%), as you will see, and rare species contribute very little. Therefore, this index is relatively insensitive to the random inclusion or omission of rare species that happens with any sampling effort.

The equation for Shannon's index is:

(Note: our spreadsheet performs these calculations.)

$$H' = -\sum_{i=1}^{s} p_i \ln(p_i)$$

or

$$H' = -[p_1 \ln(p_1) + p_2 \ln(p_2) + p_3 \ln(p_3) + \ldots + p_s \ln(p_s)]$$

The proportion of species i relative to the total number of species (p_i) is calculated, and then multiplied by the natural logarithm of this proportion ($\ln p_i$). The resulting product is summed across all species, and multiplied by -1.

Shannon's index is unitless and has no true biological meaning. However, if we take the exponent of this index, or $e^{H'}$, we have an "equivalent number of equally common species." In other words, $e^{H'}$ is a type of weighted number of species present in your sample for which very common species contribute much more than rare species to the numerical "diversity" estimate.

Consider Example 1 below. Note that although there are a total of 5 species in the sample, species A accounts for 70% of the observations. In fact $e^{H'}$ for this 5 species sample is only about 2.74 "equivalently common species." So the diversity of this 5 species community is the same as that of a community with (not quite) 3 species with all the same abundance. This community has a fairly low diversity because it is dominated by species A.

In contrast in Example 2 below, note that all 5 species are equally abundant; consequently $e^{H'}$ equals 5 species. Thus, this is a high diversity community – as high as it can get for a 5 species community. In fact, the maximum value of eH′ will always be S, the number of species, when all species are equally abundant, and the actual value of $e^{H'}$ should always be compared against S.

Example 1

Species	**A**	**B**	**C**	**D**	**E**	**Total**
# obs.	70	10	10	5	5	100
p_i	0.7	0.1	0.1	0.05	0.05	

$H' = -[(0.7) \times \ln(0.7) + 2 \times (0.1) \times \ln(0.01) + 2 \times (0.05) \times \ln(0.05)]$

$H' = 1.01$ and $e^{H'} = 2.74$ equivalently common species

Example 2

Species	**A**	**B**	**C**	**D**	**E**	**Total**
# obs.	20	20	20	20	20	100
p_i	0.2	0.2	0.2	0.2	0.2	

$H' = -[5 \times (0.2) \times \ln(0.2)]$

$H' = 1.61$ and $e^{H'} = 5.00$ equivalently common species

Calculating species richness and evenness

It is also desirable to separately measure richness and evenness. A simple way to calculate richness is to tally the number of species. We can calculate evenness (J') from H'. To do this, we first have to calculate the maximum possible diversity (H'_{max}), given the number of species S. That would occur when individuals are distributed evenly among species, as in Example 2. The formula for H'_{max} is:

$$H'_{max} = \ln(S)$$

and the formula for evenness is:

$$J' = H'/H'_{max}$$

Thus, in Example 1, $H'_{max} = \ln(5) = 1.61$, and $J' = 1.01/1.61 = 0.627$. In Example 2, H'_{max} is also 1.61, but $J' = 1$! That is because H' in Example 2 is the maximum possible diversity, as the individuals are distributed evenly among species. Note that since we will usually miss rare species in a sample from a community, H'_{max} is usually an underestimate and so J' is an overestimate of evenness.