Package ‘qape’

September 8, 2020

Type Package

Title Quantile Absolute Prediction Error

Version 1.0

Date 2020-09-01

Author Alicja Wolny-Dominiak, Tomasz Zadlo

Maintainer Alicja Wolny-Dominiak <alicja.wolny-dominiak@ue.katowice.pl>

Imports lme4, Matrix, mvtnorm, plyr, dplyr

Depends R (>= 3.5.0)

Description Estimates Quantile Absolute Prediction Error using bootstrap procedures. The residual, parametric and double bootstrap is used.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2020-09-08 08:20:02 UTC

R topics documented:

bootPar 2
bootRes 4
correction 6
corrRancomp 7
corrRanef 8
doubleBoot 9
EBLUP 12
ebpLMMne 16
EmpCM 20
EstCM 21
invData 22
lwzl 22
plugInLMM 23
Zfun 27

Index **29**

`bootPar` *Parametric bootstrap estimators of prediction accuracy*

Description

The function computes values of parametric bootstrap estimators of RMSE and QAPE prediction accuracy measures.

Usage

```
bootPar(predictor, B, p)
```

Arguments

`predictor`	one of objects: EBLUP, ebpLMMne or plugInLMM.
`B`	number of iterations in the bootstrap procedure.
`p`	orders of quantiles in the QAPE.

Details

We use bootstrap model presented by Chatterjee, Lahiri and Li (2008) p. 1229 but assumed for all population elements. Vectors of random effects and random components are generated from the multivariate normal distribution where REML estimates of model parameters are used. Random effects are generated for all population elements even for subsets with zero sample sizes (for which random effects are not estimated). We use the MSE estimator defined as the mean of squared bootstrap errors considered by Rao and MOlina (2015) p. 141 and given by equation (6.2.22). The QAPE is a quantile of absolute prediction error which means that at least p100% of realizations of absolute prediction errors are smaller or equal to QAPE. It is estimated as a quantile of absolute bootstrap errors as proposed by Zadlo (2017) in Section 2.

Value

`estQAPE`	estimated value/s of QAPE - number of rows is equal the number of orders of quantiles to be considered (declared in *p*), number of columns is equal the number of predicted characteristics (declared in in *thetaFun*).
`estRMSE`	estimated value/s of RMSE (more than one value is computed if in *thetaFun* more than one population characteristic is defined).
`predictorSim`	bootstrapped values of the predictor/s.
`thetaSim`	bootstrapped values of the predicted population or subpopulation characteristic/s.
`YSim`	simulated values of the (possibly tranformed) variable of interest.
`error`	differences between bootstrapped values of the predictor/s and bootstrapped values of the predicted characteristic/s.

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Butar, B. F., Lahiri, P. (2003) On measures of uncertainty of empirical Bayes small-area estimators, Journal of Statistical Planning and Inference, Vol. 112, pp. 63-76.

2. Chatterjee, S., Lahiri, P. Li, H. (2008) Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models, Annals of Statistics, Vol. 36 (3), pp. 1221-1245.

3. Rao, J.N.K. and Molina, I. (2015) Small Area Estimation. Second edition, John Wiley & Sons, New Jersey.

4. Zadlo T. (2017), On asymmetry of prediction errors in small area estimation, Statistics in Transition, 18 (3), 413-432

Examples

```
library(lme4)
library(Matrix)
library(mvtnorm)


data(invData)
#data from one period are considered:
invData2018<-invData[invData$year == 2018,]
attach(invData2018)

N=nrow(invData2018) #population size

con=rep(1,N)
con[c(379,380)]<-0 # last two population elements are not observed

YS=(investments[con==1]) # log-transformed values
backTrans <- function(x) x # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
random.part <- '((1|NUTS2)+((newly_registered-1)|NUTS2))'

reg=invData2018[, - which(names(invData2018) == 'investments')]
weights=rep(1,N) #homoscedastic random components

# Characteristics to be predicted:
# values of the variable for last two population elements
thetaFun <- function(x) {x[c(379,380)]}
set.seed(123456)
```

```
# Predicted values of quartiles
# in the following subpopulation: NUTS4type==2
# in the following time period: year==2018

predictor=plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)
predictor$thetaP

# Estimation of prediction accuracy
est_accuracy=bootPar(predictor, 10, c(0.75,0.9))

# Estimation of prediction RMSE
est_accuracy$estRMSE

# Estimation of prediction QAPE
est_accuracy$estQAPE
```

---

bootRes *Residual bootstrap estimators of prediction accuracy*

Description

The function computes values of residual bootstrap estimators of RMSE and QAPE prediction accuracy measures.

Usage

```
bootRes(predictor, B, p, correction)
```

Arguments

predictor	one of objects: EBLUP, ebpLMMne or plugInLMM.
B	number of iterations in the bootstrap procedure.
p	orders of quantiles in the QAPE.
correction	logical. If TRUE, both boostrapped random effects and random components are tranformed to avoid the problem of underdispersion of residual bootstrap distributions (see Details).

