Robust Bayesian Analysis in Finite Population Sampling with Auxiliary Information

Seung A Lee[1] · Sang Hyuck Suh[2] · Dal Ho Kim[3]

Abstract

The paper considers some Bayes estimators of the finite population mean with auxiliary information under priors which are scale mixtures of normal, and thus have tail heavier than that of the normal. The proposed estimators are quite robust in general. Numerical methods of finding Bayes estimators under these heavy tailed priors are given, and are illustrated with an actual example.

Keywords : Auxiliary information, Bayes, Finite population sampling, Gibbs sampler, Robust Bayes

1. Introduction

We consider a finite population sampling with units labeled $1, 2, ..., N$. Let y_i denote the value of a single characteristic attached to the ith unit $(i = 1, ..., N)$. The vector $\boldsymbol{y} = (y_1, ..., y_N)^T$ is the unknown state of nature, and is assumed to belong to $\Theta = R^N$. A subset s of $1, 2, ..., N$ is called a sample. Let $n(s)$ denote the number of elements belonging to s. The set of all possible samples is denoted by S. A design is a function p on S such that $p(s) \in [0, 1]$ for all $s \in S$ and $\sum_{s \in S} p(s) = 1$. Given $\boldsymbol{y} \in \Theta$ and $s = \{i_1, ..., i_{n(s)}\}$ with $1 \leq i_1 \leq \cdots \leq i_{n(s)} \leq N$, let $y(s) = \{y_{i_1}, ..., y_{i_{n(s)}}\}$. One of the main objectives in sample surveys is to draw inference about or some function (real- or vector-valued) $\gamma(\boldsymbol{y})$ of on the basis of $_s$ and $y(s)$. In this article, we will be

---

1) Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
2) Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
3) Corresponding Author : Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
   E-mail: dalkim@knu.ac.kr

concerned exclusively with $\gamma(\boldsymbol{y}) = N^{-1} \sum_{i=1}^{N} y_i$, the finite population mean, although the general methods to be described later are applicable to other parameters of interest as well.

For most sample surveys, for every unit i in the finite population, information is available for one or more auxiliary characteristics, characteristics other than the one of direct interest. We consider the simplest situation when for every unit i in the population, value of a certain auxiliary characteristic, say $x_i(>0)$ is known $(i=1,...,N)$.

The classical estimator of the finite population mean in such cases is the ratio estimator which seems to incorporate the auxiliary information in a very natural manner. Moreover, this estimator can be justified both from the design and model based approach.

We consider a Bayesian viewpoint. For this, we assume the superpopulation model $y_i = \beta x_i + \varepsilon_i$, where ε_i are independent $N(0, \sigma^2 x_i)$, $i = 1,...,N$, while β has a $N(\beta_0, \tau_0^2)$ prior (see Ericson, 1969; Lohr, 1999).

In this article, we consider the idea of using priors which are scale-mixtures of normal when the superpopulation model is assumed to be normal. This is in contrast to Ericson(1969) and others who assumed the conjugate normal priors for normal superpopulation models. From past studies (e.g. Box and Tiao(1968, 1973), Dawid(1973), O'Hagan(1979, 1989) and West(1987)) substantial evidence exists that priors with tails that are flatter than those of the likelihood function tends to be fairly robust. Priors which are scale-mixtures of normal have flatter tails than those of the normal automatically by construction. Also, this class of priors is sufficiently rich since it includes the Student t-family, double exponential, logistic, and the exponential power family among others.

Datta and Lahiri(1995) considered general scale mixtures of normals primarily with the end of outlier detection in the context of small area estimation. Ghosh and Kim(1996) used such priors for Bayes estimation of the finite population mean without auxiliary information. The purpose of this paper is to develop inherently robust Bayes procedures to overcome the problem associated with outliers in the context of finite population sampling in the presence of auxiliary information.

The price to be paid for utilization of inherently robust procedures is computational; closed form calculation is no longer possible. However, the Markov chain Monte Carlo integration techniques, in particular the Gibbs sampling (e.g. Gelfand and Smith(1990) and Gelfand *et al.*(1990)) has proved to be a simple yet powerful tool for performing robust Bayes computations.

