BAYES ESTIMATORS FOR FINITE POPULATION TOTAL AND MEAN UNDER SOME SPECIFIED PRIOR DISTRIBUTIONS

XIAOPING HU

A THESIS

IN

THE DEPARTMENT

OF

MATHEMATICS AND STATISTICS

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science
at
Concordia University
Montreal, Quebec, Canada

March 1991

ISBN 0-315-64742-6

ABSTRACT

BAYES ESTIMATORS FOR

FINITE POPULATION TOTAL AND MEAN

UNDER SOME SPECIFIED PRIOR DISTRIBUTIONS

Xiaoping Hu

This thesis considers estimation of a finite population total and mean by adopting the Bayesian statistical viewpoint. Many results have been found based on different methods and different assumptions on prior distributions.

A linearity assumption on posterior distribution lets us have a set of estimators, called Ericson's estimators. A linear Bayesian estimator derived from Hartigan's theory is also introduced in Chapter 2. In some cases, both methods give the same results.

Chapter 3 deals with subjective Bayesian estimation based on different priors. We obtain estimators which differ with classical estimators in many situations. It is shown, though, that for large samples, our results are very close to the classical ones.

A regression based estimation procedure has been described in Chapter 4, and a numerical example illustrates the results obtained in previous chapters.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor Professor Y. P. Chaubey (Ph.D) for his guidance, encouragement and support in the preparation of this thesis.

TABLE OF CONTENTS

Chapter 1

1. INTRODUCTION

In this chapter we introduce some definitions and notations used frequently in the sequel. We also want to record some classical sampling techniques and conclusions [see Cochran (1977)] without proof for the sake of completeness.

Define a finite population U of N distinguishable elements labelled by the integers 1, 2,..., N, U={1, 2,..., N}, and define $Y=(Y_1, Y_2,..., Y_N)$ as the unknown values of some characteristic corresponding to the N units. The Y_i's can be also taken as vector valued, a case of considerable practical importance, which we will deal with in later chapters.

We seek to estimate a function of Y, which is called a parameter of the population, for example, the population mean $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N} Y_i$, the population total, $Y=\sum_{i=1}^{N} Y_i$, the ratio $R=(\sum_{i=1}^{N} Y_i)/(\sum_{i=1}^{N} X_i)$ where x_i is the value of another variable X related to Y, and $S^2=\sum_{i=1}^{N} (Y_i-\bar{Y})^2$, the population variance.

Traditional sampling texts (such as Cochran,1977) give some estimators for Y, $\bar{Y}$, R, and S^2. For similicity we consider SRS with repacement and describe some notations below.

Let $y=\{y_1, y_2,...,y_n\}$ be a simple random sample with replacement of size n drawn from a population U, then we have some standard results given below;

i) $\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i,$ is an unbiased estimator of the population mean with variance $Var(\bar{y})=(1-f)\frac{S^2}{n}$ where f= n/N.

ii) $y=\frac{N}{n}\sum_{i=1}^{n} y_i,$ is an unbiased estimator of the population total Y with variance $Var(\hat{Y})=\frac{N^2}{n}S^2(1-f)$.

iii) $\hat{S}^2=\frac{1}{(n-1)}\sum_{i=1}^{n}(y_i-\bar{y})^2$ is an unbiased estimator of the population variance.

iv) $\hat{R}=\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$ is an approximately unbiased estimator of R with variance $Var(\hat{R}) \doteq \frac{(1-f)}{n\bar{X}^2}\frac{\sum_{i=1}^{N}(Y_i-RX_i)^2}{(N-1)}$.

There are many sampling schemes in sampling theory other than simple random sampling(SRS) with replacement or without replacement. But in this thesis only SRS will be adopted.

The Bayesian estimation for finite population parameters is the main topic we will deal with in this thesis, thus we present same basic elements of Bayesian estimatation in the following section.

1.1 Bayesian Approach to Estimation in a Continuous Population

Theorem 1.2.1(Bayes' Theorem) Suppose that Y=(Y_1, Y_2,

..., Y_n)' is a vector of n observations whose probability distribution, p(Y| θ), depends on the values of k parameters $\theta' = (\theta_1, \theta_2, \ldots, \theta_k)$. Also suppose that θ itself has a probability distribution p(θ), then given the observed data y, the conditional distribution of θ is

$$p(\theta \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \theta)\, p(\theta)}{p(\boldsymbol{y})} \propto l(\theta \mid \boldsymbol{y})\, p(\theta)$$

where $p(\boldsymbol{y}) = \int_{\theta} p(\boldsymbol{y} \mid \theta)\, dF(\theta)$, F(θ) being the distribution function corresponding to p(θ), and l(θ |y)=p(y| θ) which is called the likelihood function of θ for given y.

A prior distribution , p(θ), which is supposed to represent what is known about unknown parameters before the data is available, plays an important role in Bayesian analysis.

In the subjective Bayesian viewpoint, [Jeffreys (1967), Box and Tiao (1973), Berger (1980)] a prior distribution could be any one depending on the knowledge about the parameters even if it is not a proper distribution. In this case, we call it as improper prior.

A basic property of a proper probability density function f(x) is that it integrates over its admissible range D to 1, that is $\int_D f(x)\, dx = 1$, but for an improper prior, we relax this constraint on f(x) and then it is possible that $\int_D f(x)\, dx = \infty$. When we do not have any a priori information about the parameters, we could apply an improper prior.

In many cases, the prior distribution will be dominated by the likelihood function when n, the number of observations, is large.

If we assume that the observations Y_1, Y_2, ...,Y_n in the sample are distributed independently of each other, then

$$p(\theta \mid Y_n, Y_{n-i}, \ldots, Y_1) \propto p(\theta) l(\theta \mid Y_n) l(\theta \mid Y_{n-1}), \ldots, l(\theta \mid Y_1)$$

$$\propto p(\theta \mid Y_{n-1}, Y_{n-2}, \ldots, Y_1) l(\theta \mid Y_n)$$

Since p(θ) is given, $p(\theta \mid Y_n, Y_{n-1}, \ldots, Y_1)$ could be mostly affected by the product of $l(\theta \mid Y_n) l(\theta \mid Y_{n-1}) \cdots l(\theta \mid Y_1)$ after n becomes large.

There are many techniques to get a prior p.d.f, such as subjective assumptions, empirical methods, maximal data information priors(MDIP) and conjugate priors, etc. For example, the following paragragh describes the procedure of obtaining MDIP.

In Bayesian analysis, it is often desirable to have a posterior distribution reflect mainly the information in a given sample of data. To achieve this objective, it is necessary to employ a prior distribution that adds little information to sample information. The basic idea underlying MDIP's is that they provide maximal prior average data information relative to the information in the prior distribution, with information being represented by

Shannon's(1948) measure[1] . Let Y be a random vector with a proper probability density function, $p(y|\theta)$, defined to be positive in region R, where θ is a scalar parameter such that $a \le \theta \le b$, with a and b finite. Then the information in the data distribution $p(y|\theta)$ is defined to be $I(\theta)$, given by:

$$I(\theta)=\int_R p(\boldsymbol{y}\mid\theta)\log(p(\boldsymbol{y}\mid\theta))\,dy$$

The prior average information in the data pdf, denoted by J, is:

$$J=\int_a^b I(\theta)\,p(\theta)\,d\theta$$

where $p(\theta)$ is a proper prior pdf defined to be non-negative for $a \le \theta \le b$ and zero elsewhere. Then the prior average information in the data pdf, J, minus the information in the prior pdf,

$$G=J-\int_a^b p(\theta)\log(p(\theta))\,d\theta$$

$$=\int_a^b p(\theta)\,I(\theta)\,d\theta-\int_a^b p(\theta)\log(p(\theta))\,d\theta$$

---

[1] Shannon's uncertainty measure is $W=-\int_R p(x)\log(p(x))\,dx$

$$=\int_a^b\int_R \log\frac{p(Y\mid\theta)}{p(\theta)}\,p(y,\theta)\,dY d\theta$$

where p(y, θ)=p(θ)p(y| θ) is the joint pdf for θ and Y.

We now present the following definition of a MDIP pdf:

Definition 1.2.1 A MDIP pdf is a proper, normalized prior pdf that maximizes G.

The following theorem provides a simple formula for MDIP pdf's.

Theorem 1.2.2 The normalized MDIP pdf, $p^*(\theta)$ which maximizes G, is given by:

$$p^*(\theta)=ce^{I(\theta)}$$

where c is a normalizing constant.

Particularly, if I(θ) is a constant, independent of θ , then the MDIP pdf is the uniform pdf, or $p^*(\theta)=1/(b-a)$, for a≤ θ ≤b.

Proof: Let g(θ) be an arbitary non-negative function with $\int_a^b g(\theta)\,d\theta=0$ Consider the integral,

$$G(\varepsilon)=\int_a^b\{I(\theta)-\log(p(\theta)+\varepsilon g(\theta))\}\{p(\theta)+\varepsilon g(\theta)\}\,d\theta$$

then, since $\int_a^b g(\theta)\,d\theta=0$, we have:

Since G=G(θ), G is maximized for prior p(θ), if

$$G'(\varepsilon)=\int_a^b g(\theta)\{I(\theta)-\log(p(\theta)+\varepsilon g(\theta))\}d\theta$$

$G'(\theta)=\theta$, i.e. $\log(p(\theta))=I(\theta)$, equivalently, $p(\theta)=c \exp\{I(\theta)\}$, where $c=1/\int_a^b \exp\{I(\theta)\}d\theta$ is the normalizing constant.

The posterior distribution of θ given **Y** is defined to be the conditional distribution of θ given the sample observation **Y**. Just like the prior distribution reflects knowledge about θ prior to the experimentation, the posterior distribution reflects the updated beliefs about θ after observing the sample **Y**. In other words, the posterior distribution combines the prior beliefs about θ with the information about θ contained in the sample. Note that the likelihood principle is implicitly assumed in the above statement, in that there is felt to be no sample information about θ other than that contained in $f(Y|\theta)$.

In general, p(**Y**) and $p(\theta|Y)$ are not algebraically obtainable knowing a prior. In many cases, however, $p(\theta|Y)$ can be evaluated numerically.

The concept of loss function initially comes from decision theory. It is one of the three basic elements of the statistical decision theory which established by A. Wald in 1950 (the remaining two are the probability space and action space).

<u>Definition 1.2.2</u> A loss function can be defined as a

real valued function $L(\hat{\theta},\theta)$ which measures possible consequences caused by a difference between the action $\hat{\theta}$ made by the decision maker and the true state of nature θ .[2]

There are mainly three kinds of loss functions used in Bayesian statistics.

1) Squared-Error Loss:

We define a squared-error loss $L(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2$ for the univariate situation and $L(\hat{\theta},\theta)=\sum_{i=1}^{k}(\hat{\theta}_i-\theta_i)^2$ for the k-variate situation.

2) Linear Loss:

$$L(\hat{\theta},\theta)=\begin{cases} K_1(\theta-\hat{\theta}) & if\ \theta-\hat{\theta}\geq 0 \\ K_2(\hat{\theta}-\theta) & if\ \theta-\hat{\theta}<0 \end{cases}$$

Where the constants K_1 and K_2 can be chosen to reflect the reletive importance of underestimation and overestimation, if K_1 equals to K_2, then the linear loss becomes: $L(\hat{\theta}-\theta)=|\theta-\hat{\theta}|$, which is also called absolute error loss.

3) "0-1" Loss:

$$L(\theta,\hat{\theta})=\begin{cases} 0 & if\ \theta\in\Theta_1 \\ 1 & if\ \theta\in\Theta_2 \end{cases}$$

The "0-1" loss may be interpreted as follows:

If a correct decision or estimation is made, then loss is zero, and if a incorrect decision or estimation is made, the

---

[2] In our case, θ is a parameter vector of our finite population and $\hat{\theta}$ is an estimator of θ

loss is one.

Squared-error loss makes the further calculations relatively straightforward and simple. Thus we have chosen to apply only this type of loss function in our analysis.

Bayesian Estimation

As in the most common classical technique, the maximum likelihood estimation, a Bayesian estimate of θ, $\hat{\theta}$, is commonly defined as the largest mode of the posterior distribution $p(\theta \mid Y)$ (i.e. the value $\hat{\theta}$, which maximizes $p(\theta \mid Y)$, considered as a function of θ. Other common Bayesian estimates of θ include the mean and the median of the posterior distribution $p(\theta \mid Y)$.

Another way to define a Bayesian estimator of θ depends on the selection of the loss function $L(\theta, \hat{\theta})$.

Theorem 1.2.3 A Bayes estimator of θ is the value $\hat{\theta}$ which minimizes the posterior expected loss

$$E(L(\theta,\hat{\theta}))=\int_{-\infty}^{\infty} L(\theta,\hat{\theta})\,p(\theta \mid Y)\,d\theta$$

1) The Bayes estimator of θ is the posterior mean, if the loss function is the squared-error loss.

2) The Bayes estimator of θ is the posterior median, if the loss function is the absolute-error loss.

3) The Bayes estimator of θ is the posterior mode, if the loss function is the "0-1" loss, where $\Theta_1=\{\hat{\theta}-\varepsilon, \hat{\theta}+\varepsilon\}$ and $\Theta_2=\Theta_1^c$.

Since we are adopting the squared-error loss in our analysis, we will use the posterior mean as our estimator.

Unlike classical statistics, a parameter of a population is treated in Bayesian statistics as a random variable, so it has its own probability distribution. All inferences are done based on the posterior distribution. Since the parameter is a random variable, it has a posterior variance which can generate a confidence region, now called a 'credible region', for the estimation of the parameter, it is defined below.

