Ecological Monographs, 84(1), 2014, pp. 45–67

Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies

Anne Chao,[1,6] Nicholas J. Gotelli,[2] T. C. Hsieh,[1] Elizabeth L. Sander,[2] K. H. Ma,[1] Robert K. Colwell,[3,4] and Aaron M. Ellison[5]

[1]Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043 Taiwan
[2]Department of Biology, University of Vermont, Burlington, Vermont 05405 USA
[3]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269 USA
[4]University of Colorado Museum of Natural History, Boulder, Colorado 80309 USA
[5]Harvard University, Harvard Forest, 324 North Main Street, Petersham, Massachusetts 01366 USA

Abstract. Quantifying and assessing changes in biological diversity are central aspects of

perspective, species richness is very difficult to estimate accurately from a finite sample.

A second problem with species richness as a measure of biodiversity is that it does not incorporate any

which places more weight on the frequencies of abundant species and discounts rare species. Investigators using Hill numbers should report, at least, the diversity of all species ($q = 0$), of "typical" species ($q = 1$), and of dominant species ($q = 2$). For a general order of q, if ${}^qD = x$, then the diversity is equivalent to that of an idealized assemblage with x equally abundant species, which is why Hill numbers are referred to as effective numbers of species or as species equivalents.

A complete characterization of the species diversity of an assemblage with S species, and relative abundances ($p_1, p_2, \ldots, p_S$) is conveyed by a diversity profile (a plot of ${}^qD$ vs. q from $q = 0$ to $q = 3$ or 4 [beyond this it changes little]; see Tóthmérész 1995). Although Hill numbers for $q < 0$ can be calculated, they are dominated by the frequencies of rare species and have poor statistical sampling properties. We thus restricted ourselves to the case $q \geq 0$ throughout the paper. An example of a diversity profile is shown in Fig. 1a.

Hill numbers can be regarded as the theoretical or asymptotic diversities at a sample size of infinity for which the true relative abundances $\{p_1, p_2, \ldots, p_S\}$ of each of i species are known. When sample size is relevant for discussion, we use the notation ${}^qD(\ )$ to denote the (asymptotic) Hill numbers. Throughout the paper, we use ${}^qD$

$$E[f_k(m)] = \sum_{i=1}^{S} \binom{m}{k} p_i^k (1 - p_i)^{m-k} \qquad k = 0, 1, \ldots, m. \tag{5}$$

Note that $E[f_0(m)] = \sum_{i=1}^{S}$

frequency counts. This idea can be extended easily to Hill numbers of any orders, as we next illustrate.

The extrapolated species richness estimator for a sample of n þ m* used in this paper is reviewed in Appendix B, and the formula (originally derived by Shen et al. 2003) is shown in Table 1. This approach requires an estimated asymptote of species richness. Any proper species richness estimator can be used. Colwell et al. (2012) suggested using the Chao1 estimator (Chao 1984) or abundance-based coverage estimator (ACE;

where A ¼ 2f$_2$/[(n □1)f$_1$ þ 2f$_2$]. As a result, the asymptotic estimator for Shannon diversity is $^{1}\hat{D}(\,`\,) = \exp[\hat{H}(\,`\,)]$. The extrapolated estimator for Shannon diversity of a sample of size n þ m* is as follows:

$$^{1}\hat{D}_{n+m^*} = \exp\left[\frac{n}{n+m^*}\sum_{i=1}^{S} \square \overline{X}\right.$$

or all sampling units). An unconditional variance measures the variation in diversity that would arise if another new sample of size m were taken from the entire assemblage (rather than from the original reference sample). Therefore, the unconditional variance does not approach 0 when sample size tends to n, and all associated confidence intervals are symmetric, which reflects the uncertainty of the new sample. In deriving an ''unconditional'' variance, the number of undetected species must be estimated because those undetected species also affect the variation of a new sample. In most applications, unconditional variance is more useful because inferences are not restricted to the reference sample.

Colwell et al. (2012) obtained an unconditional analytic variance estimator for rarefied and extrapolated species richness estimators. However, extending this analytic approach for variance estimators to a general order of q becomes mathematically intractable. Therefore, we suggest a simpler, bootstrap method (Appendix G), to obtain unconditional variances and confidence intervals for all rarefied and extrapolated estimators. In the proposed procedure, we follow Colwell et al. (2012) and use the Chao1 (for abundance data) or Chao2 (for incidence data) to estimate the number of undetected species in the reference sample (Chao 1984, 1987), although any other proper estimators can also be used. The examples in Worked examples: comparison of assemblages illustrate our proposed sampling curves and the associated confidence intervals based on the unconditional variance from our proposed bootstrap method.

