Package ‘REBayes’

March 22, 2022

Title Empirical Bayes Estimation and Inference

Description Kiefer-Wolfowitz maximum likelihood estimation for mixture models and some other density estimation and regression methods based on convex optimization. See Koenker and Gu (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, Journal of Statistical Software, 82, 1--26, <DOI:10.18637/jss.v082.i08>.

Version 2.51

Maintainer Roger Koenker <rkoenker@uiuc.edu>

Depends R (>= 2.10), Matrix

Imports methods, utils, lattice

Suggests Rmosek, knitr, digest

LazyData TRUE

Encoding UTF-8

VignetteBuilder knitr

SystemRequirements MOSEK (http://www.mosek.com) and MOSEK license.

License GPL (>= 2)

URL https://www.r-project.org

NeedsCompilation no

RoxygenNote 7.1.2

Author Roger Koenker [aut, cre],
Jiaying Gu [ctb],
Ivan Mizera [ctb]

Repository CRAN

Date/Publication 2022-03-22 18:20:02 UTC

R topics documented:

`B2mix` *Bivariate Binomial mixture estimation via Kiefer Wolfowitz MLE*

Description

Interior point solution of Kiefer-Wolfowitz NPMLE for mixture of bivariate binomials

Usage

```
B2mix(x, k, u = 40, v = 40, weights = NULL, ...)
```

Arguments

`x`	n by 2 matrix of counts of "successes" for binomial observations
`k`	n by 2 matrix of Number of trials for binomial observations
`u`	Grid Values for the mixing distribution defaults to equal spacing of length u on [eps, 1- eps], if u is scalar.
`v`	Grid Values for the mixing distribution defaults to equal spacing of length v on [eps, 1- eps], if v is scalar.
`weights`	replicate weights for x obervations, should sum to 1
`...`	Other arguments to be passed to KWDual to control optimization

Details

This function was inspired by a paper by Kline and Walters (2019) on evaluation of audit experiments for employment discrimination. An example of its usage is available with ‘demo(B2mix1)‘. There can be identification issues particularly when the numbers of trials are modest as described in Gu (2020). Caveat emptor! The predict method for B2mix objects will compute posterior means,

Value

An object of class density with components:

`u`	grid of evaluation points of the mixing density
`v`	grid of evaluation points of the mixing density
`y`	function values of the mixing density at x
`g`	estimates of the mixture density at the distinct data values
`logLik`	Log Likelihood value at the estimate
`dy`	Bayes rule estimates of binomial probabilities for distinct data values
`status`	exit code from the optimizer

Author(s)

R. Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist.* 27, (1956), 887-906.

Kline, P. and C. Walters, (2019) Audits as Evidence: Experiments, Ensembles and Enforcement, preprint.

Gu, J. (2020) A Collection of Notes on Binomial Mixtures, preprint.

See Also

'Bmix' for univariate binomial mixtures.

`bball` *U.S. Major League Batting Average Data: 2002-2012*

Description

Data frame consisting of the following variables:

Details

Data is aggregated into half seasons: so season indicates whether the observation is in the first or second half of the season of a given year. Only players who have more than 10 at bats in any half season are included, and only players who have more than three half seasons are represented. The transformed batting average is $arcsin(sqrt((H + 1/4)/(AB + 1/2)))$. Only regular seasons data are included. R programs to extract the data from the original sources are available on request.

- Name
- IdNum
- Year
- Halfseason
- Pitcher
- HA transformed batting average;
- AB at bats
- H hits
- BB walks
- YOB Year of Birth;
- age age of the player
- agesq age squared

Source

ESPN Website: https://www.espn.com/mlb/stats

References

Gu, Jiaying and Roger Koenker (2015) Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data, J. Applied Econometrics, forthcoming.

BDGLmix — *Efron Bayesian Deconvolution Estimator for Gaussian Mixtures*

Description

Efron (2016, 2019) penalized logspline density estimator for Gaussian mixture model g-modeling. Returns an object of class GLmix to facilitate prediction compatible with Kiefer-Wolfowitz GLmix estimation. In particular percentile confidence intervals can be constructed based on posterior quantiles. Assumes homoscedastic standard Gaussian noise, for the moment.

Usage

```
BDGLmix(y, T = 300, sigma = 1, df = 5, c0 = 0.1)
```

Arguments

`y`	Data: Sample Observations
`T`	Undata: Grid Values defaults equal spacing of with T bins, when T is a scalar
`sigma`	scale parameter of the Gaussian noise, may take vector value of length(y)
`df`	degrees of freedom of the natural spline basis
`c0`	penalty parameter for the Euclidean norm penalty.

Value

An object of class GLmix, density with components:

`x`	points of evaluation on the domain of the density
`y`	estimated function values at these points of the mixing density
`sigma`	returns a sigma = 1 for compatibility with GLmix

Author(s)

Adapted from a similar implementation in the R package deconvolveR of Narasimhan and Efron.

References

Efron, B. (2016) Empirical Bayes deconvolution estimates, Biometrika, 103, 1–20, Efron, B. (2019) Bayes, Oracle Bayes and Empirical Bayes, Statistical Science, 34, 177-201.

Bmix *Binomial mixture estimation via Kiefer Wolfowitz MLE*

Description

Interior point solution of Kiefer-Wolfowitz NPMLE for mixture of binomials

Usage

```
Bmix(x, k, v = 300, collapse = TRUE, weights = NULL, unique = FALSE, ...)
```

Arguments

x	Count of "successes" for binomial observations
k	Number of trials for binomial observations
v	Grid Values for the mixing distribution defaults to equal spacing of length v on [eps, 1- eps], if v is scalar.
collapse	Collapse observations into cell counts.
weights	replicate weights for x obervations, should sum to 1
unique	option to check unique of reported solution
...	Other arguments to be passed to KWDual to control optimization

Details

The predict method for `Bmix` objects will compute means, medians or modes of the posterior according to whether the `Loss` argument is 2, 1 or 0, or posterior quantiles if `Loss` is in (0,1). When the number of trials is small the NPMLE may be non-unique. This happens when there exists a vector v in the unit simplex of R^m such that $Av = f$ where $f = (n_0/n, ..., n_k/n)$ the observed frequencies, and A is the k by m matrix with typical element

$$C(k,x)p_j^x(1-p_j)^{k-x}.$$

If there exists such a solution, it follows that the maximal likelihood value is attained by any Ghat such that

$$p_j = \int C(k,j)p^j(1-p)^{k-j}dGhat(p) = n_j/n,$$

for j = 0, ... , k. There will be many such solutions, but by the Caratheodory theorem any one of them can be expressed as a linear combination of no more than k extreme points of the constraint set. In contrast, when there are no solutions inside the simplex satisfying the equation, then the NPMLE is the unique projection onto the boundary of that set. To facilitate checking this condition if the `check` parameter is `TRUE`, the linear program is feasible and the `unique` component is returned as `TRUE` if the program is infeasible, and `FALSE` is returned otherwise. This check is restricted to settings in which k is fixed, and `collapse` is `TRUE`. See Robbins (1956, p 161) for some further discussion of the binomial mixture model and a very clever alternative approach to prediction.

Value

An object of class density with components:

- xgrid midpoints of evaluation of the mixing density
- yfunction values of the mixing density at x
- gestimates of the mixture density at the distinct data values
- logLikLog Likelihood value at the estimate
- dyBayes rule estimates of binomial probabilities for distinct data values
- uniqueFlag indicating whether the solution is unique
- statusexit code from the optimizer

Author(s)

R. Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist*. 27, (1956), 887-906.

Koenker, R and I. Mizera, (2013) “Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules,” *JASA*, 109, 674–685.

Robbins, H. (1956) An Empirical Bayes Approach to Statistics, 3rd Berkeley Symposium.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

BPmix *Binomial mixtures with Poisson Trials via Kiefer Wolfowitz NPMLE*

Description

Interior point solution of Kiefer-Wolfowitz NPMLE for mixture of Poisson Binomials

Usage

```
BPmix(x, m, v = 50, weights = NULL, ...)
```

Arguments

x	Count of "successes" for binomial observations
m	Number of trials for binomial observations
v	Grid Values for the mixing distribution defaults to equal spacing of length v on [eps, 1- eps], if v is scalar.
weights	replicate weights for x obervations, should sum to 1
...	Other arguments to be passed to KWDual to control optimization

Details

The joint distribution of the probabilities of success and the number of trials is estimated. The grid specification for success probabilities is as for `Bmix` whereas the grid for the Poisson rate parameters is currently the support of the observed trials. There is no predict method as yet. See `demo(BPmix1)`.

