Package 'Jlfdr'

May 28, 2016

Type Package

Title Jlfdr: Controlling the Joint Local False Discovery Rate in Joint Analysis of Summary Statistics from Multiple GWASs

Version 1.0

Date 2016-05-26

Author Wei Jiang and Weichuan Yu

Maintainer Wei Jiang <wjiangaa@connect.ust.hk>

Description

The most powerful summary-statistics-based joint analysis method when controlling the false discovery rate at a certain level. We can use it to analyze multiple GWASs with the same phenotype together to discover associated genetic variants with higher power.

License GPL-3

Depends R (>= 2.10), mvtnorm

NeedsCompilation no

R topics documented:

`Jlfdr-package` *Jlfdr: Controlling the Joint Local False Discovery Rate in Joint Analysis of Summary Statistics from Multiple GWASs*

Description

The most powerful summary-statistics-based joint analysis method when controlling the false discovery rate at a certain level. We can use it to analyze multiple GWASs with the same phenotype together to discover associated genetic variants with higher power.

Details

Package:	Jlfdr
Type:	Package
Version:	1.0
Date:	2016-05-26
License:	GPL-3

In genome-wide association studies (GWASs) of common diseases/traits, we often analyze multiple GWASs with the same phenotype together to discover associated genetic variants with higher power. Since it is diffcult to access data with detailed individual measurements, summary-statistics-based meta-analysis methods have become popular to jointly analyze data sets from multiple GWASs.

Here we implement a novel summary-statistics-based joint analysis method based on controlling the joint local false discovery rate (Jlfdr). This method is the most powerful summary-statistics-based joint analysis method when controlling the false discovery rate (Fdr) at a certain level. Details about the method can be seen in our reference paper below.

The principal component of Jlfdr package is `FdrControl2`. Also we implement a Fdr controlling method `BayesFdr` for single-study analysis.

1. To jointly analyze summary statistics from multiple GWASs, we need to obtain the z-values of each genotyped SNPs in all studies. We have put example summary statistics of two studies (SmryStats1 and SmryStats2) in our package. You can use `data(SmryStats1)` and `data(SmryStats2)` to load the example data. You can also obtain the ground-truth parameters (allele frequencies, odds ratios) of the example data using `data(Param1)` and `data(Param2)`

2. You can use `FdrControl2` to jointly analyze summary statistics from two stuides.

```
FdrControl2(z1, z2, K=2, q=0.05, beta0=length(z1)/5, plot=T, output=T, dir=output)
```

Details about the function can be seen using `help(FdrControl2)`.

3. To analyze summary statistics from single study with controlling Fdr, `BayesFdr` can be used.

```
BayesFdr(z, q, K = 2, beta0 = length(z)/5)
```

Details about the function can be seen using `help(BayesFdr)`

Author(s)

Wei Jiang and Weichuan Yu

Maintainer: Wei Jiang <wjiangaa@connect.ust.hk>

References

Jiang, W. and Yu, W. *Controlling the joint local false discovery rate is more powerful than meta-analysis methods in joint analysis of summary statistics from multiple genome-wide association studies*. Submitted.

See Also

`FdrControl2`, `BayesFdr`

Examples

```
q<-5e-5                  #Fdr Threshold

##Load data
data(SmryStats1)        #Example of summary statistics in 1st study
data(SmryStats2)        #Example of summary statistics in 2nd study

z1<-SmryStats1$Z          #Z values in 1st study
z2<-SmryStats1$Z          #Z values in 2nd study

######  Jlfdr-based method  ######
FdrResult<-FdrControl2(z1,z2, K=2, beta0=length(z1)/5, q=q,dir=.)
rejectedIdx<-FdrResult$rejected  #The indexes of the rejected null hypotheses
```

BayesFdr-functions *Functions to control Fdr in single-study analysis*

Description

BayesFdr is a (Bayesian) Fdr controlling method used in single-study analysis. snpEM is a built-in function to infer the parameters in the one-dimensional Gaussian Mixture Model. calBayesFdr is a built-in function to calculate the Fdr after given parameters in the Gaussian Mixture Model.

Usage

```
BayesFdr(z, q, K = 2, beta0 = length(z)/5)

snpEM(z, K=2, maxIter=1000, tol=1e-4, beta0=length(z)/5, info=TRUE)

calBayesFdr(t, Pi1, sigma2)
```

Arguments

z	The vector containing z values of all SNPs.
q	The Fdr controlling level.
K	The components number of the z values in associated SNPs.
beta0	The penalty term for pi0. We add Dirichlet(beta0, 0) prior to proportions (pi0, Pi1).
maxIter	The maximum number of iterations.
tol	The relative error tolerance.
info	A logical number indicating whether the results should be printed out.
t	The critical value for z-values.
Pi1	The vector including the proportion of each associated component.
sigma2	The vector including the variance of standardized effect sizes in each associated component.

Details

These functions are the implementation of controlling the Fdr in single-study analysis. Here we assume the z-values follow the (K+1)-component Gaussian Mixture Model:

Z~pi_0 N(0,1)+sum_(k=1)^K pi_1k N(0, 1+sigma2_k).

`snpEM` is a built-in function to infer the parameters in the one-dimensional Gaussian Mixture Model using EM-algorithm.

`calBayesFdr` is a built-in function to calculate the Fdr after given parameters in the Gaussian Mixture Model.

`BayesFdr` is a Fdr controlling method used in single-study analysis. `snpEM` and `calBayesFdr` are called in `BayesFdr`.

Value

`BayesFdr` returns the critical value and the indexes of the rejected null hypotheses. The returned value is a LIST:

`thld`	The determined critical value in terms of z-value.
`idx`	The indexes of the rejected null hypotheses.

