

Package ‘rareGE’

January 27, 2015

Type Package

Title Testing Gene-Environment Interaction for Rare Genetic Variants

Version 0.1

Date 2014-07-28

Author Han Chen

Maintainer Han Chen <hanchen@hsph.harvard.edu>

Description Tests gene-environment interaction for rare genetic variants using Sequence Kernel Association Test (SKAT) type gene-based tests. Includes two tests for the interaction term only, and one joint test for genetic main effects and gene-environment interaction.

License GPL-3

Imports MASS, nlme, survey

URL <http://www.hsph.harvard.edu/han-chen/>

NeedsCompilation no

Repository CRAN

Date/Publication 2014-07-30 06:57:34

R topics documented:

rareGE-package	2
INT_FIX	2
INT_RAN	4
JOINT	5
rareGE	6
rareGEgeno	9
rareGEpheno	9
wuweights	10

Index	11
--------------	-----------

rareGE-package

Testing Gene-Environment Interaction for Rare Genetic Variants

Description

Tests gene-environment interaction for rare genetic variants using Sequence Kernel Association Test (SKAT) type gene-based tests. Includes two tests for the interaction term only, and one joint test for genetic main effects and gene-environment interaction.

Details

Package: rareGE
Type: Package
Version: 0.1
Date: 2014-07-28
License: GPL-3

Author(s)

Han Chen

Maintainer: Han Chen <hanchen@hsph.harvard.edu>

References

Chen H, Meigs JB, Dupuis J. (2014) Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered* 78, 81-90.

See Also

[INT_FIX](#), [INT_RAN](#), [JOINT](#), [rareGE](#)

INT_FIX

Gene-environment interaction test treating genetic main effects as fixed

Description

Perform the interaction term only test for gene-environment interaction, assuming fixed genetic main effects.

Usage

```
INT_FIX(phenotype, genotypes, covariates, mainweights = wuweights,
interweights = wuweights, family = "gaussian", binomialimpute = FALSE)
```

Arguments

phenotype	a numeric vector of phenotype values.
genotypes	a matrix or a data frame for all SNPs in the test gene or genomic region. The order of rows must match the order in phenotype.
covariates	a matrix, a data frame or a vector of covariates to adjust for. The interaction between SNPs and the first column of covariates is tested. The order of rows must match the order in phenotype.
mainweights	the weight function or vector of genetic main effects (default = wuweights).
interweights	the weight function or vector of gene-environment interaction effects (default = wuweights).
family	"gaussian" for quantitative traits and "binomial" for dichotomous traits (default = "gaussian").
binomialimpute	impute missing genotypes randomly using a binomial distribution with 2 trials and success probability equal to the minor allele frequency. If FALSE, then impute missing genotypes to 0 (default = FALSE).

Value

p-value.

Author(s)

Han Chen

References

Chen H, Meigs JB, Dupuis J. (2014) Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered* 78, 81-90.

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89, 82-93.

See Also

[INT_RAN](#), [rareGE](#)

Examples

```
data(rareGEgeno)
data(rareGEpheno)
# quantitative traits - testing for gene-BMI interactions
INT_FIX(rareGEpheno$y1, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")])
# dichotomous traits - testing for gene-BMI interactions
INT_FIX(rareGEpheno$y2, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")],
family = "binomial")
```

INT_RAN	<i>Gene-environment interaction test treating genetic main effects as random</i>
---------	--

Description

Perform the interaction term only test for gene-environment interaction, assuming random genetic main effects.

Usage

```
INT_RAN(phenotype, genotypes, covariates, mainweights = wuweights,
interweights = wuweights, family = "gaussian", binomialimpute = FALSE)
```

Arguments

phenotype	a numeric vector of phenotype values.
genotypes	a matrix or a data frame for all SNPs in the test gene or genomic region. The order of rows must match the order in phenotype.
covariates	a matrix, a data frame or a vector of covariates to adjust for. The interaction between SNPs and the first column of covariates is tested. The order of rows must match the order in phenotype.
mainweights	the weight function or vector of genetic main effects (default = wuweights).
interweights	the weight function or vector of gene-environment interaction effects (default = wuweights).
family	"gaussian" for quantitative traits and "binomial" for dichotomous traits (default = "gaussian").
binomialimpute	impute missing genotypes randomly using a binomial distribution with 2 trials and success probability equal to the minor allele frequency. If FALSE, then impute missing genotypes to 0 (default = FALSE).

Value

p-value.

Author(s)

Han Chen

References

Chen H, Meigs JB, Dupuis J. (2014) Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered* 78, 81-90.

