Journal of the Korean Data & Information Science Society
2013, **24**(6), 1543–1550

http://dx.doi.org/10.7465/jkdi.2013.24.6.1543
한국데이터정보과학회지

Bayesian estimation for finite population proportion under selection bias via surrogate samples

Seong Mi Choi[1] · Dal Ho Kim[2]

[12]Department of Statistics, Kyungpook National University

Received 9 October 2013, revised 26 October 2013, accepted 4 November 2013

Abstract

In this paper, we study Bayesian estimation for the finite population proportion in binary data under selection bias. We use a Bayesian nonignorable selection model to accommodate the selection mechanism. We compare four possible estimators of the finite population proportions based on data analysis as well as Monte Carlo simulation. It turns out that nonignorable selection model might be useful for weekly biased samples.

Keywords: Accept-reject algorithm, binary response, grid method, Monte Carlo method, selection bias, surrogate sample.

1. Introduction

We consider the situation in which a sample is drawn from a finite population, but the sample is not a random sample from the original finite population. The sample could be significantly perturbed by some mechanism. For example, in many complex surveys, sample units are drawn with probability proportional to some measure of size. Then the model holding for the sample could be different from the model for the rest of the population (i.e. there is selection bias). Patil and Rao (1978) formed models for this type of sample design using weighted distributions. We use a Bayesian method to infer about a finite population proportion under selection bias. Kwak and Kim (2012) studied Bayesian estimation for the finite population proportions in multinomial problem without selection bias.

If a biased sample is drawn from a finite population, one cannot make inference about the nonsampled values unless the nature of the bias is clearly understood. One example is when a sample is drawn from a finite population with probability proportional to size (PPS). In this case the sampled values are large and the nonsampled values tend to be small. It is possible to construct an appropriate selection model for the sample data, and this will turn out to be a weighted distribution. Predictive inference from the weighted distribution is not straightforward.

Nandram (2007) used surrogate sampling to convert data obtained through a selection bias mechanism to provide an equivalent simple random sample. In fact, the original sample

[1] Ph. D. candidate, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

is a random sample from a weighted distribution and one can convert this sample to a surrogate sample from the original distribution. This surrogate sample can be used to make an inference about the original finite population without any further consideration about the biased sample.

In concrete terms we briefly describe the weighted distribution. Let $y_i, i = 1, ..., N$, denote the finite population values and let $p(\boldsymbol{y}|\boldsymbol{\theta}_1)$ denote the probability distribution that describes the finite population. When a random sample is taken from this finite population, it is perturbed by the weight function $w(\boldsymbol{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ to produce a sample from the new probability distribution $q(\boldsymbol{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. That is, a representative sample is observed from

$$q(\boldsymbol{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = w(\boldsymbol{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) p(\boldsymbol{y}|\boldsymbol{\theta}_1).$$

The idea here is to create a surrogate sample from the original finite population using $p(\boldsymbol{y}|\boldsymbol{\theta}_1)$ and then make an inference about the finite population proportion. The Bayesian analysis is used to convert the biased sample into a random sample from the finite population. We use the surrogate sampler to infer about a finite population proportion using data from a possibly biased sample.

When one includes the selection probabilities in a model, there are two possible choices, an ignorable or nonignorable selection model. In an ignorable selection model the response variable is not related to the selection mechanism, but in a nonignorable selection model the response is related to the selection mechanism, at least partially. For example, for binary data there may ba a higher/lower proportion of positive responses among the sampled values than among the nonsampled values. To account for this discrepancy, one can allow the response binary variable to be correlated with the survey weights or their reciprocals.

To incorporate the selection bias into the ignorable selection model, Malec, Davis and Cao (1999) use a hierarchical Bayesian model to estimate a finite population proportion when there are binary data. Difficulty in including the selection probabilities directly in the model forces them to make an ad hoc adjustment to the likelihood function and use an empirical Bayes approach. Nandram and Choi (2010) have incorporated selection probabilities into a nonignorable nonresponse model to analyze continuous data using a full Bayesian analysis.

In this paper, we consider the problem of making inference about a finite population proportion when a possibly biased sample is available from it. We use the model of Nadram *et al.* (2013) to investigate this problem more deeply based on simulation study. In Section 2 we describe a Bayesian predictive inference of a finite population proportion under selection bias. In Section 3 we provide a simulation study to compare four possible estimators of the finite population proportion under an artificial scenario. Section 4 has summary and concluding remarks.

