Package 'SLIP'

October 15, 2021

Type Package

Title Change Testing for Large-Scale Datastreams with FDR Control

Version 0.1.0

Author Mengtao WEN

Maintainer The package maintainer <wenmengtao@pku.edu.cn>

Description A data-driven procedure for identifying abnormal datastreams each of which occurs at least one change while controling false discoveries refer to that an identified datastream, among large parallel datastreams, that actually occurs no changes.

License GPL-2

Depends R (>= 3.6.0)

Imports glmnet,
POET,
CovTools,
simex,
knitr,
rmarkdown

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Suggests markdown

VignetteBuilder knitr

R topics documented:

`arrayLowReso` *Reduce resolution of image*

Description

Reduce resolution of image

Usage

```
arrayLowReso(A, pixel, func)
```

Arguments

`A`	an array
`pixel`	a vector determining how much pixels of initial image one pixel of the output image contains; such as `c(2,2,2)` for a 3D array `A`
`func`	the operation executed to get the output image; such as `mean` or `max`

Value

an array with dim = `ceiling(dim(A)/pixel)`

Examples

```
A = array(1:27, c(3, 3, 3))
pixel = c(2, 2, 2)
B = arrayLowReso(A, pixel, mean)
```

`BH.asymp` *BH with p-values calculated by extreme value theory*

Description

BH with p-values calculated by extreme value theory

Usage

```
BH.asymp(dat, alpha, varEst = T, D_sig = NULL, outputCP = FALSE)
```

Arguments

`dat`	n x p data matrix (p features, n observations)
`alpha`	FDR nominal level
`varEst`	logical 0 or 1; estimate variance or not (TRUE for esitmation)
`D_sig`	p x 1 vector, true variances of streams, optional only when `varEst = FALSE`
`outputCP`	logical parameter `FALSE`(default); if `TRUE`, the change-point location in (0, 1) corresponding to signals will be returned.

Value

A list contains:

sig	indices of signals
FDP	estiamted FDP
cps	change-points, optional only when outputCP = TRUE

Examples

```
N = 120; p = 200
data = SLIP.scp.generator(N, p)
BH.asymp(data$dat, 0.1)
```

BH.simul *BH with p-values calculated by Bootstrap*

Description

BH with p-values calculated by Bootstrap

Usage

```
BH.simul(dat, alpha, ECDF, varEst = T, D_sig = NULL, outputCP = FALSE)
```

Arguments

dat	n x p data matrix (p features, n observations)
alpha	FDR nominal level
ECDF	empirical null distribution of CUSUM
varEst	logical 0 or 1; estimate variance or not (TRUE for esitmation)
D_sig	p x 1 vector, true variances of streams, optional only when varEst = FALSE
outputCP	logical parameter FALSE(default); if TRUE, the change-point location in (0, 1) corresponding to signals will be returned.

Value

A list contains:

sig	indices of signals
FDP	estiamted FDP
cps	change-points, optional only when outputCP = TRUE

Examples

```
N = 120; p = 200; B = 1000
data = SLIP.scp.generator(N, p)
ECDF = bootstrap.cusum(N, B)
BH.simul(data$dat, 0.1, ECDF)
```

bootstrap.cusum *Bootstrap for CUSUM statistics*

Description

Bootstrap for CUSUM statistics

Usage

```
bootstrap.cusum(N, B = 20000)
```

Arguments

N	the length of the stream
B	the number of simulations

Value

B x 1 vector; empirical distribution of the CUSUM

fmri.data *fMRI data used in our paper*

Description

A dataset contain the regions and their region-averaged hemodynamic response

Usage

```
fmri.data
```

Format

A list with 6 compoenents

dat a 360 x 264 data matrix (360 scans, 264 regions)

regions names of regions. For example, "2001xlylxl", "2001" indicates the Tzourio-Mazoyer code in the Anatomical Automatic Labeling (AAL) partition and "xlylxl" indicates the sub-region of the certain ROI.

coord a data.frame with four columns: the first three columns indicate the (x, y, z) location in 3D fMRI image and the last column indicates the corresponding region.

dimx dimension of x axis of 3D fMRI image.

dimy dimension of y axis of 3D fMRI image.

dimz dimension of z axis of 3D fMRI image.

