The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 407-416

Bayes Prediction for Small Area Estimation[1)]

Sang Eun Lee[2)]

Abstract

Sample surveys are usually designed and analyzed to produce estimates for a large area or populations. Therefore, for the small area estimations, sample sizes are often not large enough to give adequate precision. Several small area estimation methods were proposed in recent years concerning with sample sizes. Here, we will compare simple Bayesian approach with Bayesian predictions for small area estimation based on linear regression model. The performance of the proposed method was evaluated through unemployment population data form Economic Active Population(EAP) Survey.

Keywords : Simple Bayes estimation, small area estimation, Bayes predictive distribution,

1. Introduction

Small area estimation has recently received much attention in the literature due to growing demand for reliable small area estimation for prediction. Ghosh and Rao(1994) gave the appraisal of several small area estimation methodologies, such as ratio-synthetic, Bayes approaches and so on. And they also discussed several examples of small area applications Datta and Ghosh(1991) proposed the Bayesian prediction for small area estimation with linear models. And from empirical study of Lee and Park(1999), Bayesian approach was more stable than the others. In general, Bayes estimator is expressed by weighted average of two variables and the weight is usually involved in the variance of the variables. As we know, for direct estimation for the small area, that gives the huge variance of estimates because of small sample size. Therefore, if we can have the estimates which less involved in the variance of small area that may give the better result for small area estimation.

Let us have Bayesian perspective of parameter on unsampled area characteristic of interest and the small area estimation will be the weighted average of Bayes predictions for unsampled area characteristic and sampled information, in stead of depending on the variance,

1) This research was supported by the Kyonggi University Fund in 1999
2) Assistant Professor, Department of Applied and Information Statistics, Kyonggi University
E-mail : sanglee@stat.kyonggi.ac.kr

using the rate of sample selected as weight of two informations.

Furthermore, most of cases, they consider the situation where values of auxiliary variables are only available at the small area level not at a unit_level. However, in this paper, we are proposing the Bayes estimation method for a small area or finite population that takes into account observations of covariates at a unit_level and lends to finding the predictive densities for the future values of the responds variable with linear regression model and compare the simple Bayesian approach with Bayesian Prediction.

Section 2 gives the Bayes approach and Bayesian Prediction model and section 3 will show you the data explanation and finally in section 4, we will have the result and summary.

2. Bayesian Model

2.1 Simple Bayes approach

Simple Bayes approach of small area estimator was proposed when the expected values of observations are given as linear functions of some unknown parameters.

Let y_{jk} be the value of the variable of interest y measured on the k^{th} sampled unit within the j^{th} area ($j=1,2,\ldots,m$). Assumed model is as following;

$$y_{jk}|(\mu_1,\mu_2,\ldots,\mu_m) \sim N(\mu_j,\sigma^2) \quad j=1,2,\ldots,m,\ k=1,\ldots,n_j$$
$$\mu_j \sim N(\alpha+\beta x_j,\tau^2)$$

and x_j is the auxiliary variable of area j.

σ^2,τ^2 assume to be known and those are independent with $y_{jk}|(\mu_1,\mu_2,\ldots,\mu_m)$ and μ_j respectively ,

n_j is the sampled units in area j.

Therefore, likelihood function is

$$l(y_{jk}|\mu_j,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{-1}{2\sigma^2}\sum_k^{n_j}(y_{jk}-\mu_j)^2\right]$$

and prior function is

$$g(\mu_j) = \frac{1}{\sqrt{2\pi\tau^2}} \exp[\frac{-1}{2\tau^2}(\mu_j - (\alpha + \beta x_j))^2].$$

With the square error loss, Bayes estimator for sample mean is as following(Lee & park, 1999);

$$\hat{\mu}_{B(j)} = \frac{\tau^2}{\sigma^2/n + \tau^2}\overline{y_j} + \frac{\sigma^2/n}{\sigma^2/n + \tau^2}(\hat{\alpha} + \hat{\beta}x_j)$$

Here, $\overline{y_j}$ is the j^{th} small area of sample mean of y ,

$\hat{\alpha}$ and $\hat{\beta}$ are hyperparameter to be assessed and in this stusdy we will uas as

$\hat{\beta} = \sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}) / \sum_{i=1}^{N}(x_i - \overline{x})^2$, $\quad \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$,

N is the total sample size and

τ^2 and σ^2 are hyperparameters to be assessed and in this study we will as sample varianace of j^{th} area.

