Package ‘EbayesThresh’

April 17, 2009

Title Empirical Bayes thresholding and related methods

Version 1.3.0

Date 2005-03-24

Author Bernard W. Silverman <bernard.silverman@spc.ox.ac.uk>

Maintainer Bernard W. Silverman <bernard.silverman@spc.ox.ac.uk>

Description This package carries out Empirical Bayes thresholding using the methods developed by I. M. Johnstone and B. W. Silverman. The basic problem is to estimate a mean vector given a vector of observations of the mean vector plus white noise, taking advantage of possible sparsity in the mean vector. Within a Bayesian formulation, the elements of the mean vector are modelled as having, independently, a distribution that is a mixture of an atom of probability at zero and a suitable hevay-tailed distribution. The mixing parameter can be estimated by a marginal maximum likelihood approach. This leads to an adaptive thresholding approach on the original data. Extensions of the basic method, in particular to wavelet thresholding, are also implemented within the package.

URL http://www.bernardsilverman.com

Repository CRAN

Date/Publication 2005-03-26 21:42:31

R topics documented:

`beta.cauchy` *Function beta for the quasi-Cauchy prior*

Description

Given a value or vector x of values, find the value(s) of the function $\beta(x) = g(x)/\phi(x) - 1$, where g is the convolution of the quasi-Cauchy with the normal density $\phi(x)$.

Usage

```
beta.cauchy(x)
```

Arguments

`x`	a real value or vector

Value

A vector the same length as x, containing the value(s) $\beta(x)$.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`beta.laplace`

Examples

```
beta.cauchy(c(-2,1,0,-4,8,50))
```

beta.laplace *Function beta for the Laplace prior*

Description

Given a value or vector x of values, find the value(s) of the function $\beta(x) = g(x)/\phi(x) - 1$, where g is the convolution of the Laplace density with scale parameter a with with the normal density $\phi(x)$.

Usage

```
beta.laplace(x, a = 0.5)
```

Arguments

`x`	the value or vector of data values
`a`	the scale parameter of the Laplace distribution

Value

A vector the same length as `x` is returned, containing the value(s) $\beta(\mathrm{x})$.

Note

The Laplace density is given by $\gamma(u) = \frac{1}{2}ae^{-a|u|}$ and is also known as the double exponential density.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`beta.cauchy`

Examples

```
beta.laplace(c(-2,1,0,-4,8,50))
beta.laplace(c(-2,1,0,-4,8,50), a=1)
```

`ebayesthresh` *Empirical Bayes thresholding on a sequence*

Description

Given a sequence of data, performs Empirical Bayes thresholding, as discussed in Johnstone and Silverman (2004).

Usage

```
ebayesthresh(x, prior = "laplace", a = 0.5, bayesfac = FALSE,
       sdev = NA, verbose = FALSE, threshrule = "median")
```

Arguments

`x`	vector of data values
`prior`	specification of prior to be used conditional on the mean being nonzero; can be `cauchy` or `laplace`
`a`	scale factor if Laplace prior is used. Ignored if Cauchy prior is used. If, on entry, `a=NA` and `prior="laplace"`, then the scale parameter will also be estimated by marginal maximum likelihood. If `a` is not specified then the default value 0.5 will be used.
`bayesfac`	if `bayesfac=TRUE`, then whenever a threshold is explicitly calculated, the Bayes factor threshold will be used
`sdev`	the sampling standard deviation of the data `x`. If, on entry, `sdev=NA`, then the standard deviation will be estimated using the median absolute deviation from zero, as `mad(x, center=0)`.
`verbose`	controls the level of output. See below.
`threshrule`	specifies the thresholding rule to be applied to the data. Possible values are `median` (use the posterior median); `mean` (use the posterior mean); `hard` (carry out hard thresholding); `soft` (carry out soft thresholding); `none` (find various parameters, but do not carry out any thresholding).

Details

It is assumed that the data vector $(x_1, \ldots, x_n)$ is such that each x_i is drawn independently from a normal distribution with mean θ_i and variance σ^2. The prior distribution of each θ_i is a mixture with probability $1 - w$ of zero and probability w of a given symmetric heavy-tailed distribution. The mixing weight w is estimated by marginal maximum likelihood. Given the mixing weight, and possibly a scale factor in the symmetric distribution, are estimated by marginal maximum likelihood. The resulting values are used as the hyperparameters in the prior.

The parameters can be estimated as the posterior median or the posterior mean given the data, or by hard or soft thresholding using the posterior median threshold. If hard or soft thresholding is chosen, then there is the additional choice of using the Bayes factor threshold, which is the value such that the posterior probability of zero is exactly half if the data value is equal to the threshold.

