

Discussion of Papers on Resampling Methods in Small Area Estimation

Sharon L. Lohr

Department of Mathematics and Statistics, Arizona State University, Tempe AZ 85287-1804

KEY WORDS: Bootstrap, Jackknife, Numerical Properties.

1. Introduction

I am honored to be invited to discuss these very interesting papers on resampling methods in small area estimation. The authors of the papers have made great contributions to the field of small area estimation, and the papers presented here extend these.

The setup considered in Rao's paper for small area estimation is as follows: As in Rao (2003), there are m small areas, and the observations from each small area are assumed independent. We assume a model for the data with two stages. In stage 1, assume $\mathbf{y}_i | \theta_i \sim f_1(\mathbf{y}_i | \theta_i, \phi_1)$. The model in stage 2 relates the characteristics of interest θ_i to other areas and covariates \mathbf{x}_i : $\theta_i | \mathbf{x}_i \sim f_2(\phi_2)$. In the model, $\phi = (\phi_1, \phi_2)$ is a vector of unknown parameters. The goal is to predict θ_i , or a function of θ_i , and estimate its mean squared error. Under the distributional assumptions, if the parameters ϕ are known the best predictor of θ_i is

$$\tilde{\theta}_i = E(\theta_i | \mathbf{y}_i, \phi) = h(\mathbf{y}_i, \phi).$$

However, in practice, the parameters are unknown and must be estimated from the data. In that case, the empirical best predictor substitutes a consistent estimator $\hat{\phi}$ for ϕ , and is given by:

$$\hat{\theta}_i = h(\mathbf{y}_i, \hat{\phi}).$$

Because of the extra uncertainty due to estimating the parameters, $\text{MSE}(\hat{\theta}_i) < \text{MSE}(\tilde{\theta}_i)$; the mean squared error of $\hat{\theta}_i$ can be written as

$$\begin{aligned} \text{MSE}(\hat{\theta}_i) &= M_{1i}(\phi) + M_{2i}(\phi) \\ &= \text{MSE}(\tilde{\theta}_i) + \text{extra from estimating } \phi. \end{aligned}$$

In the Fay-Herriot (1979) model, for example, we can write $y_i = \mathbf{x}_i^T \beta + v_i + e_i$ where $v_i \sim N(0, \sigma_v^2)$, $e_i \sim N(0, \psi_i)$ with ψ_i assumed known, and $\phi = (\beta, \sigma_v^2)$. In this situation, Prasad and Rao (1990) showed that $M_{1i}(\phi) = g_{1i}(\phi) = O(1)$ and $M_{2i}(\phi) = g_{2i}(\phi) + g_{3i}(\phi) + o(m^{-1})$, where the $O(m^{-1})$ terms $g_{2i}(\phi)$ and $g_{3i}(\phi)$ represent the extra variability due to estimating β and σ_v^2 , respectively.

2. Resampling Methods for MSE Estimation

Numerous approaches have been proposed for estimating the mean squared error. The naive estimator plugs in the estimator of ϕ into the $O(1)$ term of the mean squared

error, using $g_{1i}(\hat{\phi})$. Since this does not pick up the extra variability due to estimating ϕ , it underestimates the mean squared error. SAS PROC MIXED, for linear mixed models, includes the variability due to estimating β , but not that due to estimating σ_v^2 , using $g_{1i}(\hat{\phi}) + g_{2i}(\hat{\phi})$. This too underestimates the mean squared error. Prasad and Rao (1990) and Datta and Lahiri (2000) considered an analytic approach to estimating the mean squared error, using

$$g_{1i}(\hat{\phi}) + g_{2i}(\hat{\phi}) + 2g_{3i}(\hat{\phi}) + \text{bias term}.$$

The analytic approach produces an estimator of mean squared error that is unbiased to $o(m^{-1})$ terms, but requires an explicit form of g_{1i} , g_{2i} , and g_{3i} . The g_{3i} and bias terms in this model differ when different estimators of σ_v^2 are used; in fact, it is this complication that makes it difficult for standard software to include the g_{3i} term in the calculations. Thus, in the linear model, a resampling method that works for any estimator of σ_v^2 could be used to give mean squared error estimators that are approximately unbiased regardless of choice of estimator of σ_v^2 .

In this session, Rao and Lahiri presented resampling methods for estimating $\text{MSE}(\hat{\theta}_i)$. Some view resampling methods as substituting computer thinking for human thinking. As these papers illustrate, this is far from the truth when it comes to employing resampling methods in small area estimation: one must think very hard indeed to be able to use these methods. The theoretical conditions required for the mean squared error estimators to be approximately unbiased are quite complicated in some of the resampling research discussed in these papers.