Then, we call it as simple Bayesian estimator(SBE).

2.2 Bayes prediction

Consider a large area which divided into m small areas. Within the j^{th} small area, there are N_j units, $j = 1, \ldots, m$. And sampled units in j^{th} area are n_j units. therefore this leaves behind $N_j - n_j$ unsampled units. Our primary goal is to predict these unsampled units and then estimate a mean or total for the unemployed population for a certain(small) area.

We adopt a model corresponding to an individual unit with, y_{jk} and X_{jk} is the auxiliary variable and that could be measured on k^{th} unit in the j^{th} small area.

$$y_{jk} = X_{jk}\beta + \varepsilon_{jk} \quad k = 1, \ldots, N_j, \ j = 1, \ldots, m, \ N = \sum_{j=1}^{m} N_j$$

where $\varepsilon_{jk} \overset{iid}{\sim} N(0, \sigma^2)$

or in matrix form,

$$Y_j = X_j\beta + \varepsilon_j, \quad \varepsilon_j \overset{iid}{\sim} N(0, \sigma^2 I_{N_j})$$

2.2.1 Likelihood Functions

Suppose a sample size of n_j units, $j=1,\ldots,m$ was drawn ($n=\sum_{j=1}^{m} n_j$). Let Y_j be partitioned into two parts, $Y_j^{(1)}$, $Y_j^{(2)}$, corresponding sampled and unsampled units respectively. so

$$Y_j=\begin{pmatrix} Y_j^{(1)} \\ Y_j^{(2)} \end{pmatrix}$$

where, $Y_j^{(1)}=(Y_{1j}^{(1)},\ldots,\ Y_{n_j j}^{(1)})'$: observations on sampled units and
$Y_j^{(2)}=(Y_{n_j-1j}^{(2)},\ldots,Y_{N_j j}^{(2)})'$: unsampled units.

From now on, for a convenience, we denote the sample data by $y_j=Y_j^{(1)}$, which is an $(n_j\times 1)$ column vector of sample data.

Then the likelihood function is given by

$$l(y_j\ |X_j^{(1)},\sigma)\ \propto\ \sigma^{-n_j/2}\exp[-1/2\sigma^2(y_j-x_j^{(1)}\beta)'(y_j-x_j^{(1)}\beta)] \tag{1}$$

where $X_j^{(1)}$ is $(n_j\times 1)$ vector of auxiliary variable in sampled units.

2.2.2 Prior Distributions

Let us assume the joint prior distribution of the parameters of the model is given by

$$p(\beta,\sigma)\ \propto\ p(\beta)p(\sigma) \tag{2}$$

where,

$$p(\beta)\ \propto\ \ constant$$
$$p(\sigma)\ \propto\ \frac{1}{\sigma}$$

Let $(Y_j^{(2)},X_j^{(2)})$ correspond to the unsampled units in response and auxiliary variable respectively and expressed as

$$Y_j^{(2)}=X_j^{(2)}\beta+\varepsilon_j^{(2)}\qquad \varepsilon_j^{(2)}\overset{iid}{\sim}N(0,\sigma^2 I_{N_j-n_j})$$

The induced prior distribution for $Y_j^{(2)}$ is then given by

$$f(Y_j^{(2)}|X_j^{(2)},\beta,\sigma) \propto \sigma^{-(N_j-n_j)/2}\exp[-1/2\sigma^2(Y_j^{(2)}-X_j^{(2)}\beta)'(Y_j^{(2)}-X_j^{(2)}\beta)] \quad (3)$$