Value

An object of class density with components:

`v`	grid points of evaluation of the success probabilities
`u`	grid points of evaluation of the Poisson rate for number of trials
`y`	function values of the mixing density at (v,u)
`g`	estimates of the mixture density at the distinct data values
`logLik`	Log Likelihood value at the estimate
`status`	exit code from the optimizer

Author(s)

R. Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist.* 27, (1956), 887-906.

`bwKW` *Bandwidth selection for KW smoothing*

Description

Bandwidth selection for KW smoothing

Usage

```
bwKW(g, k = 1, minbw = 0.1)
```

Arguments

`g`	KW fitted object
`k`	multiplicative fudge factor
`minbw`	minimum allowed value of bandwidth

Author(s)

R. Koenker

`bwKW2` *Bandwidth selection for bivariate KW smoothing*

Description

Bandwidth selection for bivariate KW smoothing

Usage

```
bwKW2(g, k = 1)
```

Arguments

`g`	bivariate KW fitted object
`k`	multiplicative fudge factor

Author(s)

R. Koenker

`Cosslett` *Kiefer-Wolfowitz estimator for Cosslett (1983) estimator*

Description

Kiefer-Wolfowitz-Cosslett estimator for binary response model.

Usage

```
Cosslett(x, y, v = 300, weights = NULL, ...)
```

Arguments

`x`	is the observed utility difference between two choices, it would be possible to extend this to make x a linear (index) function of some parameters
`y`	is the binary outcome
`v`	the unobserved utility difference taking values on a grid, by default this grid is equally spaced with 300 distinct points, however it is known that the mass points for the problem are located at the data points, x, so users may wish to set `v = sort(x)` although if the sample size is large this can be slow.
`weights`	replicate weights for x observations, should sum to 1
`...`	optional parameters to be passed to KWDual to control optimization

Details

In the primal form of the problem the pseudo log likelihood is:

$$l(f|y) = sum_i[y_i \log \sum_j (I(v_j <= x_i) * f_j) + (1 - y_i) \log \sum_j (I(v_j > x_i) * f_j)]$$

as usual the implementation used here solves the corresponding dual problem. Cumsum of the output y gives the CDF of the unobserved utility difference. See the `demo(Cosslett1)`and `demo(Cosslett2)` for illustrations without any covariate, and `demo(Cosslett3)` for an illustration with a covariate using profile likelihood. This model is also known as current status linear regression in the biostatistics literature, see e.g. Groeneboom and Hendrickx (2016) for recent results and references.

Value

an object of class density with the components:

`x`	points of evaluation of the mixing density
`y`	function values of the mixing density at x
`logL`	log likelihood of estimated model
`status`	exit code from the optimizer

Author(s)

Jiaying Gu and Roger Koenker

References

Kiefer, J. and J. Wolfowitz (1956) Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, *Ann. Math. Statist*, 27, 887-906.

Cosslett, S. (1983) Distribution Free Maximum Likelihood Estimator of the Binary Choice Model, *Econometrica*, 51, 765-782.

Groeneboom, P. and K. Hendrickx (2016) Current Status Linear Regression, preprint available from https://arxiv.org/abs/1601.00202.

`Finv` *Function inversion*

Description

Given a function, F(x, ...), and a scalar y, find x such that F(x, ...) = y. Note that there is no checking for the monotonicity of F wrt to x, or that the interval specified is appropriate to the problem. Such fine points are entirely the responsibility of the user/abuser. If the interval specified doesn't contain a root some automatic attempt to expand the interval will be made. Originally intended for use with F as `ThreshFDR`.

Usage

```
Finv(y, F, interval = c(0, 1), ...)
```

Arguments

`y`	the scalar at which to evaluate the inverse
`F`	the function
`interval`	the domain within which to begin looking
`...`	other arguments for the function F

Author(s)

R. Koenker

---

flies — *Medfly Data*

Description

Medfly data from the Carey et al (1992) experiment. There are 1,203,646 uncensored survival times!

Usage

```
flies
```

Format

A data frame with 19072 observations on the following 17 variables.

- `age`age at death in days
- `num`frequency count of age at death
- `prcurr`current proportion male
- `current`current density
- `cohort`cohort/pupal batch
- `size`pupal size
- `cage`cage number
- `female`female = 1
- `cumul`cumulative density
- `prcumu`cumulative proportion male
- `begin`initial cage density
- `prbegin`initial proportion mail
- `size4`size group 4

- `size5`size group 5
- `size6`size group 6
- `size7`size group 7
- `size8`size group 8

Details

Quoting from Carey et al (1992) "...Pupae were sorted into one of five size classes using a pupal sorter. This enabled size dimorphism to be eliminated as a potential source of sex-specific mortality differences. Approximately, 7,200 medflies (both sexes) of a given size class were maintained in each of 167 mesh covered, 15 cm by 60 cm by 90 cm aluminum cages. Adults were given a diet of sugar and water, ad libitum, and each day dead flies were removed, counted and their sex determined ..."

References

Carey, J.R., Liedo, P., Orozco, D. and Vaupel, J.W. (1992) Slowing of mortality rates at older ages in large Medfly cohorts, *Science,* 258, 457-61.

Koenker, R. and O. Geling (2001) Reappraising Medfly Longevity: A Quantile Regression Survival Analysis, *J. Am. Stat. Assoc*, 96, 458-468.

Koenker, R. and Jiaying Gu, (2013) "Frailty, Profile Likelihood and Medfly Mortality," *Contemporary Developments in Statistical Theory: A Festschrift for Hira Lal Koul*, S.N. Lahiri, A. Schick, Ashis Sengupta, and T.N. Sriram, (eds.), Springer.

`Gammamix` *NPMLE for Gamma Mixtures*

Description

A Kiefer-Wolfowitz MLE for Gamma mixture models

Usage

```
Gammamix(x, v = 300, shape = 1, weights = NULL, eps = 1e-10, ...)
```

Arguments

`x`	vector of observed variances
`v`	A vector of bin boundaries, if scalar then v equally spaced bins are constructed
`shape`	vector of shape parameters corresponding to x
`weights`	replicate weights for x obervations, should sum to 1
`eps`	tolerance for default gridding
`...`	optional parameters passed to KWDual to control optimization

Value

An object of class `density` with components:

`x`	midpoints of the bin boundaries
`y`	estimated function values of the mixing density
`g`	function values of the mixture density at the observed x's.
`logLik`	the value of the log likelihood at the solution
`dy`	Bayes rule estimates of
`status`	the Mosek convergence status.

Author(s)

J. Gu and R. Koenker

References

Gu J. and R. Koenker (2014) Unobserved heterogeneity in income dynamics: an empirical Bayes perspective, *JBES*, 35, 1-16.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

Gammamix for a general implementation for Gamma mixtures

`GLmix` *Kiefer-Wolfowitz NPMLE for Gaussian Location Mixtures*

Description

Kiefer Wolfowitz Nonparametric MLE for Gaussian Location Mixtures

Usage

```
GLmix(x, v = 300, sigma = 1, hist = FALSE, histm = 300, weights = NULL, ...)
```

Arguments

`x`	Data: Sample Observations
`v`	Undata: Grid Values defaults equal spacing of with v bins, when v is a scalar
`sigma`	scale parameter of the Gaussian noise, may take vector values of length(x)
`hist`	If TRUE then aggregate x to histogram bins, when sigma is vector valued this option is inappropriate unless there are only a small number of distinct sigma values.
`histm`	histogram bin boundaries, equally spacing with `histm` bins when scalar.
`weights`	replicate weights for x obervations, should sum to 1
`...`	other parameters to pass to KWDual to control optimization

Details

Kiefer Wolfowitz MLE as proposed by Jiang and Zhang for the Gaussian compound decision problem. The histogram option is intended for large problems, say n > 1000, where reducing the sample size dimension is desirable. When `sigma` is heterogeneous and `hist = TRUE` the procedure tries to do separate histogram binning for distinct values of `sigma`, however this is only feasible when there are only a small number of distinct `sigma`. By default the grid for the binning is equally spaced on the support of the data. This function does the normal convolution problem, for gamma mixtures of variances see `GVmix`, or for mixtures of both means and variances `TLVmix`.

