

Ecography

E6860

Dauby, G. and Hardy, O. J. 2011. Sampled-based estimation of diversity sensu stricto by transforming Hurlbert diversities into effective number of species. – *Ecography* 34: xxx–xxx.

Supplementary material

Appendix 1. Hurlbert diversities and Hurlbert's Effective Number of Species asymptotically satisfy the replication principle

Let π_i be the actual relative abundance of species i in a community of S species, then

Hurlbert's diversity is defined as:

$${}^k S = \sum_i^S [1 - (1 - \pi_i)^k] = \sum_i^S [1 - \exp(k \ln(1 - \pi_i))]$$

In a species-rich community, $1 \gg \pi_i > 0$, such that $\ln(1 - \pi_i) \cong -\pi_i$. Hence,

$${}^k S \cong \sum_i^S [1 - \exp(-k\pi_i)]$$

If each species is subdivided into x equi-frequent subspecies,

$${}^{xk} S' \cong x \sum_i^S [1 - \exp(-xk\pi_i/x)] \cong x {}^k S$$

where S' refers to subspecies and S to species. Hence, the replication principle (or doubling property) applies asymptotically (i.e. when π_i tends to 0 for all species, thus for high diversity communities) to Hurlbert diversities provided that k is multiplied by x .

By definition, Hurlbert's ENS (${}^k E$) is related to Hurlbert diversities by

$${}^k S = {}^k E (1 - (1 - 1/{}^k E)^k)$$

Using the same approximation as above ($1 \gg \pi_i > 0$, thus ${}^k E \gg 1$),

$${}^k S \cong {}^k E [1 - \exp(k \ln(1 - 1/{}^k E))] \cong {}^k E [1 - \exp(k/{}^k E)]$$

At the subspecies level,

$${}^{xk} S' \cong {}^{xk} E' [1 - \exp(xk/{}^{xk} E')] \cong x {}^k S \cong x {}^k E [1 - \exp(k/{}^k E)]$$

which implies that ${}^{kx} E' \cong x \cdot {}^k E$. Hence, the replication principle also applies asymptotically to the Effective Number of Species of Hurlbert diversities provided that k is multiplied by x .

As these demonstrations rely on an approximation, we investigated to what extent the doubling property is valid for Hurlbert's ENS. This could not be done analytically because ${}^k E$ is estimated by an iterative approach so we created artificial datasets containing up to 1000 species and checked a large array of possible SAD, including a completely arbitrary SAD, to identify which conditions generate the highest error. We evaluated the error due to our approximation as $({}^{2k} E' / {}^k E - 2) / 2$ where ${}^{2k} E'$ is Hurlbert's ENS after subdividing each species into two equi-frequent subspecies. The approximation error always decreased with k and was highest when one species dominated and all other species were rare and equi-frequent. The case where $k=2$ was not considered because the doubling property here applies exactly without adjusting parameter k . For $k=3$, the highest approximation error found was 8.6%, upward. This upward error shrank to 1.7% and 0.3% for $k=25$ and 100, respectively. For log-series SAD with 10 000 individuals, the highest approximation error was observed again when one species dominated and the other species were rare, and it reached 4%, 0.5% and 0.05% for $k=3, 25$ and 100, respectively. A negative error was sometimes observed, but the absolute error in these cases was always very small.

Appendix 2. Hill numbers, estimators of Hill numbers and estimators of Fisher's alpha

Hill numbers

Hill numbers of order q is defined as (Hill 1973; Jost 2007):

$${}^q D = \left(\sum_{i=1}^S \pi_i^q \right)^{1/(1-q)},$$

An interpretation of Hill numbers is that it is the inverse of the generalized mean of relative abundance. Parameter q defines which kind of mean it is: it becomes the harmonic mean when $q=0$, the geometric mean when q approaches 1 and the arithmetic mean when $q = 2$ (Hill 1973, Tuomisto 2010).

