Bayesian Adjustments for Nonignorable Nonresponse in Incomplete Two-way Categorical Table

Boseung Choi and Yousung Park
Department of Statistics, Korea University

KEY WORDS: Bayesian analysis; Nonignorable nonresponse; Priors; Boundary solution; EM algorithm

1. Introduction

The problem of missing data arising from nonresponse is common in most surveys and becomes a serious issue as the nonresponse rate increases (De Heer 1999). When survey data are presented in a two-way categorical table, they include fully classified counts, partially classified counts, and unclassified counts. The partially classified counts are often presented as row and column supplemental margins and called them item nonresponses, whereas the unclassified counts are presented in total, called unit nonresponses.

In the Buckeye State Poll (BSP) (Chen and Stasny, 2003) for U.S. Senator, Ohio 1998, one category involves the candidate 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), from which one can estimate the true support for each candidate. This poll was conducted prior to the November 1998 race and the May 1999 race. The goal is how to impute the individuals responding to "undecided" and to utilize the information from the successive results of polls to predict the actual election result.

Throughout this paper, we define two-way categorical table when the two categorical response variable are allowed to be missing, whereas, if only one response variable has missing, then we define this case as one-way categorical table. In this paper, we only consider a two-way categorical table where the two categorical response variables are allowed to be missing, because the extension to more than two categorical variables is straightforward.

Nonresponse can be further distinguished by three types of nonresponses (Little and Rubin 1987): missing completely at random (MCAR) which means that the probability of missing on a variable of interest is independent of all variables including itself in the survey; missing at random (MAR) in which the nonresponse depends only on observed data; nonignorable in which nonresponse depends on the unobserved values. Any model with MCAR or MAR is called ignorable nonresponse model.

When data are assumed as MCAR, cases with missing data can be removed in likelihood inferences (Little and Rubin 1987). However, when the nonrespondents follow a different response mechanism from the respondents (such as a MAR and nonignorable model), discarding the incomplete cases or misspecifying nonresponse mechanism leads to larger variances and biases in estimation (Chen 1972, Park and Brown 1994).

When the response mechanism obeys nonignorable nonresponse in categorical data analysis, the maximum likelihood estimation often yields boundary solutions where the probability of nonresponse is estimated to be zero in some cells of the table. In such a case the ML estimate is not uniquely determined and may be unstable (Park and Brown 1994). Furthermore, as indicated in Baker and Laird (1988), the boundary solution yields a positive deviance G^2 even for saturated or over-parameterized loglinear models.

The conditions that the maximum likelihood (ML) suffers from the boundary solution have been proposed in one-way categorical table (Baker and Laird 1988, Michels and Molenbergs 1997). Geometric explanation was presented for the boundary solution of the maximum likelihood (Smith, Skinner and Clarke 1999, Clark 2002). Baker, Rosenberger and Dersimonian (1992) presented close forms of ML estimates for incomplete two-way categorical tables using loglinear model. In particular, they provided a sufficient and necessary condition under which the ML estimates fall in the boundary solution in two-way categorical tables.

Park and Brown(1994) and Park (1998) proposed a Bayesian approach to avoid the boundary solution problem in a one-way categorical table. They used a loglinear model to link the cell expectation to relevant covariates and imposed a prior on the parameters related to nonresponse. The prior depends only on information of respondents. However, this respondent-driven prior contradicts to the fundamental principle that the nonrespondents have different response pattern from those of respondents in the nonignorable nonresponse model. We extend

Park and Brown and Park's empirical Bayesian approach (1994, 1998) not only to a two-way categorical table but also to the prior depending on information from both respondent and nonrespondent. This prior can reflect different response patterns between respondents and nonrespondents and produces smaller mean squared errors and biases than those of respondent-dependent prior.

Summarizing all of the above discussions, we are focusing on incomplete two-way categorical tables with nonignorable nonresponse and on Bayesian approach in estimating model parameters of loglinear models. Throughout this paper, we assume no explanatory variable (which is completely observed) in our analysis without loss of generality because we could include it in the loglinear model as a covariate.

The remainder of this paper is divided into four sections. In Section 2, we present the Bayesian models and introduce five types of Bayesian approaches to overcome the boundary solution of the ML methods. We also present generalized expectation maximization (EM) algorithm to estimate the cell probability specified by the loglinear models. In section 3, we use eight empirical data sets from the Buckeye State Poll (BSP) to demonstrate the impact of the boundary solution of the ML and to compare the five types of Bayesian approaches to the actual results of elections. Section 4 includes some concluding remarks and future works.

