Package 'RenewQIF'

October 3, 2019

Type Package

Title Real-time Regression Analysis of Streaming Clustered Data with Possible Abnormal Data Batches

Version 1.0

Date 2019-10-03

Author Lan Luo

Maintainer Lan Luo <luolsph@umich.edu>

Description This package updates the regression coefficients and their standard errors in streaming datasets that contain cluster- or longitudinal-correlated observations (note that data batches are independent of each other). Additionally, user can turn on the monitoring procedure to check if there are abnormal data batches.

License GPL-2

Imports Rcpp (>= 0.12.18), MASS, SimCorMultRes

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

R topics documented:

RenewQIF-package | *Real-time Regression Analysis of Streaming Clustered Data with Possible Abnormal Data Batches*

Description

This package updates the regression coefficients and their standard errors in streaming datasets that contain cluster- or longitudinal-correlated observations (note that data batches are independent of each other). Additionally, user can turn on the monitoring procedure to check if there are abnormal data batches.

1

Details

This package is used to conduct real-time regression analysis of streaming clustered data with possible abnormal data batches.

Author(s)

Lan Luo

Maintainer: Lan Luo <luolsph@umich.edu>

References

Real-time Regression Analysis of Streaming Clustered Data with Possible Abnormal Data Batches.

Examples

```
NB <- 10000  # total sample size
B <- 100  # number of data batches
n <- round(NB/B) # data batch size
m <- 10  # cluster size

thred <- 0.05  # threshold of statistical significance level
QC <- FALSE  # specify whether to conduct the monitoring procedure

cpt1 <- round(B * 0.25)
cpt2 <- round(B * 0.75)
beta0 <- c(1, 1)
beta1 <- beta2 <- beta0

p <- length(beta0)

type <- family <-"gaussian"
corst_x <- "cs"
rho_x <- 0.5

intercept <- TRUE
categorical <- FALSE

## True correlation matrix
corst_y <- "cs"

rho_y <- 0.7
phi_y <- 1
seed <- 1
## Working matrix
corstr <- "exchangeable"

tempdatadir <- "~/Desktop/tempdata"

datagenerator(n, m, p, B, tempdatadir, type, beta0, beta1, beta2, cpt1, cpt2,
intercept = intercept, categorical = categorical,
corst_x = corst_x, rho_x = rho_x, corst_y = corst_y, rho_y = rho_y, phi_y = phi_y,
seed = seed)

out <- renewqif(B = B, tempdatadir = tempdatadir, family = family, intercept = intercept, corstr = corstr,
     QC = QC, thred = thred)
```

```
unlink(tempdatadir)

# Output the regression coefficient estimates without abnormal data batch
out$beta
```

---

datagenerator	*Generate data streams with correlated subjects within each single data batch.*

Description

Generate input data for `renewqif`, may include two abnormal data batches.

Usage

```
datagenerator(n, m, p, B, tempdatadir, type, beta0, beta1, beta2, cpt1, cpt2,
  intercept = intercept, categorical = categorical,
 corst_x = corst_x, rho_x = rho_x, corst_y = corst_y, rho_y = rho_y, phi_y = phi_y,
  seed = seed)
```

Arguments

n	Sample size for each single data batch.
m	Number of repeated measurements or cluster size.
p	Number of covariates (including intercept).
B	Number of data batches/streams.
tempdatadir	Directory that stores the input data streams.
type	Type/Family of response variable, c("gaussian", "binomial", "poisson").
beta0	Main underlying true regression coefficients, length equals to p.
beta1	Coefficients for the first abnormal data batch, if it exists. Otherwise set beta1 = beta0.
beta2	Coefficients for the second abnormal data batch, if it exists. Otherwise set beta2 = beta0.
cpt1	Index for the first abnormal data batch.
cpt2	Index for the second abnormal data batch.
intercept	Logical, if TRUE, beta0[1] will be the coefficient for intercept.
categorical	logical, if TRUE, the first covariate of X will be changed to categorical variable.
corst_x	Correlation structure of X, c("ind", "cs", "ar1").
rho_x	Correlation coefficient for the covariance of X.
corst_y	Correlation structure of y, c("ind", "cs", "ar1").
rho_y	Correlation coefficient for the covariance of y (outcome vector for each cluster).
phi_y	Dispersion parameter for generating y.
seed	random seed for simulation.

Index