Package 'QQperm'

October 9, 2016

Type Package

Title Permutation Based QQ Plot and Inflation Factor Estimation

Version 1.0.1

Date 2016-10-08

Author Slave Petrovski and Quanli Wang

Maintainer Quanli Wang <quanli@stat.duke.edu>

Description Provides users the necessary utility functions to generate permutation-based QQ plots and also estimate inflation factor based on the empirical NULL distribution. While it has general utility, it is particularly helpful when the skewness of the Fisher's Exact test in sparse data situations with imbalanced case-control sample sizes renders the reliance on the uniform chi-square expected distribution inappropriate.

License GPL (>= 2)

Depends R (>= 3.1.0)

LazyData TRUE

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-10-09 09:15:16

R topics documented:

---

estlambda2   *Estimate the inflation factor for a distribution of observed P-values or 1-df chi-square test.*

Description

Estimate the inflation factor for a distribution of P-values or 1-df chi-square test using a permutation-based NULL distribution of P-values provided by the package/user. The implementation is similar to the default (estlambda) implementation in R package GenABEL but here it does not assume the NULL distribution of P-values to be uniform. Rather it estimates lambda inflation factor by comparing to the permutation-based expected NULL distribution as described for QQ plots. This is thus more representative of the true NULL distribution of Fisher's Exact p-values for the given case-control configuration in a study.

Usage

```
estlambda2(p.o, p.e, plot = FALSE, filter = TRUE, adjust.xy = FALSE, ...)
```

Arguments

p.o	Observed P-values from the data (true case/control assignments).
p.e	Expected P-values from the NULL distribution, usually obtained through label permutations of the matrix data.
plot	Indicate if a plot should be produced.
filter	Indicate if the filter should be applied. This parameter behaves the same as in estlambda in GenABEL.
adjust.xy	Indicate if the x-axis and y-axis should be adjusted to their own range.
...	Additional arguments passed to the plot function.

Value

Returns a list containing the permutation-based estimated lambda value (estimate) and its standard error (se).

Author(s)

Slave Petrovski and Quanli Wang

Examples

```
#load pre-computed p-values for IGM dataset
library(QQperm)
data("example.Ps")

#print output to pdf file only if not running in interactive mode
if (!interactive()) {
 pdf("lambda.pdf")
```

```
}

#estimate inflation factor and generate plot.
lambda <-estlambda2(example.Ps$observed,example.Ps$perm, plot = TRUE, adjust.xy = TRUE)

if (!interactive()) {
 dev.off()
}
```

---

example.data	*An example collapsing data matrix and the associated case/control status*

Description

An example collapsing data matrix and the associated case/control status

Usage

```
example.data
```

Format

A list containing both the genotype data matrix and the case/control statuses:

data a named data matrix, with a row for each gene and a column for each sample. Each cell in the matrix then represents an indicator function that reflects whether the given sample had none (0) or at least one (1) qualifying variant in the corresponding gene. It also accepts counts of qualifying variants for a given (gene,sample) pair, but will treat these as a (0/1) indicator.

is.case case/control indicator. The order of the indicators here must match that of the order of the samples defined in the data matrix.

---

example.Ps	*Distributions of expected and observed P-values from the igm.data dataset.*

Description

Distributions of expected and observed P-values from the igm.data dataset.

Usage

```
example.Ps
```

Format

A list containing both the observed and the permutation-based expected P-Values.

observed A vector for observed P-values from the true case/control assignments.

perm A vector for expected P-values (permutation-based expected p-value distribution).

---

igm.get.pvalues   *The permutation-based empirical NULL distribution of P-values is generated through label switching and permutation of the true case/control assignment. To achieve this, for a given matrix it randomly permutes the case and control labels of the original configuration and then recomputes the two-tail Fisher's Exact test for all genes. This is repeated n.permutation (e.g., 1000) times. For each of the n.permutations the p-values are ordered and then the mean of each rank-ordered estimate is taken across the n.permutations, i.e., the average 1st order statistic, the average 2nd order statistic, etc. These then represent the empirical estimates of the expected ordered p-values (expected -log10(p-values)). This empirical-based expected p-value distribution no longer depends on an assumption that the p-values are uniformly distributed under the null.*

Description

The permutation-based empirical NULL distribution of P-values is generated through label switching and permutation of the true case/control assignment. To achieve this, for a given matrix it randomly permutes the case and control labels of the original configuration and then recomputes the two-tail Fisher's Exact test for all genes. This is repeated n.permutation (e.g., 1000) times. For each of the n.permutations the p-values are ordered and then the mean of each rank-ordered estimate is taken across the n.permutations, i.e., the average 1st order statistic, the average 2nd order statistic, etc. These then represent the empirical estimates of the expected ordered p-values (expected -log10(p-values)). This empirical-based expected p-value distribution no longer depends on an assumption that the p-values are uniformly distributed under the null.