2. Bayesian estimation under selection bias

We consider a finite population of N units, and we view this finite population as a random sample from a superpopulation which is a hypothetical infinite population. However, the sample from the finite population can be biased. That is, a probability sample of size n is taken with selection probabilities. The selection probabilities are observed only for the sampled values. These selection probabilities are adjusted by the design scientists because of various reasons such as nonresponse and different weights from various sources.

2.1. Modeling

Suppose each individual does, $y = 1$, or does not, $y = 0$, have a characteristic. Thus, let $y_i|p \overset{iid}{\sim} \text{Bernoulli}(p), i = 1, \ldots, N$ and $\pi_i, i = 1, \ldots, N$, be the corresponding selection probabilities. Note that p is the proportion of ones in the entire superpopulation. A sample S of size n is taken from the finite population; also let $\bar{S}$ denote the set of nonsampled values. Letting $\boldsymbol{y}_s$ denote the vector of sampled values and $\boldsymbol{y}_{\bar{s}}$ denote the vector of nonsampled values. Let the sampled values be $y_1, \ldots, y_n$. Let $P = \sum_{i=1}^{N} y_i/N$ denote the finite population proportion. In design-based survey analysis, P is a fixed unknown quantity, but in Bayesian inference P is a random variable which is to be predicted. Our main interest is to predict P when a biased sample is available.

We assume that the sample selection probabilities $(\pi_1, \ldots, \pi_n)$ have support over the set $\pi_u^*,\ u = 1, \ldots, U$. That is, $\pi_i, i = 1, \ldots, n$, have a histogram where the midpoints of the categories are the π_u^*. Throughout these π_u^* are assumed known and the π_i are assumed to be random quantities. The distribution of the selection probabilities, given the binary response y_i, is

$$Pr(\pi_i = \pi_u^* \mid \boldsymbol{\theta}, y_i = y) = \theta_{uy},\ u = 1, \ldots, U,\ y = 0, 1,\ i = 1, \ldots, n$$

and

$$y_i \mid p \overset{iid}{\sim} \text{Bernoulli}(p), i = 1, \ldots, N.$$

Following Malec, Davis and Cao (1999), it is easy to show that

$$P(Y = y \mid \pi = \pi_u^*, \boldsymbol{\theta}, p) = \frac{\theta_{uy} p^y (1-p)^{1-y}}{\sum_y \theta_{uy} p^y (1-p)^{1-y}} \tag{2.1}$$

and

$$P(Y = y|\boldsymbol{\theta}, p) = \frac{\sum_u \pi_u^* \theta_{uy} p^y (1-p)^{1-y}}{\sum_y \sum_u \pi_u^* \theta_{uy} p^y (1-p)^{1-y}}. \tag{2.2}$$

The sampled data actually come from the probability mass function in (2.2) and the entire population is described by $P(Y = y \mid p) = p^y(1-p)^{1-y}, y = 0, 1$, thereby showing how the selection bias enters into the model. Note that in (2.1), $P(Y = y|\pi = \pi^*, \boldsymbol{\theta}, p)$ is a weighted distribution with weights $w = \frac{\theta_{uy}}{\sum_y \theta_{uy} p^y (1-p)^{1-y}}$ and $P(Y = y \mid p)$ is distribution of the population without selection bias.

Since the sampling units are independent, the joint density of the entire sample is

$$P(\boldsymbol{y}_s, \boldsymbol{\pi} \mid \boldsymbol{\theta}, p) = \frac{\prod_{u=1}^{U} (\pi_u^* \theta_{u0})^{g_{u0}} \prod_{u=1}^{U} (\pi_u^* \theta_{u1})^{g_{u1}}}{[p \sum_u \pi_u^* \theta_{u1} + (1-p) \sum_u \pi_u^* \theta_{u0}]^n} p^s (1-p)^{(n-s)}, \tag{2.3}$$

where $s = \sum_{i \in S} y_i$, g_{u0} is the cell count for category u for $y = 0$ and g_{u1} is the cell counts for category u for $y = 1$. Note that $\sum_{u=1}^{U} g_{u0} = n - s$, $\sum_{u=1}^{U} g_{u1} = s$ and $\sum_{u=1}^{U} (g_{u0} + g_{u1}) = n$. This likelihood includes the selection bias.