For simplicity, we shall assume that $n(s) \neq n \Rightarrow p(s) = 0$, that is, we effectively consider only samples of fixed size n. Also, for notational simplicity, we shall assume that $s = \{i_1,...,i_n\}$ where $1 \leq i_1 \leq \cdots \leq i_n \leq N$. Let $\bar{s} = \{1, 2, ..., N\} - s = \{j_1,...,j_{N-n}\}$ (say), where $1 \leq j_1 \leq \cdots \leq j_{N-n} \leq N$. We

shall write $y(s) = (y_{i_1}, ..., y_{i_n})^T$, $y(\overline{s}) = (y_{j_1}, ..., y_{j_{N-n}})^T$, $x(s) = (x_{i_1}, ..., x_{i_n})^T$, $D(s) = \text{Diag}(x_{i_1}, ..., x_{i_n})$, $x(\overline{s}) = (x_{j_1}, ..., x_{j_{N-n}})^T$, and $D(\overline{s}) = \text{Diag}(x_{j_1}, ..., x_{j_{N-n}})$.

The outline of the remaining sections is as follows. In Section 2, we provide the robust Bayes estimators of finite population mean based on heavy-tailed prior distributions using scale mixtures of normals for unknown σ^2. Finally in Section 3, a numerical example is provided to illustrate the results of the preceding sections.

2. Robust Bayes Estimation of the Finite Population Mean

We consider the normal superpopulation model with unknown mean βx_i and unknown variance $r^{-1}x_i$. Ericson(1969) used a normal-gamma prior on (β, r) in this setting. That is, $\beta \mid r \sim N(\beta_0, r^{-1}\tau_0^2)$ and $r \sim \text{Gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$. But in this case the ratio of model variance and prior variance is known.

Suppose now, more generally, that $y_i \mid \beta, r \overset{ind}{\sim} N(\beta x_i, r^{-1}x_i)$ $(i = 1, ..., N)$, and β and r are independently distributed with $\beta \sim N(\beta_0, \tau_0^2)$ and $\gamma \sim \text{Gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$. We shall use the notations $\overline{y}(s) = n^{-1}\sum_{i \in s} y_i$ and $\overline{y}(\overline{s}) = (N-n)^{-1}\sum_{i \in \overline{s}} y_i$. Then the posterior distribution of $y(\overline{s})$ given s and $y(s)$ is obtained via the following two steps:

(i) conditional on s, $y(s)$ and γ, $y(\overline{s})$ has

$$N\left[\left\{B(r)\beta_0 + (1 - B(r))\frac{\overline{y}(s)}{\overline{x}(s)}\right\}x(\overline{s}), r^{-1}\left\{D(\overline{s}) + \frac{\tau_0^2 x(\overline{s})x^T(\overline{s})}{r^{-1} + \tau_0^2 n\overline{x}(s)}\right\}\right], \qquad (2.1)$$

where $B(r) = r^{-1}/(r^{-1} + \tau_0^2 n\overline{x}(s))$;

(ii) the conditional distribution of r given s and $y(s)$ has probability density function

$$f(r \mid s, y(s)) \propto \frac{r^{(n+g_0-1)/2-1}}{(r^{-1} + \tau_0^2 n\overline{x}(s))^{1/2}} \exp\left[-\frac{r}{2}\left\{a_0 + \sum_{i \in s}\left(y_i - \frac{\overline{y}(s)}{\overline{x}(s)}x_i\right)^2 \frac{1}{x_i}\right.\right.$$
$$\left.\left. + \frac{r^{-1}n\overline{x}(s)}{r^{-1} + \tau_0^2 n\overline{x}(s)}\left(\frac{\overline{y}(s)}{\overline{x}(s)} - \beta_0\right)^2\right\}\right]. \qquad (2.2)$$

Note that under the posterior distribution given (2.1) and (2.2), the Bayes estimator of $\gamma(\boldsymbol{y})$ is given by

$$\delta^B(s,\,y(s)) = f\overline{y}(s) + (1-f)\overline{x}(\overline{s})\left\{E[B(r)\mid s,\,y(s)]\beta_0 + (1-E[B(r)\mid s,\,y(s)])\frac{\overline{y}(s)}{\overline{x}(s)}\right\},$$

where $f = n/N$. Also, one gets

$$V[\gamma(\boldsymbol{y})\mid s,\,y(s)] \;=\; N^{-2}(N-n)\overline{x}(\overline{s})E\left[r^{-1}\left\{1 + (N-n)\frac{\overline{x}(\overline{s})}{n\overline{x}(s)}(1-B(r))\right\}\mid s,\,y(s)\right]$$
$$+\;(1-f)^2(\overline{x}(\overline{s}))^2\left(\frac{\overline{y}(s)}{\overline{x}(s)} - \beta_0\right)^2 V[B(r)\mid s,y(s)].$$

