Package 'SpadeR'

July 21, 2025

Type Package

Title Species-Richness Prediction and Diversity Estimation with R

Version 0.1.1

Date 2016-09-06

Author Anne Chao, K. H. Ma, T. C. Hsieh and Chun-Huo Chiu

Maintainer Anne Chao <chao@stat.nthu.edu.tw>

Description Estimation of various biodiversity indices and related (dis)similarity measures based on individual-based (abundance) data or sampling-unit-based (incidence) data taken from one or multiple communities/assemblages.

License GPL (>= 3)

Imports stats

Depends R (>= 2.14)

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-06 14:53:57

Contents

SpadeR-package	*Species-richness prediction and diversity estimation with R*

Description

Provides simple functions to compute various biodiversity indices and related (dis)similarity measures based on individual-based (abundance) data or sampling-unit-based (incidence) data taken from one or multiple communities/assemblages.

This package contains six main functions:

1. `ChaoSpecies` (estimating species richness for one community).

2. `Diversity` (estimating a continuous diversity profile and various diversity indices in one community including species richness, Shannon diversity and Simpson diversity). This function also features plots of empirical and estimated continuous diversity profiles.

3. `ChaoShared` (estimating the number of shared species between two communities).

4. `SimilartyPair` (estimating various similarity indices between two assemblages). Both richness- and abundance-based two-community similarity indices are included.

5. `SimilarityMult` (estimating various similarity indices among N communities). Both richness- and abundance-based N-community similarity indices are included.

6. `Genetics` (estimating allelic dissimilarity/differentiation among sub-populations based on multiple-subpopulation genetics data).

Except for the `Genetics` function, there are at least three types of data are supported for each function.

Details

Data are generally classified as abundance data and incidence data and there are five types of data input formats options (datatype="abundance", "abundance_freq_count", "incidence_freq", "incidence_freq_count", "incidence_raw").

A. Individual-based abundance data when a sample of individuals is taken from each community.
   Type (1) abundance data (datatype = "abundance"): Input data consist of species (in rows) by community (in columns) matrix. The entries of each row are the observed abundances of a species in N communities.

Type (1A) abundance-frequency counts data only for a single community (datatype = "abundance_freq_count"): input data are arranged as (1 f_1 2 f_2 ... r f_r)(each number needs to be separated by at least one blank space or separated by rows), where r denotes the maximum frequency and f_k denotes the number of species represented by exactly k individuals/times in the sample. Here the data (f_1, f_2, ..., f_r) are referred to as "abundance-frequency counts".

B. Sampling-unit-based incidence data when a number of sampling units are randomly taken from each community. Only the incidence (detection/non-detection) of species is recorded in each sampling unit. There are three data formats options.

Type (2) incidence-frequency data (datatype="incidence_freq"): The first row of the input data must be the number of sampling units in each community. Beginning with the second row, input data consist of species (in rows) by community (in columns) matrix. The entries of each row are the observed incidence frequencies (the number of detections or the number of sampling units in which a species are detected) of a species in N communities.

Type (2A) incidence-frequency counts data only for a single community (datatype="incidence _freq_count"): input data are arranged as (T 1 Q_1 2 Q_2 ... r Q_r) (each number needs to be separated by at least one blank space or separated by rows), where Q_k denotes the number of species that were detected in exactly k sampling units, while r denotes the number of sampling units in which the most frequent species were found. The first entry must be the total number of sampling units, T. The data ($Q_1, Q_2, ..., Q_r$) are referred to as "incidence frequency counts".

Type (2B) incidence-raw data (datatype="incidence_raw"): Data consist of a species-by-sampling-unit incidence (detection/non-detection) matrix; typically "1" means a detection and "0" means a non-detection. Each row refers to the detection/non-detection record of a species in T sampling units. Users must specify the number of sampling units in the function argument "units". The first T_1 columns of the input matrix denote species detection/non-detection data based on the T_1 sampling units from Community 1, and the next T_2 columns denote the detection/non-detection data based on the T_2 sampling units from Community 2, and so on, and the last T_N columns denote the detection/non-detection data based on T_N sampling units from Community N, $T_1 + T_2 + ... + T_N = T$.

An Online version of SpadeR is also available for users without an R background: http://chao.stat.nthu.edu.tw/wordpress/software_download/softwarespader_online/. In the detailed Online SpadeR User's Guide, we illustrate all the running procedures in an easily accessible way through numerical examples with proper interpretations of portions of the output. All the data of those illustrative examples are included in this package.

functions: ChaoSpecies, Diversity, ChaoShared, SimilarityPair, SimilarityMult, Genetics

Author(s)

Anne Chao, K. H. Ma, T. C. Hsieh and Chun-Huo Chiu

Maintainer: Anne Chao <chao@stat.nthu.edu.tw>

ChaoShared   *Estimation of the number of shared species between two communities/assemblages*

Description

ChaoShared: Estimation of shared species richness between two communities/assemblages based on three types of data: Type (1) abundance data (datatype="abundance"), Type (2) incidence-frequency data (datatype="incidence_freq"), and Type (2B) incidence-raw data (datatype="incidence _raw"); see SpadeR-package details for data input formats.