The predict method for `GLmix` objects will compute means, medians or modes of the posterior according to whether the `Loss` argument is 2, 1 or 0, or posterior quantiles if `Loss` is in (0,1).

Value

An object of class density with components:

`x`	points of evaluation on the domain of the density
`y`	estimated function values at the points v, the mixing density
`g`	the estimated mixture density function values at x
`logLik`	Log likelihood value at the proposed solution
`dy`	prediction of mean parameters for each observed x value via Bayes Rule
`status`	exit code from the optimizer

Author(s)

Roger Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist*. Volume 27, Number 4 (1956), 887-906.

Jiang, Wenhua and Cun-Hui Zhang General maximum likelihood empirical Bayes estimation of normal means *Ann. Statist.*, Volume 37, Number 4 (2009), 1647-1684.

Koenker, R and I. Mizera, (2013) “Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules,” *JASA*, 109, 674–685.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

GLVmix *NPMLE of Gaussian Location-Scale Mixture Model*

Description

A Kiefer-Wolfowitz procedure for ML estimation of a Gaussian model with possibly dependent mean and variance components. This version differs from `WGLVmix` in that it doesn't assume the data is in longitudinal form. This version assumes a general bivariate distribution for the mixing distribution. The defaults use a rather coarse bivariate gridding.

Usage

```
GLVmix(t, s, m, u = 30, v = 30, ...)
```

Arguments

`t`	A vector of location estimates
`s`	A vector of variance estimates
`m`	A vector of sample sizes of the same length as t and s, or if scalar a common sample size length
`u`	A vector of bin boundaries for the location effects
`v`	A vector of bin boundaries for the variance effects
`...`	optional parameters to be passed to KWDual to control optimization

Value

A list consisting of the following components:

`u`	midpoints of mean bin boundaries
`v`	midpoints of variance bin boundaries
`fuv`	the function values of the mixing density.
`logLik`	log likelihood value for mean problem
`du`	Bayes rule estimate of the mixing density means.
`dv`	Bayes rule estimate of the mixing density variances.
`A`	Constraint matrix
`status`	Mosek convergence status

Author(s)

R. Koenker and J. Gu

References

Gu, J. and R. Koenker (2014) Heterogeneous Income Dynamics: An Empirical Bayes Perspective, *JBES*,35, 1-16.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

WTLVmix for an implementation assuming independent heterogeneity, and WGLVmix for a version that requires access to a full longitudinal data structure.

Gompertzmix *NPMLE for Gompertz Mixtures*

Description

Kiefer-Wolfowitz NPMLE for Gompertz Mixtures of scale parameter

Usage

```
Gompertzmix(
  x,
  v = 300,
  u = 300,
  alpha,
  theta,
  hist = FALSE,
  weights = NULL,
  ...
)
```

Arguments

x	Survival times
v	Grid values for mixing distribution
u	Grid values for mixing distribution
alpha	Shape parameter for Gompertz distribution
theta	Scale parameter for Gompertz Distribution
hist	If TRUE aggregate to histogram counts
weights	replicate weights for x obervations, should sum to 1
...	optional parameters passed to KWDual to control optimization

Details

Kiefer Wolfowitz NPMLE density estimation for Gompertz scale mixtures. The histogram option is intended for relatively large problems, say n > 1000, where reducing the sample size dimension is desirable. By default the grid for the binning is equally spaced on the support of the data. Parameterization: f(t|alpha,theta,v) = theta * exp(v) * exp(alpha * t) * exp(-(theta/alpha) * exp(v) * (exp(alpha*t)-1))

Value

An object of class density with components

`x`	points of evaluation on the domain of the density
`y`	estimated function values at the points x, the mixing density
`logLik`	Log likelihood value at the proposed solution
`dy`	Bayes rule estimates of theta at observed x
`status`	exit code from the optimizer

Author(s)

Roger Koenker and Jiaying Gu

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist.* Volume 27, Number 4 (1956), 887-906.

See Also

`Weibullmix`

`Gosset` *Gosset Criminal Finger Data*

Description

This data was generated by dithering the cell counts in the `crimtab` available in the base **stats** package.

Usage

```
Gosset
```

Format

A data frame with 3000 observations on 2 variables.

- `LMFinger`Length of Left Middle Finger (cm).
- `k`Height (cm)

Source

see the man page for `crimtab`

`Guvenen` *Annual Increments in Log Income*

Description

Kernel density estimates of the log density of annual increments in log income for U.S. individuals over the period 1994-2013, as estimated by *Guvenen*.

Usage

```
Guvenen
```

Format

A data frame with 279 observations on two variables.

- `earnings`annual increment in log income
- `logdensity`estimated log density values

Source

Fatih Guvenen, Fatih Karahan, Serdar Ozkan and Jae Song, (2016) What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics? https://www.nber.org/system/files/working_papers/w20913/w20913.pdf

---

`GVmix` *NPMLE for Gaussian Variance Heterogeneity*

Description

A Kiefer-Wolfowitz MLE for Gaussian models with independent variances. This can be viewed as a general form for χ^2 mixtures, see `Gammamix` for a more general form for Gamma mixtures.

Usage

```
GVmix(x, m, v = 300, weights = NULL, ...)
```

Arguments

`x`	vector of observed variances
`m`	vector of sample sizes corresponding to x
`v`	A vector of bin boundaries, if scalar then v equally spaced bins are constructed
`weights`	replicate weights for x obervations, should sum to 1
`...`	optional parameters passed to KWDual to control optimization

Value

An object of class `density` with components:

`x`	midpoints of the bin boundaries
`y`	estimated function values of the mixing density
`g`	function values of the mixture density at the observed x's.
`logLik`	the value of the log likelihood at the solution
`dy`	Bayes rule estimates of
`status`	the Mosek convergence status.

Author(s)

R. Koenker

References

Koenker, R and I. Mizera, (2013) "Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules," *JASA*, 109, 674–685.

Gu J. and R. Koenker (2014) Unobserved heterogeneity in income dynamics: an empirical Bayes perspective, *JBES*, 35, 1-16.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

Gammamix for a general implementation for Gamma mixtures

`KW2smooth` *Smooth a bivariate Kiefer-Wolfowitz NPMLE*

Description

Smooth a bivariate Kiefer-Wolfowitz NPMLE

Usage

```
KW2smooth(f, bw = NULL, k = 2)
```

Arguments

`f`	bivariate KW fitted object as from GLVmix
`bw`	bandwidth defaults to bwKW2(f),
`k`	kernel 2 for biweight, 3 for triweight

Author(s)

R. Koenker

KWDual *Dual optimization for Kiefer-Wolfowitz problems*

Description

Interface function for calls to optimizer from various REBayes functions There is currently only one option for the optimization that based on Mosek. It relies on the **Rmosek** interface to R see installation instructions in the Readme file in the inst directory of this package. This version of the function is intended to work with versions of Mosek after 7.0. A more experimental option employing the **pogs** package available from https://github.com/foges/pogs and employing an ADMM (Alternating Direction Method of Multipliers) approach has been deprecated, those interested could try installing version 1.4 of REBayes, and following the instructions provided there.

Usage

```
KWDual(A, d, w, ...)
```

Arguments

`A`	Linear constraint matrix
`d`	constraint vector
`w`	weights for `x` should sum to one.
`...`	other parameters passed to control optimization: These may include `rtol` the relative tolerance for dual gap convergence criterion, `verb` to control verbosity desired from mosek, `verb = 0` is quiet, `verb = 5` produces a fairly detailed iteration log, `control` is a control list consisting of sublists `iparam`, `dparam`, and `sparam`, containing elements of various mosek control parameters. See the Rmosek and Mosek manuals for further details. A prime example is `rtol` which should eventually be deprecated and folded into `control`, but will persist for a while for compatibility reasons. The default for `rtol` is 1e-6, but in some cases it is desirable to tighten this, say to 1e-10. Another example that motivated the introduction of `control` would be `control = list(iparam = list(num_threads = 1))`, which forces Mosek to use a single threaded process. The default allows Mosek to uses multiple threads (cores) if available, which is generally desirable, but may have unintended (undesirable) consequences when running simulations on clusters.

Value

Returns a list with components:

`f`	dual solution vector, the mixing density
`g`	primal solution vector, the mixture density evaluated at the data points
`logLik`	log likelihood
`status`	return status from Mosek

. Mosek termination messages are treated as warnings from an R perspective since solutions producing, for example, MSK_RES_TRM_STALL: The optimizer is terminated due to slow progress, may still provide a satisfactory solution, especially when the return status variable is "optimal".

