On Parametric Bootstrap Methods for Small Area Prediction

Authors: Peter Hall and Tapabrata Maiti

Outline

Introduction

- In recent years, the demand for **small area statistics** has greatly increased worldwide.
- Examples of growing demand from public and private, such as:
  - Formulating policies and programs
  - Allocation of government funds
  - Regional planning
  - Small businesses' decisions relying heavily on local socio-economic conditions.

Introduction

Problem:

- The **stochastic variability** of small area estimators which are based solely on data from the given area, can be unduly *large*.
- Partly as a result, **the estimation of mean-squared error** (MSE), and **the correction of bias** of those estimators has become a centre-piece of small area inference.
- In particular, there is strong demand for reliable, simple-to-apply **methods for MSE esimtation**. (See Rao (2003) for a detailed discussion)

Introduction

- In this presentation we propose **bootstrap methods** for MSE estimation.

- Advantages of the new techniques are:
  - Do not require derivation of analytical expansions
  - They give non-negative, bias-corrected estimators of MSE
  - Are at least as accurate as existing techniques

Background

- Resampling methods arguably have their roots in survey sampling
  - Work of Hubback and Mahalanobis in India
  - Work of Gurney and McCarthy in the USA
- More recent resampling contributions have been surveyed by
  - Lahiri (2003a) who gives an authoritative account of the bootstrap for small area inference
  - Lahiri (2003b) which reviews jackknife methods in the small area estimation problem.

Background

Bayesian vs Frequentist approaches

- Much of the existing small area bootstrap work (Butar Butar and Lahiri (2003) and Meza (2003)) is founded on **empirical Bayes arguments**.
- The approach in this presentation is distinctly **frequentist**.

Background

- Standard small area models usually have two levels.
  1. **Step 1:** describes **sampling**.
  2. **Step 2:** describes the **population**.

- Methods for inference also have two steps.
  1. **Step 1:** Here the predictor involves **unknown parameters.**
  2. **Step 2:** Here these are replaced by **suitable estimators.** (This final form is commonly known as an *empirical predictor*.)

- Below we consider the following approaches:
  - **BLUP:** The *best linear unbiased predictor*
  - **BP:** The *best predictor*

Methodology

Model for two-stage sampling

- Let $Q(\mu, \xi)$ and $R(\mu, \eta)$ represent **univariate distributions** that are determined by a scalar parameter μ, denoting expected value in each case, and by other, possibly vector parameters ξ and η.
- For an integer $n_i \geq 1$, denote by $f_i(\beta)$ a known **smooth function** of explanatory variables $X_{i1}, ..., X_{in_i}$ and the vector β.

Methodology

Model for two-stage sampling

Data pairs (X_{ij}, Y_{ij}), for $1 \leq i \leq n_i$, with Y_{ij} a scalar, are observed and generated as follows:

1. Random variables Θ_i are drawn from the distribution $Q\{f_i(\beta), \xi\}$
2. Given the values of $X_i = (X_{i1}, ..., X_{in_i})$ and Θ_i, Y_{ij} for $1 \leq j \leq n_i$ are independent and drawn from $R\{\psi(\Theta_i), \eta_i\}$, where ψ is a known link function.
3. Inference is conducted conditionally on the set $\chi = \{X_{ij} : 1 \leq i \leq n, 1 \leq j \leq n_i\}$.

Methodology

Remarks:

- The parameters β and ξ are unknown.
- In many instances η_i is known, typically as a known function of X_i, although in general η_i would consist of both known and unknown components, the latter being the same for each i.
- A few examples include the Fay-Herriot model and Battesse-Harter-Fuller model which were mentioned in previous presentations.

Methodology

We may write:

$$Y_{ij} = \{f_i(\beta) + U_i\} + V_{ij}$$

where $U_i = \Theta_i - f_i(\beta)$ and $V_{ij} = Y_{ij} - (\Theta_i)$.

The variable U_i has zero mean and conditional on U_i, V_{ij} has zero mean.

Parameter estimation

Let η' be the common unknown part of each vector η_i.

- If $\psi(x) \equiv x$, then:
  - β can be estimated root n consistently by *least squares*.
  - ξ and η' can be estimated by *moment methods*.
- Maximum likelihood can also be employed to estimate all three parameters.
  - Note that this can be extended to the case when ψ is *not* the identity.