Value

If `verbose=FALSE`, a vector giving the values of the estimates of the underlying mean vector.

If `verbose=TRUE`, a list with the following elements:

- `muhat` the estimated mean vector (omitted if `threshrule="none"`)
- `x` the data vector as supplied
- `threshold.sdevscale` the threshold as a multiple of the standard deviation `sdev`
- `threshold.origscale` the threshold measured on the original scale of the data
- `prior` the prior that was used
- `w` the mixing weight as estimated by marginal maximum likelihood
- `a` (only present if Laplace prior used) the scale factor as supplied or estimated
- `bayesfac` the value of the parameter `bayesfac`, determining whether Bayes factor or posterior median thresholds are used
- `sdev` the standard deviation of the data as supplied or estimated
- `threshrule` the thresholding rule used, as specified above

Author(s)

Bernard Silverman

References

Johnstone, I. M. and Silverman, B. W. (2004) Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Annals of Statistics*, **32**, 1594–1649.

Johnstone, I. M. and Silverman, B. W. (2004) EbayesThresh: R software for Empirical Bayes thresholding. *Journal of Statistical Software*. To appear.

Johnstone, I. M. (2004) 'Function Estimation and Classical Normal Theory' 'The Threshold Selection Problem'. The Wald Lectures I and II, 2004. Available from `http://www-stat.stanford.edu/~imj/`.

Johnstone, I. M. and Silverman, B. W. (2005) Empirical Bayes selection of wavelet thresholds. *Annals of Statistics*, **33**, to appear.

The papers by Johnstone and Silverman are available from `http://www.bernardsilverman.com`.

See also `http://www-stat.stanford.edu/~imj/` for further references, including the draft of a monograph by I. M. Johnstone.

See Also

`tfromx`, `threshld`

Examples

```
ebayesthresh(x=rnorm(100, c( rep(0,90), rep(5,10))), prior="cauchy", sdev=NA)
```

`ebayesthresh.wavelet` *Empirical Bayes thresholding on the levels of a wavelet transform*

Description

Apply an Empirical Bayes thresholding approach level by level to the detail coefficients in a wavelet transform.

Usage

```
ebayesthresh.wavelet(xtr, vscale = "independent", smooth.levels = Inf,
      prior = "laplace", a = 0.5, bayesfac = FALSE,
      threshrule = "median")
```

Arguments

- `xtr` The wavelet transform of a vector of data. The transform is obtained using one of the wavelet transform routines in R or in S+WAVELETS. Any choice of wavelet, boundary condition, etc provided by these routines can be used.
- `vscale` Controls the scale used at different levels of the transform. If `vscale` is a scalar quantity, then it will be assumed that the wavelet coefficients at every level have this standard deviation. If `vscale = "independent"`, the standard deviation will be estimated from the highest level of the wavelet transform and will then be used for all levels processed. If `vscale="level"`, then the standard deviation will be estimated separately for each level processed, allowing standard deviation that is level-dependent.
- `smooth.levels` the number of levels to be processed, if less than the number of levels of detail calculated by the wavelet transform.
- `prior` specification of prior to be used for the coefficients at each level, conditional on their mean being nonzero; can be `cauchy` or `laplace`
- `a` scale factor if Laplace prior is used. Ignored if Cauchy prior is used. If, on entry, `a=NA` and `prior="laplace"`, then the scale parameter will also be estimated at each level by marginal maximum likelihood. If `a` is not specified then the default value 0.5 will be used.
- `bayesfac` if `bayesfac=TRUE`, then whenever a threshold is explicitly calculated, the Bayes factor threshold will be used
- `threshrule` specifies the thresholding rule to be applied to the coefficients. Possible values are `median` (use the posterior median); `mean` (use the posterior mean); `hard` (carry out hard thresholding); `soft` (carry out soft thresholding);

Details

The routine `ebayesthresh.wavelet` can process a wavelet transform obtained using the routine `wd` in the WaveThresh R package, the routines `dwt` or `modwt` in the waveslim R package, or one of the routines (either dwt or nd.dwt) in S+WAVELETS. Note that the wavelet transform must be calculated before the routine ebayesthresh.wavelet is called; the choice of wavelet, boundary conditions, decimated vs nondecimated wavelet, and so on, are made when the wavelet transform is calculated. Apart from some housekeeping to estimate the standard deviation if necessary, and to determine the number of levels to be processed, the main part of the routine is a call, for each level, to the smoothing routine `ebayesthresh`. The basic notion of processing each level of detail coefficients is easily transferred to transforms constructed using other wavelet software. Similarly, it is straightforward to modify the routine to give other details of the wavelet transform, if necessary using the option `verbose = TRUE` in the calls to `ebayesthresh`.

