Parametric Bootstrap in Small Area Estimation

Partha Lahiri

Joint Program in Survey Methodology
University of Maryland, College Park, USA

(Based on joint work with Snigdhansu Chatterjee, School of Statistics, University of Minnesota, USA)

A Simple Two-level Model

Efron and Morris (JASA, 1975)

For $i = 1, \ldots, m$,

$$\text{Level 1: (Sampling Distribution)} \quad Y_i|\theta_i \stackrel{ind}{\sim} N(\theta_i, 1),$$
$$\text{Level 2: (Prior Distribution)} \quad \theta_i \stackrel{ind}{\sim} N(\mu, A).$$

The above model can be also viewed as a simple linear mixed model:

$$Y_i = \mu + v_i + e_i,$$

where $\{v_i\}$ and $\{e_i\}$ are independent with $v_i \stackrel{iid}{\sim} N(0, A)$ and $e_i \stackrel{iid}{\sim} N(0, 1)$ $i = 1, \cdots, m$.

The best prediction (BP) (Bayes) estimator of θ_i under the above model and squared error loss:

$$\hat{\theta}_i^B = (1 - B)Y_i + B\mu,$$

where $B = \frac{1}{1+A}$.
An empirical Bayes estimator of θ_i is given by

$$\hat{\theta}_i^{EB} = (1 - \tilde{B}) + \tilde{B}\bar{Y},$$

where $\bar{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j$,

$$\tilde{B} = \begin{cases} \hat{B} = \frac{m-3}{\sum_{j=1}^{m}(Y_j - \bar{Y})^2} & \text{if } \hat{B} < 1, \\ \frac{m-3}{m-1} & \text{otherwise.} \end{cases}$$

See Morris (1983).

A Measure of Uncertainty due to Morris (1983)

Using a flat improper prior distributions on μ and B, Morris (1983) suggested an approximation to the posterior variance of θ_i as a measure of uncertainty of $\hat{\theta}_i^{EB}$. The measure is given by

$$\begin{aligned} V(\theta_i|y) &= E\left[V(\theta_i|y,\mu,B)|y\right] + V\left[E(\theta_i|y,\mu,B)|y\right] \\ &\approx (1-\tilde{B}) + \frac{\tilde{B}}{m} + \frac{2\tilde{B}^2}{m-3}(Y_i - \bar{Y})^2 \\ &= V_i^M \text{ (say)} \end{aligned}$$

Parametric Bootstrap due to Laird and Louis (1987)

$$
\begin{aligned}
V_i^{LL} &= E_* V(\theta_i|y;\hat{\mu}^*,\hat{B}^*) + V_* E(\theta_i|y,\hat{\mu}^*,\hat{B}^*) \\
&\approx \frac{1}{R}\sum_{r=1}^{R} V(\theta_i|y;\hat{\mu}^{(r)},\hat{B}^{(r)}) + \frac{1}{R-1}\sum_{r=1}^{R}\left[\hat{\theta}_i^{EB(r)} - \bar{\hat{\theta}}_i^{EB}\right]^2 \\
&\approx (1-\hat{B}_0) + \frac{m-1}{m-5}\frac{\hat{B}_0}{m} + \frac{2\hat{B}_0^2}{m-5}(Y_i-\bar{Y})^2,
\end{aligned}
$$

where

- $V(\theta_i|y;\hat{\mu}^{(r)},\hat{B}^{(r)}) = 1 - \hat{B}^{(r)}$
- $\hat{\theta}_i^{EB(r)} = (1-\hat{B}^{(r)})Y_i + \hat{B}^{(r)}\hat{\mu}^{(r)}$
- $\bar{\hat{\theta}}_i^{EB} = R^{-1}\sum_{r=1}^{R}\hat{\theta}_i^{EB(r)}$
- the estimates $\hat{\mu}^{(r)}$ and $\hat{B}^{(r)}$ of μ and B, respectively, are based on the rth parametric bootstrap sample ($r = 1,\cdots,R$).
- $\hat{B}_0 = \frac{m-1}{m-3}\tilde{B}$.
- The difference between V_i^M and V_i^{LL} is of order $O_p(m^{-1})$.