2.2.3 Predictive Distributions

The joint posterior distribution of the parameter of the model condition on the data is obtained by combining (1),(2) and (3) via Bayes Theorem as :

$$f(Y_j^{(2)},\beta,\sigma|y_j\ X_j) \propto f(Y_j^{(2)}|\beta,\sigma,X_j^{(2)})\cdot h(\beta,\sigma|y_j,X_j^{(1)},)$$

where,

$$h(\beta,\sigma|y_j,X_j^{(1)}) \propto l(y_j|X_j^{(1)},\beta,\sigma)\cdot p(\beta,\sigma)$$

Therefore, joint posterior of $(Y_j^{(2)},\beta,\sigma|y_j,X_j)$ is as following ;

$$\begin{aligned} &f(Y_j^{(2)},\beta,\sigma|y_j X_j) \propto \\ &\quad \sigma^{-(N_j+2)/2}\exp[-1/2\sigma^2[(y_j-X_j^{(1)}\beta)'(y_j-X_j^{(1)}\beta)+(Y_j^{(2)}-X_j^{(2)}\beta)'(Y_j^{(2)}-X_j^{(2)}\beta)]] \end{aligned} \quad (4)$$

Here, our goal is to find the predictive density for unsampled units, $(Y_j^{(2)}|y_j, X_j)$. To find it, we eliminate the parameter β and σ by integrating out from the equation(4). From Press(1982, p257) we have following result.

For the connivence of calculation, we let $h=1/\sigma^2$ then we can rewrite equation(4) as following.

$$\begin{aligned} &f(Y_j^{(2)}\ \beta,h|y_j X_j) \propto \\ &\quad h^{(N_j+2)}\exp[-\frac{h}{2}[(y_j-X_j^{(1)}\beta)'(y_j-X_j^{(1)}\beta)+(Y_j^{(2)}-X_j^{(2)}\beta)'(Y_j^{(2)}-X_j^{(2)}\beta)]] \end{aligned} \quad (5)$$

Now $(Y_j^{(2)}|y, X_j)$ can be written arranging with β and h (Press 1982, p258)

$$f(Y_j^{(2)}|y_j\, X_j) \propto \int\int h^{(N_j+1)/2-1}\exp[-\frac{h}{2}[(\beta-\hat{\beta})'(X_j'X_j)(\beta-\hat{\beta})+ (N-1)\,\hat{\sigma}^2+(Y_j^{(2)}-X_j^{(2)}\beta)^2]]\partial\beta\partial h$$

Integrating with respect to h gives

$$f(Y_j^{(2)}|y_j\, X_j) \propto \int\frac{\partial\beta}{[(N_j-1)\,\hat{\sigma}^2+(Y_j^{(2)}-X_j^{(2)}\beta)^2+(\beta-\hat{\beta})'X_j'X_j(\beta-\hat{\beta})]^{(N_j+1)/2}}$$

Completing the square on β gives

$$f(Y_j^{(2)}|y_j\, X_j) \propto \int\frac{\partial\beta}{[w+(\beta-Q)(X_j'X_j+X_j^{(2)'}X_j^{(2)})(\beta-Q)]^{(N_j+1)/2}} \tag{6}$$

where,

$$Q=(X_j'X_j+X_j^{(2)'}X_j^{(2)})^{-1}(\hat{\beta}'X_j'X_j+Y_j^{(2)'}X_j^{(2)})$$

and

$$w=Y_j^{(2)'}Y_j^{(2)}+\hat{\beta}'X_j'X_j\hat{\beta}+(N-1)\,\hat{\sigma}^2 -(\hat{\beta}'X_j'X_j+Y_j^{(2)'}X_j^{(2)})(X_j'X_j+X_j^{(2)'}X_j^{(2)})^{-1}(\hat{\beta}'X_j'X_j+Y_j^{(2)'}X_j^{(2)})$$

and $\hat{\beta}=(X_j'X_j)^{-1}X_j'Y_j$ and $\hat{\sigma}^2=\frac{1}{n_j-1}(y_j-x_j\hat{\beta})'(y_j-x_j\hat{\beta})$.

