Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 68(3) 2014 351-358

Bayesian Predictive Inference for the Mean and Variance of a Finite Population Proportion: Two Stage Cluster Sampling with Non-Sampled Cluster Sizes Unknown

Michael J. Racz[1] and J. Sedransk[2]
[1]*Albany College of Pharmacy and Health Sciences, Albany, NY 12208*
[2]*Case Western Reserve University, Cleveland, OH 44106-7054*

Received 25 August 2012; Revised 26 December 2013; Accepted 06 June 2014

SUMMARY

With a complex survey design Bayesian predictive inference for finite population quantities may be difficult to carry out in organizations where skill in applying sampling-based methods is limited. Here we investigate the feasibility of approximating part of the analysis. Motivated by a survey of the quality of care that radiation therapy patients receive we assume a two-stage cluster sample design where the cluster sizes are known only for the clusters in the sample. We propose an exact analysis conditional on the cluster sizes for all units in the population, but an approximate analysis to take account of the unknown cluster sizes for the nonsampled clusters. Successful approximation will greatly simplify the analysis, and suggests the value of similar approximations when there are more complex sample designs.

Keywords: Binary variable, Radiation therapy, Prostate cancer.

1. INTRODUCTION

With a complex survey design Bayesian predictive inference for finite population quantities is cumbersome. Here we investigate the feasibility of approximating part of the analysis. Successful approximation will greatly simplify the analysis, important for a staff not facile with sampling-based methods. We do this by assuming a two-stage cluster sample design with simple random sampling at each stage. The cluster sizes are known only for the clusters in the sample. Our work is motivated by the Patterns of Care Study (PCS), a large survey to investigate the quality of care that radiation therapy patients receive. Within each stratum a simple random sample of radiation therapy facilities is selected and then a simple random sample of patients is chosen within each selected facility. We propose an approximate analysis to take account of the unknown cluster sizes for the nonsampled clusters. The remainder of the inference, conditional on the cluster sizes for all units in the population, may be exact or sampling-based. Our work also suggests the value of similar approximations when there are several stages of cluster sampling and more complex designs.

Assume that n clusters are sampled from the N clusters in the population. Denote by M_k and m_k the number of units and sample size in cluster k ($0 \le m_k \le M_k$). Let Y_{ki} denote the random variable corresponding to the i-th unit in cluster k. Given θ_k it is assumed that $\{Y_{ki}: i = 1, \ldots, M_k\}$ are independent with common

Corresponding author: Michael J. Racz
E-mail address : Michael.Racz@acphs.edu

distribution, $g(\cdot|\theta_k)$, and there is independence across the N clusters. Let $\underline{M}_s$ and $\underline{M}_{ns}$ denote, respectively, the sets of M_k corresponding to the sampled and non-sampled clusters. Finally, $\underline{y}_s$ and $\underline{Y}_{ns}$ denote the vectors of observed and unobserved Y's. Thus, we wish to make inference about $\underline{Y}_{ns}$ and $\underline{M}_{ns}$ given $\underline{y}_s$ and $\underline{M}_s$. For $\underline{Y}_{ns}$ we have the marginal posterior distribution

$$f(\underline{Y}_{ns}|\underline{y}_s, \underline{M}_s) = \sum_{\underline{M}_{ns}} \int f_1\left(\underline{Y}_{ns}\middle|\underline{y}_s, \underline{M}, \underline{\theta}\right) f_2\left(\underline{\theta}\middle|\underline{y}_s, \underline{M}\right) \times f_3\left(\underline{M}_{ns}\middle|\underline{M}_s, \underline{y}_s\right) d\underline{\theta} \quad (1)$$

where $\underline{\theta} = \{\theta_k : k = 1, ..., N\}$ and $\underline{M} = \left(\underline{M}_s^t, \underline{M}_{ns}^t\right)^t$. In this paper we consider a specification where exact inference for f_1 and f_2 is possible but we approximate f_3. In other cases sampling-based inference may be needed for f_1 or f_2.

In Section 2 we describe the distribution assumed for the M_k and the sampling distribution for the Y_{ki} that we use for illustration. Section 3 summarizes the exact inference (f_1 and f_2 in (1)), and presents the inference for $\underline{M}_{ns}$ in f_3. Section 4 has a numerical example, *i.e.*, one of those which motivated this work.

