Package 'CorrectedFDR'

October 12, 2022

Type Package

Title Correcting False Discovery Rates

Version 1.1

Date 2021-10-06

Author Abbas Rahal, Anna Akpawu, Justin Chitpin and David R. Bickel

Maintainer Abbas Rahal <Abbas.Rahal13@gmail.com>

Description

There are many estimators of false discovery rate. In this package we compute the Nonlocal False Discovery Rate (NFDR) and the estimators of local false discovery rate: Corrected False discovery Rate (CFDR), Re-ranked False Discovery rate (RFDR) and the blended estimator.
Bickel, D.R., Rahal, A. (2019) <https://tinyurl.com/kkdc9rk8>.

Depends R(>= 2.14.2)

Suggests LFDR.MLE, LFDREmpiricalBayes, ProData

biocViews Bayesian Statistics, MathematicalBiology

License LGPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2021-10-07 13:50:06 UTC

R topics documented:

CorrectedFDR-package 2
BlendedLFDR 3
EstimatorsFDR 4

Index **7**

`CorrectedFDR-package` *Correcting False Discovery Rates*

Description

There are many estimators of false discovery rate. In this package we compute the Nonlocal False Discovery Rate (NFDR) and the estimators of local false discovery rate: Corrected False discovery Rate (CFDR), Re-ranked False Discovery rate (RFDR) and the blended estimator. Bickel, D.R., Rahal, A. (2019) <https://tinyurl.com/kkdc9rk8>.

Details

The DESCRIPTION file:

Package:	CorrectedFDR
Type:	Package
Version:	1.1
Date:	2021-10-06
License:	GPL-3
Depends:	R(>= 2.14.2)
Suggests:	LFDR.MLE, LFDREmpiricalBayes, ProData

Two functions in `CorrectedFDR` package to compute the LFDR estimators. The function `EstimatorsFDR` computes the nonlocal false discovery rate (NFDR), the CFDR and the RFDR. The function `BlendedLFDR` uses a Benchmark of FDR, and other estimators of LFDR in order to get an estimate of LFDR.

Author(s)

Abbas Rahal, Anna Akpawu, Justin Chitpin and David R. Bickel

Maintainer: Abbas Rahal <Abbas.Rahal13@gmail.com>

References

Bickel, D.R., Rahal, A. (2019). Correcting false discovery rates for their bias toward false positives. Communications in Statistics - Simulation and Computation, https://tinyurl.com/kkdc9rk8.

Bickel, D. R. (2015). Corrigendum to: Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions. Statistical Applications in Genetics and Molecular Biology, 2015, 14, 225.

Bickel, D. R. (2015). Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing. Statistical Methods and Applications, 24(4), pp. 523-546.

Bickel, D. R. (2013). Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions. Statistical Applications in Genetics and Molecular Biology, 2013, 12, 529-543.

`BlendedLFDR` *Blended Estimator of Local False Discovery Rate (LFDR)*

Description

BlendedLFDR is a function used to compute the blended estimator based on a benchmark estimator, usually the nonlocal false discovery rate (NFDR), and a set of estimators of local false discovery rates (LFDR).

Usage

```
BlendedLFDR(Benchmark, EstLFDR)
```

Arguments

`Benchmark`	Input numeric vector for benchmark estimator (often NFDR).
`EstLFDR`	Input a matrix containing two or more sets of LFDR estimators.

Details

`Benchmark` is an estimator of the FDR. This is usually the nonlocal false discovery rate (NFDR).
`EstLFDR` is a matrix of several LFDR estimators such as corrected FDR (CFDR), re-ranked FDR (RFDR), MLE (Maximum Likelihood Estimator), BBE1(Binomial Based Estimator), etc.
The output returns a single numeric vector containing the blended estimator of the LFDR.

Value

The value of the blended estimator is an estimator of the LFDR.

Note

The number of rows for the `Benchmark` and `EstLFDR` must have equal lengths.

Author(s)

Code: Abbas Rahal.
Documentation: Anna Akpawu, Justin Chitpin and Abbas Rahal.
Maintainer: Abbas Rahal <Abbas.Rahal13@gmail.com>

References

Bickel, D. R. (2015). Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing. Statistical Methods and Applications, 24(4), pp. 523-546.

Examples

```
#The data used to compute the LFDR estimators (CFDR, RFDR, MLE, and BBE1)
#comes from the ER/PR breast cancer data from the "ProData" package.
#To read more about the data, visit the website: https://www.bioconductor.org/
#Test statistics were first obtain, then the estimators for the FDR and LFDR were estimated.
#Benchmark vector
NFDR<-c(0.5661106448, 0.6897735492, 0.0000288516, 0.1549745113, 0.1305508970, 0.2421032979,
0.1482335568, 1, 1, 1, 0.6602562820, 0.7034682859, 0.7036332234, 0.0071192090,
0.8204536037, 0.9757716498, 0.7379329991, 1, 0.6333245479, 0.9904389701)
#Estimators of LFDR
CFDR<- c(1, 1, 0.0000288516, 0.2841199373, 0.2980912149, 0.5931530799, 0.3088199101,
1, 1, 1, 1, 1, 1, 0.0106788135, 1, 1, 1, 1, 1, 1)

RFDR<- c(0.689773549, 1, 0.007119209, 0.130550897, 0.703633223, 0.660256282, 0.242103298,
1, 1, 1, 0.820453604, 1, 0.703468286, 0.154974511, 1, 1, 1, 1, 0.975771650,1)

MLE<- c(0.9865479126, 0.9969935995, 0.0002372158, 0.6531633437, 0.7611453549, 0.9187425383,
0.7359259207, 0.9996548155, 0.9997310453, 0.9997437131, 0.9944712582, 0.9981685029,
0.9937604664, 0.0215892618, 0.9990504315, 0.9997493086, 0.9967673540, 0.9997016985,
0.9970142319, 0.9997625673)

BBE1<- c(1,1, 0.0003169812, 0.1138333734, 1, 1, 1, 1, 1, 1, 0.3279109564, 1, 0.0504755806,
0.0091823115, 0.0182614994, 0.0165386682, 1, 0.6964403713, 0.1001337298, 0.8415641198 )

#Matrix of LFDR Estimators
Est.LFDR<- matrix(c(CFDR,RFDR,MLE,BBE1), ncol=4)
output<-BlendedLFDR(Benchmark = NFDR, EstLFDR = Est.LFDR)
output$Blended
```

