Bayes Estimation and Prediction under Informative Sampling Design

Abdulhakeem A.H. Eideh
Department of Mathematics
Al-Quds University, Al-Quds, Palestine
msabdul@staff.alquds.edu

Abstract

In this research we will deal with the problem of Bayes estimation of the parameter that characterise the superpopulation model, and Bayes prediction of finite population total, from a sample survey data selected from a finite population using informative probability sampling design, that is, the sample first order inclusion probabilities depend on the values of the model outcome variable (or the model outcome variable is correlated with design variables not included in the model). In order to achieve this we will first define the sample predictive distribution and the sample posterior distribution and then we use the sample posterior likelihood function to obtain the sample Bayes estimate of the superpopulation model parameters, and Bayes predictors of finite population total. These new predictors take into account informative sampling design. Thus, provides new justification for the broad use of best linear unbiased predictors (model-based school) in predicting finite population parameters in case of not accounting of complex sampling design. Furthermore, we show that the behaviours of the present estimators and predictors depends on the informativeness parameters. Also the use of the Bayes estimators and predictors that ignore the informative sampling design yields biased Bayes estimators and predictors. One of the most important feature of this paper is, specifying prior distribution for the parameters of the sample distribution makes life easier than the population parameters.

Keywords: Bayes estimator, Finite Population Sampling, Informative Sampling, Sample Distribution.

1. Introduction

Survey data may be viewed as the outcome of two processes: the process that generates the values of units in the finite population, often referred as the superpopulation model, (values of units in the finite population is a random sample from superpopulation model) and the process of selecting the sample units from the finite population, known as the sample selection mechanism. Analytic inference from sample survey data refers to the superpopulation model, while descriptive inference deals with estimation of finite population characteristics. When the sample selection probabilities depend on the values of the model response (or the model outcome variable is correlated with design variables not included in the model), even after conditioning on the auxiliary variables, the sampling mechanism becomes informative, and the selection effects need to be accounted for in the inference process. To overcome the difficulties associated with the use of classical inference procedures for cross sectional survey data, Pfeffermann, Krieger and Rinott (1998) proposed the use of the sample distribution induced by assumed population models, under informative sampling, and developed expressions for its calculation. Similarly, Eideh and Nathan (2006) fitted time series models for longitudinal survey data under informative sampling. Furthermore, to see examples in the field of cross sectional surveys and panel surveys that illustrating the effects of ignoring informative sampling design, please review the books edited by: Skinner, Holt and Smith (1989), and Chambers and Skinner (2003). In particular see the articles by Pfeffermann (2011), Pfeffermann and Sverchkov (1999, 2003), Sverchkov and Pfeffermann (2004), Eideh and

Nathan (2006a, 2006b, 2009), Eideh (2003, 2012a, 2012b). In these articles the authors reviewed many examples reported in the literature that illustrate the effect of ignoring the sampling process when fitting models to survey data based on complex sample and they discussed methods to deal with this problem.

Kim (2002) considers Bayesian and Empirical Bayesian approach to small area estimation under informative sampling. Sverchkov and Pfeffermann (2004) study the use of the sample distribution for the prediction of finite population totals under single-stage sampling. They propose predictors that employ the sample values of the target study variable, the sampling weights of the sample units and possibly known population values of auxiliary variables. They solve the prediction problem by estimating the expectation of the study values for units outside the sample as a function of the corresponding expectation under the sample distribution and the sampling weights. The prediction mean square error is estimated by a combination of an inverse sampling procedure and a re-sampling method. An interesting outcome of the present analysis is that several familiar estimators in common use are shown to be special cases of the proposed approach, thus providing them a new interpretation. They study the performance of the new and some old predictors in common use is evaluated and compared by a Monte Carlo simulation study using a real data set. In their article they use the sample and sample-complement distributions for developing design consistent predictors of finite population totals. Known predictors in common use are shown to be special cases of the present theory. The mean square errors (MSEs) of the new predictors are estimated by a combination of an inverse sampling algorithm and a resampling method. As supported by theory and illustrated in the empirical study, predictors of finite population totals that only require the prediction of the outcome values for units outside the sample perform better than predictors in common use even under a design based framework, unless the sampling fractions are very small. The MSE estimators are shown to perform well both in terms of bias and when used for the computation of confidence intervals for the population totals. The author pointed that "Further experimentation with this kind of predictors and MSE estimation is therefore highly recommended". Nandram et al. (2006) study the problem in which a biased sample is selected from a finite population (a random sample from a super-population), and inference is required for the finite population mean and the superpopulation mean. However, the measurements may not be normally distributed, a necessary condition for their method; a transformation, which can possibly provide a normal approximation, is needed. The authors use the Gibbs sampler and the sample importance resampling algorithm to fit the non-ignorable selection model to a simple example on natural gas production. Their non-ignorable selection model estimates the finite population mean production much closer to the true finite population mean than a model which ignores the selection probabilities, and there is improved precision of the non-ignorable selection model over this latter model. A naive 95% credible interval based on the Horvitz–Thompson estimator is too wide. Rao (2011) provides an account of both parametric and nonparametric Bayesian (and pseudo-Bayesian) approaches to inference from survey data, focusing on descriptive finite population parameters. Little (2012) characterize the prevailing philosophy of official statistics as a design/model compromise (DMC). It is design-based for descriptive inferences from large samples, and model-based for small area estimation, nonsampling errors such as nonresponse or measurement error, and some other subfields like autoregressive integrated moving average (ARIMA) modeling of time series. Little suggests that DMC involves a form of "inferential

schizophrenia", and offer examples of the problems this creates. He discuss an alternative philosophy for survey inference which is called calibrated Bayes (CB), where inferences for a particular data set are Bayesian, but models are chosen to yield inferences that have good design-based properties. Little argue that CB resolves DMC conflicts, and capitalizes on the strengths of both frequentist and Bayesian approaches. Features of the CB approach to surveys include the incorporation of survey design information into the model, and models with weak prior distributions that avoid strong parametric assumptions. Savitsky and Toth (2016) propose to construct a pseudo-posterior distribution that utilizes sampling weights based on the marginal inclusion probabilities to exponentiate the likelihood contribution of each sampled unit, which weights the information in the sample back to the population. They construct conditions on known marginal and pairwise inclusion probabilities that define a class of sampling designs where consistency of the pseudo posterior is guaranteed.

In this paper we will account for informative probability sampling design in the Bayesian approach to sample survey inference, for example Bayes estimation of the parameters of superpopulation model and prediction of finite population total.

The plan of the paper is as follows, Section 2 is devoted to the definitions of sample and sample-complement distributions. In Section 3 we define the sample posterior distribution. In Section 4, we develop the sample posterior likelihood and Bayesian inference for superpopulation parameter. The Bayesian approach are studied for normal and Bernoulli superpoulation models in Sections 5 and 6. In Section 7, we propose a new Bayes prediction of finite population total. Finally some conclusions are presented in Section 8.

2. Sample and Sample-Complement Distributions

Let $U = \{1,...,N\}$ denote a finite population consisting of N units. Let y be the study variable of interest and let y_i be the value of y for the ith population unit. We consider the population values $y_1,...,y_N$ as random variables, which are independent realizations from a distribution with probability density function (pdf) $f_p(y_i \mid \theta)$, indexed by a vector of parameters $\theta \in \Omega$, where Ω is the parameter space of θ. Let $\mathbf{x}_i = (x_{i1},...,x_{ip})'$, $i \in U$ be the values of a vector of auxiliary variables, $x_1,...,x_p$, and $\mathbf{z} = \{z_1,...,z_N\}$ be the values of known design variables, used for the sample selection process not included in the model under consideration. In what follows, we consider a sampling design with selection probabilities $\pi_i = \Pr(i \in s)$, and sampling weight $w_i = 1/\pi_i$; $i = 1,...,N$. In practice, the π_i's may depend on the population values $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. We express this dependence by writing: $\pi_i = \Pr(i \in s \mid \mathbf{x}, \mathbf{y}, \mathbf{z})$ for all units $i \in U$. Since $\pi_1,...,\pi_N$ are defined by the realizations $(\mathbf{x}_i, y_i, \mathbf{z}_i), i = 1,...,N$, therefore they are random realizations defined on the space of possible populations. The sample s consists of the subset of U selected at random by the sampling scheme with inclusion probabilities $\pi_1,...,\pi_N$.

Denote by $\mathbf{I} = (I_1,...,I_N)'$ the N by 1 sample indicator (vector) variable, such that $I_i = 1$

if unit $i \in U$ is selected to the sample and $I_i = 0$ if otherwise. The sample s is defined accordingly as $s = \{i \mid i \in U, I_i = 1\}$ and its complement by $c = \bar{s} = \{i \mid i \in U, I_i = 0\}$. We assume probability sampling, so that $\pi_i = \Pr(i \in s) > 0$ for all units $i \in U$.

Let f_p and $E_p(\cdot)$ denote the pdf and the mathematical expectation of the population distribution, respectively; f_s and $E_s(\cdot)$ denote the pdf and the mathematical expectation of the sample distribution, respectively; and f_c and $E_c(\cdot)$ denote the pdf and the mathematical expectation of the sample-complement distribution. Assume that the population pdf depends on known values of the auxiliary variables $\mathbf{x}_i$, so that $y_i \sim f_p(y_i \mid \mathbf{x}_i, \theta_p)$. According to Pfeffermann, Krieger and Rinott, Y. (1998), the (marginal) sample pdf of y_i is given by:

$$f_s(y_i \mid \mathbf{x}_i, \theta, \gamma) = f_p(y_i \mid \mathbf{x}_i, \theta, \gamma, i \in s)$$
$$= \frac{E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma)} \tag{1}$$

where

$$E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma) = \int E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta) dy_i$$

and γ is the informativeness parameter.