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89, 82-93.

See Also

[INT_FIX](#), [rareGE](#)

Examples

```
data(rareGEgeno)
data(rareGEpheno)
# quantitative traits - testing for gene-BMI interactions
INT_RAN(rareGEpheno$y1, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")])
# dichotomous traits - testing for gene-BMI interactions
INT_RAN(rareGEpheno$y2, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")],
family = "binomial")
```

JOINT

*Joint test of genetic main effects and gene-environment interactions***Description**

Perform a joint test of genetic main effects and gene-environment interactions, using a Monte Carlo approach to calculate the p-value.

Usage

```
JOINT(phenotype, genotypes, covariates, mainweights = wuweights,
interweights = wuweights, family = "gaussian", binomialimpute = FALSE,
rho = seq(0, 1, by = 0.1), B = 10000)
```

Arguments

phenotype	a numeric vector of phenotype values.
genotypes	a matrix or a data frame for all SNPs in the test gene or genomic region. The order of rows must match the order in phenotype.
covariates	a matrix, a data frame or a vector of covariates to adjust for. The interaction between SNPs and the first column of covariates is tested. The order of rows must match the order in phenotype.
mainweights	the weight function or vector of genetic main effects (default = wuweights).
interweights	the weight function or vector of gene-environment interaction effects (default = wuweights).
family	"gaussian" for quantitative traits and "binomial" for dichotomous traits (default = "gaussian").
binomialimpute	impute missing genotypes randomly using a binomial distribution with 2 trials and success probability equal to the minor allele frequency. If FALSE, then impute missing genotypes to 0 (default = FALSE).
rho	a numeric vector with values between 0 and 1 defining the searching grid for the nuisance parameter ρ in the joint test.
B	number of Monte Carlo simulations to approximate the multi-dimensional integral in calculating the p-value of the joint test (default = 10000).

Value

pJOINT	p-value of the joint test.
pJOINTmin	minimum raw p-value from the searching grid of the joint test.
pJOINTrho	ρ value where the minimum raw p-value is attained in the joint test.
pJOINTps	a vector of raw p-values in the joint test.
pJOINTinfo	a summary of the distribution of the integrand in each Monte Carlo simulation in the joint test. pJOINT is the mean excluding missing values.

Author(s)

Han Chen

References

- Chen H, Meigs JB, Dupuis J. (2014) Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered* 78, 81-90.
- Lee S, Wu MC, Lin X. (2012) Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13, 762-775.
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89, 82-93.

See Also

[rareGE](#)

Examples

```
set.seed(12345)
data(rareGEgeno)
data(rareGEpheno)
# quantitative traits - testing for gene-BMI interactions
JOINT(rareGEpheno$y1, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")],
      B = 1000)
# dichotomous traits - testing for gene-BMI interactions
JOINT(rareGEpheno$y2, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")],
      family = "binomial", B = 1000)
```

rareGE	<i>Tests for gene-environment interaction and joint test of genetic main effects and gene-environment interaction</i>
--------	---

Description

Perform either gene-environment interaction tests or joint test of genetic main effects and gene-environment interaction, or all 3 tests in one function.

Usage

```
rareGE(phenotype, genotypes, covariates, mainweights = wuweights,
interweights = wuweights, family = "gaussian", binomialimpute = FALSE,
rho = seq(0, 1, by = 0.1), B = 10000, INT_FIX = TRUE, INT_RAN = TRUE,
JOINT = TRUE)
```

Arguments

phenotype	a numeric vector of phenotype values.
genotypes	a matrix or a data frame for all SNPs in the test gene or genomic region. The order of rows must match the order in phenotype.
covariates	a matrix, a data frame or a vector of covariates to adjust for. The interaction between SNPs and the first column of covariates is tested. The order of rows must match the order in phenotype.
mainweights	the weight function or vector of genetic main effects (default = wuweights).
interweights	the weight function or vector of gene-environment interaction effects (default = wuweights).
family	"gaussian" for quantitative traits and "binomial" for dichotomous traits (default = "gaussian").
binomialimpute	impute missing genotypes randomly using a binomial distribution with 2 trials and success probability equal to the minor allele frequency. If FALSE, then impute missing genotypes to 0 (default = FALSE).
rho	a numeric vector with values between 0 and 1 defining the searching grid for the nuisance parameter ρ in the joint test.
B	number of Monte Carlo simulations to approximate the multi-dimensional integral in calculating the p-value of the joint test (default = 10000).
INT_FIX	a logical indicator of whether the interaction test treating genetic main effects as fixed should be performed (default = TRUE).
INT_RAN	a logical indicator of whether the interaction test treating genetic main effects as random should be performed (default = TRUE).
JOINT	a logical indicator of whether the joint test of genetic main effects and gene-environment interaction effects should be performed (default = TRUE).

