A Brief Introduction to iNEXT Online: software for interpolation and extrapolation of species diversity

Anne Chao, K. H. Ma, and T. C. Hsieh

Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Overview

iNEXT (iNterpolation and EXTrapolation) Online is the R-based interactive online version of iNEXT available via the link https://chao.shinyapps.io/iNEXTOnline/ or http://chao.stat.nthu.edu.tw/wordpress/software_download/. Clicking these links, you will be directed to the online interface window. Users do not need to learn/understand R to run iNEXT Online. The interactive web application was built using the Shiny (a web application framework).

iNEXT features two statistical analyses (non-asymptotic and asymptotic) for species diversity based on Hill numbers:

(1) A non-asymptotic approach based on interpolation and extrapolation

iNEXT computes the estimated diversities for standardized samples with a common sample size or sample completeness. This approach aims to compare diversity estimates for equally-large (with a common sample size) or equally-complete (with a common sample coverage) samples; it is based on the seamless rarefaction and extrapolation (R/E) sampling curves of Hill numbers for q = 0, 1 and 2. iNEXT offers three types of R/E sampling curves:

- Sample-size-based (or size-based) R/E sampling curves: This type of sampling curve plots the diversity estimates with respect to sample size.

- Coverage-based R/E sampling curves: This type of sampling curve plots the diversity estimates with respect to sample coverage.

- Sample completeness curve: This curve depicts how sample coverage varies with sample size. The sample completeness curve provides a bridge between the size- and coverage-based R/E sampling curves.

(2)   An asymptotic approach to infer asymptotic diversity

iNEXT computes the estimated asymptotic diversity profiles. It is based on statistical estimation of the true Hill number of any order $q >= 0$; see Chao and Jost (2015) for the statistical estimation detail.

How to cite

If you publish your work based on results from iNEXT Online, you should make references to the following methodology paper and Online software:

- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs, 84, 45-67.

- Chao, A., Ma, K. H., and Hsieh, T.C. (2016) iNEXT (iNterpolation and EXTrapolation) Online: Software for Interpolation and Extrapolation of Species Diversity. Program and User's Guide published at http://chao.stat.nthu.edu.tw/wordpress/software_download/inext-online/.

Data

iNEXT Online supports two types of data:

- Individual-based abundance data: data for each assemblage/site include sample species abundances in an empirical sample of individuals (a reference sample).

- Sampling-units-based incidence data: for each assemblage/site, data for a reference sample consist of species presence/absence (or detection/non-detection) in each of multiple sampling units.

See iNEXT Online User's Guide for data input formats.

Running procedures for "Interpolation/Extrapolation" analysis

Species identification names are irrelevant in species/taxonomic diversity assessment so they must be removed from your original data to conform iNEXT Online format. Steps 1, 2 and 4 are required selection procedures; Step 3 is optional.

Step 1. Select an analysis part ("Interpolation/Extrapolation" or "Asymptotic Analysis") from the top menu of iNEXT Online window.

Step 2. "Data Setting" (on the left hand side of the window screen)

(2a) Upload data: "Demo data" to see the illustrative example data, "Upload data" to load your own data, or "Key in data" by typing in your data in the input window.

(2b) Select data type: "Abundance data" or "Incidence data"; see below for data input format.

   Then the names/labels for the focal assemblages/sites will be automatically shown in the window; you can select one set or multiple sets for comparison.

(2c) For abundance data, there are two demo data sets. Select "Spiders" or "Beetles".

(2d) Then the names/labels for the uploaded assemblages/sites will be automatically shown in the window below; you can select one set or multiple sets for comparison.

Step 3. "General Setting" for statistical procedures

(3a) Optional: select a diversity order of q, q = 0 for species richness; q = 1 for Shannon diversity; q = 2 for Simpson diversity (default is q = 0).

(3b) Optional: specify the number of bootstrap replications to compute s.e. and confidence intervals for each estimator. (Default bootstrap replications = 50). You can type in "0" to skip all bootstrapping to save running time.

(3c) Optional: specify the level for confidence interval (default level = 0.95).

