Journal of the Nigerian Association of Mathematical Physics
Volume **12 (May, 2008), 455 - 460**

The Bayesian perspective of some sample survey design in survey non-response

***O. R. Oniyide and O. O. Alabi**
Department of Mathematical Sciences
Olabisi Onabanjo University, Ago-Iwoye, Nigeria

Abstract

In this work we outline the Bayesian approach to statistical inference for finite population survey. We argue that many standard design – based inferences can be derived from Bayesian perspectives using classical model with non informative prior distributions, thus the already existing methods do not need to be neglected. We show that the model for the selection mechanisms does not affect inferences about θ ***and*** $Q(y_u)$. ***The result of our finding however shows that Bayesian inferences under a carefully chosen model do enjoy good frequentist properties and a method derived under the design – based paradigm does not become any less robust under a Bayesian etiology.***

Keywords: Baye's Theorem, Super Population Model, Ignorable Bayesian, Robustness Design Consistent**.**

1.0 Introduction

Little (2003 [6]) observed that careful model specification, sensitive to the survey design, can address the concerns about model misspecification, and that Bayesian statistics provides a coherent and unified treatment of descriptive and analytic survey inference. Three main approaches to finite population inference can be distinguished (Little and Rubin, 1983).

(1) Design – based inference: In this case probability statements are based on the distribution of sample selection, with the population quantities treated as fixed.

(2) Frequentist super population modeling: In this case the population values are assumed to be generated from super population parameters, and inferences are based on repeated sampling from this model. For analytic inference about super population parameters, this is standard frequentist model based parametric inference with features of the complex sample design reflected in the model as appropriate. For descriptive inference, the emphasis is on prediction of non –sampled values and hence of finite population quantities.

(3) Bayesian inference: Here the population values are assigned a prior distribution and the posterior distribution of finite population quantities is computed using Baye's theorem.

In practice, the prior is usually specified by formulating a parametric super population model for the population values, and then assuming a prior distribution for the parameters as we shall see later. When the prior is relatively weak and dispersed compared with the likelihood, as when the sample size is large, this form of Bayesian inferences is closely related to frequentist super population modeling, except that a prior for the parameters is added to the model specification and inferences are based on the posterior distribution rather than on repeated sampling from the super population model.

1.1 Basis for Bayesian Approach

Some of the distinguishing features of Bayesian approach include the following:

*all correspondence to the author.

(1) Bayesian methods provide a unified frame work for addressing all the problems of survey inferences, including inferences for descriptive or analytical estimands, small or large sample sizes, inference under planned, ignorable sample selection methods such as probability sampling and problems where modeling assumptions play a more central role such as missing data or measurement error.
(2) Bayesian inference allows for prior information about a problem to be incorporated in the analysis in a simple and straight forward way, via the prior distribution. In surveys, non-informative priors are generally chosen for fixed parameters, reflecting absence of strong prior information; but multilevel models with random parameters also play an important role for problems such as small area estimation.
(3) Bayesian inference deals with nuisance parameters in a natural and appealing way.
(4) Bayesian inference satisfies the likelihood principle, unlike frequentist super population modeling or design-based inference.
(5) Modern computation tools make Bayesian analysis much more practically feasible than in the past (for example, Tanner 1996).
For more detailed notes on features of Bayesian approach see Little (2003 [6]).

2.0 Models for the selection pattern

We view now the system of modeling the selection pattern within Bayesian framework. This formulation enhances the impact of random sampling within the Bayesian inferential framework. For the purpose of notations, we follow the pattern as used by Gelman *et al.* (1995 [3]) with further modifications. We define as follows:

$y_u = (y_{tj})$, where y_{tj} is the value of survey variable J for unit t, $j = 1,\Lambda\ , J;\ t \in u = \{1 \Lambda\ \Lambda\ N\}$

$Q = Q(y_u)$ is the finite population quantity $i_u = (i_1 .... i_N)$ is the sample inclusion indicator, where

$$i_t = \begin{cases} 1, \ if \ \ unit \ t \, is \ included \\ \\ 0, \ otherwise \end{cases}$$

$$y_u = (y_{inc}, y_{exc}), \ y_{inc} implies \ \ included \ \ \ part \ of \ \ y_u$$
$$y_{exc} implies \ \ excluded \ \ \ part \ \ of \ \ y_u$$

Z_u is the fully observed covariates design variable.

