Package 'rareNMtests'

February 20, 2015

Type Package

Title Ecological and biogeographical null model tests for comparing rarefaction curves

Version 1.1

Date 2014-10-15

Author Luis Cayuela and Nicholas J. Gotelli

Maintainer Luis Cayuela <luis.cayuela@urjc.es>

Depends vegan

Description Randomization tests for the statistical comparison of \{ }emph{i} = two or more individual-based, sample-based or coverage-based rarefaction curves. The ecological null hypothesis is that the \{ }emph{i} samples were all drawn randomly from a single assemblage, with (necessarily) a single underlying species abundance distribution. The biogeographic null hypothesis is that the \{ }emph{i} samples were all drawn from different assemblages that, nonetheless, share similar species richness and species abundance distributions

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2014-10-18 06:36:06

R topics documented:

BiogTest.individual   *Biogeographical null model tests*

Description

Biogeographical null model tests for comparing rarefaction curves.

Usage

```
BiogTest.individual(x, MARGIN = 2, niter = 200, method = "sample-size", q = 0,
trace = TRUE, powerfun = 1, log.scale = FALSE, distr = "lnorm")
BiogTest.sample(x, by = NULL, MARGIN = 2, niter = 200, method = "sample-size",
q = 0, trace = TRUE, distr = "lnorm")
## S3 method for class 'BiogTest'
plot(x, max.poly = 50, ...)
```

Arguments

x	Community data, a matrix-like object, with entries representing abundance for `EcoTest.individual`, or presence-absence for `EcoTest.sample`, or an `'EcoTest'` class object for `plot` function.
MARGIN	If `MARGIN=1`, sampling units are rows and species are columns; if `MARGIN=2` (default), then species are rows and sampling units are columns.
niter	Number of randomizations used for the null model hypothesis test.
method	Either `"sample-size"` (default) or `"coverage"`. `"sample-size"` uses either abundance data (in individual-based rarefaction) or incidence data (in sample-based rarefaction), whereas `"coverage"` uses the estimated coverage (see Details) of either abundance or number of samples as the x-axis in the sampling curve.
q	First three Hill numbers, namely species richness ($q = 0$), the exponential Shannon index ($q = 1$), and the inverse Simpson index ($q = 2$).
trace	If trace=TRUE, information is printed during the randomization process.
powerfun	By default rarefied richness is estimated for all subsample sizes, from 1 to *n*, where *n* is the total number of individuals in the sample. If powerfun is less than 1, it decreases the total number of subsamples as a power function of {hn.
log.scale	If `FALSE` (default), subsample sizes for rarefying community are spaced apart at regular intervals. Otherwise (`log.scale = TRUE`), subsample sizes are spaced apart on a log-scale.
distr	Underlying species abundance distribution from which random assemblages are created by sampling to construct the null distribution. Available options include the log-normal (`lnorm`), geometric (`geom`), and broken-stick (`stick`) distributions.
by	Factor for grouping sampling units in sample-based rarefaction.
max.poly	Maximum number of polygons to draw in `plot` function. Each polygon represents the area between each unique pair of the *K* sample rarefaction curves.
...	Additional graphical parameters passed to plot (not used).

Details

> `BiogTest.individual` and `BiogTest.sample` present randomization tests for the statistical comparison of two or more individual-based, sample-based, or coverage-based rarefaction curves. The biogeographical null hypothesis *H0* is that is that two (or more) reference samples, represented by either abundance or incidence data, were drawn randomly from different assemblages that, nonetheless, share similar species richness and species abundance distributions.

> To calculate the `BiogTest` statistic, *Zobs*, we summed the area between all unique pairs of the *K* sample rarefaction curves. If *H0* is true, then *Zobs* should be relatively small because all of the rarefaction curves should have similar profiles, regardless of their species composition. In the limit, if all of the rarefaction curves had an identical profile, *Zobs* would equal 0. To construct the null distribution, the algorithm creates random assemblages by sampling from an underlying species abundance distribution. Of the many possible distributions which distribution should be used? The algorithm offers the possibility to choose among different underlying relative abundance distributions, including the log-normal, geometric and broken-stick distributions, although the former has some advantages, and thus its use is recommended.