A priori we assume that p, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are independent, and we take

$$p \sim \text{Uniform}(0, 1)$$

$$\boldsymbol{\theta}_0 \mid \tau \sim \text{Dirichlet}(\boldsymbol{\theta}_0^{(0)} \tau) \text{ and } \boldsymbol{\theta}_1 \mid \tau \sim \text{Dirichlet}(\boldsymbol{\theta}_1^{(0)} \tau),$$

where $\boldsymbol{\theta}_0^{(0)}$ and $\boldsymbol{\theta}_1^{(0)}$ are to be specified. Finally, we put a proper prior

$$p(\tau) = \frac{1}{(1+\tau)^2}, \tau \geq 0,$$

which is called a shrinkage prior.

2.2. Computation

Using Bayes' theorem, the joint posterior density of $p, \boldsymbol{\theta}_1, \boldsymbol{\theta}_0, \tau$ given the data, $\boldsymbol{\pi}, \boldsymbol{y}_s$, is

$$\begin{aligned}\pi(p, \boldsymbol{\theta}_1, \boldsymbol{\theta}_0, \tau \mid \boldsymbol{\pi}, \boldsymbol{y}_s) \quad \propto \quad & \frac{\prod_{u=1}^U (\pi_u^* \theta_{u0})^{g_{u0}} \prod_{u=1}^U (\pi_u^* \theta_{u1})^{g_{u1}}}{[p \sum_u \pi_u^* \theta_{u1} + (1-p) \sum_u \pi_u^* \theta_{u0}]^n} p^s (1-p)^{n-s} \\ & \times \frac{\prod_{u=1}^U \theta_{u0}^{\theta_{u0}^{(0)} \tau - 1}}{D(\boldsymbol{\theta}_0^{(0)} \tau)} \frac{\prod_{u=1}^U \theta_{u1}^{\theta_{u1}^{(0)} \tau - 1}}{D(\boldsymbol{\theta}_1^{(0)} \tau)} \frac{1}{(1+\tau)^2}, \end{aligned} \tag{2.4}$$

where $\sum_{u=1}^U g_{u0} = n - s$ and $\sum_{u=1}^U g_{u1} = s$. For convenience, we drop $\boldsymbol{\pi}$ from the conditioning.

Using the joint posterior density in (2.4) and assuming the π_u^* are fixed and known, to perform the Gibbs sampler, we need the conditional posterior densities, given by

$$\tilde{g}_1(\boldsymbol{\theta}_0 | \boldsymbol{\theta}_1, p, \tau, \boldsymbol{y}_s) \propto \frac{\prod_{u=1}^U (\pi_u^* \theta_{u0})^{g_{u0}}}{[a_1 p + a_0 (1-p)]^n} \prod_{u=1}^U \theta_{u0}^{\theta_{u0} \tau - 1},$$

$$\tilde{g}_2(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_0, p, \tau, \boldsymbol{y}_s) \propto \frac{\prod_{u=1}^U (\pi_u^* \theta_{u1})^{g_{u1}}}{[a_1 p + a_0 (1-p)]^n} \prod_{u=1}^U \theta_{u1}^{\theta_{u1} \tau - 1},$$

$$\tilde{g}_3(p | \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau, \boldsymbol{y}_s) \propto \frac{1}{[a_1 p + a_0 (1-p)]^n} p^s (1-p)^{n-s},$$

and

$$\tilde{g}_4(\tau \mid \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, p, \boldsymbol{y}_s) \propto \left(\frac{\Gamma(\tau)}{1+\tau} \right)^2 \prod_{u=1}^U \left\{ \frac{\theta_{u0}^{\theta_{u0} \tau - 1}}{\Gamma(\theta_{u0} \tau)} \frac{\theta_{u1}^{\theta_{u1} \tau - 1}}{\Gamma(\theta_{u1} \tau)} \right\},$$

where $a_y = \sum_u \pi_u^* \theta_{uy},\ y = 0, 1$.

Because of accept-reject method within the Gibbs chain, this Gibbs sampler is a bit slow. We use an alternative procedure that avoids the accept-reject algorithm within the Gibbs sampler. We make a one-to-one transformation from p to q via $q = \frac{a_1 p}{a_1 p + a_0 (1-p)}$. To accelerate the Gibbs sampler, we integrate out q from the joint posterior. So we generate samples for p independently with other parameters using the simple accept-reject algorithm. Then we execute the Gibbs sampler for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_0, \tau$ using a grid method. For details, see Nandram *et al.* (2013).

Once p is estimated, we draw the entire finite population values, $y_1, \ldots, y_N$, independently from Bernoulli(p). Here, we simply need $\sum_{i=1}^N y_i \mid p \sim \text{Binomial}(N, p)$. So really we have corrected the observed biased sample and replaced it by a surrogate sample for every p that we obtained from the nonignorable selection model.