Rao reviewed the history and recent development of resampling methods in small area estimation, and presented results for an area-specific jackknife. The jackknife takes the form of $g_{1i}(\hat{\phi})$ plus a jackknife bias correction plus a jackknife variance term. Jiang, Lahiri and Wan (2002) and Lohr and Rao (2007) give theoretical results showing that the unconditional and area-specific jackknife estimators of $\text{MSE}(\hat{\theta}_i)$ are unbiased up to $o(m^{-1})$ terms for linear and generalized linear mixed models. The jackknife does somewhat reduce the human thinking involved in that it does not require an explicit form for g_{2i} and g_{3i} . However, the jackknife estimators still require g_{1i} to be derived and calculated. While the form of g_{1i} is easy to compute in the linear model, it can be quite difficult to compute in generalized linear mixed models. In addition, the bias correction can result in negative values for the estimated mean squared error, which can be awkward when reporting results to a client.

Lahiri presented a method for using parametric bootstrap for mean squared error estimation. This bootstrap method also does not require the explicit form of g_{2i} and g_{3i} . Instead one needs to generate random variables from distributions. Hall and Maiti (2006), in related work, presented a nonparametric bootstrap method for the nested error regression model. They used a double bootstrap to avoid normality assumptions. As with the jackknife, a bias correction in this method can result in negative values for estimated mean squared errors.

A very promising development is the use of the bootstrap for constructing confidence intervals directly. The results presented by Lahiri in this session were for the linear model, but the method also has potential for use in generalized linear models; I look forward to seeing the authors' future work in that very important area.

3. Numerical Issues

One issue that has not received a great deal of attention in the small area resampling literature to date is the effect of numerical issues on the accuracy of the mean squared error estimates. These issues often do not arise in simulation studies, since many researchers use either the linear mixed model or a beta-binomial model in their simulations; efficient algorithms exist for calculating maximum likelihood estimates in both of these situations, and the functions used for calculating $\hat{\theta}_i$ and g_{1i} have closed form. But in other models such as a Poisson-lognormal model, $\hat{\theta}_i$ and g_{1i} must be calculated iteratively, and numerical errors from the iterative estimates can have devastating effect.

In the Prasad-Rao (1990) approach, one must derive analytical expressions for g_{1i} , g_{2i} , and g_{3i} . Once that is done, however, each term only needs to be calculated once. The Prasad-Rao approach is consequently relatively insensitive to numerical errors.

SAS PROC GLIMMIX uses linearization-based methods to calculate small area estimates. It is a doubly iterative procedure, first employing an approximation to the model, then finding maximum likelihood estimates in the approximate model.

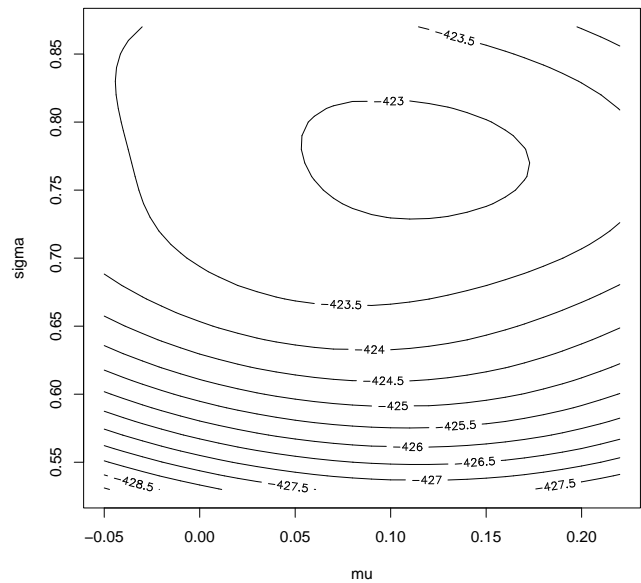
The jackknife is also doubly iterative: one must estimate ϕ and $g_{1i}(\phi)$ for every jackknife iteration. The bootstrap methods are triply iterative: they require calculating the maximum likelihood estimators, followed by two levels of bootstrap in which $\hat{\theta}_i^*$ and $\hat{\theta}_i^{**}$ are calculated. All of these steps lead to great potential for numerical errors.

