Brief Guidelines for Using the Excel Sheets, with Two Examples:

We provide two worksheets in a single Excel workbook (file) for estimating the additional sample size required to reach a fraction g of S_{est}. According to data type, select a proper sheet from the following:

Sheet 1: Abundance data
Sheet 2: Incidence data

Running Procedures:

(1) First, input data in entry cells B5-B8 of the selected sheet. (All data input cells are highlighted in light green and outlined in black.) For incidence data, an input of T (the total number of incidences, in data entry cell B4) is optional. It is needed only for calculating the value of q_0.

After data input, there are two parts: First Part (g =1) and Second Part (g < 1). *You must run the first part (g =1) even you want the sample size only for g < 1.* This is because the output for g < 1 may depend on the output obtained from the case of g = 1.

(2) To run the first part (g =1), key in an initial x_0 (data entry cell B15) to start numerical iterations for solving for a converged value of x^*. For abundance data, $x^* = m/n$ where m denotes the additional number of individuals required and n denotes original sample size; for incidence-based data, $x^* = m/t$ where m denotes the additional number of samples required and t denotes the original number of samples. A suggested initial value is $x_0 = 1$. If steps do not converge, then subsequently try $x_0 = 2$, and then $x_0 = 3$ …etc., until you obtain a convergent value for x^* in cell B27 and a converged value for m in cell B28. The additional number of individuals or samples needed is shown in cell B28.

(3) To run the second part (g < 1), key in a value for g in data entry cell B32. There is an explicit solution as long as the condition $gS_{est} > S_{obs}$ is satisfied (i.e., $g > S_{obs}/S_{est}$, this range is shown in cell B33 entry after data input). The additional number of individuals or samples required will appear in cell B35.

NOTE: For the case of $g = 1$, a figure is plotted in the Excel sheet in order to assure that the converged value x^* is the correct point at which $h(x) = v(x)$. This plot helps to examine how the two functions $h(x)$ and $v(x)$ behave in the range of a user-specified *x_min* (data entry cell E5) and *x_max* (data entry cell E6). Choose a range that covers the converged value obtained from iterations (from cell B27). If the figure does not show a clear intersection point, then narrow the range by adjusting *x_min* and *x_max* to yield a clearer picture. Then one will be able to check whether the converged value is correct. The numerical values in columns E and F are also useful for double-checking.

Two examples (from Tables 1 and 2 in the main text) demonstrate the above procedures:

Example 1: (incidence data in Table 2 of the main text)

Habitat, Taxon, and Locale	t	T	S_{obs}	S_{est}	Q_1	Q_2
Hedgerow carabid beetles in soil, litter and vegetation samples (Great Britain)	16	72	20	22	4	5

Example 2: (abundance data in table 1 of the main text)

Habitat, Taxon, and Locale	n	S_{obs}	S_{est}	f_1	f_2
Forest lizards in pitfall traps (U.S.)	161	9	11	3	2

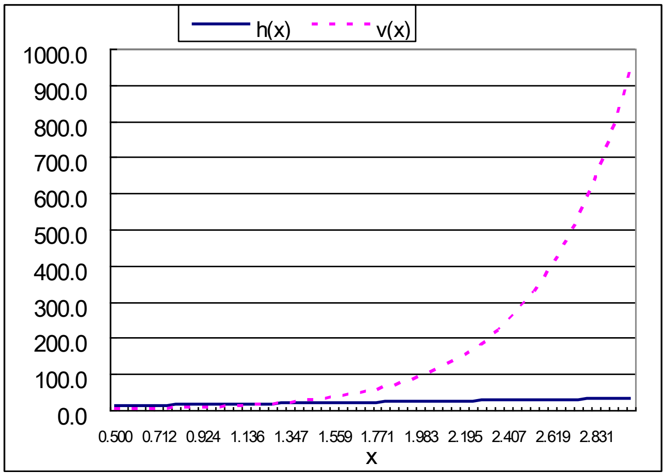
Example 1: (incidence data)

(1) Select the *Incidence Data* sheet and enter the required data (key in 4, 5, 16, and 20, respectively, in data entry cells B5, B6, B7, B8 in the Excel worksheet). If the total number of incidences is known, you may key in it in data entry cell B4 so that the value of q_0 can be computed. For these data, we have T = 72.

