Use case for bootstrap methods in saemix

Emmanuelle Comets

01/02/2024

Contents

The present document is an R markdown notebook designed to showcase the different bootstrap methods in the paper **Conditional non-parametric bootstrap for non-linear mixed effect models** by S. Kaisaridi et al (submitted to **Journal of Computational and Graphical Statistics** in February 2024). This notebook uses the saemix available on CRAN (version 3.2 when the present document was created).

In addition, several libraries need to be installed to run the present code:

- ggplot2 and its dependencies
- MASS

Loading functions

This chunk loads the necessary libraries.

```
# Libraries
library(ggplot2)
library(MASS)
library(saemix)
```

```
## Le chargement a nécessité le package : npde
```

```
## Package saemix, version 3.3, November 2023
##   please direct bugs, questions and feedback to emmanuelle.comets@inserm.fr
```

```
##
## Attachement du package : 'saemix'
```

```
## Les objets suivants sont masqués depuis 'package:npde':
##
##     kurtosis, skewness
```

Running bootstrap methods on the PD Emax dataset

Fitting the data with saemix

```
data(PD1.saemix)

saemix.data<-saemixData(name.data=PD1.saemix,header=TRUE,name.group=c("subject"),
      name.predictors=c("dose"),name.response=c("response"),
      name.covariates=c("gender"), units=list(x="mg",y="-",covariates=c("-")))
```

```
##
##
## The following SaemixData object was successfully created:
```

```
##
## Object of class SaemixData
##     longitudinal data for use with the SAEM algorithm
## Dataset PD1.saemix
##     Structured data: response ~ dose | subject
##     Predictor: dose (mg)
##     covariates: gender (-)
##       reference class for covariate gender :  0
```

```r
modelemax<-function(psi,id,xidep) {
# input:
#   psi : matrix of parameters (3 columns, E0, Emax, EC50)
#   id : vector of indices
#   xidep : dependent variables (same nb of rows as length of id)
# returns:
#   a vector of predictions of length equal to length of id
  dose<-xidep[,1]
  e0<-psi[id,1]
  emax<-psi[id,2]
  e50<-psi[id,3]
  f<-e0+emax*dose/(e50+dose)
  return(f)
}

pdmodel<-saemixModel(model=modelemax,description="Emax growth model",
       psi0=matrix(c(20,300,20,0,0,0),ncol=3,byrow=TRUE,dimnames=list(NULL,
       c("E0","Emax","EC50"))), transform.par=c(1,1,1),
       covariate.model=matrix(c(0,0,1), ncol=3,byrow=TRUE),fixed.estim=c(1,1,1))
```

```
##
##
## The following SaemixModel object was successfully created:
##
## Nonlinear mixed-effects model
##   Model function:  Emax growth model
##   Model type:  structural
## function(psi,id,xidep) {
## # input:
## #   psi : matrix of parameters (3 columns, E0, Emax, EC50)
## #   id : vector of indices
## #   xidep : dependent variables (same nb of rows as length of id)
## # returns:
## #   a vector of predictions of length equal to length of id
##   dose<-xidep[,1]
##   e0<-psi[id,1]
##   emax<-psi[id,2]
##   e50<-psi[id,3]
##   f<-e0+emax*dose/(e50+dose)
##   return(f)
## }
##   Nb of parameters: 3
##       parameter names:  E0 Emax EC50
##       distribution:
##      Parameter Distribution Estimated
## [1,] E0        log-normal   Estimated
## [2,] Emax      log-normal   Estimated
## [3,] EC50      log-normal   Estimated
##   Variance-covariance matrix:
##      E0 Emax EC50
## E0    1    0    0
## Emax  0    1    0
## EC50  0    0    1
##   Error model: constant , initial values: a.1=1
##   Covariate model:
##      E0 Emax EC50
## [1,]  0    0    1
##     Initial values
##              E0 Emax EC50
## Pop.CondInit 20  300   20
## Cov.CondInit  0    0    0
```

```
# SE not computed as not needed for the test
saemix.options<-list(algorithms=c(0,1,1),nb.chains=3,seed=765754,
       nbiter.saemix=c(500,300),save=FALSE,save.graphs=FALSE, displayProgress=FALSE)

fit.emax<-saemix(pdmodel,saemix.data,saemix.options)
```

```
