Package 'SimBPDD'

May 8, 2020

Type Package

Title Simulating differential distributions for Beta-Poisson models

Version 0.1

Author Roman Schefzik

Maintainer Roman Schefzik <roman.schefzik@medma.uni-heidelberg.de>

Description The package provides functions to generate differential distributions of different types and different degrees (from weak to strong) for Beta-Poisson models. A specific focus is given to the application in the context of single-cell RNA-sequencing data, but the procedures can in principle also be employed in other application areas.

Depends R (>= 3.6)

Imports BPSC, statmod, doParallel, foreach

Remotes github::nghiavtr/BPSC

License GPL-3

LazyData TRUE

RoxygenNote 7.1.0

R topics documented:

`bp.dd.single` *Creates control and manipulated Beta-Poisson samples for pre-specified parameters*

Description

Creates control and manipulated Beta-Poisson (BP) samples for pre-specified BP parameters

Usage

```
bp.dd.single(pars,N1,N2,case,degree,seedex)
```

Arguments

pars	vector consisting of pre-specified values for the BP parameters alpha, beta, lambda1, lambda2 and p0 (in this order!)
N1	size of the sample drawn from the control BP model
N2	size of the sample drawn from the manipulated BP model
case	case corresponding to a specific parameter manipulation in the BP model; specifically, case can be one of "DLambda", "DAlpha", "DBeta", "DAlphaBeta" or "DPZ" following the design and the respective descriptive table in Schefzik (2020)
degree	parameter to set the degree of the created difference (low to strong), see Schefzik (2020) for details and the choice of a range of possible values
seedex	seed used for sampling from the fitted BP model to ensure reproducibility

Details

Creates control and manipulated BP samples for pre-specified BP parameters

Value

A list of two:

- sim.bp.ctrl: vector of simulated control BP sample
- sim.bp.manip: vector of simulated manipulated BP sample

References

R. Schefzik (2020). Simulating differential distributions in Beta-Poisson models, in particular for single-cell RNA sequencing data.

Examples

```
#create vector consisting of pre-specified values for the BP parameters alpha, beta, lambda1,
#lambda2 and p0 (in this order!)
pars<-c(0.21,3.50,96.13,0.05218,0.02)
N1<-500
N2<-500
degree<-1/3
seedex<-24
dd1<-bp.dd.single(pars,N1,N2,case="DLambda",degree,seedex)
dd2<-bp.dd.single(pars,N1,N2,case="DAlpha",degree,seedex)
dd3<-bp.dd.single(pars,N1,N2,case="DBeta",degree,seedex)
dd4<-bp.dd.single(pars,N1,N2,case="DAlphaBeta",degree,seedex)
dd5<-bp.dd.single(pars,N1,N2,case="DPZ",degree,seedex)
```

`bp.sim.control` *Creates control samples for simulations involving Beta-Poisson models*

Description

Creates control samples for simulations involving Beta-Poisson (BP) models

Usage

```
bp.sim.control(DATA,N1,seedex)
```

Arguments

`DATA`	matrix of single-cell RNA-sequencing expression data with genes in rows and samples (cells) in columns
`N1`	size of the samples drawn from the control BP models
`seedex`	seed used for sampling from the fitted BP models to ensure reproducibility

Details

Creates control samples for simulations involving BP models. For a given single-cell RNA-sequencing data set, a five-parameter BP model (Vu et al. 2016) is fitted to each gene. Then, it is checked whether the BP model is actually a good fit following the procedure in Vu et al. (2016), and instances with a good fit are kept. For each of those well-fitted genes, a sample of size N1 is drawn from the respective fitted BP model.

Value

A list of two:

- samples.bp.wellfit: matrix with simulated control single-cell RNA-sequencing expression data of dimension GxN1, where G is the number of well-fitted genes by the BP models
- parameters.bp.wellfit: matrix with G rows containing the fitting results for the corresponding BP models in the columns, namely the parameters alpha, beta, lambda1, lambda2 and p0 as in Vu et al. (2016) or Schefzik (2020) and the Monte-Carlo-method-based p-value MCpval derived to check the validity of the BP fit (here, a fit is considered to be reasonably good if MCpval>0.05)

References

R. Schefzik (2020). Simulating differential distributions in Beta-Poisson models, in particular for single-cell RNA sequencing data.
T. N. Vu, Q. F. Wills, K. R. Kalari, N. Niu, L. Wang, M. Rantalainen, and Y. Pawitan (2016). Beta-Poisson model for single-cell RNA seq data analyses. Bioinformatics, 32:2128-2135.

Examples

```
N1<-500
seedex<-24
ctrl<-bp.sim.control(DATA.EX,N1,seedex)
```

`bp.sim.DD` *Creates control and manipulated samples based on Beta-Poisson model fits for a single-cell RNA-sequencing data set*

Description

Creates control samples based on Beta-Poisson (BP) model fits for a single-cell RNA-sequencing data set and manipulated samples based on those control samples for simulations involving BP models

Usage

```
bp.sim.DD(DATA,N1,N2,case,degree,seedex)
```

Arguments

`DATA`	matrix of single-cell RNA-sequencing expression data with genes in rows and samples (cells) in columns
`N1`	size of the samples drawn from the control BP models
`N2`	size of the samples drawn from the manipulated BP models
`case`	case corresponding to a specific parameter manipulation in the BP model; specifically, case can be one of "DLambda", "DAlpha", "DBeta", "DAlphaBeta" or "DPZ" following the design and the respective descriptive table in Schefzik (2020)
`degree`	parameter to set the degree of the created difference (low to strong), see Schefzik (2020) for details and the choice of a range of possible values
`seedex`	seed used for sampling from the fitted BP models to ensure reproducibility

Details

Creates control samples based on BP model fits for a single-cell RNA-sequencing data set and manipulated samples based on those control samples for simulations involving BP models. Details regarding the design of the manipulations can be found in Schefzik (2020). Combines the functions `bp.sim.control` and `bp.sim.manipulated`.