For specific values of q , Hill numbers can also be interpreted in terms of classical indices of diversity: ${}^{q=0}D$ is species richness S , ${}^{q=1}D$ is not defined but the limit for $q \rightarrow 1$ equals the exponential of Shannon-Wiener's index of diversity, ${}^{q=2}D$ equals one form of Simpson's diversity index. Hence:

$${}^{q=0} D = S$$

$${}^{q=1} D = \exp\left(-\sum_{i=1}^S \pi_i \log \pi_i\right)$$

$${}^{q=2} D = 1 / \sum_{i=1}^S \pi_i^2$$

Hill numbers correspond to Effective Number of Species, i.e. they quantify the number of equally frequent virtual species that would produce the value of a given diversity index (which is determined by q). Therefore, diversity indices such as Shannon-Wiener's and Simpson's indices, which are not comparable because they do not quantify the same thing, become strictly comparable once transformed into ENS. They simply differ in their sensitivity

to rare vs dominant species which is defined by parameter q : the higher parameter q , the less rare species are taken into account.

In these definitions, π_i corresponds to actual relative species abundances in the community.

Replacing π_i by their sample-based estimates, p_i , provides biased estimators of Hill numbers.

Corrected estimators of Hill numbers

Corrected (less biased) estimators are denoted by adding the symbol $\hat{\cdot}$. It has been shown by several authors that the Chao1 abundance-based estimator is among the most accurate estimators of species richness (see review in Walther and Moore (2005)). Hence,

$${}^{q=0}\hat{D} = S_{obs} + a_1(a_1 - 1)/[2(a_2 + 1)]$$

where a_1 is the number of species represented by 1 individual in the sample (singletons) and a_2 the number of species represented by 2 individuals in the sample and S_{obs} is the observed specific richness in the sample.

To estimate ${}^{q=1}D$, following Jost (2007), we used the exponential of the estimator of Shannon-

Wiener's index developed by (Chao and Shen 2003), \hat{H} :

$${}^{q=1}\hat{D} = \exp(\hat{H})$$

The estimator of ${}^{q=2}D$ is the ENS of the unbiased estimator of Gini-Simpson's index which is defined as follows:

$$\hat{\lambda} = \frac{W}{(W - 1)} \left[1 - \sum_{i=1}^S p_i^2 \right]$$

where W is the sample size. Hence,

$${}^{q=2}\hat{D} = 1/(1 - \hat{\lambda})$$

However, Nielsen et al (2003) showed that since it is not a linear transformation of $\hat{\lambda}$, ${}^{q=2}\hat{D}$

is not unbiased. They proposed another less biased estimate of ${}^{q=2}D$ as follows:

$${}^{q=2}\hat{D}_{Nielsen} = \frac{(W-1)^2}{\sum_{i=1}^S p_i^2 (W+1)(W-2) + 3 - W}$$

Estimators of Fisher's alpha

Fisher's alpha, α , is a parameter of a particular distribution of species abundances, the log-series. The number of species for an abundance of n individuals, S_n , is given by

$S_n = \alpha x^n / n$ where x is a constant ($0 < x < 1$) dependent upon the size of the sample, and the

total number of species in the sample is given by $S = \alpha \ln(1 + W / \alpha)$ (Fisher *et al.* 1943).

The first estimator of Fisher's alpha used followed Kempton and Taylor (1974) and is obtained through an iterative procedure considering only W and S . The second estimator of Fisher's alpha corresponds to the estimator of the biodiversity number of a neutral species abundance distribution developed by Etienne (2005). Here, the neutral SAD is equivalent to a log-series SAD in the special case where there is no dispersal limitation (immigration rate $m = 1$) (Hubbell 2001).

References

Chao, A., and T. J. Shen. 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. - *Environmental and Ecological Statistics* 10:429-443.

Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. - *Ecology Letters* 8:253-260.

Fisher, R. A., A. S. Corbet, et al. 1943. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. - *Journal of Animal Ecology* 12:42-58.

- Hill, M. O. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. - Ecology 54:427-432.
- Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. -. Princeton Univ. Press.
- Jost, L. 2007. Partitioning diversity into independent alpha and beta components. - Ecology 88:2427-2439.
- Kempton, R. A., and L. R. Taylor. 1974. Log-Series and Log-Normal Parameters as Diversity Discriminants for the Lepidoptera. - Journal of Animal Ecology 43:381-399.
- Nielsen, R., D. R. Tarpy, et al. 2003. Estimating effective paternity number in social insects and the effective number of alleles in a population. - Molecular Ecology 12:3157-3164.
- Tuomisto, H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. - Ecography 33:2-22.
- Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. - Ecography 28:815-829.