The main routine `ebayesthresh.wavelet` calls the relevant one of the routines `ebayesthresh.wavelet.wd` (for a transform obtained from WaveThresh), `ebayesthresh.wavelet.dwt` (for transforms obtained from either `dwt` or `modwt` in waveslim) or `ebayesthresh.wavelet.splus` (for transforms obtained from S+WAVELETS.

Value

The wavelet transform (in the same format as that supplied to the routine) of the values of the estimated regression function underlying the original data.

Author(s)

Bernard Silverman

References

Johnstone, I. M. and Silverman, B. W. (2005) Empirical Bayes selection of wavelet thresholds. *Annals of Statistics*, **33**, to appear.

See also the other references given for `ebayesthresh` and at `http://www.bernardsilverman.com`

See Also

`ebayesthresh`

`postmean` *Posterior mean estimator*

Description

Given a data value or a vector of data, find the corresponding posterior mean estimate(s) of the underlying signal value(s)

Usage

```
postmean(x, w, prior = "laplace", a = 0.5)
```

Arguments

`x`	a data value or a vector of data
`w`	the value of the prior probability that the signal is nonzero
`prior`	family of the nonzero part of the prior; can be `"cauchy"` or `"laplace"`
`a`	the scale parameter of the nonzero part of the prior if the Laplace prior is used

Value

If x is a scalar, the posterior mean $E(\theta|x)$ where θ is the mean of the distribution from which x is drawn. If x is a vector with elements $x_1, ..., x_n$, then the vector returned has elements $E(\theta_i|x_i)$, where each x_i has mean θ_i, all with the given prior.

Note

If the quasicauchy prior is used, the argument `a` is ignored. If `prior="laplace"`, the routine calls `postmean.laplace`, which finds the posterior mean explicitly, as the product of the posterior probability that the parameter is nonzero and the posterior mean conditional on not being zero. If `prior="cauchy"`, the routine calls `postmean.cauchy`; in that case the posterior mean is found by expressing the quasi-Cauchy prior as a mixture: The mean conditional on the mixing parameter is found and is then averaged over the posterior distribution of the mixing parameter, including the atom of probability at zero variance.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`postmed`

Examples

```
postmean(c(-2,1,0,-4,8,50), w=0.05, prior="cauchy")
postmean(c(-2,1,0,-4,8,50), w=0.2, prior="laplace", a=0.3)
```

postmed *Posterior median estimator*

Description

Given a data value or a vector of data, find the corresponding posterior median estimate(s) of the underlying signal value(s)

Usage

```
postmed(x, w, prior = "laplace", a = 0.5)
```

Arguments

`x`	a data value or a vector of data
`w`	the value of the prior probability that the signal is nonzero
`prior`	family of the nonzero part of the prior; can be `"cauchy"` or `"laplace"`
`a`	the scale parameter of the nonzero part of the prior if the Laplace prior is used

Details

The routine calls the relevant one of the routines `postmed.laplace` or `postmed.cauchy`. In the Laplace case, the posterior median is found explicitly, without any need for the numerical solution of an equation. In the quasi-Cauchy case, the posterior median is found by finding the zero, component by component, of the vector function `cauchy.medzero`.

Value

If x is a scalar, the posterior median med$(\theta|x)$ where θ is the mean of the distribution from which x is drawn. If x is a vector with elements $x_1, ..., x_n$, then the vector returned has elements med$(\theta_i|x_i)$, where each x_i has mean θ_i, all with the given prior.

Note

If the quasicauchy prior is used, the argument `a` is ignored. The routine calls the approprate one of `postmed.laplace` or `postmed.cauchy`.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`postmean`

Examples

```
postmed(c(-2,1,0,-4,8,50), w=0.05, prior="cauchy")
postmed(c(-2,1,0,-4,8,50), w=0.2, prior="laplace", a=0.3)
```

`tfromw` *Find threshold from mixing weight*

Description

Given a weight or vector of weights (i.e. prior probabilities that the parameter is nonzero), find the corresponding threshold(s) under the specified prior.

Usage

```
tfromw(w, prior = "laplace", bayesfac = FALSE, a = 0.5)
```

Arguments

`w`	prior weight or vector of weights
`prior`	specification of prior to be used; can be `"cauchy"` or `"laplace"`
`bayesfac`	specifies whether Bayes factor threshold should be used instead of posterior median threshold
`a`	scale factor if Laplace prior is used. Ignored if Cauchy prior is used.

