Package 'basa'

April 20, 2023

Type Package

Title Bayesian Analysis of Species Associations

Version 0.7.1

Date 2023-03-29

Author Manuel Mendoza [aut], Eduardo Gutierrez-Pena [cre].

Maintainer Eduardo Gutierrez-Pena <eduardo@sigma.iimas.unam.mx>

Description

Sampling-based Bayesian inference for some common association indices between two species.

License GPL (>= 3)

NeedsCompilation no

R topics documented:

basa *Bayesian Analysis of Species Associations*

Description

Sampling-based Bayesian inference for some common association indices between two species.

Details

This package generates simulated samples from the posterior distribution of association indices between two species based on presence-absence data. These samples are then used to conduct Bayesian inference about the indices. Built-in indices include Ochiai, Dice, Jaccard, Pearson, and Hamann (V ratio). User-defined indices can also be specified by defining suitable R functions.

Joint posterior inferences for two indices are also available.

Note

If no prior is given, the results are based on a noninformative prior for the population frequencies.

Author(s)

Manuel Mendoza [aut], Eduardo Gutierrez-Pena [cre].

Maintainer: Eduardo Gutierrez-Pena <eduardo@sigma.iimas.unam.mx>

References

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.* 57, 669–689.

Mendoza, M., Mendoza, E. and Gutierrez-Pena, E. (2021). Statistical analysis of species association indices. *Submitted*.

See Also

`cor.test`

Examples

```
# Data configuration
#          Present  Absent
# Present    a        b
# Absent     c        d

# Create the data vector
# n = c(a,b,c,d)

n <- c(3,2,19,16)

# Example 1
```

```
# Compute and display inferential summaries for the Ochiai index

psai(ochiai,n)

# Example 2

# Create a function to compute Simpson's index

simpson <- function(n){
  n[1] /  min(n[1]+n[2],n[1]+n[3])
}

# Compute and display inferential summaries for the new user-defined index

psai(simpson,n)

# Example 3

# Compute and display inferential summaries for the Pearson and Ochiai indices.

psai2(pearson,ochiai,n)
```

`con.tab` *Contingency table*

Description

Creates a 2 x 2 contingency table from presence-absence data of two species.

Usage

```
con.tab(species1, species2)
```

Arguments

`species1`	Presence-absence data of Species 1
`species2`	Presence-absence data of Species 2

Value

A vector of dimension 4 containing the contingency table, row by row.

Author(s)

Eduardo Gutierrez-Pena.

corrmat.pearson — *Pearson correlation matrix for a collection of indices*

Description

Posterior correlations for all pairs of a given set of indices.

Usage

```
corrmat.pearson(data, indices)
```

Arguments

`data`	A vector of dimension 4 containing the data table, row by row.
`indices`	A character vector containing the names of the indices.

Details

Correlations are calculated from samples of the posterior distribution of the indices.

Value

A matrix containing the Pearson correlation coefficients for all pairs of a given set of indices.

Author(s)

Eduardo Gutierrez-Pena.

corrmat.spearman — *Spearman correlation matrix for a collection of indices*

Description

Posterior correlations for all pairs of a given set of indices.

Usage

```
corrmat.spearman(data, indices)
```

Arguments

`data`	A vector of dimension 4 containing the data table, row by row.
`indices`	A character vector containing the names of the indices.

Details

Correlations are calculated from samples of the posterior distribution of the indices.

Value

A matrix containing the Spearman correlation coefficients for all pairs of a given set of indices.

Author(s)

Eduardo Gutierrez-Pena.

---

dice *Dice index*

Description

Dice index.

Usage

```
dice(n)
```

Arguments

`n`	A vector of dimension 4 containing the data table, row by row. The table must contain a cross-classification of two binary variables.

Details

Following Hubalek (1982), the Dice index is computed as

$$D = a/(a + (b + c)/2)$$

with n=c(a,b,c,d).

Value

The value of D, the Dice index. D takes values on [0,1].

Author(s)

Manuel Mendoza.

References

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.* 57, 669–689.

hamann *Hamann index*

Description

Hamann index.

Usage

```
hamann(n)
```

Arguments

`n`	A vector of dimension 4 containing the data table, row by row. The table must contain a cross-classification of two binary variables.

Details

Following Hubalek (1982), the Hamann index is computed as

$$H = ((a+d) - (b+c))/(a+b+c+d).$$

with n=c(a,b,c,d).