> The statistical parameters of the log-normal rank abundance distribution are the number of species in the assemblage and the variance of the distribution; the latter controls the differences in abundance between common and rare species. If these underlying parameters are known, then sample size effects can be estimated by random sampling of individuals from the specified distribution. However, it is very difficult to directly estimate these parameters from a sample or set of samples (O'Hara 2005). Instead, a suite of log-normal distributions is generated, that might act as a reasonable sampling universe for comparison with a set of reference samples to test the biogeographic null hypothesis. Our strategy was to specify a distribution for each of the two parameters in the log normal: species number and variance. As in a random effects model, each replicate of the null distribution reflects a single sample from a log-normal distribution in which the two model parameters were first determined by random assignment.

> For the lower boundary of species richness, the minimum possible value cannot be smaller than the maximum number of species observed in the richest single sample among a set of samples. For the upper boundary of species richness, we calculated the upper bound of the Chao1 95

> For the standard deviation of the log-normal, we sampled a random uniform value between 1.1 and 33 (0.1 and 3.5 on a log scale). For empirical assemblages, standard deviations typically fall within this range (Limpert et al. 2001). Once the null assemblage was specified by selection of parameters for species richness and the standard deviation, we sampled (with replacement) the specified number of individuals for each sample in an individual-based data set. For incidence data, we sampled the observed number of species in each sampling unit, sampling species without replacement, with sample probabilities set proportional to relative abundances in the log-normal distribution. We then used the analytic formulas in Chao et al. (2014) to construct the rarefaction curves for each of the pseudosamples, simulated a distribution of `Zsim`, and compared it to `Zobs` to estimate $p(Z_{obs}|H_0)$.

> For the broken stick, the number of species in the meta-community must be known, and then a random partition is made to define the relative abundance of each species. For the geometric series, two parameters are needed: the number of species in the meta-community and a constant ratio `D` (`D<1`), which determines the abundance of the next species in the sequence. In the geometric series, `D` was obtained by sampling a random uniform value between 0.1 and 1. In all cases, the number of species (`S`), as in the log-normal distribution, was obtained by randomly drawing from a uniform distribution that was bounded at the low end by the maximum observed `S` and at the high end by the maximum upper bound of the 95

To better mimic the sampling process, a negative binomial random error was added to the abundance counts every time a sample was randomly drawn from the simulated meta-community. The negative binomial distribution was used to generate realistic heterogeneity that often results from spatial clustering of individuals and other small-scale processes. The expectation mu of the negative binomial was represented by the abundance count of each species in the meta-community.

Analytic estimators for individual-based, sample-based, and coverage-based rarefaction on Hill numbers were derived by Chao et al. (2014), and are currently implemented in this package (see link{rarefaction.individual} and link{rarefaction.sample} functions).

Value

`BiogTest.individual` and `BiogTest.sample` return an object of class `'BiogTest'`, basically a list with the following components:

subclass	A character indicating the type of data used to build rarefaction curves, namely "Individual-based method" for individual-based rarefaction (abundance data), and "Sample-based method" for sample-based rarefaction (incidence data), respectively.
type	A character indicating the variable used for the x-axis in the sampling curve, either "Sample-size" for abundance data (in individual-based rarefaction) or incidence data (in sample-based rarefaction), or "Coverage measure" for the estimated coverage of either abundance or number of samples.
obs	A list of data frames, with as many components as individual samples. Each data frame contains two columns, one for either sample-size (this refers either to number of individuals in individual-based rarefaction, or to number of sampling units in sample-based rarefaction) or coverage (in coverage-based rarefaction), and another for the estimated Hill number in each observed sample.
sim	A list of data frames, with as many components as individual samples. Each data frame contains two columns, one for either sample-size (this refers either to number of individuals in individual-based rarefaction, or to number of sampling units in sample-based rarefaction) or coverage (in coverage-based rarefaction), and another for the estimated Hill number in each simulated sample from one artificial set of samples.
Zsim	A vector of length `niter` showing each of the cumulative areas between all unique pairs of the `K` simulated rarefaction curves.
Z	The cumulative area between all unique pairs of the `K` observed sample rarefaction curves.
pval	The probability of Z given the distribution of Zsim. Low *p-values* imply that observed differences among samples in richness, and/or relative abundance are improbable if the samples were all drawn from similar assemblages regardless of differences in species composition.