The expanded Bayesian approach specifies a model for both the survey data y_u and the sample inclusion indicator i_u. Thus the model can be formulated as

$$P(y_u, i_u \mid Z_u, \theta, \phi) = P(y_u \mid Z_u, \theta) \times P(i_u \mid Z_u, y_u, \phi), \tag{2.1}$$

Where $\theta, \ \phi$ are unknown parameters indexing the joint distribution of y_u and i_u. The likelihood of $\theta, \ \phi$ based on the observed data (Z_u, y_{inc}, i_u) is then

$$L(\theta, \phi \mid Z_u, \ y_{inc}, \ i_u) \ \ \alpha \ \ P(y_{inc}, \ i_u \mid Z_u \ \theta, \phi)$$
$$= \int P(y_u, \ i_u \mid Z_u, \theta, \phi) \ dy_{exc} \tag{2.2}$$

In the Bayesian approach the parameters $(\theta, \ and \, \phi)$ are assigned a prior distribution $P(\theta, \ \phi \mid Z_u)$. Analytical inference about the parameters is based on the posterior distribution.

$$P(\theta, \phi \mid Z_u, \ y_{inc}, i_u) \ \ \alpha \ \ P(\theta, \ \phi \mid Z_u) L(\theta, \ \phi \mid Z_u, y_{inc}, i_u) \tag{2.3}$$

Descriptive inference about $\theta = \theta(y_u)$ is based on its posterior distribution given the data, which is conveniently derived by first conditioning on, and then integrating over, the parameters:

$$P(Q(y_u) \mid Z_u, y_{inc}, i_u) = \int P(Q(y_u) \mid Z_u, y_{inc}, i_u, \theta, \phi)\; P(\theta, \phi \mid Z_u, y_{inc}, i_u) d\theta \qquad (2.4)$$

The more usual likelihood does include the inclusion indicators i_u as part of the model specifically, the likelihood ignoring the selection process is based on the model for y_u alone:

$$L(\theta \mid Z_u, y_{inc}) \alpha\, P(y_{inc} \mid Z_u, \theta) = \int P(y_u \mid Z_u, \theta) dy_{exc} \qquad (2.5)$$

Thus the corresponding posterior distribution of θ and $Q\,(y_u)$ are

$$P(\theta \mid Z_u, y_{inc}) \alpha P(\theta \mid Z_u)\;\; L(\theta \mid Z_u, y_{inc})\;\; p(\theta(y_u) \mid Z_u, y_{inc})$$

$$= \int P\;(\theta(y_u) \mid Z, y_{inc}, \theta)\;\; P\;(\theta \mid Z_u, y_{inc}) d\theta \qquad (2.6)$$

It is to be noted that when the full posterior reduces to this simple posterior equation (2.6) above, the selection mechanisms is called ignorable for Bayesian inference.

Applying Rubin's (1976) theory, the two general and simple conditions for ignoring the selection mechanism are:

(1) Selection at Random (SAR), $P(i_u \mid Z_u, y_u, \theta) = P(i_u \mid Z_u, y_{inc}, \theta)$ for all y_{exc}

(2) Bayesian distinctness: $P(\theta, \phi \mid Z_u) = P\,\theta(\theta \mid Z_u) P(\phi \mid Z_u)$

It can easily be verified that these conditions together imply that

$$P(\theta \mid Z_u, y_{inc}) = P(\theta \mid Z_u, y_{inc},\, i_u) \qquad (2.7)$$

and

$$P(Q(y_u) \mid Z_u, y_{inc}) = P(Q(y_u) \mid Z_u, y_{inc}, i_u) \qquad (2.8)$$

Thus the model for the selection mechanism does not affect inferences about θ and $\theta(y_u)$ as mentioned earlier. However, probability sample designs are generally both ignorable and known, in the sense that

$$P(i_u / Z_u, y_u, \phi) = P(i_u / Z_u, y_{inc}) \qquad (2.9)$$

where Z_u represents known sample design information, such as clustering or stratification information

***Remark* 2.1**

Many sample designs depend only on Z_u and not on y_{inc}; an exception is double sampling, where a sample of units is obtained, and inclusion into a second phase of question is restricted to a sub-sample with a design that depends on characteristics measured in the first phase.

