BAYESIAN ESTIMATION IN TWO-STAGE RANDOM SAMPLING

Ing-Tzer Wey
Professor of Statistics,
Graduate School of Business Administration

I. INTRODUCTION

In a series of articles, Ericson, W. A. [1, 2, 3, 4] has put forth a subjective Bayesian approach to problems of inference regarding characteristics of finite populations. The basic model of a finite population is taken to be that which was initially put forth and subsequently used by Godambe, V. P. [5, 6, 7] and others. The model may be summarized as follows:

The notation of this section follows that of Ericson [4]. A finite population of N distinguishable units is defined by $N = \{1, 2, ..., N\}$, the set of unit labels, and by $\mathbf{X} = (X_1, ..., X_N)$, where X_i is the unknown value of some characteristic possessed by the population element labelled or identified by the integer i. For any sample from this population we let $(s, \mathbf{x})$ denote the Bayesian sufficient statistic, comprising $s \subset N$ where s is the set of the n distinct population elements included in the sample and $\mathbf{x} = (x_{i_1}, ..., x_{i_n})$ is the vector of the observed values of X_{i_j} for $i_j \in s$. If $p(\mathbf{X})$ denotes any joint density assigned to $\mathbf{X}$, then the posterior distribution of $\mathbf{X}$ given any sample has density given by

$$p(\mathbf{X}|(s, \mathbf{x})) = \begin{cases} p(\mathbf{X})/p_s(\mathbf{x}), & \text{for } \mathbf{X} \text{ such that } X_{i_j} = x_{i_j},\ i_j \in s \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $p_s(\mathbf{x}) \neq 0$ denotes the marginal prior density of X_{i_j} evaluated at $x_{i_j}, i_j \in s$.

While many other classes of prior densities, $p(\mathbf{X})$, will be useful to study we will consider here only those which represent finitely exchangeable prior knowledge regarding the X_i's and moreover can be generated by assuming that conditional on some parameter $\theta = (\theta_1, ..., \theta_m)$ the X_i's are independent, identically distributed with density $f(X|\theta)$ and that θ is assigned a distribution with density $g(\theta)$. Thus

$$p(\mathbf{X}) = \int_\theta \prod_{i=1}^{N} f(X_i \mid \theta) g(\theta) d\theta \tag{2}$$

It follows from (1) that under the class of priors in (2) the posterior density of **X** given any sample is given by

$$p(\mathbf{X} \mid (s, \mathbf{x})) = \begin{cases} \int_\theta \prod_{i \notin s} f(X_i \mid \theta)\, g(\theta \mid (s, \mathbf{x}))\, d\theta & \text{for } \mathbf{X} \text{ such that } X_{ij} = x_{ij},\ ij \in s \\ 0, & \text{otherwise.} \end{cases} \qquad (3)$$

where $g(\theta \mid (s, \mathbf{x})) \propto \prod_{i \in s} f(x_i \mid \theta)\, g(\theta)$ is the posterior density of the parameter θ.

The mean of the finite population will be denoted by $\mu = \sum_{i=1}^{N} X_i/N$, and we will let $\mu(\theta) \equiv E(X_i \mid \theta)$ be the mean of the conditional distribution of X_i given θ. Finally we will let $\bar{X}$ denote the unknown mean of any subset of n of the distinct finite population characteristic values, X_i, and given the sample sufficient statistic, (s, x), will let $\bar{x} = \sum_{i \in s} x_i/n$ be the observed sample mean.

The theorem below given by Ericson [4] will be compared with the new results which are to be given in the present paper.

Theorem 1. (Ericson) Under the model in (2) suppose that, conditional on θ, $X_1, \ldots, X_N$ are independently distributed with common density $f(X \mid \theta)$. Let G_f be a class of distributions of θ haveing density $g(\theta \mid x', n', y')$ for $(x', n', y') \in Y$, say. Suppose that G_f has the property that if x is any observation on X and if a prior distribution of θ is taken to be $g(\theta \mid x', n', y') \in G_f$ then the posterior distribution of θ has a density given by

$$g(\theta \mid x + x', n + n', y'') \in G_f \qquad (4)$$

where a single-primed symbol denotes a prior parameter, double-primed a posterior parameter and an unprimed symbol a sample statistic. Finally, letting $m' = E(X_i)$, suppose that for every $g(\theta \mid x', n', y') \in G_f$

$$m' = E_\theta[E(X_i \mid \theta)] = E_\theta[\mu(\theta)] = (x' + a)/(n' + b) \qquad (5)$$

where a and b are any constants.

