Package 'L1pack'

September 8, 2024

Type Package
Title Routines for L1 Estimation
Version 0.50
Date 2024-09-07
Maintainer Felipe Osorio <felipe.osorios@usm.cl></felipe.osorios@usm.cl>
Description L1 estimation for linear regression using Barrodale and Roberts' method doi:10.1145/355616.361024 and the EM algorithm doi:10.1023/A:1020759012226 . Estimation of mean and covariance matrix using the multivariate Laplace distribution, density, distribution function, quantile function and random number generation for univariate and multivariate Laplace distribution doi:10.1080/03610929808832115 . Implementation of Naik and Plungpongpun doi:10.1007/0-8176-4487-3_7 for the Generalized spatial median estimator is included.
Depends $R(>=3.5.0)$, fastmatrix
LinkingTo fastmatrix
Imports stats, grDevices, graphics
License GPL-3
<pre>URL http://l1pack.mat.utfsm.cl/</pre>
NeedsCompilation yes
LazyLoad yes
Author Felipe Osorio [aut, cre] (https://orcid.org/0000-0002-4675-5201), Tymoteusz Wolodzko [aut]
Repository CRAN
Date/Publication 2024-09-08 07:20:02 UTC
Contents
confint.lad

2 confint.lad

	lad.fit										 									6
	lad.fit-methods										 									7
	Laplace										 									8
	LaplaceFit																			
	mLaplace																			
	simulate.lad .																			
	spatial.median																			
	vcov.lad										 							 		14
	WH.Laplace .										 							 		15
																				4.0
Index																				17

confint.lad

Confidence intervals from lad models

Description

Computes confidence intervals for one or more parameters in a fitted model associated to a lad object.

Usage

```
## S3 method for class 'lad'
confint(object, parm, level = 0.95, ...)
```

Arguments

object a fitted model object.

parm a specification of which parameters are to be given confidence intervals, either

a vector of numbers or a vector of names. If missing, all parameters are consid-

ered.

level the confidence level required.

... additional argument(s) for related methods.

Details

confint is a generic function. Confidence intervals associated to lad objects are asymptotic, and needs suitable coef and vcov methods to be available.

Value

A matrix (or vector) with columns giving lower and upper confidence limits for each parameter. These will be labelled as (1-level)/2 and 1 - (1-level)/2 in % (by default 2.5% and 97.5%).

See Also

```
confint.glm and confint.nls in package MASS.
```

ereturns 3

Examples

```
fm <- lad(stack.loss \sim ., data = stackloss, method = "BR") confint(fm) # based on asymptotic normality
```

ereturns

Excess returns for Martin Marietta and American Can companies

Description

Data from the Martin Marietta and American Can companies collected over a period of 5 years on a monthly basis.

Usage

data(ereturns)

Format

A data frame with 60 observations on the following 4 variables.

Date the month in which the observations were collected.

am.can excess returns from the American Can company.

m.marietta excess returns from the Martin Marietta company.

CRSP an index for the excess rate returns for the New York stock exchange.

Source

Butler, R.J., McDonald, J.B., Nelson, R.D., and White, S.B. (1990). Robust and partially adaptive estimation of regression models. *The Review of Economics and Statistics* **72**, 321-327.

11fit

Minimum absolute residual (L1) regression

Description

Performs an L1 regression on a matrix of explanatory variables and a vector of responses.

Usage

```
l1fit(x, y, intercept = TRUE, tolerance = 1e-07, print.it = TRUE)
```

4 11fit

Arguments

X	vector or matrix of explanatory variables. Each row corresponds to an observation and each column to a variable. The number of rows of x should equal the number of data values in y, and there should be fewer columns than rows. Missing values are not allowed.
у	numeric vector containing the response. Missing values are not allowed.
intercept	logical flag. If TRUE, an intercept term is included in the regression model.
tolerance	numerical value used to test for singularity in the regression.
print.it	logical flag. If TRUE, then warnings about non-unique solutions and rank deficiency are given.

Details

The Barrodale-Roberts algorithm, which is a specialized linear programming algorithm, is used.

Value

list defining the regression (compare with function lsfit).

coefficients vector of coefficients.

residuals residuals from the fit.

message character strings stating whether a non-unique solution is possible, or if the x matrix was found to be rank deficient.

References

Barrodale, I., Roberts, F.D.K. (1973). An improved algorithm for discrete L1 linear approximations. *SIAM Journal of Numerical Analysis* **10**, 839-848.

Barrodale, I., Roberts, F.D.K. (1974). Solution of an overdetermined system of equations in the L1 norm. *Communications of the ACM* **17**, 319-320.