(3d) Optional: you can either specify the endpoint of extrapolation range and/or the number of knots or specify the samples sizes for which diversity estimates will be calculated. (Default endpoint = double of the minimum reference sample size, and the default number of knots = 40). If you choose to specify the sample sizes, then type in those sample sizes in the window space.

Step 4. Press the "Run!" button to get the output.

NOTE1: We use a bootstrap resampling method to compute s.e. and confidence interval of any estimator involved in the analysis. The default number of bootstrap replications is 50 to save running time. You may specify a larger number (say, 100) to obtain more accurate results for publication purposes, but it will take a longer time to get the output.

NOTE2: If you just want to take a glance at the pattern of rarefaction/extrapolation sampling curves and diversity profiles without requiring confidence intervals, then type in "0" for the number of bootstrap replications; in this case, the computation of s.e. and confidence intervals will be skipped so that the output can be promptly shown.

NOTE3: The bootstrap resampling procedures vary with trial, meaning that two different runs for the same data may result in different s.e. estimates and different confidence intervals.

Running procedures for "Asymptotic Analysis"

When you select the "Asymptotic Analysis" from the top menu, all the procedures are the same as those for the "Interpolation/Extrapolation" part except for Step (3d):

Step (3d) Optional: you either use the default diversity orders ($q = 0$ to 3 in increments of 0.25) or you type in those diversity orders for which asymptotic estimates will be computed.

Output for "Interpolation/Extrapolation"

Along the second row (output) menu, there are five output selection tabs:

● In the "**Data Summary**" tab panel, basic data information (sample size, observed species richness and estimated sample coverage) and the first 10 abundance (or incidence) frequency counts are shown for the reference sample.

● In the "**Rarefaction and Extrapolation**" tab panel, various estimates for interpolated or extrapolated samples and their confidence intervals are supplied. These results can be downloaded by clicking "download as CSV file" at the bottom of the displayed figure.

● In the "**Figure Plots**" tab panel, three types of sampling curves are shown: sample-size-based rarefaction and extrapolation sampling curve, sample completeness curve and Coverage-based rarefaction and extrapolation sampling curve. These figures can be downloaded by clicking "download as PNG file" at the bottom of the displayed figures.

● In the "**Introduction**" panel, users can view a brief introduction to iNEXT Online and a summary of the running procedures.

● In the "**User Guide**" panel, a link will direct users to iNEXT Online User Guide.

Output for "Asymptotic Analysis"

Along the second row (output) menu, there are four output selection tabs:

● In the "**Diversity Profile (estimated)**" tab panel, the figure including estimated diversity profiles of all assemblages for diversity order between 0 and 3 are shown based on the statistical method proposed in Chao and Jost (2015). The figure can be downloaded by clicking "download as PNG file" at the bottom of the displayed figure. Below the profile, all numerical values of diversity (and Tsallis entropy) estimates are tabulated along with s.e. and confidence intervals. These results can be downloaded by clicking "download as CSV file" below the table.

● In the "**Diversity Profile (empirical)**" tab panel, the observed diversity profiles for diversity order between 0 and 3 are shown. Below the profile, all observed values of

diversity (and Tsallis entropy) are tabulated along with s.e. and confidence intervals. The figure and table can be downloaded as described earlier for estimated part.

- In the "**Introduction**" panel, users can view a brief introduction to iNEXT Online and a summary of the running procedures.

- In the "**User Guide**" panel, a link will direct users to this user guide.

The user is referred to iNEXT Online User's Guide for examples and sample output/plots. For more sophisticated plots, please use iNEXT R packages available from CRAN; see Hsieh et al. (2016).

References

The following papers for pertinent background on rarefaction/extrapolation and related statistical analyses. These papers can be directly downloaded from Anne Chao's website.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. and Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45-67.

Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, **93**, 2533-2547.

Chao, A. and Jost, L. (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution*, **6**, 873-882.

Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.-Y., Mao, C.X., Chazdon, R.L. & Longino, J.T. (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, **5**, 3-21.

Hsieh, T.C., Ma, K.H. & Chao, A. (2016) iNEXT: An R package for interpolation and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, **7**, 1451