A Non-parametric bootstrap for multilevel models

By
James Carpenter
London School of Hygiene and Tropical Medicine

Harvey Goldstein
and
Jon Rasbash
Institute of Education

1. Introduction

Bootstrapping is now a well established procedure for assessing the bias and standard error of parameters in statistical models (Davison and Hinckley, 1997). Given a fitted model and parameter estimates, the idea is to generate synthetic (termed bootstrap) data from the fitted model, and then refit the model to the synthetic data, thus obtaining a set of synthetic (termed bootstrap) parameter estimates. These synthetic parameter estimates stand in approximately the same relationship to the model parameter estimates as the model parameter estimates stand in relationship to the population parameters. Thus, we can estimate quantities of interest relating the population parameters and the estimated parameters (such as bias, confidence intervals) by looking at the relationship between the estimated parameters and the synthetic, or bootstrap, parameters.

Broadly speaking the synthetic data can be generated in one of two ways, termed the *parametric* and *non-parametric* bootstrap. The parametric bootstrap, already implemented in *MLwiN* , generates the bootstrap data from the full parametric model. For example, consider the 2-level model fitted to the tutorial data example in the *MLwiN* user's guide,

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij} \quad (i=1,2,\ldots,I_j;\ j=1,\ldots,J) \tag{1}$$

where the response is the normalised exam score, the explanatory variable is the standardised LRT score and there are I_j pupils within school j. Suppose we have fitted the model and obtained estimates of all the parameters. Then the parametric bootstrap simulates

1) $e_{ij}^* \sim \mathrm{N}(0, \hat{\sigma}_e^2)$, (i=1,2,…, I_j; j=1,…,J),

where $\hat{\sigma}_e^2$ is the estimate of $\hat{\sigma}_e^2 = \mathrm{Var}(e_{ij})$ obtained from the data

2) $\left[\begin{pmatrix} u^*_{0j} \\ u^*_{1j} \end{pmatrix}\right] \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_{u0}^2 & \\ \sigma_{u01} & \hat{\sigma}_{u1}^2 \end{pmatrix}\right]$ (j=1,…,J),

where the $\hat{\sigma}_u^2$'s are elements of the variance-covariance matrix of the u's estimated from the data.

The bootstrap data set is then (y_{ij}^*, x_{1ij}), (i=1,2,…, I_j; j=1,…,J), where

$$y_{ij}^{*} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + u^{*}_{0j} + u^{*}_{1j} x_{1ij} + e^{*}_{ij}$$

A large number, B, typically 1000, such bootstrap data sets are generated, and the model fitted to each one. We thus obtain B bootstrap estimates of each parameter in the model, which we can then use to estimate bias, standard error and confidence intervals, as described in the *MLwiN* user's guide. We can also obtain bootstrap estimates of other quantities, such as the level 2 residuals.

Here we outline a non-parametric alternative to the parametric bootstrap, and show that it can yield a substantial reduction in the coverage error of parametric bootstrap confidence intervals when the data are not truly normally distributed.

2. A Non parametric bootstrap for multilevel models

Non-parametric bootstrapping can take two forms. In the first kind, case re-sampling, we build a bootstrap data set from the original data by sampling with replacement from the (y_{ij}, x_{1ij}) pairs that make up the data. However, in a multilevel context doing this crudely would break the structure of the dataset; if, as an alternative, we resample 'blocks' of data, it is not at all obvious which 'level' the blocks should correspond to. Furthermore, work in the standard regression context suggests that while this approach might be useful for deciding between models, it does not give accurate inference for parameters within such models, which is our principal goal.

We therefore propose to generalise the residual non-parametric bootstrap for regression models to the multilevel case. A crude residual bootstrap for model (1) would be the following:

1) Fit the model (1) to the data, and calculate the set of residuals $\{e_{ij}\}_{i=1,\dots,I_j;j=1\dots J}$ and $\{(u_{0j}, u_{1j})\}_{j=1,\dots,J}$
2) Sample with replacement from these two sets, obtaining two new sets $\{e_{ij}^{*}\}_{i=1,\dots I_j;j=1,\dots J}$ and $\{(u_{0j}^{*}, u_{1j}^{*})\}_{j=1,\dots,J}$
3) The bootstrap data set is then (y_{ij}^{*}, x_{1ij}), where

$$y_{ij}^{*} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + u^{*}_{0j} + u^{*}_{1j} x_{1ij} + e^{*}_{ij}$$

The drawback of this simple approach is that we will underestimate variances in particular because the crude residuals are 'shrunk' towards zero. We therefore need to 'reflate' the residuals before passing them back through the fitted model as in step (3) above. We now outline a procedure for doing this. For convenience we shall illustrate the procedure using the level 2 residuals, but analogous operations can be carried out at all levels. Rewrite model (1) as

$$\begin{aligned} y_{ij} &= (X\beta)_{ij} + (ZU)_j + e_{ij} \\ U^T &= \{U_0, U_1 \dots\} \end{aligned} \qquad (2)$$