`snpEM` returns the inferred parameters and the iteration status. The returened value is a LIST:

`pi0`	The infered proportion of true null hypotheses.
`Pi1`	The vector including the inferred proportion of each associated component.
`sigma2`	The vector including the inferred variance of standardized effect sizes in each associated component.
`h0`	The vector including the probability of being true null hypothesis for each SNP.
`h`	A K-vector list. Each vector includes the probability of being component k (1<=k<=K) for each SNP.
`iter`	The iteration number.
`Qval`	The expected negative log-likelihood.

`calBayesFdr` returns the Fdr value for a given threshold t.

Author(s)

Wei Jiang and Weichuan Yu

Maintainer: Wei Jiang <wjiangaa@connect.ust.hk>

References

Jiang, W. and Yu, W. *Controlling the joint local false discovery rate is more powerful than meta-analysis methods in joint analysis of summary statistics from multiple genome-wide association studies*. Submitted.

See Also

`Jlfdr`, `FdrControl2`

Examples

```
q<-5e-5                  #Fdr Threshold

##Load data
data(SmryStats1)        #Example of summary statistics in 1st study

z1<-SmryStats1$Z          #Z values in 1st study

####Bayes Fdr Control###
FdrResult<-BayesFdr(z1, q=q, K=2, beta0=length(z1)/5)
rejectedIdx<-FdrResult$idx #The indexes of the rejected null hypotheses
```

`Jlfdr-functions` *Functions to control Fdr in joint analysis of multiple GWASs via Jlfdr-based method*

Description

`FdrControl2` is a novel summary-statistics-based joint analysis method to infer which SNPs have associations with traits. The method is based on controlling the joint local false discovery rate (Jlfdr). `snpEM2` is a built-in function to infer the parameters in the Gaussian Mixture Model. `calJlfdr2` is a built-in function to calculate the Jlfdr of each SNP after given parameters in the Gaussian Mixture Model.

Usage

```
FdrControl2(z1, z2, K = 2, q = 0.05, beta0 = length(z1)/5, plot = T,
 output = T, dir = "output")

snpEM2(z1, z2, K=2, maxIter=1000, tol=1e-4, beta0=length(z1)/5, info=TRUE)

calJlfdr2(z1, z2, Pi1, Sigma)
```

Arguments

`z1`	The vector containing z values of all SNPs from Study 1.
`z2`	The vector containing z values of all SNPs from Study 2.
`K`	The components number of the z values in associated SNPs.
`q`	The Fdr controlling level.
`beta0`	The penalty term for pi0. We add Dirichlet(beta0, 0) prior to proportions (pi0, Pi1).
`plot`	A logical number indicating whether to plot the discovered associations in the z-value plane.
`output, info`	A logical number indicating whether the results should be printed out.
`dir`	The directory to save the results when `output=T`
`maxIter`	The maximum number of iterations.
`tol`	The relative error tolerance.
`Pi1`	The vector including the proportion of each associated component.
`Sigma`	A K-objects list. Each object is the covariance matrix of standarized effect size vector in the corresponding associated component.

Details

These functions are the implementation of Jlfdr-based method to control the Fdr in joint analysis of multiple GWASs. Here we assume the z-value vector follow the (K+1)-component Gaussian Mixture Model:

Z~pi_0 N(0,I)+sum_(k=1)^K pi_1k N(0, I+Sigma_k).

`snpEM2` is a built-in function to infer the parameters in the Gaussian Mixture Model using EM-algorithm.

`calJlfdr2` is a built-in function to calculate the Jlfdr of each SNP after given parameters in the Gaussian Mixture Model.

`FdrControl2` is based on controlling the joint local false discovery rate (Jlfdr) in the joint analysis of multiple GWASs. `snpEM2` and `calJlfdr2` are called in `FdrControl2`.

Value

`FdrControl2` returns the following LIST:

`Pi1`	The vector including the inferred proportion of each associated component.
`Sigma`	A K-objects list. Each object is the inferred covariance matrix of standarized effect size vector in the corresponding associated component.
`rejected`	The indexes of the rejected null hypotheses.
`JlfdrThld`	The determined threshold for Jlfdr.
`Jlfdr`	The vector including the calculated Jlfdr value for each SNP.

`snpEM2` returns the infered parameters and the iteration status. The returened value is a LIST:

`pi0`	The infered proportion of true null hypotheses.
`Pi1`	The vector including the infered proportion of each associated component.
`Sigma`	A K-objects list. Each object is the inferred covariance matrix of standarized effect size vector in the corresponding associated component.
`h0`	The vector including the probability of being true null hypothesis for each SNP.
`h`	A K-vector list. Each vector includes the probability of being component k (1<=k<=K) for each SNP.
`iter`	The iteration number.
`Qval`	The expected negative log-likelihood.

`calJlfdr2` returns the calculated Jlfdr value for each SNP.

Author(s)

Wei Jiang and Weichuan Yu

Maintainer: Wei Jiang <wjiangaa@connect.ust.hk>

References

Jiang, W. and Yu, W. *Controlling the joint local false discovery rate is more powerful than meta-analysis methods in joint analysis of summary statistics from multiple genome-wide association studies*. Submitted.

See Also

Jlfdr, BayesFdr

Examples

```
q<-5e-5                  #Fdr Threshold

##Load data
data(SmryStats1)        #Example of summary statistics in 1st study
data(SmryStats2)        #Example of summary statistics in 2nd study

z1<-SmryStats1$Z          #Z values in 1st study
z2<-SmryStats1$Z          #Z values in 2nd study

######  Jlfdr-based method  ######
FdrResult<-FdrControl2(z1,z2, K=2, beta0=length(z1)/5, q=q,dir=.)
rejectedIdx<-FdrResult$rejected #The indexes of the rejected null hypotheses
```

Index