Similarly, Sverchkov and Pfeffermann (2004) define the sample-complement pdf of y_i as:

$$f_c(y_i \mid \mathbf{x}_i) = \frac{E_p(1 - \pi_i \mid \mathbf{x}_i, y_i) f_p(y_i \mid \mathbf{x}_i)}{E_p(1 - \pi_i \mid \mathbf{x}_i)} \tag{2}$$

For vector of random variables $(y_i, \mathbf{x}_i)$, the following relationships hold:

$$E_s(w_i \mid y_i) = \{E_p(\pi_i \mid y_i)\}^{-1} \tag{3a}$$
$$E_p(y_i \mid \mathbf{x}_i) = \{E_s(w_i \mid \mathbf{x}_i)\}^{-1} E_s(w_i y_i \mid \mathbf{x}_i) \tag{3b}$$
$$E_s(w_i) = \{E_p(\pi_i)\}^{-1} \tag{3c}$$
$$E_c(y_i \mid \mathbf{x}_i) = \frac{E_p\{(1 - \pi_i) y_i \mid x_i\}}{E_p\{(1 - \pi_i) \mid x_i\}} = \frac{E_s\{(w_i - 1) y_i \mid x_i\}}{E_s\{(w_i - 1) \mid x_i\}} \tag{3d}$$

See Pfeffermann and Sverchkov (1999) and Sverchkov and Pfeffermann (2004).

Having derived the sample distribution, Pfeffermann, Krieger and Rinott (1998) proved that if the population measurements y_i are independent, then as $N \to \infty$ (with n fixed) the sample measurements are asymptotically independent, so we can apply standard inference procedures to complex survey data by using the marginal sample distribution for each unit. Based on the sample data $\{y_i, \mathbf{x}_i, w_i; i \in s\}$, Pfeffermann, Krieger and Rinott (1998) proposed a two-step estimation method:

Step-one: Estimate the informativeness parameters γ using equation (3a). Denoting the resulting estimate of γ by $\tilde{\gamma}$.

Step-two: Substitute $\tilde{\gamma}$ in the sample log-likelihood function, and then maximize the resulting sample log-likelihood function with respect to the population parameters, θ:

$$l_{rs}\left(\theta,\tilde{\gamma}\right)=l_{srs}\left(\theta\right)-\sum_{i=1}^{n}\log E_{p}\left(\pi_{i}\mid \mathbf{x}_{i},\theta,\tilde{\gamma}\right)=l_{srs}\left(\theta\right)+\sum_{i=1}^{n}\log E_{s}\left(w_{i}\mid \mathbf{x}_{i},\theta,\tilde{\gamma}\right) \tag{4}$$

where $l_{rs}\left(\theta,\tilde{\gamma}\right)$ is the sample log-likelihood after substituting $\tilde{\gamma}$ in the sample log-likelihood function and where $l_{srs}\left(\theta\right)=\sum_{i=1}^{n}\log\left\{f_{p}\left(y_{i}\mid \mathbf{x}_{i},\theta\right)\right\}$ is the classical log-likelihood.

For more discussion on the use of sample and sample-complement distributions for analytic inference, see Pfeffermann and Sverchkov (1999, 2003), Sverchkov and Pfeffermann (2004), Eideh and Nathan (2009) and Eideh (2012a, 2012b).

3. Sample Posterior Distribution

In this section we describe Bayesian approach to the problem of estimation. This approach takes into account any prior knowledge of the survey that the sampler has. Let $q(\theta)$ be the prior pdf of Θ, here we look upon θ as a possible value of Θ. Assume that the sample conditional pdf of y_i given $\Theta=\theta$ is $f_s\left(y_i\mid\theta\right)$.

The informativess parameter (or the parameter of sample selection process) γ that index the selection model is a characteristic of the data collection but is not generally of scientific interestis and it is not identifiable (see Nandram et al. (2006)), therefore in this paper we treat γ as fixed in repeated sampling. So that $f_s\left(y_i\mid \mathbf{x}_i,\theta,\gamma\right)$ can be written as $f_s\left(y_i\mid\theta\right)$. For more information on estimation of γ, see Eideh (2012a).

We next define the sample joint pdf of y_i and $\Theta=\theta$, Bayes sample marginal or Bayes sample predictive pdf of y_i, and sample posterior distribution.

(a) The sample joint pdf of y_i and Θ denoted by, $m_s\left(y_i,\theta\right)$ is defined by:

$$m_s\left(y_i,\theta\right)=q(\theta)f_s\left(y_i\mid\theta\right) \tag{5}$$

(b) An important quantity in Bayesian inferences and statistical decision theory is the marginal or Bayes predictive distribution of Y, which plays an important role in various scientific applications. The Bayes sample marginal or Bayes sample predictive or Bayes sample predictive pdf of y_i, denoted by, $m_s\left(y_i\right)$ is given by:

$$m_s\left(y_i\right)=\begin{cases}\displaystyle\int_{\Theta}m_s\left(y_i,\theta\right)d\theta=\int_{\Theta}q(\theta)f_s\left(y_i\mid\theta\right)d\theta & \text{if }\Theta\text{ is continuous}\\[2mm]\displaystyle\sum_{\Theta}m_s\left(y_i,\theta\right)=\sum_{\Theta}q(\theta)f_s\left(y_i\mid\theta\right) & \text{if }\Theta\text{ is discrete}\end{cases} \tag{6}$$

The Bayes sample predictive pdf of y_i presents the evidence for a particular model, defined by prior distribution $q(\theta)$ known as prior predictive as it represents the probability of observing the data y_i that was observed before it was observed.

(c) Now we introduce the sample posterior distribution, denoted by $q_s(\theta \mid y_i)$, which is a way, after (posterior to) observing y_i, for adjusting or updating the prior distribution $q(\theta)$ (the beliefs about $\theta \in \Omega$ prior to experimentation or prior to survey) to $q_s(\theta \mid y_i)$. Accordingly, all parametric statistical inferences about the parameters of interest, say $\theta \in \Omega$, are derived from the posterior sample distributions so it represents a complete solution to the inference problem.

The sample posterior distribution can be viewed as a way of combining the prior information $q(\theta)$ and the sample information $f_s(y \mid \theta)$, in other words, $q_s(\theta \mid x)$ contains all of the information combining prior knowledge and observations. The sample posterior distribution of $\Theta = \theta$ given y_i, denoted by $q_s(\theta \mid y_i)$ is defined by:

$$
\begin{aligned}
q_s(\theta \mid y_i) &= \frac{m_s(y_i, \theta)}{m_s(y_i)} = \frac{q(\theta) f_s(y_i \mid \theta)}{m_s(y_i)} \\
&= \frac{q(\theta) E_p(\pi_i \mid y_i, \gamma) f_p(y_i \mid \theta)}{m_s(y_i) E_p(\pi_i \mid \theta)} \\
&= \frac{q(\theta) \Pr(i \in s \mid y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{m_s(y_i) \Pr(i \in s \mid \gamma)}
\end{aligned}
\tag{7}
$$

Note that:

(i) If the sampling design is noninformative, that is, $\Pr(i \in s \mid y_i, \gamma) = \Pr(i \in s \mid \gamma)$ for all y_i, then the sample joint pdf of y_i and Θ is the same as the population joint pdf of y_i and Θ, the Bayes sample predictive pdf of y_i posterior distribution is the same as the Bayes population predictive pdf of y_i, and the sample posterior distribution of $\Theta = \theta$ given y_i is the same as the population posterior distribution of $\Theta = \theta$ given y_i. So that data collection method (sampling design) does not influence Bayesian inference.

(ii) The Bayes sample predictive pdf of y_i, $m_s(y_i)$, represents the normalizing constant in Bayes theorem, so that

$$
q_s(\theta \mid y_i) = \frac{q(\theta) f_s(y_i \mid \theta)}{m_s(y_i)}
\tag{8}
$$

is called normalized sample posterior distribution, while

$$
q_s(\theta \mid y_i) \propto q(\theta) f_s(y_i \mid \theta)
\tag{9}
$$

is the unnormalized one.

(iii)   The sample posterior distribution (the distribution of Θ after the sample is drawn) can be viewed as a way of combining the prior information $q(\theta)$ (which reflects the subjective belief of Θ before the sample is drawn), the sampling design, $\Pr(i \in \text{s} \mid y_i, \gamma)$, and the population information $f_p(y_i \mid \theta)$. In other words, $q_s(\theta \mid y_i)$ contains all of the information combining prior knowledge, sampling design and observations.

(iv)   With fixed $f_p(y_i \mid \theta)$ and $q(\theta)$, the sample posterior distribution is completely determined by $E_p(\pi_i \mid y_i, \gamma)$. For more information on this issue, see Pfeffermann, Krieger and Rinott (1998), and Eideh (2010). Hence, sample posterior distribution combine: population distribution $f_p(y_i \mid \theta)$, prior distribution $q(\theta)$ and $E_p(\pi_i \mid y_i, \gamma)$.

(v)   Furthermore,

$$\frac{q_s(\theta_1 \mid y_i)}{q_s(\theta_2 \mid y_i)} = \frac{q(\theta_1)}{q(\theta_2)} \frac{q(\theta_1) f_s(y_i \mid \theta_1)}{q(\theta_2) f_s(y_i \mid \theta_2)} \tag{10}$$

That is, the sample posterior odds are equal to prior odds, $q(\theta_1)/q(\theta_2)$, multiplied by the sample likelihood ratio,

$$\frac{L_s(\theta_1 \mid y_i)}{L_s(\theta_2 \mid y_i)} = \frac{f_s(y_i \mid \theta_1)}{f_s(y_i \mid \theta_2)} \tag{11}$$

4.   Sample Posterior Likelihood and Bayesian Inference for Superpopulation Parameter

Data collected by sample surveys, are used extensively to make inferences on assumed population models. Often, survey design features (clustering, stratification, unequal probability selection, etc.) are ignored and the sample data are then analyzed using classical methods based on simple random sampling. In this section we account for informative sampling design in the analysis of survey data using Bayesian inference.