Author(s)

R. Koenker

References

Koenker, R and I. Mizera, (2013) "Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules," *JASA*, 109, 674–685.

Mosek Aps (2015) Users Guide to the R-to-Mosek Optimization Interface, https://docs.mosek.com/8.1/rmosek/index.html.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

`KWPrimal` *Primal optimization for Kiefer-Wolfowitz problems*

Description

Interface function for calls to optimizer from various REBayes functions There is currently only one option for the optimization that based on Mosek. It relies on the **Rmosek** interface to R see installation instructions in the Readme file in the inst directory of this package. This version of the function works only with versions of Mosek 9.0. This is an experimental alternative to the main KWDual which is the usual interface from fitting functions to Mosek, caveat emptor..

Usage

```
KWPrimal(A, d, w, ...)
```

Arguments

`A`	Linear constraint matrix
`d`	constraint vector
`w`	weights for `x` should sum to one.
`...`	other parameters passed to control optimization: These may include `rtol` the relative tolerance for dual gap convergence criterion, `verb` to control verbosity desired from mosek, `verb = 0` is quiet, `verb = 5` produces a fairly detailed iteration log, `control` is a control list consisting of sublists `iparam`, `dparam`, and `sparam`, containing elements of various mosek control parameters. See the Rmosek and Mosek manuals for further details. A prime example is `rtol` which should eventually be deprecated and folded into `control`, but will persist for a while for compatibility reasons. The default for `rtol` is 1e-6, but in some cases it is desirable to tighten this, say to 1e-10. Another example

that motivated the introduction of `control` would be `control = list(iparam = list(num_threads = 1))`, which forces Mosek to use a single threaded process. The default allows Mosek to uses multiple threads (cores) if available, which is generally desirable, but may have unintended (undesirable) consequences when running simulations on clusters.

Value

Returns a list with components:

`f`	primal solution vector, the mixing density
`g`	the mixture density evaluated at the data points
`logLik`	log likelihood
`status`	return status from Mosek

. Mosek termination messages are treated as warnings from an R perspective since solutions producing, for example, MSK_RES_TRM_STALL: The optimizer is terminated due to slow progress, may still provide a satisfactory solution, especially when the return status variable is "optimal".

Author(s)

R. Koenker

References

Koenker, R and I. Mizera, (2013) "Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules," *JASA*, 109, 674–685.

Mosek Aps (2015) Users Guide to the R-to-Mosek Optimization Interface, https://docs.mosek.com/8.1/rmosek/index.html.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

`KWsmooth` *Smooth a Kiefer-Wolfowitz NPMLE*

Description

Smooth a Kiefer-Wolfowitz NPMLE

Usage

```
KWsmooth(f, bw = NULL, k = 2)
```

Arguments

`f`	KW fitted object
`bw`	bandwidth defaults to 2 * mad
`k`	kernel 2 for biweight, 3 for triweight

Author(s)

R. Koenker

L1norm *L1norm for piecewise linear functions*

Description

Intended to compute the L1norm of the difference between two distribution functions.

Usage

```
L1norm(F, G, eps = 1e-06)
```

Arguments

F	A stepfunction
G	Another stepfunction
eps	A tolerance parameter

Details

Both F and G should be of class `stepfun`, and they should be non-defective distribution functions. There are some tolerance issues in checking whether both functions are proper distribution functions at the extremes of their support. For simulations it may be prudent to wrap `L1norm` in `try`.

Value

A real number.

Author(s)

R. Koenker

Examples

```
# Make a random step (distribution) function with Gaussian knots
rstep <- function(n){
        x <- sort(rnorm(n))
        y <- runif(n)
        y <- c(0,cumsum(y/sum(y)))
        stepfun(x,y)
        }
F <- rstep(20)
G <- rstep(10)
S <- L1norm(F,G)
plot(F,main = paste("||F - G|| = ", round(S,4)))
lines(G,col = 2)
```

`Lfdr` *Local False Discovery Rate Estimation*

Description

A Generic function for estimation of Local FDR

Usage

```
Lfdr(G, ...)

## S3 method for class 'GLVmix'
Lfdr(G, newdata, cnull, tail = "R", ...)

## S3 method for class 'WGLVmix'
Lfdr(G, newdata, cnull, tail = "R", ...)

## S3 method for class 'GLmix'
Lfdr(G, newdata, cnull, tail = "R", ...)
```

Arguments

`G`	A fitted object from some G-modeling function.
`...`	other arguments
`newdata`	data frame to in which to evaluate Lfdr
`cnull`	threshold for evaluation of Lfdr
`tail`	either "R" or "L" to specify tail focus

Details

Given an estimated mixing distribution, G, Lfdr computes an estimated local false discovery rate at a specified set of points and threshold value cnull. The argument G can be specified as the fitted object from one of several possible fitting routines for nonparametric mixing distributions.

`medde` *Maximum Entropy [De]Regularized Density Estimation*

Description

Density estimation based on maximum entropy methods

Usage

```
medde(
  x,
  v = 300,
  lambda = 0.5,
  alpha = 1,
  Dorder = 1,
  w = NULL,
  mass = 1,
  rtol = 1e-06,
  verb = 0,
  control = NULL
)
```

Arguments

x	Data: either univariate or bivariate, the latter is highly experimental
v	Undata: either univariate or bivariate, univariate default is an equally spaced grid of 300 values, for bivariate data there is not (yet) a default.
lambda	total variation penalty smoothing parameter, if lambda is in [-1,0], a shape constraint is imposed. see Koenker and Mizera (2010) for further details. When Dorder = 0, the shape constraint imposes that the density is monotonically decreasing, when Dorder = 1 it imposes a concavity constraint.
alpha	Renyi entropy parameter characterizing fidelity criterion by default 1 is log-concave and 0.5 is Hellinger.
Dorder	Order of the derivative operator for the penalty default is Dorder = 1, corresponding to TV norm constraint on the first derivative, or a concavity constraint on some transform of the density. Dorder = 0 imposes a TV penalty on the function itself, or when lambda < 0 a monotonicity constraint.
w	weights associated with x,
mass	normalizing constant for fitted density,
rtol	Convergence tolerance for Mosek algorithm,
verb	Parameter controlling verbosity of solution, 0 for silent, 5 gives rather detailed iteration log.
control	Mosek control list see KWDual documentation

Details

See the references for further details. And also Mosek "Manuals". The acronym, according to the urban dictionary has a nice connection to a term used in Bahamian dialect, mostly on the Family Islands like Eleuthera and Cat Island meaning "mess with" "get involved," "get entangled," "fool around," "bother:" "I don't like to medder up with all kinda people" "Don't medder with people (chirren)" "Why you think she medderin up in their business."

This version implements a class of penalized density estimators solving:

$$\min_x \phi(x_1) | A_1 x_1 - A_2 x_2 = b, 0 \leq x_1, -\lambda \leq x_2 \leq \lambda$$

where x is a vector with two component subvectors: x_1 is a vector of function values of the density x_2 is a vector of dual values, λ is typically positive, and controls the fluctuation of the Dorder derivative of some transform of the density. When alpha = 1 this transform is simply the logarithm of the density, and Dorder = 1 yields a piecewise exponential estimate; when Dorder = 2 we obtain a variant of Silverman's (1982) estimator that shrinks the fitted density toward the Gaussian, i.e. with total variation of the second derivative of $log f$ equal to zero. See demo(Silverman) for an illustration of this case. If λ is in $(-1, 0]$ then the x_2 TV constraint is replaced by $x_2 \geq 0$, which for $\alpha = 1$, constrains the fitted density to be log-concave; for $\alpha = 0.5$, $-1/\sqrt{f}$ is constrained to be concave; and for $\alpha \leq 0$, $1/f^{\alpha-1}$ is constrained to be concave. In these cases no further regularization of the smoothness of density is required as the concavity constraint acts as regularizer. As explained further in Koenker and Mizera (2010) and Han and Wellner (2016) decreasing α constrains the fitted density to lie in a larger class of quasi-concave densities. See `demo(velo)` for an illustration of these options, but be aware that more extreme α pose more challenges from an numerical optimization perspective. Fitting for $\alpha < 1$ employs a fidelity criterion closely related to Renyi entropy that is more suitable than likelihood for very peaked, or very heavy tailed target densities. For $\lambda < 0$ fitting for `Dorder != 1` proceed at your own risk. A closely related problem is illustrated in the demo Brown which imposes a convexity constraint on $0.5x^2 + log f(x)$. This ensures that the resulting Bayes rule, aka Tweedie formula, is monotone in x, as described further in Koenker and Mizera (2013).

