Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Bayesian methods for an incomplete two-way contingency table with application to the Ohio (Buckeye State) Polls

by Bo-Seung Choi, Jai Won Choi and Yousung Park

June 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2009
Vol. 35, No. 1, pp. 37-51
Statistics Canada, Catalogue No. 12-001-X

Bayesian methods for an incomplete two-way contingency table with application to the Ohio (Buckeye State) Polls

Bo-Seung Choi, Jai Won Choi and Yousung Park [1]

Abstract

We use a Bayesian method to resolve the boundary solution problem of the maximum likelihood (ML) estimate in an incomplete two-way contingency table, using a loglinear model and Dirichlet priors. We compare five Dirichlet priors in estimating multinomial cell probabilities under nonignorable nonresponse. Three priors among them have been used for an incomplete one-way table, while the remaining two new priors are newly proposed to reflect the difference in the response patterns between respondents and the undecided. The Bayesian estimates with the previous three priors do not always perform better than ML estimates unlike previous studies, whereas the two new priors perform better than both the previous three priors and the ML estimates whenever a boundary solution occurs. We use four sets of data from the 1998 Ohio state polls to illustrate how to use and interpret estimation results for the elections. We use simulation studies to compare performance of the five Bayesian estimates under nonignorable nonresponse.

Key Words: Bayesian analysis; Nonignorable nonresponse; Contingency table; Boundary solution; EM algorithm.

1. Introduction

The problem of nonresponse is common in most surveys becoming a serious issue as the nonresponse rate increases (De Heer 1999; Groves and Couper 1998). When survey data is summarized in a two-way contingency table, the table includes fully classified counts, partially classified counts (*i.e.*, item nonresponse), and unclassified counts (*i.e.*, unit nonresponse). For example, in the Ohio (Buckeye State) Poll (BSP) (Chen and Stasny 2003), one category involves the voting preference (candidates A,B,C, or undecided) and the other category is the likelihood of voting (likely to vote, not likely to vote, and undecided). First supplemental margin contains data only on the voting preference, second contains data only on the likelihood of voting, and third is only the number of unit nonresponses (both responses unknown). Our interest is to incorporate these missing observations into estimating the true support for each candidate and to present Bayesian models to predict the winner.

In some surveys, the undecided answers are treated as a valid response category when the respondents do not have strong preference for a candidate and voting intention (Smith 1984; Rubin, Stern and Vehovar 1995). Many studies, however, have shown that the voting behavior of the undecided voters can have a significant impact on the final result and that by considering these undecided voters, the accuracy of election forecasting can be improved (Perry 1979; Fenwick, Wiseman, Becker and Heiman 1982; Myers and O'Connor 1983; Kim 1995; Chen and Stasny 2003; Martin, Traugott and Kennedy 2005). Perry (1979), among them, showed that the undecided percentage in a poll is likely to be greater than the true percentage by presenting an empirical evidence using a secret ballot approach. Kim (1995) also indicated that these undecided voters are critical, especially in cases where the number of undecided voters is greater than the gap between the two leading runners in an election race. Three of our empirical studies in Section 3 belong to this critical case. Fenwick *et al.* (1982) and Kim (1995) applied a discriminant analysis to the October 1980 poll data in Massachusetts and the 1992 USA presidential election, from which they allocated the undecided voters to candidates to show that undecided voters generally do not vote in the same proportions as their decided counterparts. When the focus is on the candidate the undecided voter may vote for, undecided responses are better treated as missing data (Myers and O'Connor 1983). As indicated in Flannelly, Flannelly and McLeod (2000) and Lau (1994), the forecasting error for the actual election results increases as the rate of undecided voters increases. To overcome this problem, Monterola, Lim, Garcia and Saloma (2001) applied a neural network approach to classify undecided voters in a public opinion survey. Smith, Skinner and Clarke (1999) and Molenberghs, Kenward and Goetghebeur (2001) utilized model based imputation methods for the 1992 British General Election Panel Survey and the 1991 Slovenian plebiscite public opinion survey. Because our main goal is to obtain more accurate forecasts by allocating undecided voters to proper cell, we treat undecided voters as missing observations in the same way as these researchers handled them.

1. Bo-Seung Choi, Research Professor, Institute of Economics, Korea University, Seoul 136-701, Korea; Jai Won Choi, Professor, Department of Biostatistics, Medical College of Georgia, Augusta, GA 30912; Yousung Park, Professor, Department of Statistics, Korea University, Seoul 136-701, Korea. E-mail: yspark@korea.ac.kr.

Nonresponse (or undecided, equivalently) can be distinguished by three types of nonresponses (Little and Rubin 2002, page 11): missing completely at random (MCAR) means that the probability of a nonresponse on a variable of interest is independent of all survey variables including itself; missing at random (MAR) means that the probability of a nonresponse depends only on the observed data; missing not at random (MNAR) means that the probability of nonresponse depends on the unobserved values. Models for MCAR or MAR are called ignorable nonresponse models while models for MNAR are called nonignorable. For example, in a pre-election survey, if the respondents do not answer with their preference of a candidate, although they support a particular candidate, the pattern for candidate preference can be different between the respondents and nonrespondents. Then, the nonresponse mechanism is nonignorable. When data is assumed to be MCAR, the effect of nonresponse can be removed in likelihood inference (Little and Rubin 2002, page 11). However, when the nonrespondents follow a response pattern different from that of the respondents, discarding nonresponses or misspecifying the nonresponse mechanism leads to larger variances and biases in estimation (Chen 1972; Park and Brown 1994).

When nonresponse is nonignorable in contingency tables, ML estimation often yields boundary solutions where the probability of nonresponse is estimated to be zero in some cells. These boundary solutions often provide a local maximum of the likelihood function. In this case, the maximum likelihood (ML) estimates of the loglinear model parameters cannot have a unique solution and usually have large standard deviations (see Section 4 or Baker, Rosenberger and Dersimonian (1992) and Park and Brown (1994) for more detailed discussions).

The conditions where the ML estimate falls on the boundary solution have been proposed in a one-way contingency table (Baker and Laird 1988; Michiels and Molenbergs 1997). The geometric explanation for the boundary solution of the ML estimate was presented (Smith *et al.* 1999; Clark 2002). Baker *et al.* (1992) presented a sufficient and necessary condition under which the ML estimate can have a boundary solution in a two-way contingency table.

To overcome such a boundary problem in the ML estimate under the existence of nonignorable nonresponses, Park and Brown (1994) and Park (1998) proposed Bayesian approach using empirical priors based only on respondent information. Clogg, Rubin, Schenker and Schultz (1991) used constant prior for an incomplete one-way contingency table. Although they showed that, under nonignorable nonresponse, Bayesian methods provided smaller mean squared errors (MSE) than ML estimate in estimating cell expectations, our simulation study shows that this is generally not true in an incomplete two-way contingency table. Thus, we present two Bayesian models whose priors depend on information from both respondents and undecided. We, then, apply each to analyze incomplete two-way contingency table. An extension to a multi-way table is straightforward. We can easily apply this extension to weighted data from stratified or cluster sampling using appropriate covariates (see Section 2.2).

The remainder of this paper is divided into four sections. In Section 2, we consider Bayesian models with five different priors and present a generalized expectation maximization (EM) algorithm to estimate cell probabilities. In Section 3, we apply the Bayesian models to four empirical data sets from the Buckeye State Poll (BSP) and compare the Bayesian estimates with the ML estimate and the actual election results. In Section 4, we use simulation studies to compare MSEs and biases of the Bayesian estimates from different missing percentages and response patterns of the respondents and nonrespondents. In this section, we also calculate the coverage probability to examine the performance of the Bayesian estimates. Section 5 includes some concluding remarks.

2. Bayesian models

We discuss five Bayesian estimates to accommodate nonignorable nonresponse in an incomplete two-way contingency table. We present an EM algorithm to tackle the nonresponse problem in a two-way contingency table in Section 2.1. Then, in Section 2.2, we specify five priors and extend our approach to a multi-way contingency table.

Let X_1 and X_2 be response variables indexed by I and J categories, respectively, in a two-way contingency table. We also let $R_1 = 1$ when X_1 is observed and $R_1 = 2$ when X_1 is missing. Similarly, $R_2 = 1$ when X_2 is observed and $R_2 = 2$ when X_2 is missing. Then the full array of X_1, X_2, R_1, and R_2 constructs a $I \times J \times 2 \times 2$ contingency table in which we have completely classified counts, partially classified counts, and unclassified counts. To distinguish these three types of observations, let y_{ijkl} be the count belonging to the i^{th} category of X_1, the j^{th} category of X_2, the k^{th} value of R_1, and the l^{th} value of R_2. Thus, y_{ij11} is used for the completely classified counts, y_{i+12} and y_{+j21} for the respective column and row supplemental margins, and y_{++22} for the unclassified counts. We assume a multinomial distribution for these three types of observations to have the following log likelihood:

$$l = \sum_i \sum_j y_{ij11} \cdot \log(\pi_{ij11}) + \sum_i y_{i+12} \cdot \log(\pi_{i+12}) + \sum_j y_{+j21} \cdot \log(\pi_{+j21}) + y_{++22} \cdot \log(\pi_{++22}) \qquad (1)$$

where $\pi_{ijkl} = \Pr[X_1 = i, X_2 = j, R_1 = k, R_2 = l]$ and $N = \sum_{i,j,k,l} y_{ijkl}$ is fixed.

Since this likelihood function involves more parameters than degrees of freedom available for estimation, we link π_{ijkl} to relevant covariates using a loglinear function. Since no explanatory variable is available, we do not use any explanatory variables. However, the loglinear model can easily incorporate explanatory variables in the same way as it incorporates the categorical variables (see Baker and Laird 1988 and Park and Brown 1994 for details).

A nonignorable nonresponse model for all of the variables $X_1, X_2, R_1,$ and R_2 is defined by

$$\log(m_{ijkl}) = \beta_0 + \beta^i_{X_1} + \beta^j_{X_2} + \beta^k_{R_1} + \beta^l_{R_2} + \beta^{ik}_{X_1R_1} + \beta^{jl}_{X_2R_2} + \beta^{ij}_{X_1X_2} + \beta^{kl}_{R_1R_2}$$

$$\text{for } i = 1,\ldots,I,\ j = 1,\ldots,J,\ k = 1,2, \text{ and } l = 1,2 \quad (2)$$

where $m_{ijkl} = N \cdot \pi_{ijkl}$ is the expected cell count for the $(i, j, k, l)^{\text{th}}$ category and the sum of each β-term over any of its respective super script(s) is zero.