If $\mathbf{y}_s = (y_1, \ldots y_n)'$ are independent and identical random variables from $f_s(y_i \mid \theta)$, then we can write the sample joint conditional pdf of $\mathbf{y}_s = (y_1, \ldots y_n)'$ given $\Theta = \theta$, as:

$$f_s(\mathbf{y}_s \mid \theta) = L_s(\mathbf{y}_s \mid \theta) = \prod_{i=1}^{n} f_s(y_i \mid \theta) \tag{12}$$

Thus sample joint pdf of $\mathbf{y}_s = (y_1, \ldots y_n)'$ and Θ is:

$$m_s(\mathbf{y}_s, \theta) = q(\theta) L_s(\mathbf{y}_s \mid \theta) \tag{13}$$

If Θ is a continuous random variable, then Bayes sample predictive pdf of $\mathbf{y}_s = (y_1, \ldots y_n)'$ is:

$$m_s(\mathbf{y}_s) = \int_\Theta m_s(\mathbf{y}_s, \theta) d\theta = \int_\Theta q(\theta) L_s(\mathbf{y}_s | \theta) d\theta \tag{14}$$

Hence the conditional sample pdf of $\Theta_s | \mathbf{y}_s$ or the sample posterior likelihood:

$$\begin{aligned}
q_s(\theta | \mathbf{y}_s) &= L_s(\theta | \mathbf{y}_s) \\
&= \frac{q(\theta) L_s(\mathbf{y}_s | \theta)}{m_s(\mathbf{y}_s)} \\
&\propto q(\theta) L_s(\mathbf{y}_s | \theta)
\end{aligned} \tag{15}$$

Now we are in a position to deal with Bayesian inference, via the sample posterior distribution and the sample posterior likelihood. This provides the following representation of the sample posterior pdf of $\Theta | \mathbf{y}_s$:

Sample Posterior information = Prior information + Sampling design information + Sampling information

And in terms of pdfs:

Sample Posterior pdf = Prior pdf+ Sample Likelihood function
= Prior pdf+ sampling design+ population pdf

Therefore the sample posterior pdf of $\Theta | \mathbf{y}_s$ summarizes the total information, after viewing the sample data, collected under informative sampling design, and provides a basis for sample posterior or Bayesian inferences concerning the population parameters of interest.

The generalized maximum likelihood estimator (MLE) $\hat{\theta}_g \in \Omega$ of $\theta \in \Omega$ is the value of θ that maximizes the sample posterior likelihood function, $L_s(\theta | \mathbf{y}_s)$. That is, $\hat{\theta}_g \in \Omega$ is the most likely value of $\theta \in \Omega$, given the prior and the sample data $\mathbf{y}_s = (y_1, \ldots y_n)'$. Obviously $\hat{\theta}_g \in \Omega$ has the interpretation of being the "most likely" value of θ given the prior and the sample data $\mathbf{y}_s = (y_1, \ldots y_n)'$.

From Bayesian viewpoint, estimating the population parameter θ is obtained by choosing a decision function $\delta(\mathbf{y}_s)$, which depends on a loss function $L[\delta(\mathbf{y}_s), \theta]$, whose conditional expectation under the sample posterior distribution is minimum. That is,

$$\begin{aligned}
\delta(\mathbf{y}_s) &= \operatorname{argmin} E_s \{ L[\delta(\mathbf{y}_s), \theta] | \mathbf{y}_s \} \\
&= \operatorname{argmin} \int_\Theta L[\delta(\mathbf{y}_s), \theta] q_s(\theta | \mathbf{y}_s) d\theta
\end{aligned} \tag{16}$$

If $L[\delta(\mathbf{y}_s),\theta] = [\delta(\mathbf{y}_s) - \theta]^2$ (squared error loss function), then it is easy to show that, the sample Bayes estimate of $\theta \in \Omega$ is given by:

$$\delta(\mathbf{y}_s) = \hat{\theta}_B = E_s(\Theta|\mathbf{y}_s) \tag{17}$$

and the sample Bayes estimate of $\beta = \beta(\Theta) \in \Omega$ is given by:

$$\hat{\beta}_B = E_s\{\beta(\Theta)|\mathbf{y}_s\} \tag{18}$$

Moreover, the $100(1-\alpha)\%$ highest sample posterior density (HSPD) credible set for $\theta \in \Omega$, is the subset C of Θ of the form:

$$C = \{\theta \in \Theta : q_s(\theta \mid \mathbf{y}_s) \ge k(\alpha)\} \tag{19}$$

where $\mathbf{y}_s = (y_1,\dots y_n)'$, and $k(\alpha)$ is the largest constant such that:

$$P(C|\mathbf{y}_s) \ge 1 - \alpha \tag{20}$$

5. Application 1 – Normal Superpopulation Model

According to the definition of sample posterior distribution, to consider the Bayesian inference, we need to specify: $f_p(y_i \mid \theta)$ and $q(\theta)$, and $E_p(\pi_i \mid y_i, \gamma)$.

From now on, we denote the parameter index the population distribution by θ_p, and the parameter index the sample distribution by θ_s. Consider the following population model:

$$y_i \mid \Theta_p = \theta_p \sim N(\theta_p, \sigma^2) \tag{21}$$

be independent random variables, $i = 1,\dots,N$, where $\sigma^2 > 0$ is known, $i = 1,\dots,N$. Let $\mathbf{y}_s = (y_1,\dots y_n)'$ be the sample data. Suppose that

$$E_p(\pi_i|y_i) = \exp(a_1 y_i) \tag{22}$$

Using equation (1), it is easy to verify that the sample model is:

$$y_i \mid \Theta_s = \theta_s \sim N(\theta_s, \sigma^2) \tag{23}$$

where $\theta_s = \theta_p + a_1\sigma^2 = \theta_s(\theta_p, a_1, \sigma^2)$ is a linear function of θ_p. So in order to find the Bayes estimator of θ_p, we (for simplicity and to makes life easy) choose a prior distribution for θ_s and then derive the Bayes estimator for θ_p, via the relationship between θ_p and θ_s. We consider different cases based on prior distribution.

Case 1: Assume that the prior distribution for $\Theta_s \sim N(\mu,\tau^2)$ with known (μ,τ^2). Hence in this case and the subsequent cases, $(a_1, \sigma^2, \tau^2, \mu)$ is know, and the only unknown parameter is θ_p. Using the formulas in the previous sections, and doing some algebra we have:

(a) The sample posterior distribution of $\Theta_s|\mathbf{y}_s$ is given by:

$$\Theta_s|\mathbf{y}_s \sim N\left(\frac{\mu\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2}, \frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2}\right) \tag{24}$$

where $\bar{y}_s = n^{-1}\sum_{i=1}^{n} y_i$.

so that

$$E_s\left(\Theta_s|\mathbf{y}_s\right) = \frac{\mu\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2} \tag{25}$$

and

$$V_s\left(\Theta_s|\mathbf{y}_s\right) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2} = \frac{\left(\sigma^2/n\right)}{\left(\sigma^2/n\right) + \tau^2}\tau^2 \tag{26}$$

Now, for given $\left(a_1, \sigma^2, \tau^2, \mu\right)$ and using $\Theta_s = \Theta_p + a_1\sigma^2$, we have

$$\begin{aligned} E_s\left(\Theta_p|\mathbf{y}_s\right) &= E_s\left(\Theta_s - a_1\sigma^2|\mathbf{y}_s\right) \\ &= \frac{\mu\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2} - a_1\sigma^2 \end{aligned} \tag{27}$$

and

$$\begin{aligned} V_s\left(\Theta_p|\mathbf{y}_s\right) &= V_s\left(\Theta_s - a_1\sigma^2|\mathbf{y}_s\right) \\ &= V_s\left(\Theta_s|\mathbf{y}_s\right) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2} \end{aligned} \tag{28}$$

Furthermore, the sample posterior distribution of $\Theta_p|\mathbf{y}_s$ is given by:

$$\Theta_p|\mathbf{y}_s \sim N\left(\frac{\mu\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2} - a_1\sigma^2, \frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2}\right) \tag{29}$$

Hence the sample and population posterior pdfs of $\Theta_p|\mathbf{y}_s$ belong to the same family of normal distributions, but the mean under the sample posterior pdf is different from the mean under the population posterior pdf of $\Theta_p|\mathbf{y}_s$. Furthermore if $a_1 = 0$, that is the sampling design is ignorable, then the sample and population posterior pdfs are similar.

(b) For given $\left(a_1, \sigma^2, \tau^2, \mu\right)$, the Bayes estimate of $\theta_p \in \Omega$ under the sample posterior pdf is:

$$\begin{aligned} \hat{\theta}_{pB} = E_s\left(\Theta_p|\mathbf{y}_s\right) &= \frac{\mu\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2} - a_1\sigma^2 \\ &= \phi\mu + (1-\phi)\bar{y}_s - a_1\sigma = \phi\mu + (1-\phi)\bar{y}_s - a_1\sigma^2 \\ &= \hat{\theta}_{pB}(srs) - a_1\sigma^2 \end{aligned} \tag{30}$$

where

$$\phi = \frac{\left(\sigma^2/n\right)}{\left(\sigma^2/n\right)+\tau^2} \tag{31}$$

and

$$\hat{\theta}_{pB}\left(srs\right)= \phi\mu + \left(1-\phi\right)\bar{y}_s \tag{32}$$

is the Bayes estimate of $\theta_p \in \Omega_p$ under noninformative sampling design.

Note that:

(1)   If $a_1 = 0$ (that is, sampling design is noninformative), then the Bayes estimate of $\theta_p \in \Omega$ based on the population model and sample model are coincide.