2. NOTATION AND MODELS

Motivated by the Patterns of Care Study data we specify a model for the M_k that will be useful for many applications. Assume that each cluster size, M_k, can take on any value from $\{t_i : i = 1, 2, ..., T\}$, a set of non-negative integers. Often, each t_i is an integer that represents a range of values; *e.g.*, if M_k lies in the interval $[a_i, b_i], t_i = (a_i + b_i)/2$. Given $\underline{p} = (p_1, p_2, ..., p_T)$, such that $0 < p_i < 1$ for $i = 1, 2, ..., T$ and $\sum_{i=1}^{T} p_i = 1$, each M_k, $k = 1, 2, ..., N$, is assumed to be independent and identically distributed such that $\Pr(M_k = t_i \mid \underline{p}) = p_i$ for $i = 1, 2, ..., T$. Given $\underline{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_T)$, the prior of $\underline{p}$, $\pi(\underline{p})$, is specified as a Dirichlet distribution, *i.e.*,

$$\pi(\underline{p}) = \frac{\Gamma\left(\sum_{t=1}^{T} \alpha_i\right)}{\prod_{t=1}^{T} \Gamma(\alpha_i)} \prod_{t=1}^{T} p_i^{\alpha_i - 1}. \quad (2)$$

For the Patterns of Care Study (Section 4) we used the "noninformative" prior where $\alpha_i = 1$.

To illustrate our proposed methodology, we assume that Y_{ki} is a Bernoulli random variable, *i.e.*,

$$\Pr(Y_{ki} = 1|\theta_k) = \theta_k \quad (3)$$

We made this choice because the Patterns of Care Study has only binary variables. However, our results about approximate inference for the nonsampled cluster sizes, the objective of our paper, would be similar had we chosen the Y_{ki} to be normally distributed random variables (and used analogous prior distributions). Throughout we assume that the sample design is not informative.

Letting $M'_k = \sum_{i=1}^{M_k} Y_{ki}$, we wish to make inference about

$$P = \sum_{k=1}^{N} M'_k \bigg/ \sum_{k=1}^{N} M_k \quad (4)$$

For the prior distribution on $\underline{\theta}$, given β and τ, we take $\theta_1, ..., \theta_N$ to be distributed independently with beta density function

$$p(\theta\,|\beta, \tau) = B(\beta, \tau - \beta)^{-1} \theta^{\beta - 1}(1 - \theta)^{\tau-\beta-1} \quad (5)$$

where $B(a, b) = \Gamma(a)\Gamma(b)\Gamma(a + b)^{-1}$.

For, $\beta \in \{a_r : 0 < a_1 < a_2 < ... < a_R < \tau\}$,

$$\Pr(\beta = a_r) = \omega_r \quad (6)$$

where $\sum_{r=1}^{R} \omega_r = 1$. It is assumed throughout that τ and the ω_r are fixed quantities. Methods for choosing values for τ and ω_r are discussed in Nandram and Sedransk (1993) and Racz and Sedransk (1996). Thus, conditional on $\underline{M}$, the likelihood is given by the specification in Section 1 and (3) while the prior is given by (5) and (6).

Given $\underline{M}$, the posterior mean and variance of P are given by Nandram and Sedransk (1993) and also presented in our Appendix.

Alternatively, we could assume that the Y_{ki} are normally distributed and use the prior in Scott and Smith (1969). Then, conditional on $\underline{M}$, the posterior mean and variance of the finite population mean,

$\bar{Y} = \sum_{k=1}^{N}\sum_{i=1}^{M_k} Y_{ki} \Big/ \sum_{k=1}^{N} M_k$, have the same forms as the mean and variance of P. Thus, the approximations we consider in this paper are also applicable to this widely used specification.

3. INFERENCE

The joint density of $\underline{Y} = \left(\underline{Y}_{ns}^t, \underline{y}_s^t\right)^t$ and $\underline{M}$, conditional on both $\underline{\theta}$ and $\underline{p}$,

Pr($\{Y_{ki} = y_{ki}$: $k = 1, ..., N, i = 1, ..., M_k\}$, $\{M_k = x_k$: $k = 1, ..., N\}|\underline{\theta}, \underline{p})$, is

Pr($\{Y_{ki} = y_{ki}$: $k = 1, ..., N, i = 1, ..., M_k\}|\{M_k = x_k$: $k = 1, ..., N\}, \underline{\theta})$ **Pr**($\{M_k = x_k$:$k = 1, ..., N\}|\underline{p})$.