Value

The value of *H*, the Hamann index. *H* takes values on [-1,1].

Author(s)

Manuel Mendoza.

References

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.* 57, 669–689.

`jaccard` *Jaccard index*

Description

Jaccard index.

Usage

```
jaccard(n)
```

Arguments

`n`	A vector of dimension 4 containing the data table, row by row. The table must contain a cross-classification of two binary variables.

Details

Following Hubalek (1982), the Jaccard index is computed as

$$J = a/(a + b + c).$$

with n=c(a,b,c,d).

Value

The value of *J*, the Jaccard index. *J* takes values on [0,1].

Author(s)

Manuel Mendoza.

References

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.* 57, 669–689.

mi *Mutual information*

Description

Mutual information.

Usage

```
mi(n)
```

Arguments

`n`	A vector of dimension 4 containing the data table, row by row. The table must contain a cross-classification of two binary variables.

Details

Computes an estimate, *MI*, of the mutual information, which is defined as the Kullback-Leibler divergence between the joint distribution and the product of the respective marginal distributions of two binary variables (e.g., Cover and Thomas, 1991). mi(n) estimates these distributions using the relative frecuencies corresponding to the counts n=c(a,b,c,d).

Value

The value of *MI*, the mutual information. *MI* takes values on $[0,\infty)$.

Author(s)

Eduardo Gutierrez-Pena.

References

Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. Wiley.

Mendoza, M., Mendoza, E. and Gutierrez-Pena, E. (2021). Statistical analysis of species association indices. *Submitted*.

`ochiai` *Ochiai index*

Description

Ochiai index.

Usage

```
ochiai(n)
```

Arguments

`n`	A vector of dimension 4 containing the data table, row by row. The table must contain a cross-classification of two binary variables.

Details

Following Hubalek (1982), the Ochiai index is computed as

$$O = a/sqrt((a+b)*(a+c)).$$

with n=c(a,b,c,d).

Value

The value of *O*, the Ochiai index. *O* takes values on [0,1].

Author(s)

Manuel Mendoza.

References

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.* 57, 669–689.

`pearson` *Peason's correlation coefficient*

Description

Peason's correlation coefficient.

Usage

```
pearson(n)
```

Arguments

`n`	A vector of dimension 4 containing the data table, row by row. The table must contain a cross-classification of two binary variables.

Details

Following Hubalek (1982), Pearson's correlation coefficient is computed as

$$R = ((a*d) - (b*c))/sqrt((a+b)*(c+d)*(a+c)*(b+d)).$$

with n=c(a,b,c,d).

Value

The value of *R*, the Peason's correlation coefficient. *R* takes values on [-1,1].

Author(s)

Manuel Mendoza.

References

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.* 57, 669–689.

`posterior.mads` *Posterior median absolute deviations*

Description

Posterior median absolute deviation of a given index for each of the species.

Usage

```
posterior.mads(data.frame, index)
```

Arguments

`data.frame`	A data frame whose rows represent sites and whose columns represent species.
`index`	The name of the index

Details

Median absolute deviations are calculated from samples of the posterior distribution of the index for each of the species.

Value

A matrix containing the median absolute deviations of a given index for all pairs of species.

Author(s)

Eduardo Gutierrez-Pena.

`posterior.means` *Posterior mean*

Description

Posterior mean of a given index for each of the species.

Usage

```
posterior.means(data.frame, index)
```

Arguments

`data.frame`	A data frame whose rows represent sites and whose columns represent species.
`index`	The name of the index

Details

Means are calculated from samples of the posterior distribution of the index for each of the species.

Value

A matrix containing the means of a given index for all pairs of species.

Author(s)

Eduardo Gutierrez-Pena.

`posterior.medians` *Posterior medians*

Description

Posterior median of a given index for each of the species.

Usage

```
posterior.medians(data.frame, index)
```

Arguments

`data.frame`	A data frame whose rows represent sites and whose columns represent species.
`index`	The name of the index

Details

Medians are calculated from samples of the posterior distribution of the index for each of the species.

Value

A matrix containing the median of a given index for all pairs of species.

Author(s)

Eduardo Gutierrez-Pena.

posterior.sds *Posterior standard deviations*

Description

Posterior standard deviation of a given index for each of the species.