2.3. Estimation

Let $\pi(p \mid \boldsymbol{y}_s, \boldsymbol{\pi})$ denote the posterior density of p. Note again that p is the proportion of ones in the entire superpopulation (i.e., the selection bias has been removed). Thus,

$$\Pi(P \mid \boldsymbol{y}_s, \boldsymbol{\pi}) = \int \pi(P \mid p)\pi(p \mid \boldsymbol{y}_s, \boldsymbol{\pi})dp.$$

Once samples are obtained from the posterior density of p, it is now easy to take a census of the entire population using the composition method. To every sample of $p^{(t)}$, $t = 1, ..., T$, obtained from the Gibbs sampler, a sample of P, say $P_1^{(t)}$, is obtained by drawing $\sum_{i=1}^N y_i$ from Binomial$(N, p^{(t)})$ and dividing the result by N. Thus, we have obtained a Rao-Blackwellized estimator of the posterior density of P. The postrior mean is given by $\hat{P}_1 = \sum_{t=1}^T P_1^{(t)}/T$.

Alternatively we can use the posterior density of q as a mixing distribution in the composition method. To every sample of $q^{(t)}$, $t = 1, ..., T$, obtained from the Gibbs sampler, we obtain a sample of P, say $P_2^{(t)}$, by drawing $\sum_{i=1}^N y_i$ from Binomial$(N, q^{(t)})$ and dividing it by N. The postrior mean is given by $\hat{P}_2 = \sum_{t=1}^T P_2^{(t)}/T$.

Finally we write $P = (\sum_{i\in S} y_i + \sum_{i\in \bar{S}} y_i)/N$. Since the seen part $s = \sum_{i\in S} y_i$ is known, we obtain a sample of P, say $P_3^{(t)}$, by drawing $\sum_{i\in \bar{S}} y_i$ from Binomial$(N - n, p^{(t)})$ using every sample of $p^{(t)}$, $t = 1, ..., T$, obtained from the Gibbs sampler. Thus we obtain the postrior mean $\hat{P}_3 = \sum_{t=1}^T \hat{P}_3^{(t)}/T$.

Remark 2.1 An ignorable selection model for the binary variables $y_i, i = 1, \ldots, N$, is

$$y_i \mid p \overset{iid}{\sim} \text{Bernoulli}(p) \text{ and } p \sim \text{Uniform}(0, 1).$$

Then $p \mid s \overset{ind}{\sim} \text{Beta}(s + 1, n - s + 1)$. Since $\sum_{i\in \bar{S}} y_i | p \sim \text{Binomial}(N - n, p)$, we can draw directly $\sum_{i\in \bar{S}} y_i$ using samples of p obtained from the Beta posterior. Thus we can obtain the estimator of $P = (s + \sum_{i\in \bar{S}} y_i)/N$ easily. We call it $\hat{P}_0$.

3. Numerical studies

We illustrate the results in preceding section with an analysis of one simulated data set. We generate a finite population of size $N = 500$ with selection probability π. Then a sample of size $n = 50$ is taken from this finite population. The outline of the generation is as follows.

- Set $\tau = 100$, $p = 0.5$, $f = n/N$, $a = 0.9$, $\mu_0 = af$, $\mu_1 = f/a$.
- Generate $u \sim U(0, 1)$. If $u \leq p$ then we set $y_i = 1$; otherwise set $y_i = 0$, $i = 1, \cdots, N$.
- If $y_i = 1$ then we generate $\pi_i \sim Beta(\mu_1\tau, (1 - \mu_1)\tau)$; if $y_i = 0$ then we generate $\pi_i \sim Beta(\mu_0\tau, (1 - \mu_0)\tau)$ for $i = 1, \cdots, N$.
- Sample n units by systematic PPS sampling with probabilities $\pi_i = \frac{n\pi_i}{\sum_{i=1}^N \pi_i}$.

Using the simulated data set, we compare the four estimates for the finite population proportion $P = \sum_{i=1}^N y_i$. Specifically we calculate the posterior mean (PM), the posterior standard deviation (PSD), the 95% credible interval (CrI), and the 95% HPD interval (HPD).

The results are provided in Table 3.1 and are robust to different τ, p and a values. From Table 3.1, we may find that the estimator $\hat{P}_1$ is well behaved in the sense that it is closer to true $P = 0.48$ than others. But from Table 3.1, it appears that the estimator $\hat{P}_3$ provides shorter 95% credible interval and 95% HPD interval than others.

