Package ‘SDBP’

May 2, 2013

Type Package

Title Calculate the third-order accurate Unbiased P-values via Speedy double bootstrap method

Version 1.0

Date 2013-03-04

Author Aizhen Ren

Maintainer Aizhen Ren <ren_az@hotmail.co.jp>

Description Calculating the third-order unbiased p-values via speedy double bootstrap method.

URL http://www.bi.cs.titech.ac.jp/sdbp/

R topics documented:

`bp` *Bootstrap probabilities for phylogenetic trees*

Description

Calculate the bootstrap probabilities for phylogenetic trees

Usage

```
bpk(dat,k,nb=10000,...)
bp(dat,nb=10000,...)

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

Arguments

`dat`	a site-wise maximum log-likelihood matrix. Each column is for a tree, each row is for a site-wise maximum log-likelihood.
`k`	a specific tree number in the .mt file.
`nb`	the number of regular bootstrap replicates sampled from a multivariable normal distribution $N_T(y, \Sigma)$. T is the number of trees, y is the T-degree vector, which is the maximum log-likelihood vector consisting of the maximum log-likelihood for each tree, Σ is the unbiased covariance matrix of the maximum log-likelihood vector.
`...`	not used.
`x`	an object of class `bp`.

Value

An object of class `bpk`, basically a list including elements

`BP`	the bootstrap probability for a specific tree k.
`t`	a specific tree number k.

An object of class `bp`, basically a list including element

`pvs`	the bootstrap probabilities for the trees.

`print` method for an object of class `bp`; prints the bootstrap probabilities for the trees.

Author(s)

Ren Aizhen

Examples

```
library(SDBP)
data(mam20) ### data named mam20 stored in the file mam20.rda
result1<-bpk(mam20,2)
result1
## Not run:
result2<-bp(mam20)
result2

## End(Not run)
### the output of result1
## $BP
## [1] 0.0292

## $t
## [1] 2

### the output of result2 is in decreasing order of log-likelihood for
###   each tree
## Call:
## bp(dat = data)

## Bootstrap probabilities:
##      t1     t4     t3     t7     t2     t5     t9    t11    t10
```

```
## 0.4735 0.1934 0.1197 0.0927 0.0268 0.0422 0.0011 0.0256 0.0183
##         t6     t8    t14    t13    t12    t15
##  0.0010 0.0003 0.0000 0.0000 0.0000 0.0000
```

`dbpk` *Double bootstrap probability for a specific phylogenetic tree*

Description

Calculate the double bootstrap probability for a specific phylogenetic tree

Usage

```
dbpk(dat,k,nb1=1000,nb2=1000,nb=5000,...)
```

Arguments

`dat`	a site-wise maximum log-likelihood matrix. Each column is for a tree, each row is for a site-wise maximum log-likelihood.
`k`	a specific tree number in the .mt file.
`nb1`	the number of first-tier bootstrap replicates sampled from a multivariable normal distribution $N_T(\hat{\mu}, \Sigma)$. T is the number of trees, $\hat{\mu}$ is the T-degree vector, which is a projection of the maximum log-likelihood vector, consisting of the maximum log-likelihood for each tree, Σ is the unbiased variance matrix of the maximum log-likelihood vector.
`nb2`	the number of second-tier bootstrap replicates sampled from a multivariable normal distribution $N_T(y^*, \Sigma)$. T is the number of trees, y^* is the T-degree vector, which are first-tier bootstrap replicates, Σ is the unbiased variance matrix of the maximum log-likelihood vector.
`nb`	the number of regular bootstrap replicates sampled from a multivariable normal distribution $N_T(y, \Sigma)$. T is the number of trees, y is the T-degree vector, which is the maximum log-likelihood vector, consisting of the maximum log-likelihood for each tree, Σ is the unbiased variance matrix of the maximum log-likelihood vector.
`...`	not used.

Value

An object of class `dbpk`, basically a list including elements

`DBP`	the double bootstrap probability for a tree k.
`t`	a specific phylogenetic tree numbered k.

Author(s)

Ren Aizhen

Examples

```
## Not run:
library(SDBP)
data(mam20)### data named mam20 stored in the file mam20.rda
### 3 is the tree number ordered in file trees15.tpl
result<-dbpk(mam20,3)
result

## End(Not run)
### the output of result

## $DBP
##    t3
## 0.317

## $t
## [1] 3
```

`mam20` *Site-wise loglikelihoods scores for each of the mammalian mitochondrial amino acid sequences and the DNA sequences 12S and 16S rRNA genes*

Description

Site-wise loglikelihoods scores for each of the 20 mammalian species' mitochondrial amino acid sequences (12 individual proteins) and the DNA sequences 12S and 16S rRNA genes for 15 trees.

Usage