Details

The Bayes factor method uses a threshold such that the posterior probability of zero is exactly half if the data value is equal to the threshold. If `bayesfac` is set to `FALSE` (the default) then the threshold is that of the posterior median function given the data value.

The routine carries out a binary search over each component of an appropriate vector function, using the routine `vecbinsolv`. For the posterior median threshold, the function to be zeroed is `laplace.threshzero` or `cauchy.threshzero`. For the Bayes factor threshold, the corresponding functions are `beta.laplace` or `beta.cauchy`.

Value

The value or vector of values of the estimated threshold(s).

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

wfromx,tfromx,wandafromx

Examples

```
tfromw(c(0.05, 0.1))

tfromw(c(0.05, 0.1), prior="cauchy", bayesfac=TRUE)
```

---

tfromx *Find threshold from data*

Description

Given a vector of data, find the threshold corresponding to the marginal maximum likelihood choice of weight.

Usage

```
tfromx(x, prior = "laplace", bayesfac = FALSE, a = 0.5)
```

Arguments

`x`	vector of data
`prior`	specification of prior to be used; can be `"cauchy"` or `"laplace"`
`bayesfac`	specifies whether Bayes factor threshold should be used instead of posterior median threshold
`a`	scale factor if Laplace prior is used. Ignored if Cauchy prior is used.

Details

First, the routine `wfromx` is called to find the estimated weight. Then the routine `tfromw` is used to find the threshold. See the documentation for these routines for more details.

Value

The numerical value of the estimated threshold is returned.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and http://www.bernardsilverman.com

See Also

tfromw, wfromx

Examples

```
tfromx(x=rnorm(100, c( rep(0,90), rep(5,10))), prior="cauchy")
```

---

threshld *Threshold data with hard or soft thresholding*

Description

Given a data value or a vector of data, threshold the data at a specified value, using hard or soft thresholding

Usage

```
threshld(x, t, hard = TRUE)
```

Arguments

x	a data value or a vector of data
t	value of threshold to be used
hard	specifies whether hard or soft thresholding is applied

Value

A value or vector of values the same length as x, containing the result of the relevant thresholding rule applied to x.

Author(s)

Bernard Silverman

References

See ebayesthresh and http://www.bernardsilverman.com

See Also

ebayesthresh

Examples

```
threshld(-5:5, 1.4, FALSE)
```

wandafromx — *Find weight and scale factor from data if Laplace prior is used*

Description

Given a vector of data, find the marginal maximum likelihood choice of both weight and scale factor under the Laplace prior.

Usage

```
wandafromx(x)
```

Arguments

`x`	a vector of data

Details

The parameters are found by marginal maximum likelihood. The search is over weights corresponding to thresholds in the range $[0, \sqrt{2 \log n}]$, where n is the length of the data vector. The search uses a nonlinear optimization routine (`optim` in R) to minimize the negative log likelihood function `negloglik.laplace`. The range over which the scale factor is searched is (0.04, 3). For reasons of numerical stability within the optimization, the prior is parametrized internally by the threshold and the scale parameter.

Value

A list with values

`w`	The estimated weight
`a`	The estimated scale factor

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`wfromx`, `tfromw`

Examples

```
wandafromx(rnorm(100, c( rep(0,90), rep(5,10))))
```

`wfromt` *Mixing weight from posterior median threshold*

Description

Given a threshold value, find the mixing weight for which this is the threshold of the posterior median estimator. If a vector of threshold values is provided, the vector of corresponding weights is returned.

Usage

```
wfromt(tt, prior = "laplace", a = 0.5)
```

Arguments

`tt`	threshold value or vector of values
`prior`	specification of prior to be used; can be `"cauchy"` or `"laplace"`
`a`	scale factor if Laplace prior is used. Ignored if Cauchy prior is used.

Value

The numerical value or vector of values of the corresponding weight is returned.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`tfromw`

Examples

```
wfromt( c(2,3,5), prior="cauchy" )
```

`wfromx` *Find Empirical Bayes weight from data*

Description

Suppose the vector $(x_1, \ldots, x_n)$ is such that x_i is drawn independently from a normal distribution with mean θ_i and variance 1. The prior distribution of the θ_i is a mixture with probability $1 - w$ of zero and probability w of a given symmetric heavy-tailed distribution. This routine finds the marginal maximum likelihood estimate of the parameter w.

