

Bayesian Mixture Model for Multiple Tests

Description

This procedure estimates first the proportion of association markers based on a Bayesian mixture model (Wei et al. 2010) and uses the ordered p-values to select the significant ones. This algorithm also provides an evaluation on the individual marker level to assess by four different thresholds of Bayes Factor if any specific marker is "important."

Usage

```
Bmix<-function(no.allele, marker.name=c(1:no.allele), count.case, count.cntl,
freq.case, freq.cntl, pool.data, low.threshold=0.05, upp.threshold=0.95,
low.prior=0.1001252, upp.prior=4.129483, perce=0.1, accu=2.5*(10^(-6)),
datatype="frequency")
```

Arguments

no.allele	Number of markers. (required for allele frequency data)
marker.name	Vector of marker names or marker indexes.
count.case	Vector of sample sizes for the case group. (required for allele frequency data)
count.cntl	Vector of sample sizes for the control group. (required for allele frequency data)
freq.case	Vector of the minor allele frequencies of the case group. Each component must be in [0,0.5]. (required for allele frequency data)
freq.cntl	Vector of the minor allele frequencies of the control group. Each component must be in [0,0.5]. (required for allele frequency data)
pool.data	Vector of the t statistics based on unequal variances. (required for continuous data)
low.threshold	Lower bound of the frequency that is allowed. Must be in [0,0.5]. Default value is 0.05.

upp.threshold	Upper bound of the frequency that is allowed. Default value is 0.95.
low.prior	Lower bound for γ , denoted as t in the reference. Default value is 0.1001252.
upp.prior	Upper bound for γ , denoted as t in the reference. Default value is 4.129483.
pernt	The percentile of input statistics "pool.data" to be used as t in the lower bound of γ . Default value is the 10 th percentile.
accu	Accuracy of the global estimation of the significant proportion. Must be in $[10^{-7}, 1]$. Default value is $2.5 \cdot (10^{-6})$.
datatype	Either "frequency" or "continuous"; use the program for allele frequency data or microarray expression intensity data.

Details

If no options are selected, the values of γ 's prior is described in Wei et al. (2010).

```
aa<- list(BF=BF, Glb_Lambda = Glo_Lambda, Glb_Index = Glo_Ind, Glb_alleleIndex = Glo_allInd,
Glb.P_Value=cbind(P_Value_Sort_Value_G, P_Value_Sort_Index_G) , Ind_BF1_Lambda=
Ind_BF1_Lambda, Ind_BF1_Ind= Ind_BF1_Ind, Ind_BF1_allInd= Ind_BF1_allInd,Ind_BF3_Lambda=
Ind_BF3_Lambda, Ind_BF3_Ind= Ind_BF3_Ind, Ind_BF3_allInd= Ind_BF3_allInd,Ind_BF5_Lambda=
Ind_BF5_Lambda, Ind_BF5_Ind= Ind_BF5_Ind, Ind_BF5_allInd= Ind_BF5_allInd,Ind_BF10_Lambda=
Ind_BF10_Lambda, Ind_BF10_Ind= Ind_BF10_Ind, Ind_BF10_allInd=
Ind_BF10_allInd,SummaryMatrix= SummaryMatrix)
```

Value

A list containing:

no.allele.final	Allele number excluded allele frequency less than the lower bound.
Glb_Lambda	The global estimate of the proportion of significant markers.
Glb_Index	The allele index number of significant alleles selected by global Bayesian mixture model.

Glb_alleleIndex	The names (in marker.name) of significant alleles selected by global Bayesian mixture model.
Ind_BF1_Lambda	The proportion of significant markers by the threshold, log Bayes Factor 1.
Ind_BF1_Index	The allele index number of significant alleles selected by log Bayes Factor 1.
Ind_BF1_alleleIndex	The names (in marker.name) of significant alleles selected by log Bayes Factor 1.
Ind_BF3_Lambda	The proportion of significant markers by the threshold, log Bayes Factor 3.
Ind_BF3_Index	The allele index number of significant alleles selected by log Bayes Factor 3.
Ind_BF3_alleleIndex	The names (in marker.name) of significant alleles selected by log Bayes Factor 3.
Ind_BF5_Lambda	The proportion of significant markers by the threshold, log Bayes Factor 5.
Ind_BF5_Index	The allele index number of significant alleles selected by log Bayes Factor 5.
Ind_BF5_alleleIndex	The names (in marker.name) of significant alleles selected by log Bayes Factor 5.
Ind_BF10_Lambda	The proportion of significant markers by the threshold, log Bayes Factor 10.
Ind_BF10_Index	The allele index number of significant alleles selected by log Bayes Factor 10.
Ind_BF10_alleleIndex	The names (in marker.name) of significant alleles selected by log Bayes Factor 10.
All.P_Value	The p-values and the corresponding index numbers of markers.
BF	Bayes Factor of markers.
SummaryMatrix	Table of all markers information, included Number, marker.names, no.case, no.cntl, freq.case, freq.cntl, P_Value, BF, P_Value_label, BF_label. Significant alleles selected by global Bayesian mixture model shows "Glo" at P_Value_label column. Significant alleles selected by Bayes Factor show *,**,***,**** for log (BF) 1, 3, 5,10, respectively.

Author(s)

Yu-Chung Wei, Shu-Hui Wen, Pei-Chun Chen, Chih-Hao Wang and Chuhsing K Hsiao

References

Yu-Chung Wei, Shu-Hui Wen, Pei-Chun Chen, Chih-Hao Wang and Chuhsing K Hsiao
A simple Bayesian mixture model with a hybrid procedure for genome-wide
association studies. Eur J Hum Genet advance online publication, April 21, 2010;
doi:10.1038/ejhg.2010.51

Bmix Manual [http:// homepage.ntu.edu.tw/~ckhsiao/Bmix/manual.pdf](http://homepage.ntu.edu.tw/~ckhsiao/Bmix/manual.pdf)

Examples

#--Allele Frequency data

#--5 markers input the program. Case frequency of "rs00002" is less than low
threshold and would be deleted from the analysis. "rs00004" and "rs00005" are
significant markers in this example.

```
Bmix (no.allele=5, marker.name=c("rs00001", "rs00002", "rs00003", "rs00004",  
"rs00005"), count.case=c(500, 499, 493, 501, 498), count.cntl=c(492, 500, 493, 500,  
499), freq.case=c(0.06, 0.04, 0.2, 0.5, 0.3), freq.cntl=c(0.08, 0.3, 0.2, 0.1, 0.2) ,  
low.threshold=0.05, upp.threshold=0.5, low.prior=0.07, upp.prior=2.06,  
accu=10-5 , datatype= "frequency")
```

#--Continuous (expression intensity) data

```
Bmix (no.allele=5, marker.name =c("rs00001", "rs00002", "rs00003", "rs00004",  
"rs00005") , pool.data=c(0.04,2.35,4,1.7,0.9) ,accu=10-5),pernt=0.1,datatype=  
"continuous")
```