Package 'ComplexSim'

January 16, 2024

Title ComplexSim: Simulate high dimensional Omics data with covariance structure with time and subjects

Version 0.0.0.9000

Description This R package allows simulation using both Gaussian and Negative Binomial distribution with different set ups of the parameters, that is, mean and variance of Gaussian, mean and dispersion parameter of Negative Binomial, and one can switch between the two distributions to get equivalent simulation for benchmarking different models. The simulation functions have flexible parametrization that represent the treatment effect, time effect, correlation structure and subject level variability and correlation. One can use the simulation functions to explore different designs and the impact on the data.

License `use_mit_license()`, `use_gpl3_license()` or friends to pick a license

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Depends R (>= 3.5.0),
stats,
methods,
utils,
graphics,
rlang

Imports purrr,
emmeans,
reshape2,
dplyr,
ggplot2,
nlme,
RhpcBLASctl

Suggests ggplot2, knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

CovCor_err   *Title The correlation and covariance*

Description

Title The correlation and covariance

Usage

```
CovCor_err(data, condition, plot.line = FALSE, plot.heat = FALSE)
```

Arguments

data	Simulated data frame
condition	A specific group level
plot.line	Plot the line graph (TRUE or FALSE)
plot.heat	Plot the heatmap (TRUE or FALSE)

CovHeatPlot   *Title Plot the variance-covariance matrix in heatmap*

Description

Title Plot the variance-covariance matrix in heatmap

Usage

```
CovHeatPlot(CMatrix)
```

Arguments

CMatrix	A squared matrix with numeric elements

CovMarginal_sim *Title Generate the mariginal variance matrix*

Description

Title Generate the mariginal variance matrix

Usage

```
CovMarginal_sim(CorrSub, SigmaGroup, CorrRep, SigmaRep, sigmaRE = NULL)
```

Arguments

CorrSub	Subject level correlation matrix, dim nq*nq
SigmaGroup	Group level variance, length q
CorrRep	Correlation matrix between repeated measures, dim m*m
SigmaRep	Variance of repeated measures, length m

Value

Marginal covariance matrix cross samples, dim nqm*nqm

---

fVarAR1 *Title AR(1) correlation matrix*

Description

Title AR(1) correlation matrix

Usage

```
fVarAR1(rho, d)
```

Arguments

rho	Correlation coefficient
d	Dimension of repeated measures

Value

AR(1) correlation matrix

---

fVarCS *Title CS correaltion matrix*

Description

Title CS correaltion matrix

Usage

```
fVarCS(rho, d)
```

Arguments

rho	Correlation coefficient
d	Dimension of repeated measures

Value

CS correlation matrix

---

plot_prof *Title Plot gene profiles*

Description

Title Plot gene profiles

Usage

```
plot_prof(data)
```

Arguments

data	Simulated data frame of gene profiles

wrap_design	*Title wrap_design <- function(n,m,q, SingleGroup = FALSE, eps = FALSE)*

Description

Title wrap_design <- function(n,m,q, SingleGroup = FALSE, eps = FALSE)

Usage

```
wrap_design(n, m, q, SingleGroup = FALSE, eps = FALSE)
```

Arguments

n	Number of patients in each group. (Need >50 individuals to have a reasonable estimated covariance structure.)
m	Number of repeated visits (time points).
q	Number of groups (eg., control and multiple treatment groups).
SingleGroup	Whether the individuals are in single group or multiple groups.

Value

Model matrix with desired design.

wrap_simulation	*Title Simulate the log-scale expression of one gene*

Description

Title Simulate the log-scale expression of one gene

Usage

```
wrap_simulation(
  data,
  SingleGroup,
  beta0 = NULL,
  beta1 = NULL,
  beta2 = NULL,
  lambda = NULL,
  sigma = NULL,
  mu = NULL,
  tau = NULL,
  act.cor = NULL
)
```

Arguments

`data`	Design matrix in long format.
`beta0`	Mean of the global effect.
`beta1`	Mean of group effect.
`beta2`	Mean of time effect.
`lambda`	Coefficient of interaction terms.
`sigma`	Standard deviation of the simulated AR(1) sequence.
`mu`	Mean vector of random intercept and slope.
`tau`	Elements of variance-covariance matrix of random intercept and slope.
`act.cor`	Correlation parameter of the AR(1) structure.