Taylor series linearization method due to Prasad and Rao (1990)

Define $MSE(\hat{\theta}_i^{EB}) = E(\hat{\theta}_i^{EB} - \theta_i)^2$, where the expectation is with respect to the joint distribution of $\{(Y_i, \theta_i),\ i = 1, \cdots, m\}$ under the linear mixed model. For this simple model, their second-order approximation to $MSE(\hat{\theta}_i^{EB})$ is given by

$$MSE(\hat{\theta}_i^{EB}) = g_1(B) + g_2(B) + g_3(B) + o(m^{-1}),$$

where $g_2(B)$ and $g_3(B)$ are of order $O(m^{-1})$, but $g_1(B)$ is of order $O(1)$. This leads to the following second-order unbiased (or nearly unbiased) estimator of $MSE(\hat{\theta}_i^{EB})$:

$$\begin{aligned} V_i^{PR} &= g_1(\tilde{B}) + g_2(\tilde{B}) + g_3(\tilde{B}) \\ &= (1 - \tilde{B}) + \frac{\tilde{B}}{m} + \frac{2\tilde{B}}{m}, \end{aligned}$$

(Note we do not need a bias-correction for $g_1(\tilde{B})$ for this simple case). We have the second-order unbiasedness property of V_i^{PR}

$$E[V_i^{PR}] = MSE(\hat{\theta}_i^{EB}) + o(m^{-1}),$$

Parametric Bootstrap method of Butar (1997), Butar and Lahiri (2003), and Pfeffermann and Glickmann (2004)

$$
\begin{aligned}
& V_i^{BL} \\
= \; & g_1(\hat{B}) + g_2(\hat{B}) - E_* \left[\{g_1(\hat{B}^*) + g_1(\hat{B}^*)\} - \{g_1(\hat{B}) + g_2(\hat{B})\} \right] \\
& + E_* \left[\hat{\theta}_i^{EB}(\hat{B}^*) - \hat{\theta}_i^{EB}(\hat{B}) \right]^2 \\
\approx \; & g_1(\hat{B}) + g_2(\hat{B}) - \frac{1}{R} \sum_{r=1}^{R} \left[\{g_1(\hat{B}^{(r)}) + g_1(\hat{B}^{(r)})\} - \{g_1(\hat{B}) + g_2(\hat{B})\} \right] \\
& + \frac{1}{R} \sum_{r=1}^{R} \left[\hat{\theta}_i^{EB}(\hat{B}^{(r)}) - \hat{\theta}_i^{EB}(\hat{B}) \right]^2 \\
\approx \; & (1 - \tilde{B}) + \frac{\tilde{B}}{m} + \frac{2\tilde{B}^2}{m}(Y_i - \bar{Y})^2,
\end{aligned}
$$

- V_i^{BL} and V_i^{M} are identical up to order $O_p(m^{-1})$.
- V_i^{BL} satisfies the second-order unbiasedness property.

Other references on Parametric Bootstrap

- Efron (1982)
- Meza (2003)
- Lahiri (2003)
- Rao (2003)
- Hall and Maiti (2006a,2006b)
- Hall (2006)
- Jiang and Lahiri (2006)
- Chatterjee and Lahiri (2007)
- González-Manteiga et al. (2007)
- González-Manteiga et al. (2008)
- Efron (2012)

A Class of Area Level Models

For $i = 1, \ldots, m$,

$$\text{Level 1:} \quad Y_i = h_{1i}^{-1}(\tilde{Y}_i)|\theta_i \overset{ind}{\sim} N(\theta_i, D_i),$$
$$\text{Level 2:} \quad \theta_i \overset{ind}{\sim} N(x_i'\beta, A),$$

where

- $\tilde{Y}_i$: direct estimator of small area parameter of interest $h_{1i}(\theta_i)$ (mean, total, proportion)
- $h_{1i}(Y_i)$ is a one-to-one measurable function of Y_i
- D_i is the sampling variance of Y_i usually approximated or/and estimated using GVF; for inference from such models D_i's are assumed to be known.
- x_i^T: a $p \times 1$ column vector of known auxiliary variables.
- Two types of parameters: $h_{1i}(\theta_i)$ (high-dimensional or small area parameters) and (β, A) low dimensional hyperparameters

