

Use case for bootstrap methods in saemix

Emmanuelle Comets

17/09/2020

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** submitted by E. Comets et al (currently submitted to **Biostatistics**, September 2020). This notebook uses code from the development version of the *saemix* package, available on github: <https://github.com/saemixdevelopment/saemixextension>. An easy way to ensure all the functions are available is to clone the project using Rstudio, and follow updates by pull requests. Folders are relative to the installation folder, assuming the same architecture as on the github repository.

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

- ggplot2 and its dependencies
- MASS

The full set of simulated data and bootstrap distributions used for the paper have been uploaded to the Zenodo archive system <https://doi.org/10.5281/zenodo.4059718>

Loading functions

As we use the development version of *saemix*, this chunk loads the functions from the package, as well as necessary libraries.

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

# Sourcing saemix functions
{
  source(file.path(progDir, "aaa_generics.R"))
  #source(file.path(progDir, "global.R"))
  source(file.path(progDir, "SaemixData.R"))
  source(file.path(progDir, "SaemixRes.R"))
  source(file.path(progDir, "SaemixModel.R"))
  source(file.path(progDir, "SaemixObject.R"))
  source(file.path(progDir, "main.R"))
  source(file.path(progDir, "func_aux.R"))
  source(file.path(progDir, "main_initialiseMainAlgo.R"))
  source(file.path(progDir, "main_estep.R"))
  source(file.path(progDir, "main_mstep.R"))
  source(file.path(progDir, "func_FIM.R"))
  source(file.path(progDir, "func_plots.R"))
  source(file.path(progDir, "func_distcond.R"))
}
```

```

source(file.path(progDir,"func_simulations.R"))
source(file.path(progDir,"compute_LL.R"))
source(file.path(progDir,"func_estimParam.R"))
}

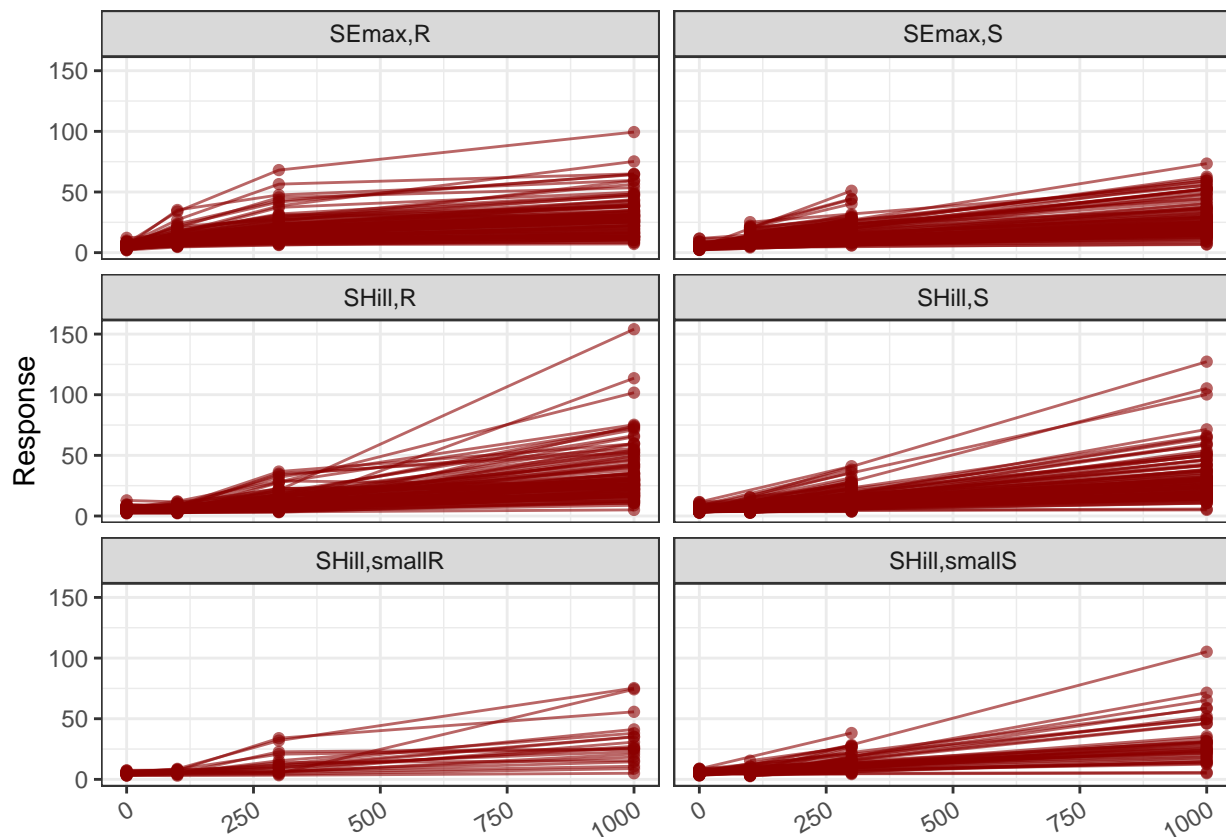
# Bootstrap functions
source(file.path(workDir,"bootstrap","saemix_bootstrap.R"))

```

Bootstrap distributions for one simulated dataset

Plot of simulated data for one simulation in each scenario

Below, we load the simulated data for the 6 scenarios presented in the paper (we randomly sampled one of the 200 simulations used to evaluate the performances of the different bootstrap approaches, simulation 87) and plot it using the *ggplot2* package (Supplemental material Figure 12).



Fitting the simulated data in replication 87, scenario SHill,R, with saemix