In order to robustify the above model, consider the case when

(i) $y_i\mid\beta,\,r \overset{ind}{\sim} N(\beta x_i,\,r^{-1}x_i)\ (i=1,...,N)$, (ii) $r \sim Gamma(\frac{1}{2}a_0,\,\frac{1}{2}g_0)$ and

(iii) $\beta \sim \frac{1}{\tau_0}p\left(\frac{\beta-\beta_0}{\tau_0}\right)$, where $p(t)=\int_0^\infty \lambda^{1/2}\phi(t\lambda^{1/2})g(\lambda)d\lambda$ and $\phi(t)$ denotes the standard normal pdf. Here $p(\cdot)$ is a scale mixture of the normal distribution with mixing distribution $g(\cdot)$. Note that the prior distribution of β does not depend on r. We can write (iii) in the following two steps; (iiia) $\beta\mid\lambda \sim N(\beta_0,\lambda^{-1})$ and (iiib) $\lambda \sim \tau_0^2\,g(\tau_0^2\lambda)$, where $\int_0^\infty g(t)dt = 1$. The following list identifies the necessary functional form for $g(\lambda)$ to obtain a wide range of densities which represent departures from normality (see West(1987)):

t priors: If $k\lambda \sim \chi_k^2$ then β is Student t with k degrees of freedom, location parameter β_0 and scale parameter τ_0.

double-exponential priors: If $1/\lambda$ has exponential distribution with mean 2 then β is double-exponential with location parameter β_0 and scale parameter τ_0.

exponential power family priors: If λ has positive stable distribution with index $\alpha/2$ then β has exponential power distribution with location parameter β_0 and scale parameter τ_0.

logistic priors: If $\sqrt{\lambda}$ had the asymptotic Kolmogorov distance distribution then β is logistic with location parameter β_0 and scale parameter τ_0. [A random variable Z is said to have an asymptotic Kolmogorov distance distribution if it has a probability density function of the form

$$f(z) = 8z\sum_{j=1}^\infty (-1)^{j-1}j^2\exp(-2j^2z^2)I_{(0,\infty)}(z)].$$

Then the posterior distribution of $y(\overline{s})$ given by s and $y(s)$ is obtained as follows:

(i) conditional on s, $y(s)$, λ and r, $y(\overline{s})$ has

$$N\left[\left\{B(\lambda,r)\beta_0 + (1-B(\lambda,r))\frac{\overline{y}(s)}{\overline{x}(s)}\right\}x(\overline{s}) , \ r^{-1}\left\{D(\overline{s}) + \frac{x(\overline{s})x^T(\overline{s})}{\lambda r^{-1} + n\overline{x}(s)}\right\}\right], \quad (2.3)$$

where $\quad B(\lambda,r) = \lambda r^{-1}/(\lambda r^{-1} + n\overline{x}(s));$

(ii) the conditional distribution of λ and r given s and $y(s)$ has probability density function

$$f(\lambda,r|s,y(s)) \propto \frac{r^{(n+g_0-1)/2-1}}{(r^{-1}+\lambda^{-1}n\overline{x}(s))^{1/2}} \exp\left[-\frac{r}{2}\left\{a_0 + \sum_{i\in S}\left(y_i - \frac{\overline{y}(s)}{\overline{x}(s)}x_i\right)^2\frac{1}{x_i}\right.\right.$$
$$\left.\left. + \frac{r^{-1}n\overline{x}(s)}{r^{-1}+\lambda^{-1}n\overline{x}(s)}\left(\frac{\overline{y}(s)}{\overline{x}(s)}-\beta_0\right)^2\right\}\right]g(\tau_0^2\lambda). \quad (2.4)$$

