Journal of the Korean Statistical Society (2002), **31**: *1, pp 45–61*

Hierarchical Bayesian Inference of Binomial Data with Nonresponse

Geunshik Han[1] and Balgobin Nandram[2]

ABSTRACT

We consider the problem of estimating binomial proportions in the presence of nonignorable nonresponse using the Bayesian selection approach. Inference is sampling based and Markov chain Monte Carlo (MCMC) methods are used to perform the computations. We apply our method to study doctor visits data from the the Korean National Family Income and Expenditure Survey (NFIES). The ignorable and nonignorable models are compared to Stasny's method (1991) by measuring the variability from the Metropolis-Hastings (MH) sampler. The results show that both models work very well.

Keywords: Beta-binomial model, ignorable nonresponse, latent variable, selection approach, Metropolis-Hastings sampler.

1. Introduction

The nonresponse rates in many surveys have been increasing steadily (De Heer 1999, and Groves and Couper 1998), making the nonresponse problem more important.

For the analysis of survey data with nonresponse, Rubin (1987) and Little and Rubin (1987) describe two types of models according to the ignorability of response. In the ignorable nonresponse model the distribution of the variable of interest for a respondent is the same as the distribution of that variable for a nonrespondent with the same values of the covariates. In addition, the parameters in the distributions of the variable and response must be distinct (Rubin, 1976). All other models are nonignorable.

We consider a hierarchical Bayes model to study nonignorable nonresponse. Recently there has been much activity in the analysis of survey nonresponse.

[1]Department of Computer Science & Statistics, Hanshin University 411, Yangsan-Dong, Osan, Korea

[2]Mathematical Sciences, Worcester Polytechnic Institute

Bradlow and Zaslavsky (1999) estimate item nonresponse in the 1992 DuPont Corporation customer satisfaction survey data by Bayesian hierarchical latent variable nonresponse model for ordinal data. Scharfstein, Rotinitzky and Robins (1999) show how to make inference using semiparametric nonresponse models. Crawford, Johnson and Laird (1993) used nonresponse models to analyze data from the Harvard Medical Practice Survey. Stasny, Kadane, and Fritsch (1998) used a Bayesian hierarchical model for the probabilities of voting guilty or not guilty on a particular trial given various death-penalty beliefs in which the views of nonrespondents may differ from those of respondents. Park and Brown (1994) used a pseudo-Bayesian method (Baker and Laird, 1988) in which prior observations are assigned to both observed and unobserved cells to estimate the missing cells of a multi-way categorical table under nonresponse.

Stasny (1991) used a hierarchical Bayesian model to study victimization in the National Crime Survey and used the selection approach developed primarily to study sample selection problems (e.g., Heckman 1976, and Olson 1980). A related method was presented by Albert and Gupta (1985), they made an approximation to obtain a Bayesian approach for a single area, see also (Kaufman and King, 1973).

Since Bayesian approach can incorporate prior information about nonrespondents, the Bayesian method is appropriate for the analysis of nonignorable nonresponse problems (Little and Rubin 1987, and Rubin 1987). However the main difficulty is how to describe the relationship between the respondents and nonrespondents. Using the selection approach within the framework of Bayes empirical Bayes (Deely and Lindley, 1981), Stasny (1991) estimated the hyper-parameters by maximum likelihood methods and assumed them as known. But we assume that the parameter for nonresponses is differnt from that of respondents and estimate hyperparameters by random priors.

The rest of the paper is organized as follows. In Section 2 we briefly illustrate the NFIES data and nonresponse problems. In Section 3 we describe the hierarchical Bayesian model for ignorable and nonignorable nonresponses and also computational procedures. In Section 4 we show how to obtain variance estimates and an empirical analysis to assess the performance of our model. Finally, Section 5 has concluding remarks.

2. National Family Income and Expenditure Survey

The National Family Income and Expenditure Survey (NFIES) has been conducted every five year since 1991 by the National Statistical Office to measure an aspect of income and expenditure structure of Korean household (National Survey of Family Income and Expenditure Vol. 2). For this survey, a preliminary survey was conducted in 1989 for urban area and in 1990 for rural area. The data used in this paper was collected between October 1996 and December 1996.

The NFIES frame is a stratified two-stage sample design of probability proportional to measure of size of population. The first stage is for the selection of primary sampling units (Enumemeration Districts : ED), and second stage for selection of segments. All households in each selected segment are the sampling unit of the survey at the initial interview, and each household is received account books to record the receipts and disbursements. The survey data was collected by using family account books which were distributed to each sample households prior to the survey so that income source, type of expenditure and their value could be recorded daily. At the final interview, the information about status of households, yearly income, savings & liabilities and durable goods were obtained upon interview and recorded on the questionnaires by the enumerators.

One of the variables of interest in the NFIES is doctor visits (at least one doctor visit during the past year or no visits) of households.

The NFIES nonresponses are refusals (to protect privacy), noncontacts. They may arise nonrandomly. For example, households of the wealthy neighborhood gave low response, and therefore, nonresponse from this source can be considered nonrandom nonresponses.

The average NFIES response rate was about 80%. The nonresponses of sample households were imputed assuming the distribution of the respondents and nonrespondents are the same within the same imputation class. The order of hierarchical imputation classes were area, household type and household size. However this hierarchical imputation method may be unreliable, thus there is a need to consider the adjustment by a method other than random hierarchical imputation. The Bayesian method is discussed as a possible alternative to impute the NFIES nonresponses. For each area in Table 1 n, r, and y denote the number of sampled households, the number of households who completed daily account books and the number of households who visited doctor at least once respectively.

3. Hierarchical Bayes Models

In this section we describe the hierarchical Bayes model for nonresponse. This model is a standard beta-binomial model for ignorable nonresponse. Let

$$x_{ij} = \begin{cases} 1, & \text{household } j \text{ in area } i \text{ visited doctor at least once} \\ 0, & \text{household } j \text{ in area } i \text{ did not visit doctor} \end{cases}$$

and

$$y_{ij} = \begin{cases} 1, & \text{household } j \text{ in area } i \text{ is a respondent} \\ 0, & \text{household } j \text{ in area } i \text{ is not a respondent,} \end{cases}$$

$i = 1, \ldots, \ell,\ \ j = 1, \ldots, n_i$. We use a probabilistic structure to model the x_{ij} and y_{ij}.