Then, given any sample yielding the values (s, **x**):

$$E[\mu(\theta) \mid (s, x)] = \frac{\bar{x}\,\mathrm{Var}[\mu(\theta)] + m' E_\theta[\mathrm{Var}(\bar{X} \mid \theta)]}{\mathrm{Var}[\mu(\theta)] + E_\theta[\mathrm{Var}(\bar{X} \mid \theta)]} \qquad (6)$$

$$E[\mu \mid (s, x)] = \frac{\bar{x}\,\mathrm{Var}(\mu) + m' E_\mu[\mathrm{Var}(\bar{X} \mid \mu)]}{\mathrm{Var}(\mu) + E_\mu[\mathrm{Var}(\bar{X} \mid \mu)]} \qquad (7)$$

$$\mathrm{Var}\,[\,\mu \mid (s, x)\,] = \frac{N - n}{N} \left\{ \frac{\mathrm{Var}\,[\,\mu(\theta) \mid (s, x)\,]}{\mathrm{Var}\,[\,\mu(\theta)\,]} \right\} \mathrm{Var}(\,\mu\,) \tag{8}$$

It is readily seen that all of the conditions of the theorem hold if $f(X \mid \theta)$ is a member of the exponential family, if a member of a perhaps extended class of natural conjugate priors is assigned to θ, and if (5) holds. For example, the forms (6) through (8) hold in the cases where $f(X \mid \theta)$ is taken to be normal with known variance, normal with unknown mean and variance, binomial, Possion, etc., and $g(\theta)$ taken to be normal, normal-gamma, beta, gamma, etc., respectively. It is well known that under a squared-error loss function, the posterior mean of μ is the Bayes estimator of μ. We note that the posterior mean of μ in (7) is a weighted average of the prior mean m' and the sample mean $\bar{x}$ with weights inversely proportional to the prior variance of μ and the prior expected conditional sampling variance of the sample mean.

II. THE POSTERIOR MEAN AND THE BEST LINEAR UNBIASED ESTIMATOR

In the present study, a further generalization of Ericson's theorem is given which provides a similar tie between the posterior mean and the best linear unbiased estimator.

Theorem 2 Suppose n elements of nxl vector $\underline{X} = (X_1, \ldots, X_n)'$ and θ are any n+1 jointly distributed random variables such that

(a) $E(\theta) = m$, $\mathrm{Var}(\theta) = v(\theta) < \infty$.

(b) $\Sigma(\underline{X})$, the dispersion matrix of $\underline{X}$ is nxn positive definite.

(c) $E(X_i \mid \theta) = \theta$, $i = 1, \ldots, n$.

Suppose also that either

(d) $E(\theta \mid \underline{X}) = \underline{X}'\underline{\beta} + \alpha$, $\underline{\beta} = (\beta_1, \ldots, \beta_n)'$

or (d') α and $\underline{\beta}$ are chosen to minimize

$$E_{\underline{X}}\,[\,E(\theta \mid \underline{X}) - \underline{X}'\underline{\beta} - \alpha\,]^2.$$

Under these conditions, then

$$E(\theta \mid \underline{X}) = \frac{\hat{\theta}\, v(\theta) + m\, E_\theta\,[\mathrm{Var}(\hat{\theta} \mid \theta)\,]}{v(\theta) + E_\theta\,[\mathrm{Var}(\hat{\theta} \mid \theta)]}$$

where $\hat{\theta} = \underline{1}'[\Sigma(\underline{X})]^{-1}\underline{X} / \underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1}$, $\underline{1} = (1, 1, \ldots, 1)'$, is the usual best linear unbiased estimator of θ with respect to the dispersion matrix $\Sigma(\underline{X})$ or $E_\theta\,[\,\Sigma(\underline{X} \mid \theta)]$.

