Package 'iGasso'

June 4, 2016

Type Package

Title Statistical Tests and Utilities for Genetic Association

Version 1.4

Date 2016-06-03

Author Dr. Kai Wang

Maintainer Kai Wang <kai-wang@uiowa.edu>

Depends lattice, CompQuadForm

Description A collection of statistical tests for genetic association studies.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2016-06-04 07:52:27

R topics documented:

iGasso-package *Statistical Tests and utilities for Genetic Association*

Description

iGasso is a collection of statistical tests developed by our group for genetic association studies. So far it contains functions for rare variants association, for association with multiple phenotypes, for linear mixed model analysis, and for model-free association analysis. There is also a function for genome plot. It will keep growing as more tests are developed. Use ?iGasso to see an introduction.

Details

Package:	iGasso
Type:	Package
Version:	1.4
Date:	2016-06-3
License:	GPL (>=2)
LazyLoad:	yes

Author(s)

Kai Wang <kai-wang@uiowa.edu>

References

Anscombe F.J. (1948) The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35(3/4)**, 246–254.

Chanter, D. O. (1975). Modifications of the angular transformation. *Journal of the Royal Statistical Society. Series B (Applied Statistics)*, **24 (3)**, 354–359.

Freeman, M. F., Tukey, J. W. (1950) Transformations related to the angular and the square root. *The Annals of Mathematical Statistics* **21(4)**, 607–611.

Wang, K. (2012) An application of the proportional odds model to genetic association studies. Submitted.

Wang K. (2012) Statistical tests of genetic association for case-control study designs. *Biostatistics*. Accepted. PMID: 22389176

Wang, K., Fingert, J. (2012) Statistical tests for detecting rare variants using variance-stabilizing transformations. *Annals of Human Genetics*. Accepted.

Zar, J. H. (1999) *Biostatistical Analysis, 4th ed.*, New Jersey:Prentice-Hall, Inc.

Examples

```
y = rnorm(100)
chr = c(rep(1, 20), rep(3, 20), rep(10, 20), rep(19, 30), rep("X", 10))
pos = c(1:20, 1:20, 1:20, 1:30, 1:10)
mydata = data.frame(y=y, chr=chr, pos=pos)
genome.plot(mydata, sig.line=c(1, -1), ylab="T Statistic")

G = rbind(c(14, 999), c(3, 1081))
VSTF.test(G)

G = rbind(c(161, 474, 489), c(231, 444, 380))
MFree.test(G)

G = matrix(sample(c(0,1,2), 200, replace=TRUE), ncol=10)
y = rnorm(10)
X = matrix(rnorm(10), ncol=1)
```

`genome.plot` *Genome-wide Plot of a Variable*

Description

`genome.plot` plots the value of a variable across the genome.

Usage

```
genome.plot(mydata, style=1, type="h", sig.line=c(4, -4),
            sig.color=c("red", "red"), ...)
```

Arguments

`mydata`	a data frame containing three variables: `y` (numeric, the value of the variable to be plotted), `chr` (character, chromosome label), and `pos` (numeric, position, for instance, in base pair or centi-Morgan). Examples of `y` include -log10 of p-values and test statistic values.
`style`	either 1 (default) or 2.
`type`	a generic graphic parameter. Recommended values are `"h"` (default) and `"b"`.
`sig.line`	vertical locations of significance lines.
`sig.color`	colors of significance lines.
`...`	other parameters to be passed to function `xyplot` in the `lattice` package.

Details

This function makes use of the function `xyplot` from package `lattice`.

Author(s)

Kai Wang <kai-wang@uiowa.edu>

Examples

```
y = rnorm(100)
chr = c(rep(1, 20), rep(3, 20), rep(10, 20), rep(19, 30), rep("X", 10))
pos = c(1:20, 1:20, 1:20, 1:30, 1:10)
mydata = data.frame(y=y, chr=chr, pos=pos)
mydata2 = data.frame(y=y^2, chr=chr, pos=pos)

genome.plot(mydata, sig.line=c(1, -1), ylab="T Statistic")
genome.plot(mydata, sig.line=c(1, -1), ylab="T Statistic", type="b")
genome.plot(mydata2, sig.line=c(2), ylab="y squared")
genome.plot(mydata, style=2, sig.line=c(1, -1), ylab="T Statistic")
genome.plot(mydata, style=2, sig.line=c(1, -1), ylab="T Statistic", type="b")
```

KAT.coin *Conditional Inference for the Kernel Association Test (KAT)*

Description

KAT.coin computes the asymptotic and the approximate conditional p-values for the kernel association test

Usage

```
KAT.coin(y, G, X = NULL, out_type = "D", distribution = "asymptotic", B = 1000)
```

Arguments

y	a vector of phenotype on n subjects.
G	an $n \times m$ matrix of SNP genotypes of n study subjects at m loci.
X	a matrix of covariates. It has n rows.
out_type	an indicator of the outcome type. "C" for the continuous outcome and "D" for the dichotomous outcome.
distribution	a character, the conditional null distribution of the test statistic can be approximated by its asymptotic distribution ("asymptotic", default) or via Monte Carlo resampling ("approximate"), as in package coin.
B	the number of permutations if distribution = "approximate"

Details

The asymptotic conditional null distribution is obtained using results in Strasser and Weber (1999). The p-value based on this distribution is computed using Davies' method.

Value

A list with class "htest" containing the following components:

statistic	the value of the kernel association test statistic.
parameter	sample size and the number of SNPs
p.value	the p-value based on the asymptotic or the approximate conditional null distribution.
method	a character string indicting the test performed.
data.name	a character string giving the name of the data.

Author(s)

Kai Wang <kai-wang@uiowa.edu>

References

Strasser, H. and Weber, C. (1999) On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics*. 8(2):220-250.