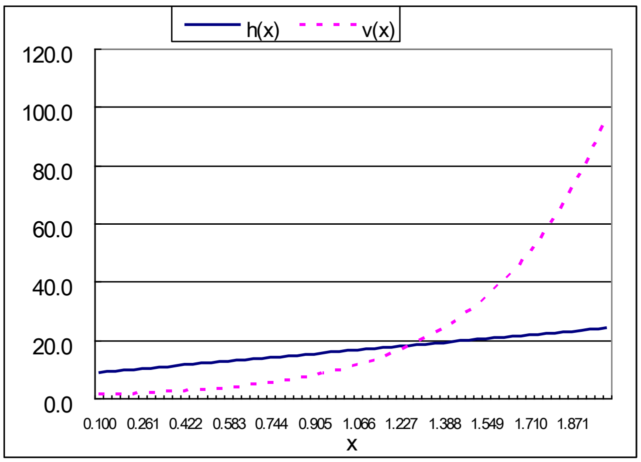
(2) As suggested, use an initial value of x_0 = 1 for starting iteration (in data entry cell B15). The iterations do converge to a value of x^* = 1.268. This implies that the required additional number of *samples* is thus $m = tx^*$ = 20.29. In other words, 20 additional samples are required to observe the estimated S_{est} = 21.5 species (i.e., to observe two previously undetected species).

(3) In this example, S_{est} = 21.5. Thus in the second part (g < 1), we must restrict $g > S_{obs}/S_{est}$ = 20/21.5 = 0.930. Therefore, for these data, it is meaningless to input a value g < 0.93. For g = 0.95, key-in 0.95 in data entry cell B32, yielding the answer 2.16 in cell B35.

NOTE: To make sure the that the converged value is the correct point at which $h(x) = v(x)$, suppose one inputs $x\_ min$ = 0.5 (in data entry cell E5), $x\_ max$ = 3 (in data entry cell E6). The resulting plot for the range of (0.5, 3) is shown below. However, this plot does *not* clearly show the intersection point.



Adjust the range of the figure, say, $x\_min = 0.1$ (in data entry cell E5) and $x\_max = 2$ (in data entry cell E6). Now you will see a detailed plot in the range (0.1, 2), as shown below. It is clearly seen that the correct intersection does occur at $x^* = 1.268$. One can further check this result by examining the numerical values for the two functions in columns E and F of the worksheet. Thus, we can assure that $x^* = 1.268$ and $m = tx^* = 20.29$ are the correct answers.



Example 2: (abundance data)

(1) Select the *Abundance Data* sheet and enter the required data (key in 3, 2, 161, 9, respectively, in data entry cells B5, B6, B7, B8 in the Excel worksheet).
(2) Try an initial value of $x_0 = 1$ for starting iteration (in data entry cell B15). However, the iterations do *not* converge. Try another initial value of $x_0 = 2$, yielding the convergent value of $x^* = 2.221$. This result indicates that the required additional number of *individuals* is $m = nx^* = 357.6$. That is, 358 additional individuals are needed to observe $S_{est} = 11.25$ species (i.e., to observe two previously undetected species).
(3) In this example, $S_{est} = 11.25$. Thus in the second part ($g < 1$), we must restrict $g > S_{obs}/S_{est} = 9/11.25 = 0.8$. If we enter 0.9 in data entry cell B32, then the answer 84 appears in cell B35. If we enter 0.95 in data entry cell B32, we get the answer 167 in cell B35.

NOTE: To make sure that the converged value x^* = 2.221 is the correct point at which $h(x)$ = $v(x)$, we may set x_min = 1.5 (in data entry cell E5), and x_max = 2.5 (in data entry cell E6). The resulting detailed plot in the range (1.5, 2.5) is shown below. It is clear that the correct intersection does occur at x^* = 2.221.