

Package ‘BSDE’

March 26, 2021

Type Package

Title A non-parametric method for Single-Cell Differential Expression for Case-Control Study based on the Barycenter

Version 0.1.0

Author Mengqi Zhang, Richard Guo

Maintainer Mengqi Zhang <meggy.q.zhang@gmail.com>

Description A single-cell RNAseq differential expression analysis approach in case-control study.

License LGPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

R topics documented:

calc_multimod_param	1
calc_zinb_param	2
cal_w2_bulk_pval	3
cal_w2_pval	3
simu_base_param	5
WaBarycenter2	6
Index	7

calc_multimod_param	<i>This function returns the overdispersion parameters(size3) for the single ZINB distribution $ZINB(\mu_3, size_3, \pi)$. This distribution has the same dropout_rate as the mixture of two ZINB with equal 50 The two ZINB have the same dropout_rate, same overdispersion (size) and the same distance of expectation from the μ_3, i.e. $ZINB(\mu_3(1+t), size, \pi_3)$ and $ZINB(\mu_3(1-t), size, \pi_3)$</i>
---------------------	---

Description

This function returns the overdispersion parameters(size3) for the single ZINB distribution $ZINB(\mu_3, size_3, \pi)$. This distribution has the same dropout_rate as the mixture of two ZINB with equal 50 The two ZINB have the same dropout_rate, same overdispersion (size) and the same distance of expectation from the μ_3 , i.e. $ZINB(\mu_3(1+t), size, \pi_3)$ and $ZINB(\mu_3(1-t), size, \pi_3)$

Usage

```
calc_multimod_param(size, t = 0.5)
```

Arguments

size	parameter theta of the zinb model, the distribution will be more close to zero-inflated poisson with larger parameter size.
parameter	t as the numeric $0 \leq t \leq 1$

Value

size3 The parameter theta in the notation

calc_zinb_param	<i>this function returns the parameters for changing mean and variance of the zero-inflated negative binomial(ZINB) distribution</i>
-----------------	--

Description

this function returns the parameters for changing mean and variance of the zero-inflated negative binomial(ZINB) distribution

Usage

```
calc_zinb_param(mu, theta, drop = 0, r_m = 1, r_v = 1)
```

Arguments

mu	the original parameter mu of ZINB distribtuion
theta	the original parameter theta of ZINB distribtuion, the distribution will be more close to zero-inflated poisson with larger parameter size.
drop	the original parameter dropout_rate of of ZINB distribution
r_m	the targeted fold change of mean
v_m	the targeted fold change of variance

Value

two number for parameter mu and theta of modified ZINB distribution

Note

For numerical meaningful calculation, is is required that $r_m/r_v < 1 + \mu/\theta$

cal_w2_bulk_pval	<i>This function calculates the w2_pval based on BSDE for a given gene, based on bulk RNAseq. The expression are represented as one numerical value per gene, per individual.</i>
------------------	---

Description

This function calculates the w2_pval based on BSDE for a given gene, based on bulk RNAseq. The expression are represented as one numerical value per gene, per individual.

Usage

```
cal_w2_bulk_pval(
  cur_ind_count,
  cur_individual,
  phenotype,
  perm_num = 500,
  unif_round_unit = 0.5
)
```

Arguments

count_per_gene:
len x 1 matrix,length=ind num, gene count

meta_individual:
len x 1 vector, length=ind num, indicator labels (contains 0,1)

perm_num:ingeter,
permutation time

Value

a list contains: pval a p-value, from the based on Monte Carlo Permutation Procedure. case_distr a vectors contains the density of case subjects distribution of Barycenter ctrl_distr a vectors contains the density of ctrl subjects distribution of Barycenter

cal_w2_pval	<i>This function calculates the w2_pval based on BSDE for a given gene</i>
-------------	--

Description

This function calculates the w2_pval based on BSDE for a given gene

Usage

```
cal_w2_pval(
  count_per_gene,
  meta_individual,
  meta_phenotype,
  perm_num = 200,
```

```

unif_round_unit = 0.2,
l_fold = 1,
merge_method = "python",
weight = 1,
shrink = FALSE
)

```

Arguments

count_per_gene is a len x 1 matrix, length=cell num. It represents gene count.