2. Bayesian models

We describe five Bayesian approaches to accommodate nonignorable nonresponse in a two-way categorical table. Let X_1 and X_2 be response variables indexed by I and J categories, respectively. 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 $I \times J \times 2 \times 2$ categorical 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 ith category of X_1, the jth category of X_2, the kth value of R_1, and the lth value of R_2. Thus, y_{ij11} is used for completely classified counts, y_{i+12} and y_{+j21} for respective column and row supplemental margins, and y_{++22} for unclassified counts.

Throughout this chapter, we assume a multinomial assumption for the three types of observations to have the following log likelihood proportional to

$$\begin{aligned} l \;\propto\; & \sum_i \sum_j y_{ij11} \cdot \log(\pi_{ij11}) + \sum_i y_{i+12} \cdot \log(\pi_{i+12}) \\ + \; & \sum_j y_{+j21} \cdot \log(\pi_{+j21}) + y_{++22} \cdot \log(\pi_{++22}) \quad (1) \end{aligned}$$

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 by a loglinear function. It is assumed that no explanatory variable is available throughout this paper. However, note that the loglinear model easily incorporates any explanatory variable in the same way that it incorporates the categorical variables.

Loglinear model is a tool frequently used for analyzing incomplete categorical tables with nonignorable nonresponse. Let p be the number of parameters to be estimated and z_{ijkl} be the $p \times 1$ covariate vector for the observation belonging to the (i, j, k, l)th category. Then the loglinear model can be written as

$$\log \mathbf{m} = \mathbf{Z}\boldsymbol{\beta} \quad (2)$$

where $\mathbf{m}$ is the cell expectation. To avoid a boundary solution of ML in model (1), we impose the Dirichlet priors to the cell probabilities (π_{ij11}, π_{ij12}, π_{ij21}, π_{ij22}) as given by

$$\prod_i \prod_j \pi_{ij11}^{\delta_{ij11}} \cdot \pi_{ij12}^{\delta_{ij12}} \cdot \pi_{ij21}^{\delta_{ij21}} \cdot \pi_{ij22}^{\delta_{ij22}}. \quad (3)$$

Together with (2), the multinomial distribution of (1) for observations and the prior distribution of (3) yield the following log posterior distribution:

$$\begin{aligned} l_{pos} = & \sum_i \sum_j y_{ij11} \cdot (\mathbf{m}_{ij11} \cdot \beta) \\ & + \sum_i y_{i+12} \cdot \log\left(\sum_j \exp(\mathbf{z}_{ij12} \cdot \beta)\right) \\ & + \sum_j y_{+j21} \cdot \log\left(\sum_i \exp(\mathbf{z}_{ij21} \cdot \beta)\right) \\ & + y_{++22} \cdot \log\left(\sum_i \sum_j \exp(\mathbf{z}_{ij22} \cdot \beta)\right) \\ & + \sum_{i,j,k,l} \delta_{ijkl} \cdot (\mathbf{z}_{ijkl} \cdot \beta) \\ & - (N + \delta_{++++}) \cdot \log\left(\sum_{i,j,k,l} \exp(\mathbf{z}_{ijkl} \cdot \beta)\right) \quad (4) \end{aligned}$$

2.1 EM Algorithm

We maximize the posterior distribution given in (4) over parameter $\boldsymbol{\beta}$ by the generalized expectation maximization (GEM) algorithm (Dempster, Laird and Rubin 1977) with the following E and M steps.
E-step : Using augmented y_{ij12} given y_{i+12}, y_{ij21} given y_{+j21}, and y_{ij22} given y_{++22} for $i = 1, \cdots, I$ and $j = 1, \cdots, J$, the posterior (4) can be written as this augmented posterior distribution

$$\begin{aligned} l_{a.pos} \quad & \propto \textstyle\sum_i \sum_j (y_{ij11} + \delta_{ij11}) \log(\pi_{ij11}) \\ & + \textstyle\sum_i \sum_j (y_{ij12} + \delta_{ij12}) \log(\pi_{ij12}) \\ & + \textstyle\sum_i \sum_j (y_{ij21} + \delta_{ij21}) \log(\pi_{ij21}) \\ & + \textstyle\sum_i \sum_j (y_{ij22} + \delta_{ij22}) \log(\pi_{ij22}). \qquad (5) \end{aligned}$$