3.1. Ignorable versus nonignorable model

For the ignorable nonresponse, we take

$$x_{ij} \mid p_i \overset{iid}{\sim} \text{Bernoulli } (p_i)$$
$$y_{ij} \mid \pi_i \overset{iid}{\sim} \text{Bernoulli } (\pi_i)$$

$$p_i \mid \mu_{11}, \tau_{11} \overset{iid}{\sim} \text{Beta } (\mu_{11}\tau_{11}, (1-\mu_{11})\tau_{11}) \tag{3.1}$$

$$\pi_i \mid \mu_{12}, \tau_{12} \overset{iid}{\sim} \text{Beta } (\mu_{12}\tau_{12}, (1-\mu_{12})\tau_{12}) \tag{3.2}$$

$$i = 1, \ldots, \ell;\ j = 1, \ldots, n_i.$$

Assumptions (3.1) and (3.2) express similarity among the areas.
For the nonignorable nonresponse, we have

$$x_{ij} \mid p_i \overset{iid}{\sim} \text{Bernoulli } (p_i)$$
$$y_{ij} \mid x_{ij} = s-1, \pi_{is} \overset{iid}{\sim} \text{Bernoulli } (\pi_{is}), \quad s = 1, 2$$
$$p_i \mid \mu_{21}, \tau_{21} \overset{iid}{\sim} \text{Beta } (\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21}) \tag{3.3}$$

$$\pi_{is} \mid \mu_{2,s+1}, \tau_{2,s+1} \overset{iid}{\sim} \text{Beta } (\mu_{2,s+1}\tau_{2,s+1}, (1-\mu_{2,s+1})\tau_{2,s+1}), \quad s = 1, 2. \tag{3.4}$$

Like assumptions (3.1) and (3.2), (3.3) and (3.4) express similarity among the areas.

Stasny did not work with the reparameterized beta distribution that we are using here. For the nonignorable nonresponse model she set $a = \mu_{21}\tau_{21}$, $b =$

$(1-\mu_{21})\tau_{21}$, $\alpha_0 = \mu_{22}\tau_{22}$, $\beta_0 = (1-\mu_{22})\tau_{22}$, $\alpha_1 = \mu_{23}\tau_{23}$, $\beta_1 = (1-\mu_{23})\tau_{23}$. There two advantages in working with this reparameterization : (a) the μ_{2s} are all in (0, 1) and this helps to get around the issue of impropriety (b) the parameters μ_{2s} and τ_{2s} are roughly orthogonal. Both (a) and (b) improve the computations.

While Stasny assumed that the hyperparameters a, b, α_s and β_s, are fixed but unknown, we take

$$\mu_{rs} \sim \text{Beta}\ (\mu_{rs}^{(0)}\ \tau_{rs}^{(0)}, (1-\mu_{rs}^{(0)})\tau_{rs}^{(0)})$$

$$\tau_{rs} \sim \Gamma\ (\eta_{rs}^{(0)}, \nu_{rs}^{(0)}),$$

where for the ignorable model $r = 1$, $s = 1, 2$ and for nonignorable model $r = 2$, $s = 1, 2, 3$. In our case the parameters $\mu_{rs}^{(0)}$, $\tau_{rs}^{(0)}$, $\eta_{rs}^{(0)}$ and $\nu_{rs}^{(0)}$ are to be specified.

Ignorable model specifies a complete Bayesian model in which the probability that a household respond does not depend on its doctor visit whereas nonignorable model has this specification. Thus, in ignorable model, but not in nonignorable model, the missing data are missing at random and the missing data are not used for inference about the p_i. In this sense our first model is ignorable.

Let r_i be the number of responded households and y_i the number of visits in area i. Then r_i and y_i are random variables, $n_i - r_i$ is the number of nonresponded households. Since the number of visits for the nonresponded households is unknown, we denote it by the latent variable z_i, and the number of non visits is $n_i - r_i - z_i$.

For ignorable model the likelihood function is

$$p(\mathbf{p}, \boldsymbol{\pi} \mid \mathbf{y}, \mathbf{r}) \propto \prod_{i=1}^{\ell} \{p_i^{y_i}(1-p_i)^{r_i - y_i}\pi_i^{r_i}(1-\pi_i)^{n_i - r_i}\}$$

and using Bayes' theorem the joint posterior density of all the parameters is

$$\begin{aligned} &f(\mathbf{p}, \boldsymbol{\pi}, \mu_{11}, \tau_{11}, \mu_{12}, \tau_{12} \mid \mathbf{y}, \mathbf{r}) \\ &\propto \prod_{i=1}^{\ell} \Bigg\{ p_i^{y_i}(1-p_i)^{r_i - y_i}\pi_i^{r_i}(1-\pi_i)^{n_i - r_i} \\ &\quad \times \frac{p_i^{\mu_{11}\tau_{11}-1}(1-p_i)^{(1-\mu_{11})\tau_{11}-1}}{B(\mu_{11}\tau_{11}, (1-\mu_{11})\tau_{11})} \frac{\pi_i^{\mu_{12}\tau_{12}-1}(1-\pi_i)^{(1-\mu_{12})\tau_{12}-1}}{B(\mu_{12}\tau_{12}, (1-\mu_{12})\tau_{12})} \Bigg\} \\ &\quad \times \prod_{s=1}^{2} \left\{ \mu_{1s}^{\mu_{1s}^{(0)}\tau_{1s}^{(0)}-1}(1-\mu_{1s})^{(1-\mu_{1s}^{(0)})\tau_{1s}^{(0)}-1}\tau_{1s}^{\eta_{1s}^{(0)}-1}\exp(-\nu_{1s}^{(0)}\tau_{1s}) \right\}, \end{aligned}$$

where $B(u,v) = \Gamma(u)\Gamma(v)/\Gamma(u+v)$ is the beta function.

