

# Package ‘qrLMM’

March 21, 2017

**Type** Package

**Title** Quantile Regression for Linear Mixed-Effects Models

**Version** 1.3

**Date** 2017-03-21

**Author** Christian E. Galarza <cgalarza88@gmail.com> and Victor H. Lachos <hlachos@ime.unicamp.br>

**Maintainer** Christian E. Galarza <cgalarza88@gmail.com>

**Imports** matrixcalc, mvtnorm, ghyp, quantreg, psych, tcltk, ald

**Description** Quantile regression (QR) for Linear Mixed-Effects Models via the asymmetric Laplace distribution (ALD). It uses the Stochastic Approximation of the EM (SAEM) algorithm for deriving exact maximum likelihood estimates and full inference results for the fixed-effects and variance components. It also provides graphical summaries for assessing the algorithm convergence and fitting results.

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-03-21 20:44:27 UTC

## R topics documented:

qrLMM-package	2
Cholesterol	3
group.plots	4
Orthodont	5
QRLMM	6

<b>Index</b>	<b>11</b>
--------------	-----------

---

qrLMM-package

*Package for Quantile Regression for Linear Mixed-Effects Model*

---

## Description

This package contains a principal function that performs a quantile regression for a Linear Mixed-Effects Model using the Stochastic-Approximation of the EM Algorithm (SAEM) for an unique or a set of quantiles.

Exploiting the nice hierarchical representation of the ALD, our classical approach follows the Stochastic Approximation of the EM(SAEM) algorithm for deriving exact maximum likelihood estimates of the fixed-effects and variance components.

## Details

Package: qrLMM  
Type: Package  
Version: 1.0  
Date: 2014-12-19  
License: What license is it under?

## Author(s)

Christian E. Galarza <<cgalarza88@gmail.com>> and Victor H. Lachos <<hlachos@ime.unicamp.br>>

Maintainer: Christian E. Galarza <<cgalarza88@gmail.com>>

## References

Delyon, B., Lavielle, M. & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, pages 94-128.

Koenker, R., Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.* 94(3):1296-1309.

Yu, K. & Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4), 437-447.

Yu, K., & Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics-Theory and Methods*, 34(9-10), 1867-1879.

## See Also

[Orthodont](#), [Cholesterol](#), [QRLMM](#), [QRNLMM](#)

**Examples**

```
#See examples for the QRLMM function linked above.
```

---

Cholesterol	<i>Framingham cholesterol study</i>
-------------	-------------------------------------

---

**Description**

The Framingham cholesterol study generated a benchmark dataset (Zhang and Davidian, 2001) for longitudinal analysis to examine the role of serum cholesterol as a risk factor for the evolution of cardiovascular disease for 200 randomly selected subjects.

**Usage**

```
data(Cholesterol)
```

**Format**

This data frame contains the following columns:

`newid` a numeric vector indicating the subject on which the measurement was made. It represents the subject number in the sample.

`ID` a numeric vector indicating the subject on which the measurement was made. It represents the subject number in the population.

`cholst` cholesterol level for patient `newid`.

`sex` a dichotomous gender (0=female, 1=male).

`age` age of the patient in years.

`year` years elapsed since the start of the study to the current measurement.

**Source**

Zhang, D., & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57(3), 795-802.

**References**

<https://www.framinghamheartstudy.org/about-fhs/background.php>

## Examples

```
## Not run:
data(Cholesterol)
attach(Cholesterol)

y = cholst           #response
x = cbind(1,sex,age) #design matrix for fixed effects
z = cbind(1,year)   #design matrix for random effects

#A median regression
median_reg = QRLMM(y,x,newid,nj,MaxIter = 500)

## End(Not run)
```

---

group.plots

*Plot function for grouped data*

---

## Description

Functions for plotting a profiles plot for grouped data.

## Usage

```
group.plot(x,y,groups,...)
group.lines(x,y,groups,...)
group.points(x,y,groups,...)
```

## Arguments

y	the response vector of dimension $N$ where $N$ is the total of observations.
x	vector of longitudinal (repeated measures) covariate of dimension $N$ . For example: Time, location, etc.
groups	factor of dimension $N$ specifying the partitions of the data over which the random effects vary.
...	additional graphical arguments passed to <a href="#">matplot</a> . See <a href="#">par</a> .

## Author(s)

Christian E. Galarza <<cgalarza88@gmail.com>> and Victor H. Lachos <<hlachos@ime.unicamp.br>>

## See Also

[Orthodont](#), [Cholesterol](#), [QRLMM](#)

**Examples**

```
## Not run:
##A full profile plot for Soybean data

data(Soybean,package = "qrNLMM")
attach(Soybean)

group.plot(x = Time,y = weight,groups = Plot,type="b",
           main="Soybean profiles",xlab="time (days)",
           ylab="mean leaf weight (gr)")

#Profile plot by genotype

group.plot(x = Time[Variety=="P"],y = weight[Variety=="P"],
           groups = Plot[Variety=="P"],type="l",col="blue",
           main="Soybean profiles by genotype",xlab="time (days)",
           ylab="mean leaf weight (gr)")

group.lines(x = Time[Variety=="F"],y = weight[Variety=="F"],
            groups = Plot[Variety=="F"],col="black")

## End(Not run)
```

---

 Orthodont

*Growth curve data on an orthodontic measurement*


---

**Description**

The Orthodont data frame has 108 rows and 4 columns of the change in an orthodontic measurement over time for several young subjects.