Usage

```
ChaoShared(data, datatype = c("abundance", "incidence_freq", "incidence_raw"),
  units, se = TRUE, nboot = 200, conf = 0.95)
```

Arguments

data   a matrix/data.frame of species abundances/incidences.

datatype   type of input data, "abundance", "incidence_freq" or "incidence_raw".

units   number of sampling units in each community. For datatype = "incidence_raw", users must specify the number of sampling units taken from each community. This argument is not needed for "abundance" and "incidence_freq" data.

se   a logical variable to calculate the bootstrap standard error and the associated confidence interval.

nboot   an integer specifying the number of bootstrap replications.

conf   a positive number ≤ 1 specifying the level of confidence interval.

Value

a list of two objects:

$Basic_data_information for summarizing data information.

$Estimation_results for showing a table of various shared richess estimates, standard errors, and the associated confidence intervals.

References

Chao, A., Hwang, W.-H., Chen, Y.-C. and Kuo. C.-Y. (2000). Estimating the number of shared species in two communities. Statistica Sinica, 10, 227-246.

Pan, H.-Y., Chao, A. and Foissner, W. (2009). A non-parametric lower bound for the number of species shared by multiple communities. Journal of Agricultural, Biological and Environmental Statistics, 14, 452-468.

Examples

```
data(ChaoSharedData)
# Type (1) abundance data
ChaoShared(ChaoSharedData$Abu,"abundance",se=TRUE,nboot=200,conf=0.95)
# Type (2) incidence-frequency data
ChaoShared(ChaoSharedData$Inci,"incidence_freq",se=TRUE,nboot=200,conf=0.95)
# Type (2B) incidence-raw data
ChaoShared(ChaoSharedData$Inci_raw,"incidence_raw",units=c(16,17),se=TRUE,nboot=200,conf=0.95)
```

ChaoSharedData *Data for Function ChaoShared*

Description

There are three data sets:

1. Type (1) abundance data (`ChaoSharedData$Abu`)

The data consist of the observed bird abundances/frequencies collected from two estuaries (Chao et al. 2000). For each species (row), the entry of the first column is the observed species frequency from Estuary I, and the second column is the observed species frequency from Estuary II. The species checklist includes 201 species, so the entry data includes a matrix of 201 rows and 2 columns.

2. Type (2) incidence-frequency data (`ChaoSharedData$Inci`)

The data consist of bird incidence (detection/non-detection) frequencies observed in 2015 (by 16 teams) and 2016 (by 17 teams) in the Hong Kong Bird Race. Each team is regarded as a sampling unit. Unlike the abundance data, the numbers of sampling units (16 and 17 for these data) are specified in the first row. Beginning with the second row, the entry of the first column is the observed incidence frequency (the total number of detections among all teams) of a given species in 2015, and the entry of the second column is the observed incidence frequency of the same species in 2016. A 280-species checklist was used, thus the input data consist of 281 rows (the first entry records the number of sampling units) and 2 columns.

3. Type (2B) incidence-raw data (`ChaoSharedData$Inci_raw`)

The data consist of raw detection/non-detection records of bird species in 2015 (by 16 teams) and 2016 (by 17 teams) in the Hong Kong Bird Race. A 280-species checklist was used. The raw data consist of a 280 x 33 (species-by-sampling-unit) matrix with element 1's (detection) or 0's (non-detection). Each row of the matrix refers to the detection/non-detection records of the same species so that the information about shared species can be computed. The first 16 columns of the matrix denote the species detection/non-detection data by 16 teams in 2015, and the next 17 columns denote the species detection/non-detection data by 17 teams in 2016.

Usage

```
data(ChaoSharedData)
```

References

Chao, A., Hwang, W.-H., Chen, Y.-C. and Kuo. C.-Y. (2000). Estimating the number of shared species in two communities. Statistica Sinica, 10, 227-246.

World Wildlife Fund (WWF) for Nature, Hong Kong. Bird Bird Race. http://www.wwf.org.hk/en/getinvolved/hkbbr/. Assessed on July 26, 2016

ChaoSpecies   *Estimation of species richness in a community*

Description

`ChaoSpecies`: Estimation of species richness in a single community based on five types of data: Type (1) abundance data (datatype="abundance"), Type (1A) abundance-frequency counts (datatype="abundance_freq_count"), Type (2) incidence-frequency data (datatype = "incidence_freq"), Type (2A) incidence-frequency counts (datatype="incidence_freq_count"), and Type (2B) incidence-raw data (datatype="incidence_raw"); see `SpadeR-package` details for data input formats.