(2)   If $a_1 > 0$ and $y_i > 0$, then $E_p\left(\pi_i \mid y_i\right)= e^{a_1 y_i}$ is an increasing function of y_i, so that larger values are more likely to be in the sample than smaller values. In this case, $\hat{\theta}_{pB}\left(srs\right)$ overestimate $\theta_p \in \Omega$. The Bayes estimate $\hat{\theta}_{pB} = \hat{\theta}_{pB}\left(srs\right)-a_1\sigma^2$ adjusts $\hat{\theta}_{pB}\left(srs\right)$ by the positive quantity $a_1\sigma^2$.

(3)   If $a_1 < 0$ and $y_i > 0$, then $E_p\left(\pi_i \mid y_i\right)= e^{a_1 y_i}$ is a decreasing function of y_i, so that smaller values are more likely to be in the sample than larger values. In this case, $\hat{\theta}_{pB}\left(srs\right)$ underestimate $\theta_p \in \Omega_p$. The Bayes estimate $\hat{\theta}_{pB} = \hat{\theta}_{pB}\left(srs\right)-a_1\sigma^2$ adjusts $\hat{\theta}_{pB}\left(srs\right)$ by the negative quantity $a_1\sigma^2$.

(4) $\hat{\theta}_{pB} + a_1\sigma^2 = \hat{\theta}_{sB}$ is a weighted average of μ (prior mean) and $\bar{y}_s$ (the maximum likelihood estimator of $\theta_s \in \Omega$), with weights ϕ and $1-\phi$, respectively.

(5) $\hat{\theta}_{pB}$ can be written as:

$$\hat{\theta}_{pB} = \frac{\tau^2}{\left(\sigma^2/n\right)+\tau^2}\bar{y}_s + \frac{\left(\sigma^2/n\right)}{\left(\sigma^2/n\right)+\tau^2}\mu - a_1\sigma^2 \tag{33}$$

so that,

$$\lim_{n\to\infty}\hat{\theta}_{pB} = \lim_{n\to\infty}\left\{\frac{\tau^2}{\left(\sigma^2/n\right)+\tau^2}\bar{y}_s + \frac{\left(\sigma^2/n\right)}{\left(\sigma^2/n\right)+\tau^2}\mu - a_1\sigma^2\right\}$$
$$= \bar{y}_s - a_1\sigma^2 \tag{34}$$

Hence, for large sample size n, the Bayes estimate of $\theta_p \in \Omega_p$, shrinkage to $\bar{y}_s - a_1\sigma^2$. That is, as the sample size n increases, the influence of the prior distribution on posterior inferences decreases.

This idea, sometimes referred to as asymptotic theory, because it refers to properties that hold in the limit as n becomes large. The large-sample results are not actually necessary for performing Bayesian data analysis but are often useful as approximations and as tools for understanding. See Gelman et al. 2009.

(c) It is easy to verify that for given $\left(a_1, \sigma^2, \tau^2, \mu\right)$, the generalized MLE of $\theta_p \in \Omega$ is given by:

$$
\begin{aligned}
\hat{\theta}_{pg} &= \bar{y}_s - \phi(\bar{y}_s - \mu) - a_1\sigma^2 \\
&= \phi\mu + (1-\phi)\bar{y}_s - a_1\sigma^2
\end{aligned}
\tag{35}
$$

which is the sample Bayes estimate of $\theta_p \in \Omega_p$.

(d) Furthermore, the $100(1-\alpha)\%$ highest sample posterior density (HSPD) credible set for θ_p, is given by:

$$
C = \left(\left(\phi\mu + (1-\phi)\bar{y}_s - a_1\sigma^2\right) \pm z_{\alpha/2} \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}} \right)
\tag{36}
$$

where is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution.

Note that if $a_1 = 0$ (that sampling design is noninformative), then the $100(1-\alpha)\%$ highest sample posterior density (HSPD) credible set for θ, is given by:

$$
C = \left(\phi\mu + (1-\phi)(\bar{y}_s) \pm z_{\alpha/2} \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}} \right)
\tag{37}
$$

which is the classical $100(1-\alpha)\%$ highest posterior density (HPD) credible set for θ_p.

Not that $\{q(\theta_s); \theta_s \in \Omega\}$ is a conjugate family for $\{f_s(y_i \mid \theta_s); \theta_s \in \Omega\}$, because $q_s(\theta_s \mid y_i)$ is in the class of $\{q(\theta_s); \theta_s \in \Omega\}$.

Case 2: Assume that the prior distribution for $\Theta_s = \Theta_p + a_1\sigma^2 \sim N\left(\mu + a_1\sigma^2, \tau^2\right)$ with known $\left(\mu, \tau^2, a_1\right)$. Under this case we have:

$$
\Theta_s \mid \mathbf{y}_s \sim N\left(\frac{\left(\mu + a_1\sigma^2\right)\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2}, \frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2} \right)
\tag{38}
$$

$$
\begin{aligned}
E_s\left(\Theta_s \mid \mathbf{y}_s\right) &= \phi\mu + (1-\phi)(\bar{y}_s) + \phi a_1\sigma^2 \\
&= E_p\left(\Theta_p \mid \mathbf{y}_s\right) + \phi a_1\sigma^2
\end{aligned}
\tag{39}
$$

$$
V_s\left(\Theta_s \mid \mathbf{y}_s\right) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2} = V_p\left(\Theta_p \mid \mathbf{y}_s\right)
\tag{40}
$$

$$
E_s\left(\Theta_p \mid \mathbf{y}_s\right) = E_s\left(\Theta_s \mid \mathbf{y}_s\right) - a_1\sigma^2 = \phi\mu + (1-\phi)(\bar{y}_s - a_1\sigma^2)
\tag{41}
$$

For given $\left(a_1, \sigma^2, \tau^2, \mu\right)$, the Bayes estimator of $\theta_s \in \Omega$ under the sample posterior pdf is:

$$\hat{\theta}_{sB} = E_s\left(\Theta_s | \mathbf{y}_s\right) = \phi\mu + (1-\phi)\left(\bar{y}_s\right) + \phi a_1 \sigma^2 \tag{42}$$

where

$$\phi = \left(\frac{\sigma^2/n}{\sigma^2/n + \tau^2}\right)$$

Hence, the Bayes estimator of $\theta_p \in \Omega$ is:

$$\begin{aligned}
\hat{\theta}_p &= \bar{y}_s - \phi\left(\bar{y}_s - \mu + n a_1 \tau^2\right) \\
&= \phi\mu + (1-\phi)\left(\bar{y}_s - a_1\sigma^2\right)
\end{aligned} \tag{43}$$

It is easy to verify that for given $\left(a_1, \sigma^2, \tau^2, \mu\right)$, the generalized MLE of $\theta_s \in \Omega_s$ is given by:

$$\hat{\theta}_{sg} = \phi\mu + (1-\phi)\left(\bar{y}_s\right) + \phi a_1 \sigma^2 \tag{44}$$

So that:

$$\begin{aligned}
\hat{\theta}_{pg} &= \bar{y}_s - \phi\left(\bar{y}_s - \mu + n a_1 \tau^2\right) \\
&= \phi\mu + (1-\phi)\left(\bar{y}_s - a_1\sigma^2\right)
\end{aligned} \tag{45}$$

The $100(1-\alpha)\%$ highest sample posterior density (HSPD) credible set for $\theta_s \in \Omega$, is given by:

$$C_s = \left(\phi\mu + (1-\phi)\left(\bar{y}_s\right) + \phi a_1 \sigma^2 \pm z_{\alpha/2}\sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}}\right) \tag{46}$$

So that for $\theta_p = \theta_s - a_1\sigma^2 \in \Omega$ is:

$$C = \left(\phi\mu + (1-\phi)\left(\bar{y}_s - a_1\sigma^2\right) \pm z_{\alpha/2}\sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}}\right) \tag{47}$$

Not that $\left\{q(\theta_s); \theta_s \in \Omega\right\}$ is a conjugate family for $\left\{f_s(y_i | \theta_s); \theta_s \in \Omega\right\}$, because $q_s\left(\theta_s | y_i\right)$ is in the class of $\left\{q(\theta_s); \theta_s \in \Omega\right\}$.

Case 3: Assume that the prior distribution for $\Theta_p \sim N\left(\mu, \tau^2\right)$ with known $\left(\mu, \tau^2\right)$. Similar to Cases 1 and 2, we have:

$$\Theta_p | \mathbf{y}_s \sim N\left(\frac{\mu\sigma^2 + n\tau^2\bar{y}_s - n a_1\tau^2\sigma^2}{\sigma^2 + n\tau^2}, \frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2}\right) \tag{48}$$

so that

$$E_s\left(\Theta_p\middle|\mathbf{y}_s\right) = \frac{\mu\sigma^2 + n\tau^2\bar{y}_s - na_1\tau^2\sigma^2}{\sigma^2 + n\tau^2}$$

$$= \frac{\mu\sigma^2 + n\tau^2\bar{y}_s}{\sigma^2 + n\tau^2} - na_1\frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2} \qquad (49)$$

$$= E_p\left(\Theta_p\middle|\mathbf{y}_s\right) - na_1V_p\left(\Theta_p\middle|\mathbf{y}_s\right)$$

and

$$V_s\left(\Theta_p\middle|\mathbf{y}_s\right) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}$$

$$= \frac{\left(\sigma^2/n\right)}{\left(\sigma^2/n\right) + \tau^2}\tau^2 = V_p\left(\Theta_p\middle|\mathbf{y}_s\right) \qquad (50)$$

which is similar to Case 2 (a).

