Package 'TwoStage'

May 7, 2014

Type HMM-based FDR control for two-stage designs

Title Compute FDR under dependence using HMM from two-stage data

Version 1.0.0

Date 2014-03-28

Author Xiaoshan Wang <xwang@forsyth.org>

Maintainer Xiaoshan Wang <xwang@forsyth.org>

Depends R (>= 2.10.0)

Description The TwoStage package implements the Hidden Markov Model based FDR control as described in Wang et al., 2014 for multiple comparison adjustment under dependence.

License GPL (version 2 or later)

LazyLoad yes

R topics documented:

`TwoStage-package` *FDR control for multiple hypothesis tests of two-stage designs.*

Description

The TwoStage package implement hidden-Markov-model-based FDR control procedure as described in Wang et al (2014). It includes Zehetmayer's approach, and Sarkar's p value combination methods.

Details

Package:	TwoStage
Type:	Package
Version:	1.0
Date:	2014-03-28
License:	GPL (version 2 or later)
LazyLoad:	yes

Author(s)

Xiaoshan Wang <xwang at forsyth dot org>

References

X Wang, J. Starr, F. Jorge (2014) Two-stage design optimizes power and false discovery rates for microbiome research.

SK Sarkar, J. Chen, W. Guo (2013). Multiple Testing in a Two-Stage Adaptive Design with Combination Tests Controlling FDR. J. Am. Stat. Assoc. 108: 1385-1401.

W. Sun and T. Cai (2009). Large-scale multiple testing under dependence. J. R.Stat. Soc B. 71, 393-424.

`PowerZ` *Sample size estimation for Zehetmayer's approach.*

Description

PowerZ implement a search for optimal allocation of sample size in a two stage design planned to apply Zehetmayer's analysis approach.

Usage

```
PowerZ(m, pi0, alpha = 0.05, delta)
```

Arguments

`m`	The total number of hypotheses to be tested.
`r`	the ratio of sample size allocation between Stage 1 and Stage 2.
`alpha`	The overall alpha level. Default is 0.05.
`pi0`	The proportion of null hypotheses.
`delta`	The effect size between two groups.

Value

A vector includes Stage 1 threshold gamma1, total sample size n, Stage 1 proportion of sample size r, and the maximum power max.power.

Author(s)

Xiaoshan Wang

References

S. Zehetmayer, P Bauer, M. Posch (2005). Two-stage designs for experiments with a large number of hypotheses. Bioinformatics. 21, 3771-3777.

Examples

```
PowerZ(m=1000,alpha=0.05,pi0=0.95,delta=1.5)
```

---

sim *Data simulation for hidden Markov model.*

Description

Generate data based on predefined transition probability matrix.

Usage

```
sim(m, pii, A, f0, pc, f1, f2)
```

Arguments

m	The totmal number of hypothesis.
pii	Initial status. The default is pii = (0,1).
A	The transition probability matrix.
f0	The distribution to generate data of null hypotheses.
pc	The proportions for non-null components. For example, if we simulate 0.3N(1,1) + 0.3N(2,1) + 0.4N(3,1), then pc = (0.3, 0.3, 0.4).
f1	The distribution to generate data of non-null hypotheses for the first stage. It should be a vector for one-component non-null, and a matrix with means of non-null components on Column 1 and corresponding standard deviations on Colum 2.
f2	The distribution to generate data of non-null hypotheses for the second stage. It's struction should be the same as f1.

Value

s	The true status.
o1	Generated data for the first stage.
o2	Generated data for the second stage.

Examples

```
m = 1000
pii<-c(1, 0)
f0<-c(0, 1)
f1<-matrix(c(2, 1, 2.5, 1),ncol=2, byrow=T)
f2<-matrix(c(2, 1, 2.5, 1),ncol=2, byrow=T)
a11 = 0.8
A<-matrix(c(0.95, 0.05, 1- a11, a11), 2, 2, byrow=T)
pc = c(0.5, 0.5)

dt<-sim(m, pii, A, f0, pc, f1, f2)
```

TwoStageF *Fisher's combination function based analysis of two-stage data.*

Description

TwoStageF implements the Fisher's combination method proposed by Sarkar (2013), which combines two-stage p values by Fisher's combination function.