Usage

```
ChaoSpecies(data, datatype = c("abundance", "abundance_freq_count",
  "incidence_freq", "incidence_freq_count", "incidence_raw"), k = 10,
  conf = 0.95)
```

Arguments

data	a matrix/data.frame of species abundances/incidences.
datatype	type of input data, "abundance", "abundance_freq_count", "incidence_freq", "incidence_freq_count" or "incidence_raw".
k	the cut-off point (default = 10), which separates species into "abundant" and "rare" groups for abundance data for the estimator ACE; it separates species into "frequent" and "infrequent" groups for incidence data for the estimator ICE.
conf	a positive number ≤ 1 specifying the level of confidence interval.

Value

a list of three objects:

`$Basic_data_information` and `$Rare_species_group/$Infreq_species_group` for summarizing data information.

`$Species_table` for showing a table of various species richness estimates, standard errors, and the associated confidence intervals.

References

Chao, A., and Chiu, C. H. (2012). Estimation of species richness and shared species richness. In N. Balakrishnan (ed). Methods and Applications of Statistics in the Atmospheric and Earth Sciences. p.76-111, Wiley, New York.

Chao, A., and Chiu, C. H. (2016). Nonparametric estimation and comparison of species richness. Wiley Online Reference in the Life Science. In: eLS. John Wiley and Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0026329.

Chao, A., and Chiu, C. H. (2016). Species richness: estimation and comparison. Wiley StatsRef: Statistics Reference Online. 1-26.

Chiu, C. H., Wang Y. T., Walther B. A. and Chao A. (2014). An improved non-parametric lower bound of species richness via the Good-Turing frequency formulas. Biometrics, 70, 671-682.

Gotelli, N. G. and Chao, A. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. Encyclopedia of Biodiversity, 2nd Edition, Vol. 5, 195-211, Waltham, MA.

Examples

```
data(ChaoSpeciesData)
# Type (1) abundance data
ChaoSpecies(ChaoSpeciesData$Abu,"abundance",k=10,conf=0.95)
# Type (1A) abundance-frequency counts data
ChaoSpecies(ChaoSpeciesData$Abu_count,"abundance_freq_count",k=10,conf=0.95)
# Type (2) incidence-frequency data
ChaoSpecies(ChaoSpeciesData$Inci,"incidence_freq",k=10,conf=0.95)
# Type (2A) incidence-frequency counts data
ChaoSpecies(ChaoSpeciesData$Inci_count,"incidence_freq_count",k=10,conf=0.95)
# Type (2B) incidence-raw data
ChaoSpecies(ChaoSpeciesData$Inci_raw,"incidence_raw",k=10,conf=0.95)
```

ChaoSpeciesData   *Data for Function ChaoSpecies*

Description

There are five data sets:

1. Type (1) abundance data (`ChaoSpeciesData$Abu`)

The data consist of 25 birds abundances/frequencies in a sample (Magurran, 1988, p.152). Their observed frequencies are respectively 752, 276, 194, 126, 121, 97, 95, 83, 72, 44, 39, 16, 15, 13, 9, 9, 9, 8, 7, 4, 2, 2, 1, 1, 1.

2. Type (1A) abundance-frequency counts data (`ChaoSpeciesData$Abu_count`)

The data consist of the observed species abundance distribution of endangered and rare vascular plant species in the central portion of the southern Appalachian region (Miller and Wiegert, 1989). A total of 188 species were recorded out of 1008 individuals compiled over a span of 150 years of field observations. The data are read as: (1 61 2 35 3 18 4 12 ... 67 1); each number needs to be separated by at least one blank space or by separated by rows. Here the first pair (1, 61) indicates that there are 61 singletons, the second pair (2, 35) indicates there are 35 doubletons, and so on, with the last pair (67, 1) indicating that there is one species that is represented by 67 individuals.

3. Type (2) incidence-frequency data (`ChaoSpeciesData$Inci`)

The data include seed-bank records taken from Butler and Chazdon (1998). There were 121 soil samples (each soil sample is regarded as a sampling unit) and species of seedlings that germinated from each soil sample were recorded. A total of 34 species of seedlings were found in the 121 soil samples. In the input data, the entry in the first row denotes the number of sampling units. Then, beginning with the second row, each row records the species incidence frequency (i.e., the number of soil samples in which the seedlings were found) of a given species in all 121 soil samples. The ordering of data entries does not affect the analysis.

