

Package ‘PoissonSeq’

February 19, 2015

Type Package

Title Significance analysis of sequencing data based on a Poisson log linear model

Version 1.1.2

Date 2012-10-9

Depends R(>= 2.10),combinat,splines

Author Jun Li

Maintainer Jun Li <jun.li@nd.edu>

Description This package implements a method for normalization, testing, and false discovery rate estimation for RNA-sequencing data. The description of the method is in Li J, Witten DM, Johnstone I, Tibshirani R (2012). Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* 13(3): 523-38. We estimate the sequencing depths of experiments using a new method based on Poisson goodness-of-fit statistic, calculate a score statistic on the basis of a Poisson log-linear model, and then estimate the false discovery rate using a modified version of permutation plug-in method. A more detailed instruction as well as sample data is available at <http://www.stanford.edu/~junli07/research.html>. In this version, we changed the way of calculating log foldchange for two-class data. The FDR estimation part remains unchanged.

License GPL (>= 2)

LazyLoad yes

Repository CRAN

Date/Publication 2012-10-10 15:36:34

NeedsCompilation no

R topics documented:

dat 2

PS.Est.Depth	2
PS.Main	3
Index	6

dat	<i>A toy RNA-Seq data.</i>
-----	----------------------------

Description

This toy data contains 1000 genes, and 6 samples in each of the two classes. The sequencing depths of the 12 samples are $\text{rep}(c(1, 2), 6)$. The first 80 genes are set to be overexpressed in Class 2, and the next 20 genes are set to be underexpressed in Class 2. The other 900 genes are null. This is a Poisson-distributed data.

Usage

```
data(dat)
```

Format

A list.

- n the count matrix
- y the outcome vector
- type the outcome type
- pair whether the samples are paired or not

PS.Est.Depth	<i>estimate the sequencing depths</i>
--------------	---------------------------------------

Description

Given a data matrix, this function will estimate the sequencing depths on the basis of the Poisson goodness-of-fit statistic. This estimate is applicable to data with any types of outcome, as it estimates under the null hypothesis.

A more detailed instruction as well as sample data is available at <http://www.stanford.edu/~junli07/research.html>.

Usage

```
PS.Est.Depth(n, iter=5, ct.sum=5, ct.mean=0.5)
```

Arguments

n	The data matrix. The rows are counts for a gene, and the columns are counts from an experiment.
iter	Number of iterations used. Default value: 5. The default value is usually a good choice.
ct.sum	if the total number of reads of a gene across all experiments \leq ct.sum, this gene will not be considered for estimating sequencing depth. Default value: 5.
ct.mean	if the mean number of reads of a gene across all experiments \leq ct.mean, this gene will not be considered for estimating sequencing depth. Default value: 0.5.

Details

The value in the data matrix does not need to be integers.

Value

estimated sequencing depth. a vector. their product is 1.

Author(s)

Jun Li.

References

Li J, Witten DM, Johnstone I, Tibshirani R (2011). Normalization, testing, and false discovery rate estimation for RNA-sequencing data. To appear, Biostatistics.

Examples

```
data(dat)
seq.depth <- PS.Est.Depth(dat$n)
```

PS.Main

detecting differentially expressed genes from RNA-Seq data.

Description

This function is the main function of this package. Given the data matrix and the outcome vector, this function returns the estimated permutation-based p-values, the estimated permutation-based false discovery rates, et al.

A more detailed instruction as well as sample data is available at <http://www.stanford.edu/~junli07/research.html>.

Usage

```
PS.Main(dat, para=list())
```

Arguments

dat	<p>The input RNA-Seq data. It <i>must</i> have the following three attributes:</p> <ol style="list-style-type: none"> (1) n: the data matrix. Rows for genes, columns for experiments (samples). (2) y: the outcome vector (3) type: 'twoclass', 'multiclass' or 'quant' <p>The following attributes are <i>optional</i>. If not specified, the default values will be used.</p> <ol style="list-style-type: none"> (4) pair: paired data or not. Default value: FALSE. Only take effect for twoclass data. (5) gname: gene names. Default value: 1 : nrow(n). That is, the i'th gene is named "i".
para	<p>A list of parameters. It can have the following attributes:</p> <ol style="list-style-type: none"> (1) trans: to tranform the data using the order transformation or not to transform it. default value: TRUE (2) npermu: number of permutations. default value: 100 (3) seed: random seed to generate the permutation indexes. default value: 10 (4) ct.sum: if the total number of reads of a gene across all experiments \leq ct.sum, this gene will not be considered for differential expression detection. Default value: 5. (5) ct.mean: if the mean number of reads of a gene across all experiments \leq ct.mean, this gene will not be considered for differential expression detection. Default value: 0.5. (6) div: the number of divisions of genes for estimating theta. default value: 10 (7) pow.file: the file to store the power transform curve ($\text{mean}(\log(\mu)) \sim 1/\theta$). default value: 'pow.txt' <p>All the above attributes are <i>optional</i>.</p>

Value

a data frame (table) containing the following columns. Each row stands for a gene. The genes are sorted from the most significant to the most insignificant.

nc	number of significant genes called. $nc = 1 : (\text{number of genes})$.
gname	the sorted gene names.
tt	The score statistics of the genes.
pval	Permutation-based p-values of the genes.
fdr	Estimated false discovery rate.
log.fc	Estimated log fold change of the genes. Only available for twoclass outcomes.

Author(s)

Jun Li.

References

Li J, Witten DM, Johnstone I, Tibshirani R (2012). Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* 13(3): 523-38.

Examples

```
data(dat)
res <- PS.Main(dat)
```

Index

*Topic **datasets**

dat, [2](#)

*Topic **iteration**

PS.Est.Depth, [2](#)

*Topic **models**

PS.Est.Depth, [2](#)

PS.Main, [3](#)

dat, [2](#)

PS.Est.Depth, [2](#)

PS.Main, [3](#)