Pharmaceutical Research manuscript No. (will be inserted by the editor)

Supplementary material for article: Conditional non-parametric bootstrap for non-linear mixed effect models

Emmanuelle Comets · Christelle Rodrigues · Vincent Jullien · Moreno Ursino

Received: date / Accepted: date

Appendix A - Bootstrap methods

Let B be the number of bootstrap samples to be drawn from the original dataset, a general bootstrap algorithm is:

- 1. Generate a bootstrap sample by resampling from the data and/or from estimated model
- 2. Compute the estimates of the parameters of the model for the bootstrap sample
- 3. Repeat steps 1-2 B times and compute the 95th bootstrap confidence interval from the bootstrap distribution as

$$\hat{\theta}^*_{(\alpha \cdot B)} \le \hat{\theta} \le \hat{\theta}^*_{((1-\alpha) \cdot B)} \tag{1}$$

where $\alpha = 0.025$ and $\hat{\theta}^*_{(\alpha \cdot B)}$ is the α -quantile of the bootstrap distribution. The mean and standard deviation of the distribution of the bootstrap estimates $\hat{\theta}^*_b$ (b = 1...B) can be viewed as the bootstrap estimates of the parameters and their SE, however the bootstrap distribution may be asymmetric contrary to the asymptotic estimate.

Case bootstrap (Case)

This method consists of resampling with replacement the entire subjects (ξ_i, \mathbf{Y}_i) where $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ from the original data before modelling. It is also called the *paired bootstrap*. It is the most obvious way to do bootstrapping and is mostly assumption-free.

Non-parametric residual bootstrap (NP)

This method resamples with replacement from the residuals obtained after model fitting for all subjects. In non-linear mixed effect models, two levels of variability

Emmanuelle Comets E-mail: emmanuelle.comets@inserm.fr

must be considered, within-subject and between different subjects. The bootstrap sample is obtained as follows:

- 1. fit the model to the data
- 2. estimate the individual parameters $\hat{\eta}_i$ (as the EBE), then center and normalise them as detailed below
- 3. draw a sample η_i^* with replacement for i = 1, ...N from the resulting set
- 4. calculate the residuals $\hat{\epsilon}_{ij} = y_{ij} f(x_{ij}, \hat{\theta}_i)$, then center and normalise them
- 5. draw a sample $\{\epsilon_i^*\} = \{\epsilon_{i^*j^*}\}$ with replacement globally from the resulting set
- 6. generate the bootstrap responses $y_{ij}^* = f(x_{ij}, \mu, \eta_i^*) + g(x_{ij}, \mu, \eta_i^*, \sigma) \epsilon_{ij}^*$

The normalisation in step 2 and 4 is necessary because, due to the small number of samples, these estimates generally suffer from regression to the mean and their variance may be considerably smaller than the true variability in the population [4]. To correct for this, [1] suggested centering the residuals, to ensure their distribution has mean 0, and inflating the variance by using the ratio between the estimated and empirical variance-covariance matrices. The transformation of random effects was carried out using the eigenvalue decomposition (EVD) proposed by Thai et al. [6], to limit numerical difficulties due to almost singular matrices. This procedure reads as follows:

- 1. center the raw estimated random effects: $\tilde{\eta}_i = \hat{\eta}_i \bar{\eta}_i$
- 2. obtain the EVD of the estimated variance-covariance matrix: $\hat{\Omega} = V_{\Omega} D_{\Omega} V_{\Omega}^{T}$ where D_{Ω} is the diagonal matrix containing the eigenvalues of $\hat{\Omega}$
- 3. obtain the EVD of S, the variance-covariance matrix of the centered random effects: $S = V_S D_S V_S^T$ where D_S is the diagonal matrix containing the eigenvalues of S
- 4. calculate the correction matrix A_{η} using these two decompositions as $A_{\eta} = V_S D_S^{-1/2} V_{\Omega} D_{\Omega}^{-1/2}$
- 5. transform the centered random effects using the ratio A_{η} : $\hat{\eta}'_{i} = \tilde{\eta}_{i} \times A_{\eta}$

Similarly, the residuals were transformed as follows:

- 1. center the raw estimated residuals: $\tilde{\epsilon}_{ij} = \hat{\epsilon}_{ij} \bar{\epsilon}_{ij}$
- 2. calculate the correction factor $A_{\sigma} = 1/\sigma_{emp}$ where σ_{emp} is the empirical standard deviation of the centered residuals
- 3. transform the centered residuals using the ratio A_{σ} : $\hat{\epsilon}'_{ij} = \tilde{\epsilon}_{ij} \times A_{\sigma}$

