Bayesian Analysis of Nonignorable Missing Categorical Data: An Application to Bone Mineral Density and Family Income

Balgobin Nandram
Department of Mathematical Sciences, Worcester Polytechnic Institute
100 Institute road, Worcester MA 01609

and

Lawrence H. Cox and Jai Won Choi
Office of research and Methodology, National Center for Health Statistics
6525 Toledo Road, Hyattsville, MD 20782

December 28, 2002

1. Introduction

It is a common practice to use two-way categorical tables to present survey data. For many surveys there are missing data and this gives rise to partial classification of the sampled individuals. Thus, for the two-way table there are both item nonresponse (one of the two categories is missing) and unit nonresponse (both categories are missing). One may not know how the data are missing, and a model that includes some difference between the observed data and missing data (i.e., nonignorable missing data) may be preferred. For a general $r \times c$ categorical table we address two issues (a) an estimation of the cell probabilities of the two-way table and (b) a test of no association between the two categories using the Bayes factor. Both problems are important because with a substantial number of nonrespondents an analysis based on only the complete observed data may be misleading.

Our application is in health statistics, and there are several problems at the National Center for Health Statistics (NCHS) in which our methodology will be important. The NCHS categorizes the sampled persons by two types of attributes, and researchers analyze such categorical tables for goodness of fit or independence. However, only partial classification of the individuals is available because some individuals are classified by only one attribute while others are not classified. We use data from the third National Health and Nutrition Examination Survey (NHANES III) to study the relation between bone mineral density (BMD) and family income (FI). While FI is a discrete variable, BMD is not; BMD is classified into three levels: normal, osteopenia and osteoporosis, and FI into three levels: low, medium and high. About 62% of the households have both FI and BMD observed, 8% with only BMD observed, 29% with only income observed, 1% with neither income nor BMD among those participated in the examination stage. The data set, we used in our study are presented in Table ?? in a 3×3 categorical table of BMD and FI. Our problem is to (a) estimate the cell probabilities and (b) test for association between BMD and FI.

Nandram, Han and Choi (2002) analyze multinomial data with one attribute using a nonignorable nonresponse model. Nandram and Choi (2002 a,b) use an expansion model to study nonignorable nonresponse binary data. The expansion model, a nonignorable nonresponse model, degenerates into an ignorable nonresponse model (in the spirit of Draper 1995), allowing an expression of uncertainty about ignorability; see also Forster and Smith (1998). In this paper, for nonignorable nonresponse we attempt a related methodology, but the issues for a two-way categorical table are more complex.

Rao and Scott (1981) investigate the effects of stratification and clustering on the asymptotic distribution of Pearson's chi-squared statistic for goodness of fit and independence. They propose new measures called generalized design effects. See also Rao and Scott (1984) who generalized the results of Rao and Scott (1981) to multi-way categorical tables. The works of Cohen (1976), Altham (1976), Brier (1978) and Choi and McHugh (1989) are also relevant.

Chen and Fienberg (1974) describe the two issues we are discussing in this paper. For the two-way categorical tables they can tolerate item nonresponse only; unit nonresponse is excluded from their analysis. However, they assume that the data are missing at random (i.e., simple random sampling), and they show how to obtain maximum likelihood estimators under multinomial and Poisson sampling schemes. See Little and Rubin (1987) for a simple illustration of the EM algorithm for the problem in which exactly one attribute is missing (ignorable). Little and Rubin (1987) also discussed the nonignorable case for this problem (see also Little 1985 for a discussion of the case in which there are both supplemental margins). It is noted in Little and Rubin (1987) that the issue of the nonignorable model for this problem is that there are too many parameters, and therefore many parameters are not identified so they resorted to hierarchical log-linear models. Finally, Chen and Fienberg (1974) show how to adjust the chi-squared and the likelihood ratio statistics for the partially classified data, an issue that has been discussed earlier by Rao and Scott (1981) and others.

More recently Wang (2001) considers a problem with the same two issues we want to investigate. However, simple random sampling and stratified random sampling (ignorable missing data mechanism) is used and unit nonresponse is not included in the discussion. They consider marginal imputation and conditional imputation. In marginal imputation the marginal probabilities are estimated using the completers, and the cell probabilities are imputed. In conditional imputation the conditional probability of each cell is estimated given the margins for the completers. He obtained asymptotic distributions for the estimated cell probabilities and an adjustment (in the spirit of Rao and Scott 1981). Finally, we note that Greene et al. (2002) describe a simple raking method for imputing the cell counts in a two-way table with missing data; our methodology goes far beyond these authors.

