SMALL AREA ESTIMATION FOR THE DISTRIBUTION OF PARAMETERS WITH COVARIATES

Michael P. Cohen, National Center for Education Statistics*
555 New Jersey Avenue NW, Washington DC 20208-5654

Key Words: Borrowing strength, Direct estimates, Distance, Ensemble

Abstract: If one wants to estimate a parameter for each of many small areas, one can generally improve the independent direct estimates by "borrowing strength" from the other small areas. Much research has been devoted to the situation in which one seeks to minimize the (possibly weighted) sums of the expected squared errors of the small area estimates. Thomas A. Louis, Malay Ghosh, and others have considered the contrasting situation in which the relationship among the small area parameters is of primary interest. For example, one might be interested in knowing the proportion of small areas where the high school dropout rate is above some level. The aim in such problems is to minimize the distance between the observed distribution of the "ensemble" (set) of small area estimates and the true distribution of the ensemble of parameters. This paper explores the situation further, expanding on Cohen (1998) by considering the effects of covariates.

1. Introduction

Suppose we are investigating the values of a certain parameter (e.g. average income or an average measure of the level of literacy) for each of many small areas. If the goal is the best estimates of these parameters considered individually, then empirical and hierarchical Bayes techniques have been developed that improve upon naïve estimators. What if, though, we want to know which small areas have parameter values above a fixed cutoff C and which below? A different approach is required to treat problems of this type.

Louis (1984) was the first to study these small area estimation problems although Rubin (1981) had looked at the situation in another context. Lahiri (1990) and Ghosh (1992, 1994) built on the work of Louis, extending it to non-normal and multivariate situations. Our aim is to build on the work of these authors and, in particular, to investigate the use of loss functions that measure the distance between the distribution of the estimates and the distribution of the parameters. This paper continues the work of Cohen (1998) by considering covariates.

For a general appraisal of small area estimation, Ghosh and Rao (1994) is highly recommended.

The recent and interesting work of Shen and Louis (1998) studies and compares the different approaches to small area estimation in a two-stage hierarchical setting.

The organization of this paper is as follows: This introduction is Section 1. Section 2 provides background information. Section 3 treats a normal model with covariates. Some concluding remarks are given in Section 4. An Appendix discusses the estimation of ranks.

2. Background

2.1 Previous Work

Consider the estimation of m parameters $\theta_1, \dots, \theta_m$ under squared error loss. Let $\hat{\theta}_1^B, \dots, \hat{\theta}_m^B$ denote Bayes estimates of these parameters based on data $\mathbf{Y} = (Y_1, \dots, Y_m)$. Let $\theta_{\bullet} = \frac{1}{m}\sum_{i=1}^m \theta_i$ and $\hat{\theta}_{\bullet}^B = \frac{1}{m}\sum_{i=1}^m \hat{\theta}_i^B$. Then

$$\mathrm{E}(\theta_{\bullet}|\mathbf{Y}) = \hat{\theta}_{\bullet}^B$$

but

$$\mathrm{E}\left[\sum_{i=1}^m (\theta_i - \theta_{\bullet})^2 \middle| \mathbf{Y}\right] > \sum_{i=1}^m \left(\hat{\theta}_i^B - \hat{\theta}_{\bullet}^B\right)^2.$$

This was shown by Louis (1984) under a normality assumption and, in general, by Ghosh (1992).

The point is that the Bayes estimates of the parameters (under squared error loss) have the same mean as the parameters themselves, but are on average less "spread out." If we are trying to use the collection of Bayes estimates to study the distribution of the parameters, we will have the distorted view that the parameters are more concentrated about their mean than they really are. We have been discussing Bayes estimates, but *empirical* Bayes estimates face the same problem.

In the context of small area estimation, the θ_i are parameters associated with small area i, say mean household income. If we use the $\hat{\theta}_i^B$ to study the θ_i, we will underestimate the diversity in the parameters.

Louis (1984) tackled this problem by investigating the class of estimators $\tilde{\theta}_i$ that satisfy

$$\mathrm{E}(\theta_{\bullet}|\mathbf{Y}) = \tilde{\theta}_{\bullet} \qquad \text{where } \tilde{\theta}_{\bullet} = \frac{1}{m}\sum_{i=1}^m \tilde{\theta}_i$$

and

$$\mathrm{E}\left[\sum_{i=1}^{m}(\theta_i-\theta_{\bullet})^2\middle|\,\mathbf{Y}\right]=\sum_{i=1}^{m}\left(\tilde{\theta}_i-\tilde{\theta}_{\bullet}\right)^2.$$

He still used squared error loss but it was minimized subject to these constraints. The constraints force a match on the first two moments between the distribution of the estimates and the distribution of the parameters.