Details

We use interaction tests and the joint test for different hypotheses. For interaction tests, genetic main effects are included in the null model, either as fixed effects (INT_FIX) or random effects (INT_RAN). We do not recommend treating genetic main effects as fixed when the number of SNPs in the gene is not small. For the joint test (JOINT), genetic main effects are not included in the null model and we are testing genetic effects, allowing for effect modification by the environmental variable. In the joint test, raw p-values are calculated for each value of the nuisance parameter ρ , then the minimum p-value is taken (similar to the optimal test proposed by Lee et al., 2012) and the actual p-value is calculated by multi-dimensional integration, approximated by a Monte Carlo method. This function is designed to perform two or more tests. If performing only one test, please use INT_FIX, INT_RAN or JOINT functions instead to get faster speed (only relevant models are fitted and relevant statistics are computed).

Value

pINT_FIX	p-value from the interaction test treating genetic main effects as fixed. NULL if INT_FIX is FALSE.
pINT_RAN	p-value from the interaction test treating genetic main effects as random. NULL if INT_RAN is FALSE.
pJOINT	p-value from the joint test of genetic main effects and gene-environment interaction effects. NULL if JOINT is FALSE.
pJOINTmin	minimum raw p-value from the searching grid of the joint test. NULL if JOINT is FALSE.
pJOINTrho	ρ value where the minimum raw p-value is attained in the joint test. NULL if JOINT is FALSE.
pJOINTps	a vector of raw p-values in the joint test. NULL if JOINT is FALSE.
pJOINTinfo	a summary of the distribution of the integrand in each Monte Carlo simulation in the joint test. pJOINT is the mean excluding missing values. NULL if JOINT is FALSE.

Author(s)

Han Chen

References

- Chen H, Meigs JB, Dupuis J. (2014) Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered* 78, 81-90.
- Lee S, Wu MC, Lin X. (2012) Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13, 762-775.
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89, 82-93.

See Also

[INT_FIX](#), [INT_RAN](#), [JOINT](#)

Examples

```
set.seed(12345)
data(rareGEgeno)
data(rareGEpheno)
# quantitative traits - testing for gene-BMI interactions
rareGE(rareGEpheno$y1, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")],
B = 1000)
# dichotomous traits - testing for gene-BMI interactions
rareGE(rareGEpheno$y2, rareGEgeno, rareGEpheno[, c("bmi", "age", "sex")],
family = "binomial", B = 1000)
```

`rareGEgeno`*Example genotype data*

Description

Example genotype data for rareGE.

Usage

```
data(rareGEgeno)
```

Format

A data matrix for 500 individuals and a gene with 20 SNPs. Each row represents an individual and each column represents a SNP.

`rareGEpheno`*Example phenotype data*

Description

Example phenotype data for rareGE.

Usage

```
data(rareGEpheno)
```

Format

A data frame with 500 observations on the following 5 variables.

`y1` a quantitative trait.

`y2` a dichotomous trait.

`age` a quantitative covariate.

`sex` a dichotomous covariate.

`bmi` a quantitative covariate.

wuweights

Wu weight function

Description

Calculate beta density weights with parameters 1 and 25 as proposed by Wu et al. (2011)

Usage

```
wuweights(maf)
```

Arguments

`maf` a numeric vector of minor allele frequencies. Values must be between 0 and 1.

Value

A numeric vector of weights, with the same length as `maf`.

Author(s)

Han Chen

References

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89, 82-93.

Examples

```
maf <- seq(0, 1, by=0.1)
wuweights(maf)
```

Index

*Topic **datasets**

rareGEgeno, 9
rareGEpheno, 9

*Topic **gene-environment interaction**

INT_FIX, 2
INT_RAN, 4
JOINT, 5
rareGE, 6

*Topic **joint test**

JOINT, 5
rareGE, 6

*Topic **package**

rareGE-package, 2

*Topic **rare variant**

INT_FIX, 2
INT_RAN, 4
JOINT, 5
rareGE, 6

INT_FIX, 2, 2, 5, 7, 8

INT_RAN, 2, 3, 4, 7, 8

JOINT, 2, 5, 7, 8

rareGE, 2, 3, 5, 6, 6

rareGE-package, 2

rareGEgeno, 9

rareGEpheno, 9

wuweights, 10