Source

original data from the subject P2 on http://www.cs.cmu.edu/~fmri/science2008/data.html

References

"Predicting Human Brain Activity Associated with the Meanings of Nouns," T. M. Mitchell, S. V. Shinkareva, A. Carlson, K.M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, Science, 320, 1191, May 30, 2008. DOI: 10.1126/science.1152876.

`SLIP.indep` *SLIP without incorporating dependence*

Description

Use SLIP without incorporating dependence to detect abnormal data streams each of which occurs at least one change.

Usage

```
SLIP.indep(
  dat,
  alpha,
  r = 3,
  varEst = T,
  D_sig = NULL,
  outputW = FALSE,
  outputCP = FALSE
)
```

Arguments

`dat`	n x p matrix (p features, n observations)
`alpha`	FDR nominal level
`r`	splitting ratio, (r-1) pieces versus 1 piece
`varEst`	logical 0 or 1; estimate variance or not (TRUE for esitmation)
`D_sig`	p x 1 vector, true variances of streams, optional only when `varEst = FALSE`
`outputW`	a logical parameter `FALSE`(default); if `TRUE`, the W-statistics and the threshold will be returned.
`outputCP`	logical parameter `FALSE`(default); if `TRUE`, the change-point location in (0, 1) corresponding to signals will be returned.

Value

A list contains:

`sig`	indices of signals
`FDP`	estiamted FDP
`W`	W-statistic, optional only when `W = TRUE`
`L`	threshold, optional only when `W = TRUE`
`cps`	change-points, optional only when `outputCP = TRUE`

Examples

```
N = 120; p = 200
data = SLIP.scp.generator(N, p)
SLIP.indep(data$dat, 0.1)
```

`SLIP.lasso` *SLIP with the Lasso*

Description

Use SLIP with mean screening to detect abnormal data streams each of which occurs at least one change.

Usage

```
SLIP.lasso(
  dat,
  alpha,
  r = 3,
  covEst = T,
  estMthd = "Cholesky",
  trueCov = NULL,
  upperPi = 0.5,
  outputW = FALSE,
  outputCP = FALSE
)
```

Arguments

`dat`	n x p matrix (p features, n observations)
`alpha`	FDR nominal level
`r`	splitting ratio, (r-1) pieces versus 1 piece
`covEst`	Estimate covariance or not (logical); T for Est
`estMthd`	optional estimation methods `c("Cholesky","POET")`
`trueCov`	the true covariance matrix; only optional when `covEst=F`
`upperPi`	Assumed upper bound of the number of signals; 0.5(default)
`outputW`	a logical parameter `FALSE`(default); if `TRUE`, the W-statistics and the threshold will be returned.
`outputCP`	logical parameter `FALSE`(default); if `TRUE`, the change-point location in (0, 1) corresponding to signals will be returned.

Value

A list contains:

`sig`	indices of signals
`FDP`	estiamted FDP
`W`	W-statistic, optional only when `W = TRUE`
`L`	threshold, optional only when `W = TRUE`
`cps`	change-points, optional only when `outputCP = TRUE`

Examples

```
N = 120; p = 200
data = SLIP.scp.generator(N, p)
SLIP.lasso(data$dat, 0.1)
```

`SLIP.mcp.generator` *Generate n-by-p matrix with at least one changes in each abnormal datastream*

Description

Generate n-by-p matrix with at least one changes in each abnormal datastream

Usage

```
SLIP.mcp.generator(
  n,
  p,
  cov = "Identity",
  rho = NULL,
  dist = "Gaussian",
  ratio = 0.15,
  delta = 1,
  varrho = 0.05,
  cpn = 4,
  Sigma = NULL,
  param = NULL
)
```

Arguments

`n`	the number of observations
`p`	the number of datastreams
`cov`	covariance type `c("CS","Factor","AR1","Block","Sparse","Identity","Given")`
`rho`	optional when using `cov` from `c("CS","AR1","Block")`, rho in [0, 1]
`dist`	noise distribution `c("Gaussian","t","exp","chisq")`
`ratio`	abnormal streams/total streams = ratio, ratio in [0, 1]
`delta`	the magnitudes of changes (delta >= 0) lie in [delta-0.1, delta+0.1] with equally probable sign from +, -
`varrho`	the parameter avoiding the boundary problem, [0, 0.5)
`cpn`	the parameter of Poisson distribution related to the number of change-points at each of abnormal streams
`Sigma`	the covariance matrix, optional only when `cov=('Given')`
`param`	the parameter when using `dist` from `c("t","chisq","exp")`