A Few Examples of Transformation Used Practice

1. **SAIPE State level model for poverty rate:** $Y_i = \tilde{Y}_i$; D_i are estimated by a replication-based method.
2. **Firm Alarm Probabilities (Carter and Rolph 1974):** $Y_i = \arcsin(\sqrt{\tilde{Y}_i})$, where n_i is the sample size for area i; $D_i = \frac{1}{4n_i}$. In Chilean poverty mapping, similar transformation is used with n_i representing effective sample size to incorporate complex sample design.
3. **Baseball Data Analysis (Efron and Morris 1975):** $Y_i = \sqrt{n_i}\arcsin(2\tilde{Y}_i - 1)$, where n_i is the sample size for area i; $D_i = 1$.
4. **Per-capita income (Fay and Herriot 1979):** $Y_{1i} = \log(\tilde{Y}_i)$; $D_i = 9/N_i$, where N_i is the population size.

In practice, small area estimators are obtained using several steps. All or a subset of the following steps are generally used:

- **Step 1: The Best Prediction (Bayes) Estimator**
  Note that

$$\theta_i | Y_i; (\beta, A) \overset{ind}{\sim} N\left[(1 - B_i)Y_i + B_i x_i^T \beta, (1 - B_i)D_i\right],$$

  where

$$B_i = \frac{D_i}{A + D_i},\ i = 1, \ldots, m.$$

  This leads to the best prediction (BP) (same as the Bayes) estimator given by

$$\tilde{\theta}_i = E[\theta_i | Y_i; (\beta, A)] = (1 - B_i)Y_i + B_i x_i^T \beta.$$

- **Step 2: Empirical Best Linear Unbiased or Empirical Bayes (EB)**

$$\hat{\theta}_{mi} = (1 - \hat{B}_i)Y_i + \hat{B}_i x_i^T \hat{\beta}$$

where

$$B_i = \frac{D_i}{\hat{A} + D_i},\ i = 1, \ldots, m;$$

$\hat{\beta}$ and $\hat{A}$ are consistent estimators of β and A, respectively.

- **Step 3: Winsorization or Limited Translation**

$$\begin{aligned} T_{m1i} &= \begin{cases} L_{mi} & \text{if } \hat{\theta}_{mi} < L_{mi} < U_{mi}, \\ \hat{\theta}_{mi} & \text{if } L_{mi} < \hat{\theta}_{mi} < U_{mi}, \\ U_{mi} & \text{if } L_{mi} < U_{mi} < \hat{\theta}_{mi}. \end{cases} \\ &= \text{median}(L_{mi}, \hat{\theta}_{mi}, U_{mi}) \end{aligned}$$

- **Step 4: Benchmarking**:

$$T_{m2i} = K_{mi} \left(\sum_{i=1}^{m} h_{1i}(T_{m1i}) \right)^{-1} h_{1i}(T_{m1i}),$$

where K_{mi} is a known constant, or

$$T_{m2i} = K_{mi} - m^{-1} \left(\sum_{i=1}^{m} h_{1i}(T_{m1i}) \right) + h_{1i}(T_{m1i}).$$

In general,

$$T_{m2i} = h_{m2i}(T_{m11}, \ldots, T_{m1n}, K_{mi})$$

where h_{m2i} is a continuous, and hence measurable, function.

How do we estimate MSE of T_{m2i}?