```

## Reading data from file /home/eco/work/saemix/saemixextension/bootstrap/data/data_pdhillhigh.rich_sim
## These are the first lines of the dataset as read into R. Please check the format of the data is appropriate
##   id dose      y
## 1  1   0  4.361467
## 2  1  100  4.504739
## 3  1  300  5.761015
## 4  1 1000 20.906273

```

```

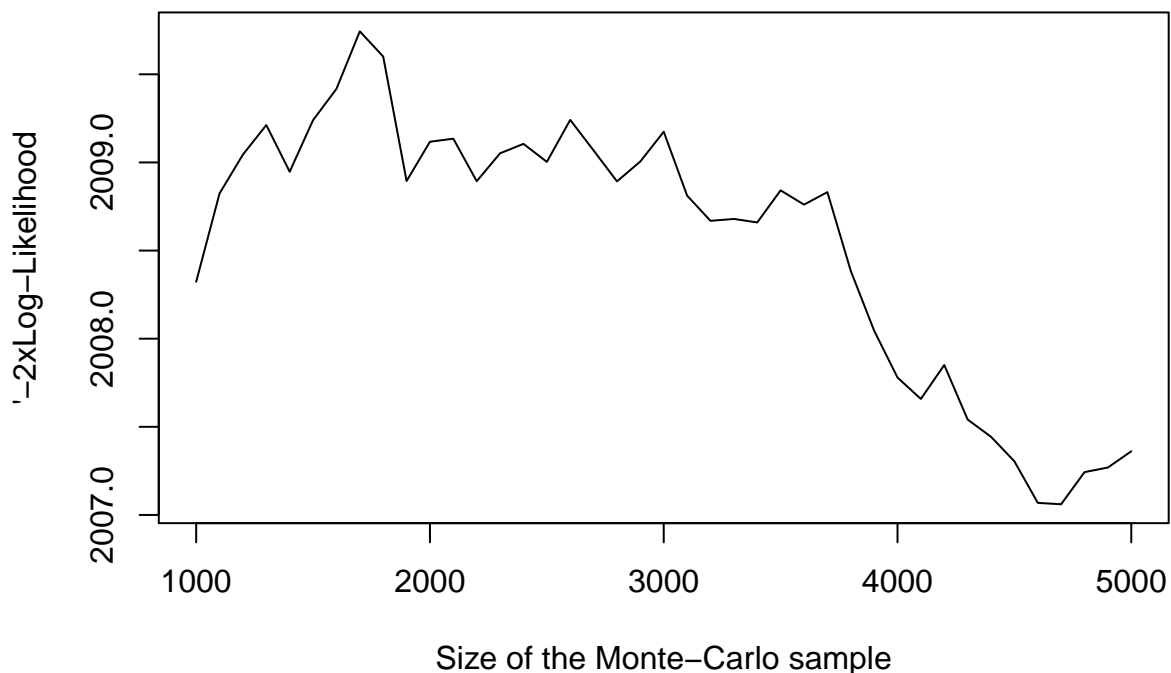
## 5 2 0 5.520022
## 6 2 100 5.802740
## Column name(s) time amt do(es) not exist in the dataset, please check
## Remove columns 1 2 ( time amt )
## No valid name given, attempting automatic recognition
## Automatic recognition of columns dose successful
## Column name(s) conc do(es) not exist in the dataset, please check
## Remove columns 1 ( conc )
## No valid name given, attempting automatic recognition
## Automatic recognition of columns y successful
##
##
## The following SaemixData object was successfully created:
##
## Object of class SaemixData
## longitudinal data for use with the SAEM algorithm
## Dataset /home/eco/work/saemix/saemixextension/bootstrap/data/data_pdhillhigh.rich_sim87.tab
## Structured data: y ~ dose | id
## Predictor: dose (hr)
##
##
## The following SaemixModel object was successfully created:
##
## Nonlinear mixed-effects model
## Model function: Hill model Model type: structural
## function(psi,id,xidep) {
## # input:
## # psi : matrix of parameters (4 columns, E0, Emax, E50, gamma)
## # 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]
## gamma<-psi[id,4]
## f<-e0+emax*dose**gamma/(e50**gamma+dose**gamma)
## return(f)
## }
## Nb of parameters: 4
## parameter names: E0 Emax ED50 Gamma
## distribution:
## Parameter Distribution Estimated
## [1,] E0 log-normal Estimated
## [2,] Emax log-normal Estimated
## [3,] ED50 log-normal Estimated
## [4,] Gamma log-normal Estimated
## Variance-covariance matrix:
## E0 Emax ED50 Gamma
## E0 1 0 0 0
## Emax 0 1 1 0
## ED50 0 1 1 0

```