3.1 Posterior Mean

We start by evaluating the posterior expected value of P. Let s denote the set of clusters in the sample, $\underline{m} = (m_1, ..., m_n)^t$ the vector of sample sizes, $m'_k = \sum_{i=1}^{m_k} Y_{ki}$, $\underline{m}' = (m'_1, ..., m'_n)^t$, $\underline{M}_s = \{M_k$:$k \in s\}$, $\underline{M}_{ns} = \{M_k : k \not\in s\}$, and $\underline{Y}_{ns}$ the vector of unobserved Y's.

Using iterated expectations

$$E_{\underline{Y}_{ns}, \underline{M}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}_s)$$

$$= E_{\underline{M}_{ns}}[E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{m}, \underline{m}', \underline{M}_s]$$

$$= E_{\underline{p}}\{E_{\underline{M}_{ns}}[E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{m}, \underline{m}', \underline{M}_s, \underline{p}] \mid \underline{m}, \underline{m}', \underline{M}_s\}.$$

To simplify, the notation ignores the dependence on τ and $\underline{\alpha}$. After some algebraic manipulation it can be shown that

$$E_{\underline{Y}_{ns}, \underline{M}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}_s) = E_{\underline{M}_{ns}}\left[\frac{c_1 + \hat{\eta}_1 X_1}{c_2 + X_1}\middle| \underline{M}_s, \underline{m}, \underline{m}'\right] \quad (7)$$

where $X_1 = \sum_{k \not\in s} M_k \Big/ (N-n)$, and c_1, c_2 and $\hat{\eta}_1$ are constants (defined in the Appendix).

3.2 Approximations for the Posterior Mean

The Patterns of Care Study has hundreds of binary variables and we wanted to provide a simple way to compute posterior means and variances. Thus, we developed approximations for these moments that can be computed by a staff not facile with sampling-based methods. To approximate (7), we use a second order Taylor series expansion about $E(X_1|\underline{M}_s)$, *i.e.*,

$$E_{\underline{M}_{ns}}\left[\frac{c_1 + \hat{\eta}_1 X_1}{c_2 + X_1}\middle| \underline{M}_s\right] \approx \frac{c_1 + \hat{\eta}_1 E(X_1 \mid \underline{M}_s)}{c_2 + E(X_1 \mid \underline{M}_s)}$$

$$+ \frac{c_1 - c_2\hat{\eta}_1}{\left[c_2 + E(X_1 \mid \underline{M}_s)\right]^3} Var(X_1 \mid \underline{M}_s) \quad (8)$$

An important objective is to assess the relative sizes of the terms in (8) to see if it is sufficient to use only the first term. Given $\underline{p}$, the M_k are independent and identically distributed. Recalling that $X_1 = \sum_{k \not\in s} M_k \Big/ (N-n)$,

$$E(X_1 \mid \underline{M}_s) = E_{\underline{p}}[E_{\underline{M}_{ns}}(X_1 \mid \underline{p}) \mid \underline{M}_s]$$

$$= E_{\underline{p}}[E(M_k \mid \underline{p})) \mid \underline{M}_s] \text{ for any } M_k,\ k \not\in s. \quad (9)$$

This simplification also holds when evaluating the variance term in (8); *i.e.*, the two components of $Var(X_1 \mid \underline{M}_s)$ are

$$E_{\underline{p}}[Var_{\underline{M}_{ns}}(X_1 \mid \underline{p}) \mid \underline{M}_s] = (N-n)^{-1} E_{\underline{p}}[Var(M_k \mid \underline{p}) \mid \underline{M}_s] \quad (10)$$

and

$$Var_{\underline{p}}[E_{\underline{M}_{ns}}(X_1 \mid \underline{p}) \mid \underline{M}_s] = Var_{\underline{p}}[E(M_k \mid \underline{p}) \mid \underline{M}_s]$$

for any M_k, $k \not\in s$.

