Different Types of Bootstraps

Spenser Steele

June 19, 2014

Contents

Overall model **1**

Parametric Bootstrap **1**

Residual Bootstrap **2**

Cases Bootstrap **2**

Two-Level Semiparametric (CGR) Bootstrap **2**

Random Effect Block(REB/0) Bootstrap **2**

Random Effect Block(REB/1) Bootstrap **3**

Random Effect Block(REB/2) Bootstrap **3**

Below are three bootstraps discussed in the paper Resampling Multilevel Models by Rien van der Leeden et al. and A Random Effect Block Bootstrap for Clustered Data by chambers et al..

Overall model

$$y_i = X_i\beta + Z_i b + \epsilon_i$$
$$\epsilon_i \sim (iid)N(0, \sigma^2_{\epsilon_i})$$
$$b_i \sim (iid)N(0, D)$$
$$Cov(\epsilon_i, b) = 0$$

These are two equations used in the following outlines:

(1) $\hat{\theta}_B = \hat{\theta} - B\hat{ias}_B = 2\hat{\theta} - \theta^*_{(.)}$

(2) $\hat{se}_B = \sqrt{V\hat{a}r(\theta^*)}$

Parametric Bootstrap

1. Draw J vectors of level-2 residuals from a multivariate normal distribution with mean zero and covariance matrix $\hat{D}$.
2. Draw J vectors ϵ^*_j of sizes n_j containing level-1 residuals from a normal distribution with means zero and covariance matrices $\hat{\sigma}^2 I_{n_j}$.

3. Generate the bootstrap sample y_j^* from $y_j^* = X_j\hat{\beta} + Z_j\delta_j^* + \epsilon_j^*$.
4. Compute estimates for all parameters of the two-level model.
5. Repeat steps 1-4 B times and compute bias-corrected estimates and bootstrap standard errors using formulas 1 and 2.

Residual Bootstrap

1. Draw a sample δ_j^* of size J with replacement from the set of estimated level-2 residuals.
2. Draw J samples ϵ_{ij}^* of size n_j with replacement from the elements $\hat{\epsilon}_{ij}$.
3. Generate the bootstrap samples y_j^* from $y_j^* = X_j\hat{\beta} + Z_j\delta_j^* + \epsilon_j^*$.
4. Compute estimates for all parameters of the two-level model.
5. Repeat steps 1-4 B times and compute bias-corrected estimates and bootstrap standard errors using formulas 1 and 2.

Cases Bootstrap

1. Draw a sample of size J with replacement from the level-2 units(unit numbers)
2. For each k, draw a sample of entire cases, with replacement, from the original level-2 unit. This sample has the same size as the original unit from which the cases are drawn. Then, for each k, we have a set of data
3. Compute estimates for all parameters of the two-level model
4. Repeat steps 1-3 B times and compute bias-corrected estimates and bootstrap standard errors using formulas 1 and 2.

Two-Level Semiparametric (CGR) Bootstrap

1. The D EBLUPs $\hat{u}_i$ of the random effects u_i and the corresponding n level-1 residuals $\hat{e}_{ij} = y_{ij} - x_{ij}^T\hat{\beta} - \hat{u}_i$ are first scaled to ensure that they have empirical variances equal to $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$, respectively. The scaled level-2 residuals are $\hat{u}_i^c = \hat{\sigma}_u\hat{u}_i\{D^{-1}\sum_{k=1}^{D}\hat{u}_k^2\}^{-1/2}$ and the scaled level-1 residuals are $\hat{e}_{ij}^c = \hat{\sigma}_e\hat{e}_{ij}\{n^{-1}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\hat{e}_{kl}^2\}^{-1/2}$. Both sets of scaled residuals are then centered at zero.
2. Sample independently with replacement from $\hat{u}^c$ and $\hat{e}^c$ to get bootstrap samples u^* and e^* of D level-2 residuals and n level-11 residuals, respectively. That is, $u^* = srswr\{\hat{u}^c, D\}$ and $e^* = srswr\{\hat{e}^c, n\}$.
3. Simulate bootstrap sample data (y_{ij}^*, x_{ij}) using the model $y_{ij}^* = x_{ij}^T\hat{\beta} + u_i^* + e_{ij}^*$.
4. Fit the two-level random effects model (1) to the bootstrap sample data generated in Step (3) to obtain bootstrap parameter estimates $\hat{\theta}^* = (\hat{\beta}^*, \hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})$.
5. Repeat Steps 2-4 B times to obtain B sets of bootstrap parameter estimates.