For nonignorable model the likelihood function is

$$
\begin{aligned}
&p(\mathbf{p}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \mathbf{z} \mid \mathbf{y}, \mathbf{r}) \\
&\propto \prod_{i=1}^{\ell} \left\{ \binom{n_i}{y_i, r_i - y_i, z_i, n_i - r_i - z_i} (\pi_{i2} p_i)^{y_i} \right. \\
&\quad \left. \times (\pi_{i1}(1-p_i))^{r_i - y_i} ((1-\pi_{i2})p_i)^{z_i} ((1-\pi_{i1})(1-p_i))^{n_i - r_i - z_i} \right\}
\end{aligned}
$$

and using Bayes' theorem the joint posterior density of all the parameters is

$$
\begin{aligned}
&f(\mathbf{p}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \mathbf{z}, \mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}, \tau_{23} \mid \mathbf{y}, \mathbf{r}) \\
&\propto \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} (\pi_{i2} p_i)^{y_i} (\pi_{i1}(1-p_i))^{r_i - y_i} ((1-\pi_{i2})p_i)^{z_i} \right. \\
&\quad \times ((1-\pi_{i1})(1-p_i))^{n_i - r_i - z_i} \frac{p_i^{\mu_{21}\tau_{21}-1}(1-p_i)^{(1-\mu_{21})\tau_{21}-1}}{B(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21})} \\
&\quad \left. \times \prod_{s=2}^{3} \left\{ \frac{\pi_{is}^{\mu_{2s}\tau_{2s}-1}(1-\pi_{is})^{(1-\mu_{2s})\tau_{2s}-1}}{B(\mu_{2s}\tau_{2s}, (1-\mu_{2s})\tau_{2s})} \right\} \right\} \\
&\quad \times \prod_{s=1}^{3} \left\{ \mu_{2s}^{\mu_{2s}^{(0)}\tau_{2s}^{(0)}-1} (1-\mu_{2s})^{(1-\mu_{2s}^{(0)})\tau_{2s}^{(0)}-1} \tau_{2s}^{\eta_{2s}^{(0)}-1} \exp(-\nu_{2s}^{(0)}\tau_{2s}) \right\}.
\end{aligned}
$$

Inference about p_i, π_{i1} and π_{i2} is required, and all information about these parameters is contained in this posterior density which is complex. We develope the conditional densities used in the Gibbs sampling algorithm. Gibbs sampling is a MCMC technique. The transitional density of this Markov chain is the product of several conditional densities. The stationary distribution of the chain is the posterior density that we desire. We list the conditional densities for the followings.

3.2. Computations

For ignorable model it is convenient to represent the joint posterior density function as

$$
\begin{aligned}
&f(\mathbf{p}, \boldsymbol{\pi}_1, \mu_{11}, \tau_{11}, \mu_{12}, \tau_{12} \mid \mathbf{y}, \mathbf{r}) \\
&\quad = f_1(\mathbf{p} \mid \mathbf{y}, \mathbf{r}, \mu_{11}, \tau_{11}) f_2(\boldsymbol{\pi}_1 \mid \mathbf{y}, \mathbf{r}, \mu_{12}, \tau_{12}) f_3(\mu_{11}, \tau_{11}, \mu_{12}, \tau_{12} \mid \mathbf{y}, \mathbf{r}),
\end{aligned}
$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are respectively the densities of $\mathbf{p}$ and $\boldsymbol{\pi}_1$ which have the following distributions

$$p_i \mid y_i, r_i, \mu_{11}, \tau_{11} \overset{ind}{\sim} \text{Beta}(y_i + \mu_{11}\tau_{11}, r_i - y_i + (1-\mu_{11})\tau_{11}), \quad i = 1, \ldots, \ell$$

$$\pi_i \mid y_i, r_i, \mu_{12}, \tau_{12} \overset{ind}{\sim} \text{Beta}\ (r_i + \mu_{12}\tau_{12}, n_i - r_i + (1-\mu_{12})\tau_{12}), \quad i = 1, \ldots, \ell$$

and

$$\begin{aligned} &f_3(\mu_{11}, \tau_{11}, \mu_{12}, \tau_{12} \mid \mathbf{y}, \mathbf{r}) \\ &\propto \prod_{i=1}^{\ell} \left\{ \frac{B(y_i + \mu_{11}\tau_{11},\ r_i - y_i + (1-\mu_{11})\tau_{11})}{B(\mu_{11}\tau_{11}, (1-\mu_{11})\tau_{11})} \right. \\ &\quad \left. \times \frac{B(r_i + \mu_{12}\tau_{12}, n_i - r_i + (1-\mu_{12})\tau_{12})}{B(\mu_{12}\tau_{12}, (1-\mu_{12})\tau_{12})} \right\} p(\mu_{11}, \tau_{11}, \mu_{12}, \tau_{12}) \end{aligned}$$

with $p(\mu_{11}, \tau_{11}, \mu_{12}, \tau_{12})$ the joint prior distribution. For nonignorable model it is convenient to represent the joint posterior density function as

$$\begin{aligned} &f(\mathbf{p}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \mathbf{z}, \mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}, \tau_{23} \mid \mathbf{y}, \mathbf{r}) \\ &= f_1(\mathbf{p} \mid \mathbf{y}, \mathbf{r}, \mathbf{z}, \mu_{21}, \tau_{21}) \left\{ \prod_{s=2}^{3} f_s(\boldsymbol{\pi}_s \mid \mathbf{y}, \mathbf{r}, \mathbf{z}, \mu_{2s}, \tau_{2s}) \right\} \\ &\quad \times f_4(\mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}, \tau_{23}, \mathbf{z} \mid \mathbf{y}, \mathbf{r}), \end{aligned}$$

where $f_1(\cdot)$, $f_2(\cdot)$ and $f_3(\cdot)$ are respectively the densities of $\mathbf{p}, \boldsymbol{\pi}_1$ and,$\boldsymbol{\pi}_2$ which have the following distributions