Details

Residual bootstrap considered by Carpener, Goldstein and Rasbash (2003), Chambers and Chandra (2013) and Thai et al. (2013) is used. To built one bootstrap realization of the population vector of the variable of interest: (i) from the sample vector of predicted random components the simple random sample with replacement of population size is drawn at random, (ii) from the vector of predicted random effects the simple random sample with replacement of size equal the number of random effects in the whole population is drawn at random. If *correction* is *TRUE*, then predicted random effects are transformed as described in Carpener, Goldstein and Rasbash (2003) in Section 3.2 and predicted random components as presented in Chambers and Chandra (2013) in Section

2.2. We use the MSE estimator defined as the mean of squared bootstrap errors considered by Rao and Molina (2015) p. 141 given by equation (6.2.22). The QAPE is a quantile of absolute prediction error which means that at least p100% of realizations of absolute prediction errors are smaller or equal to QAPE. It is estimated as a quantile of absolute bootstrap errors as proposed by Zadlo (2017) in Section 2.

Value

`estQAPE`	estimated value/s of QAPE - number of rows is equal the number of orders of quantiles to be considered (declared in *p*), number of columns is equal the number of predicted characteristics (declared in in *thetaFun*).
`estRMSE`	estimated value/s of RMSE (more than one value is computed if in *thetaFun* more than one population characteristic is defined).
`predictorSim`	bootstrapped values of the predictor/s.
`thetaSim`	bootstrapped values of the predicted population or subpopulation characteristic/s.
`YSim`	simulated values of the (possibly tranformed) variable of interest.

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society: Series C (Applied Statistics), 52, 431-443.

2. Chambers, R. and Chandra, H. (2013) A Random Effect Block Bootstrap for Clustered Data, Journal of Computational and Graphical Statistics, 22(2), 452-470.

3. Thai, H.-T., Mentre, F., Holford, N.H., Veyrat-Follet, C. and Comets, E. (2013), A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharmaceutical Statistics, 12, 129-140.

Examples

```
library(lme4)
library(Matrix)
library(mvtnorm)


data(invData)
#data from one period are considered:
invData2018<-invData[invData$year == 2018,]
attach(invData2018)
```

```
N=nrow(invData2018) #population size

con=rep(1,N)
con[c(379,380)]<-0 # last two population elements are not observed

YS=(investments[con==1]) # log-transformed values
backTrans <- function(x) x # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
random.part <- '((1|NUTS2)+((newly_registered-1)|NUTS2))'

reg=invData2018[, - which(names(invData2018) == 'investments')]
weights=rep(1,N) #homoscedastic random components

# Characteristics to be predicted:
# values of the variable for last two population elements
thetaFun <- function(x) {x[c(379,380)]}
set.seed(123456)

# Predicted values of quartiles
# in the following subpopulation: NUTS4type==2
# in the following time period: year==2018

predictor=plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)
predictor$thetaP

# Estimation of prediction accuracy
est_accuracy=bootRes(predictor, 10, c(0.75,0.9), correction=TRUE)

# Estimation of prediction RMSE
est_accuracy$estRMSE

# Estimation of prediction QAPE
est_accuracy$estQAPE
```

`correction` *Correction term for predicted random effects*

Description

The function computes the list of matrices used to correct predicted random effects as presented in Carpenter, Goldstein and Rasbash (2003) in Section 3.2 to avoid the problem of underdispersion of residual bootstrap distributions.

Usage

```
correction(model)
```

Arguments

`model` *lmer*) object.

Value

a list of square matrices used to correct predicted random effects. The length of the list is equal the number of grouping variables used in case of random effects. Each matrix is of order equal the number of random effects at the considered level of grouping.

Author(s)

Tomasz Zadlo

References

Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society: Series C (Applied Statistics), 52, 431-443.

Examples

```
library(lme4)
data(invData)
attach(invData)
model=lmer(investments~newly_registered+((newly_registered)|NUTS4))
correction(model)
```

`corrRancomp` *Correction of predicted random components*

Description

The function computes the corrected predicted random components as presented in Chambers and Chandra (2013) in Section 2.2 to avoid the problem of underdispersion of residual bootstrap distributions.

Usage

```
corrRancomp(model)
```

Arguments

`model`	*lmer*) object .

Value

the vector of corrected predicted random components.

Author(s)

Tomasz Zadlo

References

Chambers, R. and Chandra, H. (2013) A Random Effect Block Bootstrap for Clustered Data, Journal of Computational and Graphical Statistics, 22(2), 452-470.

Examples

```
library(lme4)
data(invData)
attach(invData)
model=lmer(investments~newly_registered+(1|NUTS4))
corrRancomp(model)
```

---

corrRanef *Correction of predicted random effects*

Description

The function computes the corrected predicted random effects as presented in Carpenter, Goldstein and Rasbash (2003) in Section 3.2 to avoid the problem of underdispersion of residual bootstrap distributions.

Usage

```
corrRanef(model)
```

Arguments

`model`	*lmer*) object .

Value

a list of corrected predicted random effects (of the same form as *ranef(model)*)

Author(s)

Tomasz Zadlo

References

Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society: Series C (Applied Statistics), 52, 431-443.