Our methodology differs from those of Chen and Fienberg (1974) and Wang (2001) in several ways. The major difference is that we use a Bayesian approach. This permits us to (a) use a method that does not rely on asymptotic theory, (b) incorporate nonignorable missingness into the modeling and (c) obtain an alternative to Pearson's chi-squared statistic for testing for no association. A Bayesian method permits modeling different patterns of missingness under two different assumptions (i.e., ignorable and nonignorable missingness). Our idea is to start with an ignorable model, which is then expanded into a nonignorable model. Note that we also include unit nonresponse in our modeling which the other researchers can do as a separate problem using weighting adjustment (e.g., see discussion in Kalton and Kasprzyk 1986). However, there can be nonignorability here as well, and one would need to include unit and item nonresponses simultaneously.

In (c) we use the Bayes factor (see Kass and Raftery 1995) to quantify the difference between a model with association and one without. This is the ratio of the prior odds of one model to the other to their posterior odds (obtained through the use of Bayes' theorem). This is the same as the ratio of the marginal likelihoods of the data under two models, one without association and the other with association. See Kass and Raftery (1995) for a rule of thumb for quantifying the degree of evidence. There are several methods to compute the marginal likelihood (e.g., see Section 1 of Chib and Jeliazkov 2001), and we note that one standard method is importance sampling.

In this paper, we introduce a Bayesian method to analyze data from an $r \times c$ categorical table when there are both item nonresponse and unit nonresponse, and missing data mechanism can be nonignorable. In Section 2, we describe the methodology to obtain estimates of the cell probabilities incorporating the three types of missing data, and we show how to expand an ignorable nonresponse model into a nonignorable nonresponse model. We also show how to use Markov chain Monte Carlo methods to fit the models. In Section 3, we use the Bayes factor to test for association of the two attributes and assess whether the ignorable or nonignorable model prevails. We show how to use importance sampling to compute the marginal likelihoods under different models. In Section 4, we analyze the NCHS data to demonstrate our methods. Finally, Section 5 has concluding remarks.

2. Methodology for Nonignorable Nonresponse

We construct a nonignorable nonresponse model by expanding an ignorable nonresponse model. We show how to fit both models using the Gibbs sampler.

2.1 Nonignorable Nonresponse Model

Let $I_{jk\ell},\ j=1,\ldots,r,\ k=1,\ldots,c,\ \ell=1,\ldots,n$ denote the characteristic (observed or missing) of an individual in the two-way table (i.e., the row and column the individual belongs to). Let $\underset{\sim}{J}_{jk\ell}=(1,0,0,0),\ (0,1,0,0),\ (0,0,1,0)$, or $(0,0,0,1)$ which of the cases ((a), (b), (c) and (d)) the ℓ^{th} individual belongs (e.g., $\underset{\sim}{r}_{jk\ell}=(1,0,0,0)$ indicates that the individual belongs to the completely observed table). We use the almost nonparametric assumption that there is multinomial sampling in the survey. Thus, for nonignorable missing data we take

$$\underset{\sim}{J}_{jk\ell}\mid\{I_{jk\ell}=1,I_{j'k'}=0,j\neq j',k\neq k',\underset{\sim}{\pi}_{jk}\}\overset{iid}{\sim}\text{Multinomial}\{1,\underset{\sim}{\pi}_{jk}\} \tag{1}$$

and

$$\underset{\sim}{I}_{\ell}\mid\underset{\sim}{p}\overset{iid}{\sim}\text{Multinomial}\{1,\ \underset{\sim}{p}\}, \tag{2}$$

where $\sum_{s=1}^{4}\pi_{sjk}=1,\ \pi_{sjk}\geq 0,\ j=1,\ldots,r,\ k=1,\ldots,c$ and $\sum_{j=1}^{r}\sum_{j=1}^{c}p_{jk}=1,\ p_{jk}\geq 0,\ j=1,\ldots,r,\ k=1,\ldots,c$. Assumption (1) specifies that the cases (a), (b), (c) or (d) an individual belongs to depends on the two characteristics of the individual. In this manner we incorporate the assumption that the missing data is nonignorable. Also, assumption (2) is standard when there are no missing data.