$$p_i \mid y_i, r_i, z_i, \mu_{21}, \tau_{21} \overset{ind}{\sim} \text{Beta}(z_i + y_i + \mu_{21}\tau_{21}, n_i - z_i - y_i + (1-\mu_{21})\tau_{21}), \quad i = 1, \ldots, \ell$$

$$\pi_{i1} | \, y_i, r_i, z_i, \mu_{22}, \tau_{22} \overset{ind}{\sim} \text{Beta}(r_i - y_i + \mu_{22}\tau_{22}, n_i - r_i - z_i + (1-\mu_{22})\tau_{22}), \quad i = 1, \ldots, \ell$$

$$\pi_{i2} \mid y_i, r_i, z_i, \mu_{23}, \tau_{23} \overset{ind}{\sim} \text{Beta}(y_i + \mu_{23}\tau_{23}, z_i + (1-\mu_{23})\tau_{23}), \quad i = 1, \ldots, \ell$$

and

$$\begin{aligned} &f_4(\mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}, \tau_{23}, \mathbf{z} \mid \mathbf{y}, \mathbf{r}) \\ &\propto \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} \frac{B(z_i + y_i + \mu_{21}\tau_{21},\ n_i - z_i - y_i + (1-\mu_{21})\tau_{21})}{B(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21})} \right. \\ &\quad \times \frac{B(r_i - y_i + \mu_{22}\tau_{22}, n_i - r_i - z_i + (1-\mu_{22})\tau_{22})}{B(\mu_{22}, \tau_{22})} \\ &\quad \left. \times \frac{B(y_i + \mu_{23}\tau_{23}, z_i + (1-\mu_{23})\tau_{23})}{B(\mu_{23}\tau_{23}, (1-\mu_{23})\tau_{23})} \right\} p(\mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}, \tau_{23}), \end{aligned}$$

where $p(\mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}, \tau_{23})$ is the joint prior distribution. Thus, random samples from p_i, π_{i1} and π_{i2} can be easily obtained. However, the conditional densities of $\mu_{21}, \tau_{21}, \mu_{22}, \tau_{22}, \mu_{23}$, and τ_{23} are not easily defined, so we use to sample these parameters. (see Appendix 1).

The MH algorithm of Nandram (1998) requires some dexterity to obtain tuning constants.

We have used trace plots, autocorrelations, and a maintenance of good jumping probabilities to assess convergence. We draw 5,500 iterates, threw out the first 500 and took every fifth. However, this is very conservative because convergence is very rapid. For our computation, first we set the hyper-parameters $\mu_{rs}^{(0)}$ and $\tau_{rs}^{(0)}$ equal 0. Then we ran our MH algorithm to obtain posterior samples of τ_{rs}. To ensure propriety of the posterior we estimate $\eta_{11}^{(0)}$, $\nu_{11}^{(0)}$, $\eta_{12}^{(0)}$, $\nu_{12}^{(0)}$, $\eta_{21}^{(0)}$, $\nu_{21}^{(0)}$, $\eta_{22}^{(0)}$, $\nu_{22}^{(0)}$, $\eta_{23}^{(0)}$ and $\nu_{23}^{(0)}$. Finally, we ran our algorithm with these proper priors, and we found virtually no change in the posterior distributions of the p_i.

4. Comparison of Variance Estimates

We compare our approach with that of Stasny. We note here that she did not obtain posterior standard deviations of the proportions. Within her framework, we show how to obtain reasonable estimates of the posterior standard estimates, and we compare them with the corresponding estimates from our method. We perform these activities by using data on the NFIES.

We present the data for the 24 areas in Table 1 where y_i is the number of doctor visits in area i, r_i is the number of responded households and n_i is the number of households sampled. The observed doctor visit rate is $\hat{p}_i = y_i/r_i$ and the response rate is $\hat{\pi}_i = r_i/n_i$.

4.1. Approximating the variance

Stasny used the standard beta priors with the hyperparameters $a = \mu_{21}\tau_{21}$, $b = (1-\mu_{21})\tau_{21}$, $\alpha_0 = \mu_{22}\tau_{22}$, $\beta_0 = (1-\mu_{22})\tau_{22}$, $\alpha_1 = \mu_{23}\tau_{23}$ and $\beta_1 = (1-\mu_{23})\tau_{23}$ being fixed but unknown.

We consider a simple procedure to approximate the posterior standard deviations of the proportions. This is a good idea if it turns out to be similar to the ones obtained by using the MCMC sampler.

Let q_i denote one of p_i, π_{i1} and $\pi_{i2}, \mathcal{D} = (\mathbf{r}, \mathbf{y})$ and $\Omega = (a, b, \alpha_0, \beta_0, \alpha_1, \beta_1)$. Also let (ϕ_0, ϕ_1) denote one of $(a, b), (\alpha_0, \beta_0)$ and (α_1, β_1), and $(\hat{\phi}_0, \hat{\phi}_1)$ be reason-

TABLE 1. *National family income and expenditure survey data*

Area	y_i	$n_i - r_i$	Naive Estimator $\hat{\pi}_i$	$\hat{p}_i$
area1	1127	692	.747	.551
area2	1004	474	.793	.553
area3	686	238	.840	.547
area4	891	420	.784	.584
area5	723	131	.911	.540
area6	665	324	.778	.585
area7	797	337	.800	.590
area8	357	147	.798	.616
area9	352	42	.938	.557
area10	175	29	.920	.524
area11	343	130	.812	.609
area12	133	44	.841	.571
area13	280	79	.860	.579
area14	137	49	.838	.539
area15	349	92	.864	.598
area16	77	29	.826	.558
area17	284	35	.934	.569
area18	117	12	.948	.534
area19	301	65	.886	.595
area20	180	33	.907	.563
area21	531	132	.876	.570
area22	169	34	.903	.531
area23	169	42	.879	.556
area24	63	27	.789	.624

NOTE : *The sample size from each area is n_i, r_i and y_i are the numbers of responded households and doctor visits respectively; $\hat{p}_i = y_i/r_i$ and $\hat{\pi}_i = r_i/n_i$ are the observed proportions of doctor visits and responded households.*

able estimates (e.g., posterior modes or means of maximum likelihood or method of moments estimation) of (ϕ_0, ϕ_1). Let (c_{it}, d_{it}) denote one of $(t + y_i, n_i)$ for p_i, $(r_i - y_i, n_i - y_i - t)$ for π_{i1} and $(y_i, y_i + t)$ for π_{i2}, $t = 0, 1, \ldots, n_i - r_i$. We also let $\hat{\Omega}$ be an estimator (e.g., posterior mean, mode or maximum likelihood estimator).