- Apply parametric bootstrap method . This is quite straightforward – we imitate the above steps on data generated using $\hat{\beta}$ and $\hat{A}$.
- Mean squared error (MSE) of T_{m2i} is estimated by $E_*\left[T_{m2i*} - h_{1i}(\theta_{i*})\right]^2$, where E_* is expectation with respect to the parametric bootstrap distribution of $[T_{m2i*}, h_{1i}(\theta_{i*})]$.
- In practice, we use the Monte Carlo approximation to $E_*\left[T_{m2i*} - h_{1i}(\theta_{i*})\right]^2$:

$$\widehat{MSE}_{B1.Direct} = \frac{1}{B}\sum_{b=1}^{B}[T_{m2i,b} - h_{1i}(\theta_{i,b})]^2,$$

  where $T_{m2i,b}$ and $h_{1i}(\theta_{i,b})$ are based on the bth parametric bootstrap sample ($b = 1, \cdots, B$).
- Under regularity conditions,

$$E\left[\widehat{MSE}_{B1.Direct}\right] = MSE + O(m^{-1}) + O(B^{-1/2}).$$

Here is why parametric bootstrap works: the functions h_{1i} and h_{2i} present no challenges, either they are 1-1 or continuous. The Winsorization step is about computing a median of three random variables for each small area. So, we are looking to estimate the distribution of a smooth function of

$$\begin{aligned} Z_{mi} &= (Y_1, \ldots, Y_m, \theta_1, \ldots, \theta_m, \\ &\quad K_{mi}, L_{m1}, \ldots, L_{mm}, U_{m1}, \ldots, U_{mm}) \in \mathbb{R}^{4m+1}, \end{aligned}$$

whose distribution is $F_Z(\cdot; \beta, A)$. Our assumptions make sure that F_Z is a smooth function of the parameters, and hence the parametric bootstrap approximation $F_Z(\cdot; \hat{\beta}, \hat{A})$ converges. There is some more details here that we are skipping.

General MSPE Estimation: Double bootstrap

- Estimate parameters $\xi = (\beta, A)$ with $\hat{\xi} = (\hat{\beta}, \hat{A})$.
- For $b = 1, \ldots, B$, generate parametric bootstrap results.
- Within each first-layer parametric bootstrap step $b = 1, \ldots, B$, implement a further round of bootstrap steps based on $\hat{\xi}_b$ to get $Y_{ib\tilde{b}}$'s, $\theta_{ib\tilde{b}}$'s, $T_{ib\tilde{b}}$'s etc, for $\tilde{b} = 1, \ldots, \tilde{B}$.
- For each $b = 1, \ldots, B$ and for each $\tilde{b} = 1, \ldots, \tilde{B}$ within each b, obtain $\left(\theta_{ib\tilde{b}} - T_{ib\tilde{b}}\right)^2$, and correct the bias in $\widehat{MSPE}_{B1.Direct}$ using these to get $\widehat{MSPE}_{B2.Direct}$.

$$E\left[\widehat{MSPE}_{B2.Direct}\right] = MSPE + O(n^{-3/2}) + O(B^{-1/2}) + O(\tilde{B}^{-1/2}).$$

- There is no unique, good way of combining the $\widehat{MSPE}$'s. Several choices available (see Hall (1988, 1992), others).
- This is intense computation, $O(B\tilde{B})$. Two seconds of PB is 1.5 days of DPB!! ($n = 15 \rightarrow B = \tilde{B} \approx 3000 \rightarrow B\tilde{B} \approx 10^7$).

A Monte Carlo Simulation

- For the simulation, we consider the following area level model:

$$Y_i = h_{1i}^{-1}(\tilde{Y}_i) = \mu + v_i + e_i,\ i = 1, \cdots, m,$$

  where $\{v_i\}$ and $\{e_i\}$ are independent with $e_i \sim (0, D_i)$ and $v_i \sim (0, A)$, (a, b) denoting a distribution with mean a and variance b.
- Consider 5 groups of small areas, each with 3 small areas with the same D_i. That is $m = 15$.
- D_i pattern: $(2, 0.6, 0.5, 0.4, 0.2)$. That is, each of the 3 small areas in the first group has $D_i = 2$, each of the 3 small areas in the second group has $D_i = 0.6$, and so on.
- To generate $\{(Y_i, \theta_i),\ i = 1, \cdots, m\}$, we set unknown hyperparameters as $A = 1$ and $\mu = 0$.