Examples

```
library(lme4)
data(invData)
attach(invData)
model=lmer(investments~newly_registered+(1|NUTS4))
corrRanef(model)
```

`doubleBoot` *Double bootstrap estimators of prediction accuracy*

Description

The function computes values of double bootstrap estimators of MSE and QAPE prediction accuracy measures.

Usage

```
doubleBoot(predictor, B1, B2, p, q)
```

Arguments

`predictor`	one of objects: EBLUP, ebpLMMne or plugInLMM.
`B1`	number of first-level bootstrap iterations.
`B2`	number of second-level bootstrap iterations.
`p`	orders of quantiles in the QAPE.
`q`	estimator bounds assumed for *estMSE_db_1_EF* and *estMSE_db_telesc_EF* (which are corrected versions of *estMSE_db_1* and *estMSE_db_telesc*)

Details

Double-bootstrap method considered by Hall and Maiti (2006) and Erciulescu and Fuller (2013) is used. Vectors of random effects and random components are generated from the multivariate normal distribution and REML estimates of model parameters are used. Random effects are generated for all population elements even for subsets with zero sample sizes (for which random effects are not estimated). Double-bootstrap MSE estimator presented in Hall and Maiti (2006) and Erciulescu and Fuller (2013) are taken into account. The QAPE is a quantile of absolute prediction error which means that at least p100% of realizations of absolute prediction errors are smaller or equal to QAPE.

Value

`estMSE_param`	value/s of the parametric bootstrap MSE estimator. More than one value is computed if in *thetaFun* more than one population characteristic is defined.
`estMSE_db_B2`	value/s of the double bootstrap MSE estimator computed as the difference of doubled value of *estMSE_param* and the second-level MSE estimator based on B2 iterations. More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_B2_WDZ`
: value/s of the double bootstrap MSE estimator computed as the mean of squared first-level boostraped errors, each corrected by the mean of squared second-level boostraped errors based on B2 iterations (where correction is made only if their difference is non-negative). More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_B2_HM`
: value/s of the double bootstrap MSE estimator proposed by Hall and Maiti (2006) equation (2.17). More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_1`
: value/s of the double bootstrap MSE estimator computed as the difference of doubled value of *estMSE_param* and the second-level MSE estimator based on B2=1 iteration. More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_1_WDZ`
: value/s of the double bootstrap MSE estimator computed as the mean of squared first-level boostraped errors, each corrected by the squared second-level boostraped error based on 1 iteration (where correction is made only if their difference is non-negative). More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_1_EF`
: value/s of the double bootstrap MSE estimator proposed by Erciulescu and Fuller (2014) given by equation (13) with correction (17), where the bound for the correction is declared as *q*. More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_telesc`
: value/s of the telescoping double bootstrap MSE estimator proposed by Erciulescu and Fuller (2014) given by equation (15). More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_telesc_WDZ`
: value/s of the double bootstrap MSE estimator computed as the mean of the sums of the following elements: squared first-level boostraped error, squared first-level boostrap error for the next iteration and the opposite of second-level boostraped error based on 1 iteration (but negative sums are replaced by squared first-level boostraped error). More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estMSE_db_telesc_EF`
: value/s of the telescoping double bootstrap MSE estimator proposed by Erciulescu and Fuller (2014) given by equation (15) with correction (17), where the bound for the correction is declared as *q*. More than one value is computed if in *thetaFun* more than one population characteristic is defined.

`estQAPE_param=estQAPE`
: value/s of parametric bootstrap estimator of QAPE (Quantile of Absolute Prediction Errors) given by a quantile of absolute parametric bootstrap errors. Number of rows is equal the number of orders of quantiles to be considered (declared in *p*), number of columns is equal the number of predicted characteristics (declared in in *thetaFun*).

estQAPE_db_B2 value/s of double-bootstrap estimator of QAPE (Quantile of Absolute Prediction Errors) given by a quantile of square roots of squared first-level boostraped errors, each corrected by the mean of squared second-level boostraped errors based on B2 iterations (where correction is made only if their difference is non-negative). Number of rows is equal the number of orders of quantiles to be considered (declared in *p*), number of columns is equal the number of predicted characteristics (declared in in *thetaFun*).

estQAPE_db_1 value/s of double-bootstrap estimator of QAPE (Quantile of Absolute Prediction Errors) given by a quantile of square roots of squared first-level boostraped errors, each corrected by the squared second-level boostraped error based on 1 iteration (where correction is made only if their difference is non-negative). Number of rows is equal the number of orders of quantiles to be considered (declared in *p*), number of columns is equal the number of predicted characteristics (declared in in *thetaFun*).

estQAPE_db_telesc
value/s of double-bootstrap estimator of QAPE (Quantile of Absolute Prediction Errors) given by a quantile of square roots of the sums of the following elements: squared first-level boostraped error, squared first-level boostrap error for the next iteration and the opposite of second-level boostraped error based on 1 iteration (but negative sums are replaced by squared first-level boostraped error). Number of rows is equal the number of orders of quantiles to be considered (declared in *p*), number of columns is equal the number of predicted characteristics (declared in in *thetaFun*).

