It is often required to define functions that estimate or merely summarize the entropy or degree of randomness in the distribution of observations into categories.

Typically there would be n observations in total divided into k categories with nonnegative frequencies f_i, for $i = 1, 2, \ldots, k$, so that $n = \sum_{i=1}^{k} f_i$, as for example with this extreme sample where there is clearly no evidence of differences between the groups

Group	Frequency
1	5
2	5
3	5
4	5

or this equally extreme example where the non-homogeneity is obvious.

Group	Frequency
1	1
2	1
3	1
4	17

To analyze such data choose [Statistics] from the main SimFiT menu then [Statistical calculations] followed by selecting [Shannon/Brilllouin indices of diversity]. Note that data sets must be supplied as samples of frequencies and these can be as vector files, or columns of frequencies pasted in from the clipboard. However, it is often the case when repetitive analysis of small data sets is required, that it is useful to temporarily switch off the speed-up option that suppresses input from the terminal and simply enter the frequencies manually.

Of course such problems arise constantly in data analysis but especially in ecology where several well known indices of diversity are used, as illustrated in the next table for these two extreme cases.

Data: 5,5,5,5
Number of groups	4
Total sample size	20
Pielou J' evenness	1.0000 [complement = 0.0000]
Brillouin J evenness	1.0000 [complement = 0.0000]
Shannon H'	$0.6021(\log_{10})$ $1.386(\log_e)$ $2.000(\log_2)$
Brillouin H	$0.5035(\log_{10})$ $1.159(\log_e)$ $1.672(\log_2)$
Simpson λ	0.2500 [complement = 0.7500]
Simpson λ'	0.2105 [complement = 0.7895]

Data: 1,1,1,17
Number of groups	4
Total sample size	20
Pielou J' evenness	0.4238 [complement = 0.5762]
Brillouin J evenness	0.3809 [complement = 0.6191]
Shannon H'	$0.2551 (\log_{10})$ $0.5875(\log_e)$ $0.8476(\log_2)$
Brillouin H	$0.1918 (\log_{10})$ $0.4415(\log_e)$ $0.6370(\log_2)$
Simpson λ	0.7300 [complement = 0.2700]
Simpson λ'	0.7158 [complement = 0.2842]

Definitions

Given positive integer frequencies $f_i > 0$ in $k > 1$ groups with n observations in total, then proportions $p_i = f_i/n$ can be defined, leading to the Shannon H', Brillouin H, and Simpson λ and λ' indices, and the evennness parameters J and J' defined as follows.

$$\text{Shannon diversity } H' = -\sum_{i=1}^{k} p_i \log p_i$$

$$= [n \log n - \sum_{i=1}^{k} f_i \log f_i]/n$$

$$\text{Pielou evenness } J' = H'/\log k$$

$$\text{Brilloin diversity } H = [\log n! - \log \prod_{i=1}^{k} f_i!]/n$$

$$\text{Brilloin evenness } J = nH/[\log n! - (k-d)\log c! - d\log(c+1)!]$$

$$\text{Simpson lambda } \lambda = \sum_{i=1}^{k} p_i^2$$

$$\text{Simpson lambda prime } \lambda' = \sum_{i=1}^{k} f_i(f_i - 1)/[n(n-1)]$$

where $c = [n/k]$ and $d = n - ck$. Note that H and H' are given using logarithms to bases ten, e, and two, while the forms J and J' have been normalized by dividing by the corresponding maximum diversity and so are independent of the base. The complements $1 - J$, $1 - J'$, $1 - \lambda$, and $1 - \lambda'$ are also tabulated within the square brackets.

In the above tables we see that evenness is maximized when all categories are equally occupied, so that $f_i = 1/k$ and $H' = \log k$, and is minimized when one category dominates.

Of course SimF$_I$T provides numerous techniques to test hypotheses about the distribution of frequencies into groups, e.g. a χ^2 test on observed and expected frequencies.