Predictors Θ_i^{BLUP} and Θ_i^{BP} of Θ_i

Θ_i^{BLUP}

The best linear unbiased predictor Θ_i^{BLUP} is given by:

$$\Theta_i^{BLUP} = f_i(\beta) + \frac{cov(\bar{Y}_i, \Theta_i|\chi)}{var(\bar{Y}_i|\chi)}\{\bar{Y}_i - f_i(\beta)\} \tag{1}$$

where

$$\bar{Y}_i = n_i^{-1} \sum_j Y_{ij}.$$

(Ghosh and Maiti [2004])

- The empirical version, $\hat{\Theta}_i^{BLUP}$, of Θ_i^{BLUP} is obtained by replacing β, ξ and η_i by their estimators $\hat{\beta}$, $\hat{\xi}$ and $\hat{\eta}_i$ in equation 1.
- For a general link function, numerical integration methods must generally be used, since there are no closed form expressions for equation 1.

Predictors Θ_i^{BLUP} and Θ_i^{BP} of Θ_i

Θ_i^{BP}

The best predictor Θ_i^{BP} is given by

$$\Theta_i^{BP} = f_i(\beta) + E(U_i|\bar{Y}_i) \tag{2}$$

- Again, the empirical version, $\hat{\Theta}_i^{BP}$, of Θ_i^{BP} is obtained by replacing β, ξ and η_i by their estimators $\hat{\beta}$, $\hat{\xi}$ and $\hat{\eta}_i$ in equation 2.
- Similarly to the BLUP case, calculation in the general case typically involves techniques such as numerical integration, unless ψ i the identity.

Mean-square prediction error

- The mean-squared predictive error of the 'ideal' predictor, $\Theta_i^{pred} = \Theta_i^{BLUP}$ or $\Theta_i^{pred} = \Theta_i^{BP}$, is given by

$$MSEpred_i = E\{(\Theta_i^{pred} - \Theta_i)^2 | \chi\}$$

- The counterpart for the practical predictor $\hat{\Theta}_i^{pred} = \hat{\Theta}_i^{BLUP}$ or $\hat{\Theta}_i^{pred} = \hat{\Theta}_i^{BP}$ is given by

$$MSE_i = E\{(\hat{\Theta}_i^{pred} - \Theta_i)^2 | \chi\}$$

Mean-square prediction error

A **bootstrap estimator** of MSE_i may be constructed as follows.

- Conditionally on the data $\mathcal{Z} = \{(X_{ij}, Y_{ij}) : 1 \leq i \leq n, 1 \leq j \leq n_i\}$, draw Θ_i^* from the distribution $Q\{f_i(\hat{\beta}), \hat{\xi}\}$.
- Given Θ_i^*, draw Y_{ij}^* from $R\{\psi(\Theta_i^*), \hat{\eta}_i\}$.
- Compute $\hat{\beta}^*$,$\hat{\xi}^*$ and $\hat{\eta}_i^*$ from the data

$$\mathcal{Z}^* = \{(X_{ij}, Y_{ij}^*) : 1 \leq i \leq n, 1 \leq j \leq n_i\}.$$

Mean-square prediction error

- Set
$$\bar{Y}_i^* = n_i^{-1} \sum_j Y_{ij}^*,$$
and let
$$\hat{\Theta}_i^{*pred} = K_i(\bar{Y}_i^* | \hat{\beta}^*, \hat{\xi}^*, \hat{\eta}_i^*) \tag{3}$$
denote the *bootstrap* version of $\hat{\Theta}_i^{pred}$, where K_i denotes a known function.
- Then our bootstrap estimator of MSE_i is
$$M\hat{S}E_i = E\{(\hat{\Theta}_i^{*pred} - \Theta_i^*)^2 | \mathcal{Z}\}. \tag{4}$$

Bias of MSE

In regular cases,

$$E(M\hat{S}E_i|\chi) = MSE_i + \frac{b_i(\beta, \xi, \eta_i)}{n} + O(n^{-1}) \tag{5}$$

where b_i is a smooth function of its arguments.

Analytical bias correction

- It is of significant practical interest to correct for the principal bias contribution in the above equation.
- When ψ is the identity, the quantity $b_i(\beta, \xi, \eta_i)$ admits a reasonably simple formula.
- Replacing β, ξ and η_i in that formula by their estimators, we may construct an **analytically bias-corrected** estimator of MSE_i:

$$\hat{MSE}_i^{bc} = \hat{MSE}_i - \frac{b_i(\hat{\beta}, \hat{\xi}, \hat{\eta}_i)}{n} \tag{6}$$

Analytical bias correction

Since

$$E\{b_i(\hat{\beta}, \hat{\xi}, \hat{\eta}_i)\} = b_i(\beta, \xi, \eta_i) + O(n^{-1})$$

then equations 5 and 6 imply

$$E(M\hat{S}E_i^{bc}|\chi) = MSE_i + O(n^{-2})$$

In other words, bias is **reduced** from the standard level $O(n^{-1})$ (given in 5), to only $O(n^{-2})$.

