Package 'RNOmni'

May 16, 2018

Title Omnibus Test for Genetic Association Analysis using the Rank Normal Transformation

Version 0.4.0

Date 2018-05-16

Description Implementation of genetic association tests using the rank based inverse normal transformation (INT). The primary contribution is an omnibus test, which synthesizes two complementary INT-based approaches. In simulations against phenotypes with skewed and heavy tailed residuals, the omnibus test provided valid inference in the absence of genetic effects. When genetic effects were present, the omnibus test provided power comparable to the more efficient of the component methods. Under these settings, standard linear regression variously failed to control the type I error in the absence of genetic effects, and was underpowered in the presence of genetic effects.

Depends R (>= 3.2.2)

Imports abind, foreach, mvtnorm, plyr, Rcpp

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Suggests ggplot2, knitr, reshape2, rmarkdown

VignetteBuilder knitr

LinkingTo Rcpp, RcppEigen

NeedsCompilation yes

Author Zachary McCaw [aut, cre]

Maintainer Zachary McCaw <zmccaw@g.harvard.edu>

Repository CRAN

Date/Publication 2018-05-16 21:38:13 UTC

R topics documented:

`AvgCorr` *Average Correlation Estimate.*

Description

Estimates the correlation between correlated p-values on the Z-score scale. The p-values are supposed to have arisen from different tests of the same hypothesis using the same data. Since an estimate of correlation under the null is of interest, pairs where at least one of the Z scores exceeds the threshold τ are excluded.

Usage

```
AvgCorr(p1, p2, tau = 3, a = 0.001)
```

Arguments

`p1`	Fist p-value.
`p2`	Second p-value.
`tau`	Threshold Z score above which the p-value likely corresponds to a true positive.
`a`	Force correlation estimate to fall in the interval (a,1-a);

Value

A numeric correlation.

BAT *Basic Association Test*

Description

Test of association between genotype and the untransformed phenotype, adjusting for covariates and population structure.

Usage

```
BAT(y, G, X, S, parallel = F, check = T)
```

Arguments

y	Numeric phenotype vector.
G	Obs by snp genotype matrix.
X	Model matrix of covariates.
S	Model matrix of structure adjustments.
parallel	Logical indicating whether to run in parallel. Must register parallel backend first.
check	Logical indicating whether to check input formatting.

Value

A numeric matrix of score statistics and p-values, one for each locus (column) in G, assessing the null hypothesis that genotype is unrelated to the phenotype.

Examples

```
# BAT against normal phenotype
p = RNOmni::BAT(y=RNOmni::Y[,1],G=RNOmni::G[,1:10],X=RNOmni::X,S=RNOmni::S);
```

`BootCorr` *Bootstrap Correlation Estimate.*

Description

Estimates the correlation between correlated p-values on the Z-score scale. Avoids the assumption that the correlation between p-values is constant across loci. Instead, bootstrap is used to calculate locus-specific estimates of the correlation between p-values.

Usage

```
BootCorr(y, G, X, S, k = 3/8, B = 100, parallel)
```

Arguments

`y`	Numeric phenotype vector.
`G`	Obs by snp genotype matrix.
`X`	Model matrix of covariates.
`S`	Model matrix of structure adjustments.
`k`	Offset applied during rank-normalization. See `rankNormal`.
`B`	Bootstrap samples for correlation estimation.
`parallel`	Run bootstraps in parallel? Must register parallel backend first.

Value

Numeric matrix of correlation estimates, one per locus (column) in G.

`DINT` *Direct-INT*

Description

Tests of association between genotype and the rank normalized phenotype, adjusting for covariates and population structure.

Usage

```
DINT(y, G, X, S, k = 3/8, parallel = F, check = T)
```

Arguments

y	Numeric phenotype vector.
G	Obs by snp genotype matrix.
X	Model matrix of covariates.
S	Model matrix of structure adjustments.
k	Offset applied during rank-normalization. See rankNormal.
parallel	Logical indicating whether to run in parallel. Must register parallel backend first.
check	Logical indicating whether to check input formatting.

Value

A numeric matrix of score statistics and p-values, one for each locus (column) in G, assessing the null hypothesis that genotype is unrelated to the phenotype.

Examples

```
# Direct INT on the normal phenotype
p = RNOmni::DINT(y=RNOmni::Y[,1],G=RNOmni::G[,1:10],X=RNOmni::X,S=RNOmni::S);
```

---

fastDet *Matrix Determinant*

Description

Calculates $\det(A)$.