In giving a theoretical basis to his work, Louis (1984, Subsection 2.2) introduced the notion of a general loss function operating on the empirical distributions of the parameter estimates and the parameters. Our investigation will be based on such loss functions; they are described in the next subsection.

2.2 Loss Functions

Given m parameters $\theta_1,\ldots,\theta_m$, define the function

$$G_m(t)=\frac{1}{m}\sum_{i=1}^{m}\mathrm{I}(\theta_i\le t) \tag{2.1}$$

where $\mathrm{I}(\cdot)$ is 1 when its argument is true and 0 otherwise. We can regard G_m as the empirical distribution function of the parameters. From a Bayesian point of view, the parameters are random variables. It should be noted, however, that the parameters will generally *not* be identically distributed and maybe not independent.

Let $\hat{G}_m$ be an estimator of G_m. For example, given m estimates $\hat{\theta}_1,\ldots,\hat{\theta}_m$ of $\theta_1,\ldots,\theta_m$ respectively, one could estimate G_m by

$$\hat{G}_m(t)=\frac{1}{m}\sum_{i=1}^{m}\mathrm{I}(\hat{\theta}_i\le t), \tag{2.2}$$

but we do not require estimators of G_m to be of the form (2.2). If we want to study the distribution of the θ_i, we would like to find an estimate $\hat{G}_m$ that is close, in some sense, to G_m. In other words, we would like $\|\hat{G}_m-G_m\|$ to be small where $\|\cdot\|$ is a distance function or metric. Examples of such distance functions include

$$\|\hat{G}_m-G_m\|_{W,a}=\int_{-\infty}^{\infty}|\hat{G}_m(t)-G_m(t)|^a\,\mathrm{d}W(t), \tag{2.3}$$

$$\|\hat{G}_m-G_m\|_{\mathbf{t},\mathbf{w},a}=\sum_{\ell=1}^{L}w_\ell|\hat{G}_m(t_\ell)-G_m(t_\ell)|^a, \tag{2.4}$$

and

$$\|\hat{G}_m-G_m\|_\infty=\max_{-\infty<t<\infty}|\hat{G}_m(t)-G_m(t)|. \tag{2.5}$$

In (2.3), $a>0$ and $W(t)$ is a weight function that we can choose to give more weight to ranges of parameter values in which we are especially interested. In (2.4), $a>0$ and the $\mathbf{w}=(w_1,\ldots,w_L)$ are weights attached to the points $\mathbf{t}=(t_1,\ldots,t_L)$. If we adopt a general definition of integral, the second distance function is just a special case of the first. An even more special case is

$$\|\hat{G}_m-G_m\|_{t_0,a}=|\hat{G}_m(t_0)-G_m(t_0)|^a$$

that considers only a single point in the space of parameter values. For example, if θ_i corresponds to average household income in small area i and $t_0=$ \$25,000, then $|\hat{G}_m(t_0)-G_m(t_0)|$ measures how close we are in estimating the proportion of small areas with average household incomes less than or equal to \$25,000.

The distance function (2.5) is of great interest but difficult to work with analytically. There are, of course, other distance functions one might want to consider. In this paper, though, we concentrate on (2.3) with $a=2$. The goal is to minimize the (conditional) expected distance given the data.

If we are presented with a distribution function estimate $\hat{G}_m$ of the form (2.2), we can recover the set of values of the $\hat{\theta}_i$ from the jumps in the function $\hat{G}_m$, but we cannot determine uniquely which small area i is associated with which jump. In fact, any one-to-one assignment of the small areas to the jumps gives rise to the same value of $\hat{G}_m$. Letting $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_m)$ and $\hat{\boldsymbol{\theta}}=(\hat{\theta}_1,\ldots,\hat{\theta}_m)$, Louis (1984, p. 394) suggests using a loss function of the form

$$\|\hat{G}_m-G_m\|+\epsilon\,\mathcal{L}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) \tag{2.6}$$

for some small $\epsilon>0$ where $\mathcal{L}(\cdot,\cdot)$ is, for example, the sum of squared errors $\mathcal{L}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=\sum_{i=1}^{m}(\hat{\theta}_i-\theta_i)^2$. The second term in the loss function is designed to force a unique assignment of the jumps in $\hat{G}_m$ of form (2.2) to the small areas i without otherwise affecting the loss function much. Shen and Louis (1998) develop an alternative approach in which the estimation of the ranks is a separate stage in the estimation process. See the Appendix for some discussion of estimating the ranks.

