

Package ‘EMLRT’

August 7, 2014

Type Package

Title Association Studies with Imputed SNPs Using Expectation-Maximization-Likelihood-Ratio Test

LazyData yes

Version 1.0

Date 2014-08-01

Author Kuan-Chieh Huang

Maintainer Kuan-Chieh Huang <kchuang@live.unc.edu>

Description This package contains the EM-LRT-PROB and EM-LRT-DOSE for testing whether an outcome variable is associated with imputed SNPs. Specifically, EM-LRT-PROB is used when posterior probabilities of all potential genotypes are estimated while EM-LRT-DOSE is used when only one-dimensional summary statistic imputed dosages are available. Moreover, this package allows user to simulate imputed genotype data, such as dosage and posterior probabilities, with preferred Rsq and MAF. This package is still under development and subject to change.

License GPL-2

Depends MASS, gtools

R topics documented:

Assoc_EM_LRT_DOSE	2
data.gen	2
EM_LRT_DOSE	3
EM_LRT_PROB	6
Mixture	7

Index	9
--------------	----------

Assoc_EM_LRT_DOSE *Association Test with EM-LRT-DOSE method.*

Description

Conduct association testing using imputed data (output from MaCH) using EM-LRT-DOSE method.

Usage

```
Assoc_EM_LRT_DOSE(datfile, pedfile, dosefile, infofile, match = TRUE, outfile)
```

Arguments

datfile	Phenotypic Data - datfile
pedfile	Phenotypic Data - pedfile
dosefile	Imputed Allele Counts - dosefile
infofile	Imputed Allele Counts - infofile
match	Match the individual ID among pedfile and dosefile. Default is TRUE.
outfile	Output a file.
filename	Specific filename for output file.
hideinfo	Hide program information or not.
nSNPs	The number of SNPs for association testing from the 1st SNPs.
ncore	The number of cores to use, i.e. at most how many child processes will be run simultaneously.

Author(s)

Kuan-Chieh Huang

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--or do help(data=index) for the standard data sets.
```

data.gen *Generate Data*

Description

Generate data of phenotype (Y), genotype (G), posterior probabilities (P), and imputed dosage (D) with user specified R_{sq} and MAF.

Usage

```
data.gen(n, r2, q, beta, sde, gamma = NULL)
```

Arguments

n	Sample size of the data generated.
r2	Squared Pearson's correlation between true genotype (G) and dosage (D).
q	Minor allele frequency.
beta	Coefficients of intercept and genotype.
gamma	Coefficients of covariates.
sde	Standard deviation of error term.

Author(s)

Kuan-Chieh Huang

Examples

```
##-----##
## 1 Generate data ##
##-----##
n<-2000
r2<-0.3
q<-0.4
beta<-c(1,1.5) ##under the alternative
gamma<-c(1,2,3) ##three covariates
sde<-1
data<-data.gen(n,r2,q,beta,sde,gamma)

##-----##
## 2 Confirm data generated ##
##-----##
lr<-lm(Y~G+Z,data)
lr$coef ##beta; gamma
summary(lr)$sigma ##sde
cor(data$G,data$D)^2 ##r2
mean(data$G)/2 ##q
```

EM_LRT_DOSE

*EM-LRT-Dose***Description**

An expectation-maximization (EM) likelihood-ratio test (LRT) for association based on imputed dosages.

Usage

```
EM_LRT_DOSE(param0, r2, q, RejSampling, seed, MI, Intermediate, hideinfo, data)
```

Arguments

param0	Initial values for parameters (beta0, beta1, gamma1, ..., sde).
r2	Squared Pearson's correlation between true genotype (G) and dosage (D).
q	Minor allele frequency.
RejSampling	Default rejection sampling approach will be used to sample probability of having one copy of minor allele. If false, dosage sampling will be used instead.
seed	Seed for rejection sampling. If not specified, seed will be randomly selected from 1 to 2000.
MI	Number of multiple imputation. Default is 1. i.e., single imputation.
Intermediate	Interested intermediate chi-square test statistic(s) and p-value(s) will be shown in the output. For example, if Intermediate == c(1,3), the chi-square test statistic and p-value calculated based on the first and the first three data sets will be shown in the result.
hideinfo	Rejection sampling information. If FALSE, number of rejection sampling and the corresponding sampling time will be shown.
data	Data contains phenotype (Y), dosage (D), and covariates (Z). See data.gen() for more details.

Value

chisq	Chi-square test statistic for association test based on multiple data sets.
chisq_intermediate_steps	Chi-square test statistic(s) for association test based on first interested intermediate data sets.
pval	P-value for association test w.r.t. chisq.
pval_intermediate_steps	P-value(s) for association test w.r.t. chisq_intermediate_steps.

Author(s)

Kuan-Chieh Huang

See Also

[data.gen](#)

Examples

```
##-----##
## 1 Generate data ##
##-----##
n<-2000
r2<-0.5
q<-0.4
beta<-c(1, 1.5)
gamma<-c(1, 2, 3)
sde<-1
data<-data.gen(n, r2, q, beta, sde, gamma)
```

```

##-----##
## 2 Mask true genotype ##
##-----##
data_scenario2<-list(Y=data$Y,D=data$D,Z=data$Z)

##-----##
## 3 Generate initial values ##
##-----##
lr2<-lm(Y~D+Z,data_scenario2)
beta_gamma0<-lr2$coef
sde0<-summary(lr2)$sigma
param0<-c(beta_gamma0,sde0)

##-----##
## 4 Apply method ##
##-----##
##Scenario II

##### Dosage Sampling #####
##Example 1
dose1<-EM_LRT_DOSE(param0,r2,q,RejSampling=FALSE,seed=NULL,MI=1,Intermediate=NULL,
hideinfo=FALSE,data=data_scenario2)

##Dosage sampling approach is used to sample 1 set of probability
##of having one copy of minor allele.