Definition 1.2.3 A 100(1-α)% credible region for the parameter θ is a minimal subset C of the parameter space such that:

$$1-\alpha \le P(C \mid \mathbf{Y}) = \int_C p(\theta \mid \mathbf{Y})\, d\theta$$

Chapter 2

Adaptation of Bayes' Approach to Finite Populations

2.1 Ericson's Model

In 1969, Dr. W.A. Ericson published his 130 pages(with discussion) article "subjective Bayesian models in sampling finite populations", and gave us a general and basic subjective Bayesian approach to the finite population sampling. This paper gave a new direction to researches in this field. A model for the posterior expectation was the main topic in his paper. We will consider his model in this chapter and compare its results with other models later.

Let $s=\{i_1, i_2, \ldots, i_n\}$ be a sample and $y=\{y_{i_1}, y_{i_2}, \ldots, y_{i_n}\}$ be their associated observed characteristic values, where $i_j \in \{1, 2, \ldots, N\}$. It is therefore assumed that if element j is included in the sample, then the value of Y_j becomes known with certainty.

Definition 2.1 Sample Design

A sample design is a pair (s, P) where s is an element of S of all possible samples, and P is a probability measure defined on S such that $P(s) \geq 0$ and $\sum_{s \in S} P(s) = 1$.

Definition 2.2 The likelihood function of Y given (s,y)

The likelihood function of Y given sample design (s, y) is given by

$$l\{Y;(s,y)\} = \begin{cases} kP(s) & \text{for } Y \ni S(Y)=y \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where k > 0 is an arbitrary constant, S is a matrix operator, such that $S(Y)=(Y_{i_1},Y_{i_2},\ldots Y_{i_n})$.

<u>Theorem 2.1</u>

If a joint N-dimensional prior on Y is given with density p'(Y), then the posterior distribution of Y given the sample (s,y) is

$$p\{Y\,|(s,y)\} \propto \begin{cases} p(s)p'(Y), & \text{for } Y \ni S(Y)=y \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

If p(s) is independent of Y, then this posterior distribution of Y is

$$p\{Y\,|(S,Y)\} \propto \begin{cases} \dfrac{p'(Y)}{p'_{S(Y)}(Y)} & \text{for } Y \ni S(Y)=y \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

where $p'_{S(Y)}(y) \neq 0$ is the marginal prior distribution of S(Y).

<u>Definition 2.3 Exchangeable random variables</u>

Random variables Y_1, Y_2, ... ,Y_N are said to be exchangeable if each of the N! permutations $Y_{i_1},Y_{i_2},\ldots,Y_{i_N}$, has the same joint probability distribution.

In particular, independent random variables are exchangeable. One such family of exchangeable priors is given by

$$p'(Y)=\prod_{i=1}^{N} p'_i(y_i)=\int_{\gamma\epsilon\Gamma}\prod_{i=1}^{N} p(Y_i \mid \gamma)\,dF(\gamma) \qquad (2.4)$$

where γ may be interpreted as a secondary parameter with distribution F(γ), and Γ is the secondary parameter space.

In this case, the posterior distribution will be

$$p\{Y \mid (S,Y)\} \propto \begin{cases} \int_{\theta\epsilon\Theta}\prod_{i\epsilon s} p(Y_i \mid \theta)\prod_{i\notin s} p(Y_i \mid \theta)\,dF(\theta) & Y \ni S(Y)=y \\ 0 & otherwise \end{cases}$$

$$=\begin{cases} \int_{\theta\epsilon\Theta}\prod_{i\notin s} p(Y_i \mid \theta)\,dF(\theta \mid Y) & Y \ni S(Y)=y \\ 0 & otherwise \end{cases} \qquad (2.5)$$

where F(θ |Y) is the posterior distribution function of θ .

2.2 Estimation of Posterior Mean under the Squared-Error Loss

In this section, we show that under reasonably general conditions the Bayesian estimator of population mean with the square-error loss, i.e. the posterior mean of μ as we mentioned in last chapter, can be written as a weighted average of sample mean $\bar{y}$ and the prior expectation of μ , E(μ), with weights proptional to the prior variance of μ , Var(μ), and the prior expectation of the conditional sampling variance of $\bar{Y}$. We mention below some important theorems from Ericson. [1]

Theorem 2.2 i) Under the exchangeable prior (2.4), the

[1] μ=E(Y_i), i=1,2,...,N

posterior expectation of $\mu=\frac{1}{N}\sum_{i=1}^{N} Y_i$ is given by

$$\boldsymbol{E}\{\mu \mid (S,Y)\}=\frac{1}{N}[n\bar{y}+(N-n)\boldsymbol{E}\{Y_i \mid (S,Y)\}$$

Let the prior expectation be $\mu(\theta)$ where θ is a set of parameters and $\mu(\theta)=\boldsymbol{E}(Y_i \mid \theta)$, if $\boldsymbol{E}\{\mu(\theta) \mid (S,Y)\}=\alpha\bar{y}+\beta$, then

$$\boldsymbol{E}\{\mu \mid (S,Y)\}=\frac{\bar{y}\boldsymbol{Var}(\mu)+m'\boldsymbol{E}_\mu\boldsymbol{Var}(\bar{y} \mid \mu)}{\boldsymbol{Var}(\mu)+\boldsymbol{E}_\mu\boldsymbol{Var}(\bar{y} \mid \mu)} \qquad (2.6)$$

where α and β are all constants, $m'=\boldsymbol{E}(\mu)=\boldsymbol{E}_\theta\boldsymbol{E}(\mu \mid \theta)=\boldsymbol{E}\{\mu(\theta)\}$, the prior mean of μ , and Var(μ) is the prior variance of μ , $\boldsymbol{E}_\mu\boldsymbol{Var}(\bar{y} \mid \mu)$ is the prior expectation of the conditional variance of $\bar{y}$ given μ.

ii) The Bayes estimator of population total under the squared-error loss is given by

$$\boldsymbol{E}\{\sum_{i=1}^{N} Y_i \mid (S,Y)\}=[n\bar{y}+(N-n)\boldsymbol{E}\{Y_i \mid (S,Y)\}] \qquad (2.7)$$

<u>Example 2.1</u> Assume we have a model $Y=\theta\mathbf{1}+\varepsilon$, and $Y_1, Y_2, \ldots, Y_N$ given θ are i.i.d. with normal distribution N(θ,τ^2), and the prior distribution of θ is also a normal distribution with expectation μ and variance σ^2. Since $\boldsymbol{E}(Y)=\boldsymbol{E}(\theta\mathbf{1}_N)+\boldsymbol{E}(\varepsilon)=\mu\mathbf{1}_N$ *and* $\boldsymbol{Var}(Y)=\sigma^2\mathbf{1}\mathbf{1}'+\tau^2 I_{N\times N}$ so, the joint marginal of (Y_1, Y_2, ...,Y_N) is $N(\mu\mathbf{1}_N, \tau^2 I_{N\times N}+\sigma^2\mathbf{1}_N\mathbf{1}_N')$, and the conditional joint posterior distribution of Y_j ($j\notin s$) given

Y_i ($i \in \bar{s}$) is $N(\frac{M\mu+n\bar{y}}{M+n}\mathbf{1}_{N-n}, \tau^2(I_{(N-n)\times(N-n)}+\frac{1}{M+n}\mathbf{1}_{N-n}\mathbf{1}'_{N-n})$, where $M=\tau^2/\sigma^2$, $\bar{y}=\frac{1}{n}\sum_{i\in s} y_i$, $\mathbf{1}=(1,1,\ldots,1)'$.

The above normality result follows from the well known theorem in multivariate analysis.

<u>Theorem 2.3</u> If Y is a random vector coming from a N dimensional normal distribution with mean vector μ and variance-covariance matrix Σ, and

$$Y=\begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix}, \mu=\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $Y_{(1)}$ is a n dimensional vector and $Y_{(2)}$ is a N-n dimensional vector. Then the conditional distribution of $Y_{(2)}$ given $Y_{(1)}$ is: $N(\mu_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}(Y_{(1)}-\mu_{(1)}), \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

Since our variables are exchangeable, we can assume the distribution of our sample same as that of $Y_{(1)}$. Thus according theorem 2.3, the joint distribution of the remaining variables given $Y_{(1)}$ is a N-n dimensional normal distribution. It is not hard to get:

$$\Sigma_{11}^{-1}=\begin{pmatrix} \tau^2+\sigma^2 & \tau^2 & \ldots & \tau^2 \\ \tau^2 & \tau^2+\sigma^2 & \ldots & \tau^2 \\ \vdots & \vdots & \ldots & \vdots \\ \tau^2 & \tau^2 & \ldots & \tau^2+\sigma^2 \end{pmatrix}^{-1}$$

$$=\frac{1}{n\sigma^2+\tau^2}\begin{pmatrix}\tau^2+(n-1)\sigma^2 & -\sigma^2 & \cdots & -\sigma^2\\ -\sigma^2 & \tau^2+(n-1)\sigma^2 & \cdots & -\sigma^2\\ \vdots & \vdots & \cdots & \vdots\\ -\sigma^2 & -\sigma^2 & \cdots & \tau^2+(n-1)\sigma^2\end{pmatrix} \qquad (2.8)$$

So the mean of $\mathbf{Y}_{(2)}$ given $\mathbf{Y}_{(1)}$ is

$$\mu\mathbf{1}_{N-n}+\frac{\tau^2}{(n\sigma^2+\tau^2)\tau^2}\begin{pmatrix}1 & \cdots & 1\\ \vdots & \cdots & \vdots\\ 1 & \cdots & 1\end{pmatrix}\begin{pmatrix}\tau^2+(n-1)\sigma^2 & \cdots & \sigma^2\\ \vdots & \cdots & \vdots\\ -\sigma^2 & \cdots & \tau^2+(n-1)\sigma^2\end{pmatrix}\begin{pmatrix}Y_1-\mu\\ \vdots\\ Y_n-\mu\end{pmatrix}$$

$$=\frac{\tau^2\mu+\sigma^2\sum_{i=1}^{n}y_i}{\tau^2+n\sigma^2}\mathbf{1}_{N-n}$$

$$=\frac{M\mu+n\bar{y}}{M+n}\mathbf{1}_{N-n} \qquad (2.9)$$

Similarly, as claimed before, we can also show that the variance-covariance matrix of $\mathbf{Y}_{(2)}$ given $\mathbf{Y}_{(1)}$ is $\tau^2(I_{(N-n)\times(N-n)}+\frac{1}{M+n}\mathbf{1}_{N-n}\mathbf{1}'_{N-n})$.

Therefore, Bayes estimator of population total under the squared-error loss is:

$$\boldsymbol{E}\left(\sum_{i=1}^{N}Y_i \mid s, Y(s)\right)=\sum_{i\in s}y_i+\sum_{j\notin s}\boldsymbol{E}(Y_j \mid s, Y(s))$$

$$=\sum_{i\in s}y_i+(N-n)\left(\frac{M+n\bar{y}}{M+n}\right)$$

$$=n\bar{y}+(N-n)\left(\frac{M}{M+n}\mu+\left(1-\frac{M}{M+n}\right)\bar{y}\right) \qquad (2.10)$$

Remark: In theorems 2.1 and 2.2, we have assumed that the random variables have the exchangeable property and linearity of the posterior mean. These two conditions may not always be reasonable, but in such cases, they could be satisfied by dividing the population into k strata and the elements within each stratum could be exchangeable and linearity of the posterior stratum mean may be satisfied (see Ghosh & Lahiri, 1987).

2.1.2 Linear Bayes' Estimators and Properties

Hartigan (1969) proposed another estimation procedure called linear Bayes' method which only requires the first and second moments of the posterior to be specified. The resulting estimators have the property of minimizing posterior squared-error loss among all estimators that are linear in the data and can be thought of as approximations to posterior means.

Since a linearity of the posterior mean about data has been assumed in Ericson's model, therefore, Hartigan's estimation is a particular case of Ericson's approch. Hartigan's work is further explored by Brunk (1980), who named it as Bayesian least square.

Definition 2.4 A Bayesian least square estimator of V given U is called the linear expectation vector of V given U, where V and U are Nx1 vectors, and is written as

$$\hat{V}_{BLS}=E_L(V \mid U)=E(V)+Cov(V,U)\,VAR^{-1}(U)\,(U-E(U)) \qquad (2.11)$$

Also, The linear covariance matrix of V given U is defined as

$$Var_{BLS}(V)=Var_L(V \mid U)=E_L(V-\hat{V}_{BLS})\,(V-\hat{V}_{BLS})'$$

$$=Cov(V)-Cov(V,U)\,Var^{-1}(U)\,Cov(V,U)' \qquad (2.12)$$

The following theorem is proved in Brunk (1980).

<u>Theorem 2.4</u>

The Bayesian least squares estimator has the smallest posterior square error loss.

We now use the above to estimate the finite population mean and total.

Let $Y=(Y_1,Y_2, \ldots, Y_N)'$ be a random vector, $g(Y)=a'Y$ be a linear function of Y, Where $a'=(a_1,a_2, \ldots, a_N)$ is a constant vector. Thus the population mean μ can be written as $g(Y)$ by setting $a'=(\frac{1}{N},\frac{1}{N},\ldots,\frac{1}{N})$, and the total by setting $\mathbf{a}'=\{1,1,\ldots,1\}$. The Bayesian least square estimator of the population total $\sum_{i=1}^{N} Y_i$ and mean $\bar{Y}$ can be all written in the form of $\mathbf{a}'\hat{Y}_{BLS}$.

Assume that $Y=(Y_1,Y_2, \ldots, Y_N)$ given θ are i.i.d., come from $f(\theta, \Sigma)$, where f is an arbitray p.d.f. which possesses the first moment θ, and Σ is known, positive definite matrix of variance and covariance of Y. Also we assume $Y_1,Y_2,\ldots,Y_N$ to be exchangeable. Decomposing Y as in theorem (2.3) with possible rearrangement s.t. $Y_{(1)}=(Y_{i_1},Y_{i_2},\ldots Y_{i_n})$, and $Y_{(2)}$ contains the remaining elements of Y. We have

$$Cov(Y_{(1)})=\sum_{11}, Cov(Y_{(2)})=\sum_{22}, Cov(Y_{(1)},Y_{(2)})=\sum_{12}=\sum_{21}'$$

We also have $EY_{(1)}=\theta 1_n$, and $EY_{(2)}=\theta 1_{N-n}$.