Appendix 3. Bias and precision of multiplicative beta diversity based on Effective Number of Species

Scaled mean error (SME) and scaled root mean square error (SRMSE) (Walther and Moore 2005) for multiplicative beta diversity measures applied on samples from two communities with four levels of differentiation. ${}^q\hat{\beta}$ is based on estimators of Hill numbers. ${}^k\hat{\beta}$ is based on Hurlbert's ENS. ${}^{q=2}\hat{\beta}$ is strictly identical to ${}^{k=2}\hat{\beta}$ and is here computed using Nielsen's estimator (2003). For each sample size and each level of differentiation, the minimal absolute value is given in bold.

		SME				SRMSE			
		1000	500	100	50	1000	500	100	50
Full differentiation	${}^{q=0}\hat{\beta}$	-0.003	-0.003	0.001	0.011	-0.003	-0.003	0.001	0.011
	${}^{q=1}\hat{\beta}$	-0.007	-0.012	-0.016	-0.018	-0.007	-0.012	-0.016	-0.018
	${}^{q=2}\hat{\beta} = {}^{k=2}\hat{\beta}$	-0.002	-0.004	-0.017	-0.019	-0.002	-0.004	-0.017	-0.019
	${}^{k=100}\hat{\beta}$	0.000	0.000	0.003	-	0.000	0.000	0.003	-
100 species shared	${}^{q=0}\hat{\beta}$	0.068	0.075	0.096	0.104	0.102	0.145	0.248	0.365
	${}^{q=1}\hat{\beta}$	0.007	0.011	-0.003	-0.019	0.017	0.029	0.063	0.135
	${}^{q=2}\hat{\beta} = {}^{k=2}\hat{\beta}$	-0.003	-0.002	-0.014	-0.012	0.036	0.051	0.138	0.219
	${}^{k=100}\hat{\beta}$	0.001	0.001	0.003	-	0.039	0.023	0.078	-

300 species shared		0.288	0.317	0.322	0.341	0.336	0.373	0.440	0.508
		0.022	0.031	0.057	0.051	0.032	0.046	0.109	0.163
	=	-0.002	-0.005	-0.019	-0.020	0.031	0.046	0.120	0.204
		-0.001	-0.001	0.005	-	0.013	0.021	0.090	-
Identical communities		0.419	0.459	0.527	0.627	0.353	0.389	0.488	0.631
		0.036	0.062	0.137	0.113	0.037	0.064	0.151	0.174
	=	-0.002	-0.005	-0.015	-0.013	0.006	0.013	0.068	0.135
		0.000	0.000	0.009	-	0.004	0.009	0.065	-

References

- Nielsen, R., D. R. Tarpay, et al. 2003. Estimating effective paternity number in social insects and the effective number of alleles in a population. - *Molecular Ecology* 12:3157-3164.
- Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. - *Ecography* 28:815-829.

Appendix 4.

Scaled mean error (SME) and scaled relative mean square error (SRMSE) for the three species abundance distributions (SAD) and the four different sample sizes (250, 100, 50, 25) in the different diversity measures studied. For each sample size and each SAD, the minimal value is given in bold. Similar to Table 2 but supplemented by two estimators of Fisher's alpha (see Appendix 2 for definition).

Sample size		SME				SRMSE				
		250	100	50	25	250	100	50	25	
Neutral SAD $\theta = 100$	Estimators of Hill numbers	$q=0 D$	-0.737	-0.855	-0.915	-0.952	0.738	0.855	0.915	0.952
		$q=1 D$	-0.431	-0.638	-0.769	-0.864	0.433	0.639	0.770	0.864
		$q=2 D$	-0.209	-0.404	-0.583	-0.741	0.227	0.412	0.586	0.742
	Corrected estimators of Hill numbers	$q=0 \hat{D}$	-0.610	-0.701	-0.764	-0.824	0.447	0.572	0.648	0.725
		$q=1 \hat{D}$	-0.175	-0.269	-0.322	-0.193	0.193	0.298	0.394	0.768
		$q=2 \hat{D}$	0.019	0.070	0.199	0.491	0.158	0.300	0.595	1.200
	Hill number $q=2$ Hurlbert's ENS $k=2$	$q=2 \hat{D}_{Nielsen}$	0.016	0.049	0.097	0.074	0.153	0.271	0.474	0.698
	Hurlbert's ENS	$k=25 \hat{E}$	0.008	0.038	0.103	0.124	0.123	0.230	0.476	0.429
	Estimators of Fisher's alpha	Kempton and Taylor (1974)	-0.072	-0.107	-0.068	0.098	0.121	0.215	0.372	0.742
		Etienne (2005)	-0.068	-0.145	-0.498	-0.748	0.119	0.196	0.498	0.748