To determine the expected augmented log posterior of (5), we average over missing counts y_{ij12}, y_{ij21}, y_{ij22} conditioning on the current parameter estimates, π^{old}_{ijkl}, and observed counts y_{i+12}, y_{+j21}, and y_{++22}:

$$\begin{aligned} E_{old}[l_{a.pos}] \propto & \sum_i \sum_j (y_{ij11} + \delta_{ij11}) \log(\pi_{ij11}) \\ + & \sum_i \sum_j (E_{old}[y_{ij12} | \pi^{old}_{ijkl}, y_{i+12}] + \delta_{ij12}) \log(\pi_{ij12}) \\ + & \sum_i \sum_j (E_{old}[y_{ij21} | \pi^{old}_{ijkl}, y_{+j21}] + \delta_{ij21}) \log(\pi_{ij21}) \\ + & \sum_i \sum_j (E_{old}[y_{ij22} | \pi^{old}_{ijkl}, y_{++22}] + \delta_{ij22}) \log(\pi_{ij22}). \qquad (6) \end{aligned}$$

Since y_{ij12}, y_{ij21}, and y_{ij22} are multinomial random variates conditioned on marginal sum y_{i+12}, y_{+j21}, and y_{++22}, respectively, the conditional expectations in (6) are given by

$$\begin{aligned} E_{old}\big(y_{ij12} | \pi^{old}_{ijkl}, y_{i+12}\big) &= y_{i+12} \frac{\pi^{old}_{ij12}}{\pi^{old}_{i+12}} \\ &= y_{i+12} \frac{m^{old}_{ij12}}{m^{old}_{i+12}}, \\ E_{old}\big(y_{ij21} | \pi^{old}_{ijkl}, y_{+j21}\big) &= y_{+j21} \frac{m^{old}_{ij21}}{m^{old}_{+j21}}, \\ \text{and} & \\ E_{old}\big(y_{ij22} | \pi^{old}_{ijkl}, y_{++22}\big) &= y_{++22} \frac{m^{old}_{ij22}}{m^{old}_{++22}} \end{aligned}$$

M-step : In this step, we maximize the expected log posterior (6) using the pseudo observations $\tilde{y}_{ij11} = y_{ij11} + \delta_{ij11}$, $\tilde{y}_{ij12} = y_{i+12} \frac{m^{old}_{ij12}}{m^{old}_{i+12}} + \delta_{ij12}$, $\tilde{y}_{ij21} = y_{+j21} \frac{m^{old}_{ij21}}{m^{old}_{+j21}} + \delta_{ij21}$, and $\tilde{y}_{ij22} = y_{++22} \frac{m^{old}_{ij22}}{m^{old}_{++22}} + \delta_{ij22}$. 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

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

Then, the expected log posterior function 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 (Agresti 2002), 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 $\gamma^{(t)}_{ijkl} = \log m^{(t)}_{ijkl} + (y_{ijkl} - m^{(t)}_{ijkl}) / m^{(t)}_{ijkl}$ and $\hat{V}_t = [diag(\mathbf{m}^{(t)})]^{-1}$. We finally iterate these E and M-steps until a convergence criterion is achieved.

2.2 Five Types of Bayesian Methods

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 Clogg et al. (1991) did. Under this constraint (i.e., $\sum_{i,j,k,l} \delta_{ijkl} = p$), 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 nonresponse. That is, we determine δ_{ijkl} depending only on the respondent counts y_{ij11}, y_{i+12}, y_{+j21}, and y_{++22}. We call these priors respondent-driven priors and classify them into two types as below.

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

$$\delta_{ijkl} = \nabla_{kl} \frac{y_{ij11}}{y_{++11}}, \qquad (7)$$

where $\nabla_{kl} = p \cdot \frac{y_{++kl}}{y_{++++}}$ for $k = 1, 2$ and $l = 1, 2$.