Further, the expected value of the posterior variance of θ is given by

$$E_{\underline{X}}[Var(\theta|\underline{X})] = v(\theta)[\ 1 - v(\theta)\underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1}\ ]$$

$$= v(\theta)\left[\frac{E_\theta[Var(\hat{\theta}|\theta)]}{v(\theta) + E_\theta[Var(\hat{\theta}|\theta)]}\right] \quad (11)$$

Proof: It suffices to prove the results (10) and (11) under conditions (a), (b), (c) and (d') since condition (d) is a special case of (d'). Note that $E(\underline{X}|\theta) = \theta\underline{1}$ and

$$\begin{aligned}\Sigma(\underline{X}) &= E[\ \underline{X} - m\underline{1}\ ][\ \underline{X} - m\underline{1}\ ]' \\ &= E[\ (\underline{X} - \theta\underline{1}) + (\theta - m)\underline{1}\ ][\ (\underline{X} - \theta\underline{1}) + (\theta - m)\underline{1}]' \\ &= E_\theta[\ \Sigma(\underline{X}|\theta)\ ] + v(\theta)\underline{1}\ \underline{1}' \end{aligned} \quad (12)$$

Now, from the generalized least squares theory, we have the best linear unbiased estimator of θ with respect to the dispersion matrix $E_\theta[\Sigma(\underline{X}|\theta)]$ as follows:

$$\hat{\theta} = \frac{1'[E_\theta\Sigma(\underline{X}|\theta)]^{-1}\underline{X}}{\underline{1}'[E_\theta\Sigma(\underline{X}|\theta)]^{-1}\underline{1}} = \frac{1'[\ \Sigma(\underline{X}) - v(\theta)\underline{1}\ \underline{1}'\ ]^{-1}\underline{X}}{\underline{1}'[\ \Sigma(\underline{X}) - v(\theta)\underline{1}\ \underline{1}'\ ]^{-1}\underline{1}} \quad (13)$$

Since $E_\theta[\Sigma(\underline{X}|\theta)]$ differs from $\Sigma(\underline{X})$ by a matrix of constants, $v(\theta)\underline{1}\ \underline{1}'$, θ must also be the best linear unbiased estimator of θ with respect to the dispersion matrix $\Sigma(\underline{X})$, hence

$$\hat{\theta} = \underline{1}'[\Sigma(\underline{X})]^{-1}\underline{X}\ /\ \underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1} \quad (14)$$

Note that the equivalence of (13) and (14) may also be shown by direct computation. The expected value of the conditional variance of θ, given θ is as follows:

$$E_\theta[Var(\hat{\theta}|\theta)] = E_\theta\left\{\frac{\underline{1}'[\Sigma(\underline{X})]^{-1}\Sigma(\underline{X}|\theta)[\Sigma(\underline{X})]^{-1}\underline{1}}{(\underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1})^2}\right\}$$

$$= \frac{1}{\underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1}} - v(\theta) \quad (15)$$

It is easily to show that the following identity holds:

$$E_{\underline{X},\theta}(\ \theta - \underline{X}'\underline{\beta} - \alpha\ )^2 = E_{\underline{X}}[Var(\theta|\underline{X})] + E_{\underline{X}}[\ E(\theta|\underline{X}) - \underline{X}'\underline{\beta} - \alpha\ ]^2 \quad (16)$$

From this identity, we see that the selection of α and $\underline{\beta}$ to minimize (9) is equivalent to the choice of α and $\underline{\beta}$ to minimize the left-hand side of (16). Computing the expectation of the left-hand side of (16) and minimizing it by straight forward method yields

$$\alpha = m - m\underline{1}'\underline{\beta} \tag{17}$$

and
$$\underline{\beta} = v(\theta)[\Sigma(\underline{X})]^{-1}\underline{1} \tag{18}$$

Hence,
$$\underline{X}'\underline{\beta} + \alpha = m(1 - v(\theta)\underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1}) + v(\theta)\underline{1}'[\Sigma(\underline{X})]^{-1}\underline{X} \tag{19}$$

Substituting (14) and (15) into (19), we have

$$\underline{X}'\underline{\beta} + \alpha = \frac{\hat{\theta}\, v(\theta) + m\, E_\theta[\,Var(\hat{\theta}|\theta)]}{v(\theta) + E_\theta[\,Var(\hat{\theta}|\theta)]}$$

From the right-hand side of (16), we see that the minimum is attained at $E(\theta|\underline{X}) = \underline{X}'\underline{\beta} + \alpha$.