Usage

```
wfromx(x, prior = "laplace", a = 0.5)
```

Arguments

`x`	vector of data
`prior`	specification of prior to be used; can be `"cauchy"` or `"laplace"`
`a`	scale factor if Laplace prior is used. Ignored if Cauchy prior is used.

Details

The weight is found by marginal maximum likelihood. The search is over weights corresponding to thresholds in the range $[0, \sqrt{2 \log n}]$, where n is the length of the data vector.

The search is by binary search for a solution to the equation $S(w) = 0$, where S is the derivative of the log likelihood. The binary search is on a logarithmic scale in w.

If the Laplace prior is used, the scale parameter is fixed at the value given for `a`, and defaults to 0.5 if no value is provided. To estimate `a` as well as `w` by marginal maximum likelihood, use the routine `wandafromx`.

Value

The numerical value of the estimated weight.

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`wandafromx`, `tfromx`, `tfromw`, `wfromt`

Examples

```
wfromx(x=rnorm(100, c( rep(0,90), rep(5,10))), prior="cauchy")
```

---

wmonfromx — *Find monotone Empirical Bayes weights from data*

Description

Given a vector of data, find the marginal maximum likelihood choice of weight sequence subject to the constraints that the weights are monotone decreasing.

Usage

```
wmonfromx(xd, prior = "laplace", a = 0.5, tol = 1e-08, maxits = 20)
```

Arguments

`xd`	a vector of data
`prior`	specification of the prior to be used; can be `cauchy` or `laplace`
`a`	scale parameter in prior if `prior="laplace"`. Ignored if `prior="cauchy"`
`tol`	absolute tolerance to within which estimates are calculated
`maxits`	maximum number of weighted least squares iterations within the calculation

Details

The weights is found by marginal maximum likelihood. The search is over weights corresponding to thresholds in the range $[0, \sqrt{2 \log n}]$, where n is the length of the data vector.

An iterated least squares monotone regression algorithm is used to maximize the log likelihood. The weighted least squares monotone regression routine `isotone` is used.

To turn the weights into thresholds, use the routine `tfromw`; to process the data with these thresholds, use the routine `threshld`.

Value

The vector of estimated weights is returned

Author(s)

Bernard Silverman

References

See `ebayesthresh` and `http://www.bernardsilverman.com`

See Also

`wfromx`, `isotone`

`zetafromx` *Estimation of a parameter in the prior weight sequence in the EbayesThresh paradigm*

Description

Suppose a sequence of data has underlying mean vector with elements θ_i. Given the sequence of data, and a vector of scale factors `cs` and a lower limit `pilo`, this routine finds the marginal maximum likelihood estimate of the parameter `zeta` such that the prior probability of θ_i being nonzero is of the form `median(pilo, zeta*cs, 1)`.

Usage

```
zetafromx(xd, cs, pilo = NA, prior = "laplace", a = 0.5)
```

Arguments

`xd`	a vector of data
`cs`	a vector of scale factors, of the same length as `x`
`pilo`	the lower limit for the estimated weights. If `pilo=NA` it is calculated according to the sample size to be the weight corresponding to the universal threshold $\sqrt{2 \log n}$.
`prior`	specification of prior to be used conditional on the mean being nonzero; can be `cauchy` or `laplace`
`a`	scale factor if Laplace prior is used. Ignored if Cauchy prior is used. If, on entry, `a=NA` and `prior="laplace"`, then the scale parameter will also be estimated by marginal maximum likelihood. If `a` is not specified then the default value 0.5 will be used.

Details

An exact algorithm is used, based on splitting the range up for `\zeta` into subintervals over which no element of $\zeta *$ `cs` crosses either `pilo` or 1. Within each of these subintervals, the log likelihood is concave and its maximum can be found to arbitrary accuracy; first the derivatives at each end of the interval are checked to see if there is an internal maximum at all, and if there is this can be found by a binary search for a zero of the derivative. Finally, the maximum of all the local maxima over these subintervals is found.

Value

A list with the following elements

`zeta`	The value of `zeta` that yields the marginal maximum likelihood
`w`	The weights (prior probabilities of nonzero) yielded by this value of `zeta`
`cs`	The factors as supplied to the program
`pilo`	The lower bound on the weight, either as supplied or as calculated internally

Note

Once the maximizing zeta and corresponding weights have been found, the thresholds can be found using the program tfromw, and these can be used to process the original data using the routine threshld.

Author(s)

Bernard Silverman

References

See ebayesthresh and http://www.bernardsilverman.com

See Also

tfromw, threshld, wmonfromx, wfromx

Index