Value

data with changes in some datastreasm

dat	n-by-p data matrix
index	those datastreams containing changes
cpnum	a vector of the number of change-points
mu	a list of values of changes
loc	a list of locations of changes

`SLIP.scp.generator` *Generate n-by-p matrix with one changes in each abnormal datastream*

Description

Generate n-by-p matrix with one changes in each abnormal datastream

Usage

```
SLIP.scp.generator(
  n,
  p,
  cov = "Identity",
  dist = "Gaussian",
  rho = NULL,
  ratio = 0.15,
  delta = 1,
  varrho = 0.05,
  Sigma = NULL,
  param = NULL
)
```

Arguments

n	the number of observations
p	the number of datastreams
cov	covariance type; c("CS","Factor","AR1","Block","Sparse","Identity","Given")
dist	noise distribution c("Gaussian","t","exp","chisq")
rho	optional when using cov from c("CS","AR1","Block"), rho in [0, 1]
ratio	abnormal streams/total streams = ratio, ratio in [0, 1]
delta	the magnitudes of changes (delta >= 0) lie in [delta-0.1, delta+0.1] with equally probable sign from +, -
varrho	the parameter avoiding the boundary problem, [0, 0.5)
Sigma	the covariance matrix, optional only when cov=("Given")
param	the parameter when using dist from c("t","chisq","exp")

Value

data with changes in some datastreasm

dat	n-by-p data matrix
index	those datastreams containing changes
mu	changes
loc	locations of changes

SLIP.thresh.c *SLIP with thresholding (c version)*

Description

Use SLIP with mean screening to detect abnormal data streams each of which occurs at least one change.

Usage

```
SLIP.thresh.c(
  dat,
  alpha,
  C = 1.5,
  r = 3,
  covEst = T,
  estMthd = "Cholesky",
  trueCov = NULL,
  outputW = FALSE,
  outputCP = FALSE
)
```

Arguments

dat	n x p matrix (p features, n observations)
alpha	FDR nominal level
C	tuning parameter for mean screening; [1.4, 1.6] is suitable; larger p, bigger C
r	splitting ratio, (r-1) pieces versus 1 piece
covEst	Estimate covariance or not (logical); T for Est
estMthd	optional estimation methods c("Cholesky","POET")
trueCov	the true covariance matrix; only optional when covEst=F
outputW	logical parameter FALSE(default); if TRUE, the W-statistics and the threshold will be returned.
outputCP	logical parameter FALSE(default); if TRUE, the change-point location in (0, 1) corresponding to signals will be returned.

Value

A list contains:

sig	indices of signals
FDP	estiamted FDP
W	W-statistic, optional only when outputW = TRUE
L	threshold, optional only when outputW = TRUE
cps	change-points, optional only when outputCP = TRUE

Examples

```
N = 120; p = 200
data = SLIP.scp.generator(N, p)
SLIP.thresh.c(data$dat, 0.1)
```

SLIP.thresh.d *SLIP with thresholding (d version)*

Description

Use SLIP with mean screening to detect abnormal data streams each of which occurs at least one change.

Usage

```
SLIP.thresh.d(
  dat,
  alpha,
  upperPi = 0.5,
  r = 3,
  covEst = T,
  estMthd = "Cholesky",
  trueCov = NULL,
  outputW = FALSE,
  outputCP = FALSE
)
```

Arguments

dat	n x p matrix (p features, n observations)
alpha	FDR nominal level
upperPi	retained proportion after thresholding 0 < d < 1(default 0.5)
r	splitting ratio, (r-1) pieces versus 1 piece
covEst	Estimate covariance or not (logical); T for Est
estMthd	optional estimation methods c("Cholesky","POET")
trueCov	the true covariance matrix; only optional when covEst=F
outputW	a logical parameter FALSE(default); if TRUE, the W-statistics and the threshold will be returned.
outputCP	logical parameter FALSE(default); if TRUE, the change-point location in (0, 1) corresponding to signals will be returned.

Value

A list contains:

`sig`	indices of signals
`FDP`	estiamted FDP
`W`	W-statistic, optional only when `W = TRUE`
`L`	threshold, optional only when `W = TRUE`
`cps`	change-points, optional only when `outputCP = TRUE`

Examples

```
N = 120; p = 200
data = SLIP.scp.generator(N, p)
SLIP.thresh.d(data$dat, 0.1)
```

Index