This loglinear model is saturated since the number of parameters is exactly the same as the number of cells observed from the incomplete two-way contingency table. This model is also a nonignorable nonresponse model because of the interaction terms between X_1 and R_1 and between X_2 and R_2, implying that the nonresponse of each response variable depends on its own status. The loglinear model is a tool frequently used for analyzing incomplete contingency tables with nonignorable non-responses. Let p be the number of parameters (*i.e.*, β) to be estimated. We introduce the $p \times 1$ design vector $\mathbf{z}_{ijkl}$ to indicate the affiliation of the observation belonging to the $(i, j, k, l)^{\text{th}}$ category. Then the loglinear model given in (2) can be rewritten as

$$\log \mathbf{m} = Z\boldsymbol{\beta} \quad (3)$$

where the $I \times J \times 2 \times 2$ vector $\mathbf{m}$ is the cell expectation and $\boldsymbol{\beta}$ is the vector representation of the β s. To avoid a boundary solution of the ML estimate in model (2), we impose Dirichlet priors to the cell probabilities $(\pi_{ij11}, \pi_{ij12}, \pi_{ij21}, \pi_{ij22})$ as given by

$$\prod_i \prod_j \pi_{ij11}^{\delta_{ij11}} \cdot \pi_{ij12}^{\delta_{ij12}} \cdot \pi_{ij21}^{\delta_{ij21}} \cdot \pi_{ij22}^{\delta_{ij22}} \quad (4)$$

where the hyper parameters, the δ_{ijkl} s are specified in Section 2.2. These Dirichlet priors produce an explicit and convenient form of a posterior distribution because they are conjugated to a multinomial distribution (Clogg *et al.* 1991; Park and Brown 1994; Forster and Smith 1998). Together with (3), the multinomial distribution of (1) for observations, and the prior distribution (4), we have the following log posterior distribution:

$$\begin{aligned} l_{pos} = & \sum_i \sum_j y_{ij11} \cdot (\mathbf{z}_{ij11} \cdot \boldsymbol{\beta}) \\ & - \sum_i \sum_j y_{ij11} \cdot \log\Big(\sum_{i,j,k,l} \exp(\mathbf{z}_{ijkl} \cdot \boldsymbol{\beta})\Big) \\ & + \sum_i y_{i+12} \cdot \log\Big(\sum_j \exp(\mathbf{z}_{ij12} \cdot \boldsymbol{\beta})\Big) \\ & - \sum_i y_{i+12} \cdot \log\Big(\sum_{i,j,k,l} \exp(\mathbf{z}_{ijkl} \cdot \boldsymbol{\beta})\Big) \\ & + \sum_j y_{+j21} \cdot \log\Big(\sum_i \exp(\mathbf{z}_{ij21} \cdot \boldsymbol{\beta})\Big) \\ & - \sum_j y_{+j21} \cdot \log\Big(\sum_{i,j,k,l} \exp(\mathbf{z}_{ijkl} \cdot \boldsymbol{\beta})\Big) \\ & + y_{++22} \cdot \log\Big(\sum_i \sum_j \exp(\mathbf{z}_{ij22} \cdot \boldsymbol{\beta})\Big) \\ & - y_{++22} \cdot \log\Big(\sum_{i,j,k,l} \exp(\mathbf{z}_{ijkl} \cdot \boldsymbol{\beta})\Big) \\ & + \sum_{i,j,k,l} \delta_{ijkl} \cdot (\mathbf{z}_{ijkl} \cdot \boldsymbol{\beta}) \\ & - \sum_{i,j,k,l} \delta_{ijkl} \cdot \log\Big(\sum_{i,j,k,l} \exp(\mathbf{z}_{ijkl} \cdot \boldsymbol{\beta})\Big). \end{aligned} \quad (5)$$

Equation (5) is rather complex and thus we use the EM algorithm to estimate the parameters (*i.e.*, $\boldsymbol{\beta}$).

2.1 The EM algorithm

We maximize the posterior distribution given in (5) over the parameter $\boldsymbol{\beta}$ using the generalized expectation maximization (GEM) algorithm (Dempster, Laird and Rubin 1977) with the following E and M steps.

E-step: Using augmented $y_{ij12}, y_{ij21},$ and y_{ij22} for $i = 1, \ldots, I$ and $j = 1, \ldots, J,$ the posterior (5) can be written as

$$\begin{aligned} l_{a.\text{pos}} = & \sum_i \sum_j (y_{ij11} + \delta_{ij11}) \log(\pi_{ij11}) \\ & + \sum_i \sum_j (y_{ij12} + \delta_{ij12}) \log(\pi_{ij12}) \\ & + \sum_i \sum_j (y_{ij21} + \delta_{ij21}) \log(\pi_{ij21}) \\ & + \sum_i \sum_j (y_{ij22} + \delta_{ij22}) \log(\pi_{ij22}). \end{aligned} \quad (6)$$

To determine the expected augmented log posterior in (6), we average over the missing counts $y_{ij12}, y_{ij21},$ and y_{ij22} conditioning on the current parameter estimates, π^{old}_{ijkl}, and the marginal sums $y_{i+12}, y_{+j21},$ and y_{++22}:

$$E_{\text{old}}[l_{a.\text{pos}}] = \sum_i \sum_j (y_{ij11} + \delta_{ij11}) \cdot \log(\pi_{ij11})$$

$$+ \sum_i \sum_j (E_{\text{old}}[y_{ij12} \mid \pi_{ijkl}^{\text{old}}, y_{i+12}] + \delta_{ij12}) \cdot \log(\pi_{ij12})$$

$$+ \sum_i \sum_j (E_{\text{old}}[y_{ij21} \mid \pi_{ijkl}^{\text{old}}, y_{+j21}] + \delta_{ij21}) \cdot \log(\pi_{ij21})$$

$$+ \sum_i \sum_j (E_{\text{old}}[y_{ij22} \mid \pi_{ijkl}^{\text{old}}, y_{++22}] + \delta_{ij22}) \cdot \log(\pi_{ij22}). \quad (7)$$

Since y_{ij12}, y_{ij21}, and y_{ij22} are multinomial random variates conditioned on the respective marginal sum y_{i+12}, y_{+j21}, and y_{++22}, the conditional expectations in the equation (7) are given by

$$E_{\text{old}}(y_{ij12} \mid \pi_{ijkl}^{\text{old}}, y_{i+12}) = y_{i+12} \frac{m_{ij12}^{\text{old}}}{m_{i+12}^{\text{old}}},$$

$$E_{\text{old}}(y_{ij21} \mid \pi_{ijkl}^{\text{old}}, y_{+j21}) = y_{+j21} \frac{m_{ij21}^{\text{old}}}{m_{+j21}^{\text{old}}},$$

and

$$E_{\text{old}}(y_{ij22} \mid \pi_{ijkl}^{\text{old}}, y_{++22}) = y_{++22} \frac{m_{ij22}^{\text{old}}}{m_{++22}^{\text{old}}}$$

where $m_{ijkl}^{\text{old}} = N \cdot \pi_{ijkl}^{\text{old}}$.

M-step: In this step, we maximize the expected log posterior (7) using the pseudo observations $\tilde{y}_{ij11} = y_{ij11} + \delta_{ij11}, \tilde{y}_{ij12} = y_{i+12}\, m_{ij12}^{\text{old}} / m_{i+12}^{\text{old}} + \delta_{ij12}$, $\tilde{y}_{ij21} = y_{+j21}\, m_{ij21}^{\text{old}} / m_{+j21}^{\text{old}} + \delta_{ij21}$, and $\tilde{y}_{ij22} = y_{++22}\, m_{ij22}^{\text{old}} / m_{++22}^{\text{old}} + \delta_{ij22}$. We impose the constraints on these pseudo observations so that their marginal sums are the same as the corresponding marginal sums of observations: $\tilde{y}_{++11} = y_{++11}$, $\tilde{y}_{i+12} = y_{i+12}$, $\tilde{y}_{+j21} = y_{+j21}$, and $\tilde{y}_{++22} = y_{++22}$. Under these constraints, the pseudo observations are now

$$y_{ijkl}^{*} = \begin{cases} \tilde{y}_{ij11} \dfrac{y_{++11}}{y_{++11} + \delta_{++11}} & \text{for k = 1 and l = 1} \\ \tilde{y}_{ij12} \dfrac{y_{i+12}}{y_{i+12} + \delta_{i+12}} & \text{for k = 1 and l = 2} \\ \tilde{y}_{ij21} \dfrac{y_{+j21}}{y_{+j21} + \delta_{+j21}} & \text{for k = 2 and l = 1} \\ \tilde{y}_{ij22} \dfrac{y_{++22}}{y_{++22} + \delta_{++22}} & \text{for k = 2 and l = 2.} \end{cases}$$

Then, the expected log posterior function (7) becomes

$$E_{\text{old}}[l_{a.\text{pos}}] = \sum_i \sum_j y_{ij11}^{*} \cdot \log(\pi_{ij11})$$

$$+ \sum_i \sum_j y_{ij12}^{*} \cdot \log(\pi_{ij12})$$

$$+ \sum_i \sum_j y_{ij21}^{*} \cdot \log(\pi_{ij21})$$

$$+ \sum_i \sum_j y_{ij22}^{*} \cdot \log(\pi_{ij22}).$$

This equation has the same form as the likelihood obtained from a four-way contingency table with fully observed cell counts y_{ijkl}^{*} s. Thus, using the iterative re-weighted least squares method (Agresti 2002, page 342), we obtain the maximum posterior estimator (MPE) of $\boldsymbol{\beta}$ as follows:

$$\boldsymbol{\beta}^{(t+1)} = (Z^T \hat{V}_t^{-1} Z)^{-1} Z^T \hat{V}_t^{-1} \gamma^{(t)},$$

where $\gamma^{(t)}$ has element $\gamma_{ijkl}^{(t)} = \log m_{ijkl}^{(t)} + (y_{ijkl} - m_{ijkl}^{(t)}) / m_{ijkl}^{(t)}$ and $\hat{V}_t = [\text{diag}(\mathbf{m}^{(t)})]^{-1}$. We finally iterate these E and M-steps until a convergence criterion is achieved. The convergence criterion we use is $\varepsilon \le 10^{-6}$, where ε is the difference between two consecutive log posterior functions.

