

Package ‘RNOmni’

January 12, 2019

Title Rank Normal Transformation Omnibus Test

Version 0.6.0

Date 2019-01-12

Description Implementation of genetic association tests that incorporate the rank-based inverse normal transformation (INT). These tests are broadly indicated for traits with continuous residual distributions. In the presence of non-normal residuals, INT-based tests robustly control the type I error, whereas standard linear regression may not. Moreover, INT-based tests dominate standard linear regression in terms of power. There are two main strategies for incorporating the INT in association analysis. In direct INT (D-INT), the trait is directly transformed. In indirect INT (I-INT), residuals are formed prior to transformation. Neither D-INT nor I-INT is uniformly most powerful. The INT omnibus test (O-INT) adaptively combines D-INT and I-INT into a single robust and statistically powerful approach.

Depends R (>= 3.2.2)

Imports abind, foreach, plyr, Rcpp

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Suggests cowplot, 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 2019-01-12 22:21:21 UTC

R topics documented:

AvgCorr	2
BAT	3
BootCorr	4
cov	4
DINT	5
fitOLS	6
IINT	6
matInv	7
matLP	8
OINT	8
OmniP	10
rankNorm	10
RNOmni-help	11
SchurC	12
Index	13

AvgCorr	<i>Average Correlation Estimate.</i>
---------	--------------------------------------

Description

Estimates the correlation p-values that 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	First 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

Scalar.

Description

Conducts tests of association between the loci in G and the untransformed phenotype y, adjusting for the model matrix X.

Usage

```
BAT(y, G, X = NULL, test = "Score", simple = FALSE,
    parallel = FALSE)
```

Arguments

y	Numeric phenotype vector.
G	Obs by snp genotype matrix.
X	Model matrix of covariates and structure adjustments. Should include an intercept. Omit to perform marginal tests of association.
test	Either Score or Wald.
simple	Return only the p-values?
parallel	Logical indicating whether to run in parallel. Must register parallel backend first.

Value

If simple=T, returns a vector of p-values, one for each column of G. If simple=F, returns a numeric matrix, including the Wald or Score statistic, its standard error, the Z-score, and the p-value.

See Also

Direct INT [DINT](#), indirect INT [IINT](#), omnibus INT [OINT](#).

Examples

```
## Not run:
set.seed(100);
# Design matrix
X = cbind(1,rnorm(1e3));
# Genotypes
G = replicate(1e3,rbinom(n=1e3,size=2,prob=0.25));
storage.mode(G) = "numeric";
# Phenotype
y = as.numeric(X%*%c(1,1))+rnorm(1e3);
# Association test
p = BAT(y=y,G=G,X=X,simple=T);

## End(Not run)
```

BootCorr	<i>Bootstrap Correlation Estimate.</i>
----------	--

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 a locus-specific estimate.

Usage

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

Arguments

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

Value

Numeric matrix of correlation estimates, one per column of G.

cov	<i>Correlation</i>
-----	--------------------

Description

Calculates the correlation between two vectors.

Usage

```
cov(A, B, cor = FALSE)
```

Arguments

A	First matrix.
B	Second matrix.
cor	Return correlation matrix?

Value

Numeric matrix.

DINT

*Direct-INT***Description**

Applies the rank-based inverse normal transformation ([rankNorm](#)) to the phenotype y . Conducts tests of association between the loci in G and transformed phenotype, adjusting for the model matrix X .

Usage

```
DINT(y, G, X = NULL, k = 3/8, test = "Score", simple = FALSE,
     parallel = FALSE)
```

Arguments

<code>y</code>	Numeric phenotype vector.
<code>G</code>	Obs by snp genotype matrix.
<code>X</code>	Model matrix of covariates and structure adjustments. Should include an intercept. Omit to perform marginal tests of association.
<code>k</code>	Offset applied during rank-normalization. See rankNorm .
<code>test</code>	Either Score or Wald.
<code>simple</code>	Return only the p-values?
<code>parallel</code>	Logical indicating whether to run in parallel. Must register parallel backend first.