For an example, consider the Poisson-lognormal model. At stage 1, assume $y_i | \theta_i \sim \text{Poisson}(\theta_i)$. At stage 2, assume $\log(\theta_i) \sim N(\mu, \sigma^2)$. The marginal distribution of y_i , used in calculating the maximum likelihood estimates of μ and σ^2 , is given by $f(y_i) = \eta(0, y_i)/y_i!$, where

$$\eta(k, y_i) = E_Z[\exp\{-e^{\sigma Z + \mu} + (y_i + k)(\sigma Z + \mu)\}]$$

for $Z \sim N(0, 1)$. The empirical best predictor $\hat{\theta}_i$ requires calculation of $\eta(0, y_i)$ and $\eta(1, y_i)$; the first-order term

Figure 1: Approximate log likelihood function for lip cancer data.



in the mean squared error, $g_{1i}(\hat{\phi})$, requires calculating $\eta(0, y_i)$, $\eta(1, y_i)$, and $\eta(2, y_i)$. Each of these integrals must be calculated numerically.

Figure 1 displays contours of the calculated log likelihood function for the lip cancer data in Clayton and Kaldor (1987). These contours were calculated using 31-point Gauss-Hermite quadrature. But the log likelihood function, and consequently also the maximum likelihood estimates, are highly sensitive to the number of quadrature points used; if 15-point quadrature is used, the maximum likelihood estimates are different. Other methods for calculating maximum likelihood estimates can be worse; some algorithms specify calculating these only to a tolerance of 0.001, which is not accurate enough for use in resampling methods.

If we were only calculating the maximum likelihood estimates and the $g_{1i}(\hat{\phi})$ term once, the numerical errors might be relatively unimportant. But in resampling methods, these are calculated many times. The bias correction used in jackknife methods depends on $\sum_j [g_{1i}(\hat{\phi}_{-j}) - g_{1i}(\hat{\phi})]$, where $\hat{\phi}_{-j}$ is the maximum likelihood estimate of ϕ using all of the data except that in area j . The jackknife variance correction depends on $\sum_j [\hat{\theta}_i(\hat{\phi}_{-j}) - \hat{\theta}_i(\hat{\phi})]^2$. Each summand in these terms is small. If the algorithm used to find the maximum likelihood estimates has low precision, the jackknife correction terms do not accurately estimate the true corrections. It is possible in that situation for the jackknife to be less accurate than the naive estimator of mean squared error.

Similar problems can occur in bootstrap methods. Rao discussed the important issue of how many bootstrap iterations need to be calculated at each level to remove the

bias up to $o(m^{-1})$ terms. As in the jackknife, however, if numerically inaccurate methods are used to calculate estimates, the numerical errors then can propagate through the bootstrap levels. In that case, the bootstrap may not achieve the claimed reduction in bias, no matter how many bootstrap iterations are run. Consequently, great care must be taken so that numerical errors do not result in an MSE estimator that is actually worse than the naive estimator.

4. Conclusions

Resampling methods provide great promise for estimating MSE of small area predictors. As we saw in the papers of this session, the theory is highly intricate. Implementation, especially in nonlinear models, requires thought and care. Particular attention needs to be devoted to calculating numerically accurate estimates, so that the theoretical gains from the jackknife and bootstrap corrections can be achieved in practice.

Today's speakers have contributed a great deal to the development of resampling methods in small area estimation, and I look forward to their future contributions.

Acknowledgements

This work was partially supported by the National Science Foundation under grant 0604373.

References

- Clayton, D. and Kaldor, J. (1987). "Empirical Bayes Estimates of Age-standardized Relative Risks for use in Disease Mapping," *Biometrics*, **43**, 671–681.
- Datta, G. S. and Lahiri, P. (2000), "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems," *Statistica Sinica*, **10**, 613–627.
- Fay, R. E. and Herriot, R. A. (1979) "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, **74**, 269–277.
- Hall, P. and Maiti, T. (2006). "On Parametric Bootstrap Methods for Small Area Estimation," *Journal of the Royal Statistical Society B*, **68**, 221–238.
- Jiang, J., Lahiri, P. and Wan, S.-M. (2002), "A Unified Jackknife Theory for Empirical Best Prediction with M -estimation," *Annals of Statistics*, **30**, 1782–1810.
- Lohr, S. and Rao, J. N. K. (2007). "Jackknife Estimation of Mean Squared Error of Small Area Predictors in Nonlinear Mixed Models," submitted for publication.
- Prasad, N. G. N. and Rao, J. N. K. (1990), "The Estimation of the Mean Squared Error of Small-area

Estimators," *Journal of the American Statistical Association*, **85**, 163–171.

Rao, J. N. K. (2003), *Small Area Estimation*, New York: Wiley.