Let $Y_{\text{obs}} = (y_{ij11}, y_{i+12}, y_{+j21}, y_{++22})$ and $Y_{\text{mis}} = (y_{ij12}, y_{ij21}, y_{ij22})$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$ be the observed count vector and the missing count vector, respectively. Then the log posterior distribution (5) can be written as

$$l_{\text{pos}} = l(\boldsymbol{\beta} \mid Y_{\text{obs}}) = l(\boldsymbol{\beta} \mid Y_{\text{obs}}, Y_{\text{mis}}) - \log f(Y_{\text{mis}} \mid Y_{\text{obs}}, \boldsymbol{\beta}). \quad (8)$$

By taking differentiation twice with respect to $\boldsymbol{\beta}$, (8) yields

$$\frac{\partial^2 l(\boldsymbol{\beta} \mid Y_{\text{obs}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \frac{\partial^2 l(\boldsymbol{\beta} \mid Y_{\text{obs}}, Y_{\text{mis}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} - \frac{\partial^2 \log f(Y_{\text{mis}} \mid Y_{\text{obs}}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$$

$$= -\mathbf{Z}^T[\text{diag}(\mathbf{m}) - \mathbf{m}\mathbf{m}^T/N]\,\mathbf{Z} + \mathbf{Z}^T[\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T]\,AB\mathbf{Z}, \quad (9)$$

where $\boldsymbol{\pi}$ is vector expression of cell probabilities π_{ijkl} and A, B are given by

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathrm{diag}\left(\dfrac{y_{i+12}^2}{y_{i+12} + \delta_{i+12}} \dfrac{m_{ij12}}{m_{i+12}}\right) & 0 & 0 \\ 0 & 0 & \mathrm{diag}\left(\dfrac{y_{+j21}^2}{y_{+j21} + \delta_{+j21}} \dfrac{m_{ij21}}{m_{+j21}}\right) & 0 \\ 0 & 0 & 0 & \mathrm{diag}\left(\dfrac{y_{++22}^2}{y_{++22} + \delta_{++22}} \dfrac{m_{ij22}}{m_{++22}}\right) \end{pmatrix}$$

and

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & I_{IJ} - B^{12} & 0 & 0 \\ 0 & 0 & I_{IJ} - B^{21} & 0 \\ 0 & 0 & 0 & I_{IJ} - B^{22} \end{pmatrix}.$$

Here, to save the space and since there is no difficulty to extend for general i and j, B^{12}, B^{21}, and B^{22} are illustrated only for $I = 2$ and $J = 3$:

$$B^{12} = \begin{pmatrix} \dfrac{m_{1112}}{m_{1+12}} & \dfrac{m_{1212}}{m_{1+12}} & \dfrac{m_{1312}}{m_{1+12}} & 0 & 0 & 0 \\ \dfrac{m_{1112}}{m_{1+12}} & \dfrac{m_{1212}}{m_{1+12}} & \dfrac{m_{1312}}{m_{1+12}} & 0 & 0 & 0 \\ \dfrac{m_{1112}}{m_{1+12}} & \dfrac{m_{1212}}{m_{1+12}} & \dfrac{m_{1312}}{m_{1+12}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \dfrac{m_{2112}}{m_{2+12}} & \dfrac{m_{2112}}{m_{2+12}} & \dfrac{m_{2112}}{m_{2+12}} \\ 0 & 0 & 0 & \dfrac{m_{2212}}{m_{2+12}} & \dfrac{m_{2212}}{m_{2+12}} & \dfrac{m_{2212}}{m_{2+12}} \\ 0 & 0 & 0 & \dfrac{m_{2312}}{m_{2+12}} & \dfrac{m_{2312}}{m_{2+12}} & \dfrac{m_{2312}}{m_{2+12}} \end{pmatrix},$$

$$B^{21} = \begin{pmatrix} \dfrac{m_{1121}}{m_{+121}} & 0 & 0 & \dfrac{m_{2121}}{m_{+121}} & 0 & 0 \\ 0 & \dfrac{m_{1221}}{m_{+221}} & 0 & 0 & \dfrac{m_{2221}}{m_{+221}} & 0 \\ 0 & 0 & \dfrac{m_{1321}}{m_{+321}} & 0 & 0 & \dfrac{m_{2321}}{m_{+321}} \\ \dfrac{m_{1121}}{m_{+121}} & 0 & 0 & \dfrac{m_{2121}}{m_{+121}} & 0 & 0 \\ 0 & \dfrac{m_{1221}}{m_{+221}} & 0 & 0 & \dfrac{m_{2221}}{m_{+221}} & 0 \\ 0 & 0 & \dfrac{m_{1321}}{m_{+321}} & 0 & 0 & \dfrac{m_{2321}}{m_{+321}} \end{pmatrix},$$

and

$$B^{22} = \begin{pmatrix} \dfrac{m_{1122}}{m_{++22}} & \dfrac{m_{1222}}{m_{++22}} & \dfrac{m_{1322}}{m_{++22}} & \dfrac{m_{2122}}{m_{++22}} & \dfrac{m_{2222}}{m_{++22}} & \dfrac{m_{2322}}{m_{++22}} \\ \dfrac{m_{1122}}{m_{++22}} & \dfrac{m_{1222}}{m_{++22}} & \dfrac{m_{1322}}{m_{++22}} & \dfrac{m_{2122}}{m_{++22}} & \dfrac{m_{2222}}{m_{++22}} & \dfrac{m_{2322}}{m_{++22}} \\ \dfrac{m_{1122}}{m_{++22}} & \dfrac{m_{1222}}{m_{++22}} & \dfrac{m_{1322}}{m_{++22}} & \dfrac{m_{2122}}{m_{++22}} & \dfrac{m_{2222}}{m_{++22}} & \dfrac{m_{2322}}{m_{++22}} \\ \dfrac{m_{1122}}{m_{++22}} & \dfrac{m_{1222}}{m_{++22}} & \dfrac{m_{1322}}{m_{++22}} & \dfrac{m_{2122}}{m_{++22}} & \dfrac{m_{2222}}{m_{++22}} & \dfrac{m_{2322}}{m_{++22}} \\ \dfrac{m_{1122}}{m_{++22}} & \dfrac{m_{1222}}{m_{++22}} & \dfrac{m_{1322}}{m_{++22}} & \dfrac{m_{2122}}{m_{++22}} & \dfrac{m_{2222}}{m_{++22}} & \dfrac{m_{2322}}{m_{++22}} \\ \dfrac{m_{1122}}{m_{++22}} & \dfrac{m_{1222}}{m_{++22}} & \dfrac{m_{1322}}{m_{++22}} & \dfrac{m_{2122}}{m_{++22}} & \dfrac{m_{2222}}{m_{++22}} & \dfrac{m_{2322}}{m_{++22}} \end{pmatrix}.$$

We observe that the observed data information $\partial^2 l(\boldsymbol{\beta} \mid Y_{\text{obs}}) / \partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$ is equal to the augmented data information minus the missing data information. As shown in Gelman, Carlin, Stern and Rubin (2004, page 103), the inverse of the observed data information evaluated at the MPE of $\boldsymbol{\beta}$ is the variance of the MPE of $\boldsymbol{\beta}$.

2.2 Specification of priors

To complete the EM algorithm, we need to determine the hyper-parameters, δ_{ijkl} s. We set the sum of priors $\sum_{i,\,j,\,k,\,l} \delta_{ijkl}$ equal to the number of parameters involved in the loglinear model, p, as suggested by Clogg *et al*. (1991). Under this constraint, we propose five types of priors as follows. We first allocate δ_{ijkl} for the MPE of m_{ijkl} to shrink toward the MLE obtained under ignorable non-response. That is, we determine δ_{ijkl} depending only on the known response counts y_{ij11} and call them respondent-driven priors.

The first type of respondent-driven prior is, for all $i = 1, \ldots, I$ and $j = 1, \ldots, J$,

$$\delta_{ij11} = \nabla_{11} \frac{y_{ij11}}{y_{++11}}, \delta_{ij12} = \nabla_{12} \frac{y_{ij11}}{y_{++11}}, \delta_{ij21} = \nabla_{21} \frac{y_{ij11}}{y_{++11}}, \text{ and } \delta_{ij22} = \nabla_{22} \frac{y_{ij11}}{y_{++11}} \quad (10)$$

where $\nabla_{kl} = p \cdot y_{++kl} / y_{++++}$ for $k = 1, 2$ and $l = 1, 2$. The second type of respondent-driven prior gives no prior (*i.e*, no need of prior as described below) on π_{ij11} in the first type of priors. That is, the second type is the same as the first type except $\delta_{ij11} = 0$ for all i and j. In the case of a one-way contingency table (*i.e.*, either X_1 or X_2 is fully observed without missing information) and $y_{++22} = 0$, the first type is reduced to the priors used in Park (1998), whereas the second type is reduced to the priors used in Park and Brown (1994). These two types of respondent-driven priors may be too simplistic because the non-respondents are usually assumed to have a different response pattern from the respondents under a nonignorable nonresponse model. For example, the candidate preference of nonrespondents could be different from that of respondents in a pre-election survey.

In order to define the third type of prior, denote $\hat{m}_{ijkl}$ as the MLE for m_{ijkl}. The closed form of $\hat{m}_{ijkl}$ can be obtained from Baker *et al.* (1992) where some $\hat{m}_{ijkl}$ could be zero because of boundary solutions. For example, when a supplemental column margin has a boundary solution in an incomplete 2×2 table, the MLEs are $\hat{m}_{1j11} = y_{1j11}$,

$$\hat{m}_{2j11} = \frac{y_{2+11}(y_{2j11} + y_{+j21})}{y_{2+11} + y_{++21}},\ \hat{m}_{ij12} = \hat{m}_{ij11} b_j$$

where b_j is the solution of $\sum_{j=1}^{2} y_{ij11} b_j = y_{i+12}$, $\hat{m}_{1j21} = 0$,

$$\hat{m}_{2j21} = \hat{m}_{2j11} \frac{y_{++21}}{y_{2+11}},\ \hat{m}_{1j22} = 0,$$

and $\hat{m}_{2j22} = \hat{m}_{2j12}\, y_{++22} / y_{2+12}$. Therefore, these ML estimates accommodate both the information of respondents and nonrespondents, as well. The ML estimates can also be obtained from our EM algorithm in Section 2.1 by setting $\delta_{ijkl} = 0$ for all i, j, k and l. Using these ML estimates, we define the third type of prior as

$$\delta_{ij11} = \nabla_{11} \cdot \left(\frac{\hat{m}_{ij11}}{\hat{m}_{++11}}\right), \delta_{ij12} = \nabla_{12} \cdot \left(\frac{\hat{m}_{ij12}}{\hat{m}_{++12}} + \frac{1}{I \cdot J}\right) \cdot \frac{1}{2},$$

$$\delta_{ij21} = \nabla_{21} \cdot \left(\frac{\hat{m}_{ij21}}{\hat{m}_{++21}} + \frac{1}{I \cdot J}\right) \cdot \frac{1}{2},$$

and (11)

$$\delta_{ij22} = \nabla_{22} \cdot \left(\frac{\hat{m}_{ij22}}{\hat{m}_{++22}} + \frac{1}{I \cdot J}\right) \cdot \frac{1}{2}$$

where $\nabla_{kl} = p \cdot \hat{m}_{++kl} / \hat{m}_{++++}$ for $k, l = 1, 2$, and the term $1/IJ$ is the constant prior of Clogg *et al.* (1991) to prevent possible boundary solutions for m_{ij12}, m_{ij21}, and m_{ij22} (also see the fifth prior below). Thus, we allocate the third prior of δ_{ijkl} for the MPE of m_{ijkl} to shrink toward the ML obtained under the nonignorable nonresponse, whereas the first prior is obtained under an ignorable nonresponse model.