Value

If `simple=T`, returns a vector of p-values, one for each column of G . If `simple=F`, returns a numeric matrix, including the Wald or Score statistic, its standard error, the Z-score, and the p-value.

See Also

Basic association test [BAT](#), indirect INT [IINT](#), omnibus INT [OINT](#).

Examples

```
## Not run:
set.seed(100);
# Design matrix
X = cbind(1,rnorm(1e3));
# Genotypes
G = replicate(1e3,rbinom(n=1e3,size=2,prob=0.25));
storage.mode(G) = "numeric";
# Phenotype
y = exp(as.numeric(X%*%c(1,1))+rnorm(1e3));
# Association test
```

```
p = DINT(y=y,G=G,X=X,simple=T);
## End(Not run)
```

fitOLS	<i>Univariate OLS model.</i>
--------	------------------------------

Description

Fits the standard OLS model.

Usage

```
fitOLS(y, X)
```

Arguments

y	Numeric vector.
X	Numeric matrix.

Value

List containing the following:

Beta	Regression coefficient.
V	Outcome variance.
Ibb	Information matrix for beta.
Resid	Outcome residuals.

IINT	<i>Indirect-INT</i>
------	---------------------

Description

Two-stage association testing procedure. In the first stage, phenotype y and genotype G are each regressed on the model matrix X to obtain residuals. The phenotypic residuals are transformed using [rankNorm](#). In the next stage, the INT-transformed residuals are regressed on the genotypic residuals.

Usage

```
IINT(y, G, X = NULL, k = 3/8, simple = FALSE, parallel = FALSE)
```

Arguments

y	Numeric phenotype vector.
G	Obs by snp genotype matrix.
X	Model matrix of covariates and structure adjustments. Should include an intercept. Omit to perform marginal tests of association.
k	Offset applied during rank-normalization. See rankNorm .
simple	Return only the p-values?
parallel	Logical indicating whether to run in parallel. Must register parallel backend first.

Value

If simple=T, returns a vector of p-values, one for each column of G. If simple=F, returns a numeric matrix, including the Wald or Score statistic, its standard error, the Z-score, and the p-value.

See Also

Basic association test [BAT](#), direct INT [DINT](#), omnibus INT [OINT](#).

Examples

```
## Not run:
set.seed(100);
# Design matrix
X = cbind(1,rnorm(1e3));
# Genotypes
G = replicate(1e3,rbinom(n=1e3,size=2,prob=0.25));
storage.mode(G) = "numeric";
# Phenotype
y = exp(as.numeric(X%*%c(1,1))+rnorm(1e3));
# Association test
p = IINT(y=y,G=G,X=X,simple=T);

## End(Not run);
```

matInv

Matrix Inverse

Description

Calculates A^{-1} .

Usage

```
matInv(A)
```

Arguments

A Numeric matrix.

Value

A numeric matrix.

matIP *Matrix Inner Product*

Description

Calculates the inner product $A'B$.

Usage

matIP(A, B)

Arguments

A Numeric matrix.

B Numeric matrix.

Value

Numeric matrix.

OINT *Omnibus-INT*

Description

Association test that synthesizes the [DINT](#) and [IINT](#) tests. The first approach is most powerful for traits that could have arisen from a rank-preserving transformation of a latent normal trait. The second approach is most powerful for traits that are linear in covariates, yet have skewed or kurtotic residual distributions. During the omnibus test, the direct and indirect tests are separately applied. An omnibus statistic is calculated based on whichever approach provides more evidence of an association. Details of the method are discussed in the vignette.