Finally, the expected value of the posterior variance of θ is obtained as follows:

$$\begin{aligned} E_{\underline{X}}[Var(\theta|\underline{X})] &= E_{\underline{X}}E_{\theta|\underline{X}}[\,\theta - \underline{\beta}'\underline{X} - \alpha\,]^2 \\ &= E_{\underline{X}}E_{\theta|\underline{X}}[(\theta - m) - \underline{\beta}'(\underline{X} - m\underline{1})]^2 \\ &= v(\theta) + \underline{\beta}'\Sigma(\underline{X})\underline{\beta} - 2[v(\theta)]^2\underline{\beta}'\underline{1} \\ &= v(\theta) - [v(\theta)]^2\underline{\beta}'\,\underline{1} \\ &= v(\theta)[\,1 - v(\theta)\underline{1}'[\Sigma(\underline{X})]^{-1}\underline{1}] \end{aligned}$$

The theorem is of some interest in itself, namely, if the posterior mean of θ is a linear combination of $\underline{X}$, then it is a weighted average of the prior mean of θ and the generalized least squares estimator of θ, with weights inversely proportional to the prior variance of θ and the expected value of the conditional variance of the generalized least squares estimator of θ, given θ, respectively. It is obvious that the posterior mean of θ will coincide with that of Ericson's result given in (7) when the dispersion matrix of $\underline{X}$ is of the form $\sigma^2 I$, where I is the identity matrix. This result will be immeidately applicable in the Bayesian analysi of two-stage random sampling given in the succeeding sections.

III. A CLASS OF PRIOR DISTRIBUTIONS IN TWO-STAGE RANDOM SAMPLING

Suppose that the population under study consists of N primary sampling units (psu's) of unequal size and let M_i denote the number of secondary sampling units (ssu's) in the i-th psu. Correspondingly let X_{ij} denote the characteristic value of interest associated with the j-th ssu within the i-th psu. Let

$$\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i,\ w_i = M_i \,/\, \bar{M} \tag{20}$$

We define the population parameters as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} w_i \mu_i, \quad \mu_i = \frac{1}{M_i} \sum_{j=1}^{M_i} X_{ij} \tag{21}$$

$$\sigma_i^2 = \frac{1}{M_i} \sum_{j=1}^{M_i} (X_{ij} - \mu_i)^2 \tag{22}$$

$$\sigma_b^2 = \frac{1}{N} \sum_{i=1}^{N} w_i \, (\mu_i - \mu)^2 \tag{23}$$

and

$$\sigma_w^2 = \frac{1}{N} \sum_{i=1}^{N} w_i \, \sigma_i^2 \tag{24}$$

All of these population parameters are assumed unknown and μ is the target of inference.

The simplest useful class of prior distributions is that under which the random variables are taken to be exchangeable. N random variables are said to be exchangeable if the joint probability distribution of each of the N! permutations of the variables is the same. The class of exchangeable prior distributions given in (2) was extensively used in Ericson's papers. The main results of the present study will be demonstrated under a fairly general class, C, of prior distributions. This class of prior distributions is briefly that under which, conditional on the psu mean μ_i and variance σ_i^2, the X_{ij}'s are taken to be exchangeable and independent over the psu's and, further, the psu means and variances are exchangeable.

The class C of prior distributions is defined by the following two assumptions:

A1. Conditional on (μ_i, σ_i^2), the X_{ij}'s are exchangeable for all j and also, for $i \neq h$, X_{ij} and X_{hk} are independent for all j and k.

A2. For i = 1, 2, ..., N the ordered pairs (μ_i, σ_i^2) are exchangeable random two-tuples such that the following moments exist:

$$E(\mu_i) = m, \; Var\,(\mu_i) = v, \; Cov\,(\mu_i\,,\,\mu_h) = \rho v, \; E(\sigma_i^2) = \pi$$

For making Bayesian inference about μ, we require to specify four parameters of the prior distribution in the class C as by A2, namely, m, v, ρ and π.

This specification does not define a unique prior but suffices for determining the form of the posterior mean of μ, the Bayes estimator of μ under the squared-error loss function.

Let $\bar{X}_i$ denote the mean of a sample of size m_i drawn from the i-th psu by simple random sampling without replacement. The following results are easily shown.