Random Effect Block(REB/0) Bootstrap

1. Using the marginal residuals: $r_{ij} = y_{ij} - x_{ij}^T\hat{\beta}, \quad j = 1, \ldots, D$, calculate the level-2 average residuals for each of the D groups: $\bar{r}_h = \frac{1}{n_h}\sum_{j=1}^{n_h} r_{hj}, \quad h = 1, \ldots, D$, and the level-1 residuals within each group h as $r_{hj}^{(1)} = r_{hj} - \bar{r}_h, j = 1, \ldots, n_h$.
2. Sample independently and with replacement from these two sets of residuals to define bootstrap errors for levels 1 and 2. In particular, level-2 boostrap errors are given by $r^{*(2)} = srswr(\bar{r}^{(2)}, D)$, while level-1 errors in cluster i are given by $r^{*(1)} = srswr(r_{h(i)}^{(1)}, n_i)$, where $h(i) = srswr(\{1, \ldots, D\}, 1)$.

3. Simulate bootstrap sample data (y_{ij}^*, x_{ij}) using the model $y_{ij}^* = x_{ij}^T\hat{\beta} + r_i^{*(2)} + r_{ij}^{*(1)}$.
4. Fit the two-level random effects model (1) to the bootstrap sample data generated in Step 3 to obtain bootstrap parameter estimates $\hat{\theta}^* = (\hat{\beta}^*, \hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})$.
5. Repeat steps 2-4 B times to obtain B sets of bootstrap parameter estimates.

Random Effect Block(REB/1) Bootstrap

Variation 1 of the REB Bootstrap where we use centered and scaled residuals before bootstrapping.

Random Effect Block(REB/2) Bootstrap

Variation 2 of the REB Bootstrap where we apply scaling and bias adjustments after bootstrapping.

1. We first modify the REB bootstrap distributions of the logarithms of the variance components estimates so that they are empirically uncorrelated. The steps in this process are as follows:

- Let $(\log \hat{\sigma}_u^{2*})$ and $(\log \hat{\sigma}_e^{2*})$ denote the vectors of B bootstrap values $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$, respectivley. Define the $B \times 2$ matrices

$$\mathbf{S}^* = [(\log \hat{\sigma}_u^{2*}), (\log \hat{\sigma}_e^{2*})]$$

$$\mathbf{M}^* = [av(\log \hat{\sigma}_u^{*2}) \times 1_B, av(\log \hat{\sigma}_e^{*2}) \times 1_B]$$

$$\mathbf{D}^* = [sd(\log \hat{\sigma}_u^{*2}) \times 1_B, sd(\log \hat{\sigma}_e^{*2}) \times 1_B]$$

  Here, av$S$ and sd$S$ denote the average and the standard deviation of the values in the vector S, 1_B denotes a vector of ones of size B, and $\times$ denotes componentwise multiplication.
- Calculate the 2×2 covariance matrix $\mathbf{C}^* = cov(\mathbf{S}^*)$ and put

$$\mathbf{L}^* = \mathbf{M}^* + \{(\mathbf{S}^* - \mathbf{M}^*)\mathbf{C}^{*-1/2}\} \times \mathbf{D}^*.$$

- The modified bootstrap values of $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ are then obtained by exponentiating the elements of $\mathbf{L}^*$.

2. All bootstrap distributions of model parameter estimates (including the modified bootstrap distributions of the estimated variance components) are then centered at the original estimate values, using a mean correction for regression coefficients, that is,

$$(\hat{\beta}_k^{**}) = [\hat{\beta}_k 1_B + (\hat{\beta}_k^*) - av(\hat{\beta}_k^*)],$$

   and a ratio correction for variance components.