Given any estimator $\hat{G}_m$ of G_m, not necessarily of form (2.2), we can estimate the ensemble

$\{\theta_1, \theta_2, \dots, \theta_m\}$ by

$$\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m\} = \left\{\hat{G}_m^{-1}\left(\frac{1-\frac{1}{2}}{m}\right), \hat{G}_m^{-1}\left(\frac{2-\frac{1}{2}}{m}\right), \dots, \hat{G}_m^{-1}\left(\frac{m-1-\frac{1}{2}}{m}\right), \hat{G}_m^{-1}\left(\frac{m-\frac{1}{2}}{m}\right)\right\}.$$

We use (2.6) to determine which $\hat{\theta}_i$ corresponds to

$$\hat{G}_m^{-1}\left(\frac{j-\frac{1}{2}}{m}\right),$$

and so forth. A justification for the use of $\frac{j-\frac{1}{2}}{m}$ (rather than $\frac{j}{m+1}$ or some similar expression) is given by Theorem 1 of Shen and Louis (1998).

In the next subsection, we make use of some of the loss functions described in this subsection to investigate a normal model.

2.3 Normal Model with Known Variances

Suppose that each $\theta_i \sim \mathsf{N}(\mu, \tau^2)$, that is, suppose each θ_i is normally distributed with mean μ and variance τ^2. Suppose further that the θ_i are independent. Let Y_i given θ_i be $\mathsf{N}(\theta_i, \sigma_i^2)$ and let the Y_i be independent and $\sigma_i^2 > 0$, $i = 1, \dots, m$. We shall use this simple model as a starting point.

Let

$$\gamma_i = \frac{\tau^2}{\sigma_i^2 + \tau^2}.$$

For known μ, τ^2, and σ_i^2, the posterior distribution of θ_i given $\mathbf{Y}$ is normal with mean

$$\mathrm{E}(\theta_i|\mathbf{Y}) = \mu + \gamma_i(Y_i - \mu)$$

and variance

$$\mathrm{var}(\theta_i|\mathbf{Y}) = \gamma_i\sigma_i^2.$$

The $\theta_i|\mathbf{Y}$ are independent. We have

$$\begin{aligned} \mathrm{E}\{G_m(t)|\mathbf{Y}\} &= \frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\{\mathrm{I}(\theta_i \le t)|\mathbf{Y}\} \\ &= \frac{1}{m}\sum_{i=1}^{m}\Pr(\theta_i \le t|\mathbf{Y}) \\ &= \frac{1}{m}\sum_{i=1}^{m}\Phi\left(\frac{t-\mu-\gamma_i(Y_i-\mu)}{\sigma_i\sqrt{\gamma_i}}\right) \end{aligned} \tag{2.7}$$

where Φ is the standard normal distribution function.

Let us consider the distance function

$$\|\hat{G}_m - G_m\|_{W,2} = \int_{-\infty}^{\infty}(\hat{G}_m(t) - G_m(t))^2\,\mathrm{d}W(t)$$

where $W(t) \ge 0$ and $\int_{-\infty}^{\infty}\mathrm{d}W(t) < \infty$. The conditional expected distance given $\mathbf{Y}$ is

$$\begin{aligned} &\mathrm{E}(\|\hat{G}_m - G_m\|_{W,2} \mid \mathbf{Y}) \\ &= \mathrm{E}\left\{\int_{-\infty}^{\infty}(\hat{G}_m(t) - G_m(t))^2\,\mathrm{d}W(t)\middle|\,\mathbf{Y}\right\} \\ &= \int_{-\infty}^{\infty}\mathrm{E}\left\{(\hat{G}_m(t) - G_m(t))^2\middle|\,\mathbf{Y}\right\}\mathrm{d}W(t). \end{aligned}$$

The last step is justified because the integrand is nonnegative and bounded. But the last integral can be minimized by minimizing

$$\mathrm{E}\left\{(\hat{G}_m(t) - G_m(t))^2\,\middle|\,\mathbf{Y}\right\} \tag{2.8}$$

for each t. Note that the solution does not depend on $W(t)$. It is known from standard results in Bayes estimation that (2.8) is minimized by the choice $\hat{G}_m(t) = \mathrm{E}\{G_m(t)|\mathbf{Y}\}$. For the simple normal model, the latter quantity is given by (2.7).