Table 3.1 Posterior means, associated posterior standard deviations, and credible intervals

Estimator	PM	PSD	CrI	HPD
$\hat{P}_1$	.490	.071	(.352, .632)	(.332, .640)
$\hat{P}_2$	.538	.070	(.404, .674)	(.378, .688)
$\hat{P}_3$	.494	.064	(.374, .618)	(.364, .628)
$\hat{P}_0$	.536	.069	(.416, .664)	(.398, .668)

The four estimators of the posterior density of P are shown in Figure 3.1. Notice that the solid line represents the posterior density for the finite population proportion and the dashed vertical line represents the true finite population proportion $P = 0.48$ in Figure 3.1. As one might expect, the posterior densities corresponding to $\hat{P}_1$ and $\hat{P}_3$ looks much better than others in the closeness of the true P.

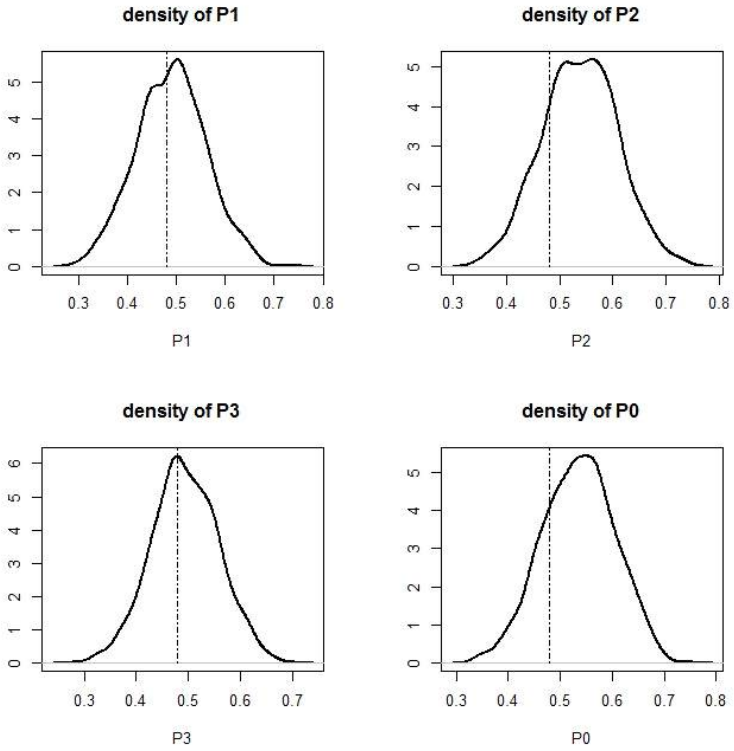

Figure 3.1. The four posterior densities of the finite population proportion

Next, in order to assess the performance of the estimators based on nonignorable selection model in the selection bias problem, we design a simulation study. Using the same algorithm as the above, we generate 1,000 binary populations of size $N = 500$ with selection probability

π with $a = 1.0, 0.9, 0.95, 0.98, 0.99, 0.995$. Then a sample of size $n = 50$ is taken from each finite population.

To compare the performance of four possible estimators, we compute the several frequentist measures. First, we calculate the finite population proportion $P^{(h)}$, the posterior mean $PM^{(h)}$ and the posterior standard deviation $PSD^{(h)}$, $h = 1, ..., 1000$. Then we compute the absolute bias $AB^{(h)} = |PM^{(h)} - P^{(h)}|$ and the root mean squared error $RMSE^{(h)} = \sqrt{PSD^{(h)2} + AB^{(h)2}}$, $h = 1, ..., 1000$. Using these frequentist quantities we obtain $AB = \frac{1}{1000}\sum_{h=1}^{1000} AB^{(h)}$ and $RMSE = \frac{1}{1000}\sum_{h=1}^{1000} RMSE^{(h)}$. Also we compute the 95% credible interval for each of the 1,000 simulated runs. Then we look at the width ($W^{(h)}$) and the credible incidence ($I^{(h)}$), where $W^{(h)}$ means the length of the 95% credible interval and $I^{(h)} = 1$ if the 95% credible interval contains the true value P and $I^{(h)} = 0$ otherwise. Then we calculate $C = \sum_{h=1}^{1000} I^{(h)}/1000$ and $W = \sum_{h=1}^{1000} W^{(h)}/1000$. The results with numerical standard errors (NSE) are reported in Table 3.2.