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Erciulescu, A. L. and Fuller, W. A. (2013) Parametric Bootstrap Procedures for Small Area Prediction Variance. JSM 2014 - Survey Research Methods Section, pp. 3307-3318.

2. Hall, P. and Maiti, T. (2006) On Parametric Bootstrap Methods for Small Area Prediction. Journal of the Royal Statistical Society. Series B, 68(2), 221-238.

Examples

```
data(invData)
#data from one period are considered:
invData2018 <- invData[invData$year == 2018,]
attach(invData2018)

N <- nrow(invData2018) #population size

con <- rep(1,N)
con[c(379,380)]<-0 # last two population elements are not observed
```

```
YS <- (investments[con==1]) # log-transformed values
backTrans <- function(x) x # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
random.part <- '((1|NUTS2)+((newly_registered-1)|NUTS2))'

reg=invData2018[, - which(names(invData2018) == 'investments')]
weights=rep(1,N) #homoscedastic random components

# Characteristics to be predicted:
# values of the variable for last two population elements
thetaFun <- function(x) {x[c(379,380)]}
set.seed(123456)

# Predicted values of quartiles
# in the following subpopulation: NUTS4type==2
# in the following time period: year==2018

predictor <- plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)
predictor$thetaP


# Estimation of prediction accuracy
# (q=0.77 is assumed below as in Erciulescu and Fuller (2014) eq. (17))
doubleBoot(predictor, 3, 3, c(0.5,0.9), 0.77)
```

EBLUP — *Empirical Best Linear Unbiased Predictor*

Description

The function computes the value of the EBLUP of the linear combination of the variable of interest under linear mixed model estimated using REML.

Usage

```
EBLUP(YS, fixed.part, random.part, reg, con, gamma, weights)
```

Arguments

YS	values of the variable of interest observed in the sample.
fixed.part	fixed-effects terms declared as in *lmer* object.
random.part	random-effects terms declared as in *lmer* object.
reg	the population matrix of auxiliary variables named in *fixed.part* and *random.part*.
con	the population 0-1 vector with 1s for elements in the sample and 0s for elements which are not in the sample.

`gamma` the population vector which transpose multiplied by the population vector of the variable of interest gives the predicted characteristic. For example, if *gamma* is the population vector on 1s, the sum of the values of the variable of interest in the whole dataset is predicted.

`weights` the population vector of weights, defined as in *lmer* object, allowing to include heteroscedasticity of random components in the mixed linear model.

Details

The function computes the value of the EBLUP of the linear combination of the variable of interest based on the formula (21) in Zadlo (2017) (see Remark 5.1 in the paper for further explanations). Predicted values for unsampled population elements in subsets for which random effects are not observed in the sample are computed based only on fixed effects.

Value

The function returns a list with the following objects:

`fixed.part` the fixed part of the formula of model.

`random.part` the random part of the formula of model.

`thetaP` the value of the predictor.

`beta` the estimated vector of fixed effects.

`Xbeta` the product of two matrices: the population model matrix of auxiliary variables X and the estimated vector of fixed effects.

`sigma2R` the estimated variance parameter of the distribution of random components.

`R` the estimated covariance matrix of random components for sampled elements.

`G` the estimated covariance matrix of random effects.

`model` the formula of the model (as in *lmer* object).

`mEst` *lmer* object with the estimated model.

`YS` the sample vector of the variable of interest.

`reg` the population matrix of auxiliary variables named in *fixed.part* and *random.part*.

`con` the population 0-1 vector with 1s for elements in the sample and 0s for elements which are not in the sample.

`regS` the sample matrix of auxiliary variables named in *fixed.part* and *random.part*.

`regR` the matrix of auxiliary variables named in *fixed.part* and *random.part* for population elements which are not observed in the sample.

`gamma` the population vector which transpose multiplied by the population vector of the variable of interest gives the predicted characteristic.

`gammaS` the subvector of *gamma* for sampled elements.

`gammaR` the subvector of *gamma* for population elements which are not observed in the sample.

`weights` the population vector of weigts, defined as in *lmer* object, allowing to include the heteroscedasticity of random components in the mixed linear model.

Z	the population model matrix of auxiliary variables associated with random effects.
X	the population model matrix of auxiliary variables associated with fixed effects.
ZS	the submatrix of Z matrix where the number of rows equals the number of sampled elements and the number of columns equals the number of estimated random effects.
XR	the submatrix of X matrix (with the same number of columns) for population elements which are not observed in the sample.
ZR	the submatrix of Z matrix where the number of rows equals the number of population elements which are not observed in the sample and the number of columns equals the number of estimated random effects.
eS	the sample vector of estimated random components.
vS	the estimated vector of random effects.

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Henderson, C.R. (1950) Estimation of Genetic Parameters (Abstract). Annals of Mathematical Statistics 21, 309-310.
2. Royall, R.M. (1976) The Linear Least Squares Prediction Approach to Two-Stage Sampling. Journal of the American Statistical Association 71, 657-473.
3. Zadlo, T. (2017) On prediction of population and subpopulation characteristics for future periods, Communications in Statistics - Simulation and Computation 461(10), 8086-8104.