Based on the prior in (2), and letting n_i be the number of sampled clusters with assigned size t_i, the approximation in (8) is easily obtained using (9), (10) and

$$E_{\underline{p}}[E_{\underline{M}_{ns}}(M_k \mid \underline{p}, \underline{M}_s) \mid \underline{M}_s] = \sum_{i=1}^{T} t_i \frac{n_i + \alpha_i}{n + \alpha_0},$$

$$E_{\underline{p}}[Var_{\underline{M}_{ns}}(M_k \mid \underline{p}, \underline{M}_s) \mid \underline{M}_s]$$

$$= \left(\frac{\alpha_0 + n}{\alpha_0 + n + 1}\right)\left[\sum_{i=1}^{T} t_i^2 \frac{n_i + \alpha_i}{n + \alpha_0} - \left(\sum_{i=1}^{T} t_i \frac{n_i + \alpha_i}{n + \alpha_0}\right)^2\right]$$

and

$$Var_{\underline{p}}[E_{\underline{M}_{ns}}(M_k \mid \underline{p}, \underline{M}_s) \mid \underline{M}_s]$$

$$= \frac{1}{\alpha_0 + n + 1}\left[\sum_{i=1}^{T} t_i^2 \frac{n_i + \alpha_i}{n + \alpha_0} - \left(\sum_{i=1}^{T} t_i \frac{n_i + \alpha_i}{n + \alpha_0}\right)^2\right]$$

where $\alpha_0 = \sum_{i=1}^{T} \alpha_i$ and, by definition, $\sum_{i=1}^{T} n_i = n.$

To evaluate the order of each of the terms in (8) we use the structure for asymptotic arguments for finite populations described in Fuller (2009) which includes the assumption that $n = [f.N], f \in (0,1)$, and $[f N]$ is the largest integer less than or equal to $f N$. It can be shown that c_1, c_2, $\hat{\eta}_1$ and $E(X_1|\underline{M}_s)$ are each $O(1)$ and that $\{c_2 + E(X_1 \mid \underline{M}_s)\}^{-1}$ is also $O(1)$. Thus, the first term in (8) is $O(1)$. Similarly, $(c_1 - c_2\hat{\eta}_1)/\{c_2 + E(X_1 \mid \underline{M}_s)\}^3$ is $O(1)$. Finally, both components of $Var(X_1|\underline{M}_s)$ are $O(n^{-1})$ so, as expected, the second term is $O(n^{-1})$.

3.3 Posterior Variance

We next evaluate the posterior variance of P, *i.e.*,

$$Var_{\underline{Y}_{ns},\underline{M}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}_s)$$

$$= E_{\underline{M}_{ns}}[Var_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{M}_s]$$

$$+ Var_{\underline{M}_{ns}}[E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{M}_s] \qquad (11)$$

We start by evaluating the second term in (11), writing $Var_{\underline{M}_{ns}}[E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{M}_s]$ as $Var_{\underline{M}_{ns}}[g(X_1) \mid \underline{M}_s]$. A Taylor series expansion of $g(X_1)$ about $E(X_1 \mid \underline{M}_s)$ yields

$$Var_{\underline{M}_{ns}}[g(X_1) \mid \underline{M}_s]$$

$$= Var_{\underline{M}_{ns}}\left[\sum_{l=1}^{\infty} \frac{(-1)^l (c_1 - c_2\hat{\eta}_1)}{[c_2 + E(X_1 \mid \underline{M}_s)]^{l+1}} [X_1 - E(X_1 \mid \underline{M}_s)]^l \mid \underline{M}_s\right] \qquad (12)$$

Because the terms in (12) are expected to be very small (discussed later) only one term from (12) will be used to approximate $Var_{\underline{M}_{ns}}[E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{M}_s]$, *i.e.*,

$$Var_{\underline{M}_{ns}}[E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{M}_s]$$

$$\approx \left[\frac{(c_1 - c_2\hat{\eta}_1)}{[c_2 + E(X_1 \mid \underline{M}_s)]^2}\right]^2 Var(X_1 \mid \underline{M}_s) \qquad (13)$$

The right side of (13) is $O(n^{-1})$, easily seen from the results presented above.