Note: Carlin and Louis (1996, p. 238) and Shen and Louis (1998) obtain

$$\hat{G}_m(t) = \mathrm{E}\{G_m(t)|\mathbf{Y}\} = \frac{1}{m}\sum_{i=1}^{m}\Pr(\theta_i \le t|\mathbf{Y})$$

for a two-stage hierarchical model.

It is of interest to compute the (conditional) expected loss because this provides a measure of the closeness of estimation, analogous to mean squared error. If $\hat{G}_m(t) = \mathrm{E}\{G_m(t)|\mathbf{Y}\}$, then

$\mathrm{E}\left\{(\hat{G}_m(t) - G_m(t))^2\,\middle|\,\mathbf{Y}\right\} = \mathrm{var}\{G_m(t)|\mathbf{Y}\}$, so

$$\mathrm{E}(\|\hat{G}_m - G_m\|_{W,2} \mid \mathbf{Y}) = \int_{-\infty}^{\infty}\mathrm{var}\{G_m(t)|\mathbf{Y}\}\,\mathrm{d}W(t). \tag{2.9}$$

But

$$\begin{aligned} &\mathrm{var}\{G_m(t)|\mathbf{Y}\} \\ &= \frac{1}{m^2}\sum_{i=1}^{m}\mathrm{var}\{\mathrm{I}(\theta_i \le t)|\mathbf{Y}\} \\ &= \frac{1}{m^2}\sum_{i=1}^{m}\Pr(\theta_i \le t|\mathbf{Y})\{1 - \Pr(\theta_i \le t|\mathbf{Y})\}; \end{aligned}$$

thus

$$\begin{aligned}\operatorname{var}\{G_m(t)|\mathbf{Y}\} \\ = \frac{1}{m^2}\sum_{i=1}^{m}\left[\Phi\left(\frac{t-\mu-\gamma_i(Y_i-\mu)}{\sigma_i\sqrt{\gamma_i}}\right)\right. \\ \left.\times\left\{1-\Phi\left(\frac{t-\mu-\gamma_i(Y_i-\mu)}{\sigma_i\sqrt{\gamma_i}}\right)\right\}\right].\end{aligned} \tag{2.10}$$

From (2.9) and (2.10), $\mathrm{E}(\|\hat{G}_m - G_m\|_{W,2} \mid \mathbf{Y})$ can be computed.

3. Normal Model with Covariates

In applications, it is usually the case that we have background data on each small area i. We shall represent this background data as $\mathbf{x}_i'$, a $1\times r$ row vector. The model of Fay and Herriot (1979) specifies that θ_i has mean $\mathbf{x}_i'\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is an $r\times 1$ column vector of regression parameters. Because $\boldsymbol{\beta}$ is the same for each small area i, we are able to "borrow strength" across small areas in estimating it.

Suppose that each $\theta_i \sim \mathsf{N}(\mathbf{x}_i'\boldsymbol{\beta}, \tau^2)$, that is, suppose each θ_i is normally distributed with mean $\mathbf{x}_i'\boldsymbol{\beta}$ and variance τ^2. Suppose further that the θ_i are independent. Let Y_i given θ_i be $\mathsf{N}(\theta_i, \sigma_i^2)$ and let the Y_i be independent and $\sigma_i^2 > 0$, $i = 1, \dots, m$.

3.1 Known $\boldsymbol{\beta}$

For *known* $\mathbf{x}_i'$, $\boldsymbol{\beta}$, τ^2, and σ_i^2, the posterior distribution of θ_i given $\mathbf{Y}$ is normal with mean $\mathrm{E}(\theta_i|\mathbf{Y}) = \mathbf{x}_i'\boldsymbol{\beta} + \gamma_i(Y_i - \mathbf{x}_i'\boldsymbol{\beta})$ and variance $\operatorname{var}(\theta_i|\mathbf{Y}) = \gamma_i\sigma_i^2$, where $\gamma_i = \tau^2/(\sigma_i^2+\tau^2)$. The assumption that $\boldsymbol{\beta}$ is known is, of course, unrealistic, but we shall make it anyway for the time being. Then the $\theta_i|\mathbf{Y}$ are independent. As in (2.7), we have

$$\begin{aligned}\mathrm{E}\{G_m(t)|\mathbf{Y}\} \\ = \frac{1}{m}\sum_{i=1}^{m}\Phi\left(\frac{t-\mathbf{x}_i'\boldsymbol{\beta}-\gamma_i(Y_i-\mathbf{x}_i'\boldsymbol{\beta})}{\sigma_i\sqrt{\gamma_i}}\right).\end{aligned} \tag{3.1}$$