3.0 Applications

3.1 Bayesian derivation for some sample designs

We now attempt to model the mechanism for selection for some sample designs, namely; Simple random sampling and Stratified random sampling

3.2 Simple random sampling

The distribution of the simple random sampling selection mechanism is

$$P(i_u / Z_u, y_u, \phi) = \begin{cases} \binom{N}{n}^{-1}, & if \sum_{t=1}^{N} i_t = n \\ 0, & otherwise \end{cases} \qquad (3.1)$$

This mechanism does not depend on y_u or unknown parameters and hence is ignorable. A basic model for a single continuous survey outcome with simple random sampling is

$$\lfloor y_t/\theta_1 \sigma^2 \rfloor \sim ind \quad N(\theta,\sigma^2) \tag{3.2}$$

where $N(\theta,\sigma^2)$ denotes the normal distribution with mean θ, variance σ^2. For simplicity, assume initially that σ^2 is known and assign an improper uniform prior for the mean.

$$P(\theta|Z_u) \ \alpha \ const \tag{3.3}$$

Standard Bayesian calculations (Gelman *et al* 1999) then yield the posterior distribution of θ to be $N(\bar{y}_s, \sigma^2/n)$ where $\bar{y}_s$, is the sample mean.

To derive the posterior distribution of the population mean, $\bar{y}_u$, note that the posterior distribution of $\bar{y}_u$ given the data and θ is normal with mean given as $\bar{y}$

$$(n\bar{y}_s + (N-n)\theta)/N \tag{3.4}$$

and variance:

$$(N-n)\theta^2/N^2 \tag{3.5}$$

Integrating over θ, the posterior distribution of $\bar{y}_u$, given the data, is normal with mean

$$(n\,\bar{y}_s + (N-n)E(\theta/\bar{y}_s))/N = (n\bar{y}_s + (N-n)\bar{y}_s)/N = \bar{y}_s \tag{3.6}$$

and variance

$$E[\text{var.}(\bar{y}_u/\bar{y}_s,\theta)] + \text{var.}[E(\bar{y}_u/\bar{y}_s,\theta)] = (N-n)\sigma^2/N^2 + (1-n/N)^2\sigma^2/n = (1-n/N)\sigma^2/n \tag{3.7}$$

Hence a 95% posterior probability interval for $\bar{y}_u$ is $\bar{y}_s \pm 1.96\sqrt{\sigma^2(1-n/N)}$ which is identical to the 95% confidence interval from design based theory for a simple random sample.

***Remark* 3.1**

This correspondence between Baye's and design-based results also applies asymptotically to nonparametric multinomial models with Dirichlet priors (Scott, 1977b : Binder 1982 [1])

***Remark* 3.2**

With σ^2 known, the standard design-based approach estimates σ^2 by the sample variance s^2, and assumes large samples. The Bayesian approach yields small sample t corrections, under normality assumptions. In particular, if the variance is assigned Jeffrey's prior $P(\sigma^2) \ \alpha \ 1/\sigma^2$, the posterior distribution of $\bar{y}_u$ is student's *t* with mean $\bar{y}_s$, scale $\sqrt{S^2(1-n/N)}$ and degrees of freedom $n-1$. The resulting 95% posterior probability interval is $\bar{y} \pm t_{n-1}, 0.975 \ \sqrt{S^2(1-n/N)/n}$, where $t_{n-1}, 0.975$ is the 97.5th percentile of the t distribution with $n-1$ degrees of freedom.

***Remark* 3.3**

The Bayesian approach automatically incorporates the finite population correction $(1-n/N)$ in the inference

3.3 Stratified random sampling:

In general, under stratified random sampling the population is divided into H strata and n_h units are selected from the population of N_H units in stratum *h*.

