Measuring Biodiversity

Biodiversity is the level of variation of life-forms in an ecosystem. Evidence suggests that greater biodiversity leads to a greater ability of ecosystems to survive shocks from climate change or human intervention. Biodiversity is also valuable economically, since the more varied species that are found, the greater the likelihood that some may be commercially or medically useful. As important as biodiversity is, it is a complex concept. There are a number of ways to measure biodiversity and it is important to be specific about the question that is being asked and the level at which diversity is being measured. For instance, we may be interested in the diversity of canopy tree species in the lowland rainforests of Belize. We would collect data on these trees, and not include data on lizards or species of insects. In this lecture we will learn about a particular measure of biodiversity called Simpson's diversity index. First, we give some definitions:

(i) **Species richness** - the number of species that are found in an ecosystem.

(ii) **Species abundance** - the number of individuals of a particular species.

Species richness may seem like an obvious measure of biodiversity. After all, the more diverse an ecosystem, the more species it should contain. That may be true, but look around at your classmates. Imagine everyone has to choose a t-shirt that is either blue, green, yellow, orange, or red. If there are twenty students in the classroom, sixteen may all choose blue shirts, while one may wear green, one yellow, one orange, and one red. This class has *five* different kinds of t-shirts, which we can imagine as species. Compare this with another class where four students wear blue, four wear green, four wear yellow, four wear orange, and four wear red. Again, there are five different kinds of t-shirts in this class as well. But is it more or less diverse? It would seem that the first class, where nearly everyone is wearing a blue shirt is less diverse than the other class which is more evenly mixed. That's why the number of individuals wearing each type of shirt matters. If instead of t-shirts we are talking about species, then the number of individuals is called *species abundance*.

In the previous example, it was clear that the second class was more evenly mixed, since every color was represented equally. In nature, we rarely collect data that is so clear. Often only a few species are heavily represented in a sample, while most species appear in very small amounts. In order to compare samples from different ecosystems, we need a precise *quantitative* measure of how evenly spread the individuals are across the different species–this quantity is called **species evenness**.

Species A, Species B, Species C, Species D,
Species E

Figure 1: Representation of an ecosystem as a jar of jellybeans. Each of the five colors represents a species (we're going to ignore the purple jellybeans at the very top!).

Simpson's diversity index

One important quantitative measure of species evenness is Simpson's diversity index. Most simply, Simpson's diversity index is the probability that two randomly chosen individuals in a sample are of different species. We'll get more specific in a moment, but first, let's talk about jellybeans.

Jar of jellybeans

Imagine that an ecosystem is a jar. Each species is represented by a different color, and individuals are represented by each jellybean. We have two ecosystems, one called Habitat 1, and the other Habitat 2. Both of them contain the same species, which we'll call A, B, C, D, and E. We've collected 100 individuals in each habitat and here's the data on how many of each species we found.

Species	Habitat 1	Habitat 2
Species A (blue)	59	26
Species B (green)	19	23
Species C (yellow)	11	18
Species D (orange)	6	17
Species E (red)	5	16

Both jars contain the same number of colors (species), but the number of each color is different. Let's say that species A are the blue jellybeans in the jar. If we choose a jellybean at random from the jar representing Habitat 1, what is the probability that it is blue (species A)? That's a simple problem, it's just the number of blue jellybeans divided by the total number of jellybeans in the jar, or P[grabbing a blue jellybean]=$59/100 = .59$. Similarly, the probability of grabbing a blue jellybean from the Habitat 2 jar is $26/100 = .26$.

If the jar is large and full of lots of jellybeans, then the probability that **two** randomly selected jellybeans from the Habitat 1 jar are both blue is

$$P[choosing\ \ two\ \ blue\ \ jellybeans] = P[blue\ \ jellybean] * P[blue\ \ jellybean] = (.59)^2 = .3481 \quad (1)$$

In a jar where the individuals are more evenly spread across the different species, then we should expect that the probability of choosing two individuals from the same species should get lower, since we'd have a better chance of choosing another species on the second draw. The idea of the Simpson's diversity index, is that a measure of diversity is the probability that two individuals (jellybeans) chosen at random from the sample (jar) are from *different* species.

Think of it this way: if most of the jellybeans are one color, say red, and only a few of them are the other four colors, then it's most likely that our first draw will be a red. When we select the next jellybean, again, it is most likely that we'd choose a red. As individuals are more evenly spread, the chance of choosing two red individuals gets smaller.

Diversity index formula

Simpson's diversity index is the probability that two randomly chosen individuals are of *different* species. If we draw two individuals at random, they will either be the same or different. Therefore we can say that these two outcomes are *mutually exclusive*, in that they cannot *both* happen. One of the laws of probability is that if two events, event 1 and event 2 are mutually exclusive, then $P[event\ \ 1] + P[event\ \ 2] = 1$. This means that one way to calculate the probability that two individuals are different, is to calculate the probability that two randomly chosen individuals are the same and subtract that from 1. Since there are five species, there are five ways that two individuals can be the same. They can both be from A, B, C, D, or E. We have to calculate the probability for each of these species and add them together:

$$P[two\ \ individuals\ \ are\ \ the\ \ same] = P[two\ \ A] + P[two\ \ B] + P[two\ \ C] + P[two\ \ D] + P[two\ \ E] \tag{2}$$

where,

$$P[two\ \ A] = P[A] * P[A], \quad P[two\ \ B] = P[B] * P[B], \ldots \tag{3}$$

Now if we call p_A the probability of drawing an A, then the probability of drawing two A's in a row is $(p_A)^2$. Since the probability of two individuals being the same species is the sum of all the probabilities of choosing two individuals from each particular species, we can use the *sigma notation* to write the formula down. In sigma notation we use the capital greek letter Σ as a way of saying, 'add things up'. Here's what we mean:

$$P[two\ \ individuals\ \ are\ \ the\ \ same] = \sum_{i=1}^{5} (p_i)^2 \tag{4}$$

is another way of saying

$$P[two\ \ individuals\ \ are\ \ the\ \ same] = (p_1)^2 + (p_2)^2 + (p_3)^2 + (p_4)^2 + (p_5)^2 \tag{5}$$

where the i is what we call an index. The $i = 1$ says 'start with 1 and keep counting upwards until you get to 5'. Now we can write Simpson's diversity index as

$$I = 1 - \sum_{i=1}^{5} (p_i)^2 \tag{6}$$

again, just a compact way of writing

$$I = 1 - [(p_1)^2 + (p_2)^2 + (p_3)^2 + (p_4)^2 + (p_5)^2] \tag{7}$$

But there's one problem. In the real world, we only ever collect a 'sample' from an ecosystem. That means we never know the *total* number of individuals of each species, or even the total number of species. Values that are calculated from samples are known as *statsitics* or *estimators* and to get good answers, we'd like to make sure our estimators are *unbiased*. Since we are calculating Simpson's diversity index from a sample, we need make sure that it is also unbiased, and we do this by multiplying by a factor $N/(N-1)$, so we now rewrite it as

$$I = \left(\frac{N}{N-1}\right) * \left(1 - \sum_{i=1}^{S} (p_i)^2\right) \tag{8}$$

where S = the number of species (species richness); p_i = the probability of drawing species i; N = the number of individuals in the sample.

Since Simpson's diversity index is a probability, it has a minimum value of zero and a maximum value of 1. Values that are higher mean greater biodiversity, and values that are closer to zero mean less. Now you're ready to take your own data and compute Simpson's diversity index.