Author(s)

Luis Cayuela and Nicholas J. Gotelli

References

Cayuela, L. & Gotelli, N.J. Ecological and biogeographical null hypotheses for comparing rarefaction curves. In prep.

Chao, A. (1984). Nonparametric-estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11: 265-270.

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. emphEcological Monographs 84: 45-67.

Limpert, E., Stahel, W.A. & Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. emphBioScience 51: 341-352.

O'Hara, R.B. (2005). Species richness estimators: how many species can dance on the head of a pin? emphJournal of Animal Ecology 74: 375-386.

See Also

rarefaction.individual, rarefaction.sample, rarecurve, chao1

Examples

```
## Not run:
 ## Individual-based and coverage-based rarefaction
 # Simulate a community with number of species (spn) and evenness randomly selected

 spn <- round(runif(1, 10, 200))
 evenness <- runif(1, log(1.1), log(33))
 com <- round(rlnorm(spn, 2, evenness))
 Ss <- round(runif(1, 50, 500))
 sample1 <- sample(paste("sp",1:length(com)), rpois(1, Ss), replace=TRUE, prob=com)
 sample1 <- data.frame(table(sample1))
 colnames(sample1) <- c("species", "sample1")
 sample2 <- sample(paste("sp",1:length(com)), rpois(1, Ss), replace=TRUE, prob=com)
 sample2 <- data.frame(table(sample2))
 colnames(sample2) <- c("species", "sample2")
 df <- merge(sample1, sample2, by="species", all=TRUE)
 rownames(df) <- df$species
 df[is.na(df)] <- 0
 df <- df[,2:3]

 # Biogeographical null model test using sample-based rarefaction curves
 # for species richness (q = 0)
 # The test should not reject the null hypothesis (p > 0.05)
 ibbiogq0 <- BiogTest.individual(df, MARGIN=2, powerfun=0.8, log.scale=TRUE)
 plot(ibbiogq0)

 # Biogeographical null model test using coverage-based rarefaction curves
 # for the exponential Shannon index (q = 1)
 ibbiogq1cov <- BiogTest.individual(df, MARGIN=2, method="coverage", q=1,
 powerfun=0.8, log.scale=TRUE)
 plot(ibbiogq1cov)
```

```
## Sample-based and coverage-based rarefaction
# Load the data
data(Chiapas)
Chiapas <- subset(Chiapas, Region!="El Triunfo")
str(Chiapas)

# Biogeographical null model test using sample-based rarefaction curves
# for species richness (q = 0)
sbbiogq0 <- BiogTest.sample(Chiapas[,-1], by=Chiapas[,1], MARGIN=1)
plot(sbbiogq0)

# Biogeographical null model test using coverage-based rarefaction curves
# for the inverse Simpson index (q = 2)
sbbioq2cov <- EcoTest.sample(Chiapas[,-1], by=Chiapas[,1], MARGIN=1,
method="coverage", q=2)
plot(sbbioq2cov)

## End(Not run)
```

---

chao1   *Computes the Chao 1 species richness estimator*

Description

Computes the Chao 1 richness estimator, its estimated variance, and the corresponding 95

Usage

```
chao1(x)
```

Arguments

x   Vector of species abundances.

