**Welcome to SMLIS - Large-Scale Dependent Multiple Testing via**

**Hidden Semi-Markov Models**

Pengfei Wang, and Jiangzhou Wang \*

January 17, 2022

**Description**

SMLIS is a program to carry out multiple testing via hidden Semi-Markov models, which is written in R code. Here SMLIS program models the temporal or one-dimensional sequential dependence structure among tests via a Semi-Markov chain.

**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. Wang P., Wang J. Large-scale dependent multiple testing via hidden semi-Markov models.

**Download**s

* Some core code of mvLIS procedure are available on GitHub (https://github.com/wpf19890429/Large-Scale-Dependent-Multiple-Testing-via-Hidden-Semi-Markov-Models). This repository contains the following files:

rdata.shifted.binomial.hsmm.R

rdata.shifted.Poisson.hsmm.R

rdata.shifted.nbinomial.hsmm.R

rdata.unstructed.geom.hsmm.R

bwfw.hsmm2hmm.R

em.hsmm2hmm.R

bwfw.hmm.multistates.R

mt.hmm.R.txt

**illustrations of core R functions**

1. rdata.shifted.binomial.hsmm.R

Description:

Generating the underlying states and observations for HSMM, where the dwell time distribution for state 0 is a shifted binomial distribution and the dwell time distribution for state 1 is a geometric distribution.

Usage:

rdata.binomial.hsmm(NUM, Pi, prob, f0, ...)

par\_dwell=list(n1, Pi)

Arguments

NUM: number of hypotheses

par\_dwell: the parameters of the shifted binomial distribution

pro: the parameter of the geometric distribution

Omega=(0, 1; 1, 0) is known

pc=(pc[1], ..., pc[L]): proportion of mixture components

f0: parameter of the null distribution

f1: parameter of the non-null mixture distribution

Values

o: continuous observed data

s: binary unobserved states

2. rdata.shifted.Poisson.hsmm.R

Description:

Generating the underlying states and observations for HSMM, where the dwell time distribution for state 0 is a shifted Poisson distribution and the dwell time distribution for state 1 is a geometric distribution.

Usage:

rdata.shifted.Poisson.hsmm(NUM, lambda, prob, f0, ...)

Arguments

NUM: number of hypotheses

par\_dwell: the parameter of the shifted Poisson distribution

pro: the parameter of the geometric distribution

Omega=(0, 1; 1, 0) is known

pc=(pc[1], ..., pc[L]): proportion of mixture components

f0: parameter of the null distribution

f1: parameter of the non-null mixture distribution

Values

o: continuous observed data

s: binary unobserved states

3. rdata.shifted.nbinomial.hsmm.R

Description:

Generates the underlying states and observations for HSMM, where the dwell time distribution for state 0 is a shifted negative binomial distribution and the dwell time distribution for state 1 is a geometric distribution.

Usage:

rdata.shifted.nbinomial.hsmm(NUM, n, Pi, prob, f0, ...)

Arguments:

NUM: number of hypotheses

par\_dwell: the parameters of the shifted negative binomial distribution

par\_dwell=list(n1, Pi)

pro: the parameter of the geometric distribution

Omega=(0, 1; 1, 0) is known

pc=(pc[1], ..., pc[L]): proportion of mixture components

f0: parameter of the null distribution

f1: parameter of the non-null mixture distribution

Values

o: continuous observed data

s: binary unobserved states

4. rdata.unstructed.geom.hsmm.R

Description:

Generating the underlying states and observations for HSMM, where the dwell time distribution for state 0 is a unstructed start and geometric tail distribution and the dwell time distribution for state 1 is a geometric distribution.

Usage:

rdata.unstructed.geom.hsmm(NUM, dm1, prob, f0, ...)

Arguments:

NUM: number of hypotheses

par\_dwell: par\_dwell=list(dm1)

dm1: the p.m.f. of the dwell time distribution of state 0

pro: the parameter of the geometric distribution

Omega=(0, 1; 1, 0) is known

pc=(pc[1], ..., pc[L]): proportion of mixture components

f0: parameter of the null distribution

f1: parameter of the non-null mixture distribution

Values

o: continuous observed data

s: binary unobserved states

5. bwfw.hsmm2hmm.R

Description:

Calculating values for backward, forward variables, LIS variables and etc by using the HMM with expanded state space to approximate the HSMM.

Usage:

bwfw.hsmm2hmm(x, dm, f0, pc, f1)

Arguments:

x=(x[1], ..., x[m]): the observed data

dm=list(dm[[1]], dm[[2]])

dm[[1]]: the p.m.f. of the dwell time distribution of state 0

dm[[2]]: the p.m.f. of the geometric distribution dm[[2]][r]=prob\*(1-prob)^(r-1)

f0=(mu\_0, sd\_0): the parameters for null distribution

pc=(pc[1], ..., pc[L]): proportion of mixture components

f1=(mu[1], sd[1]\\...\\ mu[L], sd[L]): the parameters for the non-null distribution

Values:

alpha: rescaled backward variables

beta: rescaled forward variables

LIS\_HSMM: the LIS variables

c0: scaling variables

6. em.hsmm2hmm.R

Description:

Calculating the MLE for an HMM which can closely approximate the HSMM. The distribution of state 0 is assumed to be known as N(0, 1) and the distribution of state 1 is assumed to be a normal mixture with L components.

Usage:

em.hsmm2hmm(x, m1, ...)

Arguments:

x=(x[1], ..., x[m]): the observed data

m1: the number of the state aggregates for state 0

m2: the number of the state aggregates for state 1

L: the number of mixture components for the non-null distribution

maxiter: the maximum number of iterations

progress: the variable indicates whether to show progress

Values:

dm=list(dm[[1]], dm[[2]]): the estimated p.m.f. for the dwell time distribution of state 0 and state 1

pc=(pc[1], ..., pc[L]): the estimated proportion of mixture components

f1=(mu[1], sd[1]\\...\\ mu[L], sd[L]): the estimated parameters for the non-null distribution

1. bwfw.hmm.multistates.R

Description:

Calculating values for backward, forward variables, LIS variables and etc by using the HMM with expanded state space to approximate the HSMM.

Usage:

bwfw.hmm.multistates(x, Gamma, m1, m2=1, f0, pc, f1)

Arguments:

x=(x[1], ..., x[m]): the observed data

Gamma: transition matrix

m1: the number of the state aggregates for state 0

m2: the number of the state aggregates for state 1

f0=(mu\_0, sd\_0): the parameters for null distribution

pc=(pc[1], ..., pc[L]): proportion of mixture components

f1=(mu[1], sd[1]\\...\\ mu[L], sd[L]): the parameters for the non-null distribution

Values:

alpha: rescaled backward variables

beta: rescaled forward variables

LIS\_HSMM: the LIS variables

c0: scaling variables

8. mt.hmm

Description:

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

Usage:

mt.hmm(SMLIS, q)

Arguements:

SMLIS: SMLIS 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