

Hornbill diversity – simulation

This exercise gives an overview of species accumulation curves and of the estimates for species richness calculated by EstimateS.

You will need the 'study area' (Appendix A) and the data sheet (Appendix B), and some confetti or coloured paper discs.

You will also need EstimateS installed on your computer.

Doing the surveys

Setting up the 'study area'

The study area (see appendix A) is divided into $9 \times 14 = 126$ plots. Spread the 'hornbills' (paper discs) evenly over the study area – this will be a haphazard arrangement, which is okay provided the plots are chosen at random.

To get random numbers between 1 and 9, open an Excel spreadsheet and type

= CEILING(RAND() * 9, 1)

RAND() produces random numbers between 0 and 1, so $\text{RAND()} * 9$ gives random numbers between 0 and 9. CEILING(..., 1) rounds that up to the nearest whole number, giving random integers between 1 and 9.

In another cell in the spreadsheet, type

= CEILING(RAND() * 14, 1)

to get random numbers between 1 and 14. Use these two random numbers to locate the first plot in the study area to survey.

Surveying the first plot

Look at the first plot indicated by the random numbers, and count the number of each 'species' of hornbill, including all those whose centres lies inside the square.

Enter the names of the species found and the numbers of each into the data sheet (appendix B). Enter the number of species observed so far (S_{obs}) and the number of individuals so far in the rows at the foot of the data sheet and plot the first point on the graph of the Collector's Curve (see the figure on the next page).

Surveying subsequent plots

Press F9 in Excel to generate two more random numbers. (Don't survey the same plot twice; if you have already surveyed the plot selected, press F9 again to get a different plot.)

Add any new species to the list in the first column of the data sheet and enter the numbers for each species, with '0' for those species not detected in the plot.

Add up the total species and individuals so far (ie. the total for plot 1 *and* plot 2) and enter at the foot of the table. Then plot the second point on the graph (see diagram above).

Repeat this until you have surveyed at least 5 plots; if you have time, do a few more.