Define Z_u as a set of stratum indicators, with components

$$Z_t = \begin{cases} 1, \ if \ unit \ t \ is \ in \ stratum \ h; \\ 0, otherwise \end{cases}$$

This selection mechanism is ignorable providing the model for y_t conditions on the stratum variables z_t. A simple model that does this is

$$\left\lfloor y_t \mid Z_t = h, \{\theta_h, \sigma_h^2\} \approx_{ind} N(\theta_h, \sigma_h^2) \right\rfloor \quad (3.8)$$

where *N (a, b)* denotes the normal distribution with mean *a*, variance *b*. For simplicity, assume σ_h^2 is known and the flat prior

$$p(\theta_h \mid z_u) \propto const. \quad (3.9)$$

on the stratum means. Bayesian calculations similar to the first illustration lead to

$$\left\lfloor \bar{y}_u \mid Z_u,\ data,\ \{\sigma_h^2\} \right\rfloor \approx N(\bar{y}_{st}, \sigma^2{}_{st}) \quad (3.10)$$

where $\bar{y}_{st} = \sum_{h=1}^{H} P_h\, \bar{y}_{sh},\ P_h = N_h/N,\ \bar{y}_{sh} = sample\ mean\ in\ stratum\ h,$

$$\sigma^2{}_{st} = \sum_{h=1}^{H} P_h^2\, (1 - f_h)\sigma_h^2/n_h$$

$$f_h = n_h/N_h\ .$$

These Bayesian results lead to Baye's probability intervals that are equivalent to standard confidence interval from design – based inference for a stratified random sample.

In particular, the posterior mean weights cases by the inverse of their inclusion probabilities, as in the Horvitz–Thompson estimator (Horvitz and Thompson, 1952).

$$\bar{y}_{st} = N^{-1} \sum_{h-1}^{H} N_h\, \bar{y}_{sh} = N^{-1} \sum_{h=1}^{H} \sum_{t:x_t = h} y_t\, i_t/\pi_h \quad (3.11)$$

$$\pi_h = n_h/N_h = selection\ probability\ in\ stratum\ k$$

The posterior variance equals the design-base variance of the stratified mean

$$\operatorname{var}(\bar{y}_u \mid Z_u,\ data) = \sum_{h=1}^{H} P_h^2 \left(1 - {n_h}/{N_h}\right) {\sigma_h^2}/{n_h} \quad (3.12)$$

Binder (1982 [1]) demonstrates a similar correspondence asymptotically for stratified one world model with Dirichlet priors. With unknown variances the posterior distribution of $\bar{y}_u$ for this model with a uniform prior on log (σ_h^2) is a mixture of t distributions, thus propagating the uncertainty from estimating the stratum variances.

***Corollary* 3.4**

Assume the model equation (3.2) and (3.3) with no stratum effects. With a flat prior on the mean, the posterior mean of $\bar{y}_u$ is then the unweighted mean

$$E(\bar{y}_u | z_u, data, \sigma^2) = \bar{y}_s \equiv \sum_{h=1}^{H} P_h\, \bar{y}_{sh}\ ,\ P_h = n_h/n,$$

which potentially is very biased for $\bar{y}_u$ if the selection rates $\pi_h = n_h/N_h$ vary across the strata. The problem is that inferences from this model are non robust to violations of the assumption of no stratum effects and stratum effects are to be expected in most settings. Robustness consideration leads to the model equation (3.8) that allows for stratum effect.

4.0 Conclusion

Bayesian approach under simple random sample automatically incorporates the finite population correction $\left(1-{}^{n}\!/_{N}\right)$ in the inference. However, the corresponding estimates observed between Baye's and Design based results as shown in section 3 portray Bayesian as been very useful since it incorporates prior posterior information and yet does not become less robust.

References

[1] Binder, D. A. (1982): Non – parametric Bayesian models for samples from finite populations: *Journal of the Royal Statistical Society, 44, 388 – 93*

[2] Box, G. E. P (1980): Sampling and Baye's inference in scientific modeling and robustness Journal of the royal statistical society series A, 143, 383 – 430.

[3] Gelman,A., Carlin,J.B., Stern,H.S,and Rubin D.B.(1995) Bayesian Data Analysis. London: Chapman and Hall.

[4] Little R A (1982): Models for non response in sample surveys. *Journal of the American statistical Association 77, 235 – 50.*

[5] Little R. A. (1991): Inference with survey weights. *Journal of official statistics 7, 405 – 24*

[6] Little R. A. (2003): Analysis of survey. Data: chambers & skinner (Editions) Wiley 2003.

[7] Oniyide O.R (2005): The comparative study of the complete case analysis and pseudo maximum likelihood estimation method for analysis of missing data. *Abacus, 379-387, vol.32 No 2B.*

Robert M. Grooves *et al* (2002): Survey nonresponse. Wiley 2002 (1st Edition) New York.