Parametric residual bootstrap (Par)

The parametric bootstrap requires the strongest assumptions as it depends both on the model and the distributions of parameters and errors. This method resamples the residuals by simulating from the estimated distributions obtained after model fitting. The bootstrap sample is obtained as follows:

- 1. Fit the model to the data
- 2. Draw a sample $\{\eta_i^*\}$ from a normal distribution with mean zero and covariance matrix Ω
- 3. Draw a sample $\{\epsilon_i^*\}$ from a normal distribution with mean zero and covariance matrix \mathcal{I}_{n_i} where \mathcal{I} denotes the identity matrix
- 4. Generate the bootstrap responses $y_{ij}^* = f(x_{ij}, \mu, \eta_i^*) + g(x_{ij}, \mu, \eta_i^*, \sigma) \epsilon_{ij}^*$

Conditional residual bootstrap (cNP)

Instead of using the empirical Bayes estimates and correcting the random effects for shrinkage, the conditional residual bootstrap uses samples from the conditional distribution. The algorithm then becomes:

- 1. fit the model to the data
- 2. obtain K samples for each vector η_i , η_{ik} (k = 1, ..., K) from the conditional distributions and center them
- 3. draw a sample η_i^* with replacement to obtain i = 1, ...N of vectors of individual parameters
- 4. calculate the residuals $\hat{\epsilon}_{ij} = y_{ij} f(x_{ij}, \hat{\theta}_i)$ and center them $\tilde{\epsilon}_{ij} = \hat{\epsilon}_{ij} \bar{\epsilon}_{ij}$
- 5. calculate the correction factor $A_{\sigma} = 1/\sigma_{emp}$ where σ_{emp} is the empirical standard deviation of the centered residuals and transform the centered residuals as $\hat{\epsilon}'_{ij} = \tilde{\epsilon}_{ij} \times A_{\sigma}$
- 6. draw a sample $\{\epsilon_{ij}^*\}$ with replacement globally from $\{\hat{\epsilon}_{ij}\}$ by assigning an equal probability $\frac{1}{\sum_{i=1}^{N} n_i}$ to each value
- 7. generate the bootstrap responses $y_{ij}^* = f(x_{ij}, \mu, \eta_i^*) + g(x_{ij}, \mu, \eta_i^*, \sigma) \epsilon_{ij}^*$

The conditional non-parametric bootstrap differs from the standard non-parametric bootstrap in the sampling and normalisation step for the random effects. In this work we implement step 3 by sampling from all conditional samples over i = 1...N, k = 1...K but we could also resample for each individual within his or her conditional samples.

Bootstrap evaluation

We computed the statistics of interest obtained from B bootstrap samples drawn for each simulated dataset k, including the bootstrap mean $\hat{\theta}_{B,k}$, the bootstrap standard error (SE) and the bootstrap confidence interval (CI) for each estimated parameter. In bootstrap theory, the bootstrap estimate of bias is usually calculated by the difference between the mean of bootstrap estimates and the estimate of parameter in the original dataset, to separate the bias of estimation procedure with the bias of bootstrap [3]:

$$\text{RBbias}(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^{K} \frac{\hat{\theta}_{B,k} - \hat{\theta}_k}{\hat{\theta}_k} \times 100$$
(2)

The bootstrap standard error for the k^{th} simulated dataset was obtained as the standard deviation of the bootstrap distribution $(SD(\hat{\theta}_{B,k}))$. We compared the bootstrap SE to the empirical SE obtained from K simulated datasets giving the "true" value observed across the simulations.

$$SE_{empirical}(\hat{\theta}) = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} (\hat{\theta}_k - \theta_0)^2}$$
(3)

We also compared this empirical SE with the asymptotic SE, given by the software as the inverse of the Fisher information matrix.

The relative bias on SE by each bootstrap method is then computed as the percentage difference between the average bootstrap SE and the empirical SE:

$$\operatorname{RBias}(SE(\hat{\theta})) = \frac{\frac{1}{K} \sum_{k=1}^{K} \widehat{SE}_{B,k} - \operatorname{SE}_{\text{empirical}}(\hat{\theta})}{\operatorname{SE}_{empirical}(\hat{\theta})} \times 100$$
(4)

The coverage rate of the 95% bootstrap CI of parameter is the probability that bootstrap CI contains the true value of the parameter.