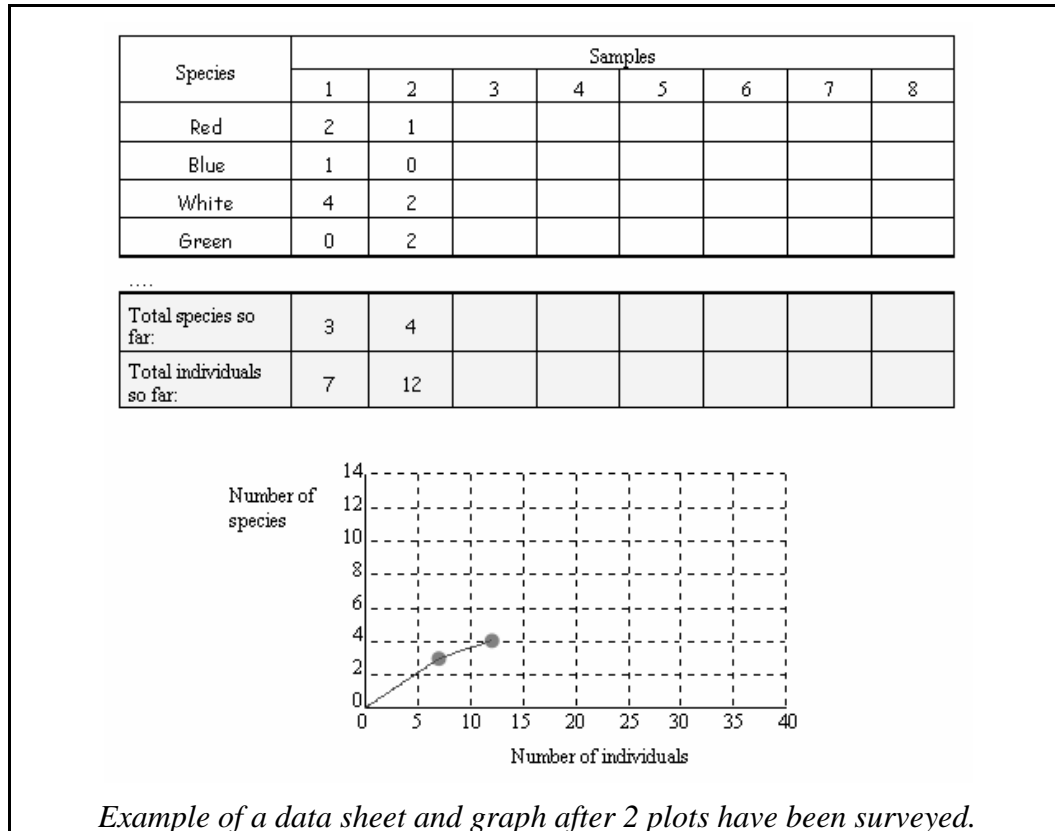
Plotting the Collector's Curve

Join the dots in the graph, starting from the origin (ie. 0 individuals, 0 species).

You will probably get a line that rises steeply at first, then flattens out. Note that the exact shape of the curve depends on the order in which the plots were selected: if you had selected the *same plots* but in a

different order, you would get a different curve (try it if you have time). But the order of doing the surveys should not matter; so which is the correct curve?

We will use EstimateS to shuffle the plots many times (the default is 50) and get an average of the number of individuals and species at each stage of the survey.



Importing the data into EstimateS

EstimateS takes its input from a text (.TXT) file, but it's easier to put it into a spreadsheet first then Save As text. But you will have to stick to EstimateS's format, so don't put in extra comments.

Open a new worksheet in Excel:

- o In cell A1 put in a title such as "Hornbill exercise".
- o In cell A2, enter the number of species you observed.
- o In cell B2, enter the number of plots you surveyed.

From the third row onwards, copy the data from your data sheet, with a row for each species and a column for each plot.

Go to File > Save As... and select "Text (tab delimited) (*.txt)" from the options at the bottom of the dialogue box. Give the file a short name and click Save. Click 'OK' to save as a single sheet, then 'Yes' to keep the workbook in this format. Finally close the file without saving it.

Open the file you just saved in Notepad to see what it looks like.

Open EstimateS. When you first open EstimateS, a file navigation window may pop up asking you to select a "Data File," choose the file called Statistics.4DD, or click on "New" and then "Save" to create a new data file. A licence agreement appears: click 'OK' to agree not to distribute EstimateS commercially.

In EstimateS, click on 'File > Load Data Input File' (or press Ctrl-I). In the dialogue box, navigate to the file with the "hornbill" data and click Open.

A confirmation screen appears with the title of the data set and the number of species and samples. There's also a list of 'Optional' items: don't worry about these, as we can set options later.

Then a dialogue appears asking for the format of the input file: this is Format 1 and there are no rows with sample information or columns with species names.

You should then see a box telling you that the data have been loaded successfully.

Running the analysis in EstimateS

Go to 'Diversity > Compute Diversity Stats' (or press Ctrl+D).

If a box appears asking if it's OK to erase some old Diversity Statistics, click on 'OK'.

A box pops up briefly with a progress bar, then a table of results appears.

Understanding EstimateS's output

The output looks like a spreadsheet with the same number of rows as you had plots in your survey, and lots of columns for the different estimators calculated.

Species accumulation curve

Look first at the columns headed "Individuals (computed)" and "Sobs Mean (runs)".

The first row corresponds to a survey with only one plot, which could be any of the plots in the input file. EstimateS takes a plot at random and notes the number of individuals and number of species (S_{obs}); it does this 50 times and computes the average.

For the second row, EstimateS takes 2 plots at random and gets the number of individuals and species; it does this 50 times and again computes the average.

Similarly, the remaining rows have averages for 50 randomisations of more and more plots, with the last row corresponding to the data for all the plots surveyed.

Plot the values for "Individuals (computed)" and "Sobs Mean (runs)" on the same graph as the collector's curve.

The result is a smoothed version of the collector's curve, usually referred to as the **species accumulation curve**.