Bloomfield, P., Steiger, W.L. (1983). *Least Absolute Deviations: Theory, Applications, and Algorithms*. Birkhauser, Boston, Mass.

```
l1fit(stack.x, stack.loss)
```

lad 5

|--|

Description

This function is used to fit linear models considering Laplace errors.

Usage

```
lad(formula, data, subset, na.action, method = "BR", tol = 1e-7, maxiter = 200, x = FALSE, contrasts = NULL)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted.
data	an optional data frame containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which lad is called.
subset	an optional expression indicating the subset of the rows of data that should be used in the fit.
na.action	a function that indicates what should happen when the data contain NAs.
method	character string specifying the fitting method to be used; the options are "BR" Barrodale and Roberts' method (the default) and "EM" for an EM algorithm using IRLS.
tol	the relative tolerance for the iterative algorithm. Default is tol = 1e-7.
maxiter	The maximum number of iterations for the EM method. Default to 200.
x, y	logicals. If TRUE the corresponding components of the fit (the model matrix, the response) are returned.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.

Value

An object of class lad representing the linear model fit. Generic function print, show the results of the fit.

The functions print and summary are used to obtain and print a summary of the results. The generic accessor functions coefficients, fitted.values and residuals extract various useful features of the value returned by lad.

Author(s)

The design was inspired by the R function 1m.

6 lad.fit

References

Barrodale, I., Roberts, F.D.K. (1974). Solution of an overdetermined system of equations in the L1 norm. *Communications of the ACM* **17**, 319-320.

Phillips, R.F. (2002). Least absolute deviations estimation via the EM algorithm. *Statistics and Computing* **12**, 281-285.

Examples

```
fm <- lad(stack.loss ~ ., data = stackloss, method = "BR")
summary(fm)

data(ereturns)
fm <- lad(m.marietta ~ CRSP, data = ereturns, method = "EM")
summary(fm)
# basic observations
fm$basic</pre>
```

lad.fit

Fitter functions for least absolute deviation (LAD) regression

Description

This function is a *switcher* among various numerical fitting functions (lad.fit.BR, and lad.fit.EM). The argument method does the switching: "BR" for lad.fit.BR, etc. This should usually *not* be used directly unless by experienced users.

Usage

```
lad.fit(x, y, method = "BR", tol = 1e-7, maxiter = 200)
```

Arguments

x design matrix of dimension $n \times p$. y vector of observations of length n.

method currently, methods "BR" (default), and "EM" are supported.

tol the relative tolerance for the iterative algorithm. Default is tol = 1e-7.

maxiter The maximum number of iterations for the EM method. Default to 200.

Value

a list with components:

coefficients a named vector of coefficients.

scale final scale estimate of the random error.

residuals the residuals, that is response minus fitted values.

fitted.values the fitted values.

lad.fit-methods 7

SAD	the sum of absolute deviations.
0710	the sum of absolute deviations.

weights estimated EM weights.

basic basic observations, that is observations with zero residuals.

logLik the log-likelihood at convergence.

See Also

```
lad.fit.BR, lad.fit.EM.
```

Examples

```
x <- cbind(1, stack.x)
fm <- lad.fit(x, stack.loss, method = "BR")
fm</pre>
```

lad.fit-methods

Fit a least absolute deviation (LAD) regression model

Description

Fits a linear model using LAD methods, returning the bare minimum computations.

Usage

```
lad.fit.BR(x, y, tol = 1e-7)
lad.fit.EM(x, y, tol = 1e-7, maxiter = 200)
```

Arguments

x, y	numeric vectors or matrices for the predictors and the response in a linear model. Typically, but not necessarily, x will be constructed by one of the fitting functions.
tol	the relative tolerance for the iterative algorithm. Default is tol = 1e-7.
maviter	The maximum number of iterations for the EM method. Default to 200

Value

The bare bones of a lad object: the coefficients, residuals, fitted values, and some information used by summary.lad.

See Also

```
lad, lad.fit, lm
```

```
x <- cbind(1, stack.x)
z <- lad.fit.BR(x, stack.loss)
z</pre>
```

8 Laplace

Laplace

The symmetric Laplace distribution

Description

Density, distribution function, quantile function and random generation for the Laplace distribution with location parameter location and scale parameter scale.

Usage