4. Type (2A) incidence-frequency counts data (`ChaoSpeciesData$Inci_freq_count`)

The data consist of cottontail capture-recapture data provided in Edwards and Eberhardt (1967) to illustrate that species richness estimation can be applied to estimate the size of a population. An "individual" animal in capture-recapture studies corresponds to a "species" in the richness estimation. A total of 142 captures were recorded for 76 distinct rabbits in 18 trapping nights. For these data, the incidence frequency counts (Q_1 to Q_7) were 43, 16, 8, 6, 0, 2, 1. The input data are read as follows: (18 1 43 2 16 3 8 4 6 6 2 7 1); each number needs to be separated by at least one blank space or separated by rows. Here the pair (1, 43) indicates that there are 43 unique species, the next pair (2, 16) indicates there are 16 duplicate species, and so on.

5. Type (2B) incidence-raw data (`ChaoSpeciesData$Inci_raw`)

In the cottontail capture-recapture experiments conducted by Edwards and Eberhardt (1967), a

total of 76 distinct individuals (regarded as 76 "species") were found in 18 trapping nights. The incidence-raw data consist of a capture/non-capture matrix (where "1" means a capture and "0" means a non-capture) with 76 rows and 18 columns. If we regard this capture-recapture matrix as a species-by-sampling-unit matrix, then species richness estimation can be applied to estimate the size of the cottontail population.

Usage

```
data(ChaoSpeciesData)
```

References

Magurran, A. E. (1988). Ecological Diversity and Its Measurement. Princeton University Press, Princeton, New Jersey.

Miller, R. I. and Wiegert, R. G. (1989). Documenting completeness, species-area relations, and the species-abundance distribution of a regional flora. Ecology, 70, 16-22.

Butler, B. J., and Chazdon, R. L. (1998). Species richness, spatial variation, and abundance of the soil seed bank of a secondary tropical rain forest. Biotropica, 30, 214-222.

Edwards, W. R. and Eberhardt, L. (1967). Estimating cottontail abundance from live trapping data. The Journal of Wildlife Management, 31, 87-96.

Diversity	*Estimation of species diversity (Hill numbers)*

Description

`Diversity`: Estimating a continuous diversity profile in one community including species richness, Shannon diversity and Simpson diversity). This function also supplies plots of empirical and estimated continuous diversity profiles. Various estimates for Shannon entropy and the Gini-Simpson index are also computed. All five types of data are supported: Type (1) abundance data (datatype="abundance"), Type (1A) abundance-frequency counts (datatype="abundance_freq_count"), Type (2) incidence-frequency data (datatype = "incidence_freq"), Type (2A) incidence-frequency counts (datatype="incidence_freq_count"), and Type (2B) incidence-raw data (datatype="incidence_raw"); see SpadeR-package details for data input formats.

Usage

```
Diversity(data, datatype = c("abundance", "abundance_freq_count",
  "incidence_freq", "incidence_freq_count", "incidence_raw"), q = NULL)
```

Arguments

`data`	a matrix/data.frame of species abundances/incidences.
`datatype`	type of input data, "abundance", "abundance_freq_count", "incidence_freq", "incidence_freq_count" or "incidence_raw".
`q`	a vector of nonnegative numbers specifying the diversity orders for which Hill numbers will be estimated. If NULL, then Hill numbers will be estimated at order q from 0 to 3 with increments of 0.25.

Value

a list of seven objects:

`$Basic_data` for summarizing data information.

`$Species_richness` for showing various species richness estimates along with related statistics.

`$Shannon_index` and `$Shannon_diversity` for showing various Shannon index/diversity estimates.

`$Simpson_index` and `$Simpson_diversity` for showing two Simpson index/diversity estimates.

`$Hill_numbers` for showing Hill number (diversity) estimates of diversity orders specified in the argument q.

References

Chao, A., and Chiu, C. H. (2012). Estimation of species richness and shared species richness. In N. Balakrishnan (ed). Methods and Applications of Statistics in the Atmospheric and Earth Sciences. p.76-111, Wiley, New York.

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.

Chao, A., Wang, Y. T. and Jost, L. (2013). Entropy and the species accumulation curve: a novel estimator of entropy via discovery rates of new species. Methods in Ecology and Evolution 4, 1091-1110.

Examples

```
## Not run:
data(DiversityData)
# Type (1) abundance data
```

```
Diversity(DiversityData$Abu,"abundance",q=c(0,0.5,1,1.5,2))
# Type (1A) abundance-frequency counts data
Diversity(DiversityData$Abu_count,"abundance_freq_count",q=seq(0,3,by=0.5))
# Type (2) incidence-frequency data
Diversity(DiversityData$Inci,"incidence_freq",q=NULL)
# Type (2A) incidence-frequency counts data
Diversity(DiversityData$Inci_freq_count,"incidence_freq_count",q=NULL)
# Type (2B) incidence-raw data
Diversity(DiversityData$Inci_raw,"incidence_raw",q=NULL)

## End(Not run)
```

DiversityData   *Data for Function Diversity*

Description

There are five data sets:

1. Type (1) abundance data (`DiversityData$Abu`)

The data include a column of the observed tree abundances/frequencies from an old-growth rain forest in Costa Rica (Chao et al. 2005, 2008). There were 69 tree species among 557 individuals.