Examples

```
library(lme4)
library(Matrix)


### Prediction of the subpopulation mean based on the cross-sectional data

data(invData)
#data from one period are considered:
invData2018<-invData[invData$year == 2018,]
attach(invData2018)

N=nrow(invData2018) #population size
n=100 # sample size
#subpopulation of interest: NUTS4type==2
Nd=sum(NUTS4type==2) #subpopulation size

set.seed(123456)
sampled_elements=sample(N,n)
```

```
con=rep(0,N)
con[sampled_elements]=1 #elements in the sample
YS=investments[sampled_elements]
fixed.part <- 'newly_registered'
random.part <- '(newly_registered| NUTS2)'
reg=invData2018[, - which(names(invData2018) == 'investments')]

gamma=rep(0,N)
gamma[NUTS4type==2]=1/Nd

weights=rep(1,N) #homoscedastic random components

#Predicted value of the mean in the following subpopulation: NUTS4type==2
EBLUP(YS, fixed.part, random.part, reg, con, gamma, weights)$thetaP

#All results
EBLUP(YS, fixed.part, random.part, reg, con, gamma, weights)

detach(invData2018)

##########################################################

### Prediction of the subpopulation total based on longitudinal data

data(invData)
attach(invData)

N=nrow(invData[(year==2013),]) #population size in the first period
n=38 # sample size in the first period
#subpopulation and time period of interest: NUTS2=='02' & year==2018
Ndt=sum(NUTS2=='02' & year==2018) #subpopulation size in the period of interest

set.seed(123456)
sampled_elements_in_2013=sample(N,n)
con2013=rep(0,N)
con2013[sampled_elements_in_2013]=1 #elements in the sample in 2013

#balanced panel sample - the same elements in all 6 periods:
con=rep(con2013,6)

YS=investments[con==1]
fixed.part <- 'newly_registered'
random.part <- '(newly_registered  | NUTS4)'
reg=invData[, - which(names(invData) == 'investments')]

gamma=rep(0,nrow(invData))
gamma[NUTS2=='02' & year==2018]=1

weights=rep(1,nrow(invData)) #homoscedastic random components

# Predicted value of the total
# in the following subpopulation: NUTS4type==2
# in the following time period: year==2018
```

```
EBLUP(YS, fixed.part, random.part, reg, con, gamma, weights)$thetaP

#All results
EBLUP(YS, fixed.part, random.part, reg, con, gamma, weights)

detach(invData)
```

---

ebpLMMne *Empirical Best Predictor based on nested error linear mixed model*

Description

The function computes the value of the EBP under nested error linear mixed model estimated using REML aaumed for possibly transformed variable of interest.

Usage

```
ebpLMMne(YS, fixed.part, division, reg, con, backTrans, thetaFun, L)
```

Arguments

YS	values of the variable of interest (already transformed if necessary) observed in the sample and used in the model as the dependent variable.
fixed.part	fixed-effects terms declared as in *lmer* object.
division	the variable dividing the population dataset into subsets (the nested error linear mixed model with 'division'-specific random components is estimated).
reg	the population matrix of auxiliary variables named in *fixed.part* and *random.part*.
con	the population 0-1 vector with 1s for elements in the sample and 0s for elements which are not in the sample.
backTrans	back-transformation function of the variable of interest (e.g. if YS is log-tranformed, then backTrans <- function(x) exp(x)).
thetaFun	the predictor function (e.g. mean or sd)
L	the number of iterations used to compute the value of the predictor.

Details

The function computes the value of the EBP based on the algorithm described in Molina and Rao (2010) in Section 4.

Value

The function returns a list with the following objects:

- `thetaP` — the value/s of the predictor (more than one value is computed if in *thetaFun* more than one population characteristic is defined).
- `fixed.part` — the fixed part of the formula of model.
- `random.part` — the random part of the formula of model.
- `division` — the variable dividing the population dataset into subsets (the nested error linear mixed model with 'division'-specific random components is estimated).
- `thetaFun` — the function of the population values of the variable of interest (on the original scale) which defines at least one population or subpopulation characteristic to be predicted.
- `backTrans` — back-transformation function of the variable of interest (e.g. if YS is log-tranformed, then backTrans <- function(x) exp(x).
- `L` — the number of iterations used to compute the value of the predictor.
- `beta` — the estimated vector of fixed effects.
- `Xbeta` — the product of two matrices: the population model matrix of auxiliary variables X and the estimated vector of fixed effects.
- `sigma2R` — the estimated variance parameter of the distribution of random components.
- `R` — the estimated covariance matrix of random components for sampled elements.
- `G` — the estimated covariance matrix of random effects.
- `model` — the formula of the model (as in *lmer* object).
- `mEst` — *lmer* object with the estimated model.
- `YS` — values of the variable of interest (already transformed if necessary) observed in the sample and used in the model as the dependent variable.
- `reg` — the population matrix of auxiliary variables named in *fixed.part* and *random.part*.
- `con` — the population 0-1 vector with 1s for elements in the sample and 0s for elements which are not in the sample.
- `regS` — the sample matrix of auxiliary variables named in *fixed.part* and *random.part*.
- `regR` — the matrix of auxiliary variables named in *fixed.part* and *random.part* for unsampled population elements.
- `weights` — the population vector of weigts, defined as in *lmer* object, allowing to include the heteroscedasticity of random components in the mixed linear model.
- `Z` — the population model matrix of auxiliary variables associated with random effects.
- `X` — the population model matrix of auxiliary variables associated with fixed effects.
- `ZS` — the submatrix of Z matrix where the number of rows equals the number of sampled elements and the number of columns equals the number of estimated random effects.
- `XR` — the submatrix of X matrix (with the same number of columns) for unsampled population elements.