Moreover, as in (2.9) and (2.10),

$$\mathrm{E}(\|\hat{G}_m - G_m\|_{W,2} \mid \mathbf{Y}) = \int_{-\infty}^{\infty}\operatorname{var}\{G_m(t)|\mathbf{Y}\}\,\mathrm{d}W(t) \tag{3.2}$$

where

$$\begin{aligned}\operatorname{var}\{G_m(t)|\mathbf{Y}\} \\ = \frac{1}{m^2}\sum_{i=1}^{m}\left[\Phi\left(\frac{t-\mathbf{x}_i'\boldsymbol{\beta}-\gamma_i(Y_i-\mathbf{x}_i'\boldsymbol{\beta})}{\sigma_i\sqrt{\gamma_i}}\right)\right. \\ \left.\times\left\{1-\Phi\left(\frac{t-\mathbf{x}_i'\boldsymbol{\beta}-\gamma_i(Y_i-\mathbf{x}_i'\boldsymbol{\beta})}{\sigma_i\sqrt{\gamma_i}}\right)\right\}\right].\end{aligned} \tag{3.3}$$

3.2 Estimated $\boldsymbol{\beta}$

Now let us suppose the regression parameters $\boldsymbol{\beta}$ are unknown and must be estimated from the data. We still suppose that $\mathbf{x}_i'$, τ^2, and σ_i^2 are known. Then, treating $\boldsymbol{\beta}$ as random, the $\theta_i|\mathbf{Y},\boldsymbol{\beta}$ are independent normal with conditional mean

$$\mathrm{E}(\theta_i|\mathbf{Y},\boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta} + \gamma_i(Y_i - \mathbf{x}_i'\boldsymbol{\beta})$$

where $\gamma_i = \tau^2/(\sigma_i^2+\tau^2)$. Moreover,

$$\begin{aligned}\mathrm{E}\{G_m(t)|\mathbf{Y},\boldsymbol{\beta}\} \\ = \frac{1}{m}\sum_{i=1}^{m}\Phi\left(\frac{t-\mathbf{x}_i'\boldsymbol{\beta}-\gamma_i(Y_i-\mathbf{x}_i'\boldsymbol{\beta})}{\sigma_i\sqrt{\gamma_i}}\right).\end{aligned}$$

Now $\hat{G}_m(t)$ can be computed by

$$\begin{aligned}\hat{G}_m(t) &= \mathrm{E}\{G_m(t)|\mathbf{Y}\} \\ &= \mathrm{E}_{\boldsymbol{\beta}}\,\mathrm{E}\{G_m(t)|\mathbf{Y},\boldsymbol{\beta}\}.\end{aligned}$$

To implement a Bayesian or empirical Bayesian approach, one needs to put a prior distribution on $\boldsymbol{\beta}$. One might use, for example, $\boldsymbol{\beta} \sim \mathsf{N}(\mathbf{0}, \lambda\mathbf{I})$ for large λ, a vague prior. Computation of $\hat{G}_m$ is feasible with Markov Chain Monte Carlo techniques (e.g., Gibbs sampling). One would also want to compute

$$\mathrm{E}\left\{\|\hat{G}_m - G_m\|_{W,2}\,\middle|\,\mathbf{Y}\right\}$$

to assess error (analogous to mean squared error).

4. Concluding Remarks

This paper has built upon the work of Louis (1984), Ghosh (1992), and others that study ways of estimating the distribution of small area parameters. Our focus has been on using loss functions that measure the distance between the distribution of the estimates of the parameters and the distribution of the parameters themselves. This paper extends Cohen (1998) by considering covariates. There are many aspects of this problem that have yet to be explored.

Acknowledgment: The author thanks Professor Malay Ghosh for acquainting him with this area of research in his 1994 Conference Board for the Mathematical Sciences lectures at the University of Connecticut, Storrs.

