Package 'EBrank'

October 12, 2022

Title Empirical Bayes Ranking

Version 1.0.0

Description Empirical Bayes ranking applicable to parallel-estimation settings where the estimated parameters are asymptotically unbiased and normal, with known standard errors. A mixture normal prior for each parameter is estimated using Empirical Bayes methods, subsequentially ranks for each parameter are simulated from the resulting joint posterior over all parameters (The marginal posterior densities for each parameter are assumed independent). Finally, experiments are ordered by expected posterior rank, although computations minimizing other plausible rank-loss functions are also given.

Depends R (>= 3.2.4)

License CC0

Encoding UTF-8

LazyData true

RoxygenNote 5.0.1

NeedsCompilation no

Author John Ferguson [aut, cre]

Maintainer John Ferguson <john.ferguson@nuigalway.ie>

Repository CRAN

Date/Publication 2017-01-12 15:13:29

R topics documented:

EBrank *EBrank: Empirical Bayes Ranking*

Description

Empirical Bayes ranking applicable to parallel-estimation settings where the estimated parameters are asymptotically unbiased and normal, with known standard errors. A mixture normal prior for the parameter is estimated, subsequentially ranks for each parameter are simulated from the resulting posterior. Finally, experiments are ordered by expected posterior rank, although computations minimizing other plausible rank-loss functions are also given.

EBrank functions

rankEM

rankEM *Empirical Bayes parameter ranking for parallel estimation scenarios*

Description

Empirical Bayes ranking applicable to parallel-estimation settings where the estimated parameters are asymptotically unbiased and normal, with known standard errors. A mixture normal prior for each parameter is estimated using Empirical Bayes methods, subsequentially ranks for each parameter are simulated from the resulting joint posterior over all parameters (The marginal posterior densities for each parameter are assumed independent). Finally, experiments are ordered by expected posterior rank, although computations minimizing other plausible rank-loss functions are also given.

Usage

```
rankEM(betahat, sebeta, Jmin = 1, Jmax = 4, maxiter = 200, tol = 1e-05,
  nsim = 10000, cutoff = 0.5, maxpar = 40000, multiplestart = FALSE,
  sigmabig = 10, fixedcluster2 = TRUE, penfactor = 5000, fudge = 0.001,
  alpha = 0.05, FDR_BH = 0.05, topvec = c(10, 100, 1000, 10000))
```

Arguments

betahat	estimated effect sizes for each experiment
sebeta	standard error of estimated effect sizes
Jmin	minimum number of non-null clusters fit
Jmax	maximum number of non-null clusters fit
maxiter	maximum number of iterations for EM algorithm

tol	EM algorithm is considered to have converged if the sum of the squared Euclidean distances between the parameter estimates on 2 iterations is less than tol
nsim	number of simulations from posterior distribution
cutoff	controls which experiments are included for posterior rank simulation. If a numeric between 0 and 1, it specifies the minimum posterior probability for inclusion in posterior rank simulations. If equal to 'f' then experiements in posterior rank simulation had p-values that were significant according to a Benjamini Hochberg correction at BH_FDR, if equal to 'b' posterior simulations correspond to experiments with Bonferoni significant p-values at level alpha.
maxpar	maximum number of experiments to simulate
multiplestart	if TRUE, multiple start points are used for the EM-algorithm based fitting of the mixture normals (for a given number of clusters)
sigmabig	the standard deviation for the 1st non-null cluster component
fixedcluster2	TRUE if the standard deviation for the 1st non-null cluster of the marginal distribution is fixed at sigmabig and its mean is fixed at 0. If set to FALSE, the estimated mean and standard deviation of cluster 2 are free to vary.
penfactor	factor for dirichlet penalization for cluster probabilities at each step of the EM algorithm. The larger this is, the smaller the Dirichlet penalization
fudge	small constant added to cluster probabilies at each EM step to ensure stability
alpha	represents Bonferroni-corrected significance threshold when cutoff="b"
FDR_BH	represents FDR-corrected significance threshold when cutoff="f"
topvec	a vector representing values for K such that posterior probabilities that the parameter for each experiment is within the set of K parameters having the largest absolute values are given.

Value

A list of the top ranked experiments

Examples

```
truetheta <- c(rep(0,900),rnorm(100))
setheta <- pmax(rexp(1000,1),.1)
esttheta <- rnorm(length(truetheta),mean=truetheta,sd=setheta)
# just rank experiments that are significant at 5% FDR
stuff <- rankEM(esttheta,setheta,cutoff='f',FDR_BH=.05)
# rank all experiments (slower)
# stuff <- rankEM(esttheta,setheta,cutoff='f',FDR_BH=1)
```

Index