`EstimatorsFDR` *Estimators of Local False Discovery Rate (LFDR)*

Description

EstimatorsFDR is an R function that computes the Nonlocal False Discovery Rate (NFDR) and the estimators of local false discovery rate: Corrected False discovery Rate (CFDR) and Re-ranked False Discovery rate (RFDR).

Usage

```
EstimatorsFDR(pvalue)
```

Arguments

`pvalue`	Input numeric vector of pvalues.

Details

The input is a list of pvalues. The pvalues can be obtained for example by performing Student's t-test between two datasets. The two groups can be data from healthy and disease states. Let $i = 1, 2, ..., N$, where i represents the ith feature (SNP or gene, for example). Then, for each i, the hypothesis indicator A_i can have two possible values.
$A_i = 0$, if the ith null hypothesis is true, or
$A_i = 1$, if the ith null hypothesis is not true,
where the null hypothesis is defined by: the ith feature is unaffected by a treatment, unassociated with a disease, etc. The values for each estimator (NFDR, CFDR, RFDR) indicate the probability that the null hypothesis of the ith feature is true ($A_i = 0$) given the statistics T_i. The alternative hypothesis is true if $A_i = 1$. For example, in gene expression data analysis, if the null hypothesis is true, this would mean that the genes are not differentially expressed.

Value

The output returns three lists. It returns the NFDR, CFDR, and RFDR estimators:

NFDR	nonlocal FDR
CFDR	corrected FDR
RFDR	re-ranked FDR

Author(s)

Code: Abbas Rahal.
Documentation: Anna Akpawu, Justin Chitpin and Abbas Rahal.
Maintainer: Abbas Rahal <Abbas.Rahal13@gmail.com>

References

Bickel, D.R., Rahal, A. (2019). Correcting false discovery rates for their bias toward false positives. Communications in Statistics - Simulation and Computation, https://tinyurl.com/kkdc9rk8.

Bickel, D. R. (2015). Corrigendum to: Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions. Statistical Applications in Genetics and Molecular Biology, 2015, 14, 225.

Bickel, D. R. (2013). Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions. Statistical Applications in Genetics and Molecular Biology, 2013, 12, 529-543.

Examples

```
#The examples below are from the "ProData" package.
#In order to use the "Prodata" input you would first need to install the ProData package.
#You will also need the function exprs in this package.
#First, make sure that the ProData package is properly installed:
#source("https://bioconductor.org/biocLite.R")
#biocLite("ProData")
#library(ProData)
#data("f45cbmk")
```

```
#q1<- quantile(as(exprs(f45cbmk[, pData(f45cbmk)$GROUP == "B"]), "numeric"), probs = 0.25)
#logish<- function(x){log(x + q1)}
#Vectors of proteins for 20 patients ER/PR-positive and Healthy
#Y<- logish(exprs(f45cbmk[, pData(f45cbmk)$GROUP == "B"])) # Control (Healthy)
#X.ER<- logish(exprs(f45cbmk[, pData(f45cbmk)$GROUP == "C"])) # Case ER/PR-positive
#pvalue<- NULL
#for (i in 1:nrow(X.ER))
#{
# t<-t.test(x=X.ER[i,], y=Y[i,], alternative = "two.sided")
# pvalue[i]<- t$p.value
#}

#The pvalues obtained from the t-test:
pvalue<- c(0.1981, 0.3794, 0.000001443, 0.02325, 0.03264, 0.07263, 0.02965, 0.8016, 0.8888,
       0.9133, 0.2971, 0.4573, 0.2815, 0.0007119, 0.5743, 0.927, 0.369, 0.8478, 0.38, 0.9904)
output<- EstimatorsFDR(pvalue)
#Three lists
output$NFDR
output$CFDR
output$RFDR
```

Index

∗ **BBE1**
BlendedLFDR, 3
∗ **Blended**
BlendedLFDR, 3
∗ **CFDR**
BlendedLFDR, 3
EstimatorsFDR, 4
∗ **False Discovery Rate**
BlendedLFDR, 3
EstimatorsFDR, 4
∗ **LFDR**
BlendedLFDR, 3
EstimatorsFDR, 4
∗ **MLE**
BlendedLFDR, 3
∗ **NFDR**
BlendedLFDR, 3
EstimatorsFDR, 4
∗ **Null**
EstimatorsFDR, 4
∗ **RFDR**
BlendedLFDR, 3
EstimatorsFDR, 4
∗ **hypothesis**
EstimatorsFDR, 4
∗ **package**
CorrectedFDR-package, 2

BlendedLFDR, 3

CorrectedFDR-package, 2

EstimatorsFDR, 4