ZR	the submatrix of Z matrix where the number of rows equals the number of unsampled population elements and the number of columns equals the number of estimated random effects.
eS	the sample vector of estimated random components.
vS	the estimated vector of random effects.

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Chwila, A., Zadlo, T. (2019) On properties of empirical best predictors. Communications in Statistics - Simulation and Computation, available online: https://doi.org/10.1080/03610918.2019.1649422
2. Molina, I., Rao, J.N.K. (2010) Small area estimation of poverty indicators. Canadian Journal of Statistics 38(3), 369-385.
3. Zadlo, T. (2017). On prediction of population and subpopulation characteristics for future periods, Communications in Statistics - Simulation and Computation 461(10) 8086-8104.

Examples

```
library(lme4)
library(Matrix)


### Prediction of the subpopulation median and subpopulation standard deviation
### based on the cross-sectional data

data(invData)
#data from one period are considered:
invData2018<-invData[invData$year == 2018,]
attach(invData2018)

N=nrow(invData2018) #population size
n=100 # sample size

set.seed(123456)
sampled_elements=sample(N,n)
con=rep(0,N)
con[sampled_elements]=1 #elements in the sample
YS=log(investments[sampled_elements]) # log-transformed values
backTrans <- function(x) exp(x) # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
division='NUTS2' # NUTS2-specific random effects are taken into account
reg=invData2018[, - which(names(invData2018) == 'investments')]


# Characteristics to be predicted - median and standard deviation
# in subpopulation of interest: NUTS4type==2
```

```
thetaFun <- function(x) {c(median(x[NUTS4type==2]),sd(x[NUTS4type==2]))}

L=5

#Predicted values of the median and the standard deviation
# in the following subpopulation: NUTS4type==2

ebpLMMne(YS, fixed.part, division, reg, con, backTrans, thetaFun, L)$thetaP

#All results
ebpLMMne(YS, fixed.part, division, reg, con, backTrans, thetaFun, L)

detach(invData2018)

##########################################################

### Prediction of the subpopulation quartiles based on longitudinal data

data(invData)
attach(invData)

N=nrow(invData[(year==2013),]) #population size in the first period
n=38 # sample size in the first period
#subpopulation and time period of interest: NUTS2=='02' & year==2018
#Ndt=sum(NUTS2=='02' & year==2018) #subpopulation size in the period of interest

set.seed(123456)
sampled_elements_in_2013=sample(N,n)
con2013=rep(0,N)
con2013[sampled_elements_in_2013]=1 #elements in the sample in 2013

#balanced panel sample - the same elements in all 6 periods:
con=rep(con2013,6)

YS=log(investments[con==1]) #log-transformed values
backTrans <- function(x) exp(x) # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
division='NUTS4' # NUTS4-specific random effects are taken into account
reg=invData[, - which(names(invData) == 'investments')]
thetaFun <- function(x) {quantile(x[NUTS2=='02' & year==2018],probs =c(0.25,0.5,0.75))}

L=5

# Predicted values of quartiles
# in the following subpopulation: NUTS4type==2
# in the following time period: year==2018

ebpLMMne(YS, fixed.part, division, reg, con, backTrans, thetaFun, L)$thetaP

#All results
predictor=ebpLMMne(YS, fixed.part, division, reg, con, backTrans, thetaFun, L)
```

```
detach(invData)
```

EmpCM *Emprirical covariance matrix of predicted random effects*

Description

A list of empirical covariance matrices of predicted random effects, where the length of the list equals the number of grouping variables used to define random effects as described in Carpenter, Goldstein and Rasbash (2003) in Section 3.2 and in Thai et al. (2013) in Section 2.3.3.

Usage

```
EmpCM(model)
```

Arguments

model	*lmer*) object.

Value

list of empirical covariance matrices of predicted random effects.

Author(s)

Tomasz Zadlo

References

1. Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society: Series C (Applied Statistics), 52, 431-443.

2. Thai, H.-T., Mentre, F., Holford, N.H., Veyrat-Follet, C. and Comets, E. (2013), A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharmaceutical Statistics, 12, 129-140.

Examples

```
library(lme4)
data(invData)
attach(invData)
model=lmer(investments~newly_registered+(1|NUTS4))
EmpCM(model)
```

EstCM — *Estimated covariance matrix of predicted random effects*

Description

A list of estimated covariance matrices of predicted random effects, where the length of the list equals the number of grouping variables used to define random effects as described in Carpenter, Goldstein and Rasbash (2003) in Section 3.2 and in Thai et al. (2013) in Section 2.3.3.