2. Type (1A) abundance-frequency counts data (`DiversityData$Abu_count`)

The data consist of the observed beetles species abundance-frequency counts collected from the Osa old-growth forest site in Costa Rica (Janzen, 1973). There were 112 species among 237 individuals. The input abundance-frequency counts data are arranged as = (1 84 2 10 3 4 4 3 ... 42 1); each number needs to be separated by at least one blank space or separated by rows. Here the first pair (1, 84) indicates that there are 84 singletons, the second pair (2, 10) indicates there are 10 doubletons, and so on, with the last pair (42, 1) indicating that there is one species that is represented by 42 individuals.

3. Type (2) incidence-frequency data (`DiversityData$Inci`)

The single-column data include the observed incidence-based frequencies of tropical rainforest ants collected by Berlese extraction of soil samples (217 sampling units) in Costa Rica (Longino et al. 2002). In the input data, the entry in the first row denotes the number of sampling units (217); the subsequent 117 rows denote species incidence frequencies of the observed species.

4. Type (2A) incidence-frequency counts data (`DiversityData$Inci_freq_count`)

The seed-bank data consist of the observed species incidence-based frequency counts of seedlings that germinated from soil samples (Butler and Chazdon, 1998); here each soil sample is regarded as a sampling unit. A total of 34 species of seedlings were found in the 121 soil samples. The incidence frequency counts are read as = (121 1 3 2 2 3 3 ... 61 1); each number needs to be separated by at least one blank space or by separated by rows. The first entry, indicating that there are 121

soil samples, is followed by the 18 pairs (1, 3), (2, 2), (3, 3), (4, 3), (5, 1), (6, 5), and so on, up to (61, 1). Here (1, 3) indicates that there are 3 unique species, (2, 2) indicates there are 2 duplicate species, and so on, with (61, 1) indicating that there is one species found in 61 soil samples.

5. Type (2B) incidence-raw data (`DiversityData$Inci_raw`)

The data consist of raw incidence data of the seed-bank records, described above for the incidence frequency counts data. The input data include a 34 x 121 (species-by-sampling-unit) matrix. For each element of the matrix, "1" means a detection and "0" means a non-detection.

Usage

```
data(DiversityData)
```

References

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters, 8, 148-159.

Chao, A., Jost, L., Chiang, S.-C., Jiang, Y.-H. and Chazdon, R. L. (2008). A Two-stage probabilistic approach to multiple-community similarity indices. Biometrics, 64, 1178-1186.

Janzen, D. H. (1973) Sweep samples of tropical foliage insects: description of study sites, with data on species abundances and size distributions . Ecology, 54, 659-686.

Longino, J. T., Coddington, J. A. and Colwell, R. K. (2002). The ant fauna of a tropical rain forest: estimating species richness three different ways. Ecology, 83, 689-702.

Butler, B. J., and Chazdon, R. L. (1998). Species richness, spatial variation, and abundance of the soil seed bank of a secondary tropical rain forest. Biotropica, 30, 214-222.

Genetics	*Estimation of genetic differentiation measures*

Description

`Genetics`: Estimation allelic differentiation among subpopulations based on multiple-subpopulation genetics data. The richness-based indices include the classic Jaccard and Sorensen dissimilarity indices; the abundance-based indices include the conventional Gst measure, Horn, Morisita-Horn and regional species-differentiation indices.

Only Type (1) abundance data (datatype="abundance") is supported; input data for each sub-population include sample frequencies in an empirical sample of individuals. When there are multiple subpopulations, input data consist of an allele-by-subpopulation frequency matrix.

Usage

```
Genetics(X, q = 2, nboot = 200)
```

Arguments

X	a matrix, or a data.frame of allele frequencies.
q	a specified order to use to compute pairwise dissimilarity measures. If q = 0, this function computes the estimated pairwise Jaccard and Sorensen dissimilarity indices. If q = 1, this function computes the estimated pairwise equal-weighted and size-weighted Horn indices; If q = 2, this function computes the estimated pairwise Morisita-Horn and regional species-diffrentiation indices.
nboot	an integer specifying the number of bootstrap replications.

Value

a list of ten objects:

`$info` for summarizing data information.

`$Empirical_richness` for showing the observed values of the richness-based dis-similarity indices including the classic Jaccard and Sorensen indices.

`$Empirical_relative` for showing the observed values of the equal-weighted dis-similarity indices for comparing allele relative abundances including Gst, Horn, Morisita-Horn and regional differentiation measures.