We approximate the posterior distribution of z_i as follows. $z_1, z_2, \ldots, z_l$ are iid. For $t = 0, 1, \ldots, n_i - r_i$,

$$P(z_i = t \mid \mathcal{D}) = \hat{\omega}_{it}$$

where $\hat{\omega}_{it} = \hat{\omega}^*_{it} / \sum_{s=0}^{n_i - r_i} \hat{\omega}^*_{is}$ and

$$\begin{aligned} \hat{\omega}^*_{it} &= \Big\{ B(y_i + t + \hat{a},\ n_i - t - y_i + \hat{b}) B(r_i - y_i + \hat{\alpha}_0,\ n_i - r_i - t + \hat{\beta}_0) \\ &\quad \times B(y_i + \hat{\alpha}_1, t + \hat{\beta}_1) \Big\} \div B(t + 1,\ n_i - r_i - t + 1) \end{aligned}$$

$t = 0, 1, \ldots, n_i - r_i$, $i = 1, \ldots, \ell$.

First, consider $E(q_i \mid \mathcal{D})$. It is true that

$$E(q_i \mid \mathcal{D}, z_i = t, \Omega) = \frac{c_{it} + \phi_0}{d_{it} + \phi_0 + \phi_1} = \mu_{it}.$$

Then

$$E(q_i \mid \mathcal{D}) \approx \sum_{t=0}^{n_i - r_i} \hat{\omega}_{it} \hat{\mu}_{it} = \bar{\hat{\mu}}_i,$$

where $\hat{\mu}_{it} = \dfrac{c_{it} + \hat{\phi}_0}{d_{it} + \hat{\phi}_0 + \hat{\phi}_1}$.

Second, we consider $\mathrm{var}(q_i \mid \mathcal{D})$. We use the conditional variance formula,

$$\mathrm{var}(q_i \mid \mathcal{D}) = E\{\mathrm{var}(q_i \mid \mathcal{D}, \mathbf{z}, \Omega)\} + \mathrm{var}\{E(q_i \mid \mathcal{D}, \mathbf{z}, \Omega)\},$$

where $\mathbf{z} = (z_1, \ldots, z_\ell)'$. Now

$$\begin{aligned} E\{\mathrm{var}(q_i \mid \mathcal{D}, \mathbf{z}, \Omega)\} &= E\{\mu_{it}(1 - \mu_{it})/(d_{it} + \phi_0 + \phi_1 + 1) \mid \mathcal{D}, \mathbf{z} = \mathbf{t}, \Omega\} \\ &\approx \sum_{t=0}^{n_i - r_i} \hat{\omega}_{it} \hat{\mu}_{it}(1 - \hat{\mu}_{it})/(d_{it} + \hat{\phi}_0 + \hat{\phi}_1 + 1). \end{aligned}$$

It is more difficult to obtain an approximation for $\mathrm{var}\{E(q_i \mid \mathcal{D}, \mathbf{z} = \mathbf{t}, \Omega)\}$. Letting $u(t, \Omega) = (c_{it} + \phi_0)/(d_{it} + \phi_0 + \phi_1)$ we have

$$\begin{aligned} \mathrm{var}(u(z, \Omega) \mid \mathcal{D}) &= E\{\mathrm{var}(u(z, \Omega) \mid \mathcal{D}, \mathbf{z})\} + \mathrm{var}\{E(u(z, \Omega) \mid \mathcal{D}, \mathbf{z})\} \\ &\approx \mathrm{var}\{u(\hat{t}_i, \Omega) \mid \mathcal{D}, \mathbf{z} = \hat{\mathbf{t}}\} + \mathrm{var}\{u(z_i, \hat{\Omega} \mid \mathcal{D}\}, \end{aligned}$$

TABLE 2. *Comparison of the posterior means and standard deviations for ignorable model based on the national family income and expenditure survey data*

	Stasny Estimates				*New Estimates*			
Area	p_i	$std(p_i)$	π_i	$std(\pi_i)$	p_i	$std(p_i)$	π_i	$std(\pi_i)$
area1	.553	.011	.749	.008	.552	.011	.749	.008
area2	.555	.011	.794	.008	.555	.011	.794	.008
area3	.550	.014	.840	.010	.550	.013	.841	.009
area4	.582	.012	.786	.010	.582	.012	.785	.009
area5	.542	.013	.909	.008	.543	.013	.909	.007
area6	.584	.014	.780	.011	.583	.014	.780	.011
area7	.588	.013	.802	.009	.588	.013	.801	.010
area8	.605	.019	.801	.015	.604	.018	.800	.014
area9	.559	.019	.933	.010	.560	.017	.933	.010
area10	.539	.024	.914	.014	.540	.022	.914	.014
area11	.599	.019	.814	.015	.600	.018	.815	.014
area12	.569	.027	.842	.021	.570	.024	.843	.021
area13	.576	.020	.860	.014	.576	.019	.859	.014
area14	.552	.027	.840	.020	.552	.024	.840	.020
area15	.591	.019	.864	.013	.591	.018	.863	.013
area16	.563	.032	.831	.027	.564	.027	.831	.026
area17	.569	.020	.929	.011	.569	.019	.929	.011
area18	.549	.028	.935	.016	.550	.025	.935	.015
area19	.588	.019	.884	.013	.587	.019	.884	.013
area20	.565	.023	.902	.015	.565	.022	.901	.015
area21	.569	.016	.875	.010	.569	.015	.875	.010
area22	.545	.025	.898	.016	.545	.022	.898	.015
area23	.561	.024	.876	.017	.561	.022	.876	.017
area24	.591	.035	.805	.032	.588	.029	.804	.031

where $\hat{\mathbf{t}}$ is an estimator of $\mathbf{z}$. We take

$$\operatorname{var}(u(z_i, \hat{\Omega}) \mid \mathcal{D}) \approx \sum_{t=0}^{n_i - r_i} \hat{\omega}_{it}(u(t, \hat{\Omega}) - \bar{\hat{\mu}}_i)^2.$$

It is now left to approximate $\operatorname{var}\{u(\hat{t}_i, \Omega) \mid \mathcal{D}, \mathbf{z} = \hat{\mathbf{t}})\}$. We defer this issue to Appendix 2.