Sample-size- and coverage-based rarefaction and extrapolation

In comparing diversities among multiple assemblages, samples can be standardized by sample size or by sample completeness. Our proposed sample-size-based sampling curve for Hill numbers of each specific order q includes the rarefaction part (which plots ${}^{q}\hat{D}(m)$ as a function of m, where $m \leq n$; see Table 1) and the extrapolation part (which plots ${}^{q}\hat{D}(n + m^*)$ as a function of $n + m^*$ for $m^* > 0$; see Table 1) and yields a smooth sampling curve, the two parts of which join smoothly at the point of the reference sample $(n, {}^{q}D_{obs})$. To fully incorporate the effect of relative abundance on diversity estimation, we suggest plotting curves for at least the first three Hill numbers ($q = 0, 1, 2$).

When there are many ''invisible'' species (species with extremely small relative abundance that are almost undetectable in normal sampling schemes) our intuition is that the number of undetected species in samples (or equivalently, species richness in the entire assemblage) is very hard to estimate; see Colwell et al. (2012) and Gotelli and Chao (2013) for a review. On the other hand, and contrary to intuition, the notion of sample

plots ${}^{q}\hat{D}(n \text{ þ } m^*)$ with respect to $\hat{C}_{ind}(n \text{ þ } m^*)$) join smoothly at the reference point ($\hat{C}_{ind}(n)$, ${}^{q}D_{obs}$). The confidence intervals of expected diversity based on the bootstrap method also join smoothly.

Bridging sample-size- and coverage-based approaches

The sample-size-based approach plots the estimated diversity as a function of sample size, whereas the corresponding coverage-based approach plots the same diversity with respect to sample coverage. Therefore, these two approaches can be bridged by the relationship between coverage and sample size. Using the coverage estimators in Tables 1 and 2 (the last row in each table), we can construct a sample completeness curve, which reveals sample completeness for a given sample size. From the original reference sample, this curve estimates sample completeness for smaller rarified samples, as well as for larger extrapolated samples. This curve also provides an estimate of the sample size needed to achieve a fixed degree of completeness.

An optimal stopping theory derived by Rasmussen and Starr (1979) specifies that sampling stops when sample coverage reaches a predetermined value. The

TABLE 2. Extended.

Extrapolation estimator (for $T + t^*$ sampling units)¶
$\hat{m}^{0}(T + t^*) = S_{obs} + \hat{Q}_0 \cdot \left[1 - \left(1 - \frac{Q_1}{T\hat{Q}_0 + Q_1}\right)^{t^*}\right]$ (reliable if $t^* < T$)
$\hat{m}^{1}(T + t^*) = \exp\left[\frac{T}{T + t^*}\sum_{i=1}^{S}\left(-\frac{Y_i}{U}\log\frac{Y_i}{U}\right) + \frac{t^*}{T + t^*}\hat{H}_{sample}(\cdot +)\right]$ (nearly unbiased)
$\hat{m}^{2}(T + t^*) = \frac{1}{\sum_{i=1}^{S}\left(\frac{Y_i}{U}\right)^2 - \frac{1}{U}\left(\frac{T}{T + t^*}\right)\sum_{i=1}^{S}\left(\frac{Y_i}{U}\right)\left(\frac{Y_i - 1}{U - 1}\right)}$

numbers of $q = 0, 1, 2$. In Fig. 3a, we show these sample-size-based curves with 95% confidence intervals based on a bootstrap method. We extrapolated up to double the reference sample size (i.e., up to size 336 for the girdled treatment and size 504 for the logged treatment). In each plot, except for initial, small sample sizes, none of the confidence intervals for the three curves intersect, and the rank order of diversity is species richness > Shannon diversity > Simpson diversity. For any fixed sample size or completeness in the comparison range, if the 95% confidence intervals do not overlap, then significant differences at a level of 5% among the expected diversities (whether interpolated or extrapolated) are guaranteed. However, partially overlapping intervals do not guarantee nonsignificance (Schenker and Gentleman 2001). The curve for species richness ($q = 0$) increases steeply with sample size in both treatments, but the curves for Shannon and Simpson diversity ($q = 1$ and $q = 2$) level off beyond the reference sample, illustrating that higher order Hill numbers are increasingly dominated by the frequencies of the more common species and are, therefore, less sensitive to sampling effects.