Lemma 1 If the prior distribution of the population under study is in the class C, then

$$E(X_{ij} \mid \mu_i, \sigma_i^2) = E(\bar{X}_i \mid \mu_i, \sigma_i^2) = \mu_i \tag{25}$$

$$Var(X_{ij} \mid \mu_i, \sigma_i^2) = \sigma_i^2 \tag{26}$$

$$Cov(X_{ij}, X_{ik} \mid \mu_i, \sigma_i^2) = -\sigma_i^2 / (M_i - 1) \tag{27}$$

$$Var(\bar{X}_i \mid \mu_i, \sigma_i^2) = \frac{M_i - m_i}{M_i - 1} \frac{\sigma_i^2}{m_i} \tag{28}$$

The following lemma can be trivially shown by using lemma 1 and the well-known theorem that if X and Y are two random variables, then the variance of X is equal to the expected value of the conditional variance of X, given Y plus the variance of the conditional expected value of X, given Y.

Lemma 2. If the prior distribution of the population under study is in the class C, then

$$E(\mu) = E(X_i) = E(X_{ij}) = m \tag{29}$$

$$Var(\mu) = v(1-\rho)\delta + \rho v, \text{ where } \delta = \sum_{i=1}^{N} w_i^2/N^2 \tag{30}$$

$$E(\sigma_w^2) = \pi \tag{31}$$

$$E(\sigma_b^2) = v(1-\rho)(1-\delta) \tag{32}$$

$$Var(X_{ij}) = v + \pi \tag{33}$$

$$Cov.(X_{ij}, X_{ik}) = v - \pi/(M_i - 1), \ j \neq k \tag{34}$$

$$\mathrm{Cov}\,(X_{ij}\,,\,X_{hk}) = \mathrm{Cov}\,(\overline{X}_i\,,\,\overline{X}_h) = \rho v,\ i \neq h \tag{35}$$

$$\mathrm{Var}\,(\overline{X}_i) = v + \frac{M_i - m_i}{M_i - 1}\,\frac{\pi}{m_i} \tag{36}$$

The prior expectations of the conditional variance and covariance of $\overline{X}_i$'s, given μ are obtained in the following lemma.

Lemma 3. If the prior distribution of the population under study is in the class C, then

$$E\,(\mu_i \mid \mu) = E\,(\overline{X}_i \mid \mu) = \mu \tag{37}$$

$$E_\mu\,[\mathrm{Var}\,(\overline{X}_i \mid \mu)] = v(1-\rho)(1-\delta) + \frac{M_i - m_i}{M_i - 1}\,\frac{\pi}{m_i} \tag{38}$$

$$E_\mu\,[\mathrm{Cov}\,(\overline{X}_i\,,\,\overline{X}_h \mid \mu)] = -v(1-\rho)\,\delta \tag{39}$$

The proof of lemma 3 follows first by observing that by A2 the μ_i's are marginally exchangeable and remain exchangeable conditional on their sum or on μ. The result then follows from lemma 1, lemma 2 and

$$E\,(\mu \mid \mu) = \mu$$

$$E_\mu\,[\mathrm{Var}\,(\overline{X}_i \mid \mu)] = \mathrm{Var}(\overline{X}_i) - \mathrm{Var}[E(\overline{X}_i \mid \mu)].$$

Now, let n psu's be selected by simple random sampling without replacement from a population of N psu's. Let m_i (i = 1, 2, ..., n) ssu's be taken independently by simple random sampling without replacement from the M_i (i = 1, 2, ..., n) ssu's in the selected psu's. Let $\underline{\overline{X}}$ denote nxl vector whose elements are the subsample means $\overline{X}_i$'s. Then, the dispersion matrix of $\underline{\overline{X}}$, denoted by $\Sigma(\underline{\overline{X}})$, is obtained from lemma 2 as follows:

$$\Sigma(\underline{\overline{X}}) = D + \rho v\,\underline{1}\,\underline{1}' \tag{40}$$

where D is nxn diagonal matrix whose i-th diagonal element is

$$d_i = v\,(1-\rho) + \frac{M_i - m_i}{M_i - 1}\,\frac{\pi}{m_i} \tag{41}$$

The prior expectation of the dispersion matrix of $\underline{\overline{X}}$ conditional on μ, denoted by E_{μ} $[\Sigma(\underline{\bar{X}} \mid \mu)]$, is obtained from lemma 3 as follows:

$$E_{\mu} [\Sigma(\underline{\bar{X}} \mid \mu)] = D - v(1 - \rho)\, \delta\, \underline{1}\, \underline{1}' \tag{42}$$

To find the inverse of $\Sigma(\underline{\bar{X}})$ and E_{μ} $[\Sigma(\underline{\bar{X}} \mid \mu)]$, the following lemma is useful.