After some algebraic manipulation we can write the first term in (11), $E_{\underline{M}_{ns}}[Var_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}, \underline{M}) \mid \underline{M}_s]$, as

$$E_{\underline{M}_{ns}}\left[\frac{c_3 + c_4 X_1 + \hat{\eta}_2 X_1^2 + c_5 X_2}{(c_2 + X_1)^2} \middle| \underline{M}_s\right] \qquad (14)$$

where $X_2 = (N - n)^{-1} \sum_{k \not\subset s} M_k^2$ and c_2, c_3, c_4, c_5 and $\hat{\eta}_2$ are constants. To approximate (14), use a second order Taylor series expansion about $[E(X_1 \mid \underline{M}_s), E(X_2 \mid \underline{M}_s)]$; *i.e.*,

$$E_{\underline{M}_{ns}}[Var_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \underline{M}) \mid \underline{M}_s]$$

$$\approx \frac{c_3 + c_4 E(X_1 \mid \underline{M}_s) + \hat{\eta}_2 [E(X_1 \mid \underline{M}_s)]^2 + c_5 E(X_2 \mid \underline{M}_s)}{(c_2 + E(X_1 \mid \underline{M}_s))^2}$$

$$+ \frac{c_2^2 \hat{\eta}_2 - 2c_2 c_4 + 3c_3(c_4 - 2c_2\hat{\eta}_2) E(X_1 \mid \underline{M}_s) + \mathrm{A}}{(c_2 + E(X_1 \mid \underline{M}_s))^4}$$

$$- \frac{2c_5}{(c_2 + E(X_1 \mid \underline{M}_s))^3} Cov(X_1, X_2 \mid \underline{M}_s). \qquad (15)$$

Now,

$$Cov(X_1, X_2 \mid \underline{M}_s) = E_{\underline{p}}[Cov_{\underline{M}_{ns}}(X_1, X_2 \mid \underline{p}) \mid \underline{M}_s]$$

$$+ Cov_{\underline{p}}[E_{\underline{M}_{ns}}(X_1 \mid \underline{p}), E_{\underline{M}_{ns}}(X_2 \mid \underline{p}) \mid \underline{M}_s]$$

It follows from the assumption that the M_k, $k \not\subset s$, are independent and identically distributed that

$$E(X_2 \mid \underline{M}_s) = E_{\underline{p}}[E_{\underline{M}_{ns}}(X_2 \mid \underline{p}) \mid \underline{M}_s]$$

$$= E_{\underline{p}}[E(M_k^2 \mid \underline{p}) \mid \underline{M}_s]$$

for any M_k, $k \in s$, and from (2),

$$E_{\underline{p}}[E_{\underline{M}_{ns}}(M_k^2 \mid \underline{p}) \mid \underline{M}_s] = \sum_{i=1}^{T} t_i^2 \frac{n_i + \alpha_i}{n + \alpha_0},$$

and

$$Cov(X_1, X_2 \mid \underline{M}_s)$$

$$= (N-n)^{-1}\left\{\sum_{i=1}^{T} t_i^3 \left(\frac{n_i+\alpha_i}{n+\alpha_0}\right) - \left[\sum_{i=1}^{T} t_i \left(\frac{n_i+\alpha_i}{n+\alpha_0}\right)\right]\right.$$

$$\left. \times \left[\sum_{i=1}^{T} t_i^2 \left(\frac{n_i+\alpha_i}{n+\alpha_0}\right)\right]\right\} + [1-(N-n)^{-1}]$$

$$\times \left\{\sum_{i=1}^{T} t_i^3 \left[\frac{(n_i+\alpha_i)(n+\alpha_0-n_i-\alpha_i)}{(n+\alpha_0)^2(n+\alpha_0+1)}\right]\right.$$

$$\left. -2\sum_{i=1}^{T}\sum_{j<i} t_i t_j^2 \left[\frac{(n_i+\alpha_i)(n_j+\alpha_j)}{(n+\alpha_0)^2(n+\alpha_0+1)}\right]\right\}$$

where $\alpha_0 = \sum_{i=1}^{T} \alpha_i$.

The terms $\hat{\eta}_2, c_3, c_4$, and $E(X_2|\underline{M}_s)$ are all $O(1)$, and c_5 and $Cov(X_1, X_2|\underline{M}_s)$ are $O(n^{-1})$. Then the order of the first term in (15) is $O(1)$, the order of the second term is $O(n^{-1})$, and the order of the third term is $O(n^{-1})$.

4. PATTERNS OF CARE STUDY EXAMPLE

The approximations were applied to the Patterns of Care Study (PCS) data used in Racz and Sedransk (1996). This illustration uses the binary variable indicating whether or not an intravenous pyelogram (IVP) was done for a prostate cancer patient. There were $n = 46$ sampled facilities with M_k ranging in size from 3 to 61, $k = 1, 2, \ldots, 46$. The value of T was chosen to be 5 with $t_i = 10(i-1) + 5$. Each t_i represents the group $(t_i - 5, t_i + 5]$ with t_5 representing $(40, \infty)$. We used a "noninformative" prior in (2) by taking $\alpha_i = 1$. Finally, we compare the approximations to "exact" results from a sampling-based evaluation.