`$Empirical_WtRelative` for showing the observed value of the dis-similarity index for comparing size-weighted allele relative abundances, i.e., Horn size-weighted measure based on Shannon-entropy under equal-effort sampling.

The corresponding three objects for showing the estimated dis-similarity indies are:
`$estimated_richness`, `$estimated_relative` and `$estimated_WtRelative`.

`$pairwise` and `$dissimilarity.matrix` for showing respectively the pairwise dis-similarity estimates (with related statistics) and the dissimilarity matrix for various measures depending on the diversity order q specified in the function argument.

`$q` for showing which diversity order q to compute pairwise dissimilarity.

References

Chao, A., and Chiu, C. H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. Methods in Ecology and Evolution, 7, 919-928.

Chao, A., Jost, L., Hsieh, T. C., Ma, K. H., Sherwin, W. B. and Rollins, L. A. (2015). Expected Shannon entropy and Shannon differentiation between subpopulations for neutral genes under the finite island model. Plos One, 10:e0125471.

Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. Molecular Ecology, 17, 4015-4026.

Examples

```
## Not run:
# Type (1) abundance data
data(GeneticsDataAbu)
Genetics(GeneticsDataAbu,q=2,nboot=200)

## End(Not run)
```

GeneticsDataAbu   *Human allele frequency data for Function Genetics*

Description

The data taken from Rosenberg et al. (2002) consist of allele frequencies from four human sub-populations (BiakaPyg, Palestin, Bedouin and Druze). The data are formatted as an allele (row) by subpopulation (column) matrix file. Entries in each row denote the frequencies of each allele found in the four subpopulations. The data include an observed allele frequency table with 27 rows and 4 columns.

Usage

```
data(GeneticsDataAbu)
```

References

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. and Feldman, M. W. (2002). Genetic structure of human populations. Science, 298, 2381-2385.

SimilarityMult   *Estimation of multiple-community similarity measures*

Description

`SimilarityMult`: Estimation various N-community similarity indices. The richness-based indices include the classic N-community Jaccard and Sorensen indices; the abundance-based indices include the Horn, Morisita-Horn, regional species-overlap, and the N-community Bray-Curtis indices. Three types of data are supported: Type (1) abundance data (datatype="abundance"), Type (2) incidence-frequency data (datatype="incidence_freq"), and Type (2B) incidence-raw data (datatype="incidence_raw"); see `SpadeR-package` details for data input formats.

Usage

```
SimilarityMult(X, datatype = c("abundance", "incidence_freq",
  "incidence_raw"), units, q = 2, nboot = 200, goal = "relative")
```

Arguments

X	a matrix/data.frame of species abundances/incidences.
datatype	type of input data, "abundance", "incidence_freq" or "incidence_raw".
units	number of sampling units in each community. For `datatype = "incidence_raw"`, users must specify the number of sampling units taken from each community. This argument is not needed for "abundance" and "incidence_freq" data.
q	a specified order to use to compute pairwise similarity measures. If `q = 0`, this function computes the estimated pairwise richness-based Jaccard and Sorensen similarity indices. If `q = 1` and goal=relative, this function computes the estimated pairwise equal-weighted and size-weighted Horn indices based on Shannon entropy; If `q = 1` and goal=absolute, this function computes the estimated pairwise Shannon-entropy-based measure for comparing absolute abundances. If `q = 2` and goal=relative, this function computes the estimated pairwise Morisita-Horn and regional species-overlap indices based on species relative abundances. If `q = 2` and goal=absolute, this function computes the estimated pairwise Morisita-Horn and regional species-overlap indices based on species absolute abundances.
nboot	an integer specifying the number of bootstrap replications.
goal	a specified estimating goal to use to compute pairwise similarity measures:comparing species relative abundances (goal=relative) or comparing species absolute abundances (goal=absolute).

Value

a list of fourteen objects:

`$datatype` for showing the specified data types (abundance or incidence).

`$info` for summarizing data information.

`$Empirical_richness` for showing the observed values of the richness-based similarity indices include the classic N-community Jaccard and Sorensen indices.

`$Empirical_relative` for showing the observed values of the equal-weighted similarity indices for comparing species relative abundances including Horn, Morisita-Horn and regional overlap measures.

`$Empirical_WtRelative` for showing the observed value of the Horn similarity index for comparing size-weighted species relative abundances based on Shannon entropy under equal-effort sampling.

`$Empirical_absolute` for showing the observed values of the similarity indices for comparing absolute abundances. These measures include the Shannon-entropy-based measure, Morisita-Horn and the regional species-overlap measures based on species absolute abundance, as well as the N-community Bray-Curtis index. All measures are valid only under equal-effort sampling.