(b) For given $\left(a_1, \sigma^2, \tau^2, \mu\right)$, the Bayes estimator of $\theta_p \in \Omega$ under the sample posterior pdf is:

$$\hat{\theta}_{pB} = E_s\left(\Theta_p\middle|\mathbf{y}_s\right) = \phi\mu + (1-\phi)\bar{y}_s - (1-\phi)a_1\sigma^2$$

$$= \phi\mu + (1-\phi)\left(\bar{y}_s - a_1\sigma^2\right) \qquad (51)$$

where

$$\phi = \left(\frac{\sigma^2/n}{\sigma^2/n + \tau^2}\right)$$

which is similar to Case 2 (b).

Note that, here, $\hat{\theta}_{pB}$ is a weighted average of μ (prior mean) and $\left(\bar{y}_s - a_1\sigma^2\right)$ (the maximum likelihood estimator of $\theta_p \in \Omega_p$), with weights ϕ and $1-\phi$, respectively.

(c) It is easy to verify that for given $\left(a_1, \sigma^2, \tau^2, \mu\right)$, the generalized MLE of $\theta_p \in \Omega$ is given by:

$$\hat{\theta}_{pg} = \bar{y}_s - \phi\left(\bar{y}_s - \mu + na_1\tau^2\right) \qquad (52)$$

which is the sample Bayes estimate of $\theta_p \in \Omega$ - similar to Case 2 (c).

(d) The $100(1-\alpha)\%$ highest sample posterior density (HSPD) credible set for θ_p, is given by:

$$C = \left(\phi\mu + (1-\phi)\left(\bar{y}_s - a_1\sigma^2\right) \pm z_{\alpha/2}\sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}}\right) \qquad (53)$$

which is similar to Case 2 (d).

340

Pak.j.stat.oper.res. Vol.XIII No.2 2017 pp327-353

Note that if $a_1 = 0$ (that sampling design is noninformative), then the $100(1-\alpha)\%$ highest sample posterior density (HSPD) credible set for θ_p, is given by

$$C = \left(\phi\mu + (1-\phi)(\bar{y}_s) \pm z_{\alpha/2} \sqrt{\frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2}} \right) \tag{54}$$

which is the $100(1-\alpha)\%$ highest posterior density (HPD) credible set for θ_p.

Hence, the prior distribution for $\Theta_p \sim N(\mu, \tau^2)$ with known (μ, τ^2) is equivalent to the prior distribution $\Theta_s = \Theta_p + a_1\sigma^2 \sim N(\mu + a_1\sigma^2, \tau^2)$ with known (μ, τ^2, a_1), in the sense that, they give the same predictor of $\theta_p \in \Omega$.

Not that $\{q(\theta_p); \theta_p \in \Omega\}$ is a conjugate family for $\{f_s(y_i \mid \theta_s); \theta_s \in \Omega\}$, because $q_s(\theta_p \mid y_i)$ is in the class of $\{q(\theta_p); \theta_p \in \Omega\}$.

It should be noted that, the similarity between Case 2 and Case 3 is interpreted as follows: since under Case 3, $\Theta_p \sim N(\mu, \tau^2)$, therefore $\Theta_s = \Theta_p + a_1\sigma^2 \sim N(\mu + a_1\sigma^2, \tau^2)$, which is the prior distribution under Case 2. The situation is different under Case 1, where $\Theta_s = \Theta_p + a_1\sigma^2 \sim N(\mu, \tau^2)$.

Case 4: Suppose that the prior distribution of θ_s is noninformative or improper:

$$q(\theta_s) = 1, \ -\infty < \theta_s < \infty \tag{55}$$

(a) Since $\bar{y}_s$ is a sufficient statistics for $\theta_s = \theta_p + a_1\sigma^2$, therefore, the conditional sample pdf of $\Theta_s \mid y_1, \ldots, y_n$ or the sample posterior likelihood is:

$$\begin{aligned} q_s(\theta_s \mid \mathbf{y}_s) &= q(\theta_s) L_s(\theta_s \mid \mathbf{y}_s) \\ &= q_s(\theta \mid \bar{y}_s) q(\theta_s) = L_s(\theta \mid \bar{y}_s) \\ &= \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left[-\frac{n}{2\sigma^2} (\bar{y}_s - \theta_s)^2 \right] \end{aligned} \tag{56}$$

Hence the posterior distribution of $\Theta_s \mid \mathbf{y}_s$ is given by:

$$\Theta_s \mid \mathbf{y}_s \sim N\left(\bar{y}_s, \frac{\sigma^2}{n} \right) \tag{57}$$

so that

$$E_s(\Theta_s \mid \mathbf{y}_s) = \bar{y}_s \tag{58}$$

and

$$V_s(\Theta_s \mid \mathbf{y}_s) = \frac{\sigma^2}{n} \tag{59}$$

We are interested in sample posterior distribution of $\Theta_p | \mathbf{y}_s$.

Since $\theta_p = \theta_s - a_1 \sigma^2$, therefore, sample posterior distribution of $\Theta_p | \mathbf{y}_s$ is given by:

$$\Theta_p | \mathbf{y}_s \sim N\left(\bar{y}_s - a_1 \sigma^2, \frac{\sigma^2}{n} \right) \tag{60}$$

so that

$$\begin{aligned} E_s\left(\Theta_p | \mathbf{y}_s \right) &= \bar{y}_s - a_1 \sigma^2 \\ &= E_p\left(\Theta_p | \mathbf{y}_s \right) - a_1 \sigma^2 \end{aligned} \tag{61}$$

and

$$V_s\left(\Theta_p | \mathbf{y}_s \right) = \frac{\sigma^2}{n} = V_p\left(\Theta_p | \mathbf{y}_s \right) \tag{62}$$

Note that the sample and population posterior pdfs of $\Theta_p | \mathbf{y}_s$ belong to the same family of normal distributions, but the mean under the sample posterior pdf is different from the mean under the population posterior pdf of $\Theta_p | \mathbf{y}_s$. Furthermore if $a_1 = 0$, that is the sampling design is ignorable, then the sample and population pdfs are similar.

(b) For given $\left(a_1, \sigma^2 \right)$, the Bayes estimator of $\theta_p \in \Omega$ under the sample posterior pdf is

$$\hat{\theta}_{pB} = E_s\left(\Theta_p | \mathbf{y}_s \right) = \bar{y}_s - a_1 \sigma^2 \tag{63}$$

which is the same as the maximum likelihood estimator of $\theta_p \in \Omega$ under the sample model: $y_i | \Theta_s = \theta_s \sim N\left(\theta_s, \sigma^2 \right)$ where $\theta_p = \theta_s - a_1 \sigma^2$.

(c) It is easy to verify that for given $\left(a_1, \sigma^2 \right)$, the generalized MLE of $\theta_p \in \Omega$ is given by:

$$\hat{\theta}_{sg} = \bar{y}_s - a_1 \sigma^2 \tag{64}$$

(d) The $100(1 - \alpha)\%$ highest sample posterior density (HSPD) credible set for θ_p, is given by:

$$C = \left(\left(\bar{y}_s - a_1 \sigma^2 \right) \pm z_{\alpha/2} \sqrt{\frac{\sigma^2}{n}} \right) \tag{65}$$

Note that if $a_1 = 0$ (that sampling design is noninformative), then the $100(1 - \alpha)\%$ highest sample posterior density (HSPD) credible set for θ_p, is given by:

$$C = \left(\left(\bar{y}_s \right) \pm z_{\alpha/2} \sqrt{\frac{\sigma^2}{n}} \right) \tag{66}$$

which is the $100(1 - \alpha)\%$ highest posterior density (HPD) credible set for θ. Also it is the $100(1 - \alpha)\%$ confidence interval for θ_p.

Not that $\{q(\theta_s); \theta_s \in \Omega\}$ is a conjugate family for $\{f_s(y_i \mid \theta_s); \theta_s \in \Omega\}$. Because $q_s(\theta_s \mid y_i)$ is in the class of $\{q(\theta_s); \theta_s \in \Omega\}$.

6. Application 2 – Bernoulli Superpopulation Model

Let $y_i \mid \Theta_p = \theta_p \sim Ber(\theta_p)$ are independent random variables, where $\theta_p \in (0,1)$, $i = 1, \ldots, N$. Suppose that the sample inclusion probabilities have expectations:

$$E_p(\pi_i \mid y_i) = \Pr(i \in s \mid y_i) = \exp(a_0 + a_1 y_i), \, a_0, a_1 \neq 0$$
$$= \begin{cases} \exp(a_0) = \rho_0 & if \ y_i = 0 \\ \exp(a_0 + a_1) = \rho_1 & if \ y_i = 1 \end{cases} \tag{67}$$

So that, $a_0 = \ln \rho_0$ and $a_1 = \ln(\rho_1 / \rho_0)$. In this case ρ_0 and ρ_1 are known before sampling. We can show that the sample distribution of y_i is:

$$f_s(y_i \mid \theta_s) = \frac{\Pr(i \in s \mid y_i) f_p(y_i \mid \theta_p)}{\Pr(i \in s)}$$
$$= \left(\frac{\theta_p \rho_1}{\theta_p \rho_1 + (1 - \theta_p)\rho_0} \right)^{y_i} \left(\frac{(1 - \theta_p)\rho_0}{\theta_p \rho_1 + (1 - \theta_p)\rho_0} \right)^{1 - y_i} \tag{68}$$
$$= (\theta_s)^{y_i} (1 - \theta_s)^{1 - y_i}, \, y_i = 0,1$$

Thus the distribution of $y_i, i \in s$ is the same as the distribution of $y_i, i \in U$, except that the population parameter θ_p changes to:

$$\theta_s = \frac{\theta_p \rho_1}{\theta_p \rho_1 + (1 - \theta_p)\rho_0} = \theta_s(\theta_p, \rho_1, \rho_0) = \theta_s(\theta_p) \tag{69}$$

which is not a linear function of θ_p.

Notice that if $a_1 = 0$, or equivalently $\rho_0 = \rho_1$, that is, the sampling design is ignorable then the sample and population distributions are the same.