```
## Gamma 0 0 0 0
## Error model: proportional , initial values: b.1=0.1
## No covariate in the model.
## Initial values
##      E0 Emax ED50 Gamma
## Pop.CondInit 5 30 500 3

## Problem inverting covariance matrix, setting initial Omega to diagonal.
## Running main SAEM algorithm
## [1] "Thu Oct 1 11:32:45 2020"
## ....
## Minimisation finished
## [1] "Thu Oct 1 11:32:52 2020"
```

Estimation of the log-likelihood



```
## Overwriting files in directory current
## -----
## ----- Fixed effects -----
## -----
##      Parameter Estimate SE      CV(%)
## [1,] E0          5.021  0.1641 3.3
## [2,] Emax        32.860  2.6051 7.9
## [3,] ED50        549.137 37.8725 6.9
## [4,] Gamma        3.108  0.1029 3.3
## [5,] b.1          0.088  0.0059 6.7
## -----
## ----- Variance of random effects -----
## -----
##      Parameter      Estimate SE      CV(%)
## E0  omega2.E0      0.10     0.015 15
## Emax omega2.Emax    0.43     0.071 16
```

```

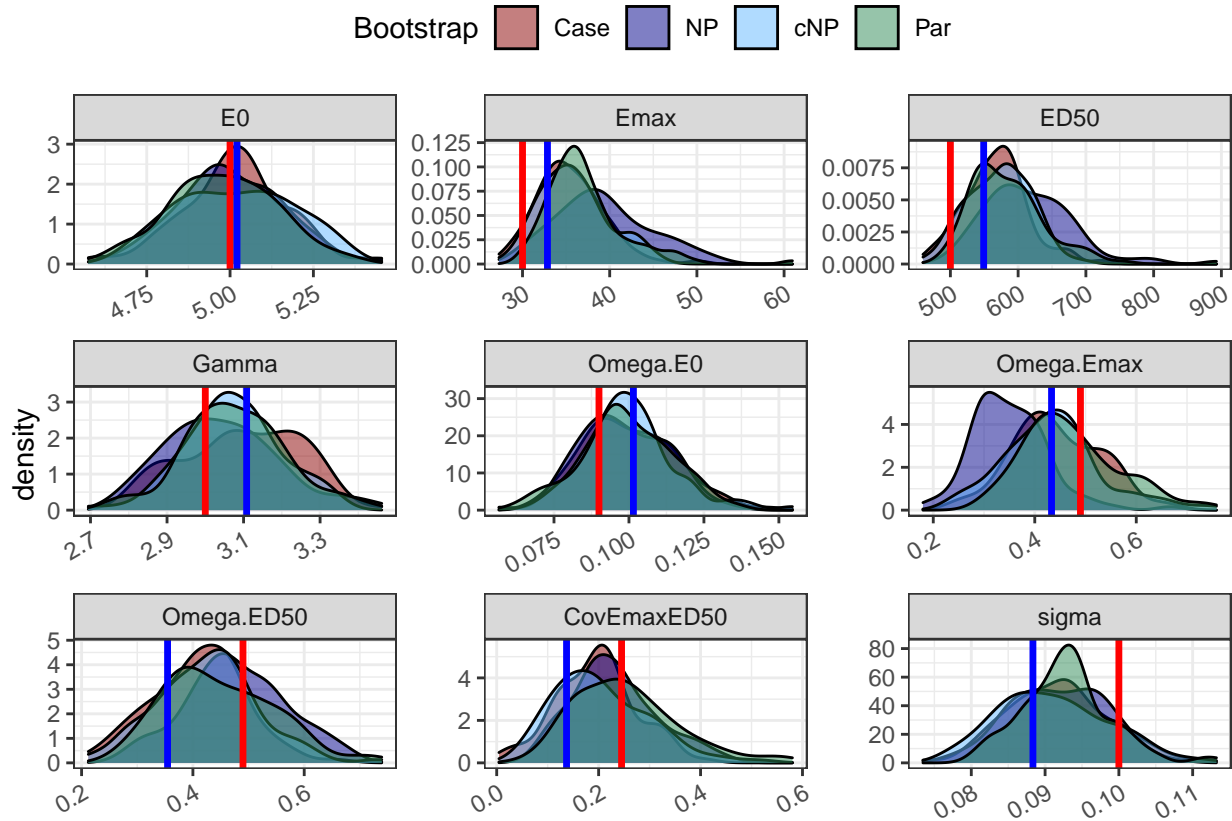
## ED50  omega2.ED50  0.35      0.055 16
## covar cov.Emax.ED50 0.14      0.048 35
## -----
## -----  Correlation matrix of random effects  -----
## -----
##              omega2.E0  omega2.Emax  omega2.ED50
## omega2.E0    1          0.00        0.00
## omega2.Emax  0          1.00        0.35
## omega2.ED50  0          0.35        1.00
## -----
## -----  Statistical criteria  -----
## -----
## Likelihood computed by linearisation
##      -2LL= 2010.044
##      AIC = 2028.044
##      BIC = 2051.491
##
## Likelihood computed by importance sampling
##      -2LL= 2007.361
##      AIC = 2025.361
##      BIC = 2048.808
## -----

```

Bootstrap distributions for the simulated data in replication 87, scenario SHill,R

The following code loads the bootstrap distributions obtained by the four tested bootstraps for the simulated data in replication 87 (shown the first plot of the second line in the previous figure) from the result files, and plots these for each parameter. This corresponds to Figure 1 in the paper (estimated values may slightly vary due to the stochastic nature of the SAEM algorithm).

Note: 500 bootstrap samples were generated for each of the first 4 scenarios, but in the paper we harmonise across the different scenarios to 100 samples through the *nboot<-100* command in the code below.