Next, we need the likelihood function. Let the cell counts be $y_{sjk}=\sum_{\ell=1}^{n}I_{jk\ell}J_{sjk\ell},\ s=1,2,3,4$ for the four cases. Here y_{1jk} are observed and $y_{sjk},\ s=2,3,4$ are missing. For y_{1jk} we know that $\sum_{j=1}^{r}\sum_{k=1}^{c}y_{1jk}=t$, the number of individuals with complete data. For y_{2jk} we know that $\sum_{k=1}^{c}y_{2jk}=u_j$, where $u_j,\ j=1,\ldots,r$ are observed. For y_{3jk} we know that $\sum_{j=1}^{r}y_{3jk}=v_k$, where $v_k,\ k=1,\ldots,c$ are observed. For y_{4jk} we know that $\sum_{j=1}^{r}\sum_{k=1}^{c}y_{4jk}=w$. In this analysis $\underset{\sim}{u}$, $\underset{\sim}{v}$ and w are held fixed (i.e., fixed margin analysis). Then, the augmented likelihood function for $\underset{\sim}{p},\underset{\sim}{\pi},\underset{\sim}{y}_s,\ s=2,3,4\mid\underset{\sim}{y}_1$ is

$$\begin{aligned} g(\underset{\sim}{p},\underset{\sim}{\pi},\underset{\sim}{y}_s,\ s=2,3,4\mid\underset{\sim}{y}_1,\underset{\sim}{u},\underset{\sim}{v},w) &\propto \prod_{j=1}^{r}\prod_{k=1}^{c}\prod_{s=1}^{4}\left\{\frac{(\pi_{sjk}p_{jk})^{y_{sjk}}}{y_{sjk}!}\right\} \\ &= \left[\prod_{j=1}^{r}\prod_{k=1}^{c}\prod_{s=1}^{4}\frac{\pi_{sjk}^{y_{sjk}}}{y_{sjk}!}\right]\left[\prod_{j=1}^{r}\prod_{k=1}^{c}p_{jk}^{\sum_{s=1}^{4}y_{sjk}}\right] \end{aligned} \tag{3}$$

subject to $\sum_{k=1}^{c}y_{2jk}=u_j, j=1,\ldots,r$, $\sum_{j=1}^{r}y_{3jk}=v_k, k=1,\ldots,c$, and $\sum_{j=1}^{r}\sum_{k=1}^{c}y_{4jk}=w$.

Observe that in (3) the parameters p_{jk} and π_{sjk} are not identifiable. Clearly, to estimate p_{jk} one needs to know $\sum_{s=1}^{4}y_{sjk}$ but only the y_{1jk} are known. Also, to estimate π_{sjk} one needs to know

$\sum_{s=1}^4 y_{sjk}$. Thus, $y_{sjk},\ s = 2, 3, 4$ are also not identifiable. Putting very informative proper priors on the π_{sjk} will help, but this is not a practical solution. If an ignorable model (i.e., $\pi_{sjk} = \pi_s$) is used, then all the parameters can be identified. Therefore, a sensible solution is to attempt to link the $\underset{\sim}{\pi}_{jk}$ using a common feature over (j, k). If the $\underset{\sim}{\pi}_{jk}$ come from a common distribution with "known" parameters, we would be able to estimate them. That is, we must attempt to "borrow strength" as in small area estimation. This permits estimation of the $y_{sjk},\ s = 2, 3, 4$ which, in turn, will facilitate estimation of the p_{jk}.

For the p_{jk} we take

$$\underset{\sim}{p} \sim \text{Dirichlet}(1, \dots, 1) \tag{4}$$

(i.e., a uniform prior density on $\underset{\sim}{p}$), and for the π_{jk} we consider "centering" the nonignorable model on the ignorable model which has $\underset{\sim}{\pi}_{jk} = \underset{\sim}{\pi}, j = 1, \dots, r,\ k = 1, \dots, c$. We assume that

$$\underset{\sim}{\pi}_{jk} \mid \underset{\sim}{\mu}, \tau \overset{iid}{\sim} \text{ Dirichlet } (\mu_1\tau,\ \mu_2\tau,\ \mu_3\tau,\ \mu_4\tau), \tau \geq 0,\quad \sum_{s=1}^4 \mu_s,\quad \mu_s \geq 0,\ s = 1, 2, 3, 4. \tag{5}$$

In (5) the parameter τ tells us about the closeness of the nonignorable model to the ignorable model. For example, if τ is small, the $\underset{\sim}{\pi}_{jk}$ will be very different, and if τ is large, the $\underset{\sim}{\pi}_{jk}$ will be very similar. Thus, inference will be sensitive to the choice of τ, and one has to be careful in choosing τ. For large τ, the nonignorable model is kept at the ignorable model in (5).