Lemma 4. Let the nxn matrix A be given by

$$A = D + \lambda \underline{a}\, \underline{b}'$$

where D is a nonsingular diagonal matrix, a and b are each nxl vectors, and λ is a scalar such that

$$\lambda \neq - [\sum_{i=1}^{n} a_i\, b_i/d_i]^{-1}$$

The inverse of A is

$$A^{-1} = D^{-1} + \gamma \underline{a}^{*} \underline{b}^{*\prime}$$

where $\gamma = - \lambda (1 + \lambda \sum_{i=1}^{n} a_i b_i/d_i)^{-1}$, $a_i^{*} = a_i/d_i$, $b_i^{*} = b_i/d_i$, and d_i is the i-th diagonal element of $\underline{D}$.

Proof: The proof is given by showing that $AA^{-1} = I$.

Now, using lemma 4 we obtain the inverse of $\Sigma(\underline{\bar{X}})$ and E_{μ} $[\Sigma(\underline{\bar{X}} \mid \mu)]$ as follows: The inverse of $\Sigma(\underline{\bar{X}})$ is

$$[\Sigma(\underline{\bar{X}})]^{-1} = D^{-1} + \gamma\, \underline{u}\, \underline{u}' \tag{43}$$

where $\gamma = -\rho v/(1 + \rho v \sum_{i=1}^{n} u_i)$ and $u_i = d_i^{-1}$.

The inverse of E_{μ} $[\Sigma(\underline{\bar{X}} \mid \mu)]$ is

$$[E_{u} \Sigma(\underline{\bar{X}} \mid \mu)]^{-1} = D^{-1} + \alpha\, \underline{u}\, \underline{u}' \tag{44}$$

where $\alpha = v(1 - \rho)\, \delta/[1 - v(1 - \rho)\, \delta \sum_{i=1}^{n} u_i]$.

The results of this section will be immediately applicable in finding the Bayes estimator of the population mean μ with two-stage random sampling given in the next section.

IV. THE BAYES ESTIMATOR OF A POPULATION MEAN

Suppose that the population of interest consists of N psu's of unequal sizes and suppose that the prior distribution of the population belongs to the class C defined in the preceding section. Let n psu's be selected by simple random sampling without replacement from the population and let m_i ($i = 1, 2, ..., n$) ssu's be taken independently by simple random sampling without replacement from the M_i ($i = 1, 2, ..., n$) ssu's in the selected psu's. Let $\underline{\bar{X}}$ denote the nxl vector whose elements are the subsample means, $\bar{X}_i$'s. Then, it is easily to see that, under the correspondence $\mu \leftrightarrow \theta$ and $\bar{X}_i \leftrightarrow X_i$, the conditions (a), (b) and (c) of Theorem 2 are met. Now, if we use the squared-error loss function to find a Bayes estimator of μ and if the posterior mean of μ is linear in $\underline{\bar{X}}$, then the Bayes estimator of μ is the posterior mean of μ which is, from Theorem 2, the weighted average of m, the prior mean of μ, and $\hat{\mu}$, the best linear unbiased estimator of μ, with weights inversely proportional to the prior variance of μ and the expectation of the conditional variance of $\hat{\mu}$, given μ, respectively. Hence, we have the following result.

Theorem 3. Suppose that a prior distribution in the class C is assigned to a population consisting of N psu's of unequal sizes from which a two-stage random sample is drawn. If the squared-error loss function is used in finding the Bayes estimator of the population mean μ and if the posterior mean of μ is linear in $\underline{\bar{X}} = (\bar{X}_1, \bar{X}_2, ..., \bar{X}_n)'$, then the Bayes estimator of μ is given by

$$\hat{\mu}_B = \frac{\hat{\mu} \text{ Var }(\mu) + m \text{ } E_\mu \text{ [Var } (\hat{\mu}|\mu)]}{\text{Var }(\mu) + E_\mu \text{ [Var } (\hat{\mu}|\mu)]} \tag{45}$$