```
dlaplace(x, location = 0, scale = 1, log = FALSE)
plaplace(q, location = 0, scale = 1, lower.tail = TRUE, log.p = FALSE)
qlaplace(p, location = 0, scale = 1, lower.tail = TRUE, log.p = FALSE)
rlaplace(n, location = 0, scale = 1)
```

Arguments

x, q	vector of quantiles.
р	vector of probabilities.
n	number of observations.
location	location parameter μ , which is the mean.
scale	scale parameter ϕ . Scale must be positive.
log, log.p	logical; if TRUE, probabilities p are given as $\log(p)$.
lower.tail	logical; if TRUE (default), probabilities are $P(X \le x)$, otherwise, $P(X > x)$.

Details

If location or scale are not specified, they assume the default values of 0 and 1 respectively.

The Laplace distribution with location μ and scale ϕ has density

$$f(x) = \frac{1}{\sqrt{2}\phi} \exp(-\sqrt{2}|x - \mu|/\phi),$$

where $-\infty < y < \infty, -\infty < \mu < \infty$ and $\phi > 0$. The mean is μ and the variance is ϕ^2 .

The cumulative distribution function, assumes the form

$$F(x) = \begin{cases} \frac{1}{2} \exp(\sqrt{2}(x - \mu)/\phi) & x < \mu, \\ 1 - \frac{1}{2} \exp(-\sqrt{2}(x - \mu)/\phi) & x \ge \mu. \end{cases}$$

The quantile function, is given by

$$F^{-1}(p) = \begin{cases} \mu + \frac{\phi}{\sqrt{2}} \log(2p) & p < 0.5, \\ \mu - \frac{\phi}{\sqrt{2}} \log(2(1-p)) & p \ge 0.5. \end{cases}$$

LaplaceFit 9

Value

dlaplace, plaplace, and qlaplace are respectively the density, distribution function and quantile function of the Laplace distribution. rlaplace generates random deviates drawn from the Laplace distribution, the length of the result is determined by n.

Author(s)

Felipe Osorio and Tymoteusz Wolodzko

References

Kotz, S., Kozubowski, T.J., Podgorski, K. (2001). *The Laplace Distributions and Generalizations*. Birkhauser, Boston.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*, 2nd Ed. Chapman & Hall, Boca Raton.

See Also

Distributions for other standard distributions and rmLaplace for the random generation from the multivariate Laplace distribution.

Examples

```
x <- rlaplace(1000)
## QQ-plot for Laplace data against true theoretical distribution:
qqplot(qlaplace(ppoints(1000)), x, main = "Laplace QQ-plot",
    xlab = "Theoretical quantiles", ylab = "Sample quantiles")
abline(c(0,1), col = "red", lwd = 2)</pre>
```

LaplaceFit

Estimation of location and scatter using the multivariate Laplace distribution

Description

Estimates the location vector and scatter matrix assuming the data came from a multivariate Laplace distribution.

Usage

```
LaplaceFit(x, data, subset, na.action, tol = 1e-6, maxiter = 200)
```

10 LaplaceFit

Arguments

x a	formula or a nun	eric matrix or	an object that can	be coerced to a numeric
-----	------------------	----------------	--------------------	-------------------------

matrix.

data an optional data frame (or similar: see model.frame), used only if x is a for-

mula. By default the variables are taken from environment(formula).

subset an optional expression indicating the subset of the rows of data that should be

used in the fitting process.

na.action a function that indicates what should happen when the data contain NAs.

tol the relative tolerance in the iterative algorithm.

maxiter maximum number of iterations. The default is 200.

Value

A list with class 'LaplaceFit' containing the following components:

call a list containing an image of the LaplaceFit call that produced the object.

center final estimate of the location vector.

Scatter final estimate of the scale matrix.

logLik the log-likelihood at convergence.

numIter the number of iterations used in the iterative algorithm.

weights estimated weights corresponding to the Laplace distribution.

distances estimated squared Mahalanobis distances.

Generic function print show the results of the fit.

References

Yavuz, F.G., Arslan, O. (2018). Linear mixed model with Laplace distribution (LLMM). *Statistical Papers* **59**, 271-289.

See Also

cov

```
fit <- LaplaceFit(stack.x)
fit

# covariance matrix
p <- fit$dims[2]
Sigma <- (4 * (p + 1)) * fit$Scatter
Sigma</pre>
```

mLaplace 11

	mLaplace	Multivariate Laplace distribution	
--	----------	-----------------------------------	--

Description

These functions provide the density and random number generation from the multivariate Laplace distribution.

Usage

