**Welcome to mvLIS - Large-Scale Multiple Testing via Multivariate Hidden Markov Models**

Tingting Cui, and Pengfei Wang \*

September 2, 2020

**Description**

mvLIS is a program to carry out multiple testing via multivariate hidden Markov models, which is written in R code. Here mvLIS program models the temporal or one-dimensional sequential dependence structure among tests via a Markov chain and incorporates auxiliary information via multi-dimensional summary statistics.

**Citations**

1. Sun W., Cai T. Large-scale multiple testing under dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009;71(2):393–424.

2. Cui T., Wang P. Large-scale multiple testing via multivariate hidden Markov models.

**Download**s

* Some core code of mvLIS procedure are available on GitHub (https://github.com/wpf19890429/Multiple-testing-via-multivariate-HMM). This repository contains the following files:

rdata1\_mvHMM.R

bwfw1\_mvHMM.R

em1\_mvHMM.R

rdata\_mvHMM.R

bwfw\_mvHMM.R

em\_mvHMM.R

mt.hmm.R.txt

**illustrations of core R functions**

1. rdata1\_mvHMM.R

Description:

Generating the observed 2-dimensional z-values and the states of hypotheses that are based on multivariate hidden Markov models (L=1).

Usage:

rdata1\_mvHMM (NUM, pii, A, mu\_0, Sigma\_0, mu\_1, Sigma\_1)

Arguements:

NUM: the number of hypotheses

pii=(pii[1], pii[2]): initial state distribution

A: transition matrix

mu\_0: mean vector under the null

Sigma\_0: covariance matrix under the null

mu\_1: mean vector under the non-null

Sigma\_1: covariance matrix under the non-null

Values:

o: observed 2-dimensional z-values

s: underlying states of hypotheses

2. bwfw1\_mvHMM.R

Description:

Calculating the mvLIS statistics via the backward-forward algorithm (L=1).

Usage:

bwfw1\_mvHMM (z, A, mu\_0, Sigma\_0, mu\_1, Sigma\_1)

Arguements:

z: observed 2-dimensional z-values

A: transition matrix

mu\_0: mean vector under the null

Sigma\_0: covariance matrix under the null

mu\_1: mean vector under the non-null

Sigma\_1: covariance matrix under the non-null

Values:

bw: rescaled backward variables

fw: rescaled forward variables

mLIS: mvLIS statistics

3. em1\_mvHMM

Description:

Calculating the parameters of multivariate hidden Markov models (L=1) via the EM algrithm.

Usage:

em1\_mvHMM (z, maxiter=200)

Arguements:

z: observed 2-dimensional z-values

maxiter: the maximum number of iterations

Values:

pii: the estimation of initial state distribution

A: the estimation of transition matrix

mu\_1: the estimation of mean vector under the non-null

Sigma\_1: the estimation of covariance matrix under the non-null

niter: the number of iterations

4. rdata\_mvHMM

Description:

Generating the observed 2-dimensional z-values and the states of hypotheses that are based on multivariate hidden Markov models (L>1).

Usage:

rdata\_mvHMM (NUM, pii, A, mu\_0, Sigma\_0, pc, mu\_1, Sigma\_1)

Arguements:

NUM: the number of hypotheses

pii=(pii[1], pii[2]): initial state distribution

A: transition matrix

mu\_0: mean vector under the null

Sigma\_0: covariance matrix under the null

pc: proportions of mixture components

mu\_1: mean vector under the non-null

Sigma\_1: covariance matrix under the non-null

Values:

o: observed 2-dimensional z-values

s: underlying states of hypotheses

5. bwfw\_mvHMM.R

Description:

Calculating the mvLIS statistics via the backward-forward algorithm (L>1).

Usage:

bwfw\_mvHMM (z, A, mu\_0, Sigma\_0, pc, mu\_1, Sigma\_1)

Arguements:

z: observed 2-dimensional z-values

A: transition matrix

mu\_0: mean vector under the null

Sigma\_0: covariance matrix under the null

pc: proportions of mixture components

mu\_1: mean vector under the non-null

Sigma\_1: covariance matrix under the non-null

Values:

bw: rescaled backward variables

fw: rescaled forward variables

mLIS: mvLIS statistics

6. em\_mvHMM

Description:

Calculating the parameters of multivariate hidden Markov models (L>1) via the EM algrithm.

Usage:

em\_mvHMM (z, L=2, maxiter=200)

Arguements:

z: observed 2-dimensional z-values

L: the number of mixture components

maxiter: the maximum number of iterations

Values:

pii: the estimation of initial state distribution

A: the estimation of transition matrix

pc: the estimation of proportions of mixture components

mu\_1: the estimation of mean vector under the non-null

Sigma\_1: the estimation of covariance matrix under the non-null

niter: the number of iterations

7. mt.hmm

Description:

Conducting mvLIS procedure when a pre-specified nominal level is given. See Sun W., and Cai T. (2009).

Usage:

mt.hmm(mvLIS, q)

Arguements:

mvLIS: mvLIS statistics

q: the pre-specified nominal level

Values:

nr: the number of rejected hypotheses

th: the threshold

re: the rejected hypotheses

ac: the accepted hypotheses

de: the decision rule

**Examples:**

## the number of observed z-values

NUM<-5000

## initialize the transition matrx

A<-matrix(c(0.9, 0.1, 0.2, 0.8), 2, 2, byrow=TRUE)

## initialize parameter set of the null and non-null distributions

mu\_0<-c(0, 0)

Sigma\_0<-matrix(c(1, 0, 0, 1), 2, 2, byrow=TRUE)

pc<-c(0.5, 0.5)

mu\_1<-matrix(c(2, 3, 1, 2), nrow = 2, ncol = 2, byrow = TRUE)

Sigma\_1<-array(c(1, 0.5, 0.5, 1, 1, 0.5, 0.5, 1), dim=c(2, 2, 2))

## initialize state distribution

pii<-c(0, 1)

## Generating the observed z-valuse and the states of hypotheses that are based on

## multivariate hidden Markov models.

rdata<-rdata\_mvHMM(NUM, pii, A, mu\_0, Sigma\_0, pc, mu\_1, Sigma\_1)

z<-rdata$o

theta<-rdata$s

## Calculating the repLIS multiple testing statistics via the backward-forward

## algorithm.

bwfw.res<-bwfw\_mvHMM(z, A, mu\_0, Sigma\_0, pc, mu\_1, Sigma\_1)

mvLIS <-bwfw.or$mLIS

## Conducting repLIS procedure given the pre-specified level is 0.1.

res.mvLIS<-mt.hmm(mvLIS,0.1)$de

N10<-length(which(res. mvLIS-theta>0))

R<-length(which(res. mvLIS==1))+0.0001

FDR<-N10/R