Details

Different equations are used to compute the classic Chao1 richness estimator, its estimated variance, and the corresponding 95

Value

A data frame with five columns including the observed number of species (Sobs), the estimated number of species with the Chao 1 estimator (S.chao1), the estimated variance (var), and the upper and lower 95

Author(s)

Luis Cayuela

References

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43, 783-791.

See Also

specpool

Examples

```
abund <- c(19, 12, 7, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
print(chao1(abund))
```

Chiapas   *Tree abundance in tropical montane forest plots*

Description

Tree abundance in 224 0.1-ha plots in three montane cloud forest regions of Chiapas, Mexico.

Usage

```
data(Chiapas)
```

Format

The Chiapas data set contains data for 375 tree species (columns) from 224 0.1-ha plots (rows). The first column is a factor with three levels with the name of the region.

Details

This data set characterizes tree communities from tropical montane cloud forests in three regions of the state of Chiapas, Mexico: El Triunfo Biosphere Reserve in Sierra Madre (100 plots), the Highlands (38 plots), and the Northern Mountains (86 plots). For these plots, incidence records were based on trees exceeding 5 cm in diameter at breast height. Data were obtained from the BIOTREE-NET website (www.biotreenet.com). Species names were standardized and typographical errors were corrected using The Plant List through the **Taxostand** package (Cayuela et al. 2012). Distance between regions ranged from ca. 50 kms (the Highlands and Northern Mountains) to ca. 250 kms (El Triunfo and Northern Mountains).

Source

http://portal.biotreenet.com

References

Cayuela, L., Golicher, D.J., Rey Benayas, J.M., Gonzalez-Espinosa, M. & Ramirez-Marcial, N. (2006). Fragmentation, disturbance and tree diversity conservation in tropical montane forests. *Journal of Applied Ecology* 43: 1172-1181. Cayuela, L., Granzow-de la Cerda, I., Albuquerque, F.S., & Golicher, J.D. (2012). Taxonstand: An R package for species names standardisation in vegetation databases. *Methods in Ecology and Evolution* 3(6): 1078-1083. Ramirez-Marcial, N., Gonzalez-Espinosa, M. & Williams-Linera, G. (2001). Anthropogenic disturbance and tree diversity in montane rain forests in Chiapas, Mexico. *Forest Ecology and Management* 154(1): 311-326.

Examples

```
data(Chiapas)
str(Chiapas)
```

EcoTest.individual   *Ecological null model tests*

Description

Ecological null model tests for comparing rarefaction curves.

Usage

```
EcoTest.individual(x, MARGIN = 2, niter = 200, method = "sample-size", q = 0,
trace = TRUE, powerfun = 1, log.scale = FALSE)
EcoTest.sample(x, by = NULL, MARGIN = 2, niter = 200, method = "sample-size",
q = 0, trace = TRUE)
## S3 method for class 'EcoTest'
plot(x, ...)
```

Arguments

x	Community data, a matrix-like object, with entries representing abundance for `EcoTest.individual`, or presence-absence for `EcoTest.sample`, or an `'EcoTest'` class object for `plot` function.
MARGIN	If `MARGIN=1`, sampling units are rows and species are columns; if `MARGIN=2` (default), then species are rows and sampling units are columns.
niter	Number of randomizations used for the null model hypothesis test.
method	Either `"sample-size"` (default) or `"coverage"`. `"sample-size"` uses either abundance data (in individual-based rarefaction) or incidence data (in sample-based rarefaction), whereas `"coverage"` uses the estimated coverage (see Details) of either abundance or number of samples as the x-axis in the sampling curve.
q	First three Hill numbers, namely species richness ($q = 0$), the exponential Shannon index ($q = 1$), and the inverse Simpson index ($q = 2$).

trace	If `trace=TRUE`, information is printed during the randomization process.
powerfun	By default rarefied richness is estimated for all subsample sizes, from 1 to *n*, where *n* is the total number of individuals in the sample. If powerfun is less than 1, it decreases the total number of subsamples as a power function of *n*.
log.scale	If FALSE (default), subsample sizes for rarefying community are spaced apart at regular intervals. Otherwise (`log.scale = TRUE`), subsample sizes are spaced apart on a log-scale.
by	Factor for grouping sampling units in sample-based rarefaction.
...	Additional graphical parameters passed to plot (not used).