Case 1: Now suppose that the prior distribution of θ_p is $\Theta_p \sim B(\alpha, \beta)$, where (α, β) are known. The conditional sample pdf of $\Theta_p \mid \mathbf{y}_s$ or the sample posterior likelihood is given by:

$$q_s(\theta_p \mid \mathbf{y}_s) \propto q(\theta_p) L_s(\mathbf{y}_s \mid \theta_p), \, \theta_p \in \Omega \tag{70}$$

where

$$q(\theta_p) L_s(\mathbf{y}_s \mid \theta_p) = q(\theta_p) \prod_{i=1}^n f_s(y_i \mid \theta_p)$$
$$= \frac{1}{Beta(\alpha, \beta)} \theta_p^{\alpha - 1} (1 - \theta_p)^{\beta - 1} \left(\frac{\theta_p \rho_1}{\theta_p \rho_1 + (1 - \theta_p)\rho_0} \right)^{n \bar{y}_s} \times \left(\frac{(1 - \theta_p)\rho_0}{\theta_p \rho_1 + (1 - \theta_p)\rho_0} \right)^{n - n \bar{y}_{si}} \tag{71}$$

which cannot be written in the form of beta distribution.

Not that $\{q(\theta_p); \theta_p \in \Omega\}$ is not a conjugate family for $\{f_s(y_i \mid \theta_p); \theta_p \in \Omega\}$, because $q_s(\theta_p \mid y_i)$ is not in the class $\{q(\theta_p); \theta_p \in \Omega\}$. Recall that $\theta_s = \{(\theta_p \rho_1)/(\theta_p \rho_1 + (1-\theta_p)\rho_0)\}$ is not a linear function of θ_p.

Case 2: To makes life easy, assume that the prior distribution of θ_s is $\Theta_s \sim B(\alpha, \beta)$. It is easy to verify that, the conditional sample pdf of $\Theta_s \mid \mathbf{y}_s$ or the sample posterior likelihood is given by:

$$q_s(\theta_s \mid \mathbf{y}_s) \propto q(\theta_s) L_s(\mathbf{y}_s \mid \theta_s), \theta_s \in \Omega \tag{72}$$

where

$$q(\theta_s) L_s(\mathbf{y}_s \mid \theta_s) = q(\theta_s) \prod_{i=1}^{n} f_s(y_i \mid \theta_s)$$

$$= \frac{1}{Beta(\alpha, \beta)} \theta_s^{\alpha + n\bar{y}_s 1} (1 - \theta_s)^{\beta + n - n\bar{y}_s - 1} \tag{73}$$

Hence,

$$q_s(\theta_s \mid \mathbf{y}_s) \propto \frac{1}{Beta(\alpha, \beta)} \theta_s^{\alpha + n\bar{y}_s - 1} (1 - \theta_s)^{\beta + n - n\bar{y}_s - 1} \tag{74}$$

After some algebra, we can show that the sample posterior distribution of $\Theta_s \mid \mathbf{y}_s$ is given by:

$$\Theta_s \mid \mathbf{y}_s B(\alpha + n\bar{y}_s, \beta + n - n\bar{y}_s) \tag{75}$$

So that

$$E_s(\Theta_s \mid \mathbf{y}_s) = \frac{\alpha + n\bar{y}_s}{\alpha + n\bar{y}_s + n - n\bar{y}_s + \beta}$$

$$= \frac{\alpha + n\bar{y}_s}{\alpha + \beta + n} \tag{76}$$

and

$$V_s(\Theta_s \mid \mathbf{y}_s) = \frac{(\alpha + n\bar{y}_s)(n - n\bar{y}_s + \beta)}{(\alpha + \beta + n)^2 (\alpha + \beta + n + 1)} \tag{77}$$

Not that $\{q(\theta_s); \theta_s \in \Omega\}$ is a conjugate family for $\{f_s(y_i \mid \theta_s); \theta_s \in \Omega\}$. Because $q_s(\theta_s \mid y_i)$ is in the class $\{q(\theta_s); \theta_s \in \Omega\}$.

Hence, for given $(\alpha, \beta, \rho_0, \rho_1)$, the Bayes estimate of $\theta_s \in \Omega$ under the sample posterior pdf is given by:

$$\hat{\theta}_{sB} = E_s(\Theta_s \mid \mathbf{y}_s) = \frac{\alpha + n\bar{y}_s}{\alpha + \beta + n} \tag{78}$$

But we are interested in the Bayes estimate of $\theta_p \in \Omega$. For this, using the relationship between θ_s and θ_p, we get:

$$
\begin{aligned}
\hat{\theta}_{pB} &= \frac{\hat{\theta}_{sB}\rho_0}{\hat{\theta}_{sB}\rho_0 + (1 - \hat{\theta}_{sB})\rho_1} \\
&= \frac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)}
\end{aligned}
\tag{79}
$$

Notice that if $\rho_0 = \rho_1$, that is, the sampling design is ignorable, then

$$
\hat{\theta}_{pB} = \frac{\alpha + n\bar{y}_s}{\alpha + \beta + n} = \hat{\theta}_{pB}(srs)
\tag{80}
$$

which is the classical Bayes estimate of θ_p under simple random sample.

The Bayes estimate $\hat{\theta}_{pB}$ can be written as:

$$
\hat{\theta}_{pB} = \left(\frac{n\rho_0}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)\bar{y}_s + \left(\frac{\rho_0(\alpha + \beta)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)\frac{\alpha}{\alpha + \beta}
\tag{81}
$$

which is a linear combination of $\bar{y}_s$ - maximum likelihood estimator of θ_p, under the population model $y_i | \Theta_p = \theta_p \sim Ber(\theta_p)$, and $(\alpha/\alpha + \beta)$ - mean of $\Theta_s : B(\alpha, \beta)$.

Furthermore, we have:

$$
\begin{aligned}
\lim_{n \to \infty} \hat{\theta}_{pB} &= \lim_{n \to \infty} \left\{ \begin{array}{l} \left(\dfrac{n\rho_0}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)\bar{y}_s + \\ \left(\dfrac{\rho_0(\alpha + \beta)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)\dfrac{\alpha}{\alpha + \beta} \end{array} \right\} \\
&= \frac{\rho_0\bar{y}_s}{\rho_0\bar{y}_s + \rho_1(1 - \bar{y}_s)}
\end{aligned}
\tag{82}
$$

Hence, for large sample size n, the Bayes estimate of $\theta_p \in \Omega_p$, shrinkage to:

$$
\frac{\rho_0\bar{y}_s}{\rho_0\bar{y}_s + \rho_1(1 - \bar{y}_s)} = \hat{\theta}_p(\inf)
\tag{83}
$$

where $\hat{\theta}_p(\inf)$ denotes the maximum likelihood estimate of θ_p under informative sampling design, treating θ_p as fixed unknown constant.

Notice that if $\rho_0 = \rho_1$, that is, the sampling design is ignorable, then $\lim_{n \to \infty} \hat{\theta}_{pB} = \bar{y}_s$.

The important feature of this formula is that, if θ_p, (consequently θ_s) is treated as fixed unknown constant (classical or frequentist statistics), that is, $\alpha = \beta = 0$, and the sampling

design is informative, then for large sample size n, the Bayes estimate of θ_p becomes $\left\{\left(\rho_0 \bar{y}_s\right)/\left(\rho_0 \bar{y}_s + \rho_1\left(1 - \bar{y}_s\right)\right)\right\}$.

It is easy to verify that for given (α, β), the generalized MLE of $\theta_s \in \Omega_s$ is given by:

$$\hat{\theta}_{sg} = \frac{\alpha + n\bar{y}_s}{\alpha + \beta + n}$$

$$= \left(\frac{n}{\alpha + \beta + n}\right)\bar{y}_s + \left(\frac{\alpha + \beta}{\alpha + \beta + n}\right)\frac{\alpha}{\alpha + \beta} \tag{84}$$

which is the sample Bayes estimate of $\theta_s \in \Omega$. Hence using the relationship between θ_s and θ_p, we have

$$\hat{\theta}_{pg} = \frac{\rho_0\left(\alpha + n\bar{y}_s\right)}{\rho_0\left(\alpha + n\bar{y}_s\right) + \rho_1\left(n - n\bar{y}_s + \beta\right)} \tag{85}$$

Notice that if $\rho_0 = \rho_1$, that is, the sampling design is ignorable then, then

$$\hat{\theta}_{sg} = \frac{\alpha + n\bar{y}_s}{\alpha + \beta + n} = \hat{\theta}_{pg} \tag{86}$$

which the Bayes estimate of $\theta_p \in \Omega_p$ under noninformative sampling design.

The important feature of this formula is that, if θ_p is treated as fixed unknown constant (classical or frequentist statistics), that is, $\alpha = \beta = 0$, and the sampling design is noninformative, then for large sample size the Bayes estimate of θ_p becomes $\bar{y}_s$ which is the MLE of θ_p.

7. Bayes Prediction of Finite Population Total

Sverchkov and Pfeffermann (2004) use sample and sample complement distributions for the prediction of finite population totals under informative sampling for single-stage sampling designs. Later Eideh and Nathan (2009) extend the theory to general linear functions of the population values and to two-stage informative cluster sampling. None of the above studies consider prediction of finite population total from Bayesian perspective. In this section we use sample posterior and sample complement posterior distributions, to predict the finite population total under single-stage informative sampling design. Let

$$T = \sum_{i=1}^{N} y_i = \sum_{i\in s} y_i + \sum_{i\in c} y_i = \sum_{i\in s} y_i + \sum_{i\notin s} y_i \tag{87}$$

The available information for the prediction process is $O = \left[\left\{\left(y_i, \pi_i\right), i \in s\right\} \cup \left\{I_i, i \in U\right\}, N, n\right]$, where $I_i = 1$ for $i \in s$ and $I_i = 0$ for $i \notin s$.

