Package ‘SSPA’
February 1, 2017

Type Package
Title General Sample Size and Power Analysis for Microarray and Next-Generation Sequencing Data
Version 2.14.0
Author Maarten van Iterson
Maintainer Maarten van Iterson <mviterson@gmail.com>
Description General Sample size and power analysis for microarray and next-generation sequencing data.
License GPL (>= 2)
LazyLoad yes
Imports graphics, stats
Depends R (>= 2.12), methods, qvalue, lattice, limma
Suggests BiocStyle, genefilter, edgeR, DESeq
URL http://www.humgen.nl/MicroarrayAnalysisGroup.html
Collate 'zzz.R' 'numericalintegration.R' 'trimmingbinning.R'
   'DistributionClass.R' 'PilotDataClass.R' 'SampleSizeClass.R'
   'bitriangular.R' 'deconvolution.R' 'conjugategradient.R'
   'Ferreira.R' 'tikhonov.R' 'powerandsamplesize.R'
biocViews GeneExpression, RNASeq, Microarray, StatisticalMethod
NeedsCompilation yes

R topics documented:

dbitri .................................................. 2
deepSAGE ............................................. 2
Nutrigenomics ........................................ 3
pbitri .................................................. 4
pilotData ............................................. 5
plot-methods ........................................ 5
predictpower ......................................... 6
qbitri .................................................. 6
rbitri .................................................. 7
sampleSize ........................................... 8
show-methods ........................................ 10
simdat ............................................... 10


**deepSAGE**

**Index**

<table>
<thead>
<tr>
<th>dbitri</th>
<th>Density function for a bi-triangular random variable.</th>
</tr>
</thead>
</table>

**Description**

Density function for a bi-triangular random variable.

**Usage**

\[ \text{dbitri}(x, a = \log_2(1.2), b = \log_2(4), m = \log_2(2)) \]

**Arguments**

- **x**: vector
- **a**: location of point \( a \). Default \( a = \log_2(1.2) \).
- **b**: location of point \( b \). Default \( b = \log_2(4) \).
- **m**: location of the midpoint of the triangle. Default \( m = \log_2(2) \).

**Details**

For more details see M. Langaas et al. JRSS B 2005.

**Value**

Gives the density function.

**Author(s)**

Maarten van Iterson

**Examples**

\[ \text{curve(dbitri, -4, 4)} \]

---

**deepSAGE**

Test statistics derived from a deepSAGE experiment

**Description**

follow

**Usage**

\[ \text{data(deepSAGE)} \]

**Format**

A vector of 44882 test statistics.

   Vector of test statistics obtained by performing a likelihood ratio test using edgeR
Details

follow

Source


Examples

data(deepSAGE)
str(deepSAGE)

| Nutrigenomics | Test statistics from a Nutrigenomics gene expression profiling experiment |

Description

There are five sets of test statistics each represents a different compound and exposure time. Test statistics were obtained by using an empirical Bayes linear model.

Usage

data(Nutrigenomics)

Format

A data frame with 16539 test statistics for five experiments.

First row indicates the effective sample size of the experiment. Column names refer to the compound and exposure time (see details).

Details

In this experiment the outcome of specific PPAR-alpha activation on murine small intestinal gene expression was examined using Affymetrix GeneChip Mouse 430 2.0 arrays. PPAR-alpha was activated by several PPAR-alpha-agonists that differed in activating potency. In this paper the data of three agonists were used, namely Wy14,643, fenofibrate and trilinolenin (C18:3). The first two compounds belong to the fibrate class of drugs that are widely prescribed to treat dyslipidemia, whereas trilinolenin is an agonist frequently found in the human diet. For intestinal PPAR-alpha, Wy14,643 is the most potent agonist followed by C18:3 and fenofibrate. Since time of exposure also affects the effect size, intestines were collected 6 hrs (all three agonists) or 5 days (Wy14,643 and fenofibrate only) after exposure.

Source

Examples
data(Nutrigenomics)
str(Nutrigenomics)

pbitri

Distribution function for a bi-triangular random variable.

Description

Distribution function for a bi-triangular random variable.