Value

Data matrix in long data frame.

`wrap_sim_corr`	*Title Simulate DGE List*

Description

Title Simulate DGE List

Usage

```
wrap_sim_corr(
  data,
  N,
  eps_Signal,
  SingleGroup,
  coef.beta0 = NULL,
  coef.beta1 = NULL,
  coef.beta2 = NULL,
  coef.int = NULL,
  sigma = NULL,
  mu = NULL,
  tau = NULL,
  CovMarginal = NULL,
  act.cor = NULL
)
```

Arguments

`data`	Design matrix in long format.
`N`	Number of simulated genes in total.
`eps_Signal`	Proportion of signal genes.
`coef.beta0`	Mean of the global effect.
`coef.beta1`	Mean of group effect.

coef.beta2	Mean of time effect.
coef.int	Coefficient of interaction terms.
sigma	Standard deviation of the simulated AR(1) sequence.
mu	Mean vector of random intercept and slope.
tau	Elements of variance-covariance matrix of random intercept and slope.
CovMarginal	Marginal covariance.
act.cor	Correlation coefficient of the time effect.

Value

DGE list

---

wrap_sim_DGEList	*Simulate Differential Gene Expression Data*

Description

Simulates gene expression data for a given number of genes and samples, and returns a list containing the normalized counts matrix, sample phenotype data, and feature (gene) data.

Usage

```
wrap_sim_DGEList(
  n,
  m,
  q,
  N_gene,
  global_eff = NULL,
  trt_eff,
  trt_slope = NULL,
  inter = NULL,
  sigma = NULL,
  mu = NULL,
  tau = NULL,
  act.cor = NULL,
  distribution = NULL,
  SingleGroup = FALSE
)
```

Arguments

n	Integer, the number of biological replicates per group.
m	Integer, the number of time points.
q	Integer, the number of groups.
N_gene	Integer, the number of genes to simulate.
global_eff	Numeric vector or NULL, global effects for each gene.
trt_eff	Numeric vector, treatment effects for each gene.
trt_slope	Numeric vector or NULL, treatment slope effects for each gene.

inter	Numeric vector or NULL, interaction effects for each gene.
sigma	Numeric vector or NULL, residual standard deviations for each gene.
mu	Numeric vector or NULL, mean expression levels for each gene.
tau	Numeric vector or NULL, overdispersion parameters for each gene (applicable if `distribution` is "NB").
act.cor	Numeric vector or NULL, actual correlations for each gene.
distribution	Character string or NULL, specifies the distribution to use for simulation. If "NB", a negative binomial distribution is used.
SingleGroup	Logical, if TRUE, simulates data for a single group.

Value

A list with the following components:

- `norm_counts`: Transposed normalized counts matrix with genes in rows and samples in columns.
- `pheno_data`: Data frame containing phenotype data for the samples.
- `feature_data`: Data frame containing feature (gene) data.

Note

The function assumes that the necessary libraries (such as `dplyr`, `edgeR`, and any other required packages) are already loaded in the R session. The actual simulation functions `wrap_simulation` and `wrap_simulation_NB` are not defined within this documentation and should be provided for the function to work.

Author(s)

Zhiwen Jiang and Tony Kam-Thong

See Also

`DGEList`

Examples

```
## Not run:
simulated_DGEList <- wrap_sim_DGEList(n = 5, m = 3, q = 2, N_gene = 1000,
                                       global_eff = rep(0, 1000), trt_eff = rep(1, 1000),
                                      trt_slope = rep(0.5, 1000), inter = rep(0.1, 1000),
                                       sigma = rep(1, 1000), mu = rep(10, 1000),
                                       tau = rep(0.2, 1000), act.cor = rep(0.3, 1000),
                                       distribution = "NB", SingleGroup = FALSE)

## End(Not run)
```

Index