Details

`EcoTest.individual` and `EcoTest.sample` present randomization tests for the statistical comparison of two or more individual-based, sample-based, or coverage-based rarefaction curves. The ecological null hypothesis *H0* is that two (or more) reference samples, represented by either abundance or incidence data, were both drawn from the same assemblage of *N** individuals and *S* species. Therefore, any differences among the samples in species composition, species richness, or relative abundance reflect only random variation, given the number of individuals (or sampling units) in each collection. The alternative hypothesis, in the event that *H0* cannot be rejected, is that the sample data were drawn from different assemblages. If *H0* is true, then pooling the samples should give a composite sample that is also a (larger) random subset of the complete assemblage. It is from this pooled composite sample that we make random draws for comparison with the actual data.

To construct the `EcoTest` metric, we begin by plotting the expected rarefaction curves for the individual samples and for the pooled composite sample *C*. Next, for each individual sample *i*, we calculate the cumulative area *Ai* between the sample rarefaction curve and the pooled rarefaction curve. For a set of *i*=1 to *K* samples, we define the observed difference index:

$$Z_{obs} = \sum_{i=1}^{K} A_i$$

Note that two identically-shaped rarefaction curves may nevertheless differ from the pooled curve. This difference can arise because species identities in the individual samples are retained in the pooled composite sample *C*, which affects the shape of the pooled rarefaction curve.

The data are next reshuffled by randomly re-assigning every individual to a sample (for abundance data) or every sampling unit to a sample (for incidence data), and preserving the original sample sizes (number of individuals for abundance data and number of samples for incidence data). From this randomization, we again construct rarefaction curves and calculate *Zsim* as the cumulative area between the rarefaction curves of the randomized samples and the composite rarefaction curve. This procedure is repeated many times, leading to a distribution of *Zsim* values and a 95

Analytic estimators for individual-based, sample-based, and coverage-based rarefaction on Hill numbers were derived by Chao et al. (2014), and are currently implemented in this package (see `link{rarefaction.individual}` and `link{rarefaction.sample}` functions).

Value

`EcoTest.individual` and `EcoTest.sample` return an object of class `'EcoTest'`, basically a list with the following components:

subclass	A character indicating the type of data used to build rarefaction curves, namely "Individual-based method" for individual-based rarefaction (abundance data), and "Sample-based method" for sample-based rarefaction (incidence data), respectively.
type	A character indicating the variable used for the x-axis in the sampling curve, either "Sample-size" for abundance data (in individual-based rarefaction) or incidence data (in sample-based rarefaction), or "Coverage measure" for the estimated coverage of either abundance or number of samples.
obs	A list of data frames, with as many components as individual samples. Each data frame contains two columns, one for either sample-size (this refers either to number of individuals in individual-based rarefaction, or to number of sampling units in sample-based rarefaction) or coverage (in coverage-based rarefaction), and another for the estimated Hill number in each observed sample.
sim	A list of data frames, with as many components as individual samples. Each data frame contains two columns, one for either sample-size (this refers either to number of individuals in individual-based rarefaction, or to number of sampling units in sample-based rarefaction) or coverage (in coverage-based rarefaction), and another for the estimated Hill number in each simulated sample from one randomized set of samples.
pooled	A data frame with entries for sample-size (this refers either to number of individuals in individual-based rarefaction, or to number of sampling units in sample-based rarefaction) or coverage (in coverage-based rarefaction), and the corresponding Hill number for the composite (i.e. pooled) rarefaction curve.
Zsim	A vector of length `niter` showing each of the cumulative areas between the rarefaction curves of one randomized sample data set and the composite rarefaction curve.
Z	The cumulative area between the observed sample rarefaction curves and the composite rarefaction curve.
pval	The probability of *Z* given the distribution of *Zsim*. Low *p-values* imply that observed differences among samples in species composition, richness, and/or relative abundance are improbable if the samples were all drawn from the same assemblage.

