한국통계학회논문집 제3권 2호
1996년 8월 pp. 161-168

Finite Population Total Estimation On Multistage Cluster Sampling[1]

Geun-Shik Han[2], Yong-Chul Kim[3], Kiheon Choi[4]

Abstract

Multistage hierarchical models and Bayesian inferences about finite population total estimations are considered. Here, Gibbs sampling approach that can be used to predict the marginal posterior means needed for Bayesian inferences is proposed.

1. Introduction

Multistage sampling is frequently used in surveys of human population. Specially a large scale household survey uses probability samples of two or three stage cluster samples. The problems of estimating linear functions of the elements of a finite population from a multi-stage sample is considered by many statisticians after Ericson (1969) who proposed a Bayesian approach. Scott and Smith (1969) proposed a super population model appropriate for two stage sampling from a finite population and carried out Bayesian predictive inferences for a linear function of finite population elements. Royall (1976) applied linear least-square prediction approach to some problems in two stage sampling from a finite population. And Royall's estimator for the finite population total was same as the Scott and Smith, who derived it using Bayesian techniques. Pfeffermann and Nathan (1981) estimated finite population total using regression approach. Kempthorn and Koch (1983) describe a general approach for the analysis of attribute data from a two stage nested design. Malec and Sedransk (1985) extended Scott and Smith (1969)'s approach to multistage hierarchical model and described Bayesian predictive inference about a population total.

To predict a finite population total, technical difficulties are encountered in the calculation of marginal posterior means needed for Bayesian inference. For hierarchical Bayesian models, Deeply and Lindley (1981) discussed approximate Bayesian inference. Resently, Gibbs sampler approach that can be used to estimate the marginal posterior means needed for Bayesian inference is proposed.

---

1. This research was supported by Duksung Women's University Research Fund, 1996.

2. Assistant Professor, Department of Computer Science and Statistics, Hansin University, Osan, 447-270, Korea.

3. Full-time Lecturer, Department of Computer Science and Statistics, Yongin University, Yongin, 449-714, Korea.

4. Assistant Professor, Department of Statistics, Duksung Women's University, Seoul, 132-714, Korea.

Geman and Geman(1984) proposed the Gibbs sampler algorithm. A great deal of work has been done in the Gibbs algorithm to get the marginal posterior densities. Gelfand and Smith(1990), Zeger and Karim(1991) have used the Gibbs sampling method for calculating marginal and predictive densities for models with normally distributed data and for generalized linear models with random effects. In this paper, Gibbs sampling method can be used to estimate the population total in multistage Bayesian hierarchical models.

2. Model

Suppose that a population consists of two stage clustering and that the last stage cluster includes a number of final units. This two stage clustering give a three stage population. Let the finite population U includes L first stage clusters, $U_1, U_2, \cdots U_i, \cdots, U_L$, with the i-th cluster decomposed into N_i second stage clusters, $U_i = (U_{i1}, \cdots, U_{ij}, \cdots, U_{iN_i})$, and finally the (ij)-th second stage cluster U_{ij} is decomposed into $U_{ij} = (U_{ij1}, U_{ij2}, \cdots, U_{ijl}, \cdots, U_{ijM_{ij}})$. The variable of interest for the (ijk)-th element is denoted $Y_{ijk}(i=1,2,\cdots,L;\ j=1,2,\cdots,N_i;\ k=1,2,3,\cdots,M_{ij})$. y_{ijk} denote the realization of Y_{ijk} when the (ijk)-th element is in the sample.

At the first stage a sample of l clusters is selected from L primary sampling unit. At the second stage a sample of n_i clusters is selected from the N_i clusters in the i-th sampled primary sampling unit. Finally, m_{ij} third stage units are sampled from the M_{ij} units in the N_i clusters of i-th primary unit.

Estimation about the finite population total are to be considered where

$$T = a'Y = \sum_{i=1}^{L}\sum_{j=1}^{N_i}\sum_{k=1}^{M_{ij}} a_{ijk} Y_{ijk}$$

and the a_{ijk} are specified constants.

To predict the finite population total, the basic assumption about the distribution of Y_{ijk} $(i=1,2,\cdots,L, j=1,2,\cdots,N_i,\ k=1,2,\cdots,M_{ij})$ and a prior distribution for the parameters the sampling distribution of Y_{ijk} are based on Malec and Sedransk's model.