Value

An object of class "medde" with components

`x`	points of evaluation on the domain of the density
`y`	estimated function values at the evaluation points x
`status`	exit status from Mosek

Author(s)

Roger Koenker and Ivan Mizera

References

Chen, Y. and R.J. Samworth, (2013) "Smoothed log-concave maximum likelihood estimation with applications", *Statistica Sinica*, 23, 1373–1398.

Han, Qiyang and Jon Wellner (2016) "Approximation and estimation of s-concave densities via Renyi divergences, *Annals of Statistics*, 44, 1332-1359.

Koenker, R and I. Mizera, (2007) "Density Estimation by Total Variation Regularization," *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell Doksum*, V.N. Nair (ed.), 613-634.

Koenker, R and I. Mizera, (2006) "The alter egos of the regularized maximum likelihood density estimators: deregularized maximum-entropy, Shannon, Renyi, Simpson, Gini, and stretched strings," *Proceedings of the 7th Prague Symposium on Asymptotic Statistics*.

Koenker, R and I. Mizera, (2010) "Quasi-Concave Density Estimation" *Annals of Statistics*, 38, 2998-3027.

Koenker, R and I. Mizera, (2013) "Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules," JASA, 109, 674–685.

Koenker, R and I. Mizera, (2014) “Convex Optimization in R.”, *Journal of Statistical Software*, 60, 1-23.

See Also

This function is based on an earlier function of the same name in the deprecated package MeddeR that was based on an R-Matlab interface. A plotting method is available, or medde estimates can be added to plots with the usual `lines(meddefit,...` invocation. For log concave estimates there is also a quantile function `qmedde` and a random number generation function `rmedde`, eventually there should be corresponding functionality for other alphas.

Examples

```
## Not run:
#Maximum Likelihood Estimation of a Log-Concave Density
set.seed(1968)
x <- rgamma(50,10)
m <- medde(x, v = 50, lambda = -.5, verb = 5)
plot(m, type = "l", xlab = "x", ylab = "f(x)")
lines(m$x,dgamma(m$x,10),col = 2)
title("Log-concave Constraint")

## End(Not run)

## Not run:
#Maximum Likelihood Estimation of a Gamma Density with TV constraint
set.seed(1968)
x <- rgamma(50,5)
f <- medde(x, v = 50, lambda = 0.2, verb = 5)
plot(f, type = "l", xlab = "x", ylab = "f(x)")
lines(f$x,dgamma(f$x,5),col = 2)
legend(10,.15,c("ghat","true"),lty = 1, col = 1:2)
title("Total Variation Norm Constraint")

## End(Not run)
```

`Norberg` *Norberg Life Insurance Data*

Description

Norwegian Life Insurance Exposures and Claims

Usage

```
Norberg
```

Format

A data frame with 72 observations on the following 3 variables.

- `OccGroup`Occupational Group
- `Exposure`Exposures
- `Deaths`Observed Deaths

Details

The data arise from 1125 original groups insured during all or part of the period 1982-85 by a major Nowegian insurance company. Exposures can be normalized by a factor of 344 as in Hastrup (2000) and then can be interpreted as the apriori expected number of claims (deaths) for each group. The original 1125 groups were aggregated into 72 as in Norberg (1989).

References

Norberg, R. (1989) Experience rating in group life insurance, Scand. Actuarial J.,194-224.

Haastrup, S. (2000) Comparison of some Bayesian analyses of heterogeneity in group life insurance, Scand. Actuarial J.,2-16.

`NPmix` *Normal mixture with Poisson sample size via Kiefer Wolfowitz NPMLE*

Description

Interior point solution of Kiefer-Wolfowitz NPMLE for mixture of Normal/Poissons

Usage

```
NPmix(x, m, v = 50, u = 50, weights = NULL, ...)
```

Arguments

`x`	observed response for Gaussian observations
`m`	Number of trials for Poisson observations
`v`	Grid Values for the Gaussian means mixing distribution defaults to equal spacing of length v on [min(x) + eps, max(x) - eps], if v is scalar.
`u`	Grid Values for the Poisson rate mixing distribution defaults to equal spacing of length u on [min(m) + eps, max(m) - eps], if u is scalar.
`weights`	replicate weights for x obervations, should sum to 1
`...`	Other arguments to be passed to KWDual to control optimization

Details

The joint distribution of the means and the number of trials determining sample standard deviations is estimated. The grid specification for means is as for `GLmix` whereas the grid for the Poisson rate parameters by default depends on the support of the observed trials. There is no predict method as yet. See `demo(NPmix1)`.

Value

An object of class density with components:

`v`	grid points of evaluation of the success probabilities
`u`	grid points of evaluation of the Poisson rate for number of trials
`y`	function values of the mixing density at (v,u)
`g`	estimates of the mixture density at the distinct data values
`logLik`	Log Likelihood value at the estimate
`status`	exit code from the optimizer

Author(s)

R. Koenker and J. Gu

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist*. 27, (1956), 887-906.

`plot.GLVmix` *Plot a GLVmix object*

Description

Given a fitted mixture model by GLVmix plot the estimated mass points

Given a fitted mixture model by GLVmix plot the estimated mass points

Usage

```
## S3 method for class 'GLVmix'
plot(x, ...)

## S3 method for class 'GLVmix'
plot(x, ...)
```

Arguments

`x`	is the fitted object
`...`	other arguments to pass to `symbols`, notably e.g. `add = TRUE`

Value

nothing (invisibly)

nothing (invisibly)

---

plot.medde — *Plotting method for medde objects*

Description

Plotting method for medde objects

Usage

```
## S3 method for class 'medde'
plot(x, ...)
```

Arguments

`x`	object obtained from medde fitting
`...`	other parameters to be passed to plot method

Pmix — *Poisson mixture estimation via Kiefer Wolfowitz MLE*

Description

Poisson mixture estimation via Kiefer Wolfowitz MLE

Usage

```
Pmix(x, v = 300, support = NULL, exposure = NULL, ...)
```

Arguments

`x`	Data: Sample observations (integer valued)
`v`	Grid Values for the mixing distribution defaults to equal spacing of length v when v is specified as a scalar
`support`	a 2-vector containing the lower and upper support points of sample observations to account for possible truncation.
`exposure`	observation specific exposures to risk see details
`...`	other parameters passed to KWDual to control optimization

Details

The predict method for `Pmix` objects will compute means, medians or modes of the posterior according to whether the `Loss` argument is 2, 1 or 0, or posterior quantiles if `Loss` is in (0,1).

In the default case `exposure = 1` it is assumed that `x` contains individual observations that are aggregated into count bins via `table`. When `exposure` has the same length as `x` then it is presumed to be individual specific risk exposure and the Poisson mixture is taken to be $x|v\ Poi(v * exposure)$ and the is not aggregated. See for example the analysis of the Norberg data in Koenker and Gu (2016).

Value

An object of class density with components:

`x`	points of evaluation of the mixing density
`y`	function values of the mixing density at x
`g`	function values of the mixture density on $0, 1, ...max(x) + 1$
`logLik`	Log Likelihood value at the estimate
`dy`	Bayes rule estimate of Poisson rate parameter at each x
`status`	exit code from the optimizer

Author(s)

Roger Koenker and Jiaying Gu

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist.* Volume 27, Number 4 (1956), 887-906.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

`predict.B2mix` *Predict Method for Bmix*

Description

Predict Method for Binomial Mixtures

Usage

```
## S3 method for class 'B2mix'
predict(object, newdata, Loss = 2, newk, ...)
```

Arguments

`object`	fitted object of class "B2mix"
`newdata`	Values at which prediction is desired an n by 2 matrix
`Loss`	Loss function used to generate prediction: Currently supported values: 2 to get mean predictions, 1 to get median predictions, 0 to get modal predictions or any tau in (0,1) to get tau-th quantile predictions.
`newk`	k values (number of trials) for the predictions an n by 2 matrix
`...`	optional arguments to predict

Details

The predict method for B2mix objects will compute posterior means.