Now, substituting Y_1 for U and Y for V in the theorem 2.3, we get the following estimator of the vector Y,

$$\hat{Y}_{BLS}=E_L(Y\mid Y_{(1)})=EY+Cov(Y,Y_{(1)})Cov^{-1}(Y_{(1)})(Y_{(1)}-EY_{(1)})$$

$$=\theta+\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\Sigma_{11}^{-1}(Y_{(1)}-\theta_{(1)})$$

$$=\begin{pmatrix}Y_{(1)}\\ \theta_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}(Y_{(1)}-\theta_{(1)})\end{pmatrix}$$

thus,

$$\hat{g}(Y)=a'\hat{Y}_{BLS}=a'_{(1)}Y_{(1)}+a'_{(2)}(\theta_{(2)}+\Sigma_{21}\Sigma_{11}^{-1}(Y_{(1)}-\theta_{(1)}))\qquad(2.13)$$

$$Var_{BLS}(Y)=Var_L(Y\mid Y_{(1)})=Cov(Y)-Cov(Y,Y_{(1)})Cov^{-1}(Y_{(1)})Cov(Y_{(1)},Y)$$

$$=\Sigma-\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\Sigma_{11}^{-1}(\Sigma_{11}\quad\Sigma_{12})\qquad(2.14)$$

$$Var_{BLS}(a'Y)=Var_L(a'Y\mid Y_{(1)})$$

$$=\Sigma-\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\ \Sigma_{21}&\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\end{pmatrix}$$

$$=\begin{pmatrix}0&0\\ 0&\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\end{pmatrix}\qquad(2.15)$$

Thus,

$$Var_{BLS}(a'y)=a'_{(2)}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})a_{(2)}\qquad(2.16)$$

where a is decomposed into $a_{(1)}$, $a_{(2)}$ similar to Y.

Now, let us assume that the prior distribution of θ is $\pi(\mu,\Omega)$, where μ is the prior mean and Ω is the prior variance, then we have

$$E(Y)=E(E(Y\mid\theta))=E(\theta)=\mu$$

$$Cov(Y)=E(Cov(Y\mid\theta))+Cov(E(Y\mid\theta))$$

$$=E(\Sigma)+Cov(\theta)$$

$$=\Sigma+\Omega \qquad (2.17)$$

and

$$Cov(Y,Y_{(1)})=E(Cov(Y,Y_{(1)}\mid\theta))+Cov(E(Y\mid\theta),E(Y_{(1)}\mid\theta))$$

$$=E\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}+Cov(\theta,\theta_{(1)})$$

$$=\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}+\begin{pmatrix}\Omega_{11}\\ \Omega_{21}\end{pmatrix}$$

$$=\begin{pmatrix}\Sigma_{11}+\Omega_{11}\\ \Sigma_{21}+\Omega_{21}\end{pmatrix} \qquad (2.18)$$

Using above equations we get the following,

$$\hat{Y}_{BLS}=E_L(Y\mid Y_{(1)})=E(Y)+Cov(Y,Y_{(1)})Cov^{-1}(Y_{(1)})(Y_{(1)}-E(Y_{(1)}))$$

$$=\mu+\begin{pmatrix}\Sigma_{11}+\Omega_{11}\\ \Sigma_{21}+\Omega_{21}\end{pmatrix}(\Sigma_{11}+\Omega_{11})^{-1}(Y_{(1)}-\mu_{(1)})$$

$$=\begin{pmatrix}\mu_{(1)}\\ \mu_{(2)}\end{pmatrix}+\begin{pmatrix}Y_{(1)}-\mu_{(1)}\\ (\Sigma_{21}+\Omega_{21})(\Sigma_{11}+\Omega_{11})^{-1}(Y_{(1)}-\mu_{(1)})\end{pmatrix}$$

$$=\begin{pmatrix}Y_{(1)}\\ \mu_{(2)}+(\Sigma_{21}+\Omega_{21})(\Sigma_{11}+\Omega_{11})^{-1}(Y_{(1)}-\mu_{(1)})\end{pmatrix}$$

We therefore have,

$$\mathbf{a}'\hat{\mathbf{Y}}=\mathbf{a}'\hat{\mathbf{Y}}_{BLS}=\mathbf{a}_{(1)}\mathbf{Y}_{(1)}+\mathbf{a}_{(2)}(\mu_{(2)}+(\Sigma_{21}+\Omega_{21})(\Sigma_{11}+\Omega_{11})^{-1}(\mathbf{Y}_{(1)}-\mu_{(1)})) \qquad (2.19)$$

$$\mathbf{Var}_{BLS}(\mathbf{Y})=\mathbf{Var}_{L}(\mathbf{Y} \mid \mathbf{Y}_{(1)})$$

$$=\begin{pmatrix} 0 & 0 \\ 0 & (\Sigma_{22}+\Omega_{22})-(\Sigma_{21}+\Omega_{21})(\Sigma_{11}+\Omega_{11})^{-1}(\Sigma_{12}+\Omega_{12}) \end{pmatrix}$$

$$\mathbf{Var}_{BLS}(\mathbf{a}'\mathbf{Y})=\mathbf{Var}_{L}(\mathbf{a}'\mathbf{Y} \mid \mathbf{Y}_{(1)})$$

$$=\mathbf{a}'_{(2)}((\Sigma_{22}+\Omega_{22})-(\Sigma_{21}+\Omega_{21})(\Sigma_{11}+\Omega_{11})^{-1}(\Sigma_{12}+\Omega_{12}))\mathbf{a}_{(2)} \qquad (2.20)$$

where $\mathbf{a}=(1, 1, \ldots, 1)'$ or $(1/N, 1/N, \ldots, 1/N)'$.

<u>Example 2.2</u> Let $\mathbf{Y}=\boldsymbol{\theta}+\boldsymbol{\varepsilon}$

Where $\boldsymbol{\theta}=\theta(1,1,\ldots,1)'$, $\mathbf{Y}=(Y_1,Y_2,\ldots,Y_N)'$, $\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N)'$ and $\boldsymbol{\theta}$, $\boldsymbol{\varepsilon}$ are independent.

Assume $\boldsymbol{\theta} \sim N(\mathbf{0},\tau^2 I_{N\times N})$, $\mathbf{Y} \mid \theta \sim N(\theta\mathbf{1}_N,\tau^2 I_{N\times N})$ and $\theta \sim N(\mu,\sigma^2)$

$$\because \ \mathbf{E}(\theta\mathbf{1}_N)=\mu\mathbf{1}_N, \text{ and } \mathbf{Var}(\theta\mathbf{1}_N)=\sigma^2\mathbf{1}_N\mathbf{1}'_N$$

$$\therefore \ \boldsymbol{\theta}=\theta\mathbf{1}_N \sim N(\mu\mathbf{1}_N,\sigma^2\mathbf{1}_N\mathbf{1}'_N)$$

$$\because \ \mathbf{EY}=\mathbf{E}\Theta=\mu\mathbf{1}_N \text{ and } \mathbf{Cov}(\mathbf{Y})=\Sigma+\Omega=\tau^2 I_{N\times N}+\sigma^2\mathbf{1}_N\mathbf{1}'_N$$

$$\therefore \ \mathbf{Y} \sim N(\mu,\tau^2 I_{N\times N}+\sigma^2\mathbf{1}_N\mathbf{1}'_N)$$

As in example 2.1, we have

$$\mathbf{Cov}(\mathbf{Y})^{-1}=(\Sigma+\Omega)^{-1}=(\tau^2 I_{N\times N}+\sigma^2\mathbf{1}_N\mathbf{1}'_N)^{-1}$$

$$=\frac{1}{(N\sigma^2+\tau^2)\tau^2}\begin{pmatrix} \tau^2+(N-1)\sigma^2 & -\sigma^2 & \cdots & -\sigma^2 \\ -\sigma^2 & \tau^2+(N-1)\sigma^2 & \cdots & -\sigma^2 \\ \vdots & \vdots & \cdots & \vdots \\ -\sigma^2 & -\sigma^2 & \cdots & \tau^2+(N-1)\sigma^2 \end{pmatrix} \qquad (2.21)$$

Therefore,

$$\mathbf{a}'\hat{\mathbf{Y}}_{BLS}=\mathbf{a}'_{(1)}\mathbf{Y}_{(1)}+\mathbf{a}'_{(2)}(\boldsymbol{\mu}_{(2)}+(\Sigma_{21}+\Omega_{21}(\Sigma_{11}+\Omega_{11})^{-1})\mathbf{Y}_{(1)}-\boldsymbol{\mu}_{(1)}))$$

$$=\mathbf{a}'_{(1)}\mathbf{Y}_{(1)}+\mathbf{a}'_{(2)}(\boldsymbol{\mu}_{(2)}+\frac{\sigma^2}{n\sigma^2+\tau^2}\mathbf{1}_{(2)}\mathbf{1}'_{(1)}(\mathbf{Y}_{(1)}-\boldsymbol{\mu}_{(1)})) \qquad (2.22)$$

By putting $\mathbf{a}=\mathbf{1}'=(1,1,\ldots,1)'$, the estimator of the population total

$$\hat{Y}_{BLS}=\boldsymbol{E}_{BLS}(\sum_{j=1}^{N}Y_j)=\sum_{j=1}^{n}y_j+\sum_{i\in s^c}\mu_i+\frac{\sum_{j=1}^{n}y_j-\sum_{j=1}^{n}\mu_j}{n\sigma^2+\tau^2}\sigma^2(N-n)$$

$$=\sum_{j=1}^{n}y_j+\mu(N-n)+\frac{\sum_{j=1}^{n}y_j-n\mu}{n\sigma^2+\tau^2}\sigma^2(N-n)$$

$$=\sum_{j=1}^{n}y_j+(N-n)\left[\frac{\tau^2\mu}{\tau^2+n\sigma^2}+\frac{\sigma^2\sum_{j=1}^{n}y_j}{\tau^2+n\sigma^2}\right]$$

$$=[n\bar{y}+(N-n)(\frac{M}{M+n}\mu+(1-\frac{M}{M+n})\bar{y})] \qquad (2.23)$$

where $M=\frac{\tau^2}{\sigma^2}$

Similarily, the estimator of the population mean μ is,

$$\hat{\mu}=\frac{1}{N}[n\bar{y}+(N-n)(\frac{M}{M+n}\mu+(1-\frac{M}{M+n})\bar{y})]$$

The above estimators are the same as those obtained by Ericson's approach, as expected.

2.2 Bayesian Regression Estimators

A well known statistical technique, regression modelling, can be used to estimate the population total and mean under the Bayesian assumptions.

Consider the finite population as a random sample coming from a infinite population called superpopulation.

Let us assume a regression model for the population characteristic as follows;

$$Y_t = X_t\beta + \varepsilon_t \qquad for \quad t=1,\ 2,\ ,\ldots,N$$

where $X_t = (x_{t,1},\ x_{t,2},\ \ldots\ ,\ x_{t,k})$ is a vector of known constants, $\beta = (\beta_1,\ \beta_2,\ \ldots,\ \beta_k)'$, and ε_t , $t=1,2,\ldots N$ are independent.

So, we can write

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon$$

where $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \ldots, \boldsymbol{X}_N')'$ *and* $\epsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$.

We now consider different important cases.

Case 1. If k=1, $x_{t,1}=1$ for all t, the model becomes the one we discussed in section 2.1.

Case 2. If k=1, $x_{t,1} \neq 1$, for some t, then the model becomes as

$$Y_t = \beta X_t + \varepsilon_t \qquad t=1,2,\ldots N$$

Let $\varepsilon \sim N(\mathbf{0}, \tau^2 I_{N\times N})$, $\beta \sim N(\mu, \sigma^2)$, $\boldsymbol{Y}|\beta \sim N(\boldsymbol{X}\beta, \tau^2 I_{N\times N})$, then since,

$$\because \quad \boldsymbol{X}\beta \sim N(\boldsymbol{X}\mu, \sigma^2 \boldsymbol{X}\boldsymbol{X}')$$

Therefore,

$$\therefore \quad \boldsymbol{Y} \sim N(\boldsymbol{X}\mu, \tau^2 I_{N\times N} + \sigma^2 \boldsymbol{X}\boldsymbol{X}') \triangleq N\left(\boldsymbol{X}\mu, \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\right)$$

where $\boldsymbol{X}\boldsymbol{X}' = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, A_{11} is the nxn, A_{22} is the (N-n)x(N-n) and $A_{12}=A_{21}'$ is the (N-n)xn submatrix.

Assuming $(\tau^2 I_{N\times N} + \sigma^2 \boldsymbol{X}\boldsymbol{X}')^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, we get

$$\mathbf{a}'\hat{\boldsymbol{Y}}_{BLS} = \mathbf{a}'_{(1)}\boldsymbol{Y}_{(1)} + \mathbf{a}'_{(2)}\left(\boldsymbol{X}_{(2)}\mu + A_{21}A_{11}^{-1}(\boldsymbol{Y}_{(1)} - \boldsymbol{X}_{(1)}\mu)\right)$$

Letting a'=(1, 1, ... ,1), we get the estimator of the population total as

$$\hat{Y}_{BLS} = \sum_{i=1}^{n} y_i + \mu \sum_{i\in s^c} x_i + \sum_{i\in s^c} x_i (x_1, x_2, \ldots, x_n)\left[A_{11}^{-1}(\boldsymbol{Y}_{(1)} - \boldsymbol{X}_{(2)})\right] \qquad (2.27)$$

<u>Case 3.</u> Consider X to consist of k columns of independent variables, i.e. $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_k)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$

and assume that

$$\varepsilon \sim N(0, \tau^2 I_{N\times N}), \beta \sim N(\mu, \Omega)$$

$$\boldsymbol{Y} | \beta \sim N(\boldsymbol{X}\beta, \tau^2 I_{N\times N}), \ \boldsymbol{X}\beta \sim N(\boldsymbol{X}\mu, \boldsymbol{X}\Omega\boldsymbol{X}')$$

then

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\mu, \tau^2 I_{N\times N} + \boldsymbol{X}\Omega\boldsymbol{X}')$$

and in this case, the Bayes' estimator of a'Y is given by

$$\mathbf{a}'\hat{\mathbf{Y}}_{BLS}=\mathbf{a}'_{(2)}(\mathbf{X}_{(2)}\mu+A_{21}A_{11}^{-1}(\mathbf{Y}_{(1)}-\mathbf{X}_{(1)}\mu)) \qquad (2.25)$$

Chapter 3

Estimators Under Some Specified Distributions

3.1 Introduction

In this chapter, we will deal with the estimators of population total and mean under some superpopulation models with several different prior distributional assumptions.