Here it is postulated that independently for all (ijk)-th variable

$$Y_{ijk} \sim N(\mu_{ij},\ \sigma_{ij}^2),$$

the prior distribution for $\{\mu_{ij}: i=1,2,\cdots,L\ ;\ j=1,2,\cdots,N_i\}$ is based on common observation that the third stage units within the same second stage unit share an effect. Thus independently for all (ij),

$$\mu_{ij} \sim N(\nu_i,\ \sigma_i^2)$$

whereas, independently for all i,

$$v_i \sim N(\theta, \gamma^2).$$

3. Prediction

In the usual sample survey the main problem is to estimate finite population total. the finite population total with a given sample can be express as sum of sampled observation total and unobserved elements total and the frequentist approach described in Cassel, Sarndal and Wretman (1977) and Royall (1976). In contrast to frequentist methods; however the Bayesian approach provides a posterior predictive distribution for non-sampled elements total. Here, we are concern with the predictive total of unobserved values taken from population based on observed sample from population modelled according to the hierarchical Bayesian structure described above.

$$T|\, y_s,\ \mu = \sum_s a_{ijk} y_{ijk} + \sum_{ns} a_{ijk} Y_{ijk}\ , \tag{1}$$

where s denotes the elements selected in the sample and ns the element not in the sample (the complement of s in the finite population). Now the problem is to predict last part of the right hand side of Eq.(1).

In the Bayesian prediction approach the posterior mean of T is just predictive finite population total;

$$E(T|\, y_s,\ \mu) = \sum_s a_{ijk} y_{ijk} + \sum_{ns} a_{ijk} \mu_{ij}\ . \tag{2}$$

the posterior mean Eq.(2) is a compromise between the naive point estimate and the first stage prior mean.

$$\begin{aligned} E(T \mid y_s) &= E_{\mu|y_s} E(T|y_s, \mu) \\ &= E_{\mu, y_s}(E(T|y_s, \mu)) \\ &= \sum_s a_{ijk} y_{ijk} + \sum_{ns} a_{ijk} E(\mu_{ij}|y_s)\ . \end{aligned} \tag{3}$$

To evaluate $E(\mu_{ij}|y_s)$ of Eq.(3), Malec and Sedransk (1985), Royall (1976), Scott and Smith (1969) and Ericson (1969) did very extensive algebraic manipulation. And above authors assumes variance terms fixed specified values.

4. Gibbs Sampling Approach

The multistage model is considered here an example of a hierarchical Bayes model. Full marginal distributions applied to Gibbs sampler can be derived by computing the partial integral of the joint distribution with respect to each parameter.

4.1 Gibbs sampling algorithm

Gelfand and Smith (1990) discuss the Gibbs sampler as a Markovian updating scheme. The procedure of the Gibbs sampling for several random variables $\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_k$ with some joint distribution $P(\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_k)$ is as follows.

Let $P(\alpha_i \mid \alpha_j, j \neq i, 1, 2, \cdots, k)$ be the full conditional distribution of α_i. Given arbitrary starting values $\alpha_1^{(0)},\ \alpha_2^{(0)}, \cdots,\ \alpha_k^{(0)}$.

$\alpha_1^{(1)}$ is drawn from $P(\alpha_1 \mid \alpha_2^{(0)}, \cdots,\ \alpha_k^{(0)})$.

$\alpha_2^{(1)}$ is drawn from $P(\alpha_2 \mid \alpha_1^{(1)},\ \alpha_3^{(0)}, \cdots,\ \alpha_k^{(0)})$.

$\alpha_i^{(1)}$ is drawn from $P(\alpha_i \mid \alpha_1^{(1)},\ \alpha_2^{(1)}, \cdots,\ \alpha_{i-1}^{(1)},\ \alpha_{i+1}^{(0)}, \cdots,\ \alpha_k^{(0)})$.

$\alpha_k^{(1)}$ is drawn from $P(\alpha_k \mid \alpha_1^{(1)},\ \alpha_2^{(1)}, \cdots,\ \alpha_{k-1}^{(1)})$.