Value

A vector of predictions

Author(s)

Jiaying Gu and Roger Koenker

`predict.Bmix` *Predict Method for Bmix*

Description

Predict Method for Binomial Mixtures

Usage

```
## S3 method for class 'Bmix'
predict(object, newdata, Loss = 2, newk, ...)
```

Arguments

`object`	fitted object of class "Bmix"
`newdata`	Values at which prediction is desired
`Loss`	Loss function used to generate prediction: Currently supported values: 2 to get mean predictions, 1 to get median predictions, 0 to get modal predictions or any tau in (0,1) to get tau-th quantile predictions.
`newk`	k values (number of trials) for the predictions
`...`	optional arguments to predict

Details

The predict method for `Bmix` objects will compute means, quantiles or modes of the posterior according to the `Loss` argument. Typically, `newdata` would be passed to `predict`

Value

A vector of predictions

Author(s)

Jiaying Gu

`predict.GLmix` *Predict Method for GLmix*

Description

Predict Method for Gaussian Location Mixtures

Usage

```
## S3 method for class 'GLmix'
predict(object, newdata, Loss = 2, newsigma = NULL, ...)
```

Arguments

`object`	fitted object of class "GLmix"
`newdata`	Values at which prediction is desired
`Loss`	Loss function used to generate prediction: Currently supported values: 2 to get mean predictions, 1 to get median predictions, 0 to get modal predictions or any tau in (0,1) to get tau-th quantile predictions.
`newsigma`	sigma values for the predictions
`...`	optional arguments to predict

Details

The predict method for `GLmix` objects will compute means, quantiles or modes of the posterior according to the `Loss` argument. Typically, `newdata` would be passed to `predict`

Value

A vector of predictions

Author(s)

Roger Koenker

predict.GLVmix — *Predict Method for GLVmix*

Description

Predict Method for Gaussian Location-scale Mixtures

Usage

```
## S3 method for class 'GLVmix'
predict(object, newdata, Loss = 2, ...)
```

Arguments

`object`	Fitted object of class "GLVmix"
`newdata`	data.frame with components(t,s,m) at which prediction is desired
`Loss`	Loss function used to generate prediction: Currently supported values: 2 to get mean predictions, 1 to get median predictions, 0 to get modal predictions or any tau in (0,1) to get tau-th quantile predictions.
`...`	optional arguments to predict

Details

The predict method for `GLmix` objects will compute means, quantiles or modes of the posterior according to the `Loss` argument. Typically, `newdata` would be passed to `predict`. Note that these predictions are for the location parameter only.

Value

A vector of predictions

Author(s)

Roger Koenker

predict.Pmix — *Predict Method for Pmix*

Description

Predict Method for Poisson Mixtures

Usage

```
## S3 method for class 'Pmix'
predict(object, newdata, Loss = 2, newexposure = NULL, ...)
```

Arguments

`object`	fitted object of class "Pmix"
`newdata`	Values at which prediction is desired
`Loss`	Loss function used to generate prediction. Currently supported values: 2 to get mean predictions, 1 to get harmonic mean predictions, 0 to get modal predictions or any tau in (0,1) to get tau-th quantile predictions. The posterior harmonic mean is the Bayes rule for quadratic loss weighted by variances as in Clevenson and Zidek (1975).
`newexposure`	exposure values for the predictions
`...`	optional arguments to predict

Details

The predict method for `Pmix` objects will compute means, quantiles or modes of the posterior according to the `Loss` argument. Typically, `newdata` would be passed to `predict`

Value

A vector of predictions

Author(s)

Jiaying Gu and Roger Koenker

References

Clevenson, M. L. and Zidek, J. V. 1975. Simultaneous Estimation of the Means of Independent Poisson Laws, Journal of the American Statistical Association, 70, 698-705.

`predict.WGLVmix` *Predict Method for WGLVmix*

Description

Predict Method for Gaussian Location-scale Mixtures (Longitudinal Version)

Usage

```
## S3 method for class 'WGLVmix'
predict(object, newdata, Loss = 2, ...)
```

Arguments

`object`	Fitted object of class "GLVmix"
`newdata`	data.frame with components(y,id,w) at which prediction is desired this data structure must be compatible with that of `WGLVmix`, if newdata$w is NULL then w is replaced by a vector of ones of length(y)
`Loss`	Loss function used to generate prediction: Currently supported values: 2 to get mean predictions, 1 to get median predictions, 0 to get modal predictions or any tau in (0,1) to get tau-th quantile predictions.
`...`	optional arguments to predict

Details

The predict method for `WGLmix` objects will compute means, quantiles or modes of the posterior according to the `Loss` argument. Typically, `newdata` would be passed to `predict`. Note that these predictions are for the location parameter only.

Value

A vector of predictions

Author(s)

Roger Koenker

`qKW` *Quantiles of KW fit*

Description

Quantiles of KW fit

Usage

```
qKW(g, q)
```

Arguments

`g`	KW fitted object
`q`	vector of quantiles to be computed

Author(s)

R. Koenker

qKW2 — *Quantiles of bivariate KW fit*

Description

Quantiles of bivariate KW fit

Usage

```
qKW2(g, q)
```

Arguments

g	KW fitted object
q	vector of quantiles to be computed

Author(s)

R. Koenker

qmedde — *Quantile function for medde estimate*

Description

Slightly modified version borrowed from the package logcondens Todo: extend this to cases with $\alpha! = 1$.

Usage

```
qmedde(p, medde)
```

Arguments

p	vector of probabilities at which to evaluate the quantiles
medde	fitted object from medde

rKW — *Random sample from KW object*

Description

Random sample from KW object

Usage

```
rKW(n, g)
```

Arguments

n	sample size
g	KW object

Author(s)

R. Koenker

RLR — *Regularized Logistic Regression*

Description

Logistic Regression with lasso like penalties

Usage

```
RLR(X, Y, D, lambda, ...)
```

Arguments

X	a design matrix for the unconstrained logistic regression model
Y	a response vector of Boolean values, or n by 2 matrix of binomials as in `glm`
D	is a matrix specifying the penalty, `diag(ncol(X))` for the conventional lasso penalty
lambda	a scalar specifying the intensity of one's belief in the prior. No provision for automatic selection has been made (yet).
...	other parameters passed to control optimization: These may include `rtol` the relative tolerance for dual gap convergence criterion, `verb` to control verbosity desired from mosek, `verb = 0` is quiet, `verb = 5` produces a fairly detailed iteration log. See the documentation for `KWDual` for further details.

Details

In some logistic regression problems, especially those with a large number of fixed effects like the Bradley-Terry rating model, it may be plausible to consider groups of effects that would be considered equivalence classes. One way to implement such prior information is to impose some form of regularization penalty. In the general formulation we are trying to solve the problem:

$$\min \ell(\theta|X, y) + \|D\theta\|_1$$

. For example in the Bradley-Terry rating model, we may consider penalties of the form,

$$\|D\theta\|_1 = \sum_{i<j} |\theta_i - \theta_j|$$

so differences in all pairs of ratings are pulled together. This form of the penalty has been used by Hocking et al (2011) for clustering, by Masarotto and Varin (2012) for estimation of the Bradley Terry model and by Gu and Volgushev (2019) for grouping fixed effects in panel data models. This is an implementation in Mosek, so the package **Rmosek** and Mosek must be available at run time. The `demo(RLR1)` illustrates use with the conventional lasso penalty and produces a lasso shrinkage plot. The `demo(RLR2)` illustrates use with the ranking/grouping lasso penalty and produces a plot of how the number of groups is reduced as lambda rises.

Value

A list with components:

`coef`	vector of coefficients
`logLik`	log likelihood value at the solution
`status`	return status from the Mosek optimizer

.

Author(s)

Roger Koenker with crucial help from Michal Adamaszek of Mosek ApS

References

Gu, J. and Volgushev, S. (2019), ‘Panel data quantile regression with grouped fixed effects’, *Journal of Econometrics*, 213, 68–91.

Hocking, T. D., Joulin, A., Bach, F. and Vert, J.-P. (2011), ‘Clusterpath: an algorithm for clustering using convex fusion penalties’, Proceedings of the 28th International Conference on International Conference on Machine Learning, 745–752.