The fourth type of prior is defined by letting $\delta_{ij11} = 0$ in (11) as we did in obtaining the second type of prior from the first type. The last type of prior is from Clogg *et al.* (1991) defined as

$$\delta_{ij11} = 0, \delta_{ij12} = \frac{p}{3} \cdot \left(\frac{1}{I \cdot J}\right), \delta_{ij21} = \frac{p}{3} \cdot \left(\frac{1}{I \cdot J}\right),$$

and (12)

$$\delta_{ij22} = \frac{p}{3} \cdot \left(\frac{1}{I \cdot J}\right).$$

These five types of priors are summarized in Table 1 and are compared in the next section using empirical data and simulation studies.

Table 1
Five types of priors δ_{ijkl} ($\hat{m}_{ijkl}$ is MLE, I and J are the numbers of row and columns in a two-way table, and p is the number of parameters)

	δ_{ij11}	δ_{ij12}	δ_{ij21}	δ_{ij22}	
Type I	$\nabla_{11} \frac{y_{ij11}}{y_{++11}}$	$\nabla_{12} \frac{y_{ij11}}{y_{++11}}$	$\nabla_{21} \frac{y_{ij11}}{y_{++11}}$	$\nabla_{22} \frac{y_{ij11}}{y_{++11}}$,	$\nabla_{kl} = p \cdot \frac{y_{++kl}}{y_{++++}}$
Type II	0	$\nabla_{12} \frac{y_{ij11}}{y_{++11}}$	$\nabla_{21} \frac{y_{ij11}}{y_{++11}}$	$\nabla_{22} \frac{y_{ij11}}{y_{++11}}$,	$\nabla_{kl} = p \cdot \frac{y_{++kl}}{y^*_{++++}}$
Type III	$\nabla_{11} \cdot \left(\frac{\hat{m}_{ij11}}{\hat{m}_{++11}}\right)$	$\frac{\nabla_{12}}{2}\left(\frac{\hat{m}_{ij12}}{\hat{m}_{++12}} + \frac{1}{IJ}\right)$	$\frac{\nabla_{21}}{2}\left(\frac{\hat{m}_{ij21}}{\hat{m}_{++21}} + \frac{1}{IJ}\right)$	$\frac{\nabla_{22}}{2}\left(\frac{\hat{m}_{ij22}}{\hat{m}_{++22}} + \frac{1}{IJ}\right)$,	$\nabla_{kl} = p \cdot \frac{\hat{m}_{++kl}}{\hat{m}_{++++}}$
Type IV	0	$\frac{\nabla_{12}}{2}\left(\frac{\hat{m}_{ij12}}{\hat{m}_{++12}} + \frac{1}{IJ}\right)$	$\frac{\nabla_{21}}{2}\left(\frac{\hat{m}_{ij21}}{\hat{m}_{++21}} + \frac{1}{IJ}\right)$	$\frac{\nabla_{22}}{2}\left(\frac{\hat{m}_{ij22}}{\hat{m}_{++22}} + \frac{1}{IJ}\right)$,	$\nabla_{kl} = p \cdot \frac{\hat{m}_{++kl}}{\hat{m}^*_{++++}}$
Type V	0	$\nabla_{12}\left(\frac{1}{I \cdot J}\right)$	$\nabla_{21}\left(\frac{1}{I \cdot J}\right)$	$\nabla_{22}\left(\frac{1}{I \cdot J}\right)$,	$\nabla_{kl} = \frac{p}{3}$

$y^*_{++++} = y_{++++} - y_{++11}$ and $\hat{m}^*_{++++} = \hat{m}_{++++} - \hat{m}_{++11}$

Up to this point, we have presented methods for a two-way table, and y_{ijkl} is defined for the count of the (i, j) cell of the i^{th} row and j^{th} column (*i.e.*, $X_1 = i, X_2 = j$), and indicator R_1 for a missing row and R_2 for a missing column (*i.e.*, $R_1 = k, R_2 = l$). This can be easily extended to the 3-way table. Denote y_{ijklmn} to be the $(i, j, k)^{\text{th}}$ cell count for the three response variables (*i.e.*, $X_1 = i$, $X_2 = j$, and $X_3 = k$) and respective missing rows and columns (*i.e.*, $R_1 = l$, $R_2 = m$, and $R_3 = n$ for $l, m, n = 1, 2$). Thus, $lmn = 111$ implies that all of the three variables are observed, $lmn = 112$ implies that X_1 and X_2 are observed but X_3 is missing; similarly for $lmn = 121, 122, 211, 212, 221, 222$; 1 is for observed and 2 designates missing. Accordingly, the EM algorithm and priors for an incomplete three-way contingency table can be defined. The conditional expectation in the E-step for the $(i, j, k)^{\text{th}}$ cell with unknown information of k margin is

$$E_{\text{old}}(y_{ijk112} \mid \pi^{\text{old}}_{ijklmn}, y_{ij+112}) = y_{ij+112}\frac{m^{\text{old}}_{ijk112}}{m^{\text{old}}_{ij+112}}.$$

Similarly,

$$E_{\text{old}}(y_{ijk122} \mid \pi^{\text{old}}_{ijklmn}, y_{i++122}) = y_{i++122}\frac{m^{\text{old}}_{ijk122}}{m^{\text{old}}_{i++122}}.$$

and

$$E_{\text{old}}(y_{ijk222} \mid \pi^{\text{old}}_{ijklmn}, y_{+++222}) = y_{+++222}\frac{m^{\text{old}}_{ijk222}}{m^{\text{old}}_{+++222}}.$$

Other expectations and five types of priors can be similarly defined.

The Buckeye state poll is a Random Digit Dialing (RDD). No modification is necessary for the Bayesian procedures if the RDD is strictly a self-weighting survey (Lavrakas 1993; Potthoff 1994). However, RDD is not always done by a self-weighting design. For example, a telephone sample comprises a sample of households, not persons. If one person is interviewed in a household, a weight should be superimposed on the response by the number of persons in the household. A weight is also needed for the households with more than one telephone number. If an accurate estimate of the total number of households is available, stratification by region or state is possible and weighting must be considered in a comprehensive analysis. RDD was used in the 1998 Ohio election surveys. In this study, our method and models do not include weighting from stratification, clustering, and other factors leading to different probabilities of selection in a telephone survey.

However, further extension can be made for such weighting. A simple extension below shows how to accommodate a typical stratification. In a three-way table, let X_3 be the third response variable indexed by h $(h = 1, \ldots, H)$ that is assumed to be always observed. The H categories can be strata in a stratified sampling. Since X_3 is always observed, the corresponding missingness variable R_3 is equal to 1 and its observation can be denoted by y_{ijhlm1}. Then, we can write the following log likelihood for each stratum h:

$$\begin{aligned} l_h = {} & \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ijh111}\log(\pi_{ijh11}) + \sum_{i=1}^{I} y_{i+h121}\log(\pi_{i+h12}) \\ & + \sum_{j=1}^{J} y_{+jh211}\log(\pi_{+jhk21}) + y_{++h221}\log(\pi_{++h22}) \end{aligned}$$

where $\pi_{ijhlm} = P[X_1 = i, X_2 = j, R_1 = l, R_2 = m \mid X_3 = h]$. Thus, the terminology X_3 used for a three-way table acts as an indicator for strata. For each stratum h, the likelihood of (13) is exactly the same as that of a 2-way table.

Then, a log linear model for the cell expectation $m_{ijhlm} = N_h \cdot \pi_{ijhlm}$ can be defined in a similar way as in (2) where $N_h = \sum_{i,j,l,m} y_{ijhlm}$ for each $h = 1, 2, \ldots, H$. A nonignorable nonresponse model is given by

$$\begin{aligned} \log(m_{ijhlm}) = {} & \beta_{0h} + \beta^{i}_{X_1h} + \beta^{j}_{X_2h} + \beta^{l}_{R_1h} \\ & + \beta^{m}_{R_2h} + \beta^{ij}_{X_1X_2h} + \beta^{il}_{X_1R_1h} + \beta^{jm}_{X_2R_2h}. \quad (13) \end{aligned}$$

To avoid a boundary solution problem as in Section 2, we use the Dirichlet priors for π_{ijhlm}

$$\prod_i\prod_j \pi_{ijh11}^{\delta_{ijh11}} \cdot \pi_{ijh12}^{\delta_{ijh12}} \cdot \pi_{ijh21}^{\delta_{ijh21}} \cdot \pi_{ijh22}^{\delta_{ijh22}}.$$

Then, we follow exactly the same procedures as shown in Section 2 to estimate the cell expectations m_{ijhlm} for each $h = 1, 2 \ldots, H$. The estimate of the $(i, j)^{\text{th}}$ cell expectation is

$$\hat{E}(y_{ij}) = \sum_{h=1}^{H} w_h \sum_{l,m}\hat{m}_{ijhlm}$$

where w_h is the known weight for the h^{th} stratum and $\hat{m}_{ijhlm}$ is the m_{ijhlm} evaluated at the MPE of $\boldsymbol{\beta}$. For example, $w_h = N_h / \sum_h N_h$ is for a stratified sample where N_h is the population size of the h^{th} stratum.