The first objective is to compare the relative sizes of the two terms in (8), the approximation for the posterior predictive expectation of P.

The first term in (8), $\dfrac{c_1 + \hat{\eta}_1 E(X_1 \mid \underline{M}_s)}{c_2 + E(X_1 \mid \underline{M}_s)}$, of $O(1)$, has a value of 0.5078 for this example. The second term in (8), $\dfrac{c_1 - c_2\hat{\eta}_1}{[c_2 + E(X_1 \mid \underline{M}_s)]^3} Var(X_1 \mid \underline{M}_s)$, of $O(n^{-1})$, has a value of 0.0000027 for this example. While the relative sizes of the $O(1)$ and $O(n^{-1})$ terms are not surprising, the reason why the $O(n^{-1})$ is small is unexpected, *i.e.*, it is not due to the large value of n. Now, $Var(X_i|\underline{M}_s)$ is the sum of two $O(n^{-1})$ terms, $(N-n)^{-1} E_{\underline{p}}[Var_{\underline{M}_{ns}}(M_k \mid \underline{p}, \underline{M}_s) \mid \underline{M}_s]$ and $Var_{\underline{p}}[E_{\underline{M}_{ns}}(M_k \mid \underline{p}, \underline{M}_s) \mid \underline{M}_s]$, whose values are 0.1593081 and 3.1112285, respectively. The $O(1)$ term, $(C_1 - C_2\hat{\eta}_1)/[c_2 + E(X_1 \mid \underline{M}_s)]^3$, is 0.00000084. The second term in (8) is small because of the $O(1)$ term, $(c_1 - c_2\hat{\eta}_1)/[c_2 + E(X_1 \mid \underline{M}_s)]^3$, not any asymptotic properties.

Initially, one might conclude that this necessitates examination of higher order terms. As it turns out, however, higher order derivatives follow a distinct pattern. The second derivative term is already small, and higher order derivative terms will decrease consistently. To see this, rewrite the left hand side of (8) as $E_{\underline{M}_{ns}}[g(X_1) \mid \underline{M}_s]$, where $g(x) = (c_1 + \hat{\eta}_1 x)/(c_2 + x)$.

Then $g^{(l)}(x) = l!(-1)^{-1}(c_1 - c_2\hat{\eta}_1)/(c_2 + x)^{l+1}$ for $t = 1, 2, \ldots$ If higher order Taylor series expansions were used to approximate (7), the $l!$ would cancel and the general form of the l^{th} derivative term would be

$$\frac{(-1)^l (c_1 - c_2\hat{\eta}_1)}{\left\{c_2 + E_{\underline{p}}[E(M_k \mid \underline{p}) \mid \underline{M}_s\right\}^{l+1}} \quad \text{for } l = 1, 2, \ldots \qquad (16)$$

Since c_1, c_2, $\hat{\eta}_1$, and $E_{\underline{p}}[E_{\underline{M}_{ns}}(M_k \mid \underline{p}) \mid \underline{M}_s]$ are all $O(1)$ terms $c_1 - c_2\hat{\eta}_1$ and $c_2 + E_{\underline{p}}[E_{\underline{M}_{ns}}(M_k \mid \underline{p},) \mid \underline{M}_s]$ are each $O(1)$. Therefore, as long as the denominator in (16) exceeds 1, the absolute value of (16) will decrease with increasing l.

In addition, the numerator term, $c_1 - c_2\hat{\eta}_1$, has features which allow us to argue heuristically that it will be small. This is seen by regrouping and showing $c_1 - c_2\hat{\eta}_1$ to be the sum of two terms, the first of which is $(N-n)^{-1}\sum_{k\in s}(m_k' - m_k\hat{\gamma}_k)$. Loosely speaking, this is $(N-n)^{-1}$ times the sum of observed values minus their posterior expected value. The second term, $(N-n)^{-1}\sum_{k\in s} M_k(\hat{\gamma}_k - \hat{\eta}_1)$, also should be small since $\hat{\gamma}_k$ and $\hat{\eta}_1$ are similar arguments. In fact, for $k \not\subset s$, they are equal. Our conclusion is, therefore, that increasing order derivative terms will decrease, and $c_1 - c_2\,\hat{\eta}_1$ is already very small.