Usage

```
posterior.sds(data.frame, index)
```

Arguments

data.frame	A data frame whose rows represent sites and whose columns represent species.
index	The name of the index

Details

Standards deviations are calculated from samples of the posterior distribution of the index for each of the species.

Value

A matrix containing the standard deviations of a given index for all pairs of species.

Author(s)

Eduardo Gutierrez-Pena.

print.psa *print method for class 'psa'*

Description

S3 method for class 'psa'

Usage

```
## S3 method for class 'psa'
print(x, ...)
```

Arguments

x	An object of class 'psa'.
...	Other arguments.

Author(s)

Eduardo Gutierrez-Pena.

`psai` *Posterior sampling for an association index*

Description

Posterior summaries and an optional histogram of the posterior distribution of the association index.

Usage

```
psai(FUN = pearson,data=c(0,0,0,0),prior=c(0.5,0.5,0.5,0.5),probability=0.95,
     M=10000,plot=TRUE)
```

Arguments

`FUN`	Function computing the index.
`data`	A vector of dimension 4 containing the data table, row by row.
`prior`	A vector of dimension 4 containing the corresponding hyperparameters of the Dirichlet prior.
`probability`	The nominal value of the probability for the posterior credible interval.
`M`	The size of the simulated sample used for posterior computations.
`plot`	Should a histogram be produced?

Details

Computes posterior summaries of the association index based on simulated samples.

Value

A list containing the posterior mean, median, standard deviation and median absolute deviation, as well as a posterior credible interval for the association index. The list also includes the simulated posterior sample used for these calculations (hidden by default).

Note

If no prior is given, the results are based on a noninformative prior for the population frequencies.

Author(s)

Manuel Mendoza; Eduardo Gutierrez-Pena.

References

Mendoza, M., Mendoza, E. and Gutierrez-Pena, E. (2021). Statistical analysis of species association indices. *Submitted*.

Examples

```
# Data configuration
#         Present  Absent
# Present    a        b
# Absent     c        d

# Create the data vector
# species.data = c(a,b,c,d)

species.data <- c(3,2,19,16)

# Compute and display inferential summaries for the Ochiai index

psai(ochiai,species.data)
```

`psai2` *Joint posterior sampling for two association indices*

Description

Posterior summaries of the joint posterior distribution of two association indices. Optionally, display a scatter plot and histograms of the joint and marginal posterior distributions, respectively.

Usage

```
psai2(FUN1 = pearson, FUN2 = ochiai, data=c(0,0,0,0),prior=c(0.5,0.5,0.5,0.5),
      probability=0.95,M=10000,plot=TRUE)
```

Arguments

`FUN1, FUN2`	Functions computing the indices.
`data`	A vector of dimension 4 containing the data table, row by row.
`prior`	A vector of dimension 4 containing the corresponding hyperparameters of the Dirichlet prior.
`probability`	The nominal value of the probability for the posterior credible interval.
`M`	The size of the simulated sample used for posterior computations.
`plot`	Should a plot be produced?

Details

Computes posterior summaries of the association indices based on simulated samples.

Value

A list containing the posterior mean, median, standard deviation and median absolute deviation, as well as a posterior credible interval for each association index. The list also includes the posterior correlation between the two indices (Pearson and Spearman), as well as the simulated posterior samples used for all these calculations (hidden by default).

Note

If no prior is given, the results are based on a noninformative prior for the population frequencies.

Author(s)

Manuel Mendoza; Eduardo Gutierrez-Pena.

References

Mendoza, M., Mendoza, E. and Gutierrez-Pena, E. (2021). Statistical analysis of species association indices. *Submitted*.

Examples

```
# Data configuration
#         Present  Absent
# Present    a       b
# Absent     c       d

# Create the data vector
# n = c(a,b,c,d)

species.data <- c(3,2,19,16)

# Compute and display inferential summaries for the Pearson and Ochiai indices.

psai2(pearson,ochiai,species.data)
```

`results.table` *Posterior summaries*

Description

Posterior summaries for a given collection of indices.

Usage

```
results.table(data, indices)
```

Arguments

`data`	A vector of dimension 4 containing the data table, row by row.
`indices`	A character vector containing the names of the indices.

Value

A table contining the posteior mean, median, standard deviation and median absolute deviation, as well as a posterior credible interval for each association index.

Author(s)

Eduardo Gutierrez-Pena.

Index