4.2. Numerical Comparisons

In Table 2 and 3 we compare the posterior means and standard deviations obtained from our methods and the Stasny method. The estimates we employ

TABLE 3. *Comparison of the posterior means and standard deviations for nonignorable model based on the national family income and expenditure survey data*

Area	Stasny Estimates p_i	Stasny Estimates π_{i1}	Stasny Estimates π_{i2}	New Estimates p_i	New Estimates π_{i1}	New Estimates π_{i2}
area1	.531 (.009)	.715 (.012)	.779 (.011)	.539 (.010)	.746 (.013)	.781 (.011)
area2	.539 (.010)	.767 (.013)	.818 (.011)	.543 (.010)	.787 (.013)	.820 (.011)
area3	.541 (.012)	.824 (.014)	.855 (.012)	.542 (.012)	.834 (.014)	.858 (.012)
area4	.554 (.011)	.735 (.015)	.827 (.011)	.563 (.011)	.767 (.015)	.821 (.012)
area5	.540 (.012)	.904 (.011)	.913 (.010)	.539 (.012)	.904 (.011)	.917 (.010)
area6	.555 (.012)	.731 (.017)	.820 (.013)	.564 (.012)	.762 (.017)	.816 (.013)
area7	.561 (.012)	.753 (.016)	.840 (.012)	.569 (.011)	.781 (.015)	.835 (.012)
area8	.571 (.016)	.739 (.024)	.846 (.017)	.581 (.016)	.772 (.023)	.838 (.017)
area9	.553 (.016)	.924 (.015)	.936 (.012)	.553 (.016)	.925 (.015)	.938 (.012)
area10	.536 (.020)	.911 (.020)	.911 (.019)	.535 (.020)	.908 (.021)	.916 (.018)
area11	.572 (.016)	.761 (.024)	.854 (.017)	.578 (.016)	.788 (.023)	.848 (.017)
area12	.554 (.021)	.817 (.032)	.862 (.025)	.555 (.021)	.828 (.031)	.862 (.025)
area13	.560 (.017)	.833 (.023)	.878 (.017)	.562 (.017)	.844 (.022)	.878 (.017)
area14	.541 (.021)	.827 (.029)	.851 (.026)	.542 (.020)	.834 (.028)	.855 (.025)
area15	.572 (.016)	.829 (.021)	.887 (.015)	.574 (.016)	.842 (.020)	.885 (.015)
area16	.549 (.023)	.811 (.039)	.848 (.032)	.550 (.023)	.819 (.038)	.850 (.032)
area17	.560 (.017)	.918 (.017)	.933 (.014)	.560 (.017)	.918 (.017)	.935 (.013)
area18	.544 (.022)	.928 (.023)	.930 (.020)	.543 (.022)	.925 (.023)	.934 (.020)
area19	.572 (.017)	.857 (.022)	.901 (.016)	.573 (.017)	.865 (.021)	.901 (.015)
area20	.554 (.020)	.887 (.024)	.908 (.019)	.554 (.020)	.889 (.023)	.910 (.019)
area21	.558 (.014)	.855 (.016)	.890 (.012)	.559 (.014)	.863 (.015)	.891 (.012)
area22	.539 (.020)	.892 (.023)	.900 (.020)	.539 (.020)	.892 (.022)	.903 (.020)
area23	.550 (.020)	.860 (.026)	.886 (.021)	.550 (.020)	.865 (.025)	.888 (.021)
area24	.563 (.024)	.757 (.049)	.841 (.035)	.565 (.024)	.774 (.047)	.839 (.035)

NOTE : *Standard deviation estimates are shown in parentheses.*

in the approximation are the posterior means from the MCMC method. For ignorable model the estimates are $\hat{\alpha}_0 = 129.6$, $\hat{\beta}_0 = 103.6$, $\hat{\alpha}_1 = 14.6$, $\hat{\beta}_1 = 2.7$. For nonignorable model $\hat{\alpha}_0 = 144.5$, $\hat{\beta}_0 = 120.9$, $\hat{\alpha}_1 = 19.2$, $\hat{\beta}_1 = 4.2$, $\hat{\alpha}_2 = 24.6$, $\hat{\beta}_2 = 3.9$, and the posterior means of $z_1, \ldots, z_{24}$ are 346, 236, 119, 209, 65, 161, 168, 74, 21, 14, 65, 22, 39, 24, 46, 14, 17, 6, 32, 16, 66, 17, 21, 13.

First, consider Table 2 in which we present the posterior means and standard deviations of the p_i and π_i for the ignorable model. Corresponding estimates from the two methods are essentially the same for all areas. However, the standard deviations of our estimates are the same or slightly smaller than those from the Stasny method.

Second, consider Table 3 in which we present the posterior means and standard deviations of the p_i, π_{i1} and π_{i2} for the nonignorable model. Here, $\hat{p_i}'^s$, the estimated proportions of doctor's visit, are the same eventhough areas 1, 4, 6, and 8 have some differences between two methods, where the differences are within the standard deviations. However the standard deviations in all areas are the same for both methods. For the response rate $\hat{\pi_{i1}}$, the new estimates are higher than those from the Stasny method, and the standard deviations of the new estimates are the same or slightly smaller than those from the Stasny method. Finally, for the response rate $\hat{\pi_{i2}}$, both estimates are almost the same up to two decimal places and the standard deviations are also the same for all areas.

For the comparison of the corresponding $\hat{p_i}$'s between the ignorable and nonignorable models, the nonignorable model estimates are lower than those from the ignorable model.