Bias correction using bootstrap methods

Analytical bias correction is usually precluded by:

1. The complexity of a formula for $b_i(\beta, \xi, \eta_i)$
2. The unattractiveness, to practitioners, of working out a formula for b_i in the context of a new model they wish to use.

The **double-bootstrap approach** to bias correction is therefore a good alternative.

Bias correction using bootstrap methods

Above, we described a method for creating the data Y_{ij}^* from which we assembled the data set $\mathcal{Z}^*$ and then calculated the bootstrap version, $(\hat{\beta}^*, \hat{\xi}^*, \hat{\eta}_i^*)$ of $(\hat{\beta}, \hat{\xi}, \hat{\eta}_i)$.

- Conditionally on $\mathcal{Z}^*$, draw Θ_i^{**} by sampling randomly from the distribution $Q\{f_i(\hat{\beta}^*), \hat{\xi}^*\}$.
- Given Θ_i^{**}, draw Y_{ij}^{**} from the distribution $R(\Theta_i^{**}, \hat{\eta}_i^*)$.
- Compute $(\hat{\beta}^{**}, \hat{\xi}^{**}, \hat{\eta}_i^{**})$ from the data in

$$\mathcal{Z}^{**} = \{(X_{ij}, Y_{ij}^{**}) : 1 \le i \le n, 1 \le j \le n_i\}.$$

- Put

$$\bar{Y}_i^{**} = n_i^{-1} \sum_j Y_{ij}^{**}.$$

Bias correction using bootstrap methods

Analogously to equations 3 and 4, define

$$\hat{\Theta}_i^{**pred} = K_i(\bar{Y}_i^{**}|\hat{\beta}^{**}, \hat{\xi}^{**}, \hat{\eta}_i^{**}), \tag{7}$$

$$\hat{MSE}_i^{*} = E\{(\hat{\Theta}_i^{**pred} - \Theta_i^{**})^2|\mathcal{Z}^*\}. \tag{8}$$

where the function K_i again denotes a known function.

- We view $\hat{MSE}_i^{*}$ as an estimator, in the bootstrap world, of $\hat{u} \equiv \hat{MSE}_i$.
- Its conditional mean, $\hat{v} = E(\hat{MSE}_i^{*}|\mathcal{Z})$ is an estimator of $E(\hat{u}) = E(\hat{MSE}_i)$.

Bias correction using bootstrap methods

Conventional additively and multiplicatively bias-corrected estimators:

- $\hat{MSE}_i^{add-bc} = 2\hat{u} - \hat{v}$
- $\hat{MSE}_i^{mult-bc} = \hat{u}^2/\hat{v}$

- The estimator $\hat{MSE}_i^{add-bc}$ is **attractive** when $\hat{MSE}_i$ is **positively biased**, otherwise, when $\hat{u} \leq \hat{v}$, it can give **too high a degree of correction** or even **a negative value**.
- When the estimator $\hat{u}$ seems to be negatively biased, using the multiplicative approach can give **unreliable results** because of the effect of **dividing by the stochastically variable quantity** $\hat{v}$.

Bias correction using bootstrap methods

This difficulty is reduced if we employ $exp\{-(\hat{v}-\hat{u})/\hat{v}\}\hat{u}$ instead of $\hat{u}^2/\hat{v}$, which is **strictly greater** but preserves the **same degree of first-order bias correction.**

- $M\hat{S}E_i^{bc1} = \begin{cases} 2\hat{u}-\hat{v} \\ exp\{-(\hat{v}-\hat{u})/\hat{v}\}\hat{u} \end{cases}$

In almost all cases, $M\hat{S}E_i^{bc1}$ gave the greatest degree of bias correction. However, **bias reduction** almost invariably **increases the variance** and can lead to a consequent **increase in MSE.** The estimator $M\hat{S}E_i^{bc1}$ sometimes suffers from this difficulty.

Bias correction using bootstrap methods

To overcome this problem, an approach was sought which gave **less bias reduction** but suffered **less** from **inflation of variance or MSE**. The final expression is the following:

- $$\hat{MSE}_i^{bc2} = \begin{cases} \hat{u} + n^{-1} tan^{-1}\{n(\hat{u} - \hat{v})\} \\ \hat{u}^2/[\hat{u} + n^{-1} tan^{-1}\{n(\hat{v} - \hat{u})\}] \end{cases}$$

Concluding words

- Double bootstrap is a good method to use to reduce the order of bias.
- Bias reduction can increase the variance and lead to MSE increase.
- The estimator $\hat{MSE}_i^{bc2}$ is suggested as one which overcomes this problem.

Thank you

Thank you for listening! :)

References

On parametric bootstrap methods for small area prediction,
Hall, P. and Maiti, T. [2006]