Usage

```
TwoStageF(p1, p2, alpha = 0.05, alpha.L0 = 0, alpha.U0 = 0.3)
```

Arguments

p1	Stage 1 p values.
p2	Stage 1 p values. Require to have the exact same order of Stage 1 p values, with NA replacing excluded tests by Stage 1 screening.
alpha	Overall FDR level.
alpha.L0	A FDR level set for declaring significant test in Stage 1. Default is 0. Its value must be smaller than overall FDR alpha.
alpha.U0	Stage 1 threshold to screen tests. p values between alpha.L0 and alpha.U0 will be re-tested in Stage 2.

Value

`Number of rejection and acceptance`
: Includes overall FDR alpha, Stage 1 FDR spending of alpha.L0, Stage 1 screening threshold alpha.U0, Stage 1 number of rejection R1, the number to be retested S1, Stage 1 number of acceptance A1,Stage 1 number of rejection R2, overall number of rejection R.

`List of rejected test`
: A list of the order of rejected tests.

Author(s)

Adapted from the code from J. Chen, the author of the reference paper.

References

SK Sarkar, J. Chen, W. Guo (2013). Multiple Testing in a Two-Stage Adaptive Design with Combination Tests Controlling FDR. J. Am. Stat. Assoc. 108: 1385-1401.

See Also

TowStageHMM,TowStageZ

Examples

```
p1 <- runif(100)
p2 <- ifelse(p1 < 0.2,runif(1,0,0.1),NA)
TwoStageF(p1, p2, alpha = 0.05, alpha.L0 = 0, alpha.U0 = 0.3)
```

TwoStageHMM *HMM-based FDR control for two-stage design based multiple hypothesis tests under dependence*

Description

The TwoStageHMM implements the hidden Markov model based FDR control as described in Wang et al., 2014 for two-stage design based multiple comparison adjustment under dependence.

Usage

```
TwoStageHMM(x1, s1keep, x2, L = 2, maxiter = 100, EM.const = "No", tol = 1e-04, alpha = 0.05)
```

Arguments

x1	A vector of the first stage observed data.
s1keep	the list of taxa or gene selected into the second stage
x2	A vector of the second stage observed data.
L	the number of components of the normal mixture of f1 in the hidden markov model. The default is 2.
maxiter	the maximum number of iteration.
EM.const	Choose unconstrained EM algorithm or constrained EM algorithm for HMM model. The default is "NO" for unconstrained EM algorithm.
tol	Tolerance for convergence. The default is 1e-10.
alpha	the nominal level of FDR. The Default is 0.05.

Value

nr	the total number of rejections among the retested hypotheses.
re	the list of rejected taxa/gene
ac	the list of accepted taxa/gene
de	test decisions of each taxa/gene. 1 = significantly differential between the case and control groups; 0 otherwise.

Author(s)

Xiaoshan Wang

References

X Wang, J. Starr, F. Jorge (2014) Two-stage design optimizes power and false discovery rates for microbiome research.

W. Sun and T. Cai (2009). Large-scale multiple testing under dependence. J. R.Stat. Soc B. 71, 393-424.

See Also

TwoStageZ

Examples

```
################
# Simulate data
################
NUM1 = 1000
pii<-c(1, 0)
f0<-c(0, 1)
f1<-c(2, 1)
f2 <-c(2.5,1)

a11 <- 0.8
A<-matrix(c(0.95, 0.05, 1- a11, a11), 2, 2, byrow=T)

dt1<-sim1(NUM1, pii, A, f0, f1, f2)

x1 <- dt1$o1
x2 <- dt1$o2

theta <- dt1$s
s1kp <- which(theta==1)
xs2 <- x2[s1kp]

two <- TwoStageHMM(x1, s1kp, xs2, L=2, maxiter=500, EM.const = 'No')
```

TwoStageS *Simes' combination function based analysis of two-stage data.*

Description

TwoStageS implements the Simes' combination method proposed by Sarkar (2013), which combines two-stage p values by Simes' combination function.