We still assume that our population is exchangeable, and for any given vector a, the estimator of $\mathbf{a}'\boldsymbol{Y}=\mathbf{a}_s'\boldsymbol{Y}_s+\mathbf{a}_{s^c}'\boldsymbol{Y}_{s^c}$ is $\mathbf{a}'\hat{\boldsymbol{Y}}$, Y_s being the vector containing sample observations and $\boldsymbol{Y}_{s^c}$ being similarily defined. Since treat all Y_i in Y as parameters, so Bayes' estimator of Y

$$\hat{\boldsymbol{Y}}=\boldsymbol{E}(\boldsymbol{Y}\mid s,\boldsymbol{y}_s) \qquad (3.1)$$

and therefore,

$$\mathbf{a}'\hat{\boldsymbol{Y}}=\mathbf{a}_s'\boldsymbol{y}_s+\mathbf{a}_{s^c}'\begin{pmatrix}\boldsymbol{E}(Y_{j_1}\mid s,y_s)\\ \vdots \\ \boldsymbol{E}(Y_{j_{N-n}}\mid s,y_s)\end{pmatrix}, \qquad (3.2)$$

where $(j_1, j_2, \ldots, j_{N-n})$ are the indices in the population U for s^c.

Thus, the estimator of the total is

$$\sum_{i=1}^{N}\hat{Y}_{aw}=\hat{Y}=\sum_{i=1}^{n}y_i+(N-n)\boldsymbol{E}(Y_j\mid s,y_s) \qquad (3.3)$$

and that of the mean is

$$\frac{1}{N}\sum_{i=1}^{N}\hat{Y}_i=\frac{1}{N}\sum_{i=1}^{n}y_i+(1-f)\,E(Y_j\,|s,y_s) \qquad (3.4)$$

where f=n/N.

In the next section we consider different superpopulation models and derive the form of (3.4).

3.2 Simple Mean Model

Let our superpopulation be

$$Y_j=\theta+\varepsilon_j \qquad j=1,\ 2,\ \ldots,\ N \qquad (3.5)$$

so for $j\notin s$, we have

$$E(Y_j\mid s,y_s)=E(\theta\mid s,y_s)+E(\varepsilon_j\mid s,y_s)$$

$$=E(\theta\mid s,y_s) \qquad (3.6)$$

thus

$$\sum_{i=1}^{N}\hat{Y}_i=\hat{Y}=\sum_{i=1}^{n}y_i+(N-n)\,E(\theta\mid s,y_s) \qquad (3.7)$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\hat{Y}_i=\frac{1}{N}\sum_{i=1}^{n}y_i+(1-f)\,E(\theta\mid s,y_s)\,. \qquad (3.8)$$

Now we would like to work with some specified distributions and priors. Normally, if we do not have any information about the parameters of our population, we will assume the prior to be an improper prior as metioned before. But, in many cases, it is not reasonable to assume the parameter range to be (-∞, ∞), like wages or ages which are always positive, so we may assume a prior to be a uniform dirtribution U(A,B), for known constants A and B or some other distribution concentrated on the range (0, ∞).

3.2.1. Estimators Under a U(A,B) Prior

Assume $\boldsymbol{Y}|\theta \epsilon N(\theta \mathbf{1}, \sigma^2 I_{N\times N})$ *and* $\theta \epsilon U(A,B)$ where U(A,B) is an uniform distribution over the interval (A,B) and A, B, σ^2 all known, (A,B > 0). In this case we have,

$$f(\theta \mid y_s) = \frac{f(y_s,\theta)}{\int_A^B f(y_s,\theta)\, d\theta}$$

$$= \frac{f(y_s|\theta)}{\int_A^B f(y_s|\theta)\, d\theta}$$

$$= \frac{\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\}}{\int_A^B \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\}\, d\theta}$$

$$= \frac{\exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\, d\theta} \quad \text{for } A\le\theta\le B \tag{3.9}$$

It is clear that the posterior distribution of θ is the truncated normal distribution over the range (A,B) with mean $\bar{y}$ and variance $\frac{\sigma^2}{n}$

Further, we have

$$\boldsymbol{E}(\theta\,|y_s)=\bar{y}_s+\frac{\int_A^B(\theta-\bar{y}_s)\,e^{-\frac{n}{2\sigma^2}(\theta-\bar{y})^2}d\theta}{\int_A^B e^{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2}d\theta}$$

$$=\bar{y}_s+\frac{\frac{\sigma}{\sqrt{n}}[e^{-\frac{n}{2\sigma^2}(A-\bar{y}_s)^2}-e^{-\frac{n}{2\sigma^2}(B-\bar{y}_s)^2}]}{\sqrt{2\pi}\,[\Phi(\frac{B-\bar{y}_s}{\sigma}\sqrt{n})-\Phi(\frac{A-\bar{y}_s}{\sigma}\sqrt{n})]}, \qquad (3.10)$$

where the use has been made of the following theorem, $\Phi(x)$ being the distribution of standard normal.

<u>Theorem 3.1</u>

A random variable X has a Truncated Normal distribution TN(0,1;A,B), the p.d.f. is

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}[\frac{1}{\sqrt{2\pi}}\int_A^B e^{-\frac{1}{2}t^2}dt]^{-1}$$

$$\hat{=}Z(x)\,[\Phi(B)-\Phi(A)]^{-1} \qquad \textit{for } A\leq x\leq B \qquad (3.11)$$

then the expectation of X is given by

$$\boldsymbol{E}(X)=\frac{Z(A)-Z(B)}{\Phi(B)-\Phi(A)} \qquad (3.12)$$

and

$$\boldsymbol{Var}(X)=1+\frac{A\boldsymbol{Z}(A)-B\boldsymbol{Z}(B)}{\Phi(B)-\Phi(A)}-\{\frac{\boldsymbol{Z}(A)-\boldsymbol{Z}(B)}{\Phi(B)-\Phi(A)}\}^2 \qquad (3.13)$$

or

$$\boldsymbol{E}X^2=\boldsymbol{Var}X+(\boldsymbol{E}X)^2=1+\frac{AZ(A)-BZ(B))}{\Phi(B)-\Phi(A)} \qquad (3.14)$$

where $Z(x)=(\sqrt{2\pi})^{-1}e^{-\frac{1}{2}x^2}$ *and* $\Phi(x)=(\sqrt{2\pi})^{-1}\int_{-\infty}^{x}e^{-\frac{1}{2}x^2}dx$.

Hence,

$$\hat{Y}=N\bar{y}_s+(N-n)\frac{\sigma}{\sqrt{2\pi n}}\frac{e^{-\frac{n}{2\sigma^2}(A-\bar{y}_s)^2}-e^{-\frac{n}{2\sigma^2}(B-\bar{y}_s)^2}}{\Phi(\frac{B-\bar{y}_s}{\sigma}\sqrt{n})-\Phi(\frac{A-\bar{y}_s}{\sigma}\sqrt{n})} \tag{3.15}$$

and

$$\bar{Y}=\bar{y}_s+(1-f)\frac{\sigma}{\sqrt{2\pi n}}\frac{e^{-\frac{n}{2\sigma^2}(A-\bar{y}_s)^2}-e^{-\frac{n}{2\sigma^2}(B-\bar{y}_s)^2}}{\Phi(\frac{B-\bar{y}_s}{\sigma}\sqrt{n})-\Phi(\frac{A-\bar{y}_s}{\sigma}\sqrt{n})} \tag{3.16}$$

3.2.2. Estimators under a $B(\alpha,\beta;A,B)$ Prior

Assume $\mathbf{Y}|\theta\sim N(\theta\mathbf{1},\sigma^2 I_{N\times N})$, $\theta\sim\mathrm{B}(\alpha,\beta;A,B)$, A,B,σ^2 all known, where $\mathrm{B}(\alpha,\beta;A,B)$ denotes a distribution with pdf $f_\theta(\theta)\propto(\theta-A)^{\alpha-1}(B-\theta)^{\beta-1}$. Thus U(A,B) is a particular case of this prior. Now, we have the posterior distribution of θ given by

$$f(\theta \mid s,y_s)=$$

$$=\frac{\frac{1}{\sqrt{2\pi\sigma}^{n}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)(B-A)^{\alpha+\beta-1}}(\theta-A)^{\alpha-1}(\beta-\theta)^{\beta-1}}{\frac{1}{\sqrt{2\pi\sigma}^{n}}\int_{A}^{B}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)(B-A)^{\alpha+\beta-1}}(\theta-A)^{\alpha-1}(\beta-\theta)^{\beta-1}d\theta}$$

$$= \frac{\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\}(\theta-A)^{\alpha-1}(\beta-\theta)^{\beta-1}}{\int_A^B \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\}(\theta-A)^{\alpha-1}(\beta-\theta)^{\beta-1}d\theta} \qquad (3.17)$$

Thus we have

$$E(\theta\,|y_s) = \frac{\int_A^B \theta\exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)^{\alpha-1}(B-\theta)^{\beta-1}d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)^{\alpha-1}(B-\theta)^{\beta-1}d\theta}$$

$$=A+\frac{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)^{\alpha}(B-\theta)^{\beta-1}d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)^{\alpha-1}(B-\theta)^{\beta-1}d\theta} \qquad (3.18)$$

It is hard to simplify the above integral any further for arbitrary α and β , though it can be computed numerically. Below we simplify it for some specific cases.

<u>Case 1. $\alpha=2, \beta=1$</u>

In this case (3.18) becomes

$$E(\theta\,|y_s) = A+\frac{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)^2 d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)\,d\theta}$$

$$=A+\frac{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)^2 d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$+2(\bar{y}_s-A)\frac{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$+\frac{(\bar{y}_s-A)^2\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)\}^2 d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}. \qquad (3.19)$$

Now, we can write

$$\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)^2 d\theta=(\frac{\sigma}{\sqrt{n}})^3\int_{A^*}^{B^*} e^{-\frac{1}{2}\xi^2}\xi^2 d\xi$$

$$=(\frac{\sigma}{\sqrt{n}})^3\sqrt{2\pi}\,[\phi(B^*)-\phi(A^*)]\,\boldsymbol{E}\xi^2$$

$$=(\frac{\sigma}{\sqrt{n}})^2\sqrt{2\pi}\,[\phi(B^*)-\phi(A^*)]\times[1+\frac{A^*z(A^*)-B^*Z(B^*)}{\phi(B^*)-\phi(A^*)}] \qquad (3.20)$$

and

$$\int_A^B (\theta-\bar{y}_s)\exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta=\frac{\sigma^2}{n}\int_{A^*}^{B^*}\xi e^{-\frac{1}{2}\xi^2}d\xi$$

$$=\frac{\sigma^2}{n}[\phi(B^*)-\phi(A^*)]\,\boldsymbol{E}\xi$$

$$=\frac{\sigma^2}{n}\sqrt{2\pi}\times[\boldsymbol{Z}(A^*)-\boldsymbol{Z}(B^*)] \qquad (3.21)$$

where $A^*=\dfrac{A-\bar{y}_s}{\sigma/\sqrt{n}}$, $B^*=\dfrac{B-\bar{y}_s}{\sigma/\sqrt{n}}$ and $\xi=\dfrac{\theta-\bar{y}_s}{\sigma/\sqrt{n}}$.