The "Sobs Mean (runs)" column is based on 50 randomisations; this is probably enough for a small number of plots. Mao Chang Xuan has developed an algorithm to calculate the values we would get for an *infinite* number of randomisations, which he labelled with the Greek letter τ (tau) : these are in the column headed "Sobs (Mao Tau)". His algorithms also allow us to calculate the confidence interval and SD for S_{obs} .

Singletons, doubletons, uniques and duplicates

Singletons are species which are represented by a single *individual* in the collection. If you look down the column headed "Singletons Mean" you will see the average number of singletons for the 50 randomisations of plots: it should go up at first and then go down as more individuals are recorded. If our sampling has been really thorough, we will have caught all the species there not just once but twice or more, and the number of singletons will drop to zero.

Doubletons are species which are represented by exactly two *individuals* in the collection. The number of doubletons rises and then falls as we add more plots, dropping to zero if we have enough observations.

Uniques are species which occur in only one *plot*, while **duplicates** are species recorded in exactly two *plots*. Uniques and duplicates are more appropriate for animals which tend to occur in groups, so that plots generally have several individuals or none at all.

Coverage estimators

If the number of singletons and doubletons (or uniques and duplicates) has dropped to zero, we can assume that **coverage** is complete, ie. that we have recorded all the species in the population.

If coverage is not complete, Anne Chao has proposed an estimator (referred to as “**Chao 1**”) for the number of species in the population based on the numbers of singletons (F_1) and doubletons (F_2). For the simplest case ($F_2 > 0$), the formula is:

$$\hat{S}_{Chao1} = S_{obs} + \frac{F_1^2}{2F_2}$$

A similar estimate (“**Chao 2**”) is based on the numbers of uniques and duplicates.

We can take this a step further by using rare or infrequent species instead of doubletons (or duplicates). The Abundance-based Coverage Estimator (**ACE**) uses the number of species with 10 or fewer *individuals*, and the Incidence-based Coverage Estimator (**ICE**) uses the number of species occurring in 10 or fewer *plots*.

Consult the online help for EstimateS for the formulae for all these.

Estimators based on resampling

The Jackknife and Bootstrap estimators are based on the idea of changing the samples, seeing the effect on S_{obs} , and then estimating how many species would be recorded with more plots.

The first-order jackknife estimator (**Jack 1**) drops one of the plots and calculates S_{obs} from the remaining plots. It does this for each plot in turn and takes the average, which is compared with the value for all plots. The second-order jackknife (**Jack 2**) drops two plots each time. In practice, Jack 1 and Jack 2 can be calculated from the number of uniques and duplicates.

The **Bootstrap** estimator works by taking new samples from the plots surveys, but doing this *with replacement*, so that some plots appear in the new sample more than once and some are missing. In practice, the bootstrap estimator can be calculated from the proportion of plots which contain each species.

Fitting theoretical curves

EstimateS fits the Michaelis-Menton equation to the observed data. This equation describes the progress of an enzyme-catalyzed chemical reaction. There is no theoretical reason to think it might also describe species accumulation, but it does have the right sort of shape, rising quickly at first and then levelling out. The equation is quite simple:

$$S_{obs} = \frac{S_{max} \times n}{B + n}$$

where S_{max} is the maximum number of species – where the curve levels off, n is the number of individuals in the set of plots so far, and B is the number of individuals needed to get half the maximum number of species, ie. when $n = B$, $S_{obs} = S_{max}/2$. The MMeans column gives us the estimate of S_{max} as we add more plots to the sample, based on the values in the “Mao Tau” column.

How good are the estimates?

Count up the actual number of species in your population and compare this with the estimates produced by EstimateS. Compare your results with those of others in the class, since you will all have different sets of data for the plots, and some may give better estimates than other.

Finishing off

EstimateS keeps all its data and results in a file called "Statistics.4DD". Each time you import a new data set, the old data and results are overwritten. If you want to keep the results, you should export them to a .TXT file before you close EstimateS. Or, of course, you can keep the input file and just run the analysis again in EstimateS.