```
dmLaplace(x, center = rep(0, nrow(Scatter)), Scatter = diag(length(center)), log = FALSE)
rmLaplace(n = 1, center = rep(0, nrow(Scatter)), Scatter = diag(length(center)))
```

Arguments

x vector or matrix of data.

n the number of samples requested.

center a $k \times 1$ vector giving the locations.

Scatter a $k \times k$ positive-definite dispersion matrix.

log logical; if TRUE, the logarithm of the density function is returned.

Details

The multivariate Laplace distribution is a multidimensional extension of the univariate symmetric Laplace distribution. There are multiple forms of the multivariate Laplace distribution. Here, a particular case of the multivarite power exponential distribution introduced by Gomez et al. (1998) is considered.

The multivariate Laplace distribution with location μ = center and Σ = Scatter has density

$$f(\boldsymbol{x}) = \frac{\Gamma(k/2)}{\pi^{k/2}\Gamma(k)2^{k+1}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}[(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})]^{1/2}\right\}.$$

The function rmLaplace is an interface to C routines, which make calls to subroutines from LA-PACK. The matrix decomposition is internally done using the Cholesky decomposition. If Scatter is not non-negative definite then there will be a warning message.

Value

If x is a matrix with n rows, then dmLaplace returns a $n \times 1$ vector considering each row of x as a copy from the multivariate Laplace.

If n = 1, then rmLaplace returns a vector of the same length as center, otherwise a matrix of n rows of random vectors.

12 simulate.lad

References

Fang, K.T., Kotz, S., Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.

Gomez, E., Gomez-Villegas, M.A., Marin, J.M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods* **27**, 589-600.

Examples

```
# dispersion parameters
Scatter <- matrix(c(1,.5,.5,1), ncol = 2)
Scatter

# generate the sample
y <- rmLaplace(n = 2000, Scatter = Scatter)

# scatterplot of a random bivariate Laplace sample with center
# vector zero and scale matrix 'Scatter'
par(pty = "s")
plot(y, xlab = "", ylab = "")
title("bivariate Laplace sample", font.main = 1)</pre>
```

simulate.lad

Simulate responses from lad models

Description

Simulate one or more responses from the distribution corresponding to a fitted 1ad object.

Usage

```
## S3 method for class 'lad'
simulate(object, nsim = 1, seed = NULL, ...)
```

Arguments

object an object representing a fitted model.

nsim number of response vectors to simulate. Defaults to 1.

seed an object specifying if and how the random number generator should be initial-

ized ('seeded'). For the "lad" method, either NULL or an integer that will be used in a call to set.seed before simulating the response vectors. If set, the value is saved as the "seed" attribute of the returned value. The default, NULL will not change the random generator state, and return .Random.seed as the "seed"

attribute, see 'Value'.

.. additional optional arguments.

spatial.median 13

Value

For the "lad" method, the result is a data frame with an attribute "seed". If argument seed is NULL, the attribute is the value of .Random. seed before the simulation was started.

Author(s)

Tymoteusz Wolodzko and Felipe Osorio

Examples

```
fm <- lad(stack.loss ~ ., data = stackloss)
sm <- simulate(fm, nsim = 4)</pre>
```

spatial.median

Computation of the generalized spatial median

Description

Computation of the generalized spatial median estimator as defined by Rao (1988).

Usage

```
spatial.median(x, data, subset, na.action, tol = 1e-6, maxiter = 200)
```

Arguments

X	a formula or a numeric matrix or an object that can be coerced to a numeric matrix.
data	an optional data frame (or similar: see model.frame), used only if x is a formula. By default the variables are taken from environment(formula).
subset	an optional expression indicating the subset of the rows of data that should be used in the fitting process.
na.action	a function that indicates what should happen when the data contain NAs.
tol	the relative tolerance in the iterative algorithm.
maxiter	maximum number of iterations. The default is 200.

Details

An interesting fact is that the generalized spatial median estimator proposed by Rao (1988) is the maximum likelihood estimator under the Kotz-type distribution discussed by Naik and Plungpongpun (2006). The generalized spatial median estimators are defined as $\hat{\mu}$ and $\hat{\Sigma}$ which minimize

$$\frac{n}{2}\log |\mathbf{\Sigma}| + \sum_{i=1}^n \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})},$$

simultaneously with respect to μ and Σ .

The function spatial.median follows the iterative reweighting algorithm of Naik and Plungpongpun (2006).

14 vcov.lad

Value

A list with class 'spatial.median' containing the following components:

call a list containing an image of the spatial median call that produced the object.

median final estimate of the location vector.

Scatter final estimate of the scale matrix.

logLik the log-likelihood at convergence.

numIter the number of iterations used in the iterative algorithm.

innerIter the total number of iterations used in the **inner** iterative algorithm.

weights estimated weights corresponding to the Kotz distribution.

distances estimated squared Mahalanobis distances.

Generic function print show the results of the fit.

References

Naik, D.N., Plungpongpun, K. (2006). A Kotz-type distribution for multivariate statistical inference. In: Balakrishnan, N., Sarabia, J.M., Castillo, E. (Eds) *Advances in Distribution Theory, Order Statistics, and Inference*. Birkhauser Boston, pp. 111-124.

Rao, C.R. (1988). Methodology based on the L1-norm in statistical inference. *Sankhya, Series A* **50**, 289-313.

See Also