Usage

```
OINT(y, G, X = NULL, method = "AvgCorr", k = 3/8, B = 100,
      set.rho = NULL, keep.rho = FALSE, keep.stats = FALSE,
      parallel = FALSE)
```


Arguments

<code>y</code>	Numeric phenotype vector.
<code>G</code>	Obs by locus genotype matrix.
<code>X</code>	Model matrix of covariates and structure adjustments. Should include an intercept. Omit to perform marginal tests of association.
<code>method</code>	Method used to estimate correlation for the omnibus test, either "AvgCorr", "Bootstrap", or "Manual".
<code>k</code>	Offset applied during rank-normalization. See rankNorm .
<code>B</code>	If using <code>method=="Bootstrap"</code> , number of bootstrap samples for correlation estimation.
<code>set.rho</code>	If using <code>method=="Manual"</code> , the fixed value of rho, either a single value or a vector with one element per column in G.
<code>keep.rho</code>	Logical indicating whether to return the correlation parameter estimated during omnibus calculation. Defaults to FALSE.
<code>keep.stats</code>	Logical indicating whether to return the interim test statistics calculated by DINT and IINT. Defaults to FALSE.
<code>parallel</code>	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 under the null. 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 when 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 column of G. If `keep.stats=T`, the D-INT and I-INT Z statistics are returned. If `keep.rho=T`, the correlation between the p-values provided is returned.

See Also

Basic association test [BAT](#), direct INT [DINT](#), indirect INT [IINT](#).

Examples

```
## Not run:
set.seed(100);
# Design matrix
X = cbind(1,rnorm(1e3));
# Genotypes
```

```

G = replicate(1e3,rbinom(n=1e3,size=2,prob=0.25));
storage.mode(G) = "numeric";
# Phenotype
y = exp(as.numeric(X%%c(1,1))+rnorm(1e3));
# Average correlation
p = OINT(y=y,G=G,X=X,method="AvgCorr");
# Bootstrap correlation
p = OINT(y=y,G=G[,1:10],X=X,method="Bootstrap",B=100);
# Manual correlation
p = OINT(y=y,G=G,X=X,method="Manual",set.rho=0.5);

## End(Not run)

```

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

Scalar.

rankNorm

Rank-Normalize

Description

Applies the rank based inverse normal transform (INT) to a numeric vector. The INT can be broken down into a two-step procedure. In the first, the observations are transformed onto the probability scale using the empirical cumulative distribution function (ECDF). In the second, the observations are transformed onto the real line, as Z-scores, using the probit function.

Usage

rankNorm(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.

See Also

Direct INT [DINT](#), indirect INT [IINT](#), omnibus INT [OINT](#).

Examples

```
## Not run:  
# Draw from chi-1 distribution  
y = rchisq(n=1e3,df=1);  
# Rank normalize  
z = rankNorm(y);  
# Plot density of transformed measurement  
plot(density(z));  
  
## End(Not run)
```

RNOmni-help

RNOmni: Rank-Normal Omnibus Association Testing

Description

Implementation of genetic association tests that incorporate the rank-based inverse normal transformation (INT) [rankNorm](#). The direct-INT [DINT](#) test directly transforms the outcome, whereas the indirect-INT [IINT](#) test forms residuals prior to transformation. The omnibus INT [OINT](#) test adaptively combines the D-INT and I-INT tests into a single robust and statistically powerful procedure.

Author(s)

Zachary R. McCaw

SchurC

Schur complement

Description

Calculates the efficient information $I_{bb} - I_{ba}I_{aa}^{-1}I_{ab}$.

Usage

SchurC(Ibb, Iaa, Iba)

Arguments

Ibb	Information of target parameter
Iaa	Information of nuisance parameter
Iba	Cross information between target and nuisance parameters

Value

Numeric matrix.

Index

AvgCorr, [2](#)

BAT, [3](#), [5](#), [7](#), [9](#)

BootCorr, [4](#)

cov, [4](#)

DINT, [3](#), [5](#), [7–9](#), [11](#)

fitOLS, [6](#)

IINT, [3](#), [5](#), [6](#), [8](#), [9](#), [11](#)

matInv, [7](#)

matIP, [8](#)

OINT, [3](#), [5](#), [7](#), [8](#), [11](#)

OmniP, [10](#)

rankNorm, [5–7](#), [9](#), [10](#), [11](#)

RNOmni-help, [11](#)

RNOmni-help-package (RNOmni-help), [11](#)

SchurC, [12](#)