Author(s)

Luis Cayuela and Nicholas J. Gotelli

References

Cayuela, L. & Gotelli, N.J. Ecological and biogeographical null hypotheses for comparing rarefaction curves. In prep.

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. emphEcological Monographs 84: 45-67.

See Also

[rarefaction.individual](), [rarefaction.sample](), [rarecurve]()

Examples

```
## Not run:
 ## Individual-based and coverage-based rarefaction
 # Simulate a community with number of species (spn) and evenness randomly selected

 spn <- round(runif(1, 10, 200))
 evenness <- runif(1, log(1.1), log(33))
 com <- round(rlnorm(spn, 2, evenness))
 Ss <- round(runif(1, 50, 500))
 sample1 <- sample(paste("sp",1:length(com)), rpois(1, Ss), replace=TRUE, prob=com)
 sample1 <- data.frame(table(sample1))
 colnames(sample1) <- c("species", "sample1")
 sample2 <- sample(paste("sp",1:length(com)), rpois(1, Ss), replace=TRUE, prob=com)
 sample2 <- data.frame(table(sample2))
 colnames(sample2) <- c("species", "sample2")
 df <- merge(sample1, sample2, by="species", all=TRUE)
 rownames(df) <- df$species
 df[is.na(df)] <- 0
 df <- df[,2:3]

 # Ecological null model test using sample-based rarefaction curves
 # for species richness (q = 0)
 # The test should not reject the null hypothesis in most cases (p > 0.05)
 ibecoq0 <- EcoTest.individual(df, MARGIN=2, powerfun=0.8, log.scale=TRUE)
 plot(ibecoq0)

 # Ecological null model test using coverage-based rarefaction curves
 # for the exponential Shannon index (q = 1)
 ibecoq1cov <- EcoTest.individual(df, MARGIN=2, method="coverage",
 q=1, powerfun=0.8, log.scale=TRUE)
 plot(ibecoq1cov)

 ## Sample-based and coverage-based rarefaction
 # Load the data
 data(Chiapas)
 Chiapas <- subset(Chiapas, Region!="El Triunfo")
 str(Chiapas)

 # Ecological null model test using sample-based rarefaction curves
 # for species richness (q = 0)
 sbecoq0 <- EcoTest.sample(Chiapas[,-1], by=Chiapas[,1], MARGIN=1)
 plot(sbecoq0)

 # Ecological null model test using coverage-based rarefaction curves
 # for the inverse Simpson index (q = 2)
 sbecoq2cov <- EcoTest.sample(Chiapas[,-1], by=Chiapas[,1],
 MARGIN=1, method="coverage", q=2)
 plot(sbecoq2cov)

## End(Not run)
```

`rarefaction.individual`

Rarefied species richness with Hill numbers

Description

Individual-based (abundance data), sample-based (incidence data), and coveraged-based (based on both abundance and incidence data) rarefaction with Hill numbers.

Usage

```
rarefaction.individual(x, method = "sample-size", q = 0, powerfun = 1,
log.scale = FALSE, inds = NULL)
rarefaction.sample(x, method = "sample-size", q = 0)
```

