Package 'ChaoEntropy'

October 21, 2013

Type Package

Title Statistical package 'ChaoEntropy'

Version 1.0

Date 2013-10-20

Author Y.T. Wang ; Y.H. Lee ; K.S. Tseng ; Anne Chao

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

Description The purpose of this R package ChaoEntropy is mainly to provide a new Shannon entropy estimator proposed by Chao et al. (2013) for both individual-based (abundance) data and sample-based (incidence) data.

License GPL-3

URL http://chao.stat.nthu.edu.tw/blog/

LazyLoad yes

R topics documented:

`ChaoEntropy-package` *Statistical package "***ChaoEntropy***"*

Description

The purpose of this R package **ChaoEntropy** is mainly to provide a new Shannon entropy estimator proposed by Chao et al. (2013) for both individual-based (abundance) data and sample-based (incidence) data.

Details

Package:	**ChaoEntropy**
Type:	Package
Version:	1.0
Date:	2013-10-20
License:	GPL-3
URL:	`http://chao.stat.nthu.edu.tw/blog/`

functions : `ChaoEntropy`, `Count2Abun`, `Count2Inci`, `Martix2Inci`

Author(s)

Y.T. Wang ; Y.H. Lee ; K.S. Tseng ; Anne Chao
Maintainer : Anne Chao `<chao@stat.nthu.edu.tw>`

`Ant_data` *Ant data*

Description

Species "incidence frequency" for ant data provided by Longino et al. (2002). The ant data which used malaise traps method is collected in a tropical rain forest of Costa Rica. In the ant data set presented 455 species occurrences are distributed among 62 samples with 103 species observed.

Usage

```
data(Ant_data)
```

Format

The first entry is the total number of sampling units, and followed by the incidences frequency. The format is:
c(62, 1, 1, 1, 1, 1, ..., 19, 19, 20, 29)

Source

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

Examples

```
data(Ant_data)
```

`Ant_data.count` *Ant data with incidence counts*

Description

Species "incidence counts" for ant data provided by Longino et al. (2002). The ant data which used malaise traps method is collected in a tropical rain forest of Costa Rica. In the ant data set presented 455 species occurrences are distributed among 62 samples with 103 species observed.

Usage

```
data(Ant_data.count)
```

Format

a numerical matrix or a data frame of two columns. The first column is the frequency j=1, 2...; and the second column is incidence counts (Qj).

The ant data is

[j]	[Qj]
1	39
2	18
3	13
...	...
19	2
20	1
29	1

Source

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

Examples

```
data(Ant_data.count)
```

`ChaoEntropy` *Estimation of Shannon entropy*

Description

`ChaoEntropy` is a function to provide a new Shannon entropy estimator proposed by Chao et al. (2013).

For individual-based (abundance) data, there are five other estimators which are Chao and Shen (2003) estimator; Grassberger (2003) estimator; Zhal (1977) jackknife estimator; Zhang (2012) Hz* estimator; and the observed provided in comparison.

For sample-based (incidence) data, new Shannon entropy estimator is in comparison with the observed entropy.

Usage

```
ChaoEntropy(data, datatype = c("abundance", "incidence"), method = c("all", "Chao",
            "ChaoShen", "Grassberger", "Jackknife", "Zhang", "Observed"),
            se = TRUE, nboot = 200, conf = 0.95)
```

Arguments

`data`	a vector of species abundance or incidence frequency. If `datatype = "incidence"`, then the input format of first entry should be total number of sampling units, and followed by species incidence frequency.
`datatype`	the data type of input data. That is individual-based abundance data (`datatype = "abundance"`) or sample-based incidence data (`datatype = "incidence"`).
`method`	the `method` constructed to estimate entropy (see Details)
`se`	calculate bootstrap standard error and show confidence interval; default is `TRUE`.
`nboot`	the number of bootstrap resampling times, default is `200`.
`conf`	a positive number ≤ 1. `"conf"` specifies the confidence level for confidence interval. The default is `0.95`.

Details

1. If `datatype = "abundance"`:
   - `method = "all"` : all estimators below involved. The default is `"all"`
   - `method = "Chao"` estimator, see Chao et al. (2013)
   - `method = "ChaoShen"` estimator, see Chao and Shen (2003)
   - `method = "Grassberger"` estimator, see Grassberger (2003)
   - `method = "Jackknife"` estimator, see Zhal (1977)
   - `method = "Zhang"` estimator, see Zhang (2012)
   - `method = "Observed"` estimator, the observed entropy estimator
2. If `datatype = "incidence"`:
   - `method = "all"` : all estimators below involved. The default is `"all"`
   - `method = "Chao"` estimator, see Chao et al. (2013) in Appendix S6
   - `method = "Observed"` estimator, the observed entropy estimator

Value

`ChaoEntropy` returns a table of various entropy estimators, their standard error and 95 % confidence interval which the `method` you choose.

Author(s)

Y.T. Wang ; Y.H. Lee ; K.S. Tseng ; Anne Chao
Maintainer : Anne Chao <chao@stat.nthu.edu.tw>

References

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

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

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

Grassberger, P. (2003) Entropy estimates from insufficient samplings. URL www.arxiv.org. arXiv:physics/0307138v2. Updated 2008.

Zahl, S. (1977) Jackknifing an index of diversity. *Ecology*, **58**, 907-913.

Zhang, Z. (2012) Entropy estimation in Turing's perspective. *Neural Computation*, **24**, 1368-1389.