Usage

```
fastDet(A)
```

Arguments

A	Numeric matrix.

fastInv *Matrix Inverse*

Description

Matrix Inverse

Usage

```
fastInv(A)
```

Arguments

A	Numeric matrix.

fastIP *Matrix Inner Product*

Description

Calculates $A'B$.

Usage

```
fastIP(A, B)
```

Arguments

A	Numeric matrix.
B	Numeric matrix.

fastMMp *Matrix matrix product*

Description

Calculates AB;

Usage

```
fastMMp(A, B)
```

Arguments

A	Numeric matrix.
B	Numeric matrix.

fastQF *Matrix Quadratic Form*

Description

Calculates $x'Ax$.

Usage

```
fastQF(X, A)
```

Arguments

X	Numeric matrix.
A	Numeric matrix.

fastT *Matrix Transpose*

Description

Constructs A' from A.

Usage

```
fastT(A)
```

Arguments

A	Numeric matrix.

fitNorm *Normal Model*

Description

Fits the standard OLS model.

Usage

```
fitNorm(y, Z)
```

Arguments

y	Outcome.
Z	Model matrix.

G *Simulated Genotypes*

Description

Simulated genotypes for 1000 subjects (rows) at 1000 loci (columns). Genotypes were simulated with use of hapgen2, and are based on the haplotype structure of human chromosome one in the CEU population of the 1000 Genomes Project.

Usage

```
G
```

Format

A double matrix with 1000 rows and 1000 columns.

s1-s1000 Each column corresponds to genotype a given locus for 1000 subjects.

IINT *Indirect-INT*

Description

Two-stage regression procedure. In the first stage, phenotype is regressed on covariates and structure adjustments to obtain residuals. Genotype is also regressed on covariates and structure adjustments to obtain residuals. In the second stage, INT-transformed phenotypic residuals are regressed on genotypic residuals.

Usage

```
IINT(y, G, X, S, k = 3/8, parallel = F, check = T)
```

Arguments

y	Numeric phenotype vector.
G	Obs by snp genotype matrix.
X	Model matrix of covariates.
S	Model matrix of structure adjustments.
k	Offset applied during rank-normalization. See rankNormal.
parallel	Logical indicating whether to run in parallel. Must register parallel backend first.
check	Logical indicating whether to check the input.

Value

A numeric matrix of Wald statistics and p-values, one for each locus (column) in G, assessing the null hypothesis that genotype is unrelated to the phenotype.

Examples

```
# IINT against normal phenotype
p = RNOmni::IINT(y=RNOmni::Y[,1],G=RNOmni::G[,1:10],X=RNOmni::X,S=RNOmni::S);
```

---

inCheck *Input Check*

Description

Function to ensure the dimensions of inputs to association methods agree.

Usage

```
inCheck(y, G, X, S)
```

Arguments

y	Numeric phenotype vector.
G	Obs by snp genotype matrix.
X	Obs by feature model matrix of covariates.
S	Obs by feature model matrix of structure adjustments.

---

olsB *Ordinary Least Squares Coeefficient*

Description

Calculate the OLS coefficient $(A'A)^{-1}A'Y$.

Usage

```
olsB(A, Y)
```

Arguments

A	Numeric matrix
Y	Numeric matrix

OmniP *Omnibus P-value*

Description

Calculates the p-value for the maximum of two correlated, standard normal random variables.

Usage

```
OmniP(u, r)
```

Arguments

u	Test statistic.
r	Correlation.

Value

Numeric p-value.

rankNormal *Rank-Normalize*

Description

Applies the rank based inverse normal transform (INT) to a numeric vector. INT is best-suited to continuous outcomes. See the vignette for the definition of INT.

Usage

```
rankNormal(u, k = 3/8)
```

Arguments

u	Numeric vector.
k	Offset. Defaults to (3/8), correspond to the Blom transform.

Value

Numeric vector of rank normalized measurements.

Examples

```
# Draw from chi-1 distribution
y = rchisq(n=1000,df=1);
# Rank normalize
z = RNOmni::rankNormal(y);
# Plot density of transformed measurement
plot(density(z));
```

`Resid` *Residual*

Description

Calculates the residual after projection of Y onto X

Usage

```
Resid(X, Y)
```

Arguments

`X`	Numeric matrix.
`Y`	Numeric matrix.

