metaArray

November 11, 2009

R topics documented:

`em.draw` *Plot of transformed expression produced by EM algorithm*

Description

Given a numeric vector, a plot of four panels is drawn: 1) fitted mixture distribution 2) transformed expression against original expression 3) histogram of original expression 4) progression of log-likelihood during the fit

Usage

```
em.draw(vec, cl, threshold=0.0001)
#em.draw(vec, cl=1-metastasis, threshold=0.0001)
```

Arguments

`vec`	A numeric vector, especially a particular row of expression matrix
`cl`	A vector of 0s and 1s. Use 1 for normal phenotype and 0 for non-normal phenotype. If left blank, all samples will be labeled as normal phenotype. Normal component of mixture is estimated using samples with normal phenotype only. POE for samples with non-normal phenotype will be calculated after EM algorithm finishes ML estimation.
`threshold`	Crieterion for convergence in likelihood.

Value

A plot of four panels will appear upon the call.

Author(s)

Debashis Ghosh <ghoshd@umich.edu> , Hyungwon Choi <hwchoi@umich.edu>

`find.init` *Initialization of EM algorithm*

Description

This function is an automated initialization of 'z' in EM algorithm.

Usage

```
find.init(z, width = 1)
```

Arguments

`z`	Unobserved probability of membership to uniform component of the mixture.
`width`	Constant factor used when assigning 0/1 labels to samples. Larger width will result in more samples initialized at 0.

Value

`z`	A vector of 0/1, initial values of the EM algorithm

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

`fit.em` *Probability of expresison from mixture distribution for a single gene.*

Description

This core function fits two-component normal-uniform mixture distribution, and extracts probability of over/under expression for all samples in all genes.

Usage

```
fit.em(x, cl, threshold=1e-06)
```

Arguments

x	A numeric vector, especially expression values for a particular gene.
cl	A vector of 0s and 1s. Use 1 for normal phenotype and 0 for non-normal phenotype. Note that this is the opposite of POE MCMC. If all samples are of unknown phenotype or of the same one, give vector of zeros. When class information is provided, conditional estimation of the mixture is applied.
threshold	Criterion for convergence in likelihood

Value

expr	Estimated POE
a	Minimum (adjusted) of Raw Expression
b	Maximum (adjusted) of Raw Expression
sigmasq	Estimated variance of normal component
mu	Estimated mean of normal component
Pi	Probability that the gene is over/under expressed on average across the samples
lik.rec	Trajectory of likelihood during EM

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

---

intcor *Integrative Correlation Analysis*

Description

This function calculates gene-specific reproducibility score based on Parmigiani et al. R implementation in MergeMaid (pairwise.cor) is a very efficient function in that it calculates the correlation matrix only once and collect appropriate elements for calculation of scores for each gene. However, in case there are more than thousands of common genes across datasets, the correlation matrix may overflow memory cells alloted to a session of R. Therefore, a replacement to the function that remedies the storage problem by brute force but fast computation in C is provided here.

Usage

```
intcor(merged)
```

Arguments

merged	mergeExprSet object that contains gene expression and class label with all datasets.

Value

avg.cor	A vector of gene-specific integrative correlation score
pair.cor	A matrix of correlations for each gene in every pair of two studies

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

References

Clinical Cancer Research, Parmigiani et al. 10(9):2922-2927, 2004

Examples

```
#intcor(merged)
```

---

logit *Logit transform*

Description

Calculates logit of a number

Usage

```
logit(p)
```

Arguments

p	Success probability

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

---

mdata *metaArray sample dataset*

Description

Three datasets from liver, lung, and prostate cancer microarrays. Please refer to the bibliography in the vignette. Chen (30 primary, 9 metastatic), Garber (30 primary, 6 metastatic), Lapointe (30 primary, 9 metastatic)

Usage

```
data(mdata)
```

poe.em *Probability of Expression from mixture distribution for multiple genes.*

Description

This function applies fit.em function to all rows of a gene expression data matrix.