The variance-covariance matrix of an approximation to the distribution of $\hat{\mathbf{m}}$ is

$$\frac{\partial\hat{\mathbf{m}}^T}{\partial\boldsymbol{\beta}}\operatorname{Var}(\hat{\boldsymbol{\beta}}_{\mathbf{MPE}})\frac{\partial\hat{\mathbf{m}}}{\partial\boldsymbol{\beta}} \quad (14)$$

where $\hat{\mathbf{m}}$ is a vector expression of the cell estimates $\hat{m}_{ijhlm}$, $\hat{\boldsymbol{\beta}}_{\mathbf{MPE}}$ is the MPE of $\boldsymbol{\beta}$ and its variance $\operatorname{Var}(\hat{\boldsymbol{\beta}}_{\mathbf{MPE}})$ is given by the inverse of (9), and $\partial\mathbf{m}/\partial\boldsymbol{\beta} = N_h \times [\operatorname{diag}(\hat{\boldsymbol{\pi}}) - \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^T]\mathbf{Z}$ where $\hat{\boldsymbol{\pi}}$ has

$$\hat{\pi}_{ijhlm} = \pi_{ijhlm}(\hat{\boldsymbol{\beta}}_{\mathbf{MPE}}) = \frac{\exp(\mathbf{z}_{ijhlm}\hat{\boldsymbol{\beta}}_{\mathbf{MPE}})}{\sum_{k\in(i,j,h,l,m)}\exp(\mathbf{z}_k\hat{\boldsymbol{\beta}}_{\mathbf{MPE}})}$$

as its typical element.

3. An application to a Buckeye State Poll

In forecasting the winner in a poll, the accuracy of the poll often depends on how to handle undecided voters who are likely to vote but who have not yet decided their preference for a candidate. We compare the Bayesian estimates based on the five types of priors with the ML estimate using the Buckeye State Poll (BSP) conducted in 1998 by the Center for Survey Research at Ohio State University. The BSP surveys produced incomplete two-way contingency tables with one category being candidate preference and the other category being the likelihood of voting in the November 1998 races for Ohio Governor, Attorney-General, Mayor of Columbus, and Treasurer. Table 2 summarizes these four polls and shows a substantial number of undecided voters.

For comparison, we consider the following ignorable Model 1 and the two nonignorable nonresponse Model 2 and Model 3.

$$\text{Model 1:}\quad \log(m_{ijkl}) = \beta_0 + \beta^i_{X_1} + \beta^j_{X_2} + \beta^k_{R_1} + \beta^l_{R_2} + \beta^{ij}_{X_1X_2} + \beta^{kl}_{R_1R_2},$$

$$\text{Model 2:}\quad \log(m_{ijkl}) = \beta_0 + \beta^i_{X_1} + \beta^j_{X_2} + \beta^k_{R_1} + \beta^l_{R_2} + \beta^{ik}_{X_1R_1} + \beta^{jl}_{X_2R_2} + \beta^{ij}_{X_1X_2} + \beta^{kl}_{R_1R_2},$$

$$\text{Model 3:}\quad \log(m_{ijkl}) = \beta_0 + \beta^i_{X_1} + \beta^j_{X_2} + \beta^k_{R_1} + \beta^l_{R_2} + \beta^{il}_{X_1R_2} + \beta^{jk}_{X_2R_1} + \beta^{ij}_{X_1X_2} + \beta^{kl}_{R_1R_2}.$$

Model 1 is missing completely at random, and cases with missing data can be ignorable in likelihood inferences. Model 2 and Model 3 are nonignorable where the probability of missing a variable depends on itself in Model 2 while the probability in Model 3 depends on the other variable. Note that the ML estimates in Model 1 and Model 3 are not on the boundary of the parameter space as shown by Baker *et al.* (1992). Moreover, since we found that, under Model 1 and Model 3, all of the five Bayesian estimates for the expected cell counts are not only fairly close to the ML estimate and their standard deviations are almost the same, we only present the ML estimates for Model 1 and Model 3.

We denote the ML estimates under ignorable Model 1, nonignorable Model 2, and nonignorable Model 3 by $IG1_{ML}$, $NON2_{ML}$, and $NON3_{ML}$, respectively. IG and NON stand for ignorable and nonignorable, respectively. We also let $NON2^{BE}_i$ be the Bayesian estimator using the i^{th} type of priors under Model 2. That is, $NON2^{BE}_1$ uses the respondent-driven priors of (10) and $NON2^{BE}_2$ is the same priors as $NON2^{BE}_1$ except for $\delta_{ij11} = 0$. Similarly, $NON2^{BE}_3$ is given by (11) and $NON2^{BE}_4$ is the same priors except for $\delta_{ij11} = 0$. $NON2^{BE}_5$ is the Bayesian estimate using the constant priors of (12). In addition, we can use the Stasny method (1986, 1988) to estimate the expected cell counts under Model 1 and Model 3 that she implicitly assumed. However, her estimates appear to be exactly the same as $IG1_{ML}$.

Table 2
Observed data for BSP pre-election surveys

	Governor race				Attorney-general race		
	Fisher	**Taft**	**Others**	**Undecided**	**Montgomery**	**Cordray**	**Undecided**
Likely to vote	112	140	23	61	197	82	57
Unlikely to vote	96	108	21	73	161	65	75
Undecided	7	11	1	4	15	4	0
	Mayor race				**Treasurer race**		
	Coleman	**Teater**	**Espy**	**Undecided**	**Deters**	**Donofrio**	**Undecided**
Likely to vote	40	32	25	30	127	119	90
Unlikely to vote	37	47	41	56	127	90	84
Undecided	0	2	1	0	10	7	0

The top table in Table 3 shows predicted values of elections using only "likely to vote" for the four races and their standard deviations in parentheses. The standard deviations are close to each other and show significant differences between the first and second leading candidates, except in the race for Mayor. This table also includes the actual election results and shows whether or not the ML estimates fall into the boundary solutions.

The bottom table shows the predictions of elections using both "likely to vote" and "unlikely to vote" to see what happens if those who responded to "unlikely to vote" actually voted. Comparing the two tables, we may conclude that the winners for Governor, Attorney-General, and the Treasurer's elections remained unchanged regardless of the likelihood of voting, whereas the winner could have changed in the Mayor's election if most of those who were "unlikely to vote" actually voted.

Based on Table 3, we can classify the 7 estimates, except $NON2_{ML}$, into two groups: $NON2_3^{BE}$, $NON2_4^{BE}$, and $NON2_5^{BE}$ to the first group and the remaining four estimates, $NON2_1^{BE}$, $NON2_2^{BE}$, $IG1_{ML}$, and $NON3_{ML}$ to the second group. As expected, since the priors δ_{ijkl} for $NON2_1^{BE}$ and $NON2_2^{BE}$ are so defined that the estimate of m_{ijkl} shrinks toward the ML under an ignorable nonresponse model, these two Bayesian estimates are very close to $IG1_{ML}$ and hence have little advantage over the $IG1_{ML}$. It is also interesting to note that $NON3_{ML}$ is almost the same as $IG1_{ML}$ although their loglinear models are differently specified.

There is no general criterion to evaluate whether an ignorable nonresponse model or a nonignorable nonresponse model is appropriate. However, as stated in Chen and Stasny (2003), the assumption of nonignorability for a nonresponse may be a reasonable assumption in the Buckeye State Poll study because people might be reluctant to express their preference for an unpopular candidate, or their current preferences are not firm or accurate at the time of the interview. In this regard, the $NON2_1^{BE}$, $NON2_2^{BE}$, and $NON3_{ML}$ may not be appropriate in these particular case studies because they are almost the same as the $IG1_{ML}$ of Model 1.

Table 3
Prediction of elections based on the October 98 and April 98 Buckeye State Polls (the unit is % and the numbers in parentheses are standard deviations)

	Governor			Mayor			Attorney-General		Treasurer	
	Fisher	Taft	Others	Coleman	Teater	Espy	Mongomery	Cordray	Deters	Donofrio
Likely to vote only used										
$NON2_{ML}$	33.2(2.75)	42.1(3.00)	24.8	31.5(4.65)	25.3(4.23)	43.2	75.6(3.71)	24.4	57.0(3.48)	43.0
$NON2_1^{BE}$	40.6(3.04)	48.5(3.27)	10.9	38.1(5.14)	34.2(4.78)	27.7	72.1(3.61)	27.9	52.7(3.36)	47.3
$NON2_2^{BE}$	40.9(3.01)	50.7(3.20)	8.40	39.9(5.04)	33.6(4.83)	26.5	71.0(3.59)	29.0	52.1(3.34)	47.9
$NON2_3^{BE}$	35.8(2.85)	44.5(3.08)	19.7	35.6(4.87)	29.3(4.51)	35.1	63.0(3.67)	37.0	54.3(3.41)	45.7
$NON2_4^{BE}$	36.3(2.87)	45.2(3.11)	18.6	35.9(4.91)	29.4(4.52)	34.6	63.0(3.64)	37.0	53.9(3.40)	46.1
$NON2_5^{BE}$	38.9(2.99)	47.4(3.20)	13.7	37.7(4.99)	33.6(4.77)	28.7	66.0(3.54)	34.0	51.5(3.32)	48.5
$IG1_{ML}$	40.6(3.03)	51.2(3.28)	8.20	40.8(5.16)	33.4(4.76)	25.8	70.9(3.59)	29.1	51.8(3.32)	48.2
$NON3_{ML}$	40.6(3.03)	51.2(3.28)	8.20	40.9(5.16)	33.3(4.75)	25.8	70.9(3.58)	29.1	51.7(3.32)	48.3
Actual result	45	50	5	39	37	24	63	37	57	43
Boundary		yes			yes		yes		no	
Likely to vote + Unlikely to vote										
$NON2_{ML}$	32.7(1.83)	39.4(1.91)	27.8	24.8(2.45)	26.2(2.49)	49.0	77.0(1.64)	23.0	60.2(1.93)	39.8
$NON2_1^{BE}$	41.3(1.93)	46.4(1.96)	12.3	30.7(2.68)	37.1(2.75)	32.2	72.8(1.74)	27.2	56.0(1.96)	44.0
$NON2_2^{BE}$	41.9(1.93)	49.2(1.95)	8.90	32.7(2.63)	36.5(2.76)	30.8	71.4(1.77)	28.6	55.3(1.96)	44.7
$NON2_3^{BE}$	35.4(1.87)	41.8(1.93)	22.7	27.8(2.55)	30.5(2.62)	41.7	61.0(1.72)	39.0	57.6(1.95)	42.4
$NON2_4^{BE}$	36.0(1.88)	42.6(1.93)	21.4	28.7(2.57)	30.6(2.62)	40.7	60.9(1.75)	39.1	57.2(1.95)	42.8
$NON2_5^{BE}$	39.1(1.91)	45.1(1.95)	15.8	30.7(2.63)	35.8(2.74)	33.5	64.8(1.88)	35.2	54.8(1.96)	45.2
$IG1_{ML}$	41.5(1.96)	49.8(1.96)	8.70	33.9(2.70)	36.1(2.74)	29.9	71.2(1.78)	28.8	55.0(1.96)	45.0
$NON3_{ML}$	41.5(1.96)	49.8(1.96)	8.70	34.1(2.71)	36.0(2.74)	29.9	71.1(1.78)	28.9	55.0(1.96)	45.0

Compared to actual election results, $NON2_{ML}$ gives the worst prediction for Governor, Mayor, and Attorney-General because the $NON2_{ML}$ lies on a boundary solution; whereas it provides the best prediction for Treasurer because it does not lie on a boundary solution. In the Attorney-General's election, $NON2_3^{BE}$ and $NON2_4^{BE}$ not only predicted the exact actual result but also are quite different from the other estimates. Since $NON2_3^{BE}$ and $NON2_4^{BE}$ have the priors to reflect different response patterns between respondents and the undecided, we can infer that the undecided voters in the Attorney-General race have quite different preference for the candidate from the respondents (*i.e.*, $NON2_3^{BE}$ and $NON2_4^{BE}$ allocate 19.4 % of the undecided voters who are likely to vote for Montgomery and 80.6% for Cordray, whereas the data in Table 2 indicates the percentage of Montgomery *vs* Cordray is 29.4% *vs* 70.6% among respondents who are likely to vote).

