Package 'RNOmni'

February 13, 2018

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

Version 0.1.5

Date 2018-02-12

Description Implementation of genetic association tests that utilize the rank based inverse normal transformation (INT). The primary contribution is an omnibus test, which synthesizes two INT-based approaches, termed direct INT (DINT) and partially indirect INT (PINT). In genome scale simulations, the omnibus test routinely provided valid inference against continuous phenotypes with skewed and heavy tailed residual distributions. Association testing against these same phenotypes via standard linear regression led to an excess of false positive associations. In addition to controlling the type I error, the omnibus approach provided power comparable to the more powerful of the component methods.

Depends R (>= 3.2.2)

Imports abind, foreach, mvtnorm, RcppEigen

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

Suggests cowplot, ggplot2, knitr, reshape2, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Zachary McCaw [aut, cre]

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

Repository CRAN

Date/Publication 2018-02-12 23:30:29 UTC

R topics documented:

`AvgCorr` *Average Correlation Estimate.*

Description

Estimate correlation using the average of qnorm(p1)*qnorm(p2) across loci, where (p1,p2) are p-values obtained via two different association tests.

Usage

```
AvgCorr(P1, P2, eps = 0.001)
```

Arguments

`P1`	Fist p-value.
`P2`	Second p-value.
`eps`	Force correlation estimate to fall in the interval (eps,1-eps);

`BAT` *Basic Association Test*

Description

Regression of the untransformed phenotype on genotype, covariates, and adjustments for population structure.

Usage

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

Arguments

y	Numeric phenotype vector.
G	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.

Value

A numeric vector of p-values assessing the null hypothesis of no genotypic effect. P-values are estimated using the Wald statistic, and correspond to the rows of G.

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

Use bootstrap to estimate correlation among Z statistics whose maximum is taken in the omnibus test.

Usage

```
BootCorr(y, G, X, S, B = 100, parallel)
```

Arguments

y	Numeric phenotype vector.
G	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.
B	Bootstrap samples for correlation estimation.
parallel	Run bootstraps in parallel? Must register parallel backend first.

DINT *Direct-INT*

Description

Rank-normalizes the phenotype, then regresses the transformed phenotype on genotype, covariates, and adjustments for population structure.

Usage

```
DINT(y, G, X, S, k = 3/8)
```

Arguments

y	Numeric phenotype vector.
G	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.
k	Offset applied during rank-normalization. See rankNormal.

Value

A numeric vector of p-values assessing the null hypothesis of no genotypic effect. P-values are estimated using the Wald statistic, and correspond to the rows of G.

Examples

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

FIINT *Fully Indirect-INT*

Description

Two-stage regression procedure. In the first stage, phenotype is regressed on covariates and adjustments for population structure to obtain residuals. In the second stage, INT-transformed residuals are regressed on genotype.

Usage

```
FIINT(y, G, X, S, k = 3/8)
```

Arguments

y	Numeric phenotype vector.
G	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.
k	Offset applied during rank-normalization. See rankNormal.

Details

Note that, in simulations, FIINT did not consistently provide valid inference. For a similar approach that did control the type I error, see PIINT.

Value

A numeric vector of p-values assessing the null hypothesis of no genotypic effect. P-values are estimated using the Wald statistic, and correspond to the rows of G.

Examples

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

G *Simulated Genotypes*

Description

Simulated genotypes for 1000 subjects at 1000 loci. 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

An integer matrix with 1000 rows and 1000 columns

s1-s1000 s[i] is an integer vector of minor allele counts for the ith subject.

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	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.

Omni *Omnibus P-value*

Description

Calculate p-value for omnibus statistic.

Usage

```
Omni(Q)
```

Arguments

Q	Numeric vector formatted as (p1,p2,rho), where p1 and p2 are estimated p-values, and rho is their correlation.

PIINT *Partially Indirect-INT*

Description

Two-stage regression procedure. In the first stage, phenotype is regressed on covariates to obtain residuals. In the second stage, INT-transformed residuals are regressed on genotype and adjustments for population structure.

Usage

```
PIINT(y, G, X, S, k = 3/8)
```

Arguments

y	Numeric phenotype vector.
G	Snp by obs genotype matrix.
X	Obs by feature covariate matrix.
S	Obs by feature structure matrix.
k	Offset applied during rank-normalization. See rankNormal.

Value

A numeric vector of p-values assessing the null hypothesis of no genotypic effect. P-values are estimated using the Wald statistic, and correspond to the rows of G.

Examples

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

rankNormal *Rank-Normalize*

Description

Applies the rank based inverse normal transform (INT) to a numeric vector. INT is indicated for 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));
```

RNOmni *Rank-Normal Omnibus Test*

Description

Omnibus association test that synthesizes the DINT and PIINT approaches. In the omnibus test, both DINT and PIINT are 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 below and in the vignette.

Usage

```
RNOmni(y, G, X, S, method = "AvgCorr", k = 3/8, B = 100, set.rho,
  keep.rho = 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.
parallel	Run bootstraps in parallel? Must register parallel backend first.

Details

Assignment of a p-value to the omnibus statistic requires an estimate of the correlation between the test statistics estimated by DINT and PIINT. When the sample size and number of loci are both large, and efficient estimate of the correlation is obtained by averaging across loci (method="AvgCorr"). When either the sample size or the number of loci is small, bootstrap (method="Bootstrap") allows for locus specific correlation estimates. If 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 with three columsn and one row per locus, i.e. row, in the genotype matrix, and three columns. The columns are p-values obtained by DINT, PIINT, and the omnibus test.

Examples

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

S *Population Structure Adjustments*

Description

First two principal components of the centered and scaled subject by locus genotype 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.

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 linear predictor was calculated based on age, sex, pc1, and pc2. The normal phenotype was generated by adding N(0,1) deviates to the linear predictor, while the T3 phenotype was generated by adding T3 deviates to the linear predictor.

Usage

```
Y
```

Format

A numeric matrix with 1000 rows and 2 columns

YN Normal phenotype.

YT3 T3 phenotype.

Index