As a conclusion, the results of our points estimates for p_i are the same as the Stasny method, yet the standard deviations are the same or slightly smaller than the Stasny.

5. Concluding Remarks

We have studied the nonignorable nonresponse for the estimation of a proportion. We extended the method of Stasny (1991) to take care of the variablilty in the hyper-parameters by using a Bayesian method facilitated by the MH sampler. We also explored the Stasny approach in a second direction ; approximate analytical expressions for the variance of the proportions are obtained by incorporating the variation of the hyperparameters. This works well for both ignorable and nonignorable nonresponse models as we illustrated for the NFIES data.

Appendix 1

Metropolis-Hastings Samplers for Ignorable and Nonignorable models

For ignorable model (μ_{11}, τ_{11}) and (μ_{12}, τ_{12}) are independent aposteriori with

$$p(\mu_{11}, \tau_{11} \mid \mathbf{y}, \mathbf{r}) \propto p(\mu_{11}, \tau_{11}) \prod_{i=1}^{\ell} \left\{ \frac{B(y_i + \mu_{11}\tau_{11},\ r_i - y_i + (1-\mu_{11})\tau_{11})}{B(\mu_{11}\tau_{11}, (1-\mu_{11})\tau_{11})} \right\} \tag{A.1}$$

and

$$p(\mu_{12}, \tau_{12} \mid \mathbf{y}, \mathbf{r}) \propto p(\mu_{12}, \tau_{12}) \prod_{i=1}^{\ell} \left\{ \frac{B(r_i + \mu_{12}\tau_{12}, n_i - r_i + (1-\mu_{12})\tau_{12})}{B(\mu_{12}\tau_{12}, (1-\mu_{12})\tau_{12})} \right\} \tag{A.2}$$

where $p(\mu_{11}, \tau_{11})$ and $p(\mu_{12}, \tau_{12})$ are the prior distributions. Samples can be obtained from each of (A.1) and (A.2) by using the algorithm of Nandram (1998).

For nonignorable model it is convenient to condition on $\mathbf{z}$ to obtain

$$\begin{aligned} & p(\mu_{21}, \tau_{21} \mid \mathbf{z}, \mathbf{y}, \mathbf{r}) \\ \propto\ & p(\mu_{21}, \tau_{21}) \prod_{i=1}^{\ell} \left\{ \frac{B(z_i + y_i + \mu_{21}\tau_{21},\ n_i - z_i - y_i + (1-\mu_{21})\tau_{21})}{B(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21})} \right\} \end{aligned} \tag{A.3}$$

$$\begin{aligned} & p(\mu_{22}, \tau_{22} \mid \mathbf{z}, \mathbf{y}, \mathbf{r}) \\ \propto\ & p(\mu_{22}, \tau_{22}) \prod_{i=1}^{\ell} \left\{ \frac{B(r_i - y_i + \mu_{22}\tau_{22}, n_i - r_i - z_i + (1-\mu_{22})\tau_{22})}{B(\mu_{22}\tau_{22}, (1-\mu_{22})\tau_{22})} \right\} \end{aligned} \tag{A.4}$$

$$p(\mu_{23}, \tau_{23} \mid \mathbf{z}, \mathbf{y}, \mathbf{r}) \propto p(\mu_{23}, \tau_{23}) \prod_{i=1}^{\ell} \left\{ \frac{B(y_i + \mu_{23}\tau_{23}, z_i + (1-\mu_{23})\tau_{23})}{B(\mu_{23}\tau_{23}, (1-\mu_{23})\tau_{23})} \right\} \tag{A.5}$$

where $p(\mu_{21}, \tau_{21})$, $p(\mu_{21}, \tau_{21})$ and $p(\mu_{21}, \tau_{21})$ are the prior distributions. Note that conditional on $\mathbf{z}$, the tuplets (μ_{2r}, τ_{2r}) are independent over r.

Also, $z_i \mid \mathbf{y}, \mathbf{r}, \mu_{2s}, \tau_{2s}, s = 1,2,3$ are independent with

$$p(z_i = t \mid \mathbf{y}, \mathbf{r}, \mu_{2s}, \tau_{2s}, s = 1,2,3) = \omega_{it} / \sum_{s=0}^{n_i - r_i} \omega_{is},\ t = 0, 1, \ldots, n_i - r_i \tag{A.6}$$

$$\omega_{it} = \frac{A_{it}^{(1)} A_{it}^{(2)} A_{it}^{(3)}}{B(t+1,\ n_i - r_i - t + 1)},$$

$$A_{it}^{(1)} = B(y_i + t + \mu_{21}\tau_{21},\ n_i - t - y_i + (1-\mu_{21})\tau_{21})$$

$$A_{it}^{(2)} = B(y_i + t + \mu_{21}\tau_{21},\ n_i - t - y_i + (1-\mu_{21})\tau_{21})$$
$$A_{it}^{(3)} = B(y_i + \mu_{23}\tau_{23},\ t + (1-\mu_{23})\tau_{23}).$$

We run the Metropolis-Hastings sampler by drawing a random deviate from each of (A.3), (A.4), (A.5) and (A.6). It is easy to draw a random deviate from (A.6). Samples can be obtained from each of (A.3), (A.4) and (A.5) by using the algorithm of Nandram (1998).

Appendix 2

An approximate form for var$\{u(\hat{t}_i, \Omega) \mid \mathcal{D}, \mathbf{z} = \hat{\mathbf{t}}\}$

We use a multivariate Taylor's series expansion.