Note that under the posterior distribution given in (2.3) and (2.4), the Bayes estimator of $\gamma(\boldsymbol{y})$ is given by

$$\delta_{SM}(s,y(s)) = f\overline{y}(s) + (1-f)\overline{x}(\overline{s})\left\{E[B(\lambda,r)|s,y(s)]\beta_0 + (1-E[B(\lambda,r)|s,y(s)])\frac{\overline{y}(s)}{\overline{x}(s)}\right\}.$$
$$(2.5)$$

Also, one gets

$$V[\gamma(\boldsymbol{y}) \mid s,y(s)] = N^{-2}(N-n)\overline{x}(\overline{s})E\left[r^{-1}\left\{1+(N-n)\frac{\overline{x}(\overline{s})}{n\overline{x}(s)}(1-B(\lambda,r))\right\}\mid s,y(s)\right]$$
$$+ (1-f)^2(\overline{x}(\overline{s}))^2\left(\frac{\overline{y}(s)}{\overline{x}(s)}-\beta_0\right)^2 V[B(\lambda,r)\mid s,y(s)]. \quad (2.6)$$

Notice that the calculations in (2.5) and (2.6) can be performed using two -dimensional numerical integration. Alternately, one can use the Monte Carlo numerical integration techniques to generate the posterior distribution and the associated means and variances. More specifically, we use Gibbs sampling recently popularized by Gelfand and Smith(1990) and Gelfand et al.(1990).

Gelman and Rubin(1992) adapt multiple sequences, with starting points drawn from an overdispersed distribution, to monitor the convergence of the Gibbs sampler. Specifically, $m \geq 2$ independent sequences are generated, each of length $2d$. But to diminish the effect of the starting distribution, the first d iterations of each sequence are discarded. Hence, we have $m \times d$ simulated values for each parameter of interest.

Using Gibbs sampling, the posterior distribution of $y(\overline{s})$ is approximated by

$$(md)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{d}[y(\overline{s}) \mid s,y(s),\beta=\beta_{ij},\lambda=\lambda_{ij},r=r_{ij}].$$

To estimate the posterior moments, we use once again the Rao-Blackwellized

estimates as in Gelfand and Smith(1991). Note that $E[\gamma(\boldsymbol{y}) \,|\, s, y(s)]$ is approximated by

$$f\overline{y}(s) + (1-f)\overline{x}(\overline{s})(md)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{d}\left\{B(\lambda_{ij}, r_{ij})\beta_0 + (1-B(\lambda_{ij}, r_{ij}))\frac{\overline{y}(s)}{\overline{x}(s)}\right\}.$$

Next one approximates $V\left[\gamma(\boldsymbol{y}) \,|\, s, y(s)\right]$ by

$$N^{-2}(N-n)\overline{x}(\overline{s})(md)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{d}\left[r_{ij}^{-1}\left\{1 + (N-n)\frac{\overline{x}(\overline{s})}{n\overline{x}(s)}(1-B(\lambda_{ij}, r_{ij}))\right\}\right]$$

$$+ (1-f)^2\left(\overline{x}(\overline{s})\right)^2\left(\frac{\overline{y}(s)}{\overline{x}(s)} - \beta_0\right)^2\left[(md)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{d}(B(\lambda_{ij}, r_{ij}))^2\right.$$

$$\left. - \left\{(md)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{d}B(\lambda_{ij}, r_{ij})\right\}^2\right].$$

The Gibbs sampling analysis is based on the following posterior distribution:

(i) $\beta \,|\, s, y(s), y(\overline{s}), \lambda, r \sim N\left[\left(\lambda\beta_0 + \sum_{i=1}^{N}y_i/r^{-1}\right)\Big/\left(\lambda + \sum_{i=1}^{N}x_i/r^{-1}\right), \left(\lambda + \sum_{i=1}^{N}x_i/r^{-1}\right)^{-1}\right];$

(ii) $f(\lambda \,|\, s, y(s), y(\overline{s}), \beta, r) \propto \sqrt{\lambda}\exp\left[-\frac{\lambda}{2}(\beta-\beta_0)^2\right]g(\tau_0^2\lambda);$

(iii) $r \,|\, s, y(s), y(\overline{s}), \beta, \lambda \sim Gamma\left(\frac{1}{2}\left\{a_0 + \sum_{i=1}^{N}(y_i - \beta x_i)^2\frac{1}{x_i}\right\}, \frac{1}{2}(N+g_0)\right);$

