

Package ‘noise’

February 29, 2024

Type Package

Title Estimation of Intrinsic and Extrinsic Noise from Single-Cell Data

Version 1.0.2

Date 2024-02-29

Author Audrey Qiuyan Fu and Lior Pachter

Maintainer Audrey Q. Fu <audreyqyfu@gmail.com>

Description Functions to calculate estimates of intrinsic and extrinsic noise from the two-reporter single-cell experiment, as in Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002) Stochastic gene expression in a single cell. *Science*, 297, 1183-1186. Functions implement multiple estimators developed for unbiasedness or min Mean Squared Error (MSE) in Fu, A. Q. and Pachter, L. (2016). Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. *Statistical Applications in Genetics and Molecular Biology*, 15(6), 447-471.

License GPL (>= 2)

Depends R (>= 2.10)

LazyLoad yes

LazyData yes

Encoding UTF-8

Imports preprocessCore

NeedsCompilation no

RoxygenNote 7.3.1

Repository CRAN

Date/Publication 2024-02-29 14:40:05 UTC

R topics documented:

computeExtrinsicNoise	2
computeExtrinsicNoiseKnownCor	3
computeIntrinsicNoise	4

computeNoiseForSubset	5
elowitz_data	6
simulateSC	7
yang_n110	9

Index	11
--------------	-----------

computeExtrinsicNoise *Compute estimates of the extrinsic noise.*

Description

This function computes several estimates of the extrinsic noise (unscaled by the mean). The estimators, described in Fu and Pachter (2016), include the original estimators developed in Elowitz et al. (2002), the unbiased estimator, an min-MSE estimator, and an asymptotic estimator for large sample sizes.

Usage

```
computeExtrinsicNoise(reporter1, reporter2)
```

Arguments

reporter1 A vector of continuous values.
reporter2 A vector of continuous values.

Value

Four (unscaled) estimates of extrinsic noise: the original estimators developed in Elowitz et al. (2002), the unbiased estimator, an min-MSE estimator, and an asymptotic estimator for large sample sizes.

Author(s)

Audrey Q. Fu

References

Fu, A. Q. and Pachter, L. (2016). Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. arXiv:1601.03334. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002) Stochastic gene expression in a single cell. *Science*, 297, 1183-1186.

See Also

[computeIntrinsicNoise](#), [simulateSC](#). See estimates for data [elowitz_data](#) and [yang_n110](#).

`computeExtrinsicNoiseKnownCor`

Compute estimates of the extrinsic noise using a known correlation in the min-MSE estimator.

Description

This function is similar to [computeExtrinsicNoise](#), and computes several estimates of the extrinsic noise (unscaled by the mean). The estimators, described in Fu and Pachter (2016), include the original estimators developed in Elowitz et al. (2002), the unbiased estimator, an min-MSE estimator, and an asymptotic estimator for large sample sizes. The only difference between this function calculates the min-MSE estimate using a given correlation.

Usage

```
computeExtrinsicNoiseKnownCor(reporter1, reporter2, true.cor)
```

Arguments

<code>reporter1</code>	A vector of continuous values.
<code>reporter2</code>	A vector of continuous values.
<code>true.cor</code>	A scalar.

Value

Four (unscaled) estimates of extrinsic noise: the original estimators developed in Elowitz et al. (2002), the unbiased estimator, an min-MSE estimator (using the given correlation), and an asymptotic estimator for large sample sizes.

Author(s)

Audrey Q. Fu

References

Fu, A. Q. and Pachter, L. (2016). Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. arXiv:1601.03334. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002) Stochastic gene expression in a single cell. Science, 297, 1183-1186.

See Also

[computeExtrinsicNoise](#), [simulateSC](#).

computeIntrinsicNoise *Compute estimates of the intrinsic noise.*

Description

This function computes several estimates of the intrinsic noise (unscaled by the mean). The estimators, described in Fu and Pachter (2016), include the original estimators developed in Elowitz et al. (2002), unbiased estimators with and without assuming equal mean of the two reporters, min-MSE estimators with and without assuming equal mean, and asymptotic estimators for large sample sizes with and without assuming equal mean.

Usage

```
computeIntrinsicNoise(reporter1, reporter2)
```

Arguments

reporter1	A vector of continuous values.
reporter2	A vector of continuous values.