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

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

See Also

Count2Abun, Count2Inci, Matrix2Inci

Examples

```
# load the individual-base (abundacne) data
data(Spider_data)

# Estimation of Shannon entropy
ChaoEntropy(Spider_data, datatype="abundance", method="all", se=TRUE, nboot=200, conf=0.95)

# load the sample-base (incidence) data
data(Ant_data)

# Estimation of Shannon entropy
ChaoEntropy(Ant_data, datatype="incidence", method="all", se=TRUE, nboot=200, conf=0.95)
```

Count2Abun *Frequency counts data transform into abundance data*

Description

Transfer frequency counts data into abundance data.

Usage

```
Count2Abun(count.data)
```

Arguments

`count.data` input species frequency counts data: a numerical matrix or a data frame of two columns. The first column is the frequency j=1, 2...; and the second column is frequency counts (fj).

Example: there are 59 singletons, 9 doubletons and so on in Insects_data:

```
[1,]  1  59
[2,]  2  9
[3,]  3  3
[4,]  4  2
[5,]  5  2
[6,]  6  2
[7,]  11  1
```

Value

a numeric vector of species abundance data.

Author(s)

Y.T. Wang ; Y.H. Lee ; K.S. Tseng ; Anne Chao
Maintainer : Anne Chao <chao@stat.nthu.edu.tw>

See Also

ChaoEntropy

Examples

```
# load the frequency counts data
data(Insects_data)

# Transform the data format
Count2Abun(Insects_data)
ChaoEntropy(data=Count2Abun(Insects_data), datatype="abundance", method="all",
            se=TRUE, nboot=200, conf=0.95)
```

Count2Inci *Incidence counts data transform into incidence frequency data*

Description

Transfer incidence counts data into incidence frequency data.

Usage

```
Count2Inci(count.data, t)
```

Arguments

count.data	input species incidence counts data: a numerical matrix or a data frame of two columns. The first column is the frequency j=1, 2...; and the second column is incidence counts (Qj). Example: there are 39 singletons, 18 doubletons and so on in Ant_data.count: [1,] 1 39 [2,] 2 18 [3,] 3 13 [..,] ... ... [16,] 19 2 [17,] 20 1 [18,] 29 1
t	the total number of sampling units. e.g. See Ant_data: t = 62

Value

a numeric vector which first entry is total number of sampling units, and followed by species incidence frequency.

Author(s)

Y.T. Wang ; Y.H. Lee ; K.S. Tseng ; Anne Chao
Maintainer : Anne Chao <chao@stat.nthu.edu.tw>

See Also

ChaoEntropy

Examples

```
# load the incidence counts data
data(Ant_data.count)
t <- 62


# Transform the data format
Count2Inci(Ant_data.count, t)
ChaoEntropy(data=Count2Inci(Ant_data.count, t), datatype="incidence", method="all",
            se=TRUE, nboot=200, conf=0.95)
```

Insects_data *Frequency Counts for Insects Data : Day-Time*

Description

To illustrate our method, we selected the data sets from Janzen (1973a, b) when he collected tropical foliage insects. The following table gives the frequencies for beetles collected respectively in daytime from the site referred to as "Osa primary-hill, dry season, 1967" in Janzen's paper (1973a).

Usage

```
data(Insects_data)
```

Format

a numerical matrix or a data frame of two columns.

[i]	[fi]
1	59
2	9
3	3
4	2
5	2
6	2
11	1

Examples

```
data(Insects_data)
```

`Matrix2Inci` *Presence/Absence data transform into incidence frequency data*

Description

Transfer prensence/absence data into incidence frequency data.

Usage

```
Matrix2Inci(mat.data)
```

Arguments

`mat.data`	a numerical matrix or a data frame, the presence/absence of each species is recorded in the matrix. If the species is presence, it is record 1. On the contrary, it is record 0. See `Seed_data`

Value

a numeric vector which first entry is total number of sampling units, and followed by species incidence frequency.

Author(s)

Y.T. Wang ; Y.H. Lee ; K.S. Tseng ; Anne Chao
Maintainer : Anne Chao <chao@stat.nthu.edu.tw>

See Also

`ChaoEntropy`

Examples

```
# load the prensence/absence data
data(Seed_data)

# Transform the data format
Matrix2Inci(Seed_data)
ChaoEntropy(data=Matrix2Inci(Seed_data), datatype="incidence", method="all",
            se=TRUE, nboot=200, conf=0.95)
```

`Seed_data` *Seed data*

Description

The seed-bank data proposed by Colwell and Coddington (1994) which contains 121 standardized soil samples with 34 species observed collected from one-hundred 10 m x 10 m grids in a Costa Rican forest.

Usage

```
data(Seed_data)
```

Format

Only the presence/absence of each species is recorded in the matrix. If the species is presence, it is record 1. On the contrary, it is record 0.

Source

Colwell, R.K. & Coddington, J. A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London B - Biological Sciences*, **345**, 101-118.

Examples

```
data(Seed_data)
```

`Spider_data` *Spider data*

Description

Sackett et al. (2011) provided species abundance data for samples of spiders from four 5 experimental forest canopy-manipulation treatments at the Harvard Forest.

Data from one treatments is used here for illustration: the 9 Hemlock Girdled treatment, in which bark and cambium of hemlock trees were cut and the trees 10 left in place to die to mimic tree mortality by adelgid infestation.

Usage

```
data(Spider_data)
```

Format

The format is: c(0, 15, 46, 2, 0, 0, 0, 1, 6, 1, ...)

Source

Sackett, T. E., S. Record, S. Bewick, B. Baiser, N. J. Sanders, & A. M. Ellison. 2011. Response of 5 macroarthropod assemblages to the loss of hemlock (Tsuga canadensis), a foundation species.

Examples

```
data(Spider_data)
```

Index