And therefore using (3.20) and (3.21), (3.19) becomes

$$\boldsymbol{E}(\theta\,|\boldsymbol{y}_s)=A+\frac{\dfrac{\sigma}{\sqrt{n}}\left[1+\dfrac{A^*\boldsymbol{Z}(A^*)-B^*\boldsymbol{Z}(B^*)}{\phi(B^*)-\phi(A^*)}\right]}{\left[\dfrac{\boldsymbol{Z}(A^*)-\boldsymbol{Z}(B^*)}{\phi(B^*)-\phi(A^*)}\right]-A^*}$$

$$+2\frac{\sigma}{\sqrt{n}}A^*\frac{1}{1-A^*\dfrac{\phi(B^*)-\phi(A^*)}{\boldsymbol{Z}(A^*)-\boldsymbol{Z}(B^*)}}$$

$$+\frac{\sigma}{\sqrt{n}}A^{*2}\frac{1}{\dfrac{\boldsymbol{Z}(A^*)-\boldsymbol{Z}(B^*)}{\phi(B^*)-\phi(A^*)}-A^*}. \qquad (3.22)$$

Case 2. $\alpha=3,\beta=1$

In this case, (3.18) becomes

$$E(\theta\,|\bar{y}_s)=A+\frac{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)^3 d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-A)\,d\theta} \qquad (3.23)$$

$$=A+\frac{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)^3 d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$+\frac{3(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)^2 d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$+\frac{3(\bar{y}_s-A)^2\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$+\frac{(\bar{y}_s-A)^3\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$=A+\frac{[\frac{\sigma^2}{n}(A-\bar{y}_s)^2+\frac{2\sigma^4}{n^2}]\exp\{-\frac{n}{2\sigma^2}(A-\bar{y}_s)^2\}}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$-\frac{[\frac{\sigma^2}{n}(B-\bar{y}_s)^2+\frac{2\sigma^4}{n^2}]\exp\{-\frac{n}{2\sigma^2}(B-\bar{y}_s)^2\}}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$+3\frac{\sigma^2}{n}A^{*2}\frac{1}{1-A^*\frac{\phi(B^*)-\phi(A^*)}{Z(A^*)-Z(B^*)}}$$

$$+3\frac{\sigma^2}{n}A^{*2}\frac{1}{1-A^*\frac{\phi(B^*)-\phi(A^*)}{Z(A^*)-Z(B^*)}}$$

$$+\frac{\sigma^2}{n}A^{*3}\frac{1}{\frac{Z(A^*)-Z(B^*)}{\phi(B^*)-\phi(A^*)}-A^*}. \qquad (3.24)$$

Since,

$$\frac{\frac{\sigma^2}{n}\{[(A-\bar{y}_s)^2+\frac{2\sigma^2}{n}]\exp\{-\frac{n}{2\sigma^2}(A-\bar{y}_s)^2\}-[(B-\bar{y}_s)^2+\frac{2\sigma^2}{n}]\exp\{-\frac{n}{2\sigma^2}(B-\bar{y}_s)^2\}\}}{\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}(\theta-\bar{y}_s)\,d\theta+(\bar{y}_s-A)\int_A^B \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{y}_s)^2\}\,d\theta}$$

$$=\frac{\frac{\sigma^3}{n^{\frac{3}{2}}}(A^{*2}+2)Z(A^*)-\frac{\sigma^3}{n^{\frac{3}{2}}}(B^{*2}+2)Z(B^*)}{\frac{\sigma^2}{n}\int_{A^*}^{B^*}\xi e^{-\frac{1}{2}\xi^2}d\xi-\frac{\sigma^2}{n}A^*\int_{A^*}^{B^*}e^{-\frac{1}{2}\xi^2}d\xi}$$

$$=\frac{\sigma}{\sqrt{2\pi n}}\frac{(A^{*2}+2)Z(A^*)-(B^{*2}+2)Z(B^*)}{Z(A^*)-Z(B^*)+\phi(A^*)-\phi(B^*)} \qquad (3.25)$$

(3.24) simplifies to

$$E(\theta\,|y_s) = A + \frac{\sigma}{\sqrt{2\pi n}} \frac{(A^{*2}+2)Z(A^*)-(B^{*2}+2)Z(B^*)}{Z(A^*)-Z(B^*)+\Phi(A^*)-\Phi(B^*)}$$

$$9+3\frac{\sigma^2}{n}A^*\frac{[1+\dfrac{A^*Z(A^*)-B^*Z(B^*)}{\Phi(B^*)-\Phi(A^*)}]}{[\dfrac{Z(A^*)-Z(B^*)}{\Phi(B^*)-\Phi(A^*)}]-A^*}. \qquad (3.26)$$

3.2.3. Estimators Based on a Truncated Normal Likelihood with Uniform Prior

Assume $Y_i|\Theta \sim TN(\Theta,\tau^2;\ A,B)$ and $\Theta \sim U(A,B)$, where A, B and τ are known.

In this case we have

$$f(\theta\mid y_s) = \frac{\exp\{-\dfrac{1}{2\tau^2}\sum_{i=1}^{n}(y_i-\partial)^2\}}{\int_A^B \exp\{-\dfrac{1}{2\tau^2}\sum_{i=1}^{n}(y_i-\theta)^2\}\,d\theta}$$

$$=\frac{\exp\{-\dfrac{n}{2\tau^2}(\theta-\bar{y}_s)^2\}\,[\Phi(\dfrac{B-\theta}{\tau})-\Phi(\dfrac{A-\theta}{\tau})]^{-1}}{\int_A^B \exp\{-\dfrac{n}{2\tau^2}(\theta-\bar{y}_s)^2\}\,[\Phi(\dfrac{B-\theta}{\tau^2})-\Phi(\dfrac{A-\theta}{\tau})]^{-1}d\theta} \qquad (3.27)$$

and therefore,

$$E(\theta\,|y_s) = \bar{y}_s + \frac{\int_A^B (\theta-\bar{y}_s)\exp\{-\dfrac{n}{2\tau^2}(\theta-\bar{y}_s)^2\}\,[\Phi(\dfrac{B-\theta}{\tau})-\Phi(\dfrac{A-\theta}{\tau})]^{-n}d\theta}{\int_A^B \exp\{-\dfrac{n}{2\tau^2}(\theta-\bar{y}_s)^2\}\,[\Phi(\dfrac{B-\theta}{\tau})-\Phi(\dfrac{A-\theta}{\tau})]^{-n}d\theta}.$$

Hence,

$$\hat{Y}=N\bar{y}_s+(N-n)\frac{\int_A^B(\theta-\bar{y}_s)\exp\{-\frac{n}{2\tau^2}(\theta-\bar{y}_s)^2\}\,[\phi(\frac{B-\theta}{\tau})-\phi(\frac{A-\theta}{\tau})]^{-n}d\theta}{\int_A^B\exp\{-\frac{n}{2\tau^2}(\theta-\bar{y}_s)^2\}\,[\phi(\frac{B-\theta}{\tau})-\phi(\frac{A-\theta}{\tau})]^{-n}d\theta}$$

It is difficult to find a simple expression for the complicated integral in (3.28). But we can get a numerical result for any given parameters A, B, N, τ^2 and sample y_1, y_2, ..., y_n. A FORTRAN program is presented in Appendix I, and table 3.1 shows the results under some choises of A, B, τ^2 and y_1, y_2, ..., y_n.

Table 3.1

A	B	N	n	τ^2	$\bar{y}_s$	$E(\theta \mid \bar{y}_s)$	$\hat{Y}$
0	1	1000	100	0.5	0.5	0.49	0.491
0	2	1000	15	0.5	0.2	0.216	0.215
0	2	1000	15	0.5	1.8	1.78	1.778
0	2	1000	10	0.5	0.2	0.231	0.228
0	2	1000	10	0.5	1.8	1.76	1.764
0	20	1000	15	0.5	19.0	19.0	19.0
0	20	1000	15	0.5	1.0	0.999	0.999
0	20	1000	15	0.5	0.2	0.212	0.211
0	20	1000	15	10.0	1.0	2.457	2.311
0	20	1000	30	10.0	1.0	1.875	1.788

From Table 3.1, we derive the following conclusions for large and small values of τ.

a) Small τ

i) If n is larger than 15 and the average of y's is closed to the center of the interval [A, B], then the Bayes' stimators of the population total and the mean are almost the same as their classical counterparts.

ii) If n is less than 15 or the average of y's is far away from the center of the interval [A, B], then the Bayes' estimators of the population total and the mean may differ with their classical counterparts.

So, adopting Bayes' estimators is suggested when we meet a situation like case ii).

b) Large τ

When τ is greater than 3, then the Bayes estimators differ significantly with their classical counterparts. In this case, adoption of Bayes' technique may be beneficial.

3.2.4. The Inverse Gaussian Distribution As a Super Population Model

The name "inverse Gaussian " was applied to a certain class of distributions by Tweedie(1957). Since then many attractive properties of this distribution have been developed in recent years.[Banerjee and Bhattacharyya (1979), Chaubey, Nebebe and Chen (1990)]. In this section, we use this distribution in a superpopulation context.

Assuming the distribution of Y given $\theta=(\mu,\lambda)$ is the inverse Gaussian, $IG(\mu,\lambda)$, the prior distribution of μ given λ is uniform and the marginal distribution of λ is proportional to λ^{-1}, we may write a model for Y_i's as

$$Y_i=\mu+\varepsilon_i, \qquad i=1,2,...N, \qquad (3.29)$$

where μ is a parameter and $\varepsilon_i \sim IG(1,\lambda)$ *i.i.d.* . The distribution IG(1, λ) and some of its properties are given in the following theorem.

Theorem 3.2

We write $Y \sim IG(\mu,\lambda)$ if the pdf of Y is diven by

$$f(y \mid \mu,\lambda)=(\frac{\mu}{2\pi\lambda^2})^{\frac{1}{2}} y^{-\frac{3}{2}} e^{-\frac{(y-\mu)^2}{2\lambda^2\mu y}}, \qquad \text{where } y>0,\ \mu>0,\ \lambda>0. \qquad (3.30)$$

the mean and the variance of Y given $\theta=(\mu,\lambda)$ are given by,

$$\boldsymbol{E}(Y \mid \theta)=\mu \text{ and } \boldsymbol{Var}(Y \mid \theta)=\frac{\mu^2}{\lambda}. \qquad (3.31)$$

Since the likelihood function of Y in the model (3.30) is

$$l(\mu,\lambda \mid y) \propto \exp\{-\frac{nu}{2}[1+\frac{\bar{y}}{u}(\mu-\frac{1}{\bar{y}})^2]\lambda\}\lambda^{\frac{n}{2}}$$

where $\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{y}_r=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i}$ *and* $u=\bar{y}_r-\frac{1}{\bar{y}}$ and the posterior of

(μ,λ) is

$$f(\mu,\lambda\,|\boldsymbol{y}) \propto \exp\{-\frac{\nu}{2}[1+\frac{\bar{y}}{u}(\mu-\frac{1}{\bar{y}})^2]\lambda\}\lambda^{\frac{n}{2}-1}. \qquad (3.32)$$

Hence, the posterior of μ given y is

$$f(\mu\,|\boldsymbol{y}\} = \int_0^\infty f(\mu,\lambda|\boldsymbol{y})\,d\lambda$$

$$\propto \frac{1}{(\frac{\nu}{2})^{\frac{n}{2}}}[1+\frac{\bar{y}}{u}(\mu-\frac{1}{\bar{y}})^2]^{-\frac{n}{2}} \qquad (3.33)$$

where $\nu=n-1$.

If we let $\xi=(\frac{u\bar{y}}{\nu})^{-\frac{1}{2}}$ and $q=(\xi\bar{y})^{-1}$, then the marginal posterior distribution of μ can be written as

$$f(\mu\,|\boldsymbol{y}) \propto (\frac{\nu}{2})^{-\frac{n}{2}}[1+\frac{1}{\nu q^2}(\mu-\frac{1}{\bar{y}})^2]^{-\frac{\nu+1}{2}} \qquad 0<\mu<\infty. \qquad (3.34)$$

This is a left truncated t-distribution with ν degree of freedom, location parameter $\frac{1}{\bar{y}}$, and scale parameter q. The point of truncation being at zero.

Further, if we let $\mu^*=(\mu-\frac{1}{\bar{y}})/q$, then its posterior distribution is a left truncated t-distribution with ν degrees of freedom, truncated at $-\xi$.

So, the estimator of population mean is

Now we consider intruducing an auxiliary variable in

$$\bar{Y}=f\bar{y}+(1-f)E(\mu \mid s,y_s) \qquad (3.35)$$

the estimation procedure.

3.3 The Univariate Regression Model (URM)

In many cases, we would like to introduce an auxiliary variate X_i, correlated with Y_i (for i=1, 2, ..., N), to estimate the population total or the population mean. The population total X of the X_i must be known. Normally X_i is the Y_i at one previous time when a complete census was taken.

If there is an approximate linear relationship between X_i and Y_i, we can consider an univariate linear regression model as

$$Y=X\theta+\varepsilon \qquad (3.36)$$

where $Y=(Y_1, Y_2, ..., Y_N)'$, $X=(X_1, X_2, ..., X_N)'$, and the distribution of error term and prior distribution of parameter θ are $\varepsilon \sim f_N(0,\Sigma)$, $\theta \sim g(\mu,\sigma^2)$, θ and ε being independent and f_N and g denote some general distributions.

So, for any $j \notin s$, we have

$$E(Y_j \mid s,y_s)=X_j E(\theta \mid s,y_s)$$

From sector 3.1, we can also get the estimators for the population mean and total as follow:

$$\hat{Y}=\sum_{i=1}^{n} y_i+\sum_{j=1}^{N-n} X_j E(\theta \mid s,y_s)$$

$$\bar{y}=\frac{1}{N}\sum_{i=1}^{n}y_i+\frac{1}{N}\sum_{j=1}^{N-n}X_j\mathbf{E}(\theta \mid s,y_s)$$

$$=f\bar{y}_s+(1-f)\bar{X}_{s^c}\ \mathbf{E}(\theta \mid s,y_s)$$

$$=f\bar{y}_s+(\bar{X}-f\bar{x}_s)\mathbf{E}(\theta \mid s,y_s) \qquad (3.38)$$

where $\bar{X}=\frac{1}{N}\sum_{i=1}^{N}X_i,\ f=\frac{n}{N}$.