Value

Six (unscaled) estimates of intrinsic noise: the original estimators developed in Elowitz et al. (2002), unbiased estimators with and without assuming equal mean of the two reporters, min-MSE estimators with and without assuming equal mean, and asymptotic estimators for large sample sizes with and without assuming equal mean.

Author(s)

Audrey Q. Fu

References

Fu, A. Q. and Pachter, L. (2016). Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. arXiv:1601.03334. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002) Stochastic gene expression in a single cell. Science, 297, 1183-1186.

See Also

[computeExtrinsicNoise](#), [simulateSC](#). See estimates for data [elowitz_data](#) and [yang_n110](#).

computeNoiseForSubset *Select a random subset of data for noise estimation.*

Description

This function randomly selects a subset of cells (rows) from the data set, computes multiple estimates of intrinsic and extrinsic noise, as well as their mean and standard deviation.

Usage

```
computeNoiseForSubset(data, sample.size, n.iter)
```

Arguments

data	A numeric matrix of two columns. Each row is a cell, and each column expression of a reporter gene.
sample.size	An integer that specifies the number of cells in the subset.
n.iter	An integer that specifies the number of iterations (for calculation of mean and standard deviation).

Value

A list that consists of the following components:

intrinsic	A numeric matrix of estimated intrinsic noise. 7 rows and n.iter columns.
extrinsic	A numeric matrix of estimated extrinsic noise. 4 rows and n.iter columns.
intrinsic.mean	A numeric vector of length 7 that contains the mean estimates of intrinsic noise.
intrinsic.sd	A numeric vector of length 7 that contains the standard deviation of the estimates of intrinsic noise.
extrinsic.mean	A numeric vector of length 7 that contains the mean estimates of extrinsic noise.
extrinsic.sd	A numeric vector of length 7 that contains the standard deviation of the estimates of extrinsic noise.

Author(s)

Audrey Q. Fu

References

Fu, A. Q. and Pachter, L. (2016). Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. arXiv:1601.03334.

See Also

[computeIntrinsicNoise](#), [computeExtrinsicNoise](#), [elowitz_data](#), [yang_n110](#).

Examples

```

data(yang_n110)

# quantile normalization on log2 transformed data
# install bioconductor package for quantile normalization
# source('http://bioconductor.org/biocLite.R')
# biocLite('preprocessCore')
library(preprocessCore)

# ignore a few values that are negative
yang_n110.pos <- yang_n110[-which (yang_n110[,1]<0),]
yang_n110.pos.log2.quant <- normalize.quantiles (as.matrix (log2 (yang_n110.pos[,c(1,3)])))

# subset the data and compute noise estimates
yang.50 <- computeNoiseForSubset (yang_n110.pos.log2.quant, sample.size=50, n.iter=1000)
summary (yang.50)

```

elowitz_data	<i>Expression of reporter genes in the D22 and M22 E. coli cells from Elowitz et al (2002).</i>
--------------	---

Description

Expression of reporter genes CFP and YFP in over 200 E. coli cells of two strains: D22 and M22. These values are displayed in a scatterplot in Elowitz et al (2002) Fig 3a.

Usage

```
data("elowitz_data")
```

Format

The format is: List of 2 \$ D22:'data.frame': 284 obs. of 2 variables: ..\$ CFP: num [1:284] 3080 3082 2893 3053 2891\$ YFP: num [1:284] 2309 2394 2145 2340 2245 ... \$ M22:'data.frame': 250 obs. of 2 variables: ..\$ CFP: num [1:250] 2438 2316 2521 2646 2830\$ YFP: num [1:250] 1409 1391 1511 1460 1638 ...

References

Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002) Stochastic gene expression in a single cell. *Science*, 297, 1183-1186.

Examples

```

data(elowitz_data)

# Normalize data such that they are
# comparable to Fig 3a in Elowitz et al. (2002).
# Normalized data have mean 1.