Usage

pbitri(q, a = log2(1.2), b = log2(4), m = log2(2))

Arguments

q vector of quantiles.
a location of point, ... Default a = log2(1.2).
b location of point, ... Default b = log2(4).
m location of the midpoint of the triangle. Default m = log2(2).

Details

For more details see M. Langaas et al. JRSS B 2005.

Value

Gives the distribution function.

Author(s)

Maarten van Iterson

Examples

curve(pbitri, -4, 4)
pilotData

User friendly interface to class "PilotData"

Description
User friendly interface to class "PilotData"

Usage

pilotData(statistics = NULL, samplesize = NULL, distribution = c("norm", "t", "f", "chisq"), ...)

Arguments

- statistics: vector of test statistics
- samplesize: total sample size of the pilot-data or effective sample size in two-group case (see Details for more information).
- distribution: type of the null/alternative distribution, one of 'norm', 't', 'f' or 'chisq'
- ...: additional arguments for the distribution like degrees of freedom

Details
In the two-group case the effective sample size is defined as the square-root of the inverse of 1/n1 + 1/n2.

Value
object of class "PilotData"

Author(s)
Maarten van Iterson

Examples

pd <- pilotData(statistics=rnorm(100), samplesize=10, distribution="norm")
pd
plot(pd)

plot-methods

Methods for Function plot in Package SSA

Description
Plot function for objects of class PilotData and SampleSize

Methods

signature(x = "PilotData") Diagostic plots of the PilotData.
signature(x = "SampleSize") Plot the estimated density of effect sizes.
predictpower

Predict power for given vector of sample sizes

Description

Predict power for given vector of sample sizes

Usage

predictpower(object, samplesizes, alpha = 0.1, verbose = FALSE, plot = FALSE)

Arguments

- object: of class 'SampleSize'
- samplesizes: vector of total sample sizes.
- alpha: FDR.
- verbose: TRUE/FALSE
- plot: TRUE/FALSE

Details

details follow.

Value

predicted power.

Author(s)

Maarten van Iterson

qbitri

Quantile function for a bi-triangular random variable.

Description

Quantile function for a bi-triangular random variable.

Usage

qbitri(p, a = log2(1.2), b = log2(4), m = log2(2))

Arguments

- p: vector of probabilities.
- a: location of point, ... Default a = log2(1.2).
- b: location of point, ... Default b = log2(4).
- m: location of the midpoint of the triangle. Default m = log2(2).
**rbitri**

**Details**
For more details see M. Langaas et al. JRSS B 2005.

**Value**
Gives the quantile function.

**Author(s)**
Maarten van Iterson

**Examples**
curve(qbitri, 0, 1)

---

**rbitri**  
Random generation of bitriangular distributed values.

**Description**
Random generation of bitriangular distributed values.

**Usage**
rbitri(n, a = log2(1.2), b = log2(4), m = log2(2))

**Arguments**
- *n*  
  number of observations.
- *a*  
  location of point, ... Default $a = \log_2(1.2)$.
- *b*  
  location of point, ... Default $b = \log_2(4)$.
- *m*  
  location of the midpoint of the triangle. Default $m = \log_2(2)$.

**Details**
For more details see M. Langaas et al. JRSS B 2005.

**Value**
Generates random deviates.

**Author(s)**
Maarten van Iterson

**Examples**
hist(rbitri(100), freq=FALSE)  
curve(dbbitri, add=TRUE)
sampleSize

User friendly interface to class 'SampleSize'

Description

User friendly interface to class "SampleSize"

Usage

```r
sampleSize(PilotData,
    method = c("deconv", "congrad", "tikhonov", "ferreira"),
    control = list(from = -6, to = 6, resolution = 2^9))
```

Arguments

- **PilotData**: object of class 'PilotData'.
- **method**: estimation method one of 'deconv', 'congrad', 'tikhonov' or 'ferreira'. See 'Details'.
- **control**: A list of control parameters. See 'Details'.

Details

The default method is 'deconv' which is an kernel deconvolution density estimator implemented using `fft`. The 'nncg' is a nonnegative conjugate gradient algorithm based on R’s implementation see `optim`. 'tikonov' implements ridge-regression with optimal penalty selection using the L-curve approach. Higher order penalties are possible as well using a transformation to standard form (see Hansen).