Appendix 4 (continued)

Neutral SAD $\theta = 10$	Hill numbers	$q=0$ D	-0.540	-0.662	-0.738	-0.813	0.541	0.663	0.738	0.814
		$q=1$ D	-0.109	-0.207	-0.329	-0.480	0.123	0.221	0.340	0.486
		$q=2$ D	-0.051	-0.119	-0.209	-0.353	0.087	0.160	0.240	0.370
	Estimators of Hill number	$q=0$ \hat{D}	-0.410	-0.557	-0.634	-0.716	0.612	0.706	0.772	0.832
		$q=1$ \hat{D}	-0.023	-0.060	-0.094	-0.122	0.071	0.121	0.177	0.286
		$q=2$ \hat{D}	0.010	0.013	0.054	0.108	0.083	0.139	0.222	0.409
	Hill number $q=2$ Hurlbert's ENS $k=2$	$k=2$ $\hat{D}_{Nielsen}$	0.006	0.015	0.032	0.049	0.081	0.141	0.211	0.372
	Hurlbert's ENS	$k=25$ \hat{E}	0.003	0.008	0.025	0.014	0.063	0.097	0.145	0.156
	Estimators of Fisher's alpha	Kempton and Taylor (1974)	-0.006	0.044	0.143	0.280	0.103	0.160	0.280	0.529
		Etienne (2005)	-0.017	0.010	0.074	0.259	0.105	0.158	0.240	0.503
Triangular SAD	Hill numbers	$q=0$ D	-0.269	-0.531	-0.714	-0.842	0.271	0.531	0.714	0.842
		$q=1$ D	-0.255	-0.492	-0.677	-0.815	0.257	0.493	0.677	0.815
		$q=2$ D	-0.286	-0.509	-0.671	-0.804	0.289	0.510	0.671	0.804
	Estimators of Hill number	$q=0$ \hat{D}	-0.114	-0.175	-0.214	-0.262	0.135	0.238	0.380	0.527
		$q=1$ \hat{D}	0.021	0.018	0.045	0.458	0.065	0.146	0.349	1.331
		$q=2$ \hat{D}	0.007	0.030	0.089	0.279	0.075	0.176	0.389	0.812
	Hill number $q=2$ Hurlbert's ENS $k=2$	$k=2$ $\hat{D}_{Nielsen}$	0.001	0.002	0.003	-0.067	0.072	0.158	0.320	0.526
	Hurlbert's ENS	$k=25$ \hat{E}	0.005	0.018	0.094	0.089	0.068	0.154	0.371	0.372
	Estimators of Fisher's alpha	Kempton and Taylor (1974)	1.834	2.722	3.538	4.675	1.846	2.807	3.947	5.748
		Etienne (2005)	1.826	2.622	1.158	0.080	1.838	2.674	1.158	0.080

Appendix 5.