```

```

D22.cfp.norm <- (elowitz_data$D22[,1]-mean (elowitz_data$D22[,1]))/sd(elowitz_data$D22[,1])/8+1
D22.yfp.norm <- (elowitz_data$D22[,2]-mean (elowitz_data$D22[,2]))/sd(elowitz_data$D22[,2])/8+1

M22.cfp.norm <- (elowitz_data$M22[,1]-mean (elowitz_data$M22[,1]))/sd(elowitz_data$M22[,1])/12+1
M22.yfp.norm <- (elowitz_data$M22[,2]-mean (elowitz_data$M22[,2]))/sd(elowitz_data$M22[,2])/12+1

# Compute noise estimates.
# Since the mean is 1, estimates with and without
# the scaling are the same.
unlist (computeIntrinsicNoise (D22.cfp.norm, D22.yfp.norm))
unlist (computeExtrinsicNoise (D22.cfp.norm, D22.yfp.norm))

unlist (computeIntrinsicNoise (M22.cfp.norm, M22.yfp.norm))
unlist (computeExtrinsicNoise (M22.cfp.norm, M22.yfp.norm))

```

simulateSC

Simulate single-cell expression levels of two reporters

Description

This function simulates expression levels of two reporters across single cells, mimicking the two-reporter assay. The hierarchical model described in Fu and Pachter (2016) is used for simulation. We further make the simplifying assumption that intrinsic noise is the same across cells.

Usage

```
simulateSC(n = 1000, intrinsic = 0.7, extrinsic = 0.8, mean = 1)
```

Arguments

n	Number of single cells (sample size).
intrinsic	Scalar. The (unscaled) intrinsic noise (or within-cell variability), denoted by σ^2 (equation 8) in Fu and Pachter (2016).
extrinsic	Scalar. The (unscaled) extrinsic noise (or between-cell variability), denoted by σ_μ^2 (equation 9) in Fu and Pachter (2016).
mean	Scalar. The overall mean of expression level, denoted by μ (equation 6) in Fu and Pachter (2016).

Value

A data frame of two columns and n rows. Each column contains the expression levels of a reporter. Each row is a single cell.

Author(s)

Audrey Q. Fu

References

Fu, A. Q. and Pachter, L. (2016). Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. arXiv:1601.03334.

See Also

[computeIntrinsicNoise](#), [computeExtrinsicNoise](#).

Examples

```
# simulation #1

# simulate 500 data sets
n.simu <- 500

# true intrinsic and extrinsic noise
int.true <- 0.7
ext.true <- 0.8

# create matrices to hold estimated intrinsic and extrinsic noise
# using different estimators
int.simu.mtx <- matrix(0, nrow=n.simu, ncol=8)
ext.simu.mtx <- matrix(0, nrow=n.simu, ncol=4)

for (i in 1:n.simu) {
  n <- 1000
  simu <- simulateSC (n=n, intrinsic=int.true, extrinsic=ext.true, mean=1)

  int.simu.mtx[i,] <- c(unlist (computeIntrinsicNoise (simu[,1], simu[,2])),
    cor (simu[,1], simu[,2]))
  ext.simu.mtx[i,] <- unlist (computeExtrinsicNoise (simu[,1], simu[,2]))
}

# add column names to simulation estimates
colnames (int.simu.mtx) <- c("ELSS", "unbiasedGeneral", "unbiasedEqualMean",
  "minMSEGeneral", "minMSEEqualMean", "asymptoticGeneral",
  "asymptoticEqualMean", "cor")
colnames (ext.simu.mtx) <- c("ELSS", "unbiased", "minMSE", "asymptotic")

# simulation #2

# simulate 500 data sets
n.simu <- 500

# true intrinsic and extrinsic noise
int.true <- 0.7
ext.true <- 0.8

# use true correlation for the min-MSE estimates of extrinsic noise
true.cor <- ext.true / (ext.true + int.true)
```



```

# create matrices to hold estimated intrinsic and extrinsic noise
# using different estimators
int.simu.mtx <- matrix (0, nrow=n.simu, ncol=8)
ext.simu.mtx <- matrix (0, nrow=n.simu, ncol=4)
ext.simu.mtx.2 <- matrix (0, nrow=n.simu, ncol=4)

for (i in 1:n.simu) {
  n <- 50
  simu <- simulateSC (n=n, intrinsic=int.true, extrinsic=ext.true, mean=1)

  int.simu.mtx[i,] <- c(unlist (computeIntrinsicNoise (simu[,1], simu[,2])),
    cor (simu[,1], simu[,2]))
  ext.simu.mtx[i,] <- unlist (computeExtrinsicNoise (simu[,1], simu[,2]))
  ext.simu.mtx.2[i,] <- c(unlist (computeExtrinsicNoiseKnownCor (simu[,1],
    simu[,2], true.cor)))
}

# add column names to simulation estimates
colnames (int.simu.mtx) <- c("ELSS", "unbiasedGeneral", "unbiasedEqualMean",
  "minMSEGeneral", "minMSEEqualMean", "asymptoticGeneral",
  "asymptoticEqualMean", "cor")
colnames (ext.simu.mtx) <- c("ELSS", "unbiased", "minMSE", "asymptotic")
colnames (ext.simu.mtx.2) <- c("ELSS", "unbiased", "minMSE", "asymptotic")

# compute the MSE of estimates
computeMSE <- function (a, t) {return (mean((a-t)^2))}
apply (int.simu.mtx[,1:7], 2, computeMSE, t=int.true)
apply (ext.simu.mtx, 2, computeMSE, t=ext.true)
apply (ext.simu.mtx.2, 2, computeMSE, t=ext.true)