3.3.1. An Example for URM model

Let us assume our URM to be described as;

$$\mathbf{Y}\mid\theta\sim N(\mathbf{X}\theta,\sigma^2 I_{N\times N}),\ \ and\ \theta\sim U(A,B)$$

then

$$f(\theta \mid y_1,\ y_2,\ \ldots,\ y_n)=\frac{f(y_1,\ y_2,\ \ldots\ y_n,\theta)}{\int_A^B f(y_1,\ y_2,\ \ldots\ y_n,\theta)\,d\theta}$$

$$=\frac{\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-x_i\theta)^2}{\int_A^B\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-x_i\theta)^2\}\,d\theta}$$

$$=\frac{\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^2-2x_iy_i\theta+x_i^2\theta^2)\}}{\int_A^B\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^2-2x_iy_i\theta+x_i^2\theta^2)\}\,d\theta}$$

$$= \frac{\exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} x_i^2\theta^2 - 2\sum_{i=1}^{n} x_i y_i \theta)\}}{\int_A^B \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} x_i^2\theta^2 - 2\sum_{i=1}^{n} x_i y_i \theta)\}\, d\theta}$$

$$= \frac{\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 (\theta - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x^2})^2\}}{\int_A^B \exp\{-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2}(\theta - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2})^2\}\, d\theta} \tag{3.39}$$

So, the posterior distribution of θ given y_1, y_2, ..., y_n is a truncated normal with mean M^*, and variance V^*. Where

$$M^* \triangleq (\theta \mid y_1, y_2, \ldots y_n) = C + \frac{Z(\frac{A-C}{\sigma}) - Z(\frac{B-C}{\sigma})}{\phi(\frac{B-C}{\sigma}) - \phi(\frac{A-C}{\sigma})} \tag{3.40}$$

and

$$V^* \triangleq \mathbf{Var}(\theta \mid y_1, y_2, \ldots, y_n) = \sigma^2 [1 + \frac{(\frac{A-C}{\sigma}) Z(\frac{A-C}{\sigma}) - (\frac{B-C}{\sigma}) Z(\frac{B-C}{\sigma})}{\phi(\frac{B-C}{\sigma}) - \phi(\frac{A-C}{\sigma})}$$

$$- \left(\frac{Z(\frac{A-C}{\sigma}) - Z(\frac{B-C}{\sigma})}{\phi(\frac{B-C}{\sigma}) - \phi(\frac{A-C}{\sigma})} \right)^2] \tag{3.41}$$

$$C = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}, \quad \mathbf{x}' = (x_1, x_2, \ldots, x_n)' \text{ and } \mathbf{y}' = (y_1, y_2, \ldots, y_n)'.$$

Finally, we have the estimator of the population total and the population mean for this model as;

$$\hat{Y}=y_s+(X-x_s)M^* \qquad (3.42)$$

$$\hat{\bar{Y}}=f\bar{y}_s+(\bar{X}-f\bar{x}_s)M^* \qquad (3.43)$$

3.4. Multivariate Regession Model (MRM)

A multivariate regression model can be obtained by extending URM to p auxiliary variables as follows;

$$\mathbf{Y}=\mathbf{X}\theta+\varepsilon \qquad (3..44)$$

where

$\mathbf{X}=(\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_p)$, $\mathbf{X}_i=(X_{i1},X_{i2},\ldots,X_{iN})'$ and $\theta=(\theta_1,\theta_2,\ldots,\theta_p)'$

and $\varepsilon|\theta\sim f_N(\mathbf{0}_N,\Sigma),\theta\sim g_p(\mu_p,\Omega)$, with obvious extension of previous notations.

For any $j\notin \mathbf{s}$, we have

$$\mathbf{E}(Y_j \mid s,y_s)=\sum_{i=1}^{p} X_{ji}\mathbf{E}(\theta_i \mid s,y_s)$$

and

$$\hat{Y}=\sum_{i=1}^{n} y_i+\sum_{j=1}^{N-n}\sum_{i=1}^{p} X_{ji}\mathbf{E}(\theta_i \mid s,y_s)$$

$$=\sum_{i=1}^{n} y_i+\sum_{i=1}^{p} X_{i,s^c}\,\mathbf{E}(\theta_i \mid s,y_s)$$

$$=\sum_{i=1}^{n} y_i+\sum_{i=1}^{p} (X_i-x_{i,s})\mathbf{E}(\theta_i \mid s,y_s) \qquad (3.45)$$

also

$$\hat{\bar{y}}=f\bar{y}_s+\sum_{i=1}^{p}\boldsymbol{E}(\theta_i \mid s,y_s)(\bar{X}_i-f\bar{x}_{i,s}) \qquad (3.46)$$

3.4.1. An Example for MRM

Let us assume that

$$\boldsymbol{Y}\mid\boldsymbol{\theta}\sim N_N(\boldsymbol{X\theta},\sigma^2 I_{N\times N}),\boldsymbol{\theta}\sim U(A,B) \text{ or } \theta_i\sim U(A_i,B_i) \text{ (for } i=1,2,\ldots,p)$$

In this case,

$$\boldsymbol{E}(\theta_i \mid s,y_s)=\int_{A_i}^{B_i}\theta_i f(\theta_i \mid s,y_s)\,d\theta_i$$

$$=\int_{A_i}^{B_i}\theta_i\int_{A_1}^{B_1}\cdots\int_{A_{i-1}}^{B_{i-1}}\int_{A_{i+1}}^{B_{i+1}}\cdots\int_{A_p}^{B_p}f(\theta_1,\theta_2,\ldots,\theta_p|y_1,y_2,\ldots,y_n)\,d\theta_1 d\theta_2\ldots d\theta_p$$

$$=\frac{\int_{A_i}^{B_i}\theta_i\int_{A_1}^{B_1}\cdots\int_{A_{i-1}}^{B_{i-1}}\int_{A_{i+1}}^{B_{i+1}}\cdots\int_{A_p}^{B_p}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}x_{ji}\theta_j)^2\}\,d\theta_1 d\theta_2\ldots d\theta_p}{\int_{A_1}^{B_1}\int_{A_2}^{B_2}\cdots\int_{A_p}^{B_p}\exp\{\frac{1}{2\sigma^2}[\sum_{i=1}^{n}(y_i-\sum_{j=1}^{p}x_{ji}\theta_j)^2\}\,d\theta_1 d\theta_2\ldots d\theta_p}$$

$$=\frac{\int_{A_i}^{B_i}\theta_i\int_{A_1}^{B_1}\cdots\int_{A_{i-1}}^{B_{i-1}}\int_{A_{i+1}}^{B_{i+1}}\cdots\int_{A_p}^{B_p}\exp\{-\frac{1}{2\sigma^2}[\sum_{i=1}^{n}y_i^2-2\sum_{i=1}^{n}\sum_{j=1}^{p}y_i x_{ji}\theta_j+\sum_{i=1}^{n}(\sum_{j=1}^{p}x_{ji}\theta_j)^2]\}\,d\boldsymbol{\theta}}{\int_{A_1}^{B_1}\int_{A_2}^{B_2}\cdots\int_{A_p}^{B_p}\exp\{\frac{1}{2\sigma^2}[\sum_{i=1}^{n}y_i^2-2\sum_{i=1}^{n}\sum_{j=1}^{p}y_i x_{ji}\theta_j+\sum_{i=1}^{n}(\sum_{j=1}^{p}x_{ji}\theta_j)^2]\}\,d\boldsymbol{\theta}}$$

$$
= \frac{\int_{A_i}^{B_i} \theta_i \exp\{\frac{1}{2\sigma^2}[-2\theta_i \sum_{k=1}^{n} y_k x_{ki} + \theta_i^2 \sum_{k=1}^{n} x_{ki}^2]\}\, d\theta_i}{\int_{A_i}^{B_i} \exp\{-\frac{1}{2\sigma^2}[-2\theta_i \sum_{k=1}^{n} y_k x_{ki} + \theta_i^2 \sum_{k=1}^{n} x_{ki}^2]\}\, d\theta_i}
$$

$$
\times \frac{\int_{A_1}^{B_1} \cdots \int_{A_{i-1}}^{B_{i-1}} \int_{A_{i+1}}^{B_{i+1}} \cdots \int_{A_p}^{B_p} \exp\{-\frac{1}{2\sigma^2}[2\sum_{t=1}^{p} (\sum_{s=1}^{n} x_{st} x_{si}) \theta_t \theta_i]\}\, d\theta_1 d\theta_2 \ldots d\theta_p}{\int_{A_1}^{B_1} \cdots \int_{A_{i-1}}^{B_{i-1}} \int_{A_{i+1}}^{B_{i+1}} \cdots \int_{A_p}^{B_p} \exp\{\frac{1}{2\sigma^2}[2\sum_{t=1}^{p} (\sum_{s=1}^{n} x_{st} x_{si}) \theta_t \theta_i]\}\, d\theta_1 d\theta_2 \ldots d\theta_p}
$$

where $d\boldsymbol{\theta}_{(-i)} = d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_p$. And therefore,

$$
\boldsymbol{E}(\theta_i \mid s, y_s) =
$$

$$
= \frac{\int_{A_i}^{B_i} \frac{1}{\theta_i^p} \exp\{\frac{1}{2\sigma^2}[-2\theta_i \sum_{k=1}^{n} y_k x_{ki} + \theta_i^2 \sum_{k=1}^{n} x_{ki}^2]\} \prod_{k=1}^{p-1} \frac{1}{a_{k,i}} [e^{-a_{k,i} B_k \theta_i} - e^{-a_{k,i} A_k \theta_i}]\, d\theta_i}{\int_{A_i}^{B_i} \frac{1}{\theta_i^{p-1}} \exp\{-\frac{1}{2\sigma^2}[-2\theta_i \sum_{k=1}^{n} y_k x_{ki} + \theta_i^2 \sum_{k=1}^{n} x_{ki}^2]\} \prod_{k=1}^{p-1} \frac{1}{a_{k,i}} [e^{-a_{k,i} B_k \theta_i} - e^{-a_{k,i} A_k \theta_i}]\, d\theta_i}
$$

$$
= \frac{\int_{A_i}^{B_i} \frac{1}{\theta_i^p} \exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{n} x_{ki}^2 (\theta_i - \frac{\sum_{k=1}^{n} y_k x_{ki}}{\sum_{k=1}^{n} x_{ki}^2})^2\} \prod_{k=1}^{p-1} [e^{-a_{k,i} B_k \theta_i} - e^{-a_{k,i} A_k \theta_i}]\, d\theta_i}{\int_{A_i}^{B_i} \frac{1}{\theta_i^{p-1}} \exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{n} x_{ki}^2 (\theta_i - \frac{\sum_{k=1}^{n} y_k x_{ki}}{\sum_{k=1}^{n} x_{ki}^2})^2\} \prod_{k=1}^{p-1} [e^{-a_{k,i} B_k \theta_i} - e^{-a_{k,i} A_k \theta_i}]\, d\theta_i}
$$

where $a_{k,i} = \frac{1}{\sigma^2} (\sum_{s=1}^{n} x_{sk} x_{si})$.

Like we did before, this integral can also be computed numerically, then the estimator of the population total or mean can be obtained.

Chapter 4

Bayes' Estimators for Small Domains

4.1 Introduction

In a large scale sample survey, sometimes we have lack of information about some sub-areas. In applying SRS on the whole population, some sub-areas are too small to be represented in the sample, thus making it difficult to estimate their subpopulation parameters by usual estimators. We have to find other ways to do it. Such estimators are called small domain estimators. The estimates of the parameters of interest (i.e. population mean or total, etc.) for those small areas can be obtained by borrowing the information from other neighboring areas. Many works on Bayesian small domain estimation have been done in recent years, e.g. Ghosh and Meeden (1988), Lui (1988), Ghosh and Lahiri (1987), Sarndal (1984, 1985) and others. In this chapter, we will discuss the linear model in general. To get estimators for the parameters of small domains one can apply the techniques discussed in previous chapters.

Let the population $U=\{1, 2, \ldots, N\}$ be divided into D non-overlapping domains $U_{1*}, U_{2*}, \ldots, U_{D*}$. Let N_{d*} be the size of U_{d*}. The population is further divided along a second dimension into G non-overlapping groups $U_{*1}, U_{*2}, \ldots, U_{*G}$. The size of U_{*g} is denoded by N_{*g}. The cross-classification of domains and groups gives D·G population cells U_{dg}, where d=1, 2, ..., D and g=1, 2, ..., G. Let N_{dg} be the size of U_{dg}.

Then the population size of N can be expressed as

$$N=\sum_{d=1}^{D} N_{d*}=\sum_{g=1}^{G} N_{*g}=\sum_{d=1}^{D}\sum_{g=1}^{G} N_{dg} \qquad (4.1)$$

Let s denote a sample of size n drawn from U by SRS. Denote by S_{d*}, S_{*g}, S_{dg} the parts of s that happen to fall, respectively, in U_{d*}, U_{*g} and U_{dg}.

The corresponding sizes, which are random variables, are denoted by n_{d*}, n_{*g} and n_{dg}. Also we have

$$n=\sum_{d\in s_{d*}} n_{d*}=\sum_{g\in s_{*g}} n_{*g}=\sum_{d\in s_{d*}}\sum_{g\in s_{*g}} n_{dg} \qquad (4.2)$$

4.1.1. Example

Let all firms in one country be our units of the industrial population. The first dimension, domain, is the standard industrial classfication, called SIC. Each firm only belongs to one domain. The second dimension, group , is measured by regions or provinces of the country. The variable of interest is the amount of gross income of the firm called Y_t (t=1, 2, ..., N). The auxiliary variables may contain other relative information about this firm, for instance, the number of employees of the firm which can be obtained by other surveys.

4.2 Estimators for domains

4.2 Estimators for domains

If we write the domain's and group's totals and means as;

$$\overline{Y}_{d*}=\frac{1}{N_{d*}}\sum_{k=1}^{N_{d*}} Y_k \qquad d=1,2,\ldots,D \qquad (4.3)$$

$$\overline{Y}_{*g}=\frac{1}{N_{*g}}\sum_{k=1}^{N_{*g}} Y_k \qquad g=1,2,\ldots,G \qquad (4.4)$$

$$Y_{d*}=\sum_{k=1}^{N_{d*}} Y_k \qquad d=1,2,\ldots,D \qquad (4.5)$$

$$Y_{*g}=\sum_{k=1}^{N_{*g}} Y_k \qquad g=1,2,\ldots,G \qquad (4.6)$$

Obviously, both Y_{d*} and Y_{*g} (or $\overline{Y}_{d*}$ and $\overline{Y}_{*g}$) are often what

we want to estimate. Domains are supposed to be smaller compared to groups. Thus we concentrate on estimating

Y_{d*} or $\overline{Y}_{d*}$ in this chapter.

If the sample size of S_{d*}, is large enough, we prefer to use the techniques we mentioned before to estimate Y_{d*}

(or $\overline{Y}_{d*}$). Otherwise , we have to adjust our estimators by adopting other information about the domains or the groups.