(iv) $y(\overline{s}) \,|\, s, y(s), \beta, \lambda, r \sim N\left[\beta x(\overline{s}), r^{-1}D(\overline{s})\right].$

Note that if $k\lambda \sim \chi_k^2$ then $f(\lambda \,|\, s, y(s), y(\overline{s}), \beta, r)$ reduces to a Gamma $(\frac{1}{2}\left\{\tau_0^2 k + (\beta-\beta_0)^2\right\}, \frac{1}{2}(k+1))$ density. [A random variable W is said to have a Gamma (α, β) distribution if it has a probability density function of the form $f(w) \propto \exp(-\alpha w)w^{\beta-1}I_{(0,\infty)}(w)$, where I denotes the usual indicator function.] Also, if $1/\lambda$ has exponential distribution with mean 2 then $f(\lambda \,|\, s, y(s), y(\overline{s}), \beta, r)$ reduces to a IGN$(1/\sqrt{\tau_0^2(\beta-\beta_0)^2}, 1/\tau_0^2)$ density. [A random variable V is said to have a IGN(η_1, η_2) distribution if it has a probability density function of the form

$$f(v) = \sqrt{\frac{\eta_2}{2\pi}}\, v^{-3/2}\exp(-\frac{\eta_2(v-\eta_1)^2}{2\eta_1^2 v})I_{(0,\infty)}(v).]$$

3. Numerical Example

We illustrate the methods of Section 2 with an analysis of data in Royall and Cumberland(1981) for an empirical study of the ratio estimator and estimates of its variance. Our population consists of the 1960 and 1970 population, in millions, of 125 US cities with 1960 population between 100,000 and 1,000,000. Here the auxiliary information is the 1960 population.

The problem is to estimate the mean (or total) number of inhabitants in those 125 cities in 1970. For the complete population in 1970, we find the population mean to be 0.29034 and the population variance 4.84844×10^{-3}. We select 20% simple random sample without replacement from this population. So the sample size is n=25. We use both 1950 and 1960 populations in 125 cities to elicit the prior in our setting so that $\beta_0 = 1.15932$ and $\tau_0^2 = 1.21097 \times 10^{-3}$.

In deriving the robust Bayes estimates based on heavy-tailed prior distributions using scale mixtures of normals, we have considered Gibbs sampler with 10 independent sequences, each with a sample of size 5000 with a burn in sample of another 5000.

Table 1 provides Bayes estimates of $\gamma(\boldsymbol{y})$ and the associated posterior standard deviations for normal, double-exponential and t priors with degree of freedom 1, 3, 4, 10 and 15. We have used $a_0 = g_0 = 0$ to ensure some form of diffuse gamma prior for the inverse of the variance component in our superpopulation model.

An Inspection of Table 1 reveals that there can be significant improvement in the estimates of $\gamma(\boldsymbol{y})$ by using heavy-tailed prior distributions rather than using the normal prior distribution in the sense of the closeness to $\gamma(\boldsymbol{y})$. For instance, using the double-exponential and the t(1), t(3), t(5), t(10) and t(15) priors, the percentage improvements over the normal are given respectively by 68.35%, 81.98%, 73.36%, 65.83%, 51.45% and 42.04%. Here the percentage improvement of e_1 over e_2 is calculated by

$$((e_2 - \text{truth})^2 - (e_1 - \text{truth})^2)/(e_2 - \text{truth})^2,$$

where e_1 is the robust Bayes predictor based on heavy-tailed prior distributions and e_2 is the Bayes predictor using the normal prior. Also as one might expect, flatter the prior, closer is the Bayes estimates to sample mean. In general, for most cases we have considered, the Cauchy prior(i.e., t prior with 1 degree of freedom) leads to an estimate which is closest to population mean.

We adopt the basic approach of Gelman and Rubin(1992) to monitor the convergence of the Gibbs sampler. For β, we simulate m=10 independent sequences of each of length $2d$=10000 with starting points drawn from a t

distribution with 2 degree of freedom. The justification for t distributions as well as the choice of the specific parameters of this distribution are given below.