`RNOmni` *Rank-Normal Omnibus Test*

Description

Association test that synthesizes the `DINT` and `IINT` approaches. The first approach directly transforms the phenotype, whereas the second approach forms residuals prior to applying the rank normal transformation (`rankNormal`). In the omnibus test, the direct and indirect tests are separately applied. An omnibus statistic is calculated based on whichever approach provides more evidence against the null hypothesis of no genotypic effect. Details of the method are discussed in the vignette.

Usage

```
RNOmni(y, G, X, S, method = "AvgCorr", k = 3/8, B = 100, set.rho,
  keep.rho = F, keep.stats = F, parallel = F)
```

Arguments

y	Numeric phenotype vector.
G	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.
method	Method used to estimate correlation for the omnibus test, either "AvgCorr", "Bootstrap", or "Manual".
k	Offset applied during rank-normalization. See rankNormal.
B	If using method=="Bootstrap", number of bootstrap samples for correlation estimation.
set.rho	If using method=="Manual", the fixed value of rho, either a single value or a vector of length==nrow(G);
keep.rho	Logical indicating whether to return the correlation parameter estimated during omnibus calculation. Defaults to FALSE.
keep.stats	Logical indicating whether to return the interim test statistics calculated by DINT and IINT. Defaults to FALSE.
parallel	Logical indicating whether to run in parallel. Must register parallel backend first.

Details

Assigning a p-value to the omnibus statistic requires an estimate of the correlation between the test statistics estimated by DINT and IINT. When many loci are under consideration, a computationally efficient approach is to take the correlation of the observed test statistics across loci (method="AvgCorr"). Alternatively, when there are fewer loci, or locus specific estimates are desired, the correlation may be estimated using bootstrap (method="Bootstrap"). When using the bootstrap approach, consider registering a parallel backend and setting parallel=T. To manually provide an estimate of the correlation between the test statistics, set (method="Manual") and specify (set.rho).

Value

A numeric matrix of p values, three for each locus in G, assessing the null hypothesis that genotype is unrelated to the outcome. If keep.stats=T, the interim test statistics are retained. If keep.rho=T, the estimated correlation between the p values provided by DINT and IINT is retained.

Examples

```
y = RNOmni::Y[,1];
Gsub = RNOmni::G[,1:10];
X = RNOmni::X;
S = RNOmni::S;
# Omnibus test against normal phenotype using the average correlation method
p = RNOmni::RNOmni(y=y,G=Gsub,X=X,S=S,method="AvgCorr");
# Omnibus test against normal phenotype using the bootstrap correlation method
p = RNOmni::RNOmni(y=y,G=Gsub,X=X,S=S,method="Bootstrap",B=10);
```

S *Population Structure Adjustments*

Description

First two principal components of the genetic relatedness matrix. Note that principal components were calculated using genotype at more loci than are provided in the example data set.

Usage

```
S
```

Format

A numeric matrix with 1000 rows and 2 columns.

pc1 The first principal component.

pc2 The second principal component.

SchurC *Schur complement*

Description

Calculates the efficient information $I_{11} - I_{12}I_{22}^{-1}I_{21}$;

Usage

```
SchurC(I11, I22, I12)
```

Arguments

I11	Information of target parameter
I22	Information of nuisance parameter
I12	Cross information between target and nuisance parameters

vecCor *Correlation*

Description

Calculates the correlation between two vectors.

Usage

```
vecCor(a, b)
```

Arguments

a	First vector.
b	Second vector.

X *Simulated Covariates*

Description

Age and sex simulated for 1000 subjects. Age was drawn from a gamma distribution with mean 50 and variance 10. Sex was drawn from a Bernoulli distribuiton with expectation 1/2.

Usage

```
X
```

Format

A numeric matrix with 1000 rows and 2 columns.

Age Age.

Sex Sex.

Y *Simulated Phenotypes*

Description

Two independent phenotypes simulated under the null hypothesis of no genotypic effect. A subject specific mean was calculated based on age, sex, pc1, and pc2. The normal phenotype was generated by adding N(0,1) deviates to the subject means. The T3 phenotype was generated by adding $t_3/sqrt3$ deviates to the subject means.

Usage

```
Y
```

Format

A numeric matrix with 1000 rows and 2 columns.

YN Normal phenotype.

YT3 T3 phenotype.

Index