The 'control' argument is a list that can supply any of the following components. Per method logical checks are performed.

- **deconv**:
  - method:'deconv', 'ferreira'
  - pi0Method:the pi0 estimation method one of 'Langaas', 'Storey', 'Ferreira', 'Userdefined'
  - pi0:if method = 'ferreira' grid pi0-value need to be supplied e.g. seq(0.1, 0.99, 0.01)
  - adjust:Default TRUE, adjust pi0 estimate if density of effect size is somewhere negative.
  - a:Adjust pi0 better approach suggested by Efron. Symmetric range around zero of size 0.5.
  - bandwidth:Default NULL uses 1/sqrt(log(length(statistics)))
  - kernel:Either 'fan', 'wand', 'sinc' kernels can be used.
  - from:Density of effect sizes should be estimated from = -6
  - to: to = 6
  - resolution:Density of effect sizes should be estimated on 2^9 points.
  - verbose:Default FALSE if TRUE additional information is printed to the console.

- **congrad**:
  - integration:'midpoint', 'trapezoidal', 'simpson'
  - scale:'pdfstat', 'cdfstat', 'cdfpval'
  - trim:0.01, 0.99
Value

object of class SampleSize.

Author(s)

Maarten van Iterson

References


See Also

optim

Examples

```r
m <- 5000  ## number of genes
J <- 10    ## sample size per group
pi0 <- 0.8 ## proportion of non-differentially expressed genes
m0 <- as.integer(m*pi0)
mu <- rbtri(m - m0, a = log2(1.2), b = log2(4), m = log2(2))  # effect size distribution
data <- simdat(mu, m=m, pi0=pi0, J=J, noise=NULL)
library(genefilter)
stat <- rowttests(data, factor(rep(c(0, 1), each=J)), tstatOnly=TRUE)$statistic
pd <- pilotData(statistics=stat, samplesize=sqrt(J/2), distribution='norm')
ss <- sampleSize(pd, method='deconv')
plot(ss)
```

Description

Methods for function show in package SSA

Methods

signature(object = "PilotData") Show the content of a PilotData-object in a userfriendly way.

signature(object = "SampleSize") Show the content of a SampleSize-object in a userfriendly way.

simdat

Generate simulated microarray data using the bitriangular distribution.

Description

Simulated microarray data.

Usage

```
simdat(mu, m, pi0, J, nullX = function(x) rnorm(x, 0, 1),
       nullY = function(x) rnorm(x, 0, 1), noise = 0.01)
```
Arguments

**mu**  
vector of effect sizes drawn from the bitriangular distribution.

**m**  
number of features (genes, tags, ...).

**pi0**  
proportion of nondifferentially expressed features.

**J**  
number of samples per group.

**nullX**  
the distribution of nondifferentially expressed features.

**nullY**  
the distribution of nondifferentially expressed features.

**noise**  
standard deviation of the additive noise.

Details

details follow

Value

Matrix of size m x (2J), containing the simulated values.

Author(s)

Maarten van Iterson

Examples

```r
## generate two-group microarray data
m <- 5000  ## number of genes
J <- 10    ## sample size per group
pi0 <- 0.8 ## proportion of non-differentially expressed genes
m0 <- as.integer(m*pi0)
mu <- rbitri(m - m0, a = log2(1.2), b = log2(4), m = log2(2)) # effect size distribution
data <- simdat(mu, m=m, pi0=pi0, J=J, noise=0.01)
```
Index

*Topic **datasets**
   deepSAGE, 2
   Nutrigenomics, 3

*Topic **methods**
   plot-methods, 5
   show-methods, 10

dbitri, 2
deepSAGE, 2

fft, 8

Nutrigenomics, 3

optim, 8, 10

pbitri, 4
pilotData, 5
plot, ANY-method (plot-methods), 5
plot, PilotData-method (plot-methods), 5
plot, SampleSize-method (plot-methods), 5
plot-methods, 5
predictpower, 6

qbitri, 6

rbitri, 7

sampleSize, 8
show, ANY-method (show-methods), 10
show, PilotData-method (show-methods), 10
show, SampleSize-method (show-methods), 10
show-methods, 10
simdat, 10