Usage

```
igm.get.pvalues(matrix, is.case, n.permutations = 1000)
```

Arguments

matrix	The input genotype matrix, with rows for genes and columns for samples.
is.case	The case(1)/control(0) indicator.
n.permutations	The number of label permutations requested. Recommended number is 1,000.

Value

Returns a list containing the observed P-values based on the correct case/control assignment (observed) and the permuted P-Values reflecting the appropriate NULL distribution for the given case/control configuration (perm).

Author(s)

Slave Petrovski and Quanli Wang

Examples

```
#Ps  <- igm.get.pvalues(matrix, is.case)
```

---

igm.read.data	*Read sample file and genotype collapsing matrix in IGM (Insitute for Genomic Medicine, Columbia University) format.*

Description

Read sample file and genotype collapsing matrix in IGM (Insitute for Genomic Medicine, Columbia University) format.

Usage

```
igm.read.data(sample.file, matrix.file, filter.list = NULL)
```

Arguments

sample.file   The input sample file.

matrix.file   The input matrix file.

filter.list   An optional list of genes to exclude from the analysis.

Value

Returns a list containing the data matrix (matrix) and case(1)/control(0) phenotype indicators (is.case).

Author(s)

Slave Petrovski and Quanli Wang

Examples

```
#igm.data <- igm.read.data(sample.file, matrix.file)
```

QQperm   *QQperm: A package for permutation-based QQ plots and inflation factor estimates.*

Description

QQperm: A package for permutation-based QQ plots and inflation factor estimates.

Author(s)

Slave Petrovski and Quanli Wang

---

qqplot   *QQ plot of observed P-values vs expected P-values, using the empirical (permutation-based) expected p-value distribution. This empirical-based expected p-value distribution no longer depends on an assumption that the Fisher's Exact two-tailed p-values are uniformly distributed under the null. For a given matrix, the permutation-based expected distribution is plotted relative to the observed order statistic to get the permutation-based QQ plot.*

Description

QQ plot of observed P-values vs expected P-values, using the empirical (permutation-based) expected p-value distribution. This empirical-based expected p-value distribution no longer depends on an assumption that the Fisher's Exact two-tailed p-values are uniformly distributed under the null. For a given matrix, the permutation-based expected distribution is plotted relative to the observed order statistic to get the permutation-based QQ plot.

Usage

```
qqplot(P.perm, P.observed, adjust.xy = TRUE, ...)
```

Arguments

P.perm	Expected P-values from NULL distribution, which is generated through permutation in our example.
P.observed	Observed P-values from true case/control assignment.
adjust.xy	An option to have the x-axis and y-axis adjusted based on their own range in the plot.
...	Additional arguments passed to the plot function.

Value

None.

Author(s)

Slave Petrovski and Quanli Wang

Examples

```
library(QQperm)
data("example.Ps")

#print output to pdf file only if not running in interactive mode
if (!interactive()) {
 pdf("qqplot.pdf")
}

qqplot(example.Ps$perm, example.Ps$observed)

if (!interactive()) {
 dev.off()
}
```

---

toy.data	*An example collapsing data matrix and the associated case/control status*

Description

An example collapsing data matrix and the associated case/control status

Usage

```
toy.data
```

Format

A list containing both the genotype data matrix and the case/control statuses:

data a named data matrix, with a row for each gene and a column for each sample. Each cell in the matrix then represents an indicator function that reflects whether the given sample had none (0) or at least one (1) qualifying variant in the corresponding gene. It also accepts counts of qualifying variants for a given (gene,sample) pair, but will treat these as a (0/1) indicator.

is.case case/control indicator. The order of the indicators here must match that of the order of the samples defined in the data matrix.

Index