To see this difference between the respondents and undecided voters in terms of parameter estimates and to examine the effect of occurrence of the boundary solution on the estimates under the nonignorable Model 2, we present the ML estimates and $NON2_3^{BE}$ estimates and their corresponding standard deviations for the Attorney-General race in Table 4. Because of a boundary solution, all of the ML estimates have too large standard deviations as expected. On the other hand, $NON2_3^{BE}$ is very stable. Since $\beta^{11}_{X_1X_2} = 0.0472$ is the smallest and its standard deviation is relatively large, we neglect $\beta^{11}_{X_1X_2}$ to avoid complexity of interpretation. Under $\beta^{11}_{X_1X_2} = 0$, it is not difficult to show that, using the estimates of $NON2_3^{BE}$ in Table 4,

$$\log \frac{m_{1jl}}{m_{2jl}} = 2(\beta^1_{X_1} + \beta^{11}_{X_1R_1}) = 0.09$$

and

$$\log \frac{m_{1j2l}}{m_{2j2l}} = 2(\beta^1_{X_1} - \beta^{11}_{X_1R_1}) = 1.3916$$

for each fixed j and l, and

$$\log \frac{m_{i1k1}}{m_{i2k1}} = 2(\beta^1_{X_2} + \beta^{11}_{X_2R_2}) = 0.8982$$

and

$$\log \frac{m_{i1k2}}{m_{i2k2}} = 2(\beta^1_{X_2} - \beta^{11}_{X_2R_2}) = -1.4942$$

for each fixed i and k. Thus, by

$$\log \frac{m_{1jl}}{m_{2jl}} = 2(\beta^1_{X_1} + \beta^{11}_{X_1R_1}) = 0.09,$$

those who are likely to vote (*i.e.*, $i = 1$) are 1.09 times (*i.e.*, $e^{0.09}$) more than those who are unlikely to vote (*i.e.*, $i = 2$) among respondents ($k = 1$), whereas, by

$$\log \frac{m_{1j2l}}{m_{2j2l}} = 2(\beta^1_{X_1} - \beta^{11}_{X_1R_1}) = 1.3916,$$

likely voters of $i = 1$ are 4.02 times (*i.e.*, $e^{1.3916}$) more than unlikely voters of $i = 2$ among undecided ($k = 2$); by

$$\log \frac{m_{i1k1}}{m_{i2k1}} = 2(\beta^1_{X_2} + \beta^{11}_{X_2R_2}) = 0.8982,$$

those who vote for Montgomery are 2.46 times more than those who vote for Cordray among respondents; whereas, by

$$\log \frac{m_{i1k2}}{m_{i2k2}} = 2(\beta^1_{X_2} - \beta^{11}_{X_2R_2}) = -1.4942,$$

unlikely voters are 4.46 times more than likely voters among the undecided. This implies that the response pattern is much different between respondents and the undecided.

Table 4
ML and the third type Bayesian Estimates under nonignorable Model 2 for Attorney-General (the standard deviations are in parentheses)

	β_0	$\beta^1_{X_1}$	$\beta^1_{X_2}$	$\beta^1_{R_1}$	$\beta^1_{R_2}$	$\beta^{11}_{X_1R_1}$	$\beta^{11}_{X_2R_2}$	$\beta^{11}_{X_1X_2}$	$\beta^{11}_{R_1R_2}$
$NON2_{ML}$	-3.3735	-1.9487	3.2134	4.8496	4.8186	2.0283	-2.7594	-0.0452	-1.5588
		(3.120)	(8.515)	(3.996)	(8.871)	(3.120)	(8.512)	(0.045)	(2.501)
$NON2_3^{BE}$	0.6860	0.3704	-0.1490	3.3024	2.2942	-0.3254	0.5981	0.0472	-1.5450
		(0.118)	(0.052)	(2.501)	(2.501)	(0.117)	(0.052)	(0.041)	(2.501)

The extent of this difference can be measured by the most important terms, $\beta^{11}_{X_1R_1}$ and $\beta^{11}_{X_2R_2}$, in the nonignorable nonresponse model, Model 2. Since

$$\beta^{11}_{X_1R_1} = \frac{1}{4}\log\frac{m_{1111}/m_{2111}}{m_{1121}/m_{2121}} = -0.3254$$

and

$$\beta^{11}_{X_2R_2} = \frac{1}{4}\log\frac{m_{1111}/m_{1211}}{m_{1112}/m_{1212}} = 0.5981,\ \beta^{11}_{X_1R_1}$$

is the log-odds ratio that shows the log difference between the ratio of the number of those "likely to vote" to that of those "unlikely to vote" among the decided voters for Montgomery and the same ratio among the undecided voters who prefer Montgomery but who do not express their likelihood of voting. Whereas, $\beta^{11}_{X_2R_2}$ is the log-odds ratio that shows the log difference between the ratio of the number of voters for Montgomery to the voters for Cordray among the decided who are likely to vote and the same ratio among the undecided voters who are likely to vote but who do not express their candidate preference. Thus, among voters for Montgomery, the possibility for the undecided voters to vote relative to not voting is about 3.67 times

$$\left(i.e.,\ \frac{m_{1111}/m_{2111}}{m_{1121}/m_{2121}} = e^{4\times-0.3254} = 3.67^{-1}\right)$$

larger than the possibility of the decided, implying that Montgomery needs a strategy to raise the turnout of voters. On the other hand, among those likely to vote, the supporting rate of the decided for Montgomery is about 10.94 times

$$\left(i.e.,\ \frac{m_{1111}/m_{1211}}{m_{1112}/m_{1212}} = e^{4\times0.5981} = 10.94\right)$$

larger than the undecided voters for Montgomery, implying that most of the undecided voters not exposing their preference of candidate are likely to vote for Cordray as the Attorney-General. This also confirms the popular account that voters are inclined to remain "undecided" in a poll if they support the candidate who is seen as inferior in a race and that the voters are inclined to abstain from voting if they support the candidate who certainly dominates the race.

4. Simulation study

We consider a 2×2 contingency table with supplemental margins to compare the performance of the five Bayesian estimates described in Section 2 for different missing percentages and different response patterns under the following nonignorable nonresponse model (*i.e.*, Model 2):

$$\log(m_{ijkl}) = \beta_0 + \beta^i_{X_1} + \beta^j_{X_2} + \beta^k_{R_1} + \beta^l_{R_2} + \beta^{ik}_{X_1R_1} + \beta^{jl}_{X_2R_2} + \beta^{ij}_{X_1X_2} + \beta^{kl}_{R_1R_2}.$$

Thus, we only compare $NON2_{ML}$ and $NON2^{BE}_i$ for $i = 1, \ldots, 5$ in this simulation study.

Since there are two levels in all of X_1, X_2, R_1, and R_2, there are 8 parameters to be determined for the simulation study. From the equations of

$$4\beta^{11}_{X_1R_1} = \log\frac{m_{1111}/m_{2111}}{m_{1121}/m_{2121}} \text{ and } 4\beta^{11}_{X_2R_2} = \log\frac{m_{1111}/m_{1211}}{m_{1112}/m_{1212}},$$

$$\beta^{11}_{X_1R_1} = \beta^{11}_{X_2R_2} = 0$$

means that there is no difference in the response pattern between respondents and undecided. The bigger $\beta^{11}_{X_1R_1}$ and $\beta^{11}_{X_2R_2}$ are, the more different the response pattern between respondents and undecided voters is. We vary these two parameters from 0.2 to 0.8 with an increment of 0.2. We set the missing percentage to 20% and 30% by adjusting $\beta^1_{X_1}$ and $\beta^1_{R_1}$ and fixing

$$\frac{m_{1111}/m_{1211}}{m_{2111}/m_{2211}} = 5,\ \frac{m_{1111}/m_{1112}}{m_{1112}/m_{1122}} = 2,$$

and

$$N = \sum\nolimits_{ijkl} m_{ijkl} = 1{,}000.$$

This implies that the size and missing percentage for the cell of $X_1 = 1$ and $X_2 = 1$ are approximately 5 times and 2 times the size of the other three cells, respectively.

We generate a large number of samples $\{y_{ijkl}, i, j, k, l = 1, 2\}$ from the above setting until we have 1,000 random samples with boundary solutions and the other 1,000 with no boundary solutions. The occurrence of a boundary solution is determined by the criterion given in Michiels and Molenberghs (1997) (also see Clarke (2002), Smith *et al*. (1999) for more details). Using $\{y_{ij11}, y_{i+12}, y_{+j21}, y_{++22}, i, j, = 1, 2\}$ obtained from the generated data, the expected cell counts m_{ijkl}'s are estimated by each of the five Bayesian estimates and the ML estimate described in Section 2.

We calculate mean squared errors (MSEs) and absolute biases of $NON2_{ML}$, $NON2^{BE}_1$, ..., $NON2^{BE}_5$ for $\{\sum_{kl} m_{ijkl}, i, j = 1, 2\}$. Then we take the mean over the four MSEs and the four absolute biases, which we obtain from each estimate to see the overall performance of the estimate. Similarly, we calculate mean MSEs and mean absolute biases for $\{m_{ij12} + m_{ij21} + m_{ij22}, i, j = 1, 2\}$ to see the performance of each estimate in imputing the nonresponses.