On the other hand, the second type of respondent-driven priors gives no prior on π_{ij11}. That is, The second type of priors are the same as those of the first type except $\delta_{ij11} = 0$ for all i and j. In case of one-way contingency table (i.e., either X_1 or X_2 is fully observed without missing) and $y_{++22} = 0$, the first type is reduced to Park (1998), whereas the second type is reduced to Park and Brown (1994). These two types of respondent-driven priors may bring a controversy because the nonrespondents are usually assumed to have different response patterns from the respondents in the nonignorable model. For example, favored candidate attitude of nonrespondents could be different from that of respondents in a pre-election survey.

In order to reflect different response patterns between the respondent and nonrespondent, we propose the following third type of priors δ_{ijkl} depending on both respondent's and nonrespondent's information. So δ_{ijkl} is assigned to be proportional to expected cell frequencies, m_{ijkl}^{old}, where calculated at the previous iteration. We distinguish this third type of priors $\tilde{\delta}_{ijkl}$ from previous priors δ_{ijkl}:

$$\tilde{\delta}_{ijkl} = \begin{cases} \nabla_{11} \cdot \left(\frac{m_{ij11}^{old}}{m_{++11}^{old}} \right) & \text{for } k=1,\, l=1 \\ \nabla_{12} \cdot \left(\frac{m_{ij12}^{old}}{m_{++12}^{old}} + \frac{1}{I \cdot J} \right) \cdot \frac{1}{2} & \text{for } k=1,\, l=2 \\ \nabla_{21} \cdot \left(\frac{m_{ij21}^{old}}{m_{++21}^{old}} + \frac{1}{I \cdot J} \right) \cdot \frac{1}{2} & \text{for } k=2,\, l=1 \\ \nabla_{22} \cdot \left(\frac{m_{ij22}^{old}}{m_{++22}^{old}} + \frac{1}{I \cdot J} \right) \cdot \frac{1}{2} & \text{for } k=2,\, l=2. \end{cases}$$

where $\nabla_{kl} = p \cdot \frac{m_{++kl}^{old}}{m_{++++}^{old}}$ for $k = 1, 2$ and $l = 1, 2$.

Therefore, these new priors depend on their respective parameters m_{ijkl}^{old} to be estimated in previous iteration. The main reason we use a weighted priors of $m_{ijkl}^{old}/m_{++kl}^{old}$ and $1/IJ$ on $\tilde{\delta}_{ij12}$, $\tilde{\delta}_{ij21}$, and $\tilde{\delta}_{ij22}$ is to prevent a boundary solution on m_{ij12}, m_{ij21}, and m_{ij22}, respectively. In addition, pseudo observations in M-step are also changed: $\tilde{y}_{ij11} = y_{ij11} + \tilde{\delta}_{ij11}$, $y_{ij12} = y_{i+12}\frac{m_{ij12}^{old}}{m_{i+12}^{old}} + \tilde{\delta}_{ij12}$, $\tilde{y}_{ij21} = y_{+j21}\frac{m_{ij21}^{old}}{m_{+j21}^{old}} + \tilde{\delta}_{ij21}$, and $\tilde{y}_{ij22} = y_{++22}\frac{m_{ij22}^{old}}{m_{++22}^{old}} + \tilde{\delta}_{ij22}$. We also define the fourth type of priors by letting $\tilde{\delta}_{ij11} = 0$ in (8) as we obtained the second type from the first type.

The last type of priors extend the constant prior of Clogg et al. (1991) used for one-way categorical table to those for two-way categorical table as follows.

$$\tilde{\delta}_{ijkl} = \begin{cases} 0 & \text{if } k=1,\, l=1 \\ \frac{p}{3} \cdot \left(\frac{1}{I \cdot J} \right) & \text{for } k \neq 1 \text{ or } l \neq 2. \end{cases} \qquad (9)$$

These five types of priors will be compared in the subsequent two sections using empirical data and simulation studies.

3. Case Studies

In a sample survey for forecasting election outcomes, the accuracy of election forecasting polls often depends on how to handle undecide voters who are likely to vote but not yet decided their preference of candidate. We compare the five Bayesian methods with the maximum likelihood estimate (ML) through the Buckeye State Poll (BSP) data conducted in 1998 by the Center for Survey Research at the Ohio State University. As competitors, we also consider two other nonresponse models (i.e., another ignorable model and another nonignorable model).

The BSP pre-election surveys conducted in October 1998 produced two-way categorical tables with one category being candidate preference and the other category being likelihood of voting for the November 1998 races of Ohio Governor, Attorney-General, and Treasurer. The BSP survey in April 1999 gives the poll data of three candidates preference for the election of Mayor in May 1999. Table (1) summarizes these four polls and shows substantial number of undecided voters.