A Monte Carlo Simulation

We generate $R = 1000$ independent sets of $\{(Y_i, \theta_i),\ i = 1, \cdots, m\}$ for each of the following two cases:

- Case 1: No transformation, that is, $Y_i = \tilde{Y}_i$; both $\{e_i\}$ and $\{v_i\}$ are independently generated from normal distributions.
- Case 2: Exactly like in Case 1 except that v_i are generated from a shifted exponential or double exponential distribution.

Target parameters: θ_i for both cases 1 and 2

For Cases 1 and 2, we consider the following two estimators

- (i) T_{1i}, EB;
- (ii) Winsorized and benchmarked version of EB, say T_{2i}. That is, we first winsorise T_{1i} so that winsorised T_{1i}, say T_{1i}^{Win}, does not deviate from the direct Y_i by more than $\sqrt{D_i}$ and then benchmark T_{1i}^{Win} so that the final estimates T_{2i} add up to $\sum_{j=1}^{m} Y_j$.

A Monte Carlo Simulation: Different MSE Estimators

For cases 1 and 2, we consider the following MSE estimators of T_1:

- Naive;
- PR Taylor series;
- BL bias-corrected parametric bootstrap;
- PG parametric bootstrap;
- Our proposed single-stage parametric bootstrap.

For T_2, we take the same.

Monte Carlo Simulation: Cases 1 and 2

D	2.000	0.60	0.50	0.40	0.200
T_1	78.16	43.51	38.81	33.54	20.24
T_2	39.071	22.13	19.83	17.23	10.665

Table : True MSE values for EBLUP (T_1) and Winsorized, benchmarked EBLUP (T_2); the Prasad-Rao parameter estimators used.

Comparison of different estimators of MSE of T_1

Percent Relative Bias = $100 \times \frac{E(\widehat{MSE}) - MSE}{MSE}$

D	2.000	0.60	0.50	0.40	0.200
True MSE	78.16	43.51	38.81	33.54	20.24
Naive	-18.16	-21.71	-22.22	-22.96	-25.08
PR	4.35	8.34	9.73	12.00	29.71
B1.Direct	-3.642	-8.272	-8.574	-8.505	-7.15
	(2.744)	(3.396)	(2.812)	(2.893)	(2.482)
B1.PG11	0.368	-2.371	-2.525	-2.763	-2.05
BL	1.965	-1.587	-1.932	-1.996	-1.47
B2	-0.921	-1.163	-3.48	-3.002	-0.03

Table : True MSE values and percentage relative biases for naive, Prasad-Rao Taylor series (PR), parametric bootstrap with no bias correction (B1.Direct), Pfeffermann-Glickmann bias corrected parametric bootstrap (B1.PG11), Butar-Lahiri bias-corrected parametric bootstrap (BL) MSE estimators. B2 is double bootstrap. Red= $< 1\%$, Magenta= $< 3\%$, Fuschia= $< 10\%$; numbers in the parentheses are for Case 2

Comparison of different estimators of MSE of T_2

D	2.000	0.60	0.50	0.40	0.200
True MSE	39.071	22.13	19.83	17.23	10.665
Naive	60.5	55.2	54.0	52.3	45.4
PR	106.0	111.3	111.5	111.7	115.2
B1.Direct	-6.424	-8.59	-8.45	-8.31	-6.654
	(4.092)	(2.532)	(2.02)	(1.372)	(1.142)
B1.PG11	99.173	96.31	95.74	94.94	92.599
BL	25.790	22.79	23.21	24.50	30.620
B2	(-4.018)	(-1.92)	(-3.003)	(-2.996)	(0.938)

Table : True MSE values and percentage relative biases for several MSE estimators. Fuschia= $< 10\%$; numbers in the parentheses are for Case 2.

Thank you