Now, following the notations in Sections 2 and 3, the sample complement posterior distribution of $\Theta_p = \theta_p$ given y_i, denoted by $q_c\left(\theta_p \middle| y_i\right)$ is defined by:

$$q_c\left(\theta_p \mid y_i\right) = \frac{q\left(\theta_p\right)f_c\left(y_i \mid \theta_p\right)}{m_c\left(y_i\right)} \tag{88}$$

where, $f_c\left(y_i \mid \theta\right)$ is the sample complement distribution of y_i given $\Theta_p = \theta_p$, which is given by:

$$f_c\left(y_i \mid \theta_p\right) = \frac{E_p\left(1 - \pi_i \mid y_i\right)f_p\left(y_i \mid \theta_p\right)}{E_p\left(1 - \pi_i \mid \theta_p\right)} \tag{89}$$

Also, we have:

$$E_c\left(y_i \mid \theta_p\right) = \frac{E_p\left\{(1 - \pi_i)y_i \mid \theta_p\right\}}{E_p\left\{(1 - \pi_i) \mid \theta_p\right\}} = \frac{E_s\left\{(w_i - 1)y_i \mid \theta_p\right\}}{E_s\left\{(w_i - 1) \mid \theta_p\right\}} \tag{90}$$

and

$$m_c\left(y_i\right) = \begin{cases} \int_{\Theta_p} q\left(\theta_p\right)f_c\left(y_i \mid \theta_p\right)d\theta_p & \text{if } \Theta_p \text{ is continuous} \\[2mm] \sum_{\Theta_p} q\left(\theta_p\right)f_c\left(y_i \mid \theta_p\right) & \text{if } \Theta_p \text{ is discrete} \end{cases} \tag{91}$$

Hence, $q_c\left(\theta_p \mid y_i\right)$ can be written as:

$$\begin{aligned} q_c\left(\theta_p \mid y_i\right) &= \frac{q\left(\theta_p\right)E_p\left(1 - \pi_i \mid y_i\right)f_p\left(y_i \mid \theta_p\right)}{m_c\left(y_i\right)E_p\left(1 - \pi_i \mid \theta\right)} \\[2mm] &\propto \frac{q\left(\theta_p\right)E_p\left(1 - \pi_i \mid y_i\right)f_p\left(y_i \mid \theta_p\right)}{E_p\left(1 - \pi_i \mid \theta_p\right)} \end{aligned} \tag{92}$$

Let $\mathbf{y}_p = \left(y_1, \ldots, y_n, y_{n+1}, \ldots y_N\right)' = \left(\mathbf{y}_s', \mathbf{y}_c'\right)'$, where $\mathbf{y}_s = \left(y_1, \ldots, y_n\right)'$ and $\mathbf{y}_c = \left(y_{n+1}, \ldots, y_N\right)'$. Let $T = T(\mathbf{y}_p)$ be the finite population quantity to be predicted. For example: the finite population total $T = \sum_{i=1}^{N} y_i$. Let $\hat{T}\left(y_1, \ldots, y_n\right) = \hat{T}(\mathbf{y}_s)$ define the predictor of $T(\mathbf{y}_p)$ based on O. The loss involved in predicting $T(\mathbf{y}_p)$ by $\hat{T}(\mathbf{y}_s)$ is $L\left(T(\mathbf{y}_p), \hat{T}(y_s)\right)$. For given $\Theta_p = \theta_p$ and $\mathbf{y}_s$, and squared error loss function, the Bayes predictor of $T(\mathbf{y}_p)$ is:

$$\hat{T}(\mathbf{y}_s) = E_p\left(T(\mathbf{y}_p) \mid \mathbf{y}_s\right) \tag{93}$$

Now we consider the following:

$$\begin{aligned} E_p\left(T(\mathbf{y}_p) \mid \mathbf{y}_s\right) &= E_p\left[\sum_{i=1}^{N} y_i \mid \mathbf{y}_s\right] = E_p\left\{\left[\sum_{i \in s} y_i + \sum_{i \in c} y_i\right] \mid \mathbf{y}_s\right\} \\[2mm] &= \sum_{i \in s} y_i + \sum_{i \in c} E_c\left(y_i \mid \mathbf{y}_s\right) \end{aligned} \tag{94}$$

Thus, the estimate of the population total, for a fully Bayesian method under squared error loss, requires the posterior predictive expectation $E_c\left(y_i \mid \mathbf{y}_s\right)$ for prediction of

$y_i, i \in c = \bar{s}$ given $\mathbf{y}_s$. This posterior expectation, under informative sampling design, can be obtained by the iterated expectations:

$$E_c(y_i|\mathbf{y}_s) = E_s\{E_c(y_i|\mathbf{y}_s, \theta_p)|\mathbf{y}_s\} \tag{95}$$

where E_s is the expectation with respect to the sample posterior distribution of Θ_p given $\mathbf{y}_s$, and E_c is the expectation with respect to the sample complement of the posterior predictive distribution of $y_i, i \in c$ given $\mathbf{y}_s$.

Illustration 1: Case 4 in Section 5. Superpopulation model: Let $y_i | \Theta_p = \theta_p \sim N(\theta_p, \sigma^2)$ be independent random variables, $i = 1,...,N$, where $\sigma^2 > 0$ is known, $i = 1,...,N$. Suppose that $E_p(\pi_i|y_i) = \exp(a_1 y_i)$. Then the sample model is: $y_i | \Theta_s = \theta_s \sim N(\theta_s, \sigma^2)$, where $\theta_s = \theta_p + a_1\sigma^2$. Suppose that the prior distribution of θ_s is $q(\theta_s) = 1, -\infty < \theta_s < \infty$. Then $\Theta_p|\mathbf{y}_s \sim N\left(\bar{y}_s - a_1\sigma^2, \dfrac{\sigma^2}{n}\right)$.

Computation of $E_c(y_i|\mathbf{y}_s) = E_{\theta_p}\{E_c(y_i|\mathbf{y}_s, \theta_p)|\mathbf{y}_s\}$:

Since $y_i, i \in c$ is independent of $\mathbf{y}_s$, therefore

$$E_c(y_i|\mathbf{y}_s, \theta_p) = E_c(y_i|\theta_p) = \frac{E_p\{(1-\pi_i)y_i | \theta_p\}}{E_p\{(1-\pi_i)| \theta_p\}} \tag{96}$$

After some algebra, and noting that $E_p\{\pi_i y_i | \theta_p\} = E_p E_p\{\pi_i y_i | y_i, \theta_p\}$ we can show that:

$$E_c(y_i|\mathbf{y}_s, \theta_p) = \theta_p - a_1\sigma^2 \frac{M_p(a_1)}{1 - M_p(a_1)} \tag{97}$$

where $M_p(a_1) = \exp(a_1\theta_p + 0.5a_1^2\sigma^2)$ is the moment generating function of $y_i | \Theta_p = \theta_p \sim N(\theta_p, \sigma^2)$.

so that,

$$\begin{aligned}
E_c(y_i|\mathbf{y}_s) &= E_s\{E_c(y_i|\mathbf{y}_s, \theta_p)|\mathbf{y}_s\} = E_s\left\{\left[\theta_p - a_1\sigma^2 \frac{M_p(a_1)}{1 - M_p(a_1)}\right]\middle|\mathbf{y}_s\right\} \\
&= E_s(\theta_p|\mathbf{y}_s) - a_1\sigma^2 E_s\left\{\left[\frac{M_p(a_1)}{1 - M_p(a_1)}\right]\middle|\mathbf{y}_s\right\}
\end{aligned} \tag{98}$$

Now,

$$E_s(\theta_p|\mathbf{y}_s) = \bar{y}_s - a_1\sigma^2 \tag{99}$$

and

$$E_s\left\{\left[\frac{M_p(a_1)}{1 - M_p(a_1)}\right]\middle|\mathbf{y}_s\right\} = \int_{-\infty}^{\infty} \frac{\exp(a_1\theta_p + 0.5a_1^2\sigma^2)}{1 - \exp(a_1\theta_p + 0.5a_1^2\sigma^2)} \times \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left[-\frac{n}{2\sigma^2}(\theta_p - (\bar{y}_s - a_1\sigma^2))^2\right]d\theta_p \tag{100}$$

Using Taylor series expansion of $M_p(a_1)/1 - M_p(a_1)$ around $\theta_p = \bar{y}_s - a_1\sigma^2$, we can show that, approximately:

$$E_s\left\{\left[\frac{M_p(a_1)}{1 - M_p(a_1)}\right]\middle|\mathbf{y}_s\right\} = \frac{\exp\left(a_1\left(\bar{y}_s - a_1\sigma^2\right) + 0.5a_1^2\sigma^2\right)}{1 - \exp\left(a_1\left(\bar{y}_s - a_1\sigma^2\right) + 0.5a_1^2\sigma^2\right)} \tag{101}$$

Hence,

$$E_c\left(y_i\middle|\mathbf{y}_s\right) = \left(\bar{y}_s - a_1\sigma^2\right) - \frac{\exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}{1 - \exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)} \tag{102}$$

Thus, for given $\left(a_1, \sigma^2\right)$ the Bayes predictor of $T = \sum_{i=1}^{N} y_i$ is:

$$
\begin{aligned}
\hat{T}(\mathbf{y}_s) &= E_p\left(T(\mathbf{y}_p)\middle|\mathbf{y}_s\right) \\
&= \sum_{i \in s} y_i + \sum_{i \in c}\left\{\left(\bar{y}_s - a_1\sigma^2\right) - a_1\sigma^2\frac{\exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}{1 - \exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}\right\} \\
&= n\bar{y}_s + (N-n)\bar{y}_s - (N-n)a_1\sigma^2\left\{1 + \frac{\exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}{1 - \exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}\right\} \\
&= N\bar{y}_s - (N-n)a_1\sigma^2\left\{1 + \frac{\exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}{1 - \exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}\right\}
\end{aligned}
\tag{103}
$$

Note that if $a_1 = 0$, that is, the sampling design is noninformative, then the Bayes predictor of $T = \sum_{i=1}^{N} y_i$ is the classical one obtained under design-based and frequentist model-based schools. Furthermore, the Bayes predictor of $T = \sum_{i=1}^{N} y_i$ under informative sampling adjust the classical predictor by the term $(N-n)a_1\sigma^2\left\{1 + \frac{\exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}{1 - \exp\left(a_1\bar{y}_s - 0.5a_1^2\sigma^2\right)}\right\}$. This adjusted term depends on the informativeness parameter a_1.

