Package 'INTRIGUE'

October 12, 2022

Type Package

Title Quantify and Control Reproducibility in High-Throughput
Experiments

Version 0.1.0

Maintainer Michael Kleinsasser <mkleinsa@umich.edu>

Description Estimate the proportions of the null and the reproducibility and
non-reproducibility of the signal group for the input data set. The Bayes factor
calculation and EM (Expectation Maximization) algorithm procedures are also
included.

License GPL-2

Encoding UTF-8

Imports SQUAREM, dplyr, rlist

Depends R (>= 3.5.0)

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Author Yi Zhao [aut],
Xiaoquan Wen [aut],
Michael Kleinsasser [cre]

Repository CRAN

Date/Publication 2020-11-05 16:50:03 UTC

R topics documented:

---

bf.approx *Bayes Factor Approximation*

Description

A function calculates the approximation for bayes factor, when the value of original bayes factor goes to infinity.

Usage

```
bf.approx(z, param, size, k2, oa2)
```

Arguments

z	The index for individual(i).
param	Input dataset.
size	Number of replicates(m).
k2	Grid value of k^2.
oa2	Grid value of ω^2.

Value

Approximation for bayes factor in log scale.

---

bf.cal.cefn *Bayes Factor Calculation Scheme for CENF prior*

Description

A function that calculates bayes factor for each data pair on each grid point in log scale.

Usage

```
bf.cal.cefn(data, hyperparam)
```

Arguments

data	A dataset which is constructed by pairs of coefficient values β and standard errors $se(\beta)$.
hyperparam	A two-dimensional vector denoting all the grid points, namely, k x ω.

Value

A list records all the log scale bayes factor values.

---

bf.cal.meta   *Bayes Factor Calculation Scheme for META prior*

Description

A function that calculates bayes factor for each data pair on each grid point in log scale.

Usage

```
bf.cal.meta(data, hyperparam = NULL, bf.only = FALSE)
```

Arguments

data
: A dataset which is constructed by pairs of coefficient values β and standard errors $se(\beta)$.

hyperparam
: A two-dimensional vector denoting all the grid points, namely, ϕ x ω.

bf.only
: A boolean, denoting whether this function is called to calculate Bayes factor for META prior only. Usually used when publication bias issue is the target.

Value

A list records all the log scale bayes factor values or a list records log scale bayes factor for null, reproducible and irreproducible model (when bf.only=TRUE).

---

bf.em   *Bayes Factor EM Updating Scheme*

Description

A function that describes the updating process in E step and M step for EM algorithm. It will be used in SQUAREM package.

Usage

```
bf.em(w, bf)
```

Arguments

w
: The weight vector in previous M step.

bf
: A vector recording all the bayes factor values in log scale.

Value

The updated weight vector in current M step(wnew).

---

bf.loglik *Bayes Factor Loglikelihood Function*

Description

Calculate the updated loglikelihood value in EM algorithm, and to evaluate whether converge or not.

Usage

```
bf.loglik(w, bf)
```

Arguments

w	The current weight vector
bf	A vector recording all the bayes factor values in log scale.

Value

Negative summation of loglikelihood values.

---

bf.weighted_sum *Bayes Factor Weighted Summation*

Description

A function calculates the weighted summation of bayes factor.

Usage

```
bf.weighted_sum(w, bf, i)
```

Arguments

w	Input weight vector.
bf	Input bayes factor vector
i	Individual index.

Value

Weighted sum for bayes factor in log scale.

---

hetero | *Heterogeneity Evaluation*

Description

Evaluating the overall and individually heterogeneity and reproducibility for the given individuals(units) shared in different replicates.

Usage

```
hetero(
  data,
  use_cefn = TRUE,
  rep = NULL,
  irre = NULL,
  phi_min = NULL,
  phi_max = NULL,
  sq_em_tol = 1e-04,
  fdr.level = NULL,
  sample_size = NULL
)
```

Arguments

data	A dataset which is constructed by pairs of coefficient values β and standard errors $se(\beta)$.
use_cefn	A boolean, denoting whether to use CEFN prior. If the value is TRUE, CEFN prior is used, else, META prior is applied. The default value is TRUE.
rep	A vector, denoting all the k^2 (under CEFN prior) or r (under META prior) values constructing the reproducible signals. If not specified, the default one is c(0.105,0.260,0.369), which corresponds to the several prior values satisfy that $Pr(\beta_{i,1}, \beta_{i,2} have a same sign) = 0.99, 0.975, 0.95$ for CEFN prior.
irre	A vector, denoting all the k^2 or r values constructing the irreproducible signals. If not specified, the default one is c(2.198, 3.636, 6.735), which corresponds to the several prior values satisfy that $Pr(\beta_{i,1}, \beta_{i,2} have a same sign) = 0.75, 0.70, 0.65$ for CEFN prior.
phi_min	A value which determines the maximum phi. If not specified, will be constructed from the input datasets.
phi_max	A value which determines the minimum phi. If not specified, will be constructed from the input datasets.
sq_em_tol	A small, positive scalar that determines when iterations should be terminated in squarem algorithm. The default value is $1e-4$.
fdr.level	The user-defined rejection level for false discovery rate.
sample_size	The user-defined sample size.

Value

A list with the following components:

gridweight	The final optimal weight vector evaluated on each grid point.
ind_prob	A matrix denoting the converged probability for each individual being inside the three different groups, namely, the null, the reproducible and the irreproducible group.
est_prop	The estimated proportion value for the three different groups, namely, the null, the reproducible and the irreproducible group.
lfdr	The local false discovery rate based on the null hyppthesis of unit belonging to H_R, reproducible group. $lfdr = 1 - Pr(H_R)$
significant	If fdr.level is specified, a significant object recording True or False will be returned

Examples

```
data("heterodata")

hetero.out<-hetero(heterodata,fdr.level=0.05)
names(hetero.out)
print(hetero.out$est_prop)


## for CRAN check
hetero.out<-hetero(heterodata[1:100,],fdr.level=0.05)
```

---

hetero.lfdr   *Local False Discovery Rate Evaluation*

Description

Local False Discovery Rate Evaluation

Usage

```
hetero.lfdr(cat, fdr.level)
```

Arguments

cat	The final individual-level probabilities of falling into three categories, separately.
fdr.level	Rejection level for Local false discovery rate, if not specified, the rejection decision procedure won't be run.

Value

A list that preserves local false discovery rate and 5tthe corresponding reject decision if called.

---

heterodata   *Simulation Dataset*

Description

This is a simulation dataset, containing n=5000 units and m=2 replicates. The true proportion for the null, the reproducible and the irreproducible group is 0.80, 0.18, 0.02 separately.

Usage

```
data("heterodata")
```

Format

An object of class matrix (inherits from array) with 1000 rows and 4 columns.

Examples

```
data("heterodata")
```

Index