Usage

```
EstCM(model)
```

Arguments

`model`	*lmer*) object.

Value

list of estimated covariance matrices of predicted random effects.

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society: Series C (Applied Statistics), 52, 431-443.

2. Thai, H.-T., Mentre, F., Holford, N.H., Veyrat-Follet, C. and Comets, E. (2013), A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharmaceutical Statistics, 12, 129-140.

Examples

```
library(lme4)
data(invData)
attach(invData)
model=lmer(investments~newly_registered+(1|NUTS4))
EstCM(model)
```

invData	*invData*

Description

A data frame with 2280 observations on the following 8 variables.

Usage

```
data("invData")
```

Source

Polsh Statistical Office GUS

Examples

```
data(invData)
## maybe str(invData) ; plot(invData) ...
```

lwzl	*Bootstrap sample of predicted random effects*

Description

The function draw at random a simple random sample with replacement from predicted random effects, where the sample size is equal the number of random effects in the whole population.

Usage

```
lwzl(listRanef, reg)
```

Arguments

listRanef	*ranef(model)* object where *model* is an *lmer*) object.
reg	the population matrix of auxiliary variables named in *fixed.part* and *random.part*.

Value

tablwzl	a vector of a simple random sample with replacement from predicted random effects, where the sample size is equal the number of random effects in the whole population.
llwzl	##

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

1. Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement. Journal of the Royal Statistical Society: Series C (Applied Statistics), 52, 431-443.

2. Chambers, R. and Chandra, H. (2013) A Random Effect Block Bootstrap for Clustered Data, Journal of Computational and Graphical Statistics, 22(2), 452-470.

3. Thai, H.-T., Mentre, F., Holford, N.H., Veyrat-Follet, C. and Comets, E. (2013), A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharmaceutical Statistics, 12, 129-140.

Examples

```
data(invData)
#data from one period are considered:
invData2018 <- invData[invData$year == 2018,]
attach(invData2018)
N=nrow(invData2018) #population size
n=100 # sample size

set.seed(123456)
sampled_elements=sample(N,n)
reg <- invData2018[, - which(names(invData2018) == 'investments')]

detach(invData2018)
invData2018sample <- invData2018[sampled_elements,]
attach(invData2018sample)

model <- lme4::lmer(investments~newly_registered+(newly_registered| NUTS2))
raneff <- lme4::ranef(model)

lwzl(raneff, reg)$tablwzl
lwzl(raneff, reg)$llwzl
```

---

plugInLMM *PLUG-IN predictor based on linear mixed model*

Description

The function computes the value of the plug-in predictor under nested error linear mixed model estimated using REML assumed for possibly transformed variable of interest.

Usage

```
plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)
```

Arguments

`YS`	values of the variable of interest (already transformed if necessary) observed in the sample and used in the model as the dependent variable.
`fixed.part`	fixed-effects terms declared as in *lmer* object.
`random.part`	the random part of the formula of model.
`reg`	the population matrix of auxiliary variables named in *fixed.part* and *random.part*.
`con`	the population 0-1 vector with 1s for elements in the sample and 0s for elements which are not in the sample.
`weights`	the population vector of weigts, defined as in *lmer* object, allowing to include the heteroscedasticity of random components in the mixed linear model.
`backTrans`	back-transformation function of the variable of interest (e.g. if YS is log-tranformed, then backTrans <- function(x) exp(x)).
`thetaFun`	the predictor function (e.g. mean or sd)

Details

The function computes the value of the plug-in estimator in two steps as presented by Chwila and Zadlo (2019) p. 20. Firstly, the population vector consisting of real values of the variable of interest for sampled elements and (possibly back-transformed) fitted values of the variable of interest based on the estimated model. Secondly, the value/s of *thetaFun* based on the population vector built in the first step is/are computed. Predicted values for unsampled population elements in subsets for which random effects are not observed in the sample are computed based only on fixed effects.

Value

The function returns a list with the following objects:

`thetaP`	the value/s of the predictor (more than one value is computed if in *thetaFun* more than one population characteristic is defined).
`fixed.part`	the fixed part of the formula of model.
`random.part`	the random part of the formula of model.
`thetaFun`	the function of the population values of the variable of interest (on the original scale) which defines at least one population or subpopulation characteristic to be predicted.
`backTrans`	back-transformation function of the variable of interest (e.g. if YS is log-tranformed, then backTrans <- function(x) exp(x)).
`YP`	predicted values of the variable of interest for unsampled elements (without back-tranformation).
`YbackTrans`	population vector of the values of the variable of interest on the orignal scale for sampled elements and back-transformed predicted values of the variable of interest for unsampled elements.
`YPbackTrans`	back-transformed predicted values of the variable of interest for unsampled elements.
`beta`	the estimated vector of fixed effects.