The bootstrap approaches were compared in terms of the average RBias on the bootstrap parameter estimates, the RBias on SE, and the coverage rate of the 95% confidence interval (CI) of all parameter estimates from one million bootstrap samples.

Implementation

Simulations were performed in R [5]. The parameters of the models were estimated using the saemix library [2] in the development version available on github (https://github.com/saemixdevelopment/saemixextension). The default options were used to run the SAEM algorithm; the initial parameters were set to the true value in all runs, and simulated annealing was used with the default number of iterations (half of the iterations in the first phase, eg 150 iterations out of the first 300). The results were analysed in R. The runs were set up using a laptop and the simulation studies were run on the CATIBioMed cluster maintained by IFR02 (Bichat hospital, Paris).

Code and simulated data used in this study

The saemix code including the bootstrap methods can be found on the github repository https://github.com/saemixdevelopment/saemixextension. An example of code used to simulate data, estimate parameters and run the different bootstrap is available in the folder bootstrap as an Rstudio notebook https://github.com/saemixdevelopment/saemixextension/blob/master/bootstrap/comets_condBoostrapSaemix. Rmd along with an example of simulated data (instructions to download the necessary files and run the code are provided in the notebook).

The full data simulated in the 6 scenarios has been uploaded to the Zenodo repository, as well as the saemix estimates and the bootstrap distributions for each dataset in the 6 scenarios (https://doi.org/10.5281/zenodo.4059718).

Appendix B - Simulated data

Figure S1 shows the data simulated in one randomly selected simulation for each scenario (simulation 87). The scenarios in the left column involved a rich design with 4 doses given to each patient, while the scenarios in the right column were simulated using a sparse design with 2 doses per patient, and the patients were split in 4 groups to ensure enough information to estimate the parameters.



Fig. S1 Data simulated in one randomly selected dataset for each scenario.

Appendix C - Relative bias on parameters and their estimation errors

Table S1 shows the relative bias for the asymptotic method with respect to the true parameters, computed as:

$$\text{RBbias}_{asym}(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^{K} \frac{\hat{\theta}_k - \hat{\theta}_0}{\hat{\theta}_0} \times 100$$
(5)

 $\label{eq:saemix} \textbf{Table S1} \ \ \ \textbf{Relative estimation errors for the parameters estimated by \textbf{saemix for the 6 scenarios.} }$

Scenario	E_0	E_{max}	ED_{50}	γ	ω_{E_0}	$\omega_{E_{\max}}$	$\omega_{ED_{50}}$	cov	σ
$S_{Emax,R}$	0.62	0.20	-0.09		-1.81	-1.06	-1.63	-0.78	0.32
$S_{Emax,S}$	0.55	5.43	8.12		-3.03	2.82	34.73	24.42	-3.54
$S_{Hill,R}$	-0.32	0.52	1.04	0.68	-1.05	1.04	-2.31	-0.51	-1.02
$S_{Hill,S}$	0.57	-0.28	1.46	2.07	-3.16	-0.31	-11.07	-15.67	9.14
$S_{Hill,smallR}$	0.21	2.03	1.88	2.50	-5.78	-0.11	-9.99	-8.85	-4.85
$S_{Hill,smallS}$	0.29	5.39	7.39	0.46	-8.97	1.09	-9.48	-13.70	29.09

Appendix D - Distributions of the SE in scenario ${\rm S}_{\rm Hill,R}$

Figure S2 shows the distribution of the SE obtained by the four bootstraps and by the asymptotic method in scenario $S_{Hill,R}$, over the 200 datasets simulated in this scenario, comparing them with the empirical SE for each parameter. This figure shows that in this scenario, most bootstrap distributions are very similar in shape and extent, while the asymptotic distribution deviates for several parameters. For σ , the case bootstrap is shifted to the left with the same mode as the asymptotic distribution, while the NP and NPc bootstrap distributions are superimposed.



Fig. S2 Distributions of the SE obtained by the 4 bootstrap methods and by the asymptotic approach in scenario $S_{\rm Hill,R}$. The red vertical line indicates the empirical SE

Appendix E - Distributions of the RSE in the 6 scenarios

Figures S3 and S4 show the relative standard errors obtained using the bootstrap and the asymptotic estimates, and compare them to the empirical SE, for fixed and random effects respectively.