where $\hat{\mu} = 1'[\Sigma(\underline{\bar{X}})]^{-1}\underline{\bar{X}} \ / \ 1'[\Sigma(\underline{\bar{X}})]^{-1} 1$ is the best linear unbiased estimator of μ with respect to either the dispersion matrix of $\underline{\bar{X}}$, $\Sigma(\underline{\bar{X}})$, or the prior expectation of the dispersion matrix of $\underline{\bar{X}}$ conditional on μ,$E[\Sigma(\underline{\bar{X}}|\mu)]$. Further, the expected value of the posterior variance of μ is given by

$$E \text{ [Var } (\mu|\underline{\bar{X}})] = \text{Var } (\mu) \text{ [} 1 - \text{Var } (\mu) \ \underline{1}' \ [\Sigma(\underline{\bar{X}})]^{-1} \underline{1}] \tag{46}$$

It can be shown by direct computation that the best linear unbiased estimator (BLUE) of μ with respect to $\Sigma(\underline{\bar{X}})$ is the same as the BLUE of μ with respect to $E_\mu \ [\Sigma(\underline{\bar{X}}|\mu)]$. From (43), we have

$$\underline{1}' [\Sigma(\underline{\overline{X}})]^{-1} \underline{1} = \underline{1}' D^{-1} \underline{1} + \gamma\, \underline{1}' \underline{u}\, \underline{u}' \underline{1}$$

$$= \sum_{i=1}^{n} u_i / (1 + \rho v \sum_{i=1}^{n} u_i) \tag{47}$$

and
$$\underline{1}' [\Sigma(\underline{\overline{X}})]^{-1} \underline{\overline{X}} = \underline{1}' D^{-1} \underline{\overline{X}} + \gamma\, \underline{1}' \underline{u}\, \underline{u}' \underline{\overline{X}}$$

$$= \sum_{i=1}^{n} u_i \overline{X}_i / (1 + \rho v \sum_{i=1}^{n} u_i) \tag{48}$$

From (44), we have

$$\underline{1}' [E_\mu \Sigma(\underline{\overline{X}} \mid \mu)]^{-1} \underline{1} = \underline{1}' D^{-1} \underline{1} + \alpha\, \underline{1}' \underline{u}\, \underline{u}' \underline{1}$$

$$= \sum_{i=1}^{n} u_i (1 + \alpha \sum_{i=1}^{n} u_i) \tag{49}$$

and
$$\underline{1}' [E_\mu \Sigma(\underline{\overline{X}} \mid \mu)]^{-1} \underline{\overline{X}} = \underline{1}' D^{-1} \underline{\overline{X}} + \alpha\, \underline{1}' \underline{u}\, \underline{u}' \underline{\overline{X}}$$

$$= \sum_{i=1}^{n} u_i \overline{X}_i (1 + \alpha \sum_{i=1}^{n} u_i) \tag{50}$$

Thus, the BLUE of μ is

$$\hat{u} = \frac{\underline{1}' [\Sigma(X)]^{-1} \underline{\overline{X}}}{\underline{1}' [\Sigma(\underline{\overline{X}})]^{-1} \underline{1}} = \frac{\underline{1}' [E_\mu \Sigma(\underline{\overline{X}} \mid \mu)]^{-1} \underline{\overline{X}}}{\underline{1}' [E_\mu \Sigma(\underline{\overline{X}} \mid \mu)]^{-1} \underline{1}} = \sum_{i=1}^{n} u_i \overline{X}_i / \sum_{i=1}^{n} u_i \tag{51}$$

From (15), we have the expected value of the conditional variance of $\hat{\mu}$, given μ as follows:

$$E_\mu [\mathrm{Var}(\hat{\mu} \mid \mu)] = \frac{1}{\underline{1}' [\Sigma(\underline{\overline{X}})]^{-1} \underline{1}} - \mathrm{Var}(\mu)$$

$$= \frac{1 + \rho v \sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} u_i} - [v(1-\rho)\delta + \rho v]$$

$$= [1 - v(1-\rho)\delta \sum_{i=1}^{n} u_i] / \sum_{i=1}^{n} u_i \tag{52}$$

Inserting (51) and (52) into (45), we obtain $\hat{\mu}_B$, the Bayes estimator of μ We note that there is an unknown parameter in $\hat{\mu}_B$, namely,

$$\delta = \sum_{i=1}^{N} w_i^2 \; / \; N^2$$

An unbiased estimator of δ is given by

$$\hat{\delta} = \frac{1}{nN} \sum_{i=1}^{n} w_i^2 \tag{53}$$

where $w_i = M_i \; / \; \bar{M}$, provided that $\bar{M}$ is known.