Now integrating equation(6) gives

$$f(Y_j^{(2)}|y_j\, X_j) \propto \frac{1}{w^{N_j/2}}$$

and from Zellener(1971) page 73, $f(Y_j^{(2)}|y_j\, X_j)$ is distributed in the t-distribution with location $X_j^{(2)}\hat{\beta}$, that is $E(Y_j^{(2)}|y_j\, X_j)=X_j^{(2)}\hat{\beta}$

By taking T sets of data from population, we have the predictive density for the unsampled units, $Y_j^{(2)}$, is obtained as the corresponding mean of the predictive distribution becomes

$$E(Y_j^{(2)}|y,X_j) = \frac{1}{T}\sum_1^T E(Y_j^{(2)}|y,X_j)$$
$$= Y_j^*$$

Thus, an Bayes prediction estimate for the j^{th} small area is obtained from

$$\hat{\mu}_j = f_j\,\overline{y_j} + (1-f_j)Y_j^*$$

where,

$f_j = n_j/N_j$ and y_j is sample mean which is obtained from observed data.

and we call it as Bayes prediction estimator(BPE).

At this point we can easily calculate the expectation value of predictive distribution. This will be discussed in section 3.

3. Data Analysis

3.1 Data Description

From Economic Active Population Survey(EAPS), we have the total of unemployment population (Y) as a response variable and total of economic active population (X) as an auxiliary variable.

we assume simple linear regression model and there are 8 small areas and is an one auxiliary variable which is active economic population . So we have

$$y_{kj} = X_{kj}\beta + \varepsilon_{kj}\;,\quad k=1,\ldots,N_j,\quad j=1,\ldots,m,\; N=\sum_{j=1}^{m} N_j$$
$$\varepsilon_{kj} \overset{iid}{\sim} N(0,\sigma^2)$$

where y_{kj} is a response variable which is unemployment population on the k^{th} unit in the j^{th} area and x_{kj} is an auxiliary variable which is economic active population for the k^{th} unit in the j^{th} area .

Our objective is to find the predictors for the unsampled unit for each small area, and thereby the true values are given by Table 3.1.

Table 3.1 : True values for the small areas

area	Economic active population(X)	Unemployment population(Y)
1	307	29
2	313	23
3	305	32
4	281	24
5	304	18
6	167	17
7	272	13
8	59	3

Now we listed the data by "Dong" which is the smallest administrative district and selected the sample systematically with P which is rate of sampling. And using the selected data, we calculated SBE and PBE as we mentioned in section 2.1 and 2.2.3.

In section 2.1, for SBE, we let the $\hat{\sigma}^2, \hat{\tau}^2$ are hyperparameters. Therefore, we use the sample variances to evaluate $\frac{\hat{\tau}^2}{\hat{\sigma}^2/n+\hat{\tau}^2}$. Finally, we can obtain the $\hat{\mu}_{B(j)}$ and it was done by 50 times and the average of them are the in table 4.1. And for PBE, in order to evaluate the $X_j^{(2)}\hat{\beta}$ which is Y_j^* , we simulate the $X_j^{(2)}$ from the selected data. And calculate the $E(Y_j^{(2)}|y, X_j) = \frac{1}{T}\sum_1^T E(Y_j^{(2)}|y, X_j) = Y_j^*$ and it was done 50 times and average of them are in table 4.1.

4. Summary and Conclusion

In early study, Bayesian approach method is superior to the classic approaches which are direct , synthetic and composite method in terms of variance of estimations. Here, in this study, Bayesian Prediction was applied to small area estimation especially concerned with unemployment population in EAP survey and compared with Bayesian approach with rate of sampling, P, which can tell the results of depending on sample size.

For small area estimation, most of time, the problems are caused by not large enough sample size for proper precision. Therefore, this study is looking for sensitiveness of estimation respect to sample size.

