Package ‘RNOmni’

October 5, 2020

**Title**  Rank Normal Transformation Omnibus Test

**Version**  1.0.0

**Date**  2020-10-04

**Description**  Inverse normal transformation (INT) based genetic association testing. These tests are recommend for continuous traits with non-normally distributed residuals. INT-based tests robustly control the type I error in settings where standard linear regression does not, as when the residual distribution exhibits excess skew or kurtosis. Moreover, INT-based tests dominate standard linear regression in terms of power. These tests may be classified into two types. In direct INT (D-INT), the phenotype is itself transformed. In indirect INT (I-INT), phenotypic residuals are transformed. The omnibus test (O-INT) adaptively combines D-INT and I-INT into a single robust and statistically powerful approach. See McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. “Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies” <doi:10.1111/biom.13214>.

**Depends**  R (>= 3.2.2)

**Imports**  plyr, Rcpp, stats

**LinkingTo**  Rcpp, RcppArmadillo

**License**  GPL-3

**Encoding**  UTF-8

**LazyData**  true

**RoxygenNote**  7.1.1

**Suggests**  R.rsp

**VignetteBuilder**  R.rsp

**NeedsCompilation**  yes

**Author**  Zachary McCaw [aut, cre] (<https://orcid.org/0000-0002-2006-9828>)

**Maintainer**  Zachary McCaw <zmcaw@alumni.harvard.edu>

**Repository**  CRAN

**Date/Publication**  2020-10-05 16:40:02 UTC
## Basic Input Checks

### Description

Basic Input Checks

### Usage

```r
BasicInputChecks(y, G, X)
```

### Arguments

- **y**: Numeric phenotype vector.
- **G**: Genotype matrix with observations as rows, SNPs as columns.
- **X**: Covariate matrix.
Basic Association Test

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)

Arguments

y Numeric phenotype vector.
G Genotype matrix with observations as rows, SNPs as columns.
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 the p-values only?

Value

If simple = TRUE, returns a vector of p-values, one for each column of G. If simple = FALSE, 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

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)
**BAT.ScoreTest**  
*Basic Association Score Test*

**Description**

Basic Association Score Test

**Usage**

```r
BAT.ScoreTest(y, G, X)
```

**Arguments**

- `y` Numeric phenotype vector.
- `G` Genotype matrix with observations as rows, SNPs as columns.
- `X` Model matrix of covariates.

**Value**

Numeric matrix, with 1 row per SNP, containing these columns:

- "score", the score statistic.
- "se", its standard error.
- "z", the Z statistic.
- "p", the p-value.

---

**BAT.WaldTest**  
*Basic Association Wald Test*

**Description**

Basic Association Wald Test

**Usage**

```r
BAT.WaldTest(y, G, X)
```

**Arguments**

- `y` Numeric phenotype vector.
- `G` Genotype matrix with observations as rows, SNPs as columns.
- `X` Model matrix of covariates.
**CauchyToP**

**Value**

Numeric matrix, with 1 row per SNP, containing these columns:

- "score", the score statistic.
- "se", its standard error.
- "z", the Z statistic.
- "p", the p-value.

**Description**

Convert Cauchy Random Variable to P

**Usage**

CauchyToP(z)

**Arguments**

z Numeric Cauchy random variable.

**Value**

Numeric p-value.

---

**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 = 0.375, test = "Score", simple = FALSE)
Arguments

- **y**: Numeric phenotype vector.
- **G**: Genotype matrix with observations as rows, SNPs as columns.
- **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](#).
- **test**: Either Score or Wald.
- **simple**: Return the p-values only?

Value

If `simple = TRUE`, returns a vector of p-values, one for each column of `G`. If `simple = FALSE`, 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 test **IINT**.
- Omnibus INT test **OINT**.

Examples

```r
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)
```

---

**fitOLS**

Ordinary Least Squares

Description

Fits the standard OLS model.

Usage

`fitOLS(y, X)`
**Arguments**

- `y`: N x 1 Numeric vector.
- `X`: N x P Numeric matrix.