Arguments

x	Community data, a vector (for individual-based rarefaction) or a matrix-like object (for sample-based rarefaction).
method	Either `"sample-size"` (default) or `"coverage"`. `"sample-size"` uses either abundance data (in individual-based rarefaction) or incidence data (in sample-based rarefaction), whereas `"coverage"` uses the estimated coverage (see Details) of either abundance or number of samples as the x-axis in the sampling curve.
q	First three Hill numbers, namely species richness (`q = 0`), the exponential Shannon index (`q = 1`), and the inverse Simpson index (`q = 2`).
powerfun	By default rarefied richness is estimated for all subsample sizes, from 1 to *n*, where *n* is the total number of individuals in the sample. If `powerfun` is less than 1, it decreases the total number of subsamples as a power function of *n*.
log.scale	If `FALSE` (default), subsample sizes for rarefying community are spaced apart at regular intervals. Otherwise (`log.scale = TRUE`), subsample sizes are spaced apart on a log-scale.
inds	Subsample size for rarefying community, either a single value or a vector. If not specified (default), rarefaction is calculated for all subsample sizes.

Details

Consider a complete assemblage for which all species and their relative abundance are known. In this complete assemblage, there are *i*=1 to *S* species and N* total individuals, with Ni individuals of species i. For individual-based (abundance) data, the reference sample consists of n individuals drawn at random from N*, with Sobs species present, each represented by Xi individuals. Individual-based data is represented as a single vector of length Sobs, the elements of which are the observed abundances Xi. For sample-based (incidence) data, the reference sample consists of a set of R standardized sampling units, such as traps, plots, transect lines, etc. Within each of these sampling units, the presence (1) or absence (0) of each species are the required data, even though

abundance data may have been collected. Sample-based incidence data can be represented as a single matrix, with `i=1` to `R` rows and `j=1` to `Sobs` columns, and entries `Wij=1` or `Wij=0` to indicate the presence or absence of species `j` in sampling unit `i`.

In the past, rarefaction curves have been estimated by repeated subsampling, but it is no longer necessary. Analytic estimators for individual-based, sample-based, and coverage-based rarefaction on Hill numbers were derived by Chao et al. (2014). In standard rarefaction curves the x-axis is either the abundance (individual-based) or number of samples (sample-based). Additionally, the x-axis can also be the estimated coverage of either abundance or number of samples. Coverage is defined as the proportion of total individuals or samples from the complete assemblage that is represented by the species present in the sample or subset of samples (Chao and Jost 2012).

Functions `rarefaction.individual` and `rarefaction.sample` rarefies a series of diversity indices known as Hill numbers (Hill 1973), which can be algebraically transformed into more familiar diversity indices. The order `q` of the Hill number determines the weighting given to more common species, with species richness defined by q=0. Other Hill numbers are the exponential Shannon index (q=1), and the inverse Simpson index (q=2).

By default, rarefaction is estimated for all subsample sizes, from 1 to `n` in individual-based rarefaction, or from 1 to `R` in sample-based rarefaction, where `n` is the total number of individuals in the sample, and `R` is the total number of sampling units, respectively. In individual-based rarefaction, computation of Hill numbers for different subsample sizes can be very time consuming when total number of individuals, n, is high (e.g. > 1000) to extremely high (e.g. > 10000). To reduce computation time, one can select a given number of subsample sizes by setting `powerfun` below 1. This decreases the total number of subsamples as a power function of n. Subsample sizes can be spaced apart at regular intervals if `log.scale=FALSE`, or at log-intervals if `log.scale=TRUE`. This allows getting the entire rarefaction curve profile at faster speed. It is not advisable to set the value of `powerfun` below 0.5, and recommended values are between 0.6 and 0.8.

Value

A data frame with entries for sample-size (this refers either to number of individuals in individual-based rarefaction, or to number of sampling units in sample-based rarefaction) or coverage (in coverage-based rarefaction), and the corresponding Hill number.

Author(s)

Luis Cayuela and Nicholas J. Gotelli

References

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. emphEcological Monographs 84: 45-67.

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

Hill, M.O. (1973). Diversity and evenness: a unifying notation and its consequences. emphEcology 54: 427-432.