```
data(mam20)
```

Format

matrix of size 5879 × 15.

Details

See `source` section for `data files`.

Source

`http://www.bi.cs.titech.ac.jp/sdbp/`

References

Ren, A. et al. (2013) "Assessing statistical reliability of phylogenetic trees via a speedy double bootstrap method", *Molecular Phylogenetics and Evolution*, 67, 429-435.

Examples

```
## Not run:
## Reading the data
library(SDBP)
data(mam20)### data named mam20 stored in the file mam20.rda

## Speedy double bootstrap probabilities for 15 trees
mam20.pv <- sdbp.default(mam20)
mam20.pv

## End(Not run)
```

`sdbp` *Speedy double bootstrap probabilities for candidate phylogenetic trees*

Description

Calculates the speedy double bootstrap probabilities for candidate trees

Usage

```
## Default S3 method:
sdbp(dat,nb=10000,...)

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

## S3 method for class 'sdbp'
summary(object,...)
```

Arguments

`dat`	a site-wise maximum log-likelihood matrix. Each column is for a tree. Each row is for a site-wise maximum log-likelihood.
`nb`	the number of first-tier replicates sampled from a multivariable normal distribution $N_T(\hat{\mu}, \Sigma)$. T is the number of trees, $\hat{\mu}$ is the T-degree vector, it is a projection of the maximum log-likelihood vector, which consists of the maximum log-likelihood for each tree, Σ is the unbiased covariance matrix of the maximum log-likelihood vector.
`x`	an object of class `sdbp`.
`object`	an object of class `sdbp`.
`...`	not used.

Value

class `sdbp` returns an object, basically a list including elements

`pvs`	a named numeric of speedy double bootstrap probabilities for all candidate trees.

`nb`	same as argument `nb`.

`print` method for an object of class `sdbp`; prints the speedy double bootstrap probabilities for the trees. `summary` method for an object of class `sdbp`; gives the speedy double bootstrap probabilities and their standard errors for the trees.

Author(s)

Ren Aizhen

See Also

`sdbpk`.

Examples

```
## Not run:
library(SDBP)
data(mam20)### data named mam20 stored in the file mam20.rda
result <- sdbp.default(mam20)
result
summary(result)

## End(Not run)

### The results are output in decreasing order of log-likelihood for each tree
## > result
## Call:
## sdbp.default(dat = data)

## Speedy double bootstrap probabilities:
##     t1     t4     t3     t7     t2     t5     t9    t11    t10     t6     t8
## 0.7575 0.4258 0.3762 0.3330 0.2984 0.2592 0.1638 0.2164 0.1801 0.0472 0.0346
##    t14    t13    t12    t15
## 0.0101 0.0107 0.0121 0.0009

## > summary(result)
## Call:
## sdbp.default(dat = data)

##     StdErr   sDBP
## t1  0.0043 0.7575
## t4  0.0049 0.4258
## t3  0.0048 0.3762
## t7  0.0047 0.3330
## t2  0.0046 0.2984
## t5  0.0044 0.2592
## t9  0.0037 0.1638
## t11 0.0041 0.2164
## t10 0.0038 0.1801
## t6  0.0021 0.0472 *
## t8  0.0018 0.0346 *
## t14 0.0010 0.0101 *
## t13 0.0010 0.0107 *
## t12 0.0011 0.0121 *
## t15 0.0003 0.0009 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`sdbpk` *Speedy double bootstrap probability for a specific phylogenetic tree*

Description

Calculate the speedy double bootstrap probability for a specific phylogenetic tree

Usage

```
sdbpk(dat,k,nb=10000,...)

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

Arguments

`dat`	a site-wise maximum log-likelihood matrix. Each column is for a tree, each row is for a site-wise maximum log-likelihood.
`k`	a specific tree number in the .mt file.
`nb`	the number of first-tier replicates is sampled from the multivariable normal distribution $N_T(\hat{\mu}, \Sigma)$. T is the number of trees, $\hat{\mu}$ is the T-degree vector, it is a projection of the maximum log-likelihood vector, which consists of the maximum log-likelihood for each tree, Σ is the unbiased covariance matrix of the maximum log-likelihood vector.
`x`	an object of class `sdbpk`.
`...`	not used.

Value

An object of class `sdbpk`, basically a list including elements

`value`	a named numeric of the speedy double bootstrap probability for a tree k.

`print` method for an object of class `sdbpk`; prints the speedy double bootstrap probability for a tree k.

Author(s)

Ren Aizhen

See Also

`sdbp.default`.

Examples

```
library(SDBP)
data(mam20)### data named mam20 stored in the file mam20.rda
### 1 is the tree number ordered in the file trees15.tpl
result<-sdbpk(mam20,1)
result

### the output
## Call:
## sdbpk(dat = data, k = 1)

##    t1
## 0.7489
```

Index