A priori we take

$$p(\underset{\sim}{\mu}) = 1 \text{ and } \tau \sim \text{Gamma}(\alpha_0, \beta_0), \tag{6}$$

where α_0 and β_0 are to be specified.

Then combining (4), (5) and (6), the joint prior density of $\underset{\sim}{\pi}$, $\underset{\sim}{p}$, $\underset{\sim}{\mu}$ and τ is

$$\pi(\underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{\mu}, \tau) \propto \left\{ \prod_{j=1}^r \prod_{k=1}^c \frac{\prod_{s=1}^4 \pi_{sjk}^{\mu_s\tau - 1}}{D(\mu_1\tau, \dots, \mu_4\tau)} \right\} \tau^{\alpha_0 - 1} e^{-\beta_0\tau}. \tag{7}$$

Note that (7) is a proper prior density. Finally, combining the likelihood function in (3) with the joint prior density in (7) via Bayes theorem, the joint posterior density of the parameters $\underset{\sim}{\pi}$, $\underset{\sim}{p}$ and the latent variables $\underset{\sim}{y}_{(1)} = \underset{\sim}{y}_2, \underset{\sim}{y}_3, \underset{\sim}{y}_4$ (e.g., $\underset{\sim}{y}_2 = \{y_{2jk},\ j = 1, \dots, r, k = 1, \dots), c\}$ is

$$\pi(\underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{\mu}, \tau \mid \underset{\sim}{y}_1) \propto \left[\prod_{j=1}^r \prod_{k=1}^c \prod_{s=1}^4 \left\{ \frac{(\pi_{sjk} p_{jk})^{y_{sjk}}}{y_{sjk}!} \right\} \right] \left\{ \prod_{j=1}^r \prod_{k=1}^c \frac{\prod_{s=1}^4 \pi_{sjk}^{\mu_s\tau - 1}}{D(\mu_1\tau, \dots, \mu_4\tau)} \right\} \tau^{\alpha_0 - 1} e^{-\beta_0\tau}, \tag{8}$$

where $D(\mu_1\tau, \dots, \mu_4\tau) = \{\prod_{s=1}^4 \Gamma(\mu_s\tau)\}/\Gamma(\tau)$ is the Dirichlet function.

2.2 Fitting the Models

We use the Gibbs sampler to obtain iterates from the joint posterior density in order to make inference about the parameters. Specifically, we need to make inference about $\underset{\sim}{p}$, π_{jk}, any of the missing cell counts, and the relationship between the two categorical variables. We need the conditional posterior density of each of the parameters given the others.

For $\underset{\sim}{p}$ we have $\underset{\sim}{p} \mid \{\underset{\sim}{y}_s, s = 1, 2, 3, 4, \underset{\sim}{u}, \underset{\sim}{v}, w\} \sim \text{Dirichlet}(\sum_{s=1}^4 y_{s11} + 1, \dots, \sum_{s=1}^4 y_{src} + 1)$. For $\underset{\sim}{\pi}_{jk}$ we have $\underset{\sim}{\pi}_{jk} \mid \{\underset{\sim}{\mu}, \tau, \underset{\sim}{y}_s, s = 1, 2, 3, 4, \underset{\sim}{u}, \underset{\sim}{v}, w\} \overset{ind}{\sim} \text{Dirichlet}(y_{1jk} + \mu_1\tau, y_{2jk} + \mu_2\tau, y_{3jk} + \mu_3\tau, y_{4jk} + \mu_4\tau)$. We need the conditional posterior probability mass functions of $\underset{\sim}{y}_s,\ s = 2, 3, 4$ given $\underset{\sim}{y}_1$, $\underset{\sim}{p}$, $\underset{\sim}{\pi}_{jk}, j = 1, \dots, r,\ k = 1, \dots, c$. From (8) it is clear that under the conditional posterior density $\underset{\sim}{y}_s,\ s = 2, 3, 4$ are independent multinomial random vectors with parameters $(u_j,\ \underset{\sim}{q}_j^{(2)}), (v_k,\ \underset{\sim}{q}_k^{(3)}), and (w,\ \underset{\sim}{q}^{(4)})$, respectively where $\underset{\sim}{q}_j^{(2)}, \underset{\sim}{q}_k^{(3)}$, and$\underset{\sim}{q}^{(4)}$ are weighted means: where $q_{jk}^{(2)} = \pi_{2jk}p_{jk} / \sum_{k'=1}^{c} \pi_{2jk'}p_{jk'}$, $q_{jk}^{(3)} = \pi_{3jk}p_{jk} / \sum_{j'=1}^{r} \pi_{3j'