Table 5 shows the ratios of the mean MSEs and mean absolute biases of the five Bayesian estimates (*i.e.*, $NON2_1^{BE}$, ..., $NON2_5^{BE}$), relative to the ML estimate (*i.e.*, $NON2_{ML}$) when the boundary solutions occur; whereas Table 5 shows the ratios when no boundary occurs. Thus, values less than 1 imply that the corresponding Bayesian estimate has a smaller mean MSE or a smaller mean absolute bias than the ML estimate. Both tables only show the cases for $\beta_{X_1R_1}^{11} < \beta_{X_2R_2}^{11}$ and for 20% of the missing percentage because the MSEs and biases are almost symmetric about the coordinate of $(\beta_{X_1R_1}^{11}, \beta_{X_2R_2}^{11})$. They increase as we increase the missing percentage to 30% while keeping the same patterns of the MSEs and biases as those of the missing 20%.

Table 5, where a boundary solution occurs, shows that $NON2_1^{BE}$, $NON2_3^{BE}$, $NON2_4^{BE}$ have smaller MSEs than the ML estimate (*i.e.*, $NON2_{ML}$) for all values of $\beta_{X_1R_1}^{11}$ and $\beta_{X_2R_2}^{11}$, except $(\beta_{X_1R_1}^{11}, \beta_{X_2R_2}^{11})$ = (0.8, 0.8). Here, $NON2_3^{BE}$ has a smaller MSE than the ML estimate. This is true for the absolute biases. On the other hand, Table 6, where no boundary solution occurs, shows that only $NON2_3^{BE}$ is comparable to the ML estimate in the MSE although it is slightly biased. In particular, $NON2_3^{BE}$ has a smaller MSE than the ML estimate as long as $\beta_{X_1R_1}^{11} \neq 0.8$ or $\beta_{X_2R_2}^{11} \neq 0.8$ (*i.e.*, The response pattern between respondents and nonrespondents is not very different.).

Table 5
Ratios of mean MSEs and mean absolute biases of Bayesian estimates relative to the ML estimate when boundary solutions occur under a 20% missing percentage (the ratios for absolute biases are in parentheses)

	$(\beta_{X_1R_1}^{11}, \beta_{X_2R_2}^{11})$	$NON2_1^{BE}$	$NON2_2^{BE}$	$NON2_3^{BE}$	$NON2_4^{BE}$	$NON2_5^{BE}$
For $\{m_{ij11} + m_{ij12} + m_{ij21} + m_{ij22},\ i,\ j = 1,\ 2\}$	(0.2, 0.2)	0.68(0.66)	0.47(0.22)	0.76(0.76)	0.65(0.48)	0.42(0.05)
	(0.2, 0.4)	0.68(0.48)	0.57(0.20)	0.77(0.68)	0.60(0.29)	0.56(0.30)
	(0.2, 0.6)	0.67(0.23)	0.73(0.66)	0.77(0.57)	0.64(0.10)	0.69(0.64)
	(0.2, 0.8)	0.77(0.26)	1.08(1.55)	0.83(0.43)	0.76(0.28)	0.95(1.34)
	(0.4, 0.4)	0.65(0.32)	0.69(0.57)	0.76(0.63)	0.61(0.17)	0.65(0.52)
	(0.4, 0.6)	0.58(0.14)	0.83(0.90)	0.71(0.56)	0.56(0.06)	0.69(0.71)
	(0.4, 0.8)	0.75(0.36)	1.46(2.07)	0.78(0.36)	0.74(0.42)	1.12(1.61)
	(0.6, 0.6)	0.66(0.22)	1.35(1.73)	0.73(0.43)	0.66(0.16)	1.01(1.29)
	(0.6, 0.8)	0.85(0.87)	2.27(3.19)	0.76(0.17)	0.83(0.81)	1.52(2.35)
	(0.8, 0.8)	1.12(1.93)	3.58(5.49)	0.83(0.24)	1.04(1.67)	2.18(3.95)
For $\{m_{ij12} + m_{ij21} + m_{ij22},\ i,\ j = 1,\ 2\}$	(0.2, 0.2)	0.57(0.63)	0.27(0.13)	0.69(0.74)	0.41(0.40)	0.28(0.31)
	(0.2, 0.4)	0.54(0.46)	0.37(0.34)	0.68(0.68)	0.42(0.24)	0.44(0.57)
	(0.2, 0.6)	0.51(0.19)	0.69(0.94)	0.65(0.55)	0.47(0.10)	0.69(0.88)
	(0.2, 0.8)	0.63(0.35)	1.39(2.08)	0.71(0.34)	0.62(0.47)	1.11(1.52)
	(0.4, 0.4)	0.49(0.35)	0.54(0.64)	0.65(0.64)	0.42(0.17)	0.57(0.76)
	(0.4, 0.6)	0.48(0.17)	0.98(1.24)	0.62(0.51)	0.45(0.17)	0.85(1.04)
	(0.4, 0.8)	0.62(0.44)	1.81(2.33)	0.67(0.35)	0.61(0.55)	1.35(1.81)
	(0.6, 0.6)	0.55(0.42)	1.70(1.90)	0.63(0.41)	0.54(0.40)	1.28(1.51)
	(0.6, 0.8)	0.78(0.92)	2.91(3.43)	0.69(0.14)	0.75(0.92)	1.96(2.64)
	(0.8, 0.8)	1.13(1.96)	4.63(5.72)	0.75(0.33)	1.02(1.77)	2.86(4.24)

Table 6
Ratios of mean MSEs and mean absolute biases of Bayesian estimates relative to the ML estimate when no boundary solution occurs under a 20% missing percentage (the ratios for absolute biases are in parentheses)

	$(\beta^{11}_{X_1R_1}, \beta^{11}_{X_2R_2})$	$NON2^{BE}_1$	$NON2^{BE}_2$	$NON2^{BE}_3$	$NON2^{BE}_4$	$NON2^{BE}_5$
For $\{m_{ij11} + m_{ij12} + m_{ij21} + m_{ij22},\ i,\ j = 1,\ 2\}$	(0.2, 0.2)	0.99(3.37)	1.05(7.00)	0.94(2.51)	0.93(4.89)	1.06(8.96)
	(0.2, 0.4)	0.98(2.57)	1.21(5.13)	0.97(1.89)	1.00(3.26)	1.24(5.56)
	(0.2, 0.6)	1.04(2.18)	1.52(3.84)	0.95(1.67)	1.06(2.38)	1.43(3.71)
	(0.2, 0.8)	1.12(2.04)	1.75(3.53)	1.00(1.48)	1.13(2.14)	1.52(3.21)
	(0.4, 0.4)	1.03(2.40)	1.49(4.66)	0.97(1.69)	1.05(2.74)	1.39(4.46)
	(0.4, 0.6)	1.20(2.17)	2.11(3.85)	1.00(1.52)	1.22(2.24)	1.78(3.42)
	(0.4, 0.8)	1.28(2.09)	2.36(3.67)	1.05(1.45)	1.26(2.09)	1.86(3.12)
	(0.6, 0.6)	1.22(2.16)	2.49(3.90)	0.96(1.48)	1.21(2.15)	1.90(3.32)
	(0.6, 0.8)	1.52(1.99)	3.19(3.39)	1.11(1.38)	1.45(1.91)	2.29(2.77)
	(0.8, 0.8)	1.66(1.96)	3.64(3.27)	1.14(1.36)	1.52(1.83)	2.43(2.59)
For $\{m_{ij12} + m_{ij21} + m_{ij22},\ i,\ j = 1,\ 2\}$	(0.2, 0.2)	0.88(2.59)	0.89(5.66)	0.87(2.26)	0.89(4.55)	1.21(8.69)
	(0.2, 0.4)	0.93(2.40)	1.27(4.86)	0.93(1.78)	1.00(3.08)	1.50(5.29)
	(0.2, 0.6)	1.09(2.11)	1.93(3.97)	0.98(1.40)	1.15(2.29)	1.85(3.61)
	(0.2, 0.8)	1.24(2.13)	2.36(3.90)	1.02(1.48)	1.27(2.18)	2.06(3.19)
	(0.4, 0.4)	1.03(2.18)	1.81(4.30)	0.96(1.60)	1.12(2.62)	1.85(4.39)
	(0.4, 0.6)	1.23(2.28)	2.62(4.28)	0.99(1.48)	1.29(2.42)	2.28(3.80)
	(0.4, 0.8)	1.42(2.05)	3.26(3.70)	1.07(1.42)	1.44(2.07)	2.53(3.09)
	(0.6, 0.6)	1.33(2.07)	3.22(3.95)	0.99(1.36)	1.36(2.14)	2.54(3.43)
	(0.6, 0.8)	1.65(2.09)	4.14(3.74)	1.13(1.43)	1.61(2.07)	2.98(3.13)
	(0.8, 0.8)	1.91(2.02)	4.48(3.50)	1.16(1.39)	1.66(1.93)	3.03(2.83)

Park and Brown (1994) used $NON2^{BE}_2$ to estimate expected cell counts in an incomplete one-way table under a nonignorable nonresponse mechanism. They showed by simulation studies that $NON2^{BE}_2$ has a smaller MSE than the ML estimate although it is biased more than the ML. However, larger values than 1 for $NON2^{BE}_2$ in Table 5 and Table 6 indicate that this is not true in an incomplete two-way table regardless of the boundary solution and that Bayesian methods are not always better than the ML even when a boundary solution occurs. A reason that our simulation results differ from those of Park and Brown (1994) when a boundary solution occurs is attributed to the choice of $(\beta^{11}_{X_1R_1}, \beta^{11}_{X_2R_2})$ where Park and Brown performed their simulation only for $\beta^{11}_{X_1R_1} = \beta^{11}_{X_2R_2} = 0.34$. As shown in Table 5, $NON2^{BE}_2$ is better than the ML when $\beta^{11}_{X_1R_1} \le 0.4$ and $\beta^{11}_{X_2R_2} \le 0.4$, whereas $NON2^{BE}_2$ is worse than the ML when the response pattern between respondents and nonrespondents is much different (*i.e.*, $\beta^{11}_{X_1R_1} \ge 0.6$ or $\beta^{11}_{X_2R_2} \ge 0.6$).

Table 7 provides the mean of the standard deviations and the 95% coverage probabilities for $\beta^{11}_{X_1R_1}$. Here, we used the variance formula given in (9) to calculate the standard deviations and the 95% coverage probabilities are the coverage rates for nominal 95% confidence intervals. When a boundary solution occurs, although the coverage probability of the ML estimate is closest to the 95% nominal coverage level, the ML estimate has too large a standard deviation to use in practice. Such large standard deviations are due to the boundary problem of the ML estimate. The coverage probabilities of $NON2^{BE}_3$ are the closest to the 95% nominal coverage level among the Bayesian estimates, while those of the other Bayesian estimates are generally smaller than the 95% nominal coverage level. This implies that the Bayesian estimates other than $NON2^{BE}_3$ underestimate their standard deviations.