Table 4.1 SBE and PBE with different P

area	P=0.05		P=0.1		P=0.5	
	SBE	BPE	SBE	BPE	SBE	BPE
1	44.76	27.70	26.46	26.60	30.39	50.72
2	30.00	30.50	25.51	29.88	25.70	52.93
3	27.00	26.67	39.40	31.70	32.98	48.33
4	17.15	26.14	25.48	24.48	24.80	48.05
5	22.07	27.15	24.01	27.09	18.51	49.36
6	17.87	16.30	14.96	16.25	13.59	23.73
7	11.18	23.07	9.68	26.92	11.92	43.09
8	3.17	4.18	2.13	5.83	2.72	11.12

* P is the rate of sample selected from population

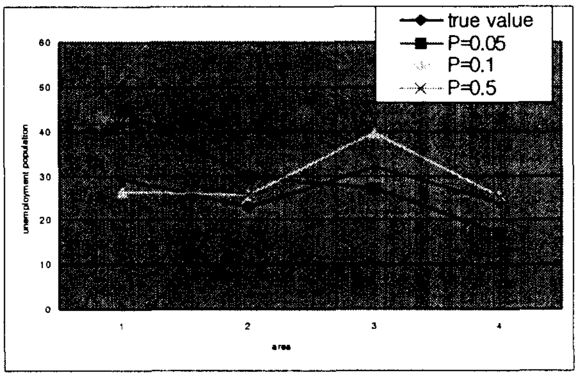

Figure 4.1 SBE with P

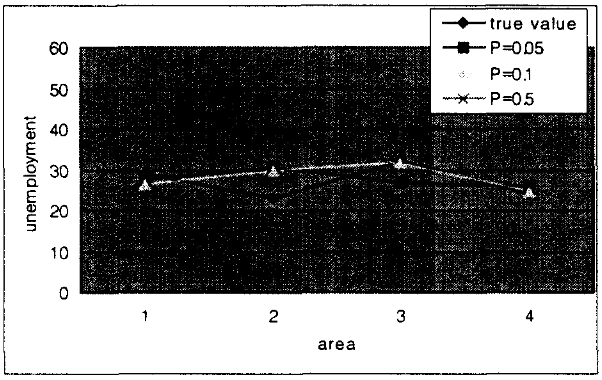

Figure 4.2 BPE with P

The table 4.1 and 4.2 give the results of both SBE and BPE small area estimations with P. Also in figure 4.1 and 4,2 , the results only area 1,2,3 and 4 was shown. Now from table 4.1 and 4.2, we can easily see when the rate of sampling is relatively big such as P=0.5 then SBE's are more likely to be better in terms of closeness based on the true value in table 3.1. That is, in SBE, the weight of likelihood and prior information depends on variance of variables, therefore, small rate of sampling such as P=0.05 will not give the relatively good estimations. On the other hand, case of BPE, it shows the better estimations when P=0.05 also in terms of true value. That is, in BPE, weight of sampled and unsampled depends on ratio of sampled selected, $f = n/N$, not the variances of variables. Therefore, it obviously SBEs are more sensitive then BPE respect to sample size. However, when selected sample rate becomes relatively larger(P=0.5), obviously PBEs give the over_estimated result .

Consequently, BPE will be use of the small area estimation. That is, especially the small area estimation problem caused by unplanned sample design which usually gives an improper(small) sample size.

Reference

[1] Datta, G.S. and Ghosh, M(1991)." Bayesian Prediction in Linear Models :Application to Small Area Estimation," *The Annals of Statistics, 19, 1748-1770.*

[2] Ghosh, M and Rao, J.N.K(1994). "Small Area Estimation : An appraisal." *Statistical Science, 9, 55-93*

[3] Lee S. E and Park J.T(1999)." An Empirical Study of small ares estimation" *51th ISI preceedings, 213-214*

[4] Lee, P. M(1988) *Bayesian Statistics: An Introduction.* University of York. English

[5] Parimal, M.(1998). *Small Area Estimation in Survey Sampling.* Narosa Publishing house

[6] Press, S. J.(1982) *Applied Multivariate Analysis:Using Bayesian and Frequentist Methods of Inference.* R.E. Krieger Publishing Co.

[7] Zellner, A.(1971) *An introduction to Bayesian Inference in Economics* John wiley & Sons, Inc