Let $\phi = (\phi_0, \phi_1)'$, we obtain an approximation to the posterior var$(\phi \mid \mathcal{D})$. We use the posterior density of ϕ given $\mathbf{z} = \hat{\mathbf{t}}$

$$p(\phi \mid \mathcal{D}, \mathbf{z} = \hat{\mathbf{t}}) \propto \prod_{i=1}^{\ell} \frac{B(c_{i\hat{t}_i} + \phi_0, d_{i\hat{t}_i} - c_{i\hat{t}_i} + \phi_1)}{B(\phi_0, \phi_1)}$$

and denote its logarithm by $\Delta(\phi)$ where

$$\begin{aligned}\Delta(\phi) &= \sum_{i=1}^{\ell}\{[\log(\Gamma(c_{i\hat{t}_i} + \phi_0)) + \log\Gamma(d_{i\hat{t}_i} - c_{i\hat{t}_i} + \phi_1) - \log\Gamma(d_{i\hat{t}_i} + \phi_0 + \phi_1)] \\ &\quad - [\log\Gamma(\phi_0) + \log\Gamma(\phi_1) - \log\Gamma(\phi_0 + \phi_1)]\}.\end{aligned}$$

We obtain an approximation for var$(\phi \mid \mathcal{D})$ by using the negative inverse Hessian matrix of $\Delta(\phi)$ at $\hat{\phi}$.

First, we let

$$\begin{aligned}g_{1it} &= (d_{it} - c_{it} + \hat{\phi}_1)/(d_{it} + \hat{\phi}_0 + \hat{\phi}_1)^2 \\ g_{2it} &= -(c_{it} + \hat{\phi}_0)/(d_{it} + \hat{\phi}_0 + \hat{\phi}_1)^2 \\ h_{1t} &= \sum_{i=1}^{\ell}\{\psi'(\hat{\phi}_0) - \psi'(c_{it} + \hat{\phi}_0)\} + h_t \\ h_{2t} &= \sum_{i=1}^{\ell}\{\psi'(\hat{\phi}_1) - \psi'(d_{it} - c_{it} + \hat{\phi}_1)\} + h_t \\ h_t &= \sum_{i=1}^{\ell}\{\psi'(d_{it} + \hat{\phi}_0 + \hat{\phi}_1) - \psi'(\hat{\phi}_0 + \hat{\phi}_1)\},\end{aligned}$$

where $\psi'(\cdot)$ is the trigamma function. When $h_{1t}h_{2t} \le h_t^2$, we take $h_t = 0$. Because $\psi'(\cdot)$ is a strictly decreasing function, $h_t = 0$ if and only if $\psi'(d_{it} + \hat{\phi}_0 + \hat{\phi}_1) = \psi'(\hat{\phi}_0 + \hat{\phi}_1)$. While this is false because $\psi'(\cdot)$ decreases very slowly for values of its arguments, especially the assumption $h_t = 0$ is reasonable. Also, letting $k_{\mathrm{t}} = (1 - h_{\mathrm{t}}^2/h_{1\mathrm{t}}h_{2\mathrm{t}})^{-1}$ we define

$$
\begin{aligned}
\nu_{1\mathrm{t}} &= k_{\mathrm{t}}/h_{1\mathrm{t}} \\
\nu_{2\mathrm{t}} &= k_{\mathrm{t}}/h_{2\mathrm{t}} \\
\nu_{\mathrm{t}} &= -k_{\mathrm{t}}, h_{\mathrm{t}}/h_{1\mathrm{t}}h_{2\mathrm{t}}.
\end{aligned}
$$

Then

$$\mathrm{var}\{u(\hat{t}_i,\Omega) \mid \mathcal{D}, \mathbf{z} = \hat{\mathrm{t}}\} = g_{1i\hat{t}_i}^2 \nu_{1\hat{\mathrm{t}}} + 2g_{1i\hat{t}_i}g_{2i\hat{t}_i}\nu_{\hat{\mathrm{t}}} + g_{2i\hat{t}_i}^2 \nu_{2\hat{\mathrm{t}}}.$$

References

Albert, J. H. and Gupta, A. K. (1985). "Bayesian methods for binomial data with applications to a nonresponse problem", *Journal of the American statistical Association*, **80**, 167-174.

Baker, S. G. and Laird, N. M. (1988)."Regression analysis for categorical variables with outcome subject to nonignorable nonresponse", *Journal of the American Statistical Association*, **83**, 62-69.

Crawford, S. L., Johnson, W. G. and Laird, N. M. (1993). "Bayes analysis of model-based methods for nonignorable nonresponse in the Harvard Medical Practice Survey (with discussions)", In *Case Studies in Bayesian Statistics* (Gatsonis C., Hodges, J. S., Kass, R. E. and Sinpurwalla, N. D., eds.), 78-117 Springer-Verlag, New York.

De Heer, W. (1999). "International response trends: results of an international survey", *Journal of Official Statistics*, **15**, 129-142.

Deely, J. J. and Smith, W. F. (1998). "Bayes empirical Bayes", *Journal of the Americal Statistical Association*, **76**, 833-841.

Groves, R. M. and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*, Wiley, New York.

Heckman, J. (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models", *Annals of Economic and Social Measurement*, **5**, 475-492.

Kaufman, G. M. and King, B. (1973). "A Bayesian analysis of nonresponse in dichotomous processes", *Journal of the American Statistical association*, **68**, 670-678.

Little R. J. A. and Rubin D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

Nandram, B. (1998). "A Bayesian analysis of the three-stage hierarchical multinomial model", *Journal of Statistical Computation and Simulation*, **61**, 97-126.

National Statistical Office of Korea (1996). *National Survey of Family Income and Expenditure, Yearly Income, Savings & Liabilities and Durable Goods*, Volumn 2, National Statistical Office of Korea.

Olson, R. L. (1980). "A least squares correction for selectivity bias", *Econometrica*, **48**, 1815-1820.

Park, T. and Brown, M. B. (1994). "Models for categorical data with nonignorable nonresponse", *Journal of the American Statistical Association*, **89**, 44-52.

Rubin D. B. (1976). "Inference and missing data", *Biometrika*, **63**, 581-590.

Rubin D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.

Stasny, E. A. (1991). "Hierarchical models for the probabilities of a survey classification and nonresponse: an example from the National Crime Survey", *Journal of the American Statistical Association*, **86**, 296-303.

Stasny, E. A., Kadane, J. B. and Fritsch, K. S. (1998). "On the fairness of death penalty jurors: a comparison of Bayesian models with different levels of hierarchy and various missing-data mechanisms", *Journal of the American Statistical Association*, **93**, 464-477.