```
cov, LaplaceFit
```

Examples

```
z <- spatial.median(stack.x)</pre>
```

vcov.lad

Calculate variance-covariance matrix from lad models

Description

Returns the variance-covariance matrix of the main parameters of a fitted model for lad objects. The "main" parameters of model correspond to those returned by coef, and typically do not contain the nuisance scale parameter.

Usage

```
## S3 method for class 'lad'
vcov(object, ...)
```

WH.Laplace 15

Arguments

object an object representing a fitted model.

... additional arguments for method functions.

Value

A matrix of the estimated covariances between the parameter estimates in the linear regression model. This should have row and column names corresponding to the parameter names given by the coef method.

WH.Laplace

Wilson-Hilferty transformation

Description

Returns the Wilson-Hilferty transformation for multivariate Laplace deviates.

Usage

WH.Laplace(x, center, Scatter)

Arguments

x object of class 'LaplaceFit' from which is extracted the estimated Maha-

lanobis distances of the fitted model. Also x can be a vector or matrix of data

with, say, p columns.

center mean vector of the distribution or data vector of length p. Not required if x have

class 'LaplaceFit'.

Scatter Scatter Matrix (p by p) of the distribution. Not required if x have class 'LaplaceFit'.

Details

Let T=D/(2p) be a Gamma distributed random variable, where \mathbb{D}^2 denotes the squared Mahalanobis distance defined as

$$D^2 = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}).$$

Thus, the Wilson-Hilferty transformation is given by

$$z = \frac{T^{1/3} - \left(1 - \frac{1}{9p}\right)}{\left(\frac{1}{9p}\right)^{1/2}}$$

and z is approximately distributed as a standard normal distribution. This is useful, for instance, in the construction of QQ-plots.

WH.Laplace

References

Osorio, F., Galea, M., Henriquez, C., Arellano-Valle, R. (2023). Addressing non-normality in multivariate analysis using the *t*-distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.

Terrell, G.R. (2003). The Wilson-Hilferty transformation is locally saddlepoint. *Biometrika* **90**, 445-453.

Wilson, E.B., Hilferty, M.M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America* **17**, 684-688.

```
Scatter <- matrix(c(1,.5,.5,1), ncol = 2)
Scatter

# generate the sample
y <- rmLaplace(n = 500, Scatter = Scatter)
fit <- LaplaceFit(y)
z <- WH.Laplace(fit)
par(pty = "s")
qqnorm(z, main = "Transformed distances QQ-plot")
abline(c(0,1), col = "red", lwd = 2)</pre>
```

Index

* datagen	lad.fit-methods,7
simulate.lad, 12	lad.fit.BR, 6, 7
* datasets	<pre>lad.fit.BR (lad.fit-methods), 7</pre>
ereturns, 3	lad.fit.EM, 6 , 7
* distribution	<pre>lad.fit.EM(lad.fit-methods), 7</pre>
Laplace, 8	Laplace, 8
mLaplace, 11	LaplaceFit, 9, <i>14</i>
* models	list, 6
confint.lad, 2	lm, 5, 7
LaplaceFit, 9	lsfit,4
simulate.lad, 12	
spatial.median, 13	mLaplace, 11
vcov.lad, 14	model.frame, 10, 13
* multivariate	1 1 (1 1) 0
LaplaceFit, 9	plaplace (Laplace), 8
mLaplace, 11	qlaplace (Laplace), 8
spatial.median, 13	qraprace (Laprace), 8
WH.Laplace, 15	rlaplace (Laplace), 8
* regression	rmLaplace, 9
11fit, 3	rmLaplace(mLaplace), 11
lad, 5	imeaprace (meaprace), 11
lad.fit,6	simulate.lad, 12
lad.fit-methods, 7	spatial.median, 13
.Random.seed, <i>12</i> , <i>13</i>	., , .
, ,	vcov, 2
coef, 2, 14, 15	vcov.lad, 14
confint.glm, 2	
confint.lad, 2	WH.Laplace, 15
confint.nls, 2	
cov, 10, 14	
Distributions, 9	
dlaplace (Laplace), 8	
dmLaplace (mLaplace), 11	
ereturns, 3	
11fit,3	
lad, 5, 7	
lad.fit, 6, 7	