The corresponding four objects for showing the estimated similarity indices are: `$estimated_richness`, `$estimated_relative`, `$estimated_WtRelative` and `$estimated_absolute`.

`$pairwise` and `$similarity.matrix` for showing respectively the pairwise dis-similarity estimates (with related statistics) and the similarity matrix for various measures depending on the diversity order q and the `goal` aspecified in the function arguments.

`$goal` for showing the goal specified in the argument goal (absolute or relative) used to compute pairwise similarity.

`$q` for showing which diversity order q specified to compute pairwise similarity.

References

Chao, A., and Chiu, C. H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. Methods in Ecology and Evolution, 7, 919-928.

Chao, A., Jost, L., Hsieh, T. C., Ma, K. H., Sherwin, W. B. and Rollins, L. A. (2015). Expected Shannon entropy and Shannon differentiation between subpopulations for neutral genes under the finite island model. Plos One, 10:e0125471.

Chiu, C. H., Jost, L. and Chao, A. (2014). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecological Monographs, 84, 21-44.

Gotelli, N. G. and Chao, A. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. Encyclopedia of Biodiversity, 2nd Edition, Vol. 5, 195-211, Waltham, MA.

Examples

```
## Not run:
data(SimilarityMultData)
# Type (1) abundance data
SimilarityMult(SimilarityMultData$Abu,"abundance",q=2,nboot=200,"relative")
# Type (2) incidence-frequency data
SimilarityMult(SimilarityMultData$Inci,"incidence_freq",q=2,nboot=200,"relative")
# Type (2B) incidence-raw data
SimilarityMult(SimilarityMultData$Inci_raw,"incidence_raw",
```

```
units=c(19,17,15),q=2,nboot=200,"relative")

## End(Not run)
```

SimilarityMultData   *Data for Function SimilarityMult*

Description

There are three data sets:

1. Type (1) abundance data (`SimilarityMultData$Abu`)

The data include the observed species frequencies of three communities: seedlings (column 1), saplings (column 2) and trees (column 3) collected from an old-growth rain forest; see Chao et al. (2005, 2008). The three entries in each row are the observed frequency (or abundance) of each species from the three communities.

2. Type (2) incidence-frequency data (`SimilarityMultData$Inci`)

The data include the observed incidence frequencies of tropical rainforest ants using three sampling techniques: (a) Berlese extraction of soil samples (217 samples), (b) fogging samples from canopy fogging (459 samples), and (c) Malaise trap samples for flying and crawling insects (62 samples); The data were collected in Costa-Rica (Longino et al. 2002). The three entries in the first row of the input data denote the number of sampling units (217, 459 and 62). Beginning with the second row, the three numbers in each row denotes incidence frequencies (the total number of detections) in the samples based on three sampling techniques.

3. Type (2B) incidence-raw data (`SimilarityMultData$Inci_raw`)

The data include the observed soil ciliate species detection/non-detection data for a total of 51 soil samples from three areas of Namibia, Africa: Etosha Pan (19 samples), Central Namib Desert (17 samples) and Southern Namib Desert (15 samples). The raw detection/non-detection data include 365 x 51 matrix of 0's and 1's (0 denotes a non-detection and 1 denotes a detection).

Usage

```
data(SimilarityMultData)
```

References

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters, 8, 148-159.

Chao, A., Jost, L., Chiang, S.-C., Jiang, Y.-H. and Chazdon, R. L. (2008). A Two-stage probabilistic approach to multiple-community similarity indices. Biometrics, 64, 1178-1186.

Longino, J. T., Coddington, J. A. and Colwell, R. K. (2002). The ant fauna of a tropical rain forest: estimating species richness three different ways. Ecology, 83, 689-702.

Foissner, W., Agatha, S. and Berger, H. (2002). Soil Ciliates (Protozoa, Ciliophora) from Namibia (Southwest Africa), with emphasis on two contrasting environments, the Etosha Region and the Namib Desert. Denisia, 5, 1-1459.

SimilarityPair	*Estimation of two-assemblage similarity measures*

Description

`SimilarityPair`: Estimation various similarity indices for two assemblages. The richness-based indices include the classic two-community Jaccard and Sorensen indices; the abundance-based indices include the Horn, Morisita-Horn, regional species-overlap, two-community Bray-Curtis and the abundance-based Jaccard and Sorensen indices. Three types of data are supported: Type (1) abundance data (datatype="abundance"), Type (2) incidence-frequency data (datatype="incidence_freq"), and Type (2B) incidence-raw data (datatype="incidence_raw"); see `SpadeR-package` details for data input formats.