A second objective of this example is to compare the relative sizes of the terms derived to approximate the posterior predictive variance of P. To approximate $Var_{\underline{M}_{ns}}\left[E_{\underline{Y}_{ns}}(P\,|\,\underline{m},\underline{m}',\underline{M})\,|\,\underline{M}_s\right]$ only one term was derived (see (13)). This is recommended since the derivatives for any higher order terms follow the pattern just described.

For this example, the computed value of this approximation is 7.1×10^{-10}. Now, the approximation for $E_{\underline{M}_{ns}}\left[Var_{\underline{Y}_{ns}}(P\,|\,\underline{m},\underline{m}',\underline{M})\,|\,\underline{M}_s\right]$ from (15) is 0.001090. Thus, as expected, the approximation for $Var_{\underline{Y}_{ns},\underline{M}_{ns}}(P\,|\,\underline{m},\underline{m}',\underline{M}_s)$ receives a negligible contribution from $Var_{\underline{M}_{ns}}\left[E_{\underline{Y}_{ns}}(P\,|\,\underline{m},\underline{m}',\underline{M})\,|\,\underline{M}_s\right]$.

Thus, we would approximate the posterior mean and variance of the population proportion of prostate cancer patients with an IVP as 0.5078 and 0.0011.

We have also evaluated the posterior mean of P in (7) and the posterior variance of P in (11) by repeated sampling of $\underline{M}_{ns}$ given $\underline{M}_s$, $\underline{m}$, $\underline{m}'$ and then evaluating the expectations $E_{\underline{M}_{ns}}(\cdot)$ in (7), $E_{\underline{M}_{ns}}(\cdot)$ and $Var_{\underline{M}_{ns}}(\cdot)$ in (11) where the dots denote quantities having analytical expressions. Letting n_i be the number of sampled observations with $M_k = t_i$, $i = 1, ..., T$, and taking $\alpha_i = 1$ in (2), the posterior distribution of $\underline{p}$ is Dirichlet as in (2) with $\alpha_i = n_i + 1$. Then, $f(M_{n+1}, ..., M_N|\underline{M}_s) = \int f_1(M_{n+1}, ..., M_N|\underline{p})f_2(\underline{p}|\underline{M}_s)d\underline{p}$, where we drop the conditioning on $\underline{m}$ and $\underline{m}'$. We draw $\underline{p}$ from f_2 and then, given $\underline{p}$, draw $M_{n+1}, ..., M_N$ independently where M_k is a multinomial random variable with $\Pr(M_k = t_i|\underline{p})$; *i.e.* if category i is selected, $M_k = t_i$. Using 20,000 iterates from the Dirchlet distribution in f_2, our estimate of the posterior mean of P is 0.5078, the same as the approximation. Proceeding in the same way our sampling based estimate of the posterior variance of P is 0.00108, essentially the same as the approximation, 0.00109.

In summary, we have described a way to simplify Bayesian inference for finite population quantities when cluster sampling is used. This is important in organizations where skill in applying sampling-based methods is limited. While we have investigated a special case, *i.e.*, two stage sampling and inference for a binary variable, one should be able to generalize this approach to additional sampling stages and continuous variables.

ACKNOWLEDGEMENTS

The authors would like to thank the referee for a helpful suggestion which improved the paper.

REFERENCES

Fuller, W.A. (2009). *Sampling Statistics.* John Wiley & Sons, Inc., New Jersey.

Nandram, B., and Sedransk, J. (1993). Bayesian predictive inference for a finite population proportion: Two-stage cluster sampling. *J. Roy. Statist. Soc. B,* **55,** 399-408.

Racz M.J., and Sedransk, J. (1996). Bayesian predictive inference for a binary random variable: Estimating the quality of care that radiation therapy patients receive. In: *Bayesian Biostatistics,* editors Berry D and Stangl D. Marcel Dekker, New York.

Scott, A., and Smith, T.M.F. (1969). Estimation in multi-stage surveys. *J. Amer. Statist. Assoc.*, 830-840.