## Nonlinear mixed-effects model fit by the SAEM algorithm
## -----------------------------------
## ----          Data             ----
## -----------------------------------
## Object of class SaemixData
##     longitudinal data for use with the SAEM algorithm
## Dataset PD1.saemix
##     Structured data: response ~ dose | subject
##     Predictor: dose (mg)
##     covariates: gender (-)
##       reference class for covariate gender :  0
## Dataset characteristics:
##     number of subjects:     100
##     number of observations: 300
##     average/min/max nb obs: 3.00  /  3  /  3
## First 10 lines of data:
##    subject dose response gender mdv cens occ ytype
## 1        1    0  11.2870      1   0    0   1     1
## 2        1   10  63.6114      1   0    0   1     1
## 3        1   90 122.9170      1   0    0   1     1
## 4        2    0  15.0514      1   0    0   1     1
## 5        2   10  39.5296      1   0    0   1     1
## 6        2   90  60.8522      1   0    0   1     1
## 7        3    0  25.5390      1   0    0   1     1
## 8        3   10  58.0035      1   0    0   1     1
## 9        3   90  81.1173      1   0    0   1     1
## 10       4    0  22.1446      1   0    0   1     1
## -----------------------------------
## ----          Model            ----
## -----------------------------------
## Nonlinear mixed-effects model
##   Model function:  Emax growth model
##   Model type:  structural
## function(psi,id,xidep) {
## # input:
## #   psi : matrix of parameters (3 columns, E0, Emax, EC50)
## #   id : vector of indices
## #   xidep : dependent variables (same nb of rows as length of id)
## # returns:
## #   a vector of predictions of length equal to length of id
##   dose<-xidep[,1]
##   e0<-psi[id,1]
##   emax<-psi[id,2]
##   e50<-psi[id,3]
##   f<-e0+emax*dose/(e50+dose)
##   return(f)
## }
## <bytecode: 0x5616e75d4928>
##   Nb of parameters: 3
##       parameter names:  E0 Emax EC50
##       distribution:
##      Parameter Distribution Estimated
## [1,] E0        log-normal   Estimated
## [2,] Emax      log-normal   Estimated
## [3,] EC50      log-normal   Estimated
##   Variance-covariance matrix:
##      E0 Emax EC50
## E0    1    0    0
## Emax  0    1    0
## EC50  0    0    1
##   Error model: constant , initial values: a.1=1
##   Covariate model:
##        [,1] [,2] [,3]
## gender    0    0    1
##     Initial values
##              E0 Emax EC50
## Pop.CondInit 20  300   20
## Cov.CondInit  0    0    0
## -----------------------------------
## ----    Key algorithm options  ----
## -----------------------------------
##     Estimation of individual parameters (MAP)
##     Estimation of standard errors and linearised log-likelihood
##     Estimation of log-likelihood by importance sampling
##     Number of iterations:  K1=500, K2=300
##     Number of chains:  3
##     Seed:  765754
##     Number of MCMC iterations for IS:  5000
##     Simulations:
##         nb of simulated datasets used for npde:  1000
##         nb of simulated datasets used for VPC:  100
##     Input/output
##         save the results to a file:  FALSE
##         save the graphs to files:  FALSE
## ----------------------------------------------------
## ----                  Results                   ----
## ----------------------------------------------------
## -----------------  Fixed effects  ------------------
## ----------------------------------------------------
##      Parameter         Estimate SE    CV(%) p-value
## [1,] E0                 23.24   1.072  4.6  -
## [2,] Emax              107.20   6.120  5.7  -
## [3,] EC50               11.45   0.980  8.6  -
## [4,] beta_gender(EC50)   0.39   0.099 25.6  9.3e-05
## [5,] a.1                 4.72   0.407  8.6  -
## ----------------------------------------------------
## -----------  Variance of random effects  -----------
## ----------------------------------------------------
##      Parameter   Estimate SE    CV(%)
## E0   omega2.E0   0.129    0.028 22
## Emax omega2.Emax 0.307    0.045 15
## EC50 omega2.EC50 0.052    0.022 43
## ----------------------------------------------------
## ------  Correlation matrix of random effects  ------
## ----------------------------------------------------
##             omega2.E0 omega2.Emax omega2.EC50
## omega2.E0   1         0           0
## omega2.Emax 0         1           0
## omega2.EC50 0         0           1
## ----------------------------------------------------
## ---------------  Statistical criteria  -------------
## ----------------------------------------------------
## Likelihood computed by linearisation
##       -2LL= 2448.635
##       AIC = 2464.635
##       BIC = 2485.477
##
## Likelihood computed by importance sampling
##       -2LL= 2452.279
##       AIC = 2468.279
##       BIC = 2489.121
## ----------------------------------------------------
```

```
om.estim<-c(diag(fit.emax@results@omega),fit.emax@results@respar[1])
par.estim<-c(fit.emax@results@fixed.effects,om.estim)
sd.estim <- c(fit.emax@results@se.fixed, fit.emax@results@se.omega, fit.emax@results@se.respar[1])
```