The set of random variables $(\ \alpha_1^{(t)},\ \alpha_2^{(t)}, \cdots,\ \alpha_k^{(t)})$ is obtained after t iterations. Independently replicating this process M times produces M sets of random variables. The marginal density of α_i then is estimates by

$$P(\alpha_i) = \frac{1}{M} \sum_{j=1}^{M} P(\alpha_i \mid \alpha_{1j}^{(t)}, \cdots,\ \alpha_{(i-1)j}^{(t)},\ \alpha_{(i+1)j}^{(t)}, \cdots,\ \alpha_{kj}^{(t)}).$$

4.2 The full conditional distributions of μ_{ij}, ν_i and θ.

Suppose that we assumed proper conjugate distributions at each stage except for θ, which is assumed by the diffuse prior. The hierarchical structure and conjugacy implies the following density

$$p(\mu_{ij} \mid\ y_s,\ \sigma,\ \nu,\ \delta,\ \theta,\ \gamma,\ \mu_{(-i,-j)}) \propto \left[\prod_{k=1}^{M_{ij}} p(y_{ijk} \mid\ \mu_{ij}, \sigma_{ij}) \right] p(\mu_{ij} \mid \nu_i, \delta_i)$$

$$\sim N \left(\frac{\delta_i \sum_{k=1}^{M_{ij}} y_{ijk} + \sigma_{ij}\nu_i}{M_{ij}\delta_i + \sigma_{ij}},\ \frac{\sigma_{ij}\delta_i}{M_{ij}\delta_i + \sigma_{ij}} \right),$$

where $\mu_{(-i,-j)}$ is vector of μ except (ij)-th term and σ, δ and γ denote variances of each hierarchical stage.

$$p(\nu_i \mid y_s, \sigma, \mu, \delta, \theta, \gamma, \nu_{(-i)}) \propto \left[\prod_{j=1}^{N_i} p(\mu_{ij} \mid \nu_i, \delta_i)\right] p(\nu_i \mid \theta, \gamma)$$

$$\sim N\left(\frac{\gamma \sum_{j=1}^{N_i} \mu_{ij} + \delta_i \theta}{\delta_i + N_i \gamma}, \frac{\delta_i \gamma}{\delta_i + N_i \gamma}\right)$$

where $\nu_{(-i)}$ is vector of ν except i-th term and

$$p(\theta \mid y_s, \sigma, \mu, \delta, \gamma, \nu) \propto \left\{\prod_{i=1}^{L} p(\nu_i \mid \theta, \gamma)\right\} p(\theta)$$

$$\sim N\left(\frac{\sum_{i=1}^{L} \nu_i}{L}, \frac{\gamma}{L}\right).$$

4.3 The posterior distribution of σ_{ij}, δ_i and γ.

As in Section 4.2, the prior distributions for hyperparameters $\{\sigma_{ij}\}$, $\{\delta_i\}$ and γ are assumed improper and independent. So we can get the full conditional distributions of $\{\sigma_{ij}\}$, $\{\delta_i\}$ and γ which are inverse gamma distribution.

$$p(\sigma_{ij} \mid y_s, \mu, \nu, \delta, \theta, \gamma, \sigma_{(-i,-j)}) \propto \left[\prod_{k=1}^{M_{ij}} p(y_{ijk} \mid \mu_{ij}, \sigma_{ij})\right] p(\sigma_{ij})$$

$$\sim IG\left(\frac{M_{ij}}{2}, \frac{\sum_{k=1}^{M_{ij}} (y_{ijk} - \mu_{ij})^2}{2}\right),$$

where $\sigma_{(-i,-j)}$ is vector of σ except (ij)-th term,

$$p(\delta_i \mid y_s, \sigma, \mu, \nu, \theta, \gamma, \delta_{(-i)}) \propto \left[\prod_{j=1}^{N_i} p(\mu_{ij} \mid \nu_i, \delta_i)\right] p(\delta_i)$$

$$\sim IG\left(\frac{N_i}{2}, \frac{\sum_{j=1}^{N_i} (\mu_{ij} - \nu_i)^2}{2}\right)$$

where $\delta_{(-i)}$ is vector of δ except i-th term and

$$p(\gamma \mid y_s, \sigma, \mu, \delta, \theta, \nu) \propto \left\{ \prod_{i=1}^{L} p(\nu_i \mid \theta, \gamma) \right\} p(\gamma)$$
$$\sim IG\left(\frac{L}{2}, -\frac{\sum_{i=1}^{L} (\nu_i - \theta)^2}{2} \right).$$

5. Empirical Study

The technique is illustrated using a Sweden Census data(Sarndal, Swensson, and Wretman, 1991; Model Assisted Survey Sampling, Appendix B and C) which are consists of 50 clusters, each clusters have a number of municipalities.

