Package 'ReAD'

July 21, 2025

Type Package

Title Powerful Replicability Analysis of Genome-Wide Association Studies

Version 1.0.1

Description A robust and powerful approach is developed for replicability analysis of two Genome-wide association studies (GWASs) accounting for the linkage disequilibrium (LD) among genetic variants. The LD structure in two GWASs is captured by a four-state hidden Markov model (HMM). The unknowns involved in the HMM are estimated by an efficient expectation-maximization (EM) algorithm in combination with a non-parametric estimation of functions. By incorporating information from adjacent locations via the HMM, this approach identifies the entire clusters of genotype-phenotype associated signals, improving the power of replicability analysis while effectively controlling the false discovery rate.

License GPL-3

Encoding UTF-8

Depends Rcpp (>= 1.0.10), qvalue

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.2.3

NeedsCompilation yes

Author Yan Li [aut, cre, cph],
Haochen lei [aut],
Xiaoquan Wen [aut],
Hongyuan Cao [aut]

Maintainer Yan Li `<yanli_@jlu.edu.cn>`

Repository CRAN

Date/Publication 2023-06-30 11:00:10 UTC

Contents

`em_hmm` *EM algorithm in combination with a non-parametric algorithm for estimation of the rLIS statistic.*

Description

Estimate the rLIS values accounting for the linkage disequilibrium across two genome-wide association studies via the four-state hidden Markov model. Apply a step-up procedure to control the FDR of replicability null.

Usage

```
em_hmm(pa_in, pb_in, pi0a_in, pi0b_in)
```

Arguments

`pa_in`	A numeric vector of p-values from study 1.
`pb_in`	A numeric vector of p-values from study 2.
`pi0a_in`	An initial estimate of the null probability in study 1.
`pi0b_in`	An initial estimate of the null probability in study 2.

Value

`rLIS`	The estimated rLIS for replicability null.
`fdr`	The adjusted values based on rLIS for FDR control.
`loglik`	The log-likelihood value with converged estimates of the unknowns.
`pi`	An estimate of the stationary probabilities of four states (0,0), (0,1), (1,0), (1,1).
`A`	An estimate of the 4-by-4 transition matrix.
`f1`	A non-parametric estimate for the non-null probability density function in study 1.
`f2`	A non-parametric estimate for the non-null probability density function in study 2.

`ReAD` *Replicability analysis across two genome-wide association studies accounting for the linkage disequilibrium structure.*

Description

Replicability analysis across two genome-wide association studies accounting for the linkage disequilibrium structure.

Usage

```
ReAD(pa, pb)
```

Arguments

`pa`	A numeric vector of p-values from study 1.
`pb`	A numeric vector of p-values from study 2.

Value

A list:

`rLIS`	The estimated rLIS for replicability null.
`fdr`	The adjusted values based on rLIS for FDR control.
`loglik`	The log-likelihood value with converged estimates of the unknowns.
`pi`	An estimate of the stationary probabilities of four states (0,0), (0,1), (1,0), (1,1).
`A`	An estimate of the 4-by-4 transition matrix.
`f1`	A non-parametric estimate for the non-null probability density function in study 1.
`f2`	A non-parametric estimate for the non-null probability density function in study 2.

Examples

```
# Simulate p-values in two studies locally dependent via a four-state hidden Markov model
data <- SimuData(J = 10000)
p1 = data$pa; p2 = data$pb; theta1 = data$theta1; theta2 = data$theta2
# Run ReAD to identify replicable signals
res.read = ReAD(p1, p2)
sig.idx = which(res.read$fdr <= 0.05)
```

`SimuData` *Simulate two sequences of p-values by accounting for the local dependence structure via a hidden Markov model.*

Description

Simulate two sequences of p-values by accounting for the local dependence structure via a hidden Markov model.

Usage

```
SimuData(
  J = 10000,
  pi = c(0.25, 0.25, 0.25, 0.25),
  A = 0.6 * diag(4) + 0.1,
  muA = 2,
  muB = 2,
  sdA = 1,
  sdB = 1
)
```

Arguments

`J`	The number of features to be tested in two studies.
`pi`	The stationary probabilities of four hidden joint states.
`A`	The 4-by-4 transition matrix.
`muA`	Mean of the normal distribution generating the p-value in study 1.
`muB`	Mean of the normal distribution generating the p-value in study 2.
`sdA`	The standard deviation of the normal distribution generating the p-value in study 1.
`sdB`	The standard deviation of the normal distribution generating the p-value in study 2.

Value

A list:

`pa`	A numeric vector of p-values from study 1.
`pb`	A numeric vector of p-values from study 2.
`theta1`	The true states of features in study 1.
`theta2`	The true states of features in study 2.

Index