To compare diversities between the girdled and logged treatments, we show in Fig. 3b, for each fixed value of q ($q = 0$, 1, and 2), the sample-size-based rarefaction and extrapolation of these two plots with 95% confidence intervals up to a base sample size. We suggest the base sample size to be double the smallest reference sample size or the maximum reference sample size, whichever is larger (the reason for our suggestion will become clearer in the second example). See Box 1 for systematic steps to determine a base sample size. In this example, the base sample size is 336 (double the smaller reference sample size). The estimated Hill numbers can then be compared across assemblages for any sample size less than the base size. In a traditional rarefaction, the data from the logged treatment would be rarefied to a sample size of 168 individuals to match the abundance in the girdled treatment. For this rarefied sample, the Hill numbers of $q = 0, 1, 2$ are estimated to be 31.71, 13.83, and 6.68, respectively. The proposed integrated sampling curve allows reliable comparisons for any sample size up to an abundance of 336. Across this range of abundance, Fig. 3b reveals that the logged treatment is more diverse for all but the smallest sample sizes for species richness ($q = 0$) and Shannon diversity ($q = 1$), although the confidence intervals overlap. In contrast, for Simpson diversity ($q = 2$), the girdled treatment is more diverse, although again the two confidence intervals overlap.

1a413e02184931270313neal39702007970219

Step 2: Construct a sample completeness curve to link sample-size- and coverage-based sampling curves (Fig. 4).—Based on Eq. 12, the coverage for the girdled treatment is estimated as 93% for the reference sample of size 168 individuals, and the coverage for the logged treatment is 94% for the reference sample of 252 individuals. It is informative to examine how the sample completeness varies with sample size (see the formulas in the last row in Table 1). In Fig. 4, we plot the sample completeness curve as a function of sample size for each of the two treatments, up to double the reference sample size. For any sample size less than 168, the curve shows that the sample completeness for the girdled treatment is estimated to be higher than that in logged treatment, although the confidence intervals overlap. When sample size is larger than 168, the estimates of sample coverages

for the two treatments differ little. If we apply a traditional rarefaction approach to standardize sample coverage, a sample size of ; 168 individuals in the logged treatment gives a sample coverage of 93%. Thus, the diversity ordering of the two treatments for 93% of the assemblage individuals is the same as that for a standardized sample of 168 individuals. The sample completeness curve figure provides a bridge between sample-size- and coverage-based sampling curves, as will be explained in the next step.

Step 3: Compare coverage-based sampling curves up to a "base coverage" (Fig. 5).—From the sample completeness curve (Fig. 4), when sample size in the girdled treatment is doubled from 168 to 336 individuals, the sample coverage is increased from 93% to 96%. In the logged treatment, when sample size is doubled from 252 to 504 individuals, the coverage is increased from 94% to 97%. In Fig. 5a, we present, for each treatment, the corresponding coverage-based rarefaction and extrapolation curves with 95% confidence intervals for diversity of q ¼ 0, 1, 2 when the coverage is extrapolated to the value for a doubling of each reference sample size.

In Fig. 5b, we compare the coverage-based diversities of the two treatments for q¼0 (left panel), q¼1 (middle panel), and q¼2 (right panel) up to the coverage of 96%. This is our "base coverage" (the lowest coverage for doubled reference sample sizes or the maximum coverage for reference samples, whichever is larger). See Box 1 for suggestions on the choice of base coverage. Because the increase in coverage for the extrapolation is small, and the estimated diversity for q¼1 and 2 hardly change beyond the reference samples, the extrapolation parts in Fig. 5b are nearly invisible for these two orders of q. Since the two confidence bands do not intersect for species richness (q¼0) if coverage exceeds 50% (Fig. 5b, left panel), species richness in the logged treatment is significantly higher than in the girdled treatment for any standardized sample coverage between 50% and 96%. For Shannon diversity (q ¼ 1), the logged treatment is more diverse, but the confidence bands overlap. For Simpson diversity (q ¼ 2), when coverage is less than 70%, both treatments have almost the same diversity, but when coverage is greater than 70%, the Simpson diversity for the girdled treatment is slightly higher.

Comparing Figs. 3b and 5b, we see that the sample-size- and coverage-based curves for q ¼ 0 and q ¼ 1 exhibit consistent diversity orderings between the two treatments. However, for q ¼ 2, the sample-size-based curves do not intersect (Fig. 3b), but the coverage-based curves have two crossing points (Fig. 5b). See Discussion for more comparisons of the two types of curves.