Having fitted the model we calculate the residuals:

$$\hat{U} = \{\hat{u}_0, \hat{u}_1 \dots\}$$

Write the empirical covariance matrix of the estimated residuals at level 2 in model (2) as

$$S = \frac{\hat{U}^T \hat{U}}{J}$$

and the corresponding model estimated covariance matrix of the random coefficients at level 2 as *R*. The empirical covariance matrix is estimated using the number of level 2 units, *J,* as divisor rather than *J-1.* We assume that the estimated residuals have been centered, although centering will only affect the overall intercept value.
We now seek a transformation of the residuals of the form

$$\hat{U}^* = \hat{U}A$$

where *A* is an upper triangular matrix of order equal to the number of random coefficients at level 2, and such that

$$\hat{U}^{*T} \hat{U}^* / J = A^T \hat{U}^T \hat{U} A = A^T S A = R \qquad (3)$$

The new set of transformed residuals $\hat{U}^*$ now have covariance matrix equal to that estimated from the model, and we sample sets of residuals with replacement from $\hat{U}^*$, as described in the residual bootstrap algorithm above.

To complete the residual bootstrap, we repeat this process at every level of the model, with sampling being independent across levels. Details of how to form *A* are given in the appendix below.

3. Example

Consider the following 2-level model fitted to the tutorial data example in the *MLwiN* User's Guide, using RIGLS estimates. The model is

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij} \,. \qquad (4)$$

We simulate data from this model using the parameter estimates given in the second column of Table 1, with residuals at level 2 simulated from

$$N\left[\begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 0.2 & \\ 0.05 & 0.2 \end{pmatrix}\right]$$

and at level 1 we simulate from a chi-squared distribution with 1 degree of freedom.

Five hundred data sets were generated from this model, containing 4059 level 1 and 65 level 2 units. For each of these data sets the bootstrap parameter estimates and confidence intervals were constructed using 500 parametric and 500 non-parametric bootstrap data sets.

Table 1 gives the parameter estimates and estimated coverage probability for a nominal 95% interval computed directly from the ranked bootstrap replications for each bootstrap set, for the parametric bootstrap and Table 2 for the nonparametric bootstrap.

Both bootstrap procedures produce unbiased estimates for all the parameters. The coverage proportions are satisfactory except for the level 1 variance in the parametric bootstrap where it is only 0.55 compared to the nominal value of 0.95.

Table 1. Parametric bootstrap estimates			
Parameter	Expected value*	Bootstrap mean	Coverage proportion
β_0	**2.00**	**2.00**	**0.95**
β_1	**0.50**	**0.500**	**0.93**
σ^2_{u0}	**0.20**	**0.200**	**0.94**
σ_{u01}	**0.05**	**0.049**	**0.96**
σ^2_{u1}	**0.20**	**0.202**	**0.95**
σ^2_e	**2.00**	**2.00**	**0.55**
***The expected value for a chi squared distribution with 1 degree of freedom (=1) is added to the intercept. The variance of a chi squared distribution with 1 degree of freedom is 2.**			

Table 2. Nonparametric bootstrap estimates			
Parameter	Expected value*	Bootstrap mean	Coverage proportion
β_0	**2.00**	**2.00**	**0.95**
β_1	**0.50**	**0.500**	**0.95**
σ^2_{u0}	**0.20**	**0.198**	**0.93**
σ_{u01}	**0.05**	**0.050**	**0.95**
σ^2_{u1}	**0.20**	**0.202**	**0.94**
σ^2_e	**2.00**	**1.99**	**0.93**
***The expected value for a chi squared distribution with 1 degree of freedom (=1) is added to the intercept. The variance of a chi squared distribution with 1 degree of freedom is 2.**			

4. Conclusions

We have briefly described a residuals non-parametric bootstrap for multilevel models. This residuals bootstrap provides a robust alternative to a fully parametric bootstrap, and can be used, for example where standardised residual plots indicate departures from normality. The bootstrap can also be used to estimate other functions. For example we can estimate residuals for each bootstrap replicate and use the resulting chains for inference about the residual estimates themselves.

This non-parametric bootstrap procedure is implemented in *MLwiN* release 1.1 (Autumn 1999).

5. References

Davison, A. C. and Hinkley, D. V. *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge, 1997.

6. Appendix

To form A we note the following.
Write the Cholesky decomposition of S, in terms of a lower triangular matrix as

$$S = L_S L_S^T$$

and the Cholesky decomposition of R as

$$R = L_R L_R^T$$

We have

$$L_R L_S^{-1} \hat{U}^T \hat{U} (L_R L_S^{-1})^T / J = L_R L_S^{-1} S (L_S^{-1})^T (L_R)^T = L_R (L_R)^T = R$$

Thus, the required matrix is

$$A = (L_R L_S^{-1})^T .$$