4.2.1 Assumptions

i) Y_k has the same distribution in each domain.

ii) Y_{dg} has one of the following linear models;

a) Y_{dg} has a linear relation with parameter θ_d , or

$$Y_{dg}=\theta_d+\varepsilon_{dg} \qquad d=1,2,\ldots,D, \qquad g=1,2,\ldots,G \qquad (4.7)$$

b) Y_{dg} has linear relation with the auxiliary variable X_{dg}

$$Y_{dg}=X_{dg}\theta_d+\varepsilon_{dg} \qquad d=1,2,\ldots,D, \qquad g=1,2,\ldots,G \qquad (4.8)$$

c) Y_{dg} has a linear relation with the auxiliary variables $X_{dg}^{(1)},X_{dg}^{(2)},\ldots,X_{dg}^{(P)}$.

$$Y_{dg}=\sum_{k=1}^{P} X_{dg}^{(k)}\theta_d^{(k)}+\varepsilon_{dg} \qquad d=1,2,\ldots,D, \qquad g=1,2,\ldots,G \qquad (4.9)$$

iii) The parameter θ_k follows the same prior distribution for each domain.

<u>4.2.2 Estimators</u>

Considering the model (4.9) which is the most general one. By the methods derived in Chapter 3, the estimator for each domain total can be written as;

$$\hat{Y}_{d\cdot}=\sum_{k\in s_{d\cdot}} y_{d\cdot}^{(k)}+\sum_{i=1}^{P} (X_{id}-x_{i,s_{d\cdot}})\,E(\theta_{id}\,|s_{d\cdot}) \qquad d=1,2,\ldots,D \qquad (4.10)$$

<u>Case 1.</u> If $n_{d\cdot}=0$, then we have

$$\hat{Y}_{d*}=\sum_{i=1}^{p} X_{id}\boldsymbol{E}(\theta_{id}) \qquad (4.11)$$

Where $\boldsymbol{E}(\theta_{id})$ is the expectation of θ_{id} 's prior

distribution.

<u>Case 2.</u> If p=1, the model is reduced to the model in (4.8). So the estimator for each domain total is given by

$$\hat{Y}_{d*}=\sum_{k\epsilon s_{d*}} y_{d}^{(k)}+(X_{d}-x_{s_d})\boldsymbol{E}(\theta_{d}\,|s_{d*}) \qquad (4.12)$$

<u>Cases 3.</u> If p=1, X_{dg}=1, The model is reduced to the model in (4.7). And the estimator for each domain total is

$$\hat{Y}_{d*}=\sum_{k\epsilon s_{d*}} y_{d*}^{(k)}+(N_{d*}-n_{s_d})\boldsymbol{E}(\theta_{d}\,|s_{d*})\,. \qquad (4.13)$$

Also, we can get the estimator for the total of the area of each domain across groups similarily. We obtain

$$\hat{Y}_{dg}=\sum_{k\epsilon s_{dg}} y_{dg}^{(k)}+\sum_{i=1}^{p}(X_{i,dg}-x_{i,dg})\boldsymbol{E}(\theta_{i,s_{dg}}\,|s_{dg}) \qquad d=1,2,\ldots,D \quad g=1,2,\ldots G \qquad (4.14)$$

Now since,

$$Y_{d*}=\sum_{g=1}^{G} Y_{dg},\ \sum_{k\epsilon s_{d*}} y_{d*}^{(k)}=\sum_{g=1}^{G}\sum_{k\epsilon s_{dg}} y_{dg}^{(k)},\ X_{i,d}=\sum_{g=1}^{G} X_{i,dg} \text{ and } x_{i,s_{dg}}=\sum_{g=1}^{G} x_{i,s_{dg}}$$

and because of assumption (iii), $\boldsymbol{E}(\theta_{i,d}\,|s_{dg})=\boldsymbol{E}(\theta_{i,d}\,|s_{d*})$ for

g=1,2,...,G. Hence, the adjusted estimator (4.14) becomes

$$\hat{Y}_{dg}=\sum\nolimits_{k\in s_{dg}} y_{dg}^{(k)}+\sum_{i=1}^{p}(X_{i,dg}-x_{i,s_{dg}})\,E(\theta_{d*}\,|s_{d*}) \qquad (4.15)$$

Also, for model (4.8) and (4.9), we have

$$\bar{Y}_{dg}=\sum\nolimits_{k\in s_{dg}} y_{dg}^{(k)}+(N_{dg}-n_{dg})\,E(\theta_{d}\,|s_{d*}) \qquad (4.16)$$

and

$$\bar{Y}_{dg}=\sum\nolimits_{k\in s_{dg}} y_{dg}^{(k)}+(X_{dg}-x_{dg})\,E(\theta_{d}\,|s_{d*}) \qquad (4.17)$$

respectively.

4.3 A Numerical Study

In this section, we will apply our methods to estimate the 1989's average and total amount of high tech imports to each province of Canada. The database including whole Canadian foreign high tech trade data between 1978 to 1989 was obtained from Statistics Canada. By the definition of high tech, about 23 kinds of goods based on three digits of Standard Industry Trade Classfication (SITC) are selected. The database includes variables of provinces, SITC codes of goods, the countries which exported the goods, etc. [see DeBresson, Hu and Cotsomitis (1990)].

According to the economic theory, the amount of high tech trade is unstable year by year. It relies on the international competition and capital flows in the country and province. Also, we found that the amount of trade shifted between different goods for each province of Canada. Totally there are about 15,870 cases in each year. We took one percent of all cases for small provinces and 0.1 percent for large provinces, such as Ontario, Quebec and B.C., as our sample. We assume that these data came from a superpopulation family. Each province is considered a domain of the population, so we have 12 domains. Table 4.1 presents the 12 provinces' high-tech trade totals, averages and number of cases.

Table 4.1

year	Province	N_i	Total (10 million$)	Average ($)
89	NF	332	1.5450877	46538
89	PEI	98	0.1119400	11422
89	NS	898	31.8	354109
89	NB	494	11.8	238835
89	Quebec	3650	509.0	1393973
89	Ontario	4685	1270.0	2710668
89	Manitoba	1043	38.2	366466
89	Sask.	493	7.7174185	156539
89	Alberta	1541	75.2	488047
89	BC	2320	207.0	890726
89	YX	61	0.4056367	66498

A one (or 0.1) percent sample is selected from the population by using SRS. The bar-charts of likelihood functions suggest that the conditional distribution of Y_i (imported value of the ith province) given parameter θ could be fitted by a log-normal distribution. Let Z_i be the 'log' transformation of Y_i, the graphs in Appendix 3 show that for most of the provinces, such as Quebec, Ontario and

Manitoba, etc. bar-charts compare very closely with normal curve.

Some Bayesian results for log-normal distributions have appeared in Kaufman (1963) and Zellner (1984). We will apply these results directly to our problem in conjunction with model (3.5). However, some details are provided in the following section.

Given that Z=ln(Y) is normally distributed with mean μ and variance σ^2, then Y has by definition a log-normal distribution. It is well known that the mean of Y is given by $\theta = \exp\{\mu+\sigma^2/2\}$, with $0 < \theta < \infty$. Given n independent observations $z' = (z_1, z_2, \ldots, z_n)$, the likelihood function for μ and σ is:

$$l(\mu,\sigma \mid z) \propto \sigma^{-n} \exp\{-\frac{1}{2\sigma^2}[\nu s^2 + n(\mu-\bar{z})^2]\} \qquad (4.18)$$

where $\nu = n-1$, $\nu s^2 = \sum_{i=1}^{n} (z_i-\bar{z})^2$, and $n\bar{z} = \sum_{i=1}^{n} z_i$.

If we employ a diffuse prior pdf for μ and σ, i.e. $f(\mu, \sigma) \propto 1/\sigma$, it is well known that:

1) The conditional posterior pdf of μ given σ is normal with mean $\bar{z}$ and variance σ^2/n,

2) The marginal posterior pdf for μ is in the univariate Student-t form with mean $\bar{z}$ and variance $\frac{\nu s^2}{(\nu-2)n}$ for $\nu>2$, and

3) The posterior pdf for θ is:

$$f(\theta \mid z) \propto \theta^{\frac{n-2}{2}} \times \int_0^{\infty} \sigma^{-(\nu+2)} \exp\left\{-\left[\frac{\nu s^2+n(\ln\theta-\bar{z})^2}{2} 2\sigma^2+\frac{n\sigma^2}{8}\right]\right\} d\sigma . \qquad (4.19)$$

Therefore, to get the estimate,

$\sum_{i=1}^{n_j} y_{ji}+(N_j-n_j)E(\theta_j \mid y_{js})$, of the population total for each

province j, we have to calculate the posterior mean of θ_j

numerically. Appendix 2 gives such a program written in FORTRAN. The estimate of the population mean for each province j can simply be obtained by dividing the above estimate by the domain size.

Table 4.2 gives average relative error of the Bayes estimator and the sample mean for various choices of sample size and priors. We have selected two choices of priors for comparison; i) uniform prior on a large interval which constitutes (approximately) no information, ii) uniform prior on an interval concentrated around the actual mean.

Table 4.2

Prov.	$\bar{Y}$	$B.T_1$	$B.T_2$	$\bar{y}$	$e._1$%	$e._2$%	e.%	f%
NF	46538	28538	45045	42383	39	3.21	8.93	1
PEI	11422	16383	11749	10951	42	2.87	4.12	1
NS	354109	469506	334956	386121	32	5.40	9.04	1
NB	238835	178464	232557	217213	25	2.62	9.05	1
Qu	1393973	862557	1452540	1949634	38	4.20	39.86	0.1
Ont	2710668	2432003	2925847	2160281	10	7.93	20.30	0.1
Man	366466	376970	363160	313795	3	0.90	14.37	1
Sas	156539	148625	155629	158166	5	0.58	1.04	1
Alb.	488047	637386	488080	478885	30	0.007	1.88	1
BC	890726	400544	886338	855631	55	0.49	3.94	0.1
YX	66498	51648	66040	67994	22	0.69	2.25	1

Table 4.3

Prov.	$\bar{Y}$	$B.T_2$	$\bar{y}$	rv_2%	rv%	f%
NF	46538	45045	42383	3.20	232.41	1
PEI	11422	11749	10951	7.99	41.62	1
NS	354109	334956	386121	0.33	2342.65	1
NB	238835	232557	217213	0.43	373.59	1
Qu	1393973	1452540	1949634	4.39	10964.93	0.1
Ont	2710668	2925847	2160281	0.26	420.94	0.1
Man	366466	363160	313795	0.11	360.56	1
Sas	156539	155629	158166	4.59	295.58	1
Alb.	488047	488080	478885	0.05	161.76	1
BC	890726	886338	855631	0.01	4154.46	0.1
YX	66498	66040	67994	3.93	381.39	1

where e.%, $e._1$% and $e._2$ are absolute relative errors, rv_2% and rv% are relative variances[1] related to sample mean and

---

[1]relative variance(rv)= $E\left(\frac{T-\bar{Y}}{\bar{Y}}\right)^2 \approx \frac{1}{J}\sum_{j=1}^{J}\left(\frac{T_j-\bar{Y}}{\bar{Y}}\right)^2$, in our case,

Bayesian estimates based on 500 with replacement simple random samples.

Conclusion: An improper prior distribution gave us very poor estimates; in many cases their performances are much worse than sample means, especially when the sample size gets larger. But once we set a proper prior which concentrates around the population mean, then the Bayes estimators show a consistent advantage over the classical estimator, especially when sample size is small (see Quebec, Ontario and Manitoba cases). It means, in this case, if we have a previous knowledge on the parameter which we want to estimate, the Bayes procedure should be adopted. The Bayes' estimator also gives a much smaller relative variance than the sample mean does (see Table 4.3).

---

j=500, T_j could be sample means or Bayes' estimates.