Computing bootstrap distributions

In the following we estimate bootstrap distributions with 100 samples using the different bootstrap methods through the *saemix.bootstrap()* function.

Warning: this code may take some time to execute. My Rstudio tends to run out of memory and crash before rendering, so I generated the pdf using the command *rmarkdown::render(“./comets_condBoostrapSaemix2024.Rmd”)* in a plain R window.

```
saemix.bootOpt<-list(fix.seed=F,directory="current",displayProgress=F, save.graphs=F, map=F, ll.is=F, p
nboot<-200

start_time <- Sys.time()
boot.case<-saemix.bootstrap(fit.emax, nboot=nboot, method="case")
boot.cNP<-saemix.bootstrap(fit.emax, nboot=nboot, method="conditional")
```

```
## Simulating data using nsim = 1000 simulated datasets
## Computing WRES and npde .........
```

```
## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN

## Warning in sqrt(varik): Production de NaN
```

```r
boot.NP<-saemix.bootstrap(fit.emax, nboot=nboot, method="residual")
```

```
## Simulating data using nsim = 1000 simulated datasets
## Computing WRES and npde .........
```

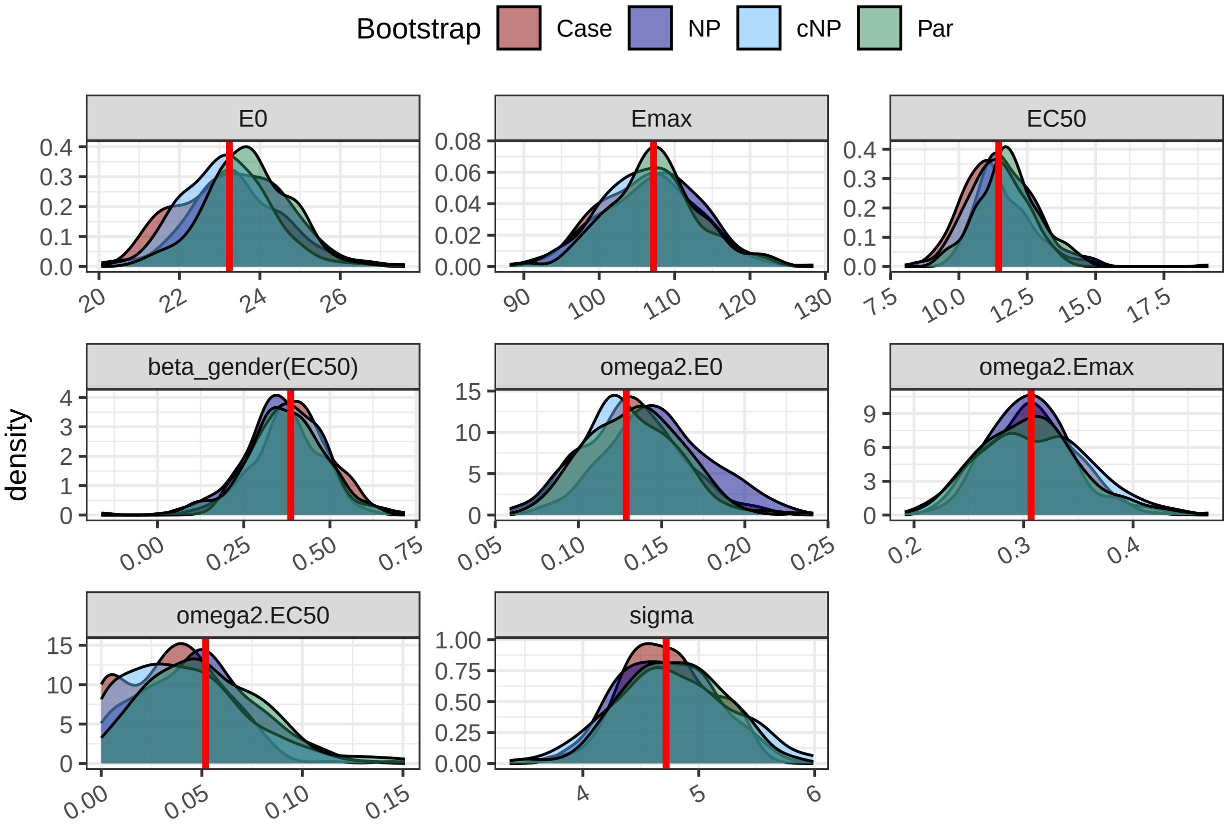
```r
boot.Par<-saemix.bootstrap(fit.emax, nboot=nboot, method="parametric")
end_time <- Sys.time()

write.table(boot.case,file.path(resDir,"condBootPaper_case.res"), quote=F, col.names = T)
write.table(boot.cNP,file.path(resDir,"condBootPaper_cond.res"), quote=F, col.names = T)
write.table(boot.NP,file.path(resDir,"condBootPaper_np.res"), quote=F, col.names = T)