Masarotto, G. and Varin, C. (2012), ‘The ranking lasso and its application to sport tournaments’, *The Annals of Applied Statistics*, 6, 1949–1970.

rmedde	*Random number generation from a medde estimate*

Description

Random number generation from a medde estimate

Usage

```
rmedde(n, medde, smooth = TRUE)
```

Arguments

n	number of observations desired in calls to rmedde
medde	fitted medde object for calls in qmedde and rmedde
smooth	option to draw random meddes from the smoothed density

Rxiv	*Archive function for auxiliary files for latex documents*

Description

Creates a tar.gz file with all of the R files needed to recreate the tables and figures that appear in the paper. Should be considered experimental at this stage. It presumes that tables are generated with something like the **Hmisc** `latex` function and included in the latex document with `input` commands. Likewise figures are assumed to be included with `includegraphics` and generated by R in pdf format. This was originally developed to sort out the files for "Empirical Bayesball Remixed". An optional side of effect of the function to create a tar.gz file with the gzipped R files required for the paper.

Usage

```
Rxiv(fname, figures = "figures", tables = "tables", tar = FALSE)
```

Arguments

fname	name of the latex file of the paper sans .tex suffix
figures	name of the directory with the files for figures
tables	name of the directory with the files for tables
tar	logical flag, if TRUE generate a gzipped tar file of .R files

Value

a list with the following components

`Rtables`	a character array with two columns: .tex files and .R files
`Rfigures`	a character array with two columns: .pdf files and .R files
`Rother`	a character vector with other R files required.
`Rcached`	a character vector with cached Rda files

Author(s)

R. Koenker

`tacks` *Beckett and Diaconis flipping tacks data*

Description

This data was generated by Beckett and Diaconis (1994). They describe it as follows: "The example involves repeated rolls of a common thumbtack. A one was recorded if the tack landed point up and a zero was recorded if the tack landed point down. All tacks started point down. Each tack was flicked or hit with the fingers from where it last rested. A fixed tack was flicked 9 times. The data are recorded in Table 1. There are 320 9-tuples. These arose from 16 different tacks, 2 "flickers," and 10 surfaces. The tacks vary considerably in shape and in proportion of ones. The surfaces varied from rugs through tablecloths through bathroom floors." Following Liu (1996), we treat the data as though they came from 320 independent binomials. See `demo(Bmix1)` for further details.

Usage

```
tacks
```

Format

A data frame with 320 observations on 2 variables.

- `x`a numeric vector giving the number of tacks landed point up.
- `k`a numeric vector giving the number of trials.

Source

Beckett, L. and Diaconis. P. (1994). Spectral analysis for discrete longitudinal data, Adv. Math., 103: 107-128.

References

Liu, J.S. (1996). Nonparametric Hierarchical Bayes via Sequential Imputations. *Annals of Statistics*, 24: 911-930.

tannenbaum *Perverse Gaussian Mixture data*

Description

Gaussian Location Mixture data to illustrate Mosek tolerance problem

Usage

```
tannenbaum
```

Format

5000 iid Gaussians This data set was randomly generated in the course of trying to understand some anomalies in estimating Gaussian location mixture problems with `GLmix`. It is used by `demo(tannenbaum)` to illustrate that sometimes it is worthwhile to tighten the default convergence tolerance for Mosek.

ThreshFDR *Thresholding for False Discovery Rate*

Description

This function approximates FDR for various values of lambda and is usually employed in conjunction with `Finv` to find an appropriate cutoff value lambda.

Usage

```
ThreshFDR(lambda, stat, v)
```

Arguments

`lambda`	is the proposed threshold
`stat`	is the statistic used for ranking
`v`	is the local false discovery statistic

TLmix *NPMLE for Student t location mixtures*

Description

Kiefer Wolfowitz NPMLE for Student t location mixtures

Usage

```
TLmix(x, v = 300, u = 300, df = 1, hist = FALSE, weights = NULL, ...)
```

Arguments

`x`	Data: Sample Observations
`v`	bin boundaries defaults to equal spacing of length v
`u`	bin boundaries for histogram binning: defaults to equal spacing
`df`	Number of degrees of freedom of Student base density
`hist`	If TRUE then aggregate x to histogram weights
`weights`	replicate weights for x obervations, should sum to 1
`...`	optional parameters passed to KWDual to control optimization

Details

Kiefer Wolfowitz MLE density estimation as proposed by Jiang and Zhang for a Student t compound decision problem. The histogram option is intended for large problems, say n > 1000, where reducing the sample size dimension is desirable. By default the grid for the binning is equally spaced on the support of the data. Equal spaced binning is problematic for Cauchy data.

Value

An object of class density with components:

`x`	midpoints of evaluation on the domain of the mixing density
`y`	estimated function values at the points x of the mixing density
`logLik`	Log likelihood value at the proposed solution
`dy`	Bayes rule estimates of location at x
`status`	Mosek exit code

Author(s)

Roger Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist*. 27, (1956), 887-906.

Jiang, Wenhua and Cun-Hui Zhang General maximum likelihood empirical Bayes estimation of normal means *Ann. Statist.*, 37, (2009), 1647-1684.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

GLmix for Gaussian version

`Tncpmix` *NPMLE for Student t non-centrality parameter mixtures*

Description

Kiefer Wolfowitz NPMLE for Student t non-centrality parameter mixtures Model: $y_{ig} = mu_g + e_{ig}, e_{ig}\ N(0, sigma_g^2)$ x is the vector of t statistics for all groups, which follows t dist if $mu_g = 0$, and noncentral t dist if $mu_g \neq 0$, with $ncp_g = \mu_g/\sigma_g$. This leads to a mixture of t distribution with ncp as the mixing parameter. df (degree of freedom) is determined by the group size in the simplest case.

Usage

```
Tncpmix(x, v = 300, u = 300, df = 1, hist = FALSE, weights = NULL, ...)
```

Arguments

`x`	Data: Sample Observations
`v`	bin boundaries defaults to equal spacing of length v
`u`	bin boundaries for histogram binning: defaults to equal spacing
`df`	Number of degrees of freedom of Student base density
`hist`	If TRUE then aggregate x to histogram weights
`weights`	replicate weights for x obervations, should sum to 1
`...`	optional parameters passed to KWDual to control optimization

Value

An object of class density with components:

`x`	midpoints of evaluation on the domain of the mixing density
`y`	estimated function values at the points x of the mixing density
`g`	estimated function values at the observed points of mixture density
`logLik`	Log likelihood value at the proposed solution
`dy`	Bayes rule estimates of location at x
`status`	Mosek exit code

Author(s)

Roger Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist*. 27, (1956), 887-906.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

GLmix for Gaussian version

`traprule` *Integration by Trapezoidal Rule*

Description

Integration by Trapezoidal Rule

Usage

```
traprule(x, y)
```

Arguments

`x`	points of evaluation
`y`	function values

Details

Crude Riemann sum approximation.

Value

A real number.

Author(s)

R. Koenker

Umix *NPMLE for Uniform Scale Mixtures*

Description

Kiefer-Wolfowitz Nonparametric MLE for Uniform Scale Mixtures

Usage

```
Umix(x, ...)
```

Arguments

`x`	Data: Sample Observations
`...`	other parameters to pass to KWDual to control optimization

Details

Kiefer-Wolfowitz MLE for the mixture model $Y \sim U[0,T], \ T \sim G$ No gridding is required since mass points of the mixing distribution, G, must occur at the data points. This formalism is equivalent, as noted by Groeneboom and Jongbloed (2014) to the Grenander estimator of a monotone density in the sense that the estimated mixture density, i.e. the marginal density of Y, is the Grenander estimate, see the remark at the end of their Section 2.2. See also `demo(Grenander)`. Note that this refers to the decreasing version of the Grenander estimator, for the increasing version try standing on your head.

Value

An object of class density with components:

`x`	points of evaluation on the domain of the density
`y`	estimated mass at the points x of the mixing density
`g`	the estimated mixture density function values at x
`logLik`	Log likelihood value at the proposed solution
`status`	exit code from the optimizer

Author(s)

Jiaying Gu and Roger Koenker

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist*. Volume 27, Number 4 (1956), 887-906.

Groeneboom, P. and G. Jongbloed, *Nonparametric Estimation under Shape Constraints*, 2014, Cambridge U. Press.

`velo` *Rotational Velocity of Stars*

Description

A sample of rotational velocities of stars from Hoffleit and Warren (1991) similar to that previosly considered by Pal, Woodroofe and Meyer (2007) and used by Koenker and Mizera (2010). The `demo(velo)` illustrates fitted densities for three relatively weak concavity constraints corresponding to $-1/sqrt(f)$, $-1/f$ and $-1/f^2$ constrained to be concave. Note that last of these pushes the optimization methods about as far as they can do.