**Format**

This data frame contains the following columns:

**distance** a numeric vector of distances from the pituitary to the pterygomaxillary fissure (mm). These distances are measured on x-ray images of the skull.

**age** a numeric vector of ages of the subject (yr).

**Subject** an ordered factor indicating the subject on which the measurement was made. The levels are labelled M01 to M16 for the males and F01 to F13 for the females. The ordering is by increasing average distance within sex.

**Sex** a factor with levels Male and Female

**Details**

Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head.

**Source**

Pinheiro, J. C. and Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York. (Appendix A.17)

Potthoff, R. F. and Roy, S. N. (1964), “A generalized multivariate analysis of variance model useful especially for growth curve problems”, *Biometrika*, **51**, 313–326.

**Examples**

```
## Not run:
data(Orthodont)
attach(Orthodont)

sex = c()
sex[Sex=="Male"]=0
sex[Sex=="Female"]=1

y = distance          #response
x = cbind(1,sex,age) #design matrix for fixed effects
z = cbind(1,age)     #design matrix for random effects

#A median regression
median_reg = QRLMM(y,x,z,Subject,MaxIter = 500)

## End(Not run)
```

---

QRLMM

*Quantile Regression for Linear Mixed-Effects Models*


---

**Description**

Performs a quantile regression for a LMEM using the Stochastic-Approximation of the EM Algorithm (SAEM) for an unique or a set of quantiles.

**Usage**

```
QRLMM(y,x,z,groups,p=0.5,precision=0.0001,MaxIter=300,M=10,cp=0.25,
      beta=NA,sigma=NA,Psi=NA,show.convergence=TRUE,CI=95)
```

**Arguments**

**y** the response vector of dimension  $N$  where  $N$  is the total of observations.

**x** design matrix for the fixed effects of dimension  $N \times d$  where  $d$  represents the number of fixed effects including the intercept, if considered.

**z** design matrix for the random effects of dimension  $N \times q$  where  $q$  represents the number of random effects.

groups	factor of dimension $N$ specifying the partitions of the data over which the random effects vary.
p	unique quantile or a set of quantiles related to the quantile regression.
precision	the convergence maximum error.
MaxIter	the maximum number of iterations of the SAEM algorithm. Default = 300.
M	Number of Monte Carlo simulations used by the SAEM Algorithm. Default = 10. For more accuracy we suggest to use $M=20$ .
cp	cut point ( $0 \leq cp \leq 1$ ) which determines the percentage of initial iterations with no memory.
beta	fixed effects vector of initial parameters, if desired.
sigma	dispersion initial parameter for the error term, if desired.
Psi	Variance-covariance random effects matrix of initial parameters, if desired.
show.convergence	if TRUE, it will show a graphical summary for the convergence of the estimates of all parameters for each quantile in order to assess the convergence.
CI	Confidence to be used for the Confidence Interval when a grid of quantiles is provided. Default=95.

## Details

This function considers a linear mixed-effects model defined as:

$$y_i = x_i * \beta_p + z_i * b_i + \epsilon_i;$$

where,  $x_i$  and  $z_i$  are the design matrices for the fixed and random effects respectively,  $\beta_p$  are the fixed effects (associated to the  $p$ -th quantile),  $b_i$  are the random (normal) effects and  $\epsilon_i$  is a random error (considered to be asymmetric Laplace).

This algorithm performs the SAEM algorithm proposed by Delyon et al. (1999), a stochastic version of the usual EM Algorithm deriving exact maximum likelihood estimates of the fixed-effects and variance components.

If the initial parameters are not provided, by default, the fixed effects parameter  $\beta$  and dispersion parameter  $\sigma$  will be the maximum Likelihood Estimates for an Asymmetric Laplace Distribution (obviating the random term). See Yu & Zhang (2005).

When a grid of quantiles is provided, a graphical summary with point estimates and Confidence Intervals for model parameters is shown and also a graphical summary for the convergence of these estimates (for each quantile), if show.convergence=TRUE.

If the convergence graphical summary shows that convergence has not be attained, it's suggested to increase  $M$  to 20, to increase the total number of iterations `MaxIter` to 500 or both.

About the cut point parameter `cp`, a number between 0 and 1 ( $0 \leq cp \leq 1$ ) will assure an initial convergence in distribution to a solution neighborhood for the first `cp*MaxIter` iterations and an almost sure convergence for the rest of the iterations. If you do not know how SAEM algorithm works, this parameter SHOULD NOT be changed.