Bayes prediction of finite population total of the other (Cases1,2,3) considered in Sections 5 can be treated similarly.

Illustration 2: Case 2 in Section 6. Let $y_i \mid \Theta_p = \theta_p \sim Ber(\theta_p)$ be independent random variables, where $\theta_p \in (0,1)$, $i = 1,...,N$. Suppose that the sample inclusion probabilities have expectations:

$$E_p\left(\pi_i \mid y_i\right) = \begin{cases} \exp(a_0) = \rho_0 & \text{if } y_i = 0 \\ \exp(a_0 + a_1) = \rho_1 & \text{if } y_i = 1 \end{cases} \tag{104}$$

Recall that the sample distribution of y_i is

$$f_s\left(y_i\middle|\theta_s\right) = \left(\theta_s\right)^{y_i}\left(1 - \theta_s\right)^{1-y_i}, \; y_i = 0,1 \tag{105}$$

where

$$\theta_s = \frac{\theta_p \rho_1}{\theta_p \rho_1 + (1 - \theta_p)\rho_0} \tag{106}$$

As in Illustration 1, the Bayes predictor of $T = \sum_{i=1}^{N} y_i$ depends on the computation of: $E_c(y_i | \mathbf{y}_s) = E_s \{E_c(y_i | \mathbf{y}_s, \theta_p) \mathbf{y}_s\}$. After some algebra, we can show that:

$$E_c(y_i | \mathbf{y}_s, \theta_p) = \frac{(1 - \rho_1)\theta_p}{(1 - \rho_1) + (\rho_0 - \rho_1)\theta_p} \tag{107}$$

so that,

$$E_c(y_i | \mathbf{y}_s) = E_s \{E_c(y_i | \mathbf{y}_s, \theta_p) \mathbf{y}_s\} = E_s \left\{ \left[\frac{(1 - \rho_1)\theta_p}{(1 - \rho_1) + (\rho_0 - \rho_1)\theta_p} \right] \mathbf{y}_s \right\} \tag{108}$$

We know that, approximately

$$E_s(\theta_p | \mathbf{y}_s) = \frac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \tag{109}$$

Expand $\left[\dfrac{(1 - \rho_1)\theta_p}{(1 - \rho_1) + (\rho_0 - \rho_1)\theta_p} \right]$ around $\theta_p = \dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)}$, we can show that, approximately:

$$E_s \left\{ \left[\frac{(1 - \rho_1)\theta_p}{(1 - \rho_1) + (\rho_0 - \rho_1)\theta_p} \right] \mathbf{y}_s \right\} = \frac{(1 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)}{(1 - \rho_1) + (\rho_0 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)} \tag{110}$$

Hence, approximately:

$$E_c(y_i | \mathbf{y}_s) = \frac{(1 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)}{(1 - \rho_1) + (\rho_0 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)} \tag{111}$$

Thus, given $(\alpha, \beta, \rho_0, \rho_1)$ the Bayes predictor of $T = \sum_{i=1}^{N} y_i$ is:

$$\hat{T}(\mathbf{y}_s) = E_p(T(\mathbf{y}_p) \mathbf{y}_s)$$

$$= \sum_{i \in s} y_i + \sum_{i \in c} \left\{ \frac{(1 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)}{(1 - \rho_1) + (\rho_0 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)} \right\} \tag{112}$$

$$= n\bar{y}_s + (N - n) \frac{(1 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)}{(1 - \rho_1) + (\rho_0 - \rho_1)\left(\dfrac{\rho_0(\alpha + n\bar{y}_s)}{\rho_0(\alpha + n\bar{y}_s) + \rho_1(n - n\bar{y}_s + \beta)} \right)}$$

Note that if $\rho_1 = \rho_0$, that is, the sampling design is noninformative, then the Bayes predictor of $T = \sum_{i=1}^{N} y_i$ becomes:

$$\hat{T}(\mathbf{y}_s) = E_p\big(T(\mathbf{y}_p)\big|\mathbf{y}_s\big) = n\bar{y}_s + (N-n)\frac{\alpha + n\bar{y}_s}{\alpha + \beta + n} \tag{113}$$

which is the classical Bayes predictor obtained under model-based school.

9. Conclusion

In this paper we introduced the posterior sample and complement-sample distributions under informative sampling, and use them in the analysis of survey data - developing Bayes predictors of finite population totals and estimating the parameters that characterize the superpoulation normal and binomial models. We have seen that several familiar estimators and predictors in common use are shown to be special cases of the present theory, thus providing them a new interpretation. Also, we have shown that many simple Bayesian analyses based on noninformative prior under informative sampling distributions give similar results to standard non-Bayesian approaches under informative sampling. Furthermore, for large sample size n, for example, the Bayes estimate of $\theta_p \in \Omega_p$, shrinkage to $\bar{y}_s - a_1\sigma^2$ which is the MLE (or Bayes estimator) under the sample distribution. That is, as the sample size n increases, the influence of the prior distribution on sample posterior inferences under informative sampling decreases.

The main features of the present estimators and predictors are their behaviours in terms of the informativeness parameters. Also the use of the Bayes estimators and predictors that ignore the informative sampling design yields biased Bayes estimators and predictors. That is, ignoring the sampling design by not using the weights in the Bayesian statistical inference of survey data can seriously bias the results of the analysis, leading to erroneous conclusions. So, we should incorporate survey design when making survey analytic inferences.

Finally, for simplicity (mathematically handle), and to makes life easy, choose a prior distribution for θ_s - the parameters that characterize the sample distribution, and then derive the Bayes estimator for θ_p - the parameters that characterize the superpoulation model, via the relationship between θ_p and θ_s is more convenient than choose a prior distribution for θ_p .

I hope that the present theory and the posed approach in this paper, will encourage survey statisticians for further investigations for this vital topic of Bayesian statistical inference from sample survey data.

Acknowledgment

References

1. Chambers, R. and Skinner, C. (2003). *Analysis of Survey Data*. New York: John Wiley.

2. Eideh, A. H. (2003). *Estimation for Longitudinal Survey Data under Informative Sampling*, PhD Thesis, Department of Statistics, Hebrew University of Jerusalem. Israel.

3. Eideh, A. H. and Nathan, G. (2006a). The analysis of data from sample surveys under informative sampling. *Acta et Commentationes Universitatis Tartuensis de Mathematica*. *Tartu 2006.* 10*, 41-51.

4. Eideh, A. H. and Nathan, G. (2006b). Fitting Time Series Models for Longitudinal Survey Data under Informative Sampling. *Journal of Statistical Planning and Inference* 136, 3052-306.

5. Eideh, A. H. and Nathan, G. (2009). Two-Stage Informative Cluster Sampling with application in Small Area Estimation. *Journal of Statistical Planning and Inference. 139*, pp 3088-3101.

6. Eideh, A.H. (2012a). Fitting Variance Components Model and Fixed Effects Model for One-Way Analysis of Variance to Complex Survey Data. *Communications in Statistics – Theory and Methods, 41*, pp 3278-3300.

7. Eideh, A.H. (2012b). Estimation and Prediction under Nonignorable Nonresponse via Response and Nonresponse Distributions. *Journal of the Indian Society of Agriculture Statistics*, 66(3) 2012, pp. 359-380.

8. Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2009). *Bayesian Data Analysis (second edition)*. CHAPMAN & HALL/CRC. New York.

9. Kim, D.H. (2002). Bayesian and Empirical Bayesian Analysis under Informative Sampling. *Sankhya*, Volume 64, Series B, Pt.3, pp.267-288.

10. Little, R.J. (2012). Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics. *Journal of Official Statistics*, Vol. 28, No. 3, 2012, pp. 309–334

11. Nandram B., Choi, J.W., Shen, G., and Burgos, C. (2006). Bayesian predictive inference under informative sampling and transformation. Appl. Stochastic Models Bus. Ind., 2006; 22:559–572.

12. Pfeffermann, D. and Sverchkov, M.(1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya*, 61, B, 166-186.

13. Pfeffermann, D. and Sverchkov, M.(2003). Fitting Generalized Linear Models under Informative Probability Sampling. In: *Analysis of Survey Data*. (Eds. R. Chambers and C. J. Skinner). New York: Wiley, pp. 175-195.

14. Pfeffermann, D., Krieger, A.M, and Rinott, Y.(1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica* 8,: 1087-1114.

15. Pfeffermann, D., Krieger, A.M, and Rinott, Y.(1998).. Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica* 8: 1087-1114.

16. Pfeffermann, D.: Modelling of complex survey data: why is it a problem? How should we approach it? Survey Methodology. 37(2), 115-136 (2011).

17. Rao, J.N.K. (2011). Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal. *Statistical Science*, 26, 240–256.

18. Savitsky, T.D. and Daniell Toth, D. (2016). Bayesian estimation under informative sampling. *Electronic Journal of Statistics.* Vol. 10, 1677–1708.

19. Skinner, C.J., Holt, D., and Smith, T.M.F (Eds.), 1989. *Analysis of Complex Surveys*, New York: Wiley.

20. Sverchkov, M. and Pfeffermann, D. (2004). Prediction of finite population totals based on the sample distribution. *Survey Methodology*, 30, 79-92.