Example 2: Incidence data—comparing species diversity

choice of base sample size is that no data are excluded from our analysis. However, a drawback is that the extrapolation range for some samples could exceed their doubled reference sample sizes. For Shannon and Simpson diversities, the prediction biases are minimal beyond the double reference sample sizes, but for species richness in such cases we should be cautious about the prediction bias. See Discussion for suggestions on extrapolation range.

Next for each specific order of q, we plot the sample-size-based interpolation and extrapolation curves with 95%

Step 3: Compare coverage-based sampling curves up to a base coverage (Fig. 8).—Fig. 8a shows, for each plot, the corresponding coverage-based rarefaction and extrapolation curves for Hill numbers of $q = 0$, 1, 2 when the coverage is extrapolated to the value for a doubling of each reference sample size. From the sample completeness curve (Fig. 7), when the sample size in each site is doubled, the sample coverage increases very slightly for all sites. There is little change in ant diversity for $q = 1$ and 2. Thus, the extrapolated portions of the curves in Fig. 8a are nearly invisible, as we also noted in Fig. 5.

When all sample sizes are doubled, the minimum value of the coverage values of these doubled sample sizes among the five sites is 99.08% (for 500 m elevation). However, it is less than the coverage 99.64% of the reference sample for 2000 m elevation. In order to use all data, we select our base coverage to be 99.64% (Box 1). Fig. 8b compares coverage-based rarefaction and extrapolation curves up to the base coverage of 99.64%. All three coverage-based diversities show the same ordering by elevation as in the sample-size-based comparison. None of the confidence intervals overlap except at very small coverage values, implying significant differences in ant diversity among the five elevational transects at comparable coverage.

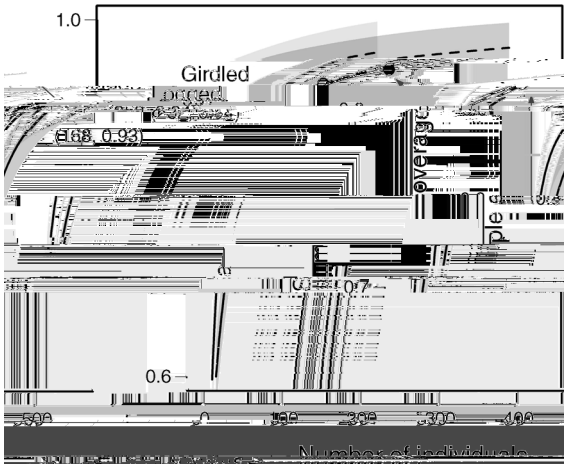

FIG. 4. Plot of sample coverage for rarefied samples (solid line) and extrapolated samples (dashed line) as a function of sample size for spider samples from the

D

FIG. 6. (a) Sample-size-based rarefaction (solid lines) and extrapolation (dashed lines, up to double the reference sample size) of tropical ant diversity from Costa Rica for Hill numbers (q ¼ 0, 1, 2) for each of the five elevations. The 95% confidence intervals were obtained by a bootstrap method based on 200 replications. Reference samples are denoted by solid dots. The numbers in parentheses are the sample size and the observed Hill numbers for each reference sample. (b) Comparison of sample-size-based rarefaction (solid line) and extrapolation (dashed line) curves with 95% confidence intervals for Hill numbers q ¼ 0 (left panel), q ¼ 1 (middle panel), and q ¼ 2 (right panel). All curves were extrapolated up to the base sample size of 599. Reference samples are denoted by solid dots. The numbers in parentheses are the sample size and the observed Hill numbers for each reference sample.

tion range. For q □ 1, the extrapolated estimator is nearly unbiased for all extrapolated sample sizes, so the extrapolation can be safely extended to the asymptote. However, for q ¼ 0, extrapolation is reliable up to no more than double the reference sample size. Beyond that, the predictor for q ¼ 0 may be subject to some bias because our asymptotic estimator for species richness (Chao1 for abundance data and Chao2 for incidence data) is a lower bound only (Chao 1984, 1987).