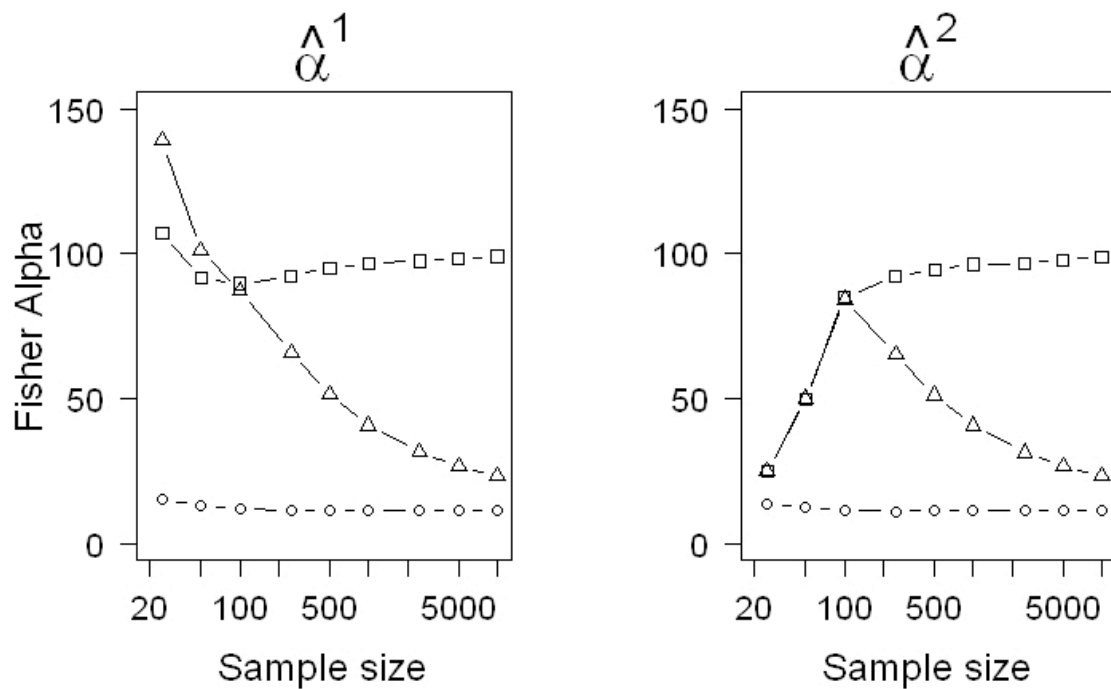
Bias profiles for Fisher's alpha diversity measures: diversity values in relation to sample size for two communities following a log-series distribution with Fisher's alpha equal to 100 (\square)

and 10 (\circ), and a third community following a triangular species abundance distribution (Δ).

An unbiased measure of diversity corresponds to a horizontal line, i.e. even the smallest sample size will give the true diversity value, which here is the diversity of the entire

community of ca. 10 000 individuals (the first value on the right of each graph). Definitions of

the two estimators are given in Appendix 2.



Appendix 6. Mathematical relationship between Hill numbers and Hurlbert ENS as provided by an anonymous referee.

As demonstrated in the article, Hurlbert diversities kS can be transformed into Hurlbert ENS through the following iterative equation

$${}^k\hat{E}_t = {}^k\hat{S} / [1 - (1 - 1/{}^k\hat{E}_{t-1})^k] \quad (1)$$

with an initial value ${}^k\hat{E}_{t=0} = {}^k\hat{S}$.

Hurlbert diversity kS can be expressed as

$${}^kS = \sum_{i=1}^S (1 - (1 - \pi_i)^k) = \sum_{i=1}^S \left[1 - \sum_{q=0}^k \binom{k}{q} (-\pi_i)^q \right] \quad (2)$$

When $q=0$, $\binom{k}{q} (-\pi_i)^q = \binom{k}{0} (-\pi_i)^0 = 1$, thus (2) becomes

$${}^kS = \sum_{i=1}^S \left[- \sum_{q=1}^k \binom{k}{q} (-\pi_i)^q \right] = \sum_{i=1}^S \sum_{q=1}^k \binom{k}{q} (-1)^{q+1} \pi_i^q \quad (3)$$

When $q=1$, $\sum_{i=1}^S \binom{k}{q} (-1)^{q+1} \pi_i^q = \sum_{i=1}^S \binom{k}{1} (-1)^2 \pi_i^1 = k$, thus (3) becomes

$${}^kS = k + \sum_{i=1}^S \sum_{q=2}^k \binom{k}{q} (-1)^{q+1} \pi_i^q = k + \sum_{q=2}^k \binom{k}{q} (-1)^{q+1} ({}^rD)^{1-q} \quad (4)$$

where ${}^qD = \left(\sum_i \pi_i^r \right)^{1/(1-q)}$ and q is the order of Hill numbers

Thus, kS is a function of ${}^2D, {}^3D, \dots, {}^kD$.

For $t=1$ in the iterative equation,

$${}^kE_{t=1} = {}^kS / [1 - (1 - 1/{}^kE_{t=0})^k] = {}^kS / [1 - (1 - 1/{}^kS)^k]$$

Since kS is a function of ${}^2D, {}^3D, \dots, {}^kD, {}^kE_{t=1}$ as well.