write.table(boot.Par,file.path(resDir,"condBootPaper_par.res"), quote=F, col.names = T)
```

The following table compares bootstrap and asymptotic estimates for this dataset. The graph shows the distributions obtained for the different bootstraps, overlaying the estimate from saemix for this dataset.

```
## Bootstrap estimates (SD)
```

```
##                          saemix          Case            NP           cNP
## E0                 23.24 (1.07)  23.17 (1.24)  23.59 (1.21)  23.14 (1.06)
## Emax              107.20 (6.12) 106.52 (6.49) 107.16 (6.46) 106.47 (6.17)
## EC50               11.45 (0.98)  11.30 (1.31)  11.86 (1.06)  11.38 (1.04)
## beta_gender(EC50)   0.39 (0.10)   0.39 (0.11)   0.35 (0.10)   0.38 (0.10)
## omega2.E0           0.13 (0.03)   0.13 (0.03)   0.15 (0.03)   0.13 (0.03)
## omega2.Emax         0.31 (0.05)   0.31 (0.04)   0.31 (0.04)   0.31 (0.05)
## omega2.EC50         0.05 (0.02)   0.04 (0.03)   0.05 (0.03)   0.04 (0.03)
## a.1                 4.72 (0.41)   4.73 (0.38)   4.73 (0.41)   4.77 (0.51)
##                             Par
## E0                 23.71 (1.04)
## Emax              107.38 (5.99)
## EC50               11.81 (1.10)
## beta_gender(EC50)   0.37 (0.11)
## omega2.E0           0.13 (0.03)
## omega2.Emax         0.31 (0.05)
## omega2.EC50         0.05 (0.03)
## a.1                 4.78 (0.43)
```

```
## Run time: 1.32199
```


Bootstrap
Case
NP
cNP
Par
density
E0
Emax
EC50
beta_gender(EC50)
omega2.E0
omega2.Emax
omega2.EC50
sigma