First note that from the posterior distribution of λ and r given s and $y(s)$ as given in (2.4), we find the posterior mode, say $\hat{\lambda}$ and $\hat{r}$ by using the Newton−Raphson algorithm. Also, we use $\overline{y}(s)$ for y_i, $i \in \overline{s}$ based on sample. We can now very well use the $N[(\hat{\lambda}\beta_0 + N\overline{y}(s)/\hat{r}^{-1})/(\hat{\lambda} + N\overline{x}(s)/\hat{r}^{-1}), (\hat{\lambda} + N\overline{x}(s)/\hat{r}^{-1})^{-1}]$ as the starting posterior distribution for β. But in order to start with an overdispersed distribution as recommended by Gelman and Rubin, we take t distribution with 2 degree of freedom. Also, note that once the initial β value has been generated, the rest of the procedure uses the posterior distributions as given in (i)−(iv) in Section 2.

Next, as in Gelman and Rubin, we compute

$B/5000 =$ the variance between the 10 sequence means $\overline{\beta}_{i.}$ each based on 5000 β values, that is $B/5000 = \sum_{i=1}^{10}(\overline{\beta}_{i.} - \overline{\beta}_{..})^2/(10-1)$, where $\overline{\beta}_{..} = \frac{1}{10}\sum_{i=1}^{10}\overline{\beta}_{i.}$,

$W =$ the average of the 10 within−sequence variance s_i^2 each based on (5000−1) degree of freedom, that is $W = \frac{1}{10}\sum_{i=1}^{10}s_i^2$.

Then, find $\widehat{\sigma^2} = \frac{5000-1}{5000}W + \frac{1}{5000}B$ and $\widehat{V} = \widehat{\sigma^2} + \frac{1}{(10)(5000)}B$. Finally, find $\widehat{R} = \widehat{V}/W$. If $\widehat{R}$ is near 1 for all scalar estimands of interest, it is reasonable to assume that the desired convergence is achieved in the Gibbs sampling algorithm.

Table 1 provides the $\widehat{R}$ values (the potential scale reduction factors) corresponding to the estimand β using the double−exponential and the t(1), t(3), t(5), t(10) and t(15) priors based on $10 \times 5000 = 50000$ simulated values.

<Table 1> Bayes Estimates and Associated Posterior Standard Deviations

Priors	Bayes Estimates	Posterior SD	Potential scale reduction factor($\hat{R}$)
Normal	0.29809	5.85559×10^{-3}	0.99981
DE	0.29470	6.61138×10^{-3}	0.99983
t(1)	0.29363	6.82615×10^{-3}	0.99982
t(3)	0.29434	6.70928×10^{-3}	0.99982
t(5)	0.29487	6.60927×10^{-3}	0.99983
t(10)	0.29574	6.42601×10^{-3}	0.99983
t(15)	0.29624	6.31247×10^{-3}	0.99981

References

1. Box, G. E. P. and Tiao, G. C. (1968). A Bayesian approach to some outlier problems. *Biometrika*, 51, 153-167.
2. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA
3. Datta, G. S. and Lahiri, P. (1995). Robust hierarchical Bayes estimation of small area characteristics in presence of covariates. *Journal of Multivariate Analysis*, 54, 310-328.
4. Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika*, 60, 664-667.
5. Ericson, W. A. (1969). Subjective Bayesian models in sampling finite populations (with discussion). *Journal of the Royal Statistical Society*, Ser. B, 31, 195-233.
6. Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M.(1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Associations*, 85, 972-985.
7. Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculation marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
8. Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations. *Communications in Statistics-Theory and Methods*, 20, 1747-1766.
9. Gelman, A. and Rubin, D. B.(1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.
10. Ghosh, M. and Kim, D.H. (1996). Bayes estimation of the finite population mean under heavy-tailed priors. *Calcutta Statistical Association Bulletin*, 46, 181-195.
11. Lohr, S. (1999). *Sampling : Design and Analysis.* Duxbury Press.
12. O'Hagan, A. (1979). On outlier rejection phenoma in Bayes inference. *Journal of the Royal Statistical Society*, Ser. B, 41, 358-367.
13. O'Hagan, A. (1989). Modeling with heavy tails. *Bayesian Statistics 3*. Eds. J.M. Bernardo, M.H. DeGroot, D.V. Linley and A.F.M. Smith, Oxford University Press, Oxford, 349-359.
14. Royall, R. M. and Cumberland, W. G. (1981). An empirical study of the ratio estimator and estimators of its variance (with discussion). *Journal of the American Statistical Association*, 76, 66-88
15. West. M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74, 646-648.