(8) For comparison, we consider the following one ignorable and two nonignorable nonresponse models.

$$\begin{aligned} &\text{Model 1}: \log(m_{ijkl}) = \beta_0 + \beta_{X_1}^i + \beta_{X_2}^j \\ &\qquad + \beta_{R_1}^k + \beta_{R_2}^l + \beta_{X_1X_2}^{ij} + \beta_{R_1R_2}^{kl}, \\ &\text{Model 2}: \log(m_{ijkl}) = \beta_0 + \beta_{X_1}^i + \beta_{X_2}^j \\ &+ \beta_{R_1}^k + \beta_{R_2}^l + \beta_{X_1R_1}^{ik} + \beta_{X_2R_2}^{jl} + \beta_{X_1X_2}^{ij} + \beta_{R_1R_2}^{kl}, \\ &\text{Model 3}: \log(m_{ijkl}) = \beta_0 + \beta_{X_1}^i + \beta_{X_2}^j \\ &+ \beta_{R_1}^k + \beta_{R_2}^l + \beta_{X_1R_2}^{il} + \beta_{X_2R_1}^{jk} + \beta_{X_1X_2}^{ij} + \beta_{R_1R_2}^{kl}. \end{aligned} \qquad (10)$$

Model 1 in (10) is missing completely at random, and cases with missing data can be ignorable in likelihood inferences. Model 2 and Model 3 are nonignorable in which the probability of missing on a variable depends on itself in Model 2, whereas the probability in Model 3 depends on the other variable which can be also not observed. Note that the ML estimates in Model 1 and Model 3 are not on the boundary of parameter space as shown by Baker, Rosenberger, and Dersimonian (1992). Moreover, since the five types of Bayesian estimates given in the previous section are almost same as the ML estimates under Model 1 and Model 3, we only present the ML estimates for these two models.

Denote the ML estimates under Model 1, Model 2, and Model 3 by IG_{ML}, $NIG1_{ML}$, and $NIG2_{ML}$, respectively. We also let $NIG1_{BE1}$, $NIG1_{BE2}$, and $NIG1_{BE3}$ be the Bayesian estimates with priors δ_{ijkl} depending on parameter m_{ijkl} given by (8), with the same priors as $NIG1_{BE1}$ except $\delta_{ij11} = 0$, and with the constant priors given by (9), respectively. Finally, let $NIG1_{BE4}$ and $NIG1_{BE5}$ be the

Table 1: 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

empirical Bayesian estimates with the respondent-driven priors given by (7) where $NIG1_{BE5}$ has the same priors as $NIG1_{BE4}$ except $\delta_{ij11} = 0$. All of these $NIG1_{BE1}$, $NIG1_{BE2}$, $NIG1_{BE3}$, $NIG1_{BE4}$, and $NIG1_{BE5}$ are obtained under Model 2 given in (10).

The first table in Table 2 shows the predictions of elections using only "likely to vote" for the four races. This table also includes the actual election results and the occurrence of boundary solutions in ML estimates. The second 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 vote. Comparing the two tables, we may conclude that the winners for Governor, Attorney-General, and Treasurer's elections are predicted to be unchanged regardless of likelihood of voting, whereas the winner could be changed in the Mayor's election if most of those who are "unlikely to vote" actually vote.

Based on Table 2, we can classify 8 estimates into two groups: $NIG1_{ML}$, $NIG1_{BE1}$, $NIG1_{BE2}$, and $NIG1_{BE3}$ consist of one group and the remaining estimates consists of one group. In the first group, $NIG1_{BE1}$ is between $NIG1_{ML}$ and $NIG1_{BE2}$ in which NIG_{BE2} is almost the same as $NIG1_{BE3}$. All estimates in the second group are almost identical although $NIG1_{BE4}$ is barely different from the second group. As expected, since the priors δ_{ijkl} for NIG_{BE4} and NIG_{BE5} are defined so that the estimate of m_{ijkl} shrinks toward the ML under ignorable nonresponse, these two empirical Bayesian estimates are almost the same as the IG_{ML} and hence have little advantage over the IG_{ML}. It is also interesting that $NIG2_{ML}$ is exactly the same as IG_{ML} although their loglinear models are differently specified.