Usage

```
velo
```

Format

A numeric vector with 3933 observations on one variable.

- `velo`a numeric vector with rotational velocities.

Source

Hoffleit, D. and Warren, W. H. (1991). The Bright Star Catalog (5th ed.). Yale University Observatory, New Haven.

References

Pal, J. K., Woodroofe, M. and Meyer, M. (2007). Estimating a Polya frequency function. In Complex Datasets and Inverse Problems: Tomography, Networks and Beyond, (R. Liu, W. Strawderman, and C.-H. Zhang, eds.). IMS Lecture Notes-Monograph Series 54 239-249. Institute of Mathematical Statistics. Koenker, R. and Mizera, I. (2010) Quasi-Concave Density Estimation, Annals of Statistics, 38, 2998-3027.

`Weibullmix` *NPMLE for Weibull Mixtures*

Description

Kiefer-Wolfowitz NPMLE for Weibull Mixtures of scale parameter

Usage

```
Weibullmix(
  x,
  v = 300,
  u = 300,
  alpha,
  lambda = 1,
  hist = FALSE,
  weights = NULL,
  ...
)
```

Arguments

`x`	Survival times
`v`	Grid values for mixing distribution
`u`	Grid values for histogram bins, if needed
`alpha`	Shape parameter for Weibull distribution
`lambda`	Scale parameter for Weibull Distribution; must either have length 1, or length equal to `length(x)` the latter case accommodates the possibility of a linear predictor
`hist`	If TRUE aggregate to histogram counts
`weights`	replicate weights for x obervations, should sum to 1
`...`	optional parameters passed to KWDual to control optimization

Details

Kiefer Wolfowitz NPMLE density estimation for Weibull scale mixtures. The histogram option is intended for relatively large problems, say n > 1000, where reducing the sample size dimension is desirable. By default the grid for the binning is equally spaced on the support of the data. Parameterization: f(t|alpha, lambda) = alpha * exp(v) * (lambda * t)^(alpha-1) * exp(-(lambda * t)^alpha * exp(v)); shape = alpha; scale = lambda^(-1) * (exp(v))^(-1/alpha)

Value

An object of class density with components

`x`	points of evaluation on the domain of the density
`y`	estimated function values at the points x of the mixing density
`logLik`	Log likelihood value at the proposed solution
`dy`	Bayes Rule estimates of mixing parameter
`status`	exit code from the optimizer

Author(s)

Roger Koenker and Jiaying Gu

References

Kiefer, J. and J. Wolfowitz Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters *Ann. Math. Statist.* Volume 27, Number 4 (1956), 887-906.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

`Gompertzmix`

`WGLVmix` *Weighted NPMLE of Longitudinal Gaussian Mean and Variances Model*

Description

A Kiefer-Wolfowitz procedure for ML estimation of a Gaussian model with dependent mean and variance components and weighted longitudinal data. This version assumes a general bivariate distribution for the mixing distribution. The defaults use a rather coarse bivariate gridding. In contrast to the function `GLVmix` the full longitudinal data structure is required for this function and the likelihood evaluation reflects this difference.

Usage

```
WGLVmix(y, id, w, u = 30, v = 30, ...)
```

Arguments

`y`	A vector of observations
`id`	A strata indicator vector of the same length
`w`	A vector of weights
`u`	A vector of bin boundaries for the mean effects
`v`	A vector of bin boundaries for the variance effects
`...`	optional parameters to be passed to KWDual to control optimization

Value

A list consisting of the following components:

`u`	midpoints of mean bin boundaries
`v`	midpoints of variance bin boundaries
`fuv`	the function values of the mixing density.
`logLik`	log likelihood value for mean problem
`du`	Bayes rule estimate of the mixing density means.
`dv`	Bayes rule estimate of the mixing density variances.
`A`	Constraint matrix
`status`	Mosek convergence status

Author(s)

R. Koenker and J. Gu

References

Gu, J. and R. Koenker (2014) Heterogeneous Income Dynamics: An Empirical Bayes Perspective, *JBES*,35, 1-16.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

WTLVmix for an implementation assuming independent heterogeneity, GLVmix for an implementation that assumes the availability of only the summary statistics but not the full longitudinal data structure.

WGVmix — *WGVmix: Weighted Generalized Maximum Likelihood for Empirical Bayes Estimation of Gamma Variances*

Description

A Kiefer-Wolfowitz procedure for ML estimation of a Gaussian model with independent variance components with weighted longitudinal data.

Usage

```
WGVmix(
  y,
  id,
  w,
  v,
  pv = 300,
  eps = 1e-06,
  rtol = 1e-06,
  verb = 0,
  control = NULL
)
```

Arguments

`y`	A vector of observations
`id`	A strata indicator vector of the same length
`w`	A vector of weights
`v`	A vector of bin boundaries for the variance effects

pv	The number of variance effect bins, if u is missing
eps	A tolerance for determining the support of the bins
rtol	A tolerance for determining duality gap convergence tolerance in Mosek
verb	A flag indicating how verbose the Mosek output should be
control	Mosek control list see KWDual documentation

Details

See Gu and Koenker (2012?)

Value

An object of class `density` consisting of the following components:

x	the variance bin boundaries
y	the function values of the mixing density for the variances.
logLik	the value of the log likelihood at the solution
status	the mosek convergence status.

Author(s)

R. Koenker

References

Gu Y. and R. Koenker (2017) Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data, *J. of Applied Econometrics*, 32, 575-599.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

`WLVmix` *NPMLE for Longitudinal Gaussian Means and Variances Model with Independent Prior*

Description

A Kiefer-Wolfowitz NPMLE procedure for estimation of a Gaussian model with independent mean and variance prior components with weighted longitudinal data. This version iterates back and forth from Gamma and Gaussian forms of the likelihood.

Usage

```
WLVmix(y, id, w, u = 300, v = 300, eps = 1e-04, maxit = 2, ...)
```

Arguments

`y`	A vector of observations
`id`	A strata indicator vector indicating grouping of y
`w`	A vector of weights corresponding to y
`u`	A vector of bin boundaries for the mean effects
`v`	A vector of bin boundaries for the variance effects
`eps`	Convergence tolerance for iterations
`maxit`	A limit on the number of allowed iterations
`...`	optional parameters to be passed to KWDual to control optimization

Value

A list consisting of the following components:

`u`	midpoints of the mean bin boundaries
`fu`	the function values of the mixing density of the means
`v`	midpoints of the variance bin boundaries
`fv`	the function values of the mixing density of the variances.
`logLik`	vector of log likelihood values for each iteration
`du`	Bayes rule estimate of the mixing density means.
`dv`	Bayes rule estimate of the mixing density variances.
`status`	Mosek convergence status for each iteration

Author(s)

J. Gu and R. Koenker

References

Gu, J. and R. Koenker (2015) Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data, *J. Applied Econometrics*, 32, 575-599.

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

WGLVmix for a more general bivariate mixing distribution version and WTLVmix for an alternative estimator exploiting a Student/Gamma decomposition

WTLVmix *NPMLE for Longitudinal Gaussian Means and Variances Model*

Description

A Kiefer-Wolfowitz NPMLE procedure for estimation of a Gaussian model with independent mean and variance components with weighted longitudinal data. This version exploits a Student t decomposition of the likelihood.

Usage

```
WTLVmix(y, id, w, u = 300, v = 300, ...)
```

Arguments

`y`	A vector of observations
`id`	A strata indicator vector indicating grouping of y
`w`	A vector of weights corresponding to y
`u`	A vector of bin boundaries for the mean effects
`v`	A vector of bin boundaries for the variance effects
`...`	optional parameters to be passed to KWDual to control optimization

Value

A list consisting of the following components:

`u`	midpoints of the mean bin boundaries
`fu`	the function values of the mixing density of the means
`v`	midpoints of the variance bin boundaries
`fv`	the function values of the mixing density of the variances.
`logLik`	log likelihood value for mean problem
`du`	Bayes rule estimate of the mixing density means.
`dv`	Bayes rule estimate of the mixing density variances.
`status`	Mosek convergence status

Author(s)

J. Gu and R. Koenker

References

Koenker, R. and J. Gu, (2017) REBayes: An R Package for Empirical Bayes Mixture Methods, *Journal of Statistical Software*, 82, 1–26.

See Also

WGLVmix for a more general bivariate mixing distribution version

Index