**Value**

List containing the following:

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

---

**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 = 0.375, simple = FALSE)
```

**Arguments**

- `y`: Numeric phenotype vector.
- `G`: Genotype matrix with observations as rows, SNPs as columns.
- `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 the p-values only?

**Value**

If `simple = TRUE`, returns a vector of p-values, one for each column of `G`. If `simple = FALSE`, 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 test DINT.
- Omnibus INT test OINT.

Examples

```r
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)
```

---

**IINT.ScoreTest**  
**Basic Association Score Test**

**Description**

Basic Association Score Test

**Usage**

```r
IINT.ScoreTest(y, G, X, k)
```

**Arguments**

- `y`  
  Numeric phenotype vector.
- `G`  
  Genotype matrix with observations as rows, SNPs as columns.
- `X`  
  Model matrix of covariates.
- `k`  
  Offset applied during rank-normalization.

**Value**

Numeric matrix, with 1 row per SNP, containing these columns:

- "score", the score statistic.
- "se", its standard error.
- "z", the Z statistic.
- "p", the p-value.
**matInv**  
*Matrix Inverse*

**Description**  
Calculates $A^{-1}$.

**Usage**  
\[ \text{matInv}(A) \]

**Arguments**  
\[ A \]  
Numeric matrix.

**Value**  
Numeric matrix.

---

**matIP**  
*Matrix Inner Product*

**Description**  
Calculates the product $A'B$.

**Usage**  
\[ \text{matIP}(A, B) \]

**Arguments**  
\[ A \]  
Numeric matrix.  
\[ B \]  
Numeric matrix.

**Value**  
Numeric matrix.
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, then the p-values are combined via the Cauchy combination method.

Usage

OINT(y, G, X = NULL, k = 0.375, simple = FALSE)

Arguments

- **y**: Numeric phenotype vector.
- **G**: Genotype matrix with observations as rows, SNPs as columns.
- **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 the OINT p-values only?

Value

A numeric matrix of p-values, three for each column of G.

See Also

- Basic association test BAT.
- Direct INT test DINT.
- Indirect INT test IINT.

Examples

```r
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))
# Omnibus
p <- OINT(y = y, G = G, X = X, simple = TRUE)
```
OmniP

Omnibus P-value.

Description

Obtains an omnibus p-value from a vector of potentially dependent p-values using the method of Cauchy combination. The p-values are converted to Cauchy random deviates then averaged. The distribution of the average of these deviates is well-approximated by a Cauchy distribution in the tails. See <https://doi.org/10.1080/01621459.2018.1554485>.

Usage

OmniP(p)

Arguments

p Numeric vector of p-values.

Value

OINT p-value.

PartitionData

Partition Data

Description

Partition y and X according to the missingness pattern of g.

Usage

PartitionData(e, g, X)

Arguments

e Numeric residual vector.
g Genotype vector.
X Model matrix of covariates.

Value

List containing:

* "g_obs”, observed genotype vector.
* "X_obs", covariates for subjects with observed genotypes.
* "X_mis", covariates for subjects with missing genotypes.
* "e_obs”, residuals for subjects with observed genotypes.
**PtoCauchy**  
*Convert P-value to Cauchy Random*

**Description**
Convert P-value to Cauchy Random

**Usage**
PtoCauchy(p)

**Arguments**
- **p**: Numeric p-value.

**Value**
Numeric Cauchy random variable.

---

**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 = 0.375)

**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 test DINT.
- Indirect INT test IINT.
- Omnibus INT test OINT.
Examples

```r
# Draw from chi-1 distribution
y <- rchisq(n = 1e3, df = 1)
# Rank normalize
z <- RankNorm(y)
# Plot density of transformed measurement
plot(density(z))
```

Description


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.
ScoreStat  

**Description**  
Score Statistics  

**Usage**  
ScoreStat(e, g, X, v)  

**Arguments**  
e  Numeric residual vector.  
g  Genotype vector.  
X  Model matrix of covariates.  
v  Residual variance.  

**Value**  
Numeric vector containing the "score" statistic, standard error "se", "z", and "p" value.

WaldStat  

**Description**  
Basic Association Score Test  

**Usage**  
WaldStat(y, g, X)  

**Arguments**  
y  Numeric phenotype vector.  
g  Genotype vector.  
X  Model matrix of covariates.  

**Value**  
Numeric matrix, with 1 row per SNP, containing these columns:  
- "score", the score statistic.  
- "se", its standard error.  
- "z", the Z statistic.  
- "p", the p-value.
Index

BasicInputChecks, 2
BAT, 3, 6, 8, 10
BAT.ScoreTest, 4
BAT.WaldTest, 4

CauchyToP, 5

DINT, 3, 5, 8, 10, 12, 13

fitOLS, 6

IINT, 3, 6, 7, 10, 12, 13
IINT.ScoreTest, 8

matInv, 9
matIP, 9

OINT, 3, 6, 8, 10, 12, 13
OmniP, 11

PartitionData, 11
PtoCauchy, 12

RankNorm, 5–7, 10, 12, 13
RNomei-help, 13

SchurC, 13
ScoreStat, 14
WaldStat, 14