

Package ‘RFdr’

July 10, 2016

Type Package

Title RFdr: Jointly determining significance levels for primary and replication studies in two-stage GWASs

Version 1.0

Date 2016-06-24

Author Wei Jiang and Weichuan Yu

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

Description In two-stage GWASs, we consider the associations showing significance in both primary and replication studies as true findings. An important question under this two-stage setting is how do we determine the significance levels in both studies? Here, we propose a novel method to determine significance levels jointly. It finds the most powerful significance levels when controlling the false discovery rate (Fdr) in the two-stage study at a certain level.

License GPL-3

Depends mvtnorm

R topics documented:

RFdr-package	1
RFdrControl	3
Index	6

RFdr-package	<i>RFdr: Jointly determining significance levels for primary and replication studies in two-stage GWASs</i>
--------------	---

Description

In two-stage GWASs, we consider the associations showing significance in both primary and replication studies as true findings. An important question under this two-stage setting is how do we determine the significance levels in both studies? Here, we propose a novel method to determine significance levels jointly. It finds the most powerful significance levels when controlling the false discovery rate (Fdr) in the two-stage study at a certain level.

Details

Package: RFdr
 Type: Package
 Version: 1.0
 Date: 2016-06-24
 License: GPL-3

In genome-wide association studies (GWASs), we normally discover associations between genetic variants and diseases/traits through the primary study, and validate findings through the replication study. We consider the associations showing significance in both studies as true findings. An important question under this two-stage setting is how do we determine the significance levels in both studies? In traditional methods, we determine significance levels of the primary and replication studies separately. We argue that this separate determination strategy will reduce the power in the overall two-stage study. Here, we propose a novel method to determine significance levels jointly. It finds the most powerful significance levels when controlling the false discovery rate (Fdr) in the two-stage study at a certain level.

The principal component of RFdr package is `RFdrControl`.

1. To jointly determine significance levels, you need obtain summary statistics of each genotyped SNPs in both the primary and replication studies. We have put a example summary statistics (`smryStats1` and `smryStats2`) in the 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(param)`.

2. You can use `RFdrControl` to determine significance levels.

```
RFdrControl(I1, I2, z1, z2, initThld=c(0,0), K=2, q=0.05, beta0=length(z1)/5, plot=T,
output=T, dir=output)
```

Details about the function can be seen using [help\(RFdrControl\)](#).

Author(s)

Wei Jiang and Weichuan Yu

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

References

Jiang, W, and Yu, W. Jointly determining significance levels for primary and replication studies by controlling the false discovery rate in two-stage genome-wide association studies. *Submitted*.

See Also

[RFdrControl](#), [snpEMM](#)

Examples

```
alpha<-5e-3
q<-5e-4
```

```
data(smryStats1) #Summary Stats in the primary study
data(smryStats2) #Summary Stats in the replication study
```

```
z1<-smryStats1$Z #Z-values
z2<-smryStats2$Z
```

```

m<-length(z1)      #SNP number

Thld1<-qnorm(1-alpha/2) #The critical value in the primary study
I1<-(1:m)
I2<-(1:m)[abs(z1)>=Thld1] #The index of Genotyped SNPs in the replication study

#The most powerful significance levels when Fdr<=q in the two-stage study
result<-RFdrControl(I1, I2, z1, z2[I2], initThld=c(Thld1,0),q=q, dir=".")
alpha1<-result$pvalThld1 #Determined significance level in the primary study
alpha2<-result$pvalThld2 #Determined significance level in the replication study

```

RFdrControl

*Determining significance levels in two-stage GWASs***Description**

RFdrControl is a novel method to determine significance levels in two-stage GWASs. The method finds the most powerful significance levels when controlling the false discovery rate (Fdr) in the two-stage study at a certain level. snpEMM is a built-in function to infer the parameters in the Gaussian Mixture Model. It is a variant of EM-algorithm.

Usage

```

RFdrControl(I1, I2, z1, z2, initThld = c(0, 0), K = 2, q = 0.05, \
beta0 = length(z1)/5, plot = T, output = T, dir = "output")

snpEMM(I1, I2, z1,z2, K=2, maxIter=1000, tol=1e-4, beta0=length(z1)/5, info=TRUE)

```

Arguments

I1	The SNP index in the primary study
I2	The SNP index in the replication study
z1	z value vector in the primary study
z2	z value vector in the replication study
initThld	A two-value vector. The determined critical values satisfy $z_{\alpha 1/2} \geq \text{initThld}[1]$ and $z_{\alpha 2/2} \geq \text{initThld}[2]$
K	The components number of the z values in associated SNPs.
q	The Fdr controlling level.
beta0	The penalty term for π_0 . We add Dirichlet(beta0, 0) prior to proportions (π_0 , π_1).
plot	A logical number indicating whether to plot the identified 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.

Details

These functions are the implementation of the joint method to determine significance levels by controlling the Fdr in two-stage GWASs. We assume the z-value vector follow the (K+1)-component Gaussian Mixture Model:

$$Z \sim \pi_0 N(0, I) + \sum_{k=1}^K \pi_k N(0, I + \Sigma_k).$$

snpEMM is a built-in function to infer the parameters in the Gaussian Mixture Model using a variant of EM-algorithm.

RFdrControl is the main function to determine significance levels. snpEMM is called in RFdrControl.

Value

RFdrControl returns the following LIST:

rejected	The indexes of the rejected null hypotheses.
BayesPower	The Bayesian power of the two-stage study with determined significant levels.
Pi1	The vector including the inferred proportion of each associated component.
Sigma	A K-objects list. Each object is the inferred covariance matrix of standardized effect size vector in the corresponding associated component.
zvalThld1	The determined critical value in the primary study.
zvalThld2	The determined critical value in the replication study.
pvalThld1	The determined significance level in the primary study
pvalThld2	The determined significance level in the replication study

snpEMM returns the inferred parameters and the iteration status. The returned value is a LIST:

pi0	The inferred proportion of true null hypotheses.
Pi1	The vector including the inferred proportion of each associated component.
Sigma	A K-objects list. Each object is the inferred covariance matrix of standardized 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 ($1 \leq k \leq K$) for each SNP.
iter	The iteration number.
Qval	The expected negative log-likelihood.

Author(s)

Wei Jiang and Weichuan Yu

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

References

Jiang, W., and Yu, W. Jointly determining significance levels for primary and replication studies by controlling the false discovery rate in two-stage genome-wide association studies. *Submitted*

See Also

[RFdr](#)

Examples

```
alpha<-5e-3
q<-5e-4

data(smryStats1) #Summary Stats in the primary study
data(smryStats2) #Summary Stats in the replication study

z1<-smryStats1$Z #Z-values
z2<-smryStats2$Z
m<-length(z1) #SNP number

Thld1<-qnorm(1-alpha/2) #The critical value in the primary study
I1<-(1:m)
I2<-(1:m)[abs(z1)>=Thld1] #The index of Genotyped SNPs in the replication study

#The most powerful significance levels when Fdr<=q in the two-stage study
result<-RFdrControl(I1, I2, z1, z2[I2], initThld=c(Thld1,0),q=q, dir=".")
alpha1<-result$pvalThld1 #Determined significance level in the primary study
alpha2<-result$pvalThld2 #Determined significance level in the replication study
```

Index

*Topic **package**

RFdr-package, 1

help(RFdrControl), 2

param (RFdr-package), 1

RFdr, 4

RFdr (RFdr-package), 1

RFdr-package, 1

RFdrControl, 2, 3

smryStats (RFdr-package), 1

snpEMM, 2

snpEMM (RFdrControl), 3