When no boundary solution occurs (the second table in Table 7), the standard deviations of the ML estimate are much more stable, compared to those for the boundary solution case. The coverage probability decreases as $\beta^{11}_{X_1R_1}$ and $\beta^{11}_{X_2R_2}$ increase. In particular, the coverage probabilities of NON^{BE}_1, NON^{BE}_2, and NON^{BE}_5 are seriously smaller than the 95% nominal coverage level when the response pattern between the respondents and undecided voters is much different (*i.e.*, $\beta^{11}_{X_1R_1} \ge 0.6$ and $\beta^{11}_{X_2R_2} \ge 0.6$).

Table 7
Mean of standard deviations and 95% coverage probabilities (in parentheses) for $\beta^{11}_{X_1 R_1}$

	$(\beta^{11}_{X_1R_1}, \beta^{11}_{X_2R_2})$	$NON2_{ML}$	$NON2^{BE}_1$	$NON2^{BE}_2$	$NON2^{BE}_3$	$NON2^{BE}_4$	$NON2^{BE}_5$
boundary	(0.2, 0.2)	89.5(0.974)	0.082(0.978)	0.064(0.978)	0.093(0.973)	0.071(0.972)	0.060(0.957)
	(0.2, 0.4)	158.3(0.959)	0.072(0.963)	0.096(0.963)	0.079(0.958)	0.066(0.958)	0.058(0.940)
	(0.2, 0.6)	135.3(0.940)	0.065(0.941)	0.057(0.941)	0.071(0.941)	0.062(0.939)	0.056(0.922)
	(0.2, 0.8)	57.4(0.930)	0.070(0.938)	0.061(0.935)	0.076(0.928)	0.066(0.928)	0.060(0.908)
	(0.4, 0.4)	153.4(0.961)	0.079(0.920)	0.061(0.913)	0.096(0.956)	0.072(0.949)	0.060(0.911)
	(0.4, 0.6)	82.2(0.955)	0.072(0.893)	0.059(0.883)	0.086(0.951)	0.069(0.940)	0.058(0.874)
	(0.4, 0.8)	51.2(0.933)	0.071(0.862)	0.059(0.849)	0.084(0.926)	0.068(0.917)	0.059(0.846)
	(0.6, 0.6)	175.5(0.946)	0.077(0.820)	0.060(0.781)	0.101(0.943)	0.074(0.921)	0.061(0.823)
	(0.6, 0.8)	159.6(0.924)	0.071(0.728)	0.057(0.657)	0.089(0.913)	0.069(0.880)	0.058(0.737)
	(0.8, 0.8)	72.8(0.920)	0.070(0.572)	0.056(0.330)	0.093(0.900)	0.070(0.842)	0.058(0.607)
no-boundary	(0.2, 0.2)	0.068(0.949)	0.060(0.959)	0.056(0.959)	0.062(0.937)	0.058(0.935)	0.055(0.922)
	(0.2, 0.4)	0.066(0.960)	0.060(0.970)	0.056(0.970)	0.061(0.935)	0.058(0.931)	0.055(0.951)
	(0.2, 0.6)	0.064(0.940)	0.058(0.945)	0.055(0.945)	0.059(0.959)	0.057(0.919)	0.054(0.909)
	(0.2, 0.8)	0.069(0.933)	0.063(0.944)	0.059(0.941)	0.065(0.926)	0.062(0.925)	0.058(0.920)
	(0.4, 0.4)	0.074(0.910)	0.061(0.836)	0.055(0.828)	0.064(0.899)	0.059(0.884)	0.055(0.824)
	(0.4, 0.6)	0.074(0.915)	0.060(0.815)	0.055(0.806)	0.064(0.922)	0.059(0.879)	0.055(0.792)
	(0.4, 0.8)	0.073(0.891)	0.061(0.786)	0.056(0.771)	0.064(0.873)	0.060(0.852)	0.056(0.763)
	(0.6, 0.6)	0.078(0.859)	0.061(0.567)	0.055(0.470)	0.067(0.853)	0.061(0.795)	0.056(0.572)
	(0.6, 0.8)	0.076(0.843)	0.060(0.515)	0.054(0.402)	0.065(0.817)	0.060(0.767)	0.055(0.556)
	(0.8, 0.8)	0.080(0.755)	0.059(0.110)	0.053(0.017)	0.065(0.728)	0.059(0.607)	0.055(0.158)

5. Concluding remarks

We investigated the Bayesian analysis for incomplete two-way contingency tables with nonignorable non-response. In this situation, the ML estimates often fall on the boundary solution. These boundary solutions can yield $G^2 > 0$ even for a saturated model (Baker *et al.* 1992; Park and Brown 1994). This means that the G^2 may not be appropriate as a statistic for model specification. To avoid the boundary solution problem and to obtain a statistic such as a Bayes factor for model specification regardless of a boundary solution, we proposed Bayesian estimation methods using five different priors. Two of them are new and the remaining three have been previously used for analyzing an incomplete one-way table. These two new priors accommodate different response patterns between respondents and nonrespondents.

Data analysis shows that these new two priors are more reasonable in the sense that they accommodate the nonignorable nonresponse mechanism better and produce estimates close to the actual results. Moreover, with the previous three priors, our simulation study shows that the Bayesian estimates can have larger MSEs than those of the ML estimates for a contingency table with no boundary solution and a boundary solution as well, contrary to the previous studies. However, when a boundary solution occurs, the two new priors perform better than the previous three priors and the ML estimates in the sense that they have generally smaller MSEs, smaller biases, and coverage probabilities closer to the nominal coverage level.

We have briefly discussed the weighting issues at Section 2.2. However, these issues need much more rigorous discussion than we did in that section. Our discussion can be further extended to include not only different weights but also response biases and other sources of biases and variations. These problems can be carefully developed on an extended paper at a later time.

Acknowledgements

This research was supported by a Korea University Grant (K0822301).

References

Agresti, A. (2002). *Categorical Data Analysis*. 2nd Edition. New York: John Wiley & Sons, Inc.

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.

Baker, S.G., Rosenberger, W.F. and Dersimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine*, 11, 643-657.

Chen, T. (1972). Mixed-up frequencies and missing data in contingency tables. Unpublished Ph.D. dissertation, University of Chicago, Dept. of Statistics.

Chen, Q.L., and Stasny, E.A. (2003). Handling undecided voters: Using missing data methods in election forecasting. *Technical Report*, Department of Statistics, The Ohio State University.

Clarke, P.S. (2002). On boundary solutions and identifiability in categorical regression with non-ignorable non-response. *Biometrical Journal*, 44, 701-717.

Clogg, C.C., Rubin, D.B., Schenker, N. and Schultz, B. (1991). Multiple imputation of industry and occupation codes in Census Public use-samples using Bayesian logistic regression. *Journal of the American Statistical Association*, 86, 68-78.

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

Dempster, A.P., Laird, N.M. and Rubin, D.M. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B, 39, 1-38.

Flannelly, K.J., Flannelly, L.T. and McLeod, M.S. Jr. (2000). Reducing undecided voters and other sources of error in election surveys. *International Journal of Market Research*, 42, 231-237.

Fenwick, I, Wiseman, F, Becker, J.F. and Heiman, J.R. (1982). Classifying undecided voters in pre-election polls. *Public Opinion Quarterly*, 46, 383-391.

Forster, J.J., and Smith, R.W.F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response. *Journal of the Royal Statistical Society B*, 60, 57-70.

Gelman, A., Carlin, J.P., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*. 2nd Edition. New York: Chapman and Hall/CRC.

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

Kim, T. (1995). Discriminant analysis as a prediction tool for uncommitted voters in pre-election polls. *International Journal of Public Opinion Research*, 7, 110-127.

Lau, R.R. (1994). An analysis of the accuracy of “trial heat” polls during the 1992 presidential elections. *Public Opinion Quarterly*, 59, 589-605.

Lavrakas, P.J. (1993). *Telephone Survey Method*: *Sampling, selection, and supervision*. 2nd Edition. Newbury Park, Calif.: Sage.

Little, J.A., and Rubin, D.B. (2002). *Statistical analysis with missing data*. 2nd Edition. New York: John Wiley & Sons, Inc.

Martin, E.A., Traugott, M.W. and Kennedy, C. (2005). A review and proposal for a new measure of poll accuracy. *The Public Opinion Quarterly*, 69, 342-369.

Michiels, B., and Molenberghs, G. (1997). Protective estimation of longitudinal categorical data with nonrandom drop-out. *Communications in Statistics*: *Theory and Methods*, 26, 65-94.

Molenberghs, G., Kenward, M.G. and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: The Slovenian plebiscite case. *Applied Statistics*, 50, 15-29.

Monterola, C., Lim, M., Garcia, F. and Saloma, C. (2001). Feasibility of a neural network as classifier of undecided respondents in a public opinion survey. *International Journal of Public Opinion Research*, 14, 222-299.

Myers, D.J., and O’Connor, R.E. (1983). The undecided respondents in mandatory voting settings: A Venezuelan exploration. *The Western Political Quarterly*, 36, 420-433.

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

Park, T. (1998). An approach to categorical data with nonignorable nonresponse. *Biometrics*, 54, 1579-1690.

Perry, P. (1979). Certain problem in election survey methodology. *Public Opinion Quarterly*, 43, 312-325.

Potthoff, R.F. (1994). Telephone sampling in epidemiologic research: To reap the benefits, avoid the pitfalls. *American Journal of Epidemiology*, 139, 967-978.

Rubin, D.B., Stern, H.S. and Vehovar, V. (1995). Handling “Don’t Know” survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association*, 90, 822-828.

Smith, P.W.F., Skinner, C.J. and Clarke, P.S. (1999). Allowing for non-ignorable nonresponse in the analysis of voting intention data. *Applied Statistics*, 48, 563-577.

Smith, T.W. (1984). Non attitudes: A review and evaluation. In *Surveying Subjective Phenomena*, (Eds. C.F. Turner and E. Martin), New York: Russell Sage Foundation, 2, 215-255.

Stasny, E.A. (1986). Estimating gross flow using panel data with nonresponse: An example from the Canadian Labor Force survey. *Journal of the American Statistical Association*, 81, 42-47.

Stasny, E.A. (1988). Modelling nonignorable nonresponse in categorical panel data with an example in estimating Gross Labor-Force flows. *Journal of Business and Economic Statistics*, 6, 207-219.