Appendix 1

A FORTRAN program for the example in §3.1.3

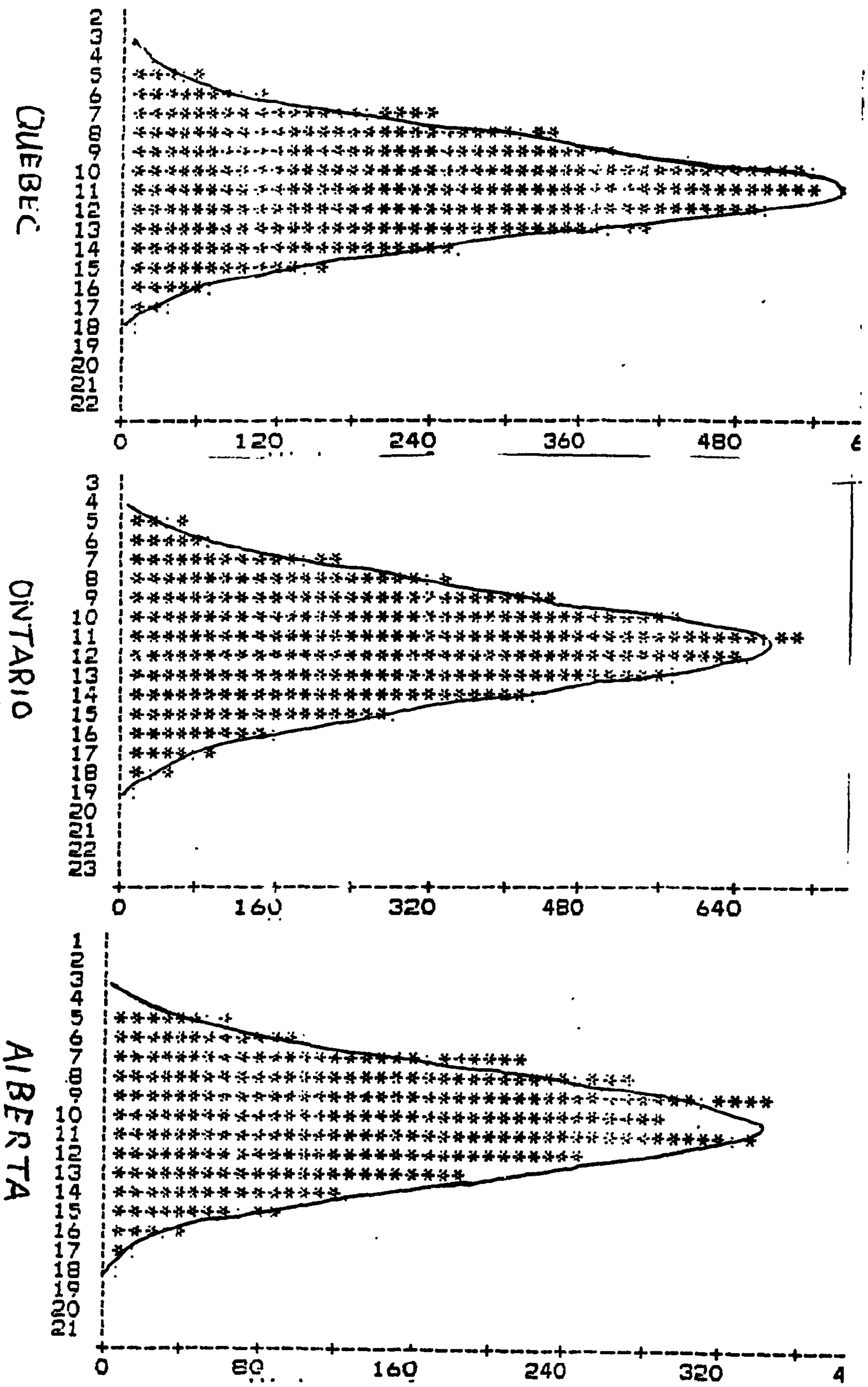
```
C       MAIN PROGRAM
        REAL N
        COMMON //TAO,N,A,B,ERRABS,ERRREL,/BL2/YBAR,/BL3/THETA
        TAO=t¹
        A=a
        B=b
        N=n
        ERRABS=0.0
        ERRREL=0.001
        YBAR=yb
        M=1000
        S=(F(A)+F(B))/2.0
        T=(A*F(A)+B*F(B))/2.0
        DO 10 T=1, M+1
        C=A+(I-1)*(B-A)/REAL(M)
        S=S+F(C)
        T=T+C*F(C)
 10     CONTINUE
        S=S*(B-A)/REAL(M)
        T=T*(B-A)/REAL(M)
        FINAL=T/S
        WRITE(*,*) FINAL, S, T, A, B, N, TAU, YBAR, M
        STOP
        END

        FUNCTION F(Y)
        REAL N
        COMMON //TAU, N, /BL2/YBAR
        F=EXP(-N*(Y-YBAR)*(Y-YBAR)/(2*TAU*TAU))/H(Y)
        RETURN
        END

        FUNCTION H(Z)
        INTEGER NOUT
        REAL N
        COMMON //TAU, N, A, B, ERRABS,ERRREL, RESULT, ERREST)
        H=RESULT**N
        RETURN
        END

        FUNCTION G(X)
        COMMON //TAU
        G=EXP(-(X-Z)*(X-Z)/(2*TAU*TAU))
        RETURN
        END
```

[1]. t, a, b, n, m and yb are all set by user according to different numbers.

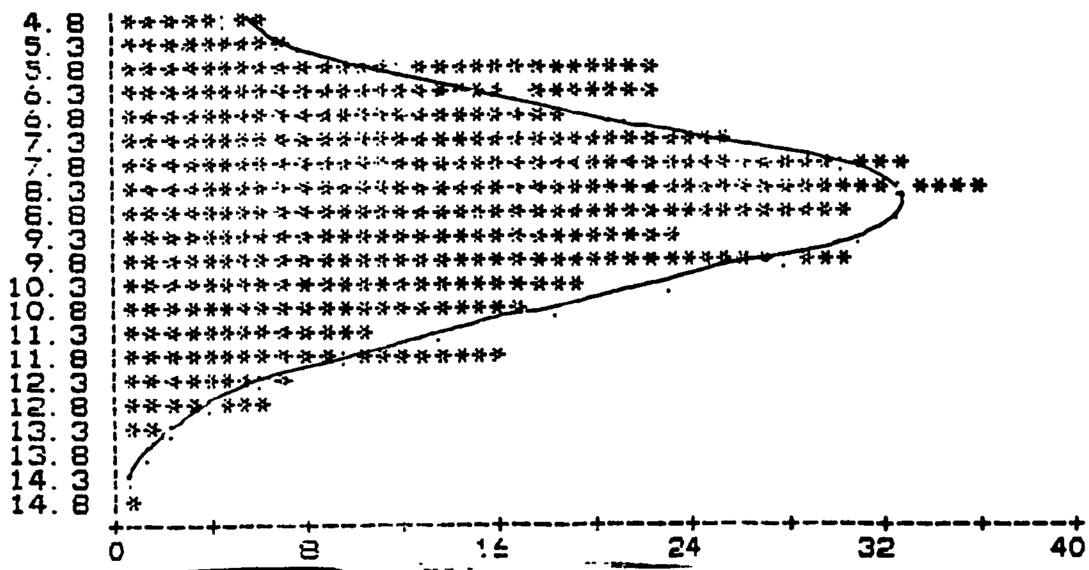
Appendix II

```
C        MAIN PROGRAM
         EXTERNAL F,U,G,H
         DIMENSION N(I),XBAR(50),S2(50),R(50),T(50),E1(50),E2(50),
     *             BX(I)
         DO 10 I=1,50
         OPEN(1,FILE='DATA.DAT',STATUS='OLD')
         READ(1,5) N(I),XBAR(I),S2(I)
  5      FORMAT (I5,2F5.2)
 10      CONTINUE
         A=1.0
         B=10000.0
         ERRABS=0.05
         ERREAL=0.05
         IRULE=1
         YBAR=11422.0
         CBX=0.0
         CX=0.0
         CALL UMACH (2,NOUT)
         DO 200 I=1,50
         CALL TWODQ(F,A,B,G,H,ERRABS,ERREAL,IRULE,R(I),E1(I))
         CALL TWODQ(U,A,B,G,H,ERRABS,ERREAL,IRULE,T(I),E2(I))
         B(I)=R(I)/T(I)
         BX(I)=0.1*XBAR(I)+0.9*B(I)
         CBX=CBX+BX(I)
         CX=CX+XBAR(I)
  200    CONTINUE
         CBX=CBX/50
         CX=CX/50
         PERCBX=(YBAR-CBX)/YBAR*100
         PERCX=(YBAR-CX)/YBAR*100
         WRITE(*,*) YBAR,CBX,PERCBX,CX,PERCX
         STOP
         END

         FUNCTION F(X,Y)
         F=((X**(N(I)/2-1))/(Y**(N(I)+1)))*EXP(-(((N(I)-1)*(S2(I)
     *    **2)+N(I)*(ALOG(X)-XBAR(I))**2)/(2*Y**2)+(N(I)*Y**2)/8))
         RETURN
         END

         FUNCTION U(X,Y)
         U=((X**(N(I)/2))/(Y**(N(I)+1)))*EXP(-(((N(I)-1)*(S2(I)**2)
     *      +N(I)*(ALOG(X)-XBAR(I))**2)/(2*Y**2)+(N(I)*Y**2)/8))
         RETURN
         END

         FUNCTION G(X)
         G=1.0
```

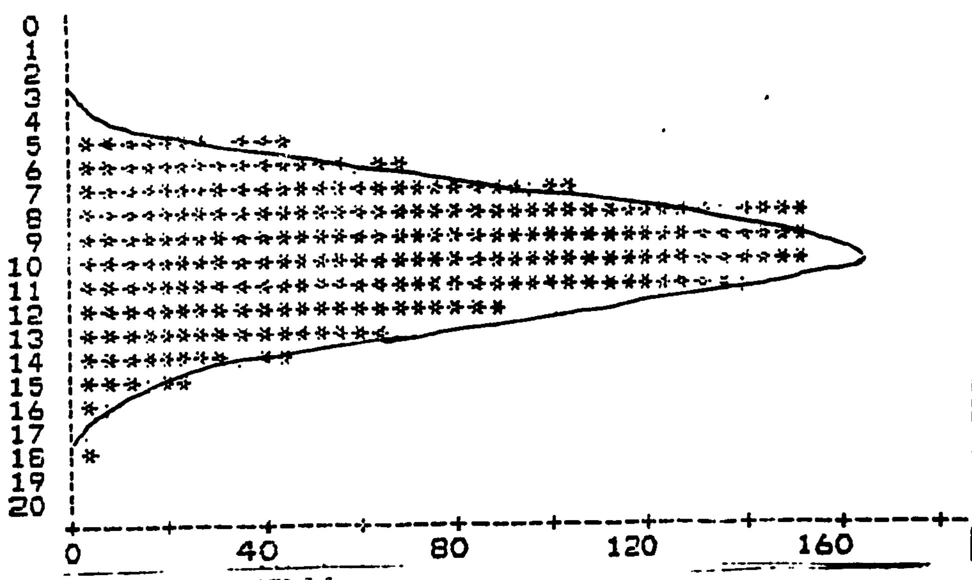
```
RETURN
END

FUNCTION H(X)
H=100.0
RETURN
END
```


QUEBEC
0
120
240
360
480
ONTARIO
0
160
320
480
640
ALBERTA
0
80
160
240
320

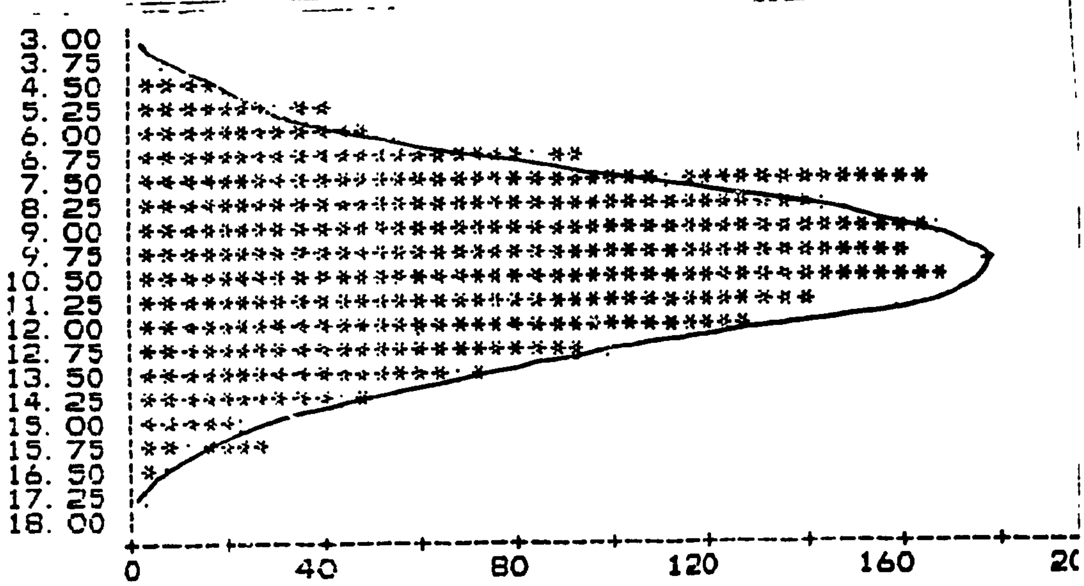
N.F.



MANITOBA



B.C.



REFERENCES

1. Banerjee, A.K. and Bhattacharyya, G.K. (1979). Bayesian results for the inverse Gaussian distribution with an application. Technometrics, vol. 21, No.2 247-251.

2. Berger, J.O. (1980). Statistical Decision Theory--foundations concepts and methods. Springer Verlag, New York, Heidelberg Berlin.

3. Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Reading, Mass.:Addison Wesley

4. Brunk, H.D. (1980), Bayesian Least Squares Estimates of Univariate Regression Functions. Conmmunications in Statistics, A9, 1101-1136.

5. Chaubey, Y.P., Nebebe, F. and Chen, P.S. (1990). Small area estimation under an inverse Gaussian Model. To appear.

6. Cochran, W.G. (1977). Sampling Techniques (3rd ed.), New York: John Wiley.

7. DeBresson, C., Hu, X. and Cotsomitis, J. (1990). Le Quebec dans le reseau mondial du commerce de produits de technologie avancee: 1978-1989. To appear.

8. Ericson, W.A. (1969). Subjective Bayesian models in sampling finite populations (with discussion). J. R. Statist. Soc. B. 31, 195-233.

9. ______, (1969). A note on the posterior mean of a population mean. J.R. Statist. Soc. B. 31, 332-335.

10. Ghosh, M. and Meeden, G. (1986). Empirical Bayes Estimation in Finite Population Sampling. J.A.S.A. Vol.

81, No. 396, 1058-1062.

11. _______, and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. J.A.S.A. Vol. 82, No. 400, 1153-1162.

12. Hartigan, J.A. (1969). Linear Bayesian Methods. J.R. Statist. Soc. B. 31, 446-454.

13. Hidiroglou, M.A. and Sarndal, C.E. (1985). An empirical study of some regression estimators for small domains. Survey Methodology, Vol.11, No.1, 65-77.

14. Jeffreys, H. (1967). Theory of Probability. (3rd rev. ed.; 1st ed. 1939) London: Oxford University Press.

15. _______, (1973). Scientific Inference. (3rd ed.; 1st ed.; 1931) Cambridge: Cambridge University Press.

16. Lui, K.J., and Cumberland, W.G. (1988). A Bayesian approach to small domain estimation.

17. Johnson, N. J., and Samuel Kotz (1970). Distributions in Statistics--continuous univariate distribution 1 & 2. Houghton Mifflin Company. Boston.

18. Sarndal, C.E. (1984). Estimation for small domains. J.A.S.A. Vol. 79, No. 387, 624-631.

19. Smouse, E.P. (1984). A note on Bayesian least squares inference for finite population models. J.A.S.A. Vol. 79, No. 386, 390-392.

20. _______, (1982). Bayesian estimation of a finite population total using auxiliary information in the presence of nonresponse. J.A.S.A. Vol. 77, 97-102.

21 Tweedie, M. (1957). Statistical properties of inverse Gaussian distribution I. Ann. Math. Stat. 28. 362-377.

22. Zellner, A. (1984). Basic Issues in Econometrics. The University of Chicago Press, Chicago and London.