```
## pdf
## 2
```

Running bootstrap methods on one dataset (replication 87, scenario SHill,R)

In the following we estimate bootstrap distributions with 100 samples using the different bootstrap methods through the *saemix.bootstrap()* function, and recreate the previous figure. Due to stochastic variability the bootstrap distributions are different from before but their shape and amplitude is similar.

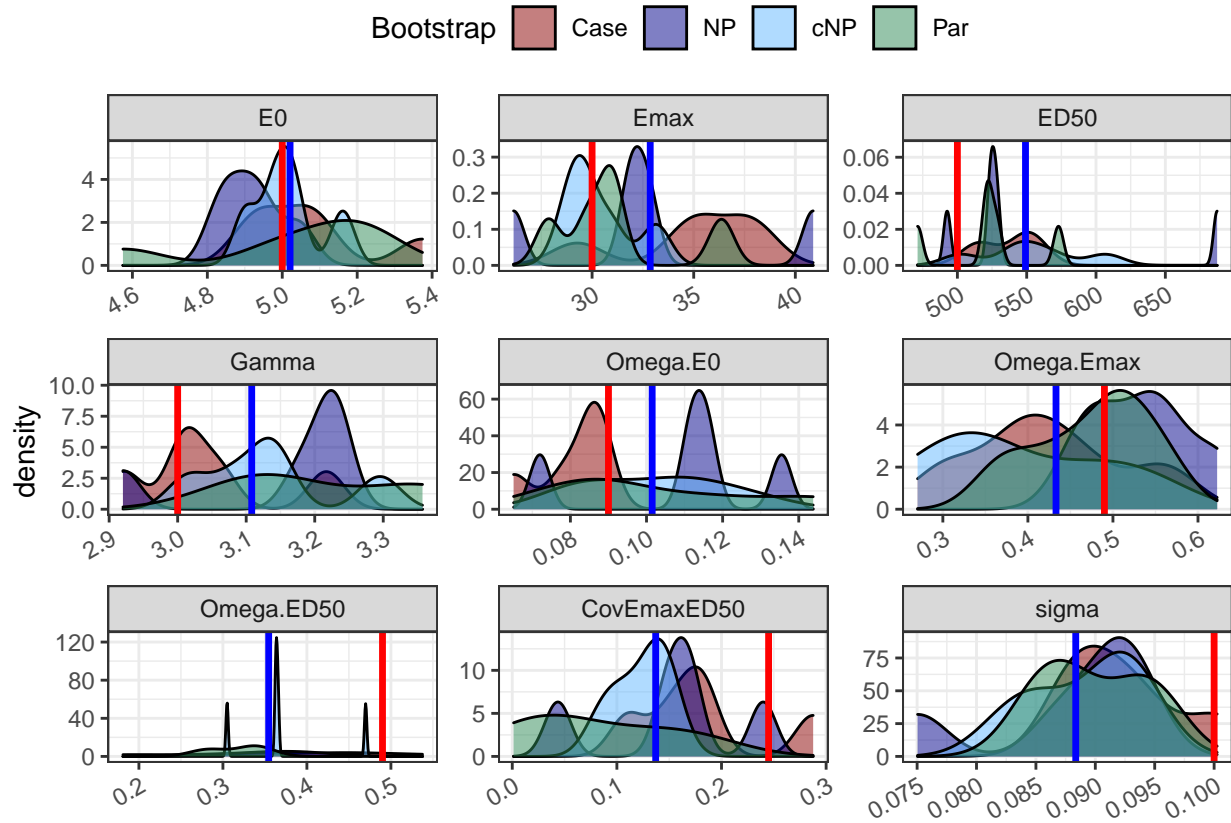
Warning: this code may take some time to execute. My Rstudio tends to run out of memory and crash before rendering, so I generated the pdf using the command `rmarkdown::render("/comets_condBootstrapSaemix.Rmd")` in a plain R window.

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

	True	Case	NP	cNP
## E0	5.000	5.08 (0.1788)	4.92 (0.0850)	5.01 (0.0963)
## Emax	30.000	34.97 (3.4781)	32.77 (5.2682)	30.38 (1.7414)
## ED50	500.000	536.91 (19.7862)	551.55 (77.3076)	553.38 (38.2279)
## Gamma	3.000	3.04 (0.1087)	3.16 (0.1337)	3.13 (0.1038)
## omega2.E0	0.090	0.08 (0.0095)	0.11 (0.0232)	0.10 (0.0180)
## omega2.Emax	0.490	0.42 (0.0951)	0.53 (0.0621)	0.39 (0.1050)
## omega2.ED50	0.490	0.42 (0.0843)	0.32 (0.0882)	0.37 (0.0597)
## cov.EmaxED50	0.245	0.18 (0.0659)	0.15 (0.0707)	0.12 (0.0262)
## b.1	0.100	0.09 (0.0050)	0.09 (0.0076)	0.09 (0.0046)
##		Par		
## E0	5.02 (0.2707)			
## Emax	31.18 (3.1687)			

```
## ED50          523.27 (36.0808)
## Gamma         3.22 ( 0.1265)
## omega2.E0     0.10 ( 0.0276)
## omega2.Emax   0.47 ( 0.0711)
## omega2.ED50   0.32 ( 0.0318)
## cov.EmaxED50  0.08 ( 0.0727)
## b.1           0.09 ( 0.0044)
```

```
## Run time: 4.778002
```



```
## pdf
## 2
```