Usage

```
SimilarityPair(X, datatype = c("abundance", "incidence_freq",
  "incidence_raw"), units, nboot = 200)
```

Arguments

X	a matrix/data.frame of species abundances/incidences.
datatype	type of input data, "abundance", "incidence_freq" or "incidence_raw".
units	number of sampling units in each community. For `datatype = "incidence_raw"`, users must specify the number of sampling units taken from each community. This argument is not needed for "abundance" and "incidence_freq" data.
nboot	an integer specifying the number of replications.

Value

a list of ten objects:

`$datatype` for showing the specified data types (abundance or incidence).

`$info` for summarizing data information.

`$Empirical_richness` for showing the observed values of the richness-based similarity indices include the classic two-community Jaccard and Sorensen indices.

`$Empirical_relative` for showing the observed values of the equal-weighted similarity indices for comparing species relative abundances including Horn, Morisita-Horn, regional overlap, Chao-Jaccard and Chao-Sorensen abundance (or incidence) measures based on species relative abundances.

`$Empirical_WtRelative` for showing the observed value of the Horn similarity index for comparing size-weighted species relative abundances based on Shannon entropy under equal-effort sampling.

`$Empirical_absolute` for showing the observed values of the similarity indices for comparing absolute abundances. These measures include the Shannon-entropy-based measure, Morisita-Horn and the regional overlap measures based on species absolute abundances, as well as the Bray-Curtis index. All measures are valid only under equal-effort sampling.

The corresponding four objects for showing the estimated similarity indices are: `$estimated_richness`, `$estimated_relative`, `$estimated_WtRelative` and `$estimated_Absolute`.

References

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters, 8, 148-159.

Chao, A., and Chiu, C. H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. Methods in Ecology and Evolution, 7, 919-928.

Chao, A., Jost, L., Hsieh, T. C., Ma, K. H., Sherwin, W. B. and Rollins, L. A. (2015). Expected Shannon entropy and Shannon differentiation between subpopulations for neutral genes under the finite island model. Plos One, 10:e0125471.

Chiu, C. H., Jost, L. and Chao, A. (2014). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecological Monographs, 84, 21-44.

Examples

```
## Not run:
data(SimilarityPairData)
# Type (1) abundance data
SimilarityPair(SimilarityPairData$Abu,"abundance",nboot=200)
# Type (2) incidence-frequency data
SimilarityPair(SimilarityPairData$Inci,"incidence_freq",nboot=200)
```

```
# Type (2B) incidence-raw data
SimilarityPair(SimilarityPairData$Inci_raw,"incidence_raw",units=c(19,17),nboot=200)

## End(Not run)
```

SimilarityPairData   *Data for Function SimilarityPair*

Description

There are three data sets:

1. Type (1) abundance data (`SimilarityPairData$Abu`)

The data include the observed species frequencies of two communities: seedlings (column 1), and trees (column 2) collected from an old-growth rain forest; see Chao et al. (2005, 2008). The two entries in each row are the observed frequency (or abundance) of each species from the two communities. (These data are subset of `SimilarityMultData$Abu` used in the function SimilarityMult.)

2. Type (2) incidence-frequency data (`SimilarityPairData$Inci`)

The data include the observed incidence frequencies of tropical rainforest ants based on two sampling techniques: (a) Berlese extraction of soil samples (217 samples), and (b) Malaise trap samples for flying and crawling insects (62 samples); see Longino et al. (2002). The two entries in first row of the input data denote the number of sampling units (217 and 62). Beginning with the second row, the two numbers in each row denotes incidence frequencies (the total number of detections) in the soil samples based on the two sampling techniques. (These data are subset of `SimilarityMultData$Inci` used in the function SimilarityMult.)

3. Type (2B) incidence-raw data (`SimilarityPairData$Inci_raw`)

The data include the observed soil ciliate species detection/non-detection data for a total of 36 soil samples from two areas of Namibia, Africa: Etosha Pan (19 samples), and Central Namib Desert (17 samples). The raw detection/non-detection data include 365 x 36 matrix of 0's and 1's (0 denotes a non-detection and 1 denotes a detection). (These data are subset of `SimilarityMultData$Inci_raw` used in the function SimilarityMult.)

Usage

```
data(SimilarityPairData)
```

References

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters, 8, 148-159.

Chao, A., Jost, L., Chiang, S.-C., Jiang, Y.-H. and Chazdon, R. L. (2008). A Two-stage proba-bilistic approach to multiple-community similarity indices. Biometrics, 64, 1178-1186.

Longino, J. T., Coddington, J. A. and Colwell, R. K. (2002). The ant fauna of a tropical rain forest: estimating species richness three different ways. Ecology, 83, 689-702.

Foissner, W., Agatha, S. and Berger, H. (2002) Soil Ciliates (Protozoa, Ciliophora) from Namibia (Southwest Africa), with emphasis on two Contrasting environments, the Etosha Region and the Namib Desert. Denisia, 5, 1-1459.

Index