Appendix: Estimating the Ranks

The development in this Appendix follows that of Subsection 2.1 of Shen and Louis (1998). The ranks $\mathbf{R} = R_1, \ldots, R_m$ are

$$R_i = \text{rank}(\theta_i) = \sum_{j=1}^{m} \mathrm{I}(\theta_i \geq \theta_j).$$

For an estimator $\tilde{\mathbf{R}}$ of $\mathbf{R}$, let us consider the loss function $\mathsf{L}(\tilde{\mathbf{R}}, \mathbf{R}) = \sum_{i=1}^{m} (\tilde{R}_i - R_i)^2$. The $\tilde{\mathbf{R}}$ that minimizes $\mathrm{E}[\mathsf{L}(\cdot, \mathbf{R})|\mathbf{Y}]$ has components

$$\tilde{R}_i = \mathrm{E}(R_i|\mathbf{Y}) = \sum_{j=1}^{m} \Pr(\theta_i \geq \theta_j|\mathbf{Y}).$$

The $\tilde{R}_i$ will generally not be integers. As noted by Shen and Louis, they will exhibit shrinkage toward $(m+1)/2$. Because of the convenience of integer ranks, Shen and Louis use the estimator

$$\hat{R}_i = \text{rank}(\tilde{R}_i),$$

the rank of the estimated rank. We supplement the results of Shen and Louis with a theorem.

Theorem. The estimator $\hat{\mathbf{R}}$ minimizes $\mathrm{E}[\mathsf{L}(\cdot, \mathbf{R})|\mathbf{Y}]$ among all estimators in the class of estimators whose components are distinct integers from the set $\{1, \ldots, m\}$.

Proof. Let $\bar{\mathbf{R}}$ be a competing estimator in the class and suppose that for some k and ℓ, $\bar{R}_k < \bar{R}_\ell$ but $\hat{R}_k > \hat{R}_\ell$. Then $\tilde{R}_k > \tilde{R}_\ell$. Moreover, for $i = 1, \ldots, m$,

$$\mathrm{E}[(\bar{R}_i - R_i)^2|\mathbf{Y}] = \mathrm{E}[(\bar{R}_i - \tilde{R}_i)^2 + (\tilde{R}_i - R_i)^2|\mathbf{Y}]$$

because $\tilde{R}_i = \mathrm{E}(R_i|\mathbf{Y})$ so it suffices to minimize $\sum_{i=1}^{m} (\bar{R}_i - \tilde{R}_i)^2$. Noting that

$$\begin{aligned}
&(\bar{R}_k - \tilde{R}_k)^2 + (\bar{R}_\ell - \tilde{R}_\ell)^2 \\
&\quad = (\bar{R}_\ell - \tilde{R}_k)^2 + (\bar{R}_k - \tilde{R}_\ell)^2 \\
&\qquad + 2(\bar{R}_\ell - \bar{R}_k)(\tilde{R}_k - \tilde{R}_\ell)
\end{aligned}$$

where the last term is positive, we see that the estimator $\bar{\mathbf{R}}'$ that switches the k^{th} and ℓ^{th} components of $\bar{\mathbf{R}}$ is better than $\bar{\mathbf{R}}$ (that is, has smaller conditional expected loss). The estimator $\hat{\mathbf{R}}$ is the only estimator in the class that cannot be improved in this way, so it minimizes the expected loss.

Note: It has been implicitly assumed that the probability of ties among the ranks is 0. The result could be modified to account for ties, but we shall not do so here.

REFERENCES

Carlin, B.P, and Louis, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis.* Boca Raton: Chapman & Hall/CRC.

Cohen, M.P. (1998). Small area estimation for the distribution of parameters, *1998 Proceedings of the Section on Survey Research Methods*, American Statistical Association 345–348.

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.

Ghosh, M. (1992). Constrained Bayes estimation with applications, *Journal of the American Statistical Association* **87** 533–540.

Ghosh, M. (1994). Bayesian methods in survey sampling, Conference Board for the Mathematical Sciences Seminar, University of Connecticut, unpublished notes.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: an appraisal (with discussion), *Statistical Science* **9** 55–93.

Lahiri, P. (1990). "Adjusted" Bayes and empirical Bayes estimation in finite population sampling, *Sankhya,* Series B **52** 50–66.

Louis, T.A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods, *Journal of the American Statistical Association* **79** 393–398.

Rubin, D.B. (1981). Estimation in parallel randomized experiments, *Journal of Educational Statistics* **6** 377–401.

Shen, W., and Louis, T.A. (1998). Triple-goal estimates in two-stage hierarchical models, *Journal of the Royal Statistical Society,* Series B **60** 455–471.

---

*This paper is intended to promote the exchange of ideas among researchers and policy makers. The views are those of the author, and no official support by the U.S. Department of Education is intended or should be inferred.