Table 3.2 Comparison of four estimates based on absolute bias, root posterior mean squared error, coverage and width of 95% credible intervals

a	Estimators	AB	NSE_{AB}	$RMSE$	NSE_{RMSE}	C	NSE_C	W	NSE_W
1.0	$\hat{P}_1$	.002	.004	.119	.001	.994	.002	.381	.001
1.0	$\hat{P}_2$	.000	.004	.093	.001	.953	.007	.276	.000
1.0	$\hat{P}_3$	.102	.004	.117	.001	.981	.004	.344	.001
1.0	$\hat{P}_0$	.100	.004	.098	.001	.896	.010	.253	.000
.90	$\hat{P}_1$	.046	.004	.121	.001	.998	.001	.383	.001
.90	$\hat{P}_2$	.052	.004	.097	.001	.951	.007	.276	.000
.90	$\hat{P}_3$	.058	.004	.110	.001	.992	.003	.345	.001
.90	$\hat{P}_0$	.051	.004	.089	.001	.944	.007	.252	.000
.95	$\hat{P}_1$	.046	.004	.121	.001	.998	.001	.383	.001
.95	$\hat{P}_2$	.052	.004	.097	.001	.951	.007	.276	.000
.95	$\hat{P}_3$	.058	.004	.110	.001	.992	.003	.345	.001
.95	$\hat{P}_0$	.051	.004	.089	.001	.944	.007	.252	.000
.98	$\hat{P}_1$	.022	.004	.120	.001	.991	.003	.384	.001
.98	$\hat{P}_2$	.022	.004	.094	.001	.963	.006	.276	.000
.98	$\hat{P}_3$	.080	.004	.113	.001	.983	.004	.347	.001
.98	$\hat{P}_0$	.079	.004	.093	.001	.917	.009	.253	.000
.99	$\hat{P}_1$	.007	.004	.121	.001	.993	.003	.385	.001
.99	$\hat{P}_2$	.012	.004	.094	.001	.964	.006	.275	.000
.99	$\hat{P}_3$	.094	.004	.117	.001	.984	.004	.348	.001
.99	$\hat{P}_0$	.089	.004	.096	.001	.915	.009	.253	.000
.995	$\hat{P}_1$	.001	.004	.118	.001	.989	.003	.383	.001
.995	$\hat{P}_2$	.001	.004	.093	.001	.961	.006	.275	.000
.995	$\hat{P}_3$	.099	.004	.116	.001	.983	.004	.345	.001
.995	$\hat{P}_0$	.099	.004	.097	.001	.911	.009	.253	.000

An inspection of Table 3.2 reveals that the estimators $\hat{P}_1$ and $\hat{P}_2$ are better than others in the sense of the closeness to the true P, but $\hat{P}_0$ and $\hat{P}_2$ is better than others in the sense of the root mean squared error. Moreover, the coverage probability of $\hat{P}_2$ is closest to the nominal value of the 95%. Also the estimator $\hat{P}_0$ is the best and $\hat{P}_2$ is the second best in the sense of the width of the credible interval. The simulation results seems to be very similar for a less than .9.

4. Concluding remarks

The analysis of one simulated data set reports that $\hat{P}_1$ and $\hat{P}_3$ are the good candidates in Bayesian estimation for the finite population proportion under the nonignorable selection model. But the simulation study indicates that there is no clear winner among four possible estimators in the sense of frequentist measures, but overall $\hat{P}_2$ would be a good choice in selection bias problem. Our simulation study reveals that the nonignorable selection model might be useful in the case of the weekly biased samples.

References

Kwak, S. and Kim, D. (2012). Bayesian estimation for finite population proportions in multinomial data. *Journal of the Korean Data & Information Science Society*, **23**, 587-593.

Malec, D., Davis, W. W. and Cao, X. (1999). Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine*, **18**, 3189-3200.

Nandram, B. (2007). Bayesian predictive inference under informative sampling via surrogate samples. In *Bayesian Statistics and Its Applications*, edited by S.K. Upadhyay et al., Anamaya, New Delhi, Chapter 25, 356-374.

Nandram, B. (2013). Bayesian predictive inference of a finite population proportion under selection bias. *Statistical Methodology*, **11**, 1-21.

Nandram, B., Bhatta, D., Bhadra, D. and Shen, G. (2013). Bayesian predictive inference of a finite population proportion under selection bias. *Statistical Methodology*, **11**, 1-21.

Nandram, B. and Choi, J. W. (2010). A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *Journal of the American Statistical Association*, **105**, 120-135.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, **34**, 179-189.