

Welcome to Covariate-Adjusted Multiple Testing in Genome-Wide Association Studies via Factorial Hidden Markov Models

Tingting Cui, Pengfei Wang, and Wensheng Zhu*

June 18, 2019

Description

CALIS-via-FHMM is a program to perform covariate-adjusted multiple testing via factorial hidden Markov models, which is written in R code. Here, CALIS-via-FHMM program contains four main files:

function 'rdata.fhmm' (rdata.fhmm.R) is the factorial HMM data generator.

function 'bwfw.fhmm' (bwfw.fhmm.R) realizes the forward-backward procedure.

function 'em.fhmm' (em.fhmm.R) realizes the E-M algorithm.

function 'mt.fhmm' (mt.fhmm.R) realizes the new multiple testing procedures.

More detailed instructions and examples on how to use these functions are given below.

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, Zhu W. Covariate-adjusted multiple testing in genome-wide association studies via factorial hidden Markov models.

Downloads

- Some core code of CALIS-via-FHMM procedure are available on GitHub (<https://github.com/wszhustat/CALIS-via-FHMM>).

This document contains the following files:

rdata.fhmm.R

bwfw.fhmm.R

em.fhmm.R

mt.fhmm.R

Illustrations of core R functions

1. rdata.fhmm

Description:

Generating the observed values and the states of hypotheses that are based on factorial hidden Markov models.

Usage:

```
rdata.fhmm (NUM, pii, piii, A, B, f00, f01, f10, f11)
```

Arguments:

NUM: the number of observations

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

piii=(piii[1], piii[2]): initial state distribution for covariate effects

A=(a00, a01 \\\ a10, a11): transition matrix for main effects

B=(b00, b01 \\\ b10, b11): transition matrix for covariate effects

f00: parameter set for the null distribution

f01,f10,f11: parameter set for the non-null distribution values

Values:

x: continuous observed data

theta: binary unobserved states for main effects

2. bwfw.fhmm

Description:

Calculating the CALIS multiple testing statistics via the backward-forward algorithm.

Usage:

```
bwfw.fhmm (x, pii, piii, A, B, f00, f01, f10, f11)
```

Arguments:

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

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

piii=(piii[1], piii[2]): initial state distribution for covariate effects

A= (a00, a01 \\\ a10, a11): transition matrix for main effects

B= (b00, b01 \\\ b10, b11): transition matrix for covariate effects

f00: parameter set for the null distribution

f01,f10,f11: parameter set for the non-null distribution

Values:

bw: rescaled backward variables

fw: rescaled forward variables

lfdr: CALIS multiple testing statistics

pr: probabilities of hidden states

ts: rescaled transition variables

3. em.fhmm

Description:

Calculating the parameters of factorial hidden Markov models via the EM algorithm.

Usage:

em.fhmm (x, maxiter)

Arguments:

x: the observed data

maxiter: the maximum number of iterations

Values:

pii: the EM estimation of initial state distribution for main effects

piii: the EM estimation of initial state distribution for covariate effects

A= (a00, a01 \ a10, a11): the EM estimation of transition matrix for main effects

B= (b00, b01 \ b10, b11): the EM estimation of transition matrix for covariate effects

f01, f10, f11: the EM estimation of parameter set for the non-null distribution

lfdr: the lfdr variables

ni: number of iterations

4. mt.fhmm

Description:

Giving a multiple testing rule in a factorial hidden markov model that controls the FDR at level q , based on sequence of lfdr

Usage:

mt.fhmm (lfdr,q)

Arguments:

lfdr: local false discovery rate sequence

q: the FDR level

Values:

nr: the number of rejected hypotheses

th: the threshold

re: the rejected hypotheses

ac: the accepted hypotheses

de: the decision rule

Examples:

```
source("rdata.fhmm.R")
source("bwfw.fhmm.R")
source("em.fhmm.R")
source("mt.fhmm.R")
# the number of observations
NUM=3000
# the prespecified FDR level
q=0.10
#####
## Example 1: the factorial HMM data generator ##
#####
# the initial state distribution
pii=c(0.95,0.05)
piii=c(0.8,0.2)
# the transition matrix
A=matrix(c(0.95,0.05,0.1,0.9),2,2,byrow=T)
B=matrix(c(0.9,0.1,0.05,0.95),2,2,byrow=T)
# the null distribution
```

```

f00=c(0,1)
# the alternative distribution
f01=c(-1, 1)
f10=c(1, 1)
f11=c(3, 1)
# the factorial HMM data
set.seed(123456)
rdata1=rdata.fhmm(NUM, pii, piii, A, B, f00, f01, f10, f11)
# the observed values
x1=rdata1$x
# the unobserved states
theta1=rdata1$theta
#####
## Example 2: the forward-backward procedure ##
#####
x1=rdata1$x
fb.res1=bwfw.fhmm(x1, pii, piii, A, B, f00, f01, f10, f11)
# the backward variable
backward.var=fb.res1$bw
# the backward variable
forward.var=fb.res1$fw
# the oracle lfdr variable (calis.or)
calis.or.var=fb.res1$lfdr
#####
## Example 3: the E-M Algorithm for calculating parameters of the factorial HMM ##
#####
# the EM algorithm
em.res1=em.fhmm(x1, maxiter=100)
# the estimates for factorial HMM parameters
em.res1$A
em.res1$B

```

```

em.res1$f01
em.res1$f10
em.res1$f11
# the data-driven lfdr variables (calis.dd)
  em.res1$lfdr
# the number of iterations
  em.res1$ni
#####
## Example 4: The CALIS.or and CALIS.dd procedures ##
#####
## (4.a) the CALIS.or procedure
# the CALIS.or values
  calis.or=fb.res1$lfdr
  calis.or.pi=mt.fhmm(calis.or, q)
# the decision rule
  calis.or.de=calis.or.pi$de
## (4.b) the CALIS.dd procedure
# the CALIS.dd variables
  calis.dd=em.res1$lfdr
  calis.dd.pi=mt.fhmm(calis.dd, q)
# the decision rule
  calis.dd.de=calis.dd.pi$de

#####
### Example 5: the analysis of the Bipolar Disorder data ####
#####
## calculate the plug-in CALIS statistic from the ten chromosomes
  chr.set=c(2,3,4,6,8,9,14,16,20,22)
  for(i in chr.set)
  {
    BD.data.chr.i=read.table(paste("BD_data_chr",i,".txt",sep=""))

```

```

    BD.data.chr.i=as.matrix(BD.data.chr.i)
    BD.chr.i.res=em.fhmm(BD.data.chr.i, maxiter=100)
    BD.chr.i.calis=BD.chr.i.res$lfd
    BD.chr.i.calis=as.matrix(BD.chr.i.calis)
    write.table(BD.chr.i.calis,paste("BD.chr",i,".calis.txt",sep=""))
  }

#####
## combine and rank the plug-in CALIS statistic from the ten chromosomes ##
#####
## combine the plug-in CALIS statistic from the ten chromosomes

    BD.chr2.calis=read.table("BD.chr2.calis.txt")
    BD.chr3.calis=read.table("BD.chr3.calis.txt")
    BD.chr4.calis=read.table("BD.chr4.calis.txt")
    BD.chr6.calis=read.table("BD.chr6.calis.txt")
    BD.chr8.calis=read.table("BD.chr8.calis.txt")
    BD.chr9.calis=read.table("BD.chr9.calis.txt")
    BD.chr14.calis=read.table("BD.chr14.calis.txt")
    BD.chr16.calis=read.table("BD.chr16.calis.txt")
    BD.chr20.calis=read.table("BD.chr20.calis.txt")
    BD.chr22.calis=read.table("BD.chr22.calis.txt")
    BD.calis.all=rbind(BD.chr2.calis,BD.chr3.calis,BD.chr4.calis,
                      BD.chr6.calis,BD.chr8.calis,BD.chr9.calis,
                      BD.chr14.calis,BD.chr16.calis,BD.chr20.calis,
                      BD.chr22.calis)

# the FDR level
    q=1e-07
# the CALIS procedure
    BD.calis.pi=mt.fhmm(BD.calis.all[,1], q)
# the threshold for calis
    BD.calis.pi$th

```

the number of rejections

BD.calis.pi\$nr

the indices of rejected hypotheses

BD.calis.pi\$re