Usage

```
TwoStageF(p1, p2, alpha = 0.05, alpha.L0 = 0, alpha.U0 = 0.3)
```

Arguments

p1	Stage 1 p values.
p2	Stage 1 p values. Require to have the exact same order of Stage 1 p values, with NA replacing excluded tests by Stage 1 screening.
alpha	Overall FDR level.
alpha.L0	A FDR level set for declaring significant test in Stage 1. Default is 0. Its value must be smaller than overall FDR alpha.
alpha.U0	Stage 1 threshold to screen tests. p values between alpha.L0 and alpha.U0 will be re-tested in Stage 2.

Value

`Number of rejection and acceptance`
: Includes overall FDR alpha, Stage 1 FDR spending of alpha.L0, Stage 1 screening threshold alpha.U0, Stage 1 number of rejection R1, the number to be retested S1, Stage 1 number of acceptance A1,Stage 1 number of rejection R2, overall number of rejection R.

`List of rejected test`
: A list of the order of rejected tests.

Author(s)

Adapted from the codes of J. Chen, the author of the reference paper.

References

SK Sarkar, J. Chen, W. Guo (2013). Multiple Testing in a Two-Stage Adaptive Design with Combination Tests Controlling FDR. J. Am. Stat. Assoc. 108: 1385-1401.

See Also

TowStageHMM,TowStageZ

Examples

```
p1 <- runif(100)
p2 <- ifelse(p1 < 0.2,runif(1,0,0.1),NA)
TwoStageF(p1, p2, alpha = 0.05, alpha.L0 = 0, alpha.U0 = 0.3)
```

TwoStageZ *Zehetmayer's method to analyze data from a two-stage design.*

Description

TwoStageZ implements the method proposed by Zehetmayer (2005), which is named as Z method in this package.

Usage

```
TwoStageZ(x1, s1keep, x2, Vx1, Vx2, gamma1, n1, n2, lambda = 0.5, alpha = 0.05)
```

Arguments

x1	Observed data from the first stage.
s1keep	the list of taxa or gene selected into the second stage.
x2	Observed data from the second stage. The order of components, say genes or bacterial species, must be the same as the order used by the first stage data.
Vx1	Variance of first-stage measurements.
Vx2	Variance of second-stage measurements.
n1	The number of subjects in the first stage.
n2	The number of subjects in the second stage.
gamma1	P value threshold for the first stage selection of hypotheses into the second stange.
lambda	A constant chosen for calculating the null proportion. 0 < lambda < 1. The default is 0.5.
alpha	The overall level of FDR. The default is 0.05.

Value

pi0	estimated null proportion
gamma2	estimated second-stage p-value threshold
de	test decisions of each taxa/gene. 1 = significantly differential between the case and control groups; 0 otherwise

Author(s)

Xiaoshan Wang

References

S. Zehetmayer, P Bauer, M. Posch (2005). Two-stage designs for experiments with a large number of hypotheses. Bioinformatics. 21, 3771-3777.

See Also

TowStageHMM

Examples

```
NUM1 <- 1000
n1 <- 10
n2 <- 20
z  <- c(rep(0,900),rep(1,100))
Vx1 <- 1
Vx2 <- 1
x1 <- rnorm(NUM1, mean=z,sd=Vx1)

p1 <- 1 - pnorm(as.vector(x1)*sqrt(Vx1),0,1)
gamma1 <- 0.3
s1keep <- which(p1 < gamma1)

z2 <- z
z2[!s1keep] <- NA
```

```
TwoStageZ(x1,s1keep,x2,Vx1,Vx2,gamma1=0.3,n1,n2)
```

Index