`Xbeta`	the product of two matrices: the population model matrix of auxiliary variables X and the estimated vector of fixed effects.
`sigma2R`	the estimated variance parameter of the distribution of random components.
`R`	the estimated covariance matrix of random components for sampled elements.
`G`	the estimated covariance matrix of random effects.
`model`	the formula of the model (as in *lmer* object).
`mEst`	*lmer* object with the estimated model.
`YS`	values of the variable of interest (already transformed if necessary) observed in the sample and used in the model as the dependent variable.
`reg`	the population matrix of auxiliary variables named in *fixed.part* and *random.part*.
`con`	the population 0-1 vector with 1s for elements in the sample and 0s for elements which are not in the sample.
`regS`	the sample matrix of auxiliary variables named in *fixed.part* and *random.part*.
`regR`	the matrix of auxiliary variables named in *fixed.part* and *random.part* for unsampled population elements.
`weights`	the population vector of weigts, defined as in *lmer* object, allowing to include the heteroscedasticity of random components in the mixed linear model.
`Z`	the population model matrix of auxiliary variables associated with random effects.
`X`	the population model matrix of auxiliary variables associated with fixed effects.
`ZS`	the submatrix of Z matrix where the number of rows equals the number of sampled elements and the number of columns equals the number of estimated random effects.
`XR`	the submatrix of X matrix (with the same number of columns) for unsampled population elements.
`ZR`	the submatrix of Z matrix where the number of rows equals the number of unsampled population elements and the number of columns equals the number of estimated random effects.
`eS`	the sample vector of estimated random components.
`vS`	the estimated vector of random effects.

Author(s)

Alicja Wolny-Dominiak, Tomasz Zadlo

References

Chwila, A., Zadlo, T. (2019) On properties of empirical best predictors. Communications in Statistics - Simulation and Computation, available online: https://doi.org/10.1080/03610918.2019.1649422

Examples

```
library(lme4)
library(Matrix)


### Prediction of the subpopulation median and subpopulation standard deviation
### based on the cross-sectional data

data(invData)
#data from one period are considered:
invData2018<-invData[invData$year == 2018,]
attach(invData2018)

N=nrow(invData2018) #population size
n=100 # sample size

set.seed(123456)
sampled_elements=sample(N,n)
con=rep(0,N)
con[sampled_elements]=1 #elements in the sample
YS=log(investments[sampled_elements]) # log-transformed values
backTrans <- function(x) exp(x) # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
random.part <- '(log(newly_registered)  | NUTS2)'
reg=invData2018[, - which(names(invData2018) == 'investments')]
weights=rep(1,N) #homoscedastic random components

# Characteristics to be predicted - median and standard deviation
# in subpopulation of interest: NUTS4type==2
thetaFun <- function(x) {c(median(x[NUTS4type==2]),sd(x[NUTS4type==2]))}

# Predicted values of the median and the standard deviation
# in the following subpopulation: NUTS4type==2

plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)$thetaP

#All results
plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)

detach(invData2018)

##########################################################

### Prediction of the subpopulation quartiles based on longitudinal data

data(invData)
attach(invData)

N=nrow(invData[(year==2013),]) #population size in the first period
n=38 # sample size in the first period
#subpopulation and time period of interest: NUTS2=='02' & year==2018
```

```
#Ndt=sum(NUTS2=='02' & year==2018) #subpopulation size in the period of interest

set.seed(123456)
sampled_elements_in_2013=sample(N,n)
con2013=rep(0,N)
con2013[sampled_elements_in_2013]=1 #elements in the sample in 2013

#balanced panel sample - the same elements in all 6 periods:
con=rep(con2013,6)

YS=log(investments[con==1]) #log-transformed values
backTrans <- function(x) exp(x) # back-transformation of the variable of interest
fixed.part <- 'log(newly_registered)'
random.part <- '(log(newly_registered)  | NUTS4)'
reg=invData[, - which(names(invData) == 'investments')]
weights=rep(1,nrow(invData)) #homoscedastic random components

thetaFun <- function(x) {quantile(x[NUTS2=='02' & year==2018],probs =c(0.25,0.5,0.75))}

# Predicted values of quartiles
# in the following subpopulation: NUTS4type==2
# in the following time period: year==2018

plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)$thetaP

#All results
plugInLMM(YS, fixed.part, random.part, reg, con, weights, backTrans, thetaFun)

detach(invData)
```

Zfun *Matrix Z creator*

Description

The function creates the Z matrix of auxiliary variables asscociatied with random effects

Usage

```
Zfun(model, data)
```

Arguments

`model`	formula of model (use formula() function)
`data`	data

Value

`Z`	Z matrix
`vNames`	labels of random effects

Author(s)

Alicja Wolny-Dominiak

Examples

```
data(invData)
modelFormula <- formula(investments~newly_registered + (newly_registered | NUTS2))
reg <- invData

Zfun(modelFormula, reg)
```

Index

bootPar, 2
bootRes, 4

correction, 6
corrRancomp, 7
corrRanef, 8

doubleBoot, 9

EBLUP, 12
ebpLMMne, 16
EmpCM, 20
EstCM, 21

invData, 22

lwzl, 22

plugInLMM, 23

Zfun, 27