This program uses progress bars that will close when the algorithm ends. They must not be closed before if not the algorithm will stop.

**Value**

The function returns a list with two objects

`conv`            A two elements list with the matrices `teta` and `se` containing the point estimates and standard error estimate for all parameters along all iterations.

The second element of the list is `res`, a list of 12 elements detailed as

<code>iter</code>	number of iterations.
<code>criteria</code>	attained criteria value.
<code>beta</code>	fixed effects estimates.
<code>weights</code>	random effects weights ( $b_i$ ).
<code>sigma</code>	scale parameter estimate for the error term.
<code>Psi</code>	Random effects variance-covariance estimate matrix.
<code>SE</code>	Standard Error estimates.
<code>table</code>	Table containing the inference for the fixed effects parameters.
<code>loglik</code>	Log-likelihood value.
<code>AIC</code>	Akaike information criterion.
<code>BIC</code>	Bayesian information criterion.
<code>HQ</code>	Hannan-Quinn information criterion.
<code>time</code>	processing time.

**Note**

If a grid of quantiles was provided, the result is a list of the same dimension where each element corresponds to each quantile as detailed above.

**Author(s)**

Christian E. Galarza <<cgalarza88@gmail.com>> and Victor H. Lachos <<hlachos@ime.unicamp.br>>

**References**

- Delyon, B., Lavielle, M. & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, pages 94-128.
- Yu, K., & Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics-Theory and Methods*, 34(9-10), 1867-1879.

**See Also**

[Orthodont](#), [Cholesterol](#), [QRNLMM](#)



**Examples**

```

#Using the Orthodontic distance growth data

data(Orthodont)
attach(Orthodont)

y = distance #response
x = cbind(1,c(rep(0,64),rep(1,44)),age) #design matrix for fixed effects
z = cbind(1,age) #design matrix for random effects
groups = Subject

## Not run:

#a quick median regression
model = QRLMM(y,x,z,groups,MaxIter=200)

beta = save$res$beta #fixed effects
weights = save$res$weight #random weights
nj = c(as.data.frame(table(groups))[2]) #obs per subject
fixed = x
random = rep(0,dim(x)[1]) #initializing random shift

for (j in 1:length(nj))
{
  z1=matrix(z[(sum(nj[1:j-1])+1):(sum(nj[1:j]))],,ncol=dim(z)[2])
  random[(sum(nj[1:j-1])+1):(sum(nj[1:j]))] = z1
}

pred = fixed + random #predictions
group.plot(age,pred,groups,type = "l")
group.points(age,distance,groups)

#####
#Fit a very quick regression for the three quartiles (Just for having an idea!)

QRLMM(y,x,z,groups,p = c(0.25,0.50,0.75),MaxIter=50,M=10)

#A full profile quantile regression (This might take some time)
QRLMM(y,x,z,groups,p = seq(0.05,0.95,0.05),MaxIter=300,M=10)

#A simple output example
-----
Quantile Regression for Linear Mixed Model
-----
Quantile = 0.75
Subjects = 27 ; Observations = 108 ; Balanced = 4
-----
Estimates
-----
- Fixed effects
Estimate Std. Error z value Pr(>|z|)
beta 1 17.08405 0.53524 31.91831 0

```

```
19
beta 2 2.15393 0.36929 5.83265 0
beta 3 0.61882 0.05807 10.65643 0
sigma = 0.38439
Random effects

i) weights
   ...

ii) Varcov matrix
z1 z2
z1 0.16106 -0.00887
z2 -0.00887 0.02839
-----
Model selection criteria
-----
Loglik AIC BIC HQ
Value -216.454 446.907 465.682 454.52
-----
Details
-----
Convergence reached? = FALSE
Iterations = 300 / 300
Criteria = 0.00381
MC sample = 10
Cut point = 0.25
Processing time = 7.590584 mins

## End(Not run)
```

# Index

- \*Topic **ALD**
    - QRLMM, 6
    - qrLMM-package, 2
  - \*Topic **SAEM**
    - QRLMM, 6
  - \*Topic **datasets**
    - Cholesterol, 3
    - Orthodont, 5
  - \*Topic **linear mixed models**
    - QRLMM, 6
  - \*Topic **mixed models**
    - group.plots, 4
  - \*Topic **nested data**
    - group.plots, 4
  - \*Topic **package**
    - qrLMM-package, 2
  - \*Topic **plot grouped data**
    - group.plots, 4
  - \*Topic **plot**
    - group.plots, 4
  - \*Topic **quantile regression**
    - QRLMM, 6
  - \*Topic **quantile**
    - QRLMM, 6
    - qrLMM-package, 2
- Cholesterol, 2, 3, 4, 8
- group.lines (group.plots), 4  
group.plot (group.plots), 4  
group.plots, 4  
group.points (group.plots), 4
- matplot, 4
- Orthodont, 2, 4, 5, 8
- par, 4
- QRLMM, 2, 4, 6  
qrLMM-package, 2  
QRNLMM, 2, 8