Appendix

Using the notation in Sections 1 and 2 the following posterior quantities are given in Nandram and Sedransk (1993).

$$\omega_r^* = \Pr(\beta = a_r \mid \underline{m}, \underline{m}', \tau) =$$

$$\omega_r \prod_{k\in s}\binom{\tau-2}{a_r-1}\binom{m_k-\tau-2}{m'_k-a_r-1}^{-1}\left\{\sum_{r=1}^{R}\omega_r\prod_{k\in s}\binom{\tau-2}{a_r-1}\binom{m_k-\tau-2}{m'_k-a_r-1}^{-1}\right\}^{-1},$$

$$\hat{\eta}_1 = E(\beta/\tau \mid \underline{m}, \underline{m}', \tau) = \sum_{r=1}^{R}\omega_r^* a_r \tau^{-1},\ \hat{\eta}_2 = \mathrm{var}(\beta/\tau \mid \underline{m}, \underline{m}', \tau) = \sum_{r=1}^{R}\omega_r^*(a_r\tau^{-1} - \hat{\eta}_1)^2,$$

$$\hat{\eta}_3 = E\{(\beta(\tau-\beta)/\tau(\tau+1) \mid \underline{m}, \underline{m}', \tau\} = (\tau+1)^{-1}\sum_{r=1}^{R}\omega_r^* a_r(1 - a_r\tau^{-1}) \text{ and}$$

$$\hat{v}_k^2 = \begin{cases}\sum_{r=1}^{R}\omega_r^*\left(a_r + m'_k\right)\{\tau + m_k - (a_r + m'_k)\}/(\tau+m_k)(\tau+m_k+1); & k\in s,\\ \hat{\eta}_3; & k\notin s.\end{cases}$$

The posterior mean of $P = \sum_{k=1}^{N} M'_k \Big/ \sum_{k=1}^{N} M_k$, is $E_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \tau, \underline{M}) = \varphi\hat{\eta}_1 + (1-\varphi)\hat{P}$

where, $\rho_k = M_k / \sum_{k=1}^{N} M_k, \varphi = 1 - \sum_{k\in s}\rho_k\lambda_k, \hat{P} = \left\{\sum_{k\in s}\rho_k\lambda_k\right\}^{-1}\sum_{k\in s}\rho_k\lambda_k(m'_k/m_k)$

and

$$\lambda_k = \begin{cases}\{1 + (\tau/M_k)\}\{1 + (\tau/m_k)\}^{-1}; & k\in s,\\ 0; & k\notin s.\end{cases}$$

The posterior variance of P is

$$Var_{\underline{Y}_{ns}}(P \mid \underline{m}, \underline{m}', \tau, \underline{M}) = \hat{\eta}_2\sum_{k=1}^{N}\sum_{l=1}^{N}\rho_k\rho_l(1-\lambda_l)(1-\lambda_k) + \sum_{k=1}^{N}\rho_k^2(1-\lambda_k)(\tau^{-1} + M_k^{-1})\hat{v}_k^2.$$

Rewriting these results from Nandram and Sedransk(1993) broken down by sampled and non-sampled clusters allows for analysis when the non-sampled cluster sizes are unknown.

Defining

$$\hat{\gamma}_k = E(Y_{kj} \mid \underline{m}, \underline{m}', \tau, \underline{M}) = \sum_{r=1}^{R}\omega_r^*\frac{a_r + I_k m'_k}{\tau + I_k m_k} \text{ for k = 1, ..., } N,\ j \notin s(k)$$

and $I_k = 0$ if $k \notin s$, $I_k = 1$ if $k \in s$, and the following constants simplify the expressions:

$$\left\{\sum_{k\in s} m'_k + \sum_{k\in s}(M_k - m_k)\hat{\gamma}_k\right\}\Big/(N-n) = c_1,\ \sum_{k\in s} M_k \Big/(N-n) = c_2,$$

$$(N-n)^{-2}\left\{\hat{\eta}_2\left[\sum_{k\in s} M_k(1-\lambda_k)\right]^2 + \sum_{k\in s} M_k^2(1-\lambda_k)(\tau^{-1}+M_k^{-1})\hat{v}_k^2\right\} = c_3,$$

$$(N-n)^{-1}\left\{2\hat{\eta}_2\sum_{k\in s} M_k(1-\lambda_k)+\hat{\eta}_3\right\} = c_4,\ \hat{\eta}_3/(\tau(N-n)) = c_5.$$