Usage

```
poe.em(mat, cl, threshold=1e-05, every = 100)
```

Arguments

mat	Gene expression data matrix
cl	A vector of 0s and 1s. Use 1 for normal phenotype and 0 for non-normal phenotype. If all samples are of unknown phenotype, give vectors of 0.
threshold	Criterion of convergence in likelihood
every	Progress of estimation is reported at every integer mode of the value 'every'

Value

A data matrix of transformed expression will result.

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

poe.mcmc *Probability of Expression (POE)*

Description

Differential expression using latent categories of up/down regulation. Three component normal-uniform mixture model under Bayesian hierarchical analysis. This is a C implementation of poe.fit function of POE package (MCMC). R portion of source code was directly adapted from POE; poe.one.iteration function was re-written in C. Some of the optional arguments available in poe.fit are suppressed here, and you cannot save the chain of samples drawn for numerical integration.

Usage

```
poe.mcmc(AA, NN = NULL, id = NULL, M = 2000, kap.min=3.0,
    logdata=FALSE, stepsize=0.5,
    centersample = FALSE, centergene = FALSE, generatestarts = TRUE, start.met
    startobject = R0, collapse.to.two = FALSE, burnin=200,
    collapse.window=50, converge.threshold=0.01,
    PR = list(alpha.mm = 0, alpha.sd = 100, mu.mm = 0, mu.sd = 100,
        pipos.mm = 0, pipos.sd = 100, pineg.mm = 0, pineg.sd = 100,
        kap.pri.rate = 1, tausqinv.aa = 1, tausqinv.bb = 0.1))
# poe.mcmc(AA = chen, NN = 1 - chen.spl$metastasis, M=2000)
```

Arguments

AA	Matrix or exprs from an ExpressionSet object.
NN	Phenotypic label of arrays. If provided, all genes from arrays with label 1 are forced e=0 in all iterations. A gene with e=0 is involved in sampling of mu and alpha. For arrays with label 0, e is assumed to be unknown, thus sampled differently at every iteration. If this slot is left NULL, all arrays will be marked 0, so that the effect of the latter above applies to all arrays.
M	Number of MCMC iterations after burn-in.

Value

poe	Probability of over/under expression. Transformed gene expression on a fixed scale of [-1,1].
Other values	Posterior median estimates of parameters. Please refer to POE package for details.

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

References

G. Parmigiani et al, JRSS, 64:717-736, 2002 or URL: http://astor.som.jhmi.edu/poe/

Examples

```
# poe.mat <- poe.mcmc(AA=exprmat, NN=clvec, M=10000)
# One can also provide different hyperparameter values.
```

Zscore *Meta-analysis of Microarray Data from Different Platforms*

Description

This function calculates Z-score for each matched gene across all datasets. In each dataset, it performs local regression smoothing of mean vs variance. Z score is constructed by taking the ratio of weighted mean difference and combined standard deviation according to Box and Tiao (1992).

Usage

```
Zscore(merged, pheno = NULL, permute = 0, verbose = TRUE)
```

Arguments

merged	mergeExprSet object that contains gene expression and class label with all datasets. Class label should consist of two unique elements. If pheno is NULL, first columns of phenoData from each ExpressionSet is sought as class labels. If a vector of particular column number in each data is specified, corresponding columns of phenoData will be considered for class labels.

pheno	A numeric vector specifying the location of class labels in phenoData from each ExpressionSet, a unit of mergeExprSet representing one dataset.
permute	If permute is 0, weighted Z-score will be referenced to standard normal distribution for two-sided p-value. Otherwise, columns of all datasets (each dataset separately) will be shuffled at random, from which a permutation distribution of Z-scores are formed and Z-scores are referenced to this distribution.
verbose	If verbose is TRUE, the progress of permutation will be reported.

Value

A data.frame with matched genes, Z-scores and p-values will result.

Author(s)

Debashis Ghosh <ghoshd@umich.edu>, Hyungwon Choi <hwchoi@umich.edu>

References

J.Wang et al, Bioinformatics 2004 Nov 22;20(17):3166-78

Examples

```
# Zscore(merged, pheno=NULL, permute=10000, verbose=FALSE)
```

Index