There is no general criterion 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 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 for the standards of the interview. In this regard, the $NIG1_{BE4}$, $NIG1_{BE5}$, and $NIG2_{ML}$ may not be appropriate because they are almost the same as the IG_{ML} obtained from Model 1 which is an ignorable nonresponse model.

Compared to actual election results, $NIG1_{ML}$ gives the worst prediction in Governor, Mayor, and Attorney-General's elections because the $NIG1_{ML}$ lies on boundary solutions, whereas the $NIG1_{ML}$ provides the best prediction in the Treasurer' election because it does not lie on a boundary solution. In Attorney-General's election, the $NIG1_{BE1}$ not only predicts the exact actual result but also is quite different from the other estimates. Since the $NIG1_{BE1}$ has the priors to reflect different response pattern between respondent and nonrespondent, we can infer that the nonrespondents who is likely to vote but not yet decided their preference of candidate have quite different preference of candidate from the observed preference of candidate by respondents (i.e., 197 (Montgomery) : 82 (Cordray) for the respondents likely to vote, while 13 (Montgomery) : 54 (cordray) for the nonrespondents likely to vote).

4. Concluding Remarks

We investigated Bayesian analysis for incomplete two-way categorical tables with nonignorable nonresponse under which the maximum likelihood estimates often fall in the boundary solution, causing the ML estimates unstable. To avoid the boundary solution problem, we proposed the five types of Bayesian methods. These Bayesian methods include the previous Bayesian models as special cases. The two among the five Bayesian models were proposed to reflect different response patterns between respondents and nonrespondents.

Table 2: Prediction of elections based on October 98 and April 99 Buckeye State Poll

	Governor			Mayor			Attorney-General		Treasurer	
	Fisher	Taft	Others	Coleman	Teater	Espy	Mongomery	Cordray	Deters	Donofrio
					Likely to vote only used					
$NIG1_{ML}$	33%	42%	25%	31%	25%	43%	76%	24%	57%	43%
$NIG1_{BE1}$	37%	45%	17%	36%	33%	30%	63%	37%	53%	47%
$NIG1_{BE2}$	39%	47%	14%	38%	33%	28%	66%	34%	51%	49%
$NIG1_{BE3}$	39%	47%	14%	38%	34%	29%	66%	34%	52%	48%
$NIG1_{BE4}$	41%	49%	11%	38%	34%	28%	72%	28%	53%	47%
$NIG1_{BE5}$	41%	51%	8%	40%	34%	26%	71%	29%	52%	48%
IG_{ML}	41%	51%	8%	41%	33%	26%	71%	29%	52%	48%
$NIG2_{ML}$	41%	51%	8%	41%	33%	26%	71%	29%	52%	48%
Actual result	45%	50%	5%	39%	37%	24%	63%	37%	57%	43%
Boundary		yes			yes		yes		no	
					Likely to vote + Unlikely to vote					
$NIG1_{ML}$	33%	39%	28%	25%	26%	49%	77%	23%	60%	40%
$NIG1_{BE1}$	37%	43%	20%	29%	36%	36%	61%	39%	56%	44%
$NIG1_{BE2}$	39%	45%	16%	31%	36%	33%	65%	35%	55%	45%
$NIG1_{BE3}$	39%	45%	16%	31%	36%	33%	65%	35%	55%	45%
$NIG1_{BE4}$	41%	46%	12%	31%	37%	32%	73%	27%	56%	44%
$NIG1_{BE5}$	42%	49%	9%	33%	37%	31%	71%	29%	55%	45%
IG_{ML}	41%	50%	9%	34%	36%	30%	71%	29%	55%	45%
$NIG2_{ML}$	41%	50%	9%	34%	36%	30%	71%	29%	55%	45%

Data analysis showed that these new Bayesian methods were more reasonable in the sense that nonignorable nonrespose mechanisms are more reflected and close to the actual results.

In this paper, we assumed an incomplete two-way categorical table without covariates. Therefore, our work can be extended to a time series of incomplete multi-way categorical tables with some useful covariates. All of these are our future works.

References

Agresti, A. (2002), *Categorical Data Analysis*, second edition. Wiley, New York.

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," unpublished paper, Ohio state university.

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.

Little, J. A. and Rubin, D. B. (2002), *Statistical analysis with missing data*, second edition. Wiley, New York.

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.

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.

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.