Value

A list of two:

- controls: a list of two:
  - samples.bp.wellfit: matrix with simulated control single-cell RNA-sequencing expression data of dimension GxN1, where G is the number of well-fitted genes by the BP models
  - parameters.bp.wellfit: matrix with G rows containing the fitting results for the corresponding BP models in the columns, namely the parameters alpha, beta, lambda1, lambda2 and p0 as in Vu et al. (2016) or Schefzik (2020) and the Monte-Carlo-method-based p-value MCpval derived to check the validity of the BP fit (here, a fit is considered to be reasonably good if MCpval>0.05)
- manipulations: matrix with simulated manipulated single-cell RNA-sequencing expression data of dimension KxN2, where K is the number of well-fitted genes by the BP models for which the desired manipulation is feasible, see Schefzik (2020) for details

References

R. Schefzik (2020). Simulating differential distributions in Beta-Poisson models, in particular for single-cell RNA sequencing data.
T. N. Vu, Q. F. Wills, K. R. Kalari, N. Niu, L. Wang, M. Rantalainen, and Y. Pawitan (2016). Beta-Poisson model for single-cell RNA seq data analyses. Bioinformatics, 32:2128-2135.

Examples

```
N1<-500
N2<-500
degree<-1/3
seedex<-24
bp1<-bp.sim.DD(DATA.EX,N1,N2,case="DLambda",degree,seedex)
bp2<-bp.sim.DD(DATA.EX,N1,N2,case="DAlpha",degree,seedex)
bp3<-bp.sim.DD(DATA.EX,N1,N2,case="DBeta",degree,seedex)
bp4<-bp.sim.DD(DATA.EX,N1,N2,case="DAlphaBeta",degree,seedex)
bp5<-bp.sim.DD(DATA.EX,N1,N2,case="DPZ",degree,seedex)
```

`bp.sim.manipulated` *Creates manipulated samples for simulations involving Beta-Poisson models*

Description

Creates manipulated samples based on the control samples for simulations involving Beta-Poisson (BP) models

Usage

```
bp.sim.manipulated(Res.par,N2,case,degree,seedex)
```

Arguments

`Res.par`	matrix including the fitted parameters for the well-fitted control instances that are to be manipulated (e.g. the output matrix parameters.bp.wellfit from the `bp.sim.control` function)
`N2`	size of the samples drawn from the manipulated BP models
`case`	case corresponding to a specific parameter manipulation in the BP model; specifically, case can be one of "DLambda", "DAlpha", "DBeta", "DAlphaBeta" or "DPZ" following the design and the respective descriptive table in Schefzik (2020)
`degree`	parameter to set the degree of the created difference (low to strong), see Schefzik (2020) for details and the choice of a range of possible values
`seedex`	seed used for sampling from the fitted BP models to ensure reproducibility

Details

Creates manipulated samples based on the control samples for simulations involving BP models. Details regarding the design of the manipulations can be found in Schefzik (2020).

Value

matrix with simulated manipulated single-cell RNA-sequencing expression data of dimension KxN2, where K is the number of well-fitted genes by the BP models for which the desired manipulation is feasible, see Schefzik (2020) for details

References

R. Schefzik (2020). Simulating differential distributions in Beta-Poisson models, in particular for single-cell RNA sequencing data.

Examples

```
N1<-500
seedex<-24
ctrl<-bp.sim.control(DATA.EX,N1,seedex)
N2<-500
degree<-1/3
man1<-bp.sim.manipulated(ctrl[[2]],N2,case="DLambda",degree,seedex)
man2<-bp.sim.manipulated(ctrl[[2]],N2,case="DAlpha",degree,seedex)
man3<-bp.sim.manipulated(ctrl[[2]],N2,case="DBeta",degree,seedex)
man4<-bp.sim.manipulated(ctrl[[2]],N2,case="DAlphaBeta",degree,seedex)
man5<-bp.sim.manipulated(ctrl[[2]],N2,case="DPZ",degree,seedex)
```

DATA.EX *Examplary single-cell RNA-sequencing data set*

Description

An examplary single-cell RNA-sequencing data set based on an excerpt of the data set by Pollen et al. (2014).

Usage

```
data(DATA.EX)
```

Format

A data matrix containing normalized expression values for 20 genes (rows) and 301 cells (samples; columns)

Source

Excerpt of the data set by Pollen et al. (2014), available at https://hemberg-lab.github.io/scRNA.seq.datasets/human/tissues/

References

A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, N. Li, L. Szpankowski,B. Fowler, P. Chen, N. Ramalingam, G. Sun, M. Thu, M. Norris, R. Lebofsky, D. Toppani, D. W. Kemp II, M. Wong, B. Clerkson, B. N. Jones, S. Wu, L. Knutsson, B. Alvarado, J. Wang, L. S. Weaver, A. P. May, R. C. Jones, M. A. Unger, A. R. Kriegstein, and J. A. A. West (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nature Biotechnology, 32:1053-1058.

Index