Finally, the expected value of the posterior variance of μ can be written in an explicit form as follows:

$$E \; [\mathrm{Var} \; (\mu \mid \underline{\bar{X}})] = \frac{1 - v(1-\rho)\,\delta \sum_{i=1}^{n} u_i}{1 + \rho v \sum_{i=1}^{n} u_i} \; [\; v(1-\rho)\delta + \rho v \;] \tag{54}$$

Note that the prior expected value of the among psu's variance, $E(\sigma_b^2)$ does not appear in the weight u_i employed in the BLUE of μ. If we consider the variance of the psu mean μ_i, defined by

$$\sigma_u^2 = \frac{1}{N} \sum_{i=1}^{N} (\mu_i - \bar{\mu})^2 \tag{55}$$

where $\bar{\mu} = \frac{1}{N} \sum_{i=1}^{N} \mu_i$,

then, the weight u_i can be written as follows:

$$u_i = [\; v(1-\rho) + \frac{M_i - m_i}{M_i - 1} \; \frac{\pi}{m_i} \;]^{-1}$$

$$= [\; \frac{N}{N-1} \; E\;(\sigma_u^2) + \frac{M_i - m_i}{M_i - 1} \; \frac{1}{m_i} \; E(\sigma_w^2) \;]^{-1} \tag{56}$$

It is, therefore, clear that the weight u_i contains two components: one arising from prior variation within the psu's and one from prior variation among the true mean of the psu's.

V. SUMMARY

In a two-stage random sampling design, suppose that the prior distribution of the population of interest is in the class C under which the primary sampling unit (psu) means and variances (μ_i, σ_i^2) are viewed as exchangeable random variables with known moments and the X_{ij}'s within each psu are exchangeable and independent. Further, suppose that the posterior mean of the population mean μ is linear in the sample means $\underline{\bar{X}} = (\bar{X}_1, \bar{X}_2, ..., \bar{X}_n)$. Then, the Bayes estimator of μ with squared-error loss function is the weighted average of the prior mean of μ and the best linear unbiased estimator (BLUE) of μ, denoted by $\hat{\mu}$, with weights inveresly proportional to the prior variance of μ and the prior expected value of the conditional variance of $\hat{\mu}$, given μ, respectively. The BLUE of μ with respect to the dispersion matrix of $\underline{\bar{X}}$, $\Sigma(\underline{\bar{X}})$, is the same as that with respect to the prior expectation of the dispersion matrix of $\underline{\bar{X}}$ conditional on μ, $E_\mu\ [\Sigma(\underline{\bar{X}}|\mu)]$. Furthermore, the BLUE of μ is the weighted average of the sample means $\underline{\bar{X}}$ with weights inversely proportional to linear combinations of the prior expected values of the variance of true psu mean, σ_u^2 and the weighted average of within variances of psu's, σ_w^2.

REFERENCES

1. Ericson, W. A. Subjective Bayesian models in sampling finite populations. Journal of Royal Statistical Society, Series B, 31:195-233, 1969.
2. Ericson, W. A. A note on the posterior mean of a population mean. Journal of Royal Statistical Society, Series B, 31:332-334, 1969.
3. Ericson, W. A. Subjective Bayesian models in sampling finite populations: Stratification. New Developments in Survey Sampling (Johnson, N. L. and Smith, H. Jr. editors) New York:John Wiley & Sons, Inc., 1969. pp 326-357.
4. Ericson, W. A. On the posterior mean and variance of a population mean. Journal of American Statistical Assoication 65:649-652, 1970.
5. Godambe, V. P. A unified theory of sampling from finite populations. Journal of Royal Statistical Society, Series B, 17:267-278, 1955.
6. Godambe, V. P. A review of the contributions towards a unified theory of sampling from finite populations. Review of Internation Statistical Institute, 33:242-258, 1965.
7. Godambe, V. P. A new approach to sampling from finite pppulations. I and II. Journal of Royal Statistical Society, Series B, 28:310-328, 1966.