To predict the total of the variable P85 (1985 total population) for the Sweden data, a small two stage sample was selected as follows. In the first stage, a PPS sample of 10 PSU's was drawn. And the drawing probabilities were proportional to value of the cluster total. From each PSU, a few units were selected using PPS sampling with drawing probabilities proportional to number of municipal employees in 1984.

The Gibbs sampling algorithm was applied to the above data and random variates were generated from the full conditional distribution which is available in Section 4.2 and 4.3. The one long-run sequences were taken to get the marginal posterior density for the i-th mean of the i-th cluster. The one long-run sequence algorithm (Zeger and Karim (1991)) was efficient in computation time which is mentioned by Albert and Chib (1993). The first 1000 random variates were droped and the estimates of cluster means were averaged with 1000 random variates in numerical work. The selected samples and estimates are shown in Table I.

Table I : Selected samples and corresponding estimates for each selected cluster

Cluster	sampled unit	cluster estimate
1	38, 42, 21	30.55
2	7, 28, 107	30.47
3	56, 60, 32	35.78
4	229, 81, 35	32.24
5	49, 31, 15	29.05
6	6, 13, 10	11.97
7	17, 12, 10	14.43
8	13, 20, 51	27.17
9	7, 17, 56	26.62
10	24, 67, 29	31.07

The results of our sample gave the predictive estimate of Sweden population, 7,859,558 and the true 1985 Sweden population was 8,285,000. After one sample analysis, we got the above estimate. It shows that the Gibbs approach is available for multistage cluster sampling purposes.

6. Concluding Remarks

To estimate the population characteristics on the multistage cluster sampling, they often use regression approach(Royall (1976), Pfeffermann and Nathan (1981)) and Bayesian predictive inference(Malec and Sedransk (1985), Scott and Smith (1969)) for hierarchical model. Recently applications of the Gibbs sampling algorithm are proposed by many authors. And the adventage of the Gibbs sampling algorithm is easy implimentation when full conditional distribution are easy to work with. The multistage model is often considered an example of a hierarchical Bayes model. There is not available to get the interesting characteristics by using the classical method such as numerical approximation. But interesting characteristic in this model can be easily estimated by using the Gibbs sampling algorithm.

References

[1] Albert, J. M. and Chib, S (1993). Bayesian analysis of Binary and Polychopomous Response Data., *Journal of the American Statistical Association*, Vol. 88, 669-679.

[2] Deely, J.J and Lindley, D.V.(1981). Bayes Empirical Bayes, *Journal of the American Statistical Association*, Vol. 76, 833-841.

[3] Ericson, W. A (1969). Subjective Bayesian Models in Sampling Finite Populations, *J. Roy. Statist. Soc.* Ser. B, 31, 195-233.

[4] Geman, S. and Geman, D (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, IEEE Transaction on Pattern Analysis and Intelligence PAMI-6, NO. 6

[5] Gelfand, A. E and Smith, A. F. M (1990). Sampling Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, Vol. 90, 972-985.

[6] Malec, D and Sedransk, J (1985). Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling *Journal of american statistical association*, Vol. 80, 897-902

[7] Pfeffermann, D and Nathan, G (1981). Regression Analysis of Data from a Cluster Sample, *Journal of the American Statistical Association*, Vol. 76, 681-689.

[8] Sarndal, C. and Swenson, B. and Wretman, J (1991). Model assisted survey sampling, Springer-Verlag.

[9] Scott, A and Smith, T. M. F (1969). Estimation in Multistage Survey, *Journal of the American Statistical Association*, Vol. 72, 516-521

[10] Royall, R. M (1976). The Linear Least Squares Prediction Approach to Two Stage Sampling, *Journal of the American Statistical Association*, Vol. 71, 657-664.

[11] Zeger, S. L and Karim, M. R (1991). Generalized Linear Models with Random Effects: A Gibbs Sampling Approach, *Journal of the American Statistical Association*, Vol. 86, 79-85.