To compare the diversities of multiple assemblages, Box 1 gives guidelines for choosing a base sample size and base coverage for comparing sample-size- and coverage-based curves. With the suggested base sample size and base coverage, all data are used for comparisons. Based on the integrated sample-size- and coverage-based rarefaction and extrapolation curves, ecologists can efficiently use all available data to make more robust and detailed inferences about the sampled assemblages for any standardized samples with sample size less than the base sample size, and for any equally complete samples with coverage less than the base coverage. However, Example 2 provides an example in which we extrapolate a sample beyond a doubling of its reference sample size, based on the suggested base sample size. For those samples, we should be cautious in estimating quantitative differences in species richness (q ¼ 0) among assemblages, although inferences about diversities of q □ 1 are reliable.

In our formulation of a diversity accumulation curve, we define the expected diversity of a finite sample of size m as the Hill numbers based on the expected abundance frequency counts $f_E[$

the sample-size- and coverage-based standardization methods. As proved in Appendix D, the expected diversity of any order obeys a replication principle only when coverage is standardized.

In biodiversity studies, ecologists are interested in measuring not only diversity, but also evenness and inequality (Ricotta 2003). Jost (2010) used partitioning theory to derive Hill's (1973) useful class of evenness measures, the ratios of Hill numbers ^{q}D and species richness, $^{q}D/S$ for $q > 0$, and he showed that the ratio of the logarithms of Hill numbers and logarithm of richness, $\log(^{q}D)/\log(S)$, including Pielou's (1975) $J' = \log(^{1}D)/\log(S)$, express the corresponding relative evenness. These two classes of measures have been difficult to accurately estimate statistically from samples due to their strong dependence on species richness, and thus on sample size. Jost (2010) suggested estimating both S and Hill numbers at fixed coverage to obtain meaningful estimates of

In addition to Hill numbers, there are two other widely used classes of measures: Renyi and Tsallis generalized entropies (Patil and Taillie 1979, 1982).

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society B 345:101–118.
Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. Ecology 85:2717–2727.
Connell, J. H. 1978. Diversity in tropical rain forests and coral reefs. Science 199:1302–1310.
Connolly, S. R., and M. Dornelas. 2011. Fitting and empirical

Smith, W., and J. F. Grassle. 1977. Sampling properties of a family of diversity measures. Biometrics 33:283–292.
Soberón, M., and J. B. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology 7:480–488.
Terborgh, J., L. Lopez, P. Nuñez, M. Rao, G. Shahabuddin, G. Orihuela, M. Riveros, R. Ascanio, G. H. Adler, and T. D. Lambert. 2001. Ecological meltdown in predator-free forest fragments. Science 294:1923–1926.
Tipper, J. C. 1979. Rarefaction and rarefiction-the use and abuse of a method in paleoecology. Paleobiology 5:423–434.
Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. Journal of Vegetation Science 6:283–290.
Walker, S. C., M. S. Poos, and D. A. Jackson. 2008. Functional rarefaction: estimating functional diversity from field data. Oikos 117:286–296.
Washington, H. G. 1984. Diversity, biotic and similarity indices: a review with special relevance to aquatic ecosystems. Water Research 18:653–694.
Wiens, J. J., and M. J. Donoghue. 2004. Historical biogeography, ecology and species richness. Trends in Ecology and Evolution 19:639–644.

Supplemental Material

Appendix A

A binomial product model can incorporate spatial aggregation for quadrat sampling (Ecological Archives M084-003-A1).

Appendix B

Rarefaction and extrapolation for species richness (abundance data) (Ecological Archives M084-003-A2).

Appendix C

Rarefaction and extrapolation for species richness (incidence data) (Ecological Archives M084-003-A3).

Appendix D

Proof details for some formulas (Eqs. 5, 8, 9b, 11a, and 11b of the main text) and a replication principle (Ecological Archives M084-003-A4).

Appendix E

Extrapolation formulas for Hill numbers of q ¼ 1 and q □ 2 based on abundance data (Ecological Archives M084-003-A5).

Appendix F

Using simulation to test the proposed analytic estimators (Ecological Archives M084-003-A6).

Appendix G

A bootstrap method to construct an unconditional variance estimator for any interpolated or extrapolated estimator (Ecological Archives M084-003-A7).

Appendix H

Rarefaction and extrapolation of Hill numbers for incidence data (Ecological Archives M084-003-A8).

Appendix I

Hill numbers and Hurlbert's indices (Ecological Archives M084-003-A9).

Appendix J

An example: sample-size- and coverage-based Shannon diversity curves may exhibit inconsistent patterns (Ecological Archives M084-003-A10).

Appendix K

Probability of an Interspecific Encounter (PIE) and rarefaction (Ecological Archives M084-003-A11).