Wang, K. (2016) Conditional Inference for the Kernel Association Test. Submitted.

Examples

```
n=1000
y = c(rep(1, n/2), rep(0, n/2))
maf = seq(0.05, 0.5, 0.05)
g = NULL
for (j in 1:10){
        geno.freq = c(maf[j]^2, 2*maf[j]*(1-maf[j]), (1-maf[j])^2)
        g = cbind(g, sample(c(0,1,2), n, replace=TRUE, prob=geno.freq))
}
KAT.coin(y, g, X=NULL, out_type="D", B=1000)
```

MFree.test *Model-free Association Tests*

Description

MFree.test performs tests on association between an SNP and case-control status. It tests whether the frequencies of an allele are the same between cases and controls. It does not require specification of an inheritance model.

Usage

```
MFree.test(G, method="score")
```

Arguments

G	a 2x3 two-dimensional contingency table in matrix form. The first row is for cases and the second one for controls. In each row, the entries are the number of subjects carrying 0, 1, and 2 copies of the reference allele, respective.
method	a character string indicating the test statistic to use. One of "score" (default), "Wald", and "LRT".

Details

Each test is named after the author(s) of the corresponding publication.

Value

A list with class "test" containing the following components:

statistic	the value of the test statistic.
p.value	the p-value for the test computed from a chi-square distribution with 1 df.
method	a character string indicting the test performed.
data.name	a character string giving the name of the data.

Author(s)

Kai Wang <kai-wang@uiowa.edu>

References

Wang K. (2012) Statistical tests of genetic association for case-control study designs. *Biostatistics*. 13(4):724-33. PMID: 22389176

Examples

```
G = rbind(c(161, 474, 489), c(231, 444, 380))
MFree.test(G)
MFree.test(G, method = "Wald")
MFree.test(G, method = "LRT")
```

SKATplus *A Gene- or Pathway-Based Test of Association*

Description

SKATplus provides enhanced power over SKAT by properly estimating the null distribution of SKAT.

Usage

```
SKATplus(y, G, X=NULL, out_type="D", tau=NULL, permutation=FALSE, B=1000)
```

Arguments

y	a vector of phenotype on n subjects.
G	an $n \times m$ matrix of SNP genotypes of n study subjects at m loci.
X	a matrix of covariates. It has n rows.
out_type	an indicator of the outcome type. "C" for the continuous outcome and "D" for the dichotomous outcome.
tau	proportion of selected subjects used for SKATplus.
permutation	an indicator. Use permutation for p-value or not.
B	the number of permutations if permutation=TRUE

Details

This version uses only subjects with lower phenotypic values for estimating the null distribution. That is, the "controls" are those of lower phenotypic values. When "controls" are of higher phenotypic values, change the sign of the phenotypic values in order to use this function.

Value

A list with class "htest" containing the following components:

statistic	the value of the test statistic, which is the same as SKAT statistic.
parameter	sample size and the number of SNPs
p.value	the p-value for SKATplus computed using Davies' method.
method	a character string indicting the test performed.
data.name	a character string giving the name of the data.

Author(s)

Kai Wang <kai-wang@uiowa.edu>

References

Wang, K. (2016) Boosting the power of the sequence kernel association test (SKAT) almost surely by properly estimating its null distribution. *A J Hum Genet*. In press.

Examples

```
n=1000
y = c(rep(1, n/2), rep(0, n/2))
maf = seq(0.05, 0.5, 0.05)
g = NULL
for (j in 1:10){
        geno.freq = c(maf[j]^2, 2*maf[j]*(1-maf[j]), (1-maf[j])^2)
        g = cbind(g, sample(c(0,1,2), n, replace=TRUE, prob=geno.freq))
}
SKATplus(y, g, X=NULL, out_type="D", tau=NULL, permutation=FALSE, B=1000)
```

VSTF.test *Association Tests for Rare Variants Based on Variance-Stabilizing Transformation*

Description

VSTF.test performs tests on association between a rare variant and case-control status using a variance-stabilizing transformation.

Usage

```
VSTF.test(G, method = "Anscombe")
```

Arguments

G	a 2x2 matrix. The first row is for cases and the second one for controls. In each row, the first element is the number of non-carriers and the second one is the number of carriers with at least 1 copy of the variant.
method	a character string indicating which transformation to use. One of "Anscombe" (default), "arcsine", "Freeman-Tukey", and "Chanter".

Details

Each test is named after the author(s) of the corresponding publication.

Value

A list with class "test" containing the following components:

statistic	the value of the test statistic.
p.value	the p-value for the test computed from a chi-square distribution with 1 df.
method	a character string indicting the test performed.
data.name	a character string giving the name of the data.

Author(s)

Kai Wang <kai-wang@uiowa.edu>

References

Anscombe, F. J. (1948) The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35(3/4)**, 246–254.

Chanter, D. O. (1975). Modifications of the angular transformation. *Journal of the Royal Statistical Society. Series B (Applied Statistics)* **24(3)**, 354–359.

Freeman, M. F., Tukey, J. W. (1950) Transformations related to the angular and the square root. *The Annals of Mathematical Statistics* **21(4)**, 607–611.

Wang, K., Fingert, J. (2012) Statistical tests for detecting rare variants using variance-stabilizing transformations. *Annals of Human Genetics*. 76(5):402-9.

Zar, J.H. (1999) *Biostatistical Analysis, 4th ed.*, New Jersey:Prentice-Hall, Inc.

Examples

```
## Example 1 of Li et al. (2010)
G = rbind(c(14, 999), c(3, 1081))
VSTF.test(G)
VSTF.test(G, method = "arcsine")
VSTF.test(G, method = "Freeman-Tukey")
```

Index