See Also

rarecurve

Examples

```
## Not run:
## Individual-based and coverage-based rarefaction
# Simulate a community with number of species (sr) and evenness randomly selected
spn <- rlnorm(runif(1, 10, 200))
evenness <- runif(1, log(1.1), log(33))
com <- round(rlnorm(spn, 3, evenness))
# Sample the community (with sample size n = 500 individuals)
sample1 <- sample(paste("sp",1:length(com)), 500, replace=TRUE, prob=com)
sample1 <- table(sample1)

# Get the individual-based and coverage-based rarefaction curves for different Hill numbers
ibr.q0 <- rarefaction.individual(sample1)
ibr.q1 <- rarefaction.individual(sample1, q=1)
ibr.q2 <- rarefaction.individual(sample1, q=2)
ibr.q0.cov <- rarefaction.individual(sample1, method="coverage")
ibr.q1.cov <- rarefaction.individual(sample1, q=1, method="coverage")
ibr.q2.cov <- rarefaction.individual(sample1, q=2, method="coverage")

# Plot the results
par(mfcol=c(1,2))
plot(ibr.q0[,1], ibr.q0[,2], lwd=2, xlab="Number of individuals", ylab="Hill numbers",
type="l", main="Individual-based rarefaction")
lines(ibr.q1[,1], ibr.q1[,2], lwd=2, lty=2)
lines(ibr.q2[,1], ibr.q2[,2], lwd=2, lty=3)
legend("bottomright", lty=c(1,2,3), lwd=2, legend=c("q = 0", "q = 1", "q = 2"), cex=1.2)
plot(ibr.q0.cov[,1], ibr.q0.cov[,2], lwd=2, xlab="Coverage", ylab="Hill numbers",
type="l", main="Coverage-based rarefaction")
lines(ibr.q1.cov[,1], ibr.q1.cov[,2], lwd=2, lty=2)
lines(ibr.q2.cov[,1], ibr.q2.cov[,2], lwd=2, lty=3)
legend("topleft", lty=c(1,2,3), lwd=2, legend=c("q = 0", "q = 1", "q = 2"), cex=1.2)

## Sample-based and coverage-based rarefaction
# Load the data
data(Chiapas)
Chiapas <- subset(Chiapas, Region=="El Triunfo")
str(Chiapas)

# Get sample-based and coverage-based rarefaction curves for different Hill numbers
sbr.q0 <- rarefaction.sample(Chiapas[,-1])
sbr.q1 <- rarefaction.sample(Chiapas[,-1], q=1)
sbr.q2 <- rarefaction.sample(Chiapas[,-1], q=2)
sbr.q0.cov <- rarefaction.sample(Chiapas[,-1], method="coverage")
sbr.q1.cov <- rarefaction.sample(Chiapas[,-1], q=1, method="coverage")
sbr.q2.cov <- rarefaction.sample(Chiapas[,-1], q=2, method="coverage")

# Plot the results
par(mfcol=c(1,2))
```

```
    plot(sbr.q0[,1], sbr.q0[,2], lwd=2, xlab="Sampling units", ylab="Hill numbers",
    type="l", main="Sample-based rarefaction")
    lines(sbr.q1[,1], sbr.q1[,2], lwd=2, lty=2)
    lines(sbr.q2[,1], sbr.q2[,2], lwd=2, lty=3)
   legend("bottomright", lty=c(1,2,3), lwd=2, legend=c("q = 0", "q = 1", "q = 2"), cex=1.2)
    plot(sbr.q0.cov[,1], sbr.q0.cov[,2], lwd=2, xlab="Coverage", ylab="Hill numbers",
    type="l", main="Coverage-based rarefaction")
    lines(sbr.q1.cov[,1], sbr.q1.cov[,2], lwd=2, lty=2)
    lines(sbr.q2.cov[,1], sbr.q2.cov[,2], lwd=2, lty=3)
    legend("topleft", lty=c(1,2,3), lwd=2, legend=c("q = 0", "q = 1", "q = 2"), cex=1.2)

## End(Not run)
```

Index