```

yang_n110

Expression of reporter genes in the NL010 E. coli cells from Yang et al (2014).

Description

Expression of reporter genes CFP and mCherry in over 40,000 E. coli cells. A subset of these values are displayed in a scatterplot in Yang et al (2014) Fig 3a rightmost panel.

Usage

```
data("yang_n110")
```

Format

A data frame with 40683 observations on the following 3 variables.

CFP a numeric vector

Venus a numeric vector
mCherry a numeric vector

References

Yang, S., S. Kim, Y. R. Lim, C. Kim, H. J. An, J.-H. Kim, J. Sung, and N. K. Lee (2014) Contribution of RNA polymerase concentration variation to protein expression noise. *Nature Communications*, 5, 4761.

Examples

```
data(yang_n110)

# compute the noise estimates

# no normalization
# unscaled by mean
unlist (computeIntrinsicNoise (yang_n110[,1], yang_n110[,3]))
unlist (computeExtrinsicNoise (yang_n110[,1], yang_n110[,3]))

# scaled by mean
unlist (computeIntrinsicNoise (yang_n110[,1], yang_n110[,3])) /
  mean (yang_n110[,1]) / mean(yang_n110[,3])
unlist (computeExtrinsicNoise (yang_n110[,1], yang_n110[,3])) /
  mean (yang_n110[,1]) / mean(yang_n110[,3])

# quantile normalization on log2 transformed data
# install bioconductor package for quantile normalization
# source('http://bioconductor.org/biocLite.R')
# biocLite('preprocessCore')
library(preprocessCore)

# ignore a few values that are negative
yang_n110.pos <- yang_n110[-which (yang_n110[,1]<0),]
yang_n110.pos.log2.quant <- normalize.quantiles (as.matrix (log2 (yang_n110.pos[,c(1,3)])))

# unscaled by mean
unlist (computeIntrinsicNoise (yang_n110.pos.log2.quant[,1], yang_n110.pos.log2.quant[,2]))
unlist (computeExtrinsicNoise (yang_n110.pos.log2.quant[,1], yang_n110.pos.log2.quant[,2]))

# scaled by mean
unlist (computeIntrinsicNoise (yang_n110.pos.log2.quant[,1], yang_n110.pos.log2.quant[,2])) /
  mean (yang_n110.pos.log2.quant[,1]) / mean(yang_n110.pos.log2.quant[,2])
unlist (computeExtrinsicNoise (yang_n110.pos.log2.quant[,1], yang_n110.pos.log2.quant[,2])) /
  mean (yang_n110.pos.log2.quant[,1]) / mean(yang_n110.pos.log2.quant[,2])
```

Index

* datasets

elowitz_data, 6

yang_n110, 9

computeExtrinsicNoise, 2, 3–5, 8

computeExtrinsicNoiseKnownCor, 3

computeIntrinsicNoise, 2, 4, 5, 8

computeNoiseForSubset, 5

elowitz_data, 2, 4, 5, 6

simulateSC, 2–4, 7

yang_n110, 2, 4, 5, 9