Fig. S3 Distributions of the relative SE on the fixed effects and on σ , obtained by the 4 bootstrap methods and by the asymptotic approach in all six scenarios. The red dot denotes the empirical RSE for each parameter. To make the figure more easy to read, and because of the difference in information between the different scenarios, each plot has its own scale.



Fig. S4 Distributions of the relative SE on the random effects obtained by the 4 bootstrap methods and by the asymptotic approach in all six scenarios. The red dot denotes the empirical RSE for each parameter. To make the figure more easy to read, and because of the difference in information between the different scenarios, each plot has its own scale.

These plots give an indication of the informativeness of the design. In the rich scenarios $S_{Emax,R}$ and $S_{Hill,R}$, the mean RSE on the fixed effects and on σ is less than 10% while the mean RSE on the random effects is less than 40%. When we move to a sparse design, the mean RSE increase to respectively 20-30% for the fixed effects and 60% for the random effects. Finally, the designs with a smaller number of subjects had a mean RSE higher than 80% for the covariance

term. Designs with even smaller sample size would not allow the estimation of a covariance term in practice (data not shown).

Appendix F - Percentage of datasets outside the 95% CI

Figure S5 shows the percentage of simulated datasets lying outside the confidence intervals defined by the various methods, stratified by parameter and scenario.



Fig. S5 Percentage of simulated datasets where the true parameter is outside the confidence interval. "U" indicates the number of datasets where the true parameter is over the 97.5th percentile of the confidence interval, and "L" the percentage of datasets where the true parameter is under the 2.5th percentile.

Appendix G - Coverage rate for the 50% interval

Figure S6 shows the coverage rates for the 50% confidence intervals in the different scenarios, showing that decreasing design information leads to overwide confidence intervals both by the asymptotic method and by the different bootstraps.



Fig. S6 Coverage rates for the 50% confidence interval in the 6 scenarios simulated in this study. Dotted lines indicate a coverage of 45 and 55%.

Appendix H - Influence of the number of bootstraps

The results shown in the main manuscript have been obtained using B=100 bootstrap datasets for each simulation. We ran the first four scenarios with up to 500 bootstrap samples, and the results of the 95% coverage are shown in Figure S7 for scenario $S_{\rm Hill,R}$, with different number of bootstrap samples used to estimate the bootstrap distributions. Similar plots were obtained for $S_{\rm Hill,S}$ (not shown).



Fig. S7 Influence of the number of bootstraps on the coverage rate for scenario $S_{Hill,R}$.

It appears that 100 bootstrap samples are therefore sufficient to obtain adequate coverage, while increasing B over 200 does not seem to have a major effect. The figures in Appendix G show the influence of the number of bootstrap samples on the estimates of the parameters and of their SE for the same scenario, and indicate the estimates stabilise after B=100. We therefore present the results of the simulations with B=100 as a trade-off between computational time and stability, although in real-life where we perform bootstrap on a single dataset we may increase this number to increase confidence in extreme percentiles of the distribution. Figures S8 and S9 show the bootstrap distributions of the parameter estimates and their SE, depending on the number of bootstraps, in scenario $S_{Hill,R}$.



Fig. S8 Influence of the number of bootstraps on the parameter estimates. The red lines show the true value of the parameters.

Appendix I - Comparison of the coverage of $\ensuremath{\mathrm{cNP}}$ and $\ensuremath{\mathrm{NP}}$

Figure S10 shows the coverage rates for the different parameters across the 6 scenarios. The lightly shaded areas represent coverages of 90 to 100%.



Fig. S9 Influence of the number of bootstraps on the standard errors of estimation. The red lines show the empirical SE.



Fig. S10 Coverage rate of cNP versus coverage of NP, coloured by parameter, over the 6 scenarios.

References for supplementary material

- J. R. Carpenter, H. Goldstein, and J. Rasbash. A novel bootstrap procedure for assessing the relationship between class size and achievement. *Applied Statistics*, 52:431–43, 2003.
- E. Comets, A. Lavenu, and M. Lavielle. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software*, 80:1–41, 2017.
- 3. A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, 1997.
- M. Karlsson and R. Savic. Diagnosing model diagnostics. Clin Pharmacol Ther, 82:17–20, 2007.
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- H. Thai, F. Mentré, N. H. Holford, C. Veyrat-Follet, and E. Comets. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. *Journal of Pharmacokinetics and Pharmacodynamics*, 41:15–33, 2014.