##### Rejection Sampling #####
##Example 1
dose1<-EM_LRT_DOSE(param0,r2,q,RejSampling=TRUE,seed=123,MI=1,Intermediate=NULL,
hideinfo=FALSE,data=data_scenario2)

##Rejection sampling approach is used to sample 1 set of probability
##of having one copy of minor allele.

##Example 2
dose2<-EM_LRT_DOSE(param0,r2,q,RejSampling=TRUE,seed=123,MI=5,Intermediate=NULL,
hideinfo=FALSE,data=data_scenario2)

##Rejection sampling approach is used to sample 5 sets of probability
##of having one copy of minor allele.

##Example 3
dose3<-EM_LRT_DOSE(param0,r2,q,RejSampling=TRUE,seed=123,MI=5,Intermediate=c(1,3),
hideinfo=FALSE,data=data_scenario2)

##Rejection sampling approach is used to sample 5 sets of probability of
##having one copy of minor allele. In addition to the chi-square test
##statistic and p-value calculated based on the 5 data sets, chi-square test
##statistics and p-values calculated based on the first data set and the first
##3 data sets will be shown in the result.

```

EM_LRT_PROB

*EM-LRT-Prob***Description**

An expectation-maximization (EM) likelihood-ratio test (LRT) for association based on posterior probabilities.

Usage

```
EM_LRT_PROB(param0, r2, q, data)
```

Arguments

param0	Initial values for parameters (beta0, beta1, gamma1, ..., sde).
r2	Squared Pearson's correlation between true genotype (G) and dosage (D).
q	Minor allele frequency.
data	Data contains phenotype (Y), dosage (D), and covariates (Z). See data.gen() for more details.

Value

chisq	Chi-squared test statistic for association test
pval	P-value for association test

Author(s)

Kuan-Chieh Huang

See Also

[Mixture](#), [data.gen](#)

Examples

```
##-----##
## 1 Generate data ##
##-----##
n<-2000
r2<-0.3
q<-0.4
beta<-c(1,1.5)
gamma<-c(1,2,3)
sde<-1
data<-data.gen(n,r2,q,beta,sde,gamma)

##-----##
## 2 Mask true genotype ##
##-----##
data_scenario1<-list(Y=data$Y,D=data$D,P=data$P,Z=data$Z)
```

```
##-----##
## 4 Generate initial values ##
##-----##
lr2<-lm(Y~D+Z,data_scenariol)
beta_gamma0<-lr2$coef
sde0<-summary(lr2)$sigma
param0<-c(beta_gamma0,sde0)

##-----##
## 5 Apply method ##
##-----##
##Scenario I
prob<-EM_LRT_PROB(param0,r2,q,data_scenariol) ##EM-LRT-PROB
```

Mixture

Mixture (Zhang et al.)

Description

Mixture of regression model that described in Zheng et al. (2011).

Usage

```
Mixture(param0, r2, q, data)
```

Arguments

param0	Initial values for parameters (beta0, beta1, gamma1, ..., sde).
r2	Squared Pearson's correlation between true genotype (G) and dosage (D).
q	Minor allele frequency.
data	Data contains phenotype (Y), probabilities (P), and covariates (Z).

Value

chisq	Chi-squared test statistic for association.
pval	P-value for association.

Author(s)

Kuan-Chieh Huang

References

Zheng, J., et al., A comparison of Approaches to account for uncertainty in analysis of imputed genotypes. *Genetic Epidemiology*, 2011. 35(2): p. 102-10.

See Also

[EM_LRT_PROB](#), [data.gen](#)

Examples

```

##-----##
## 1 Generate data ##
##-----##
n<-2000
r2<-0.5
q<-0.4
beta<-c(1,1.5)
gamma<-c(1,2,3)
sde<-1
data<-data.gen(n,r2,q,beta,sde,gamma)

##-----##
## 2 Mask true genotype ##
##-----##
data_scenariol<-list(Y=data$Y,D=data$D,P=data$P,Z=data$Z)

##-----##
## 4 Generate initial values ##
##-----##
lr2<-lm(Y~D+Z,data_scenariol)
beta_gamma0<-lr2$coef
sde0<-summary(lr2)$sigma
param0<-c(beta_gamma0,sde0)

##-----##
## 5 Apply method ##
##-----##
##Scenario I
mixture<-Mixture(param0,r2,q,data_scenariol) ##Mixture (Zhang et al.)

```


Index

*Topic \textasciitildekw1

Assoc_EM_LRT_DOSE, 2

data.gen, 2

EM_LRT_DOSE, 3

EM_LRT_PROB, 6

Mixture, 7

*Topic \textasciitildekw2

Assoc_EM_LRT_DOSE, 2

data.gen, 2

EM_LRT_DOSE, 3

EM_LRT_PROB, 6

Mixture, 7

Assoc_EM_LRT_DOSE, 2

data.gen, 2, 4, 6, 7

EM_LRT_DOSE, 3

EM_LRT_PROB, 6, 7

Mixture, 6, 7