meta_individual is a len x 1 vector, length=cell num. It is indicator labels (contains 0,1) for cases (1) and controls(0)

perm_num is a integer for times of permutation.

unif_round_unit is the unit length for frequency calculation.

l_fold is a positive number, represents the fold-change to the parameter lambda for merge_method R, the lambda is the Non-negative regularization parameter (for small lambda the Barycenter is close to the EMD).

merge_method provides the support package for barycenter calculation, the default setting is based on the python package 'ot'(merge_method="python"), if python doesn't work, the R package merge_method="R" is also available

weight is a len x 1 vector, length=subjects num. It represents the importance for subjects.

shrink is a bool value. If True, the range with zero density will be removed, for more robust performance to extreme values and reduce the computational burden. Choose TRUE if there are some extreme high expressions.

Value

a list contains: pval a p-value, from the based on Monte Carlo Permutation Procedure. case_bc_ob a vectors contains the density of case subjects distribution of Barycenter ctrl_bc_ob a vectors contains the density of ctrl subjects distribution of Barycenter

Examples

```

library("Barycenter")
library("Rcpp")
library("reticulate")
library("doRNG")
library("doParallel")
py_run_file("../inst/op_functions.py")

count_per_gene=c(rpois(60,6),c(rpois(60,4)))
meta_individual=paste0("ind",rep(1:12,each=10))
meta_phenotype=c(rep(1,60),rep(0,60))
cal_w2_pval(count_per_gene,meta_individual,meta_phenotype )
cal_w2_pval(count_per_gene,meta_individual,meta_phenotype,unif_round_unit = 1 )

```

simu_base_param	<i>This function works for the basic parameter generation of ZINB models from the given reference data.</i>
-----------------	---

Description

This function works for the basic parameter generation of ZINB models from the given reference data.

Usage

```
simu_base_param(
  t_mean,
  t_disp,
  t_drop,
  t_meta,
  nTotal = 30,
  nall = 40,
  RIN_adj = TRUE
)
```

Arguments

t_mean	a genexcell matrix, represent the mu parameter of ZINB models from reference data. DCA can generate it.
t_disp	a genexcell matrix, represent the overdispersion parameter of ZINB models from reference data. DCA can generate it.
t_drop	a genexcell matrix, represent the dropout_rate parameter of ZINB models from reference data. DCA can generate it.
nTotal	numbers of subjects needs to be simulated

Value

a list includes two objects sample_ctrl and RNA.simu

sample_ctrl is a gene x individual x param array, which gives the parameter "mean", "dispersion", and "dropout" for simulate a ZINB distribution. It also gives the residual standard deviations of means.

References

Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., & Theis, F. J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nature communications, 10(1), 1-14.

WaBarycenter2	<i>This function is modified for 1-dimensional data from WaBarycenter function of R package 'Barycenter', and used for the option "R" of merge_method at function cal_w2_pval A list of matrices satisfying the prerequisites described above.</i>
---------------	--

Description

This function is modified for 1-dimensional data from WaBarycenter function of R package 'Barycenter', and used for the option "R" of merge_method at function cal_w2_pval A list of matrices satisfying the prerequisites described above.

Usage

```
WaBarycenter2(
  images,
  maxIter = 10,
  lambda = FALSE,
  costm = FALSE,
  lambda_fold = 1
)
```

Arguments

images	The input vector, which could be the log-transformed and normalized expression counts.
maxIter	Maximum number of iterations.
lambda	Non-negative regularization parameter (for large lambda the regularized Barycenter is close to its true counterpart). If FALSE the algorithm uses a lambda depending on costm.
costm	A matrix of pairwise distances between the locations. If FALSE the algorithm uses the usual euclidean distance matrix on a [0,1]x[0,1] equidistant pixel grid.
lambda_fold	A modification for the default lambda adjustment.

Value

Value The Barycenter of the input, represented by a vector.

References

Cuturi, M.: Fast Computation of Wasserstein Barycenters, Proceedings of the International Conference on Machine Learning, Beijing, China, 2014

Index

`cal_w2_bulk_pval`, [3](#)
`cal_w2_pval`, [3](#)
`calc_multimod_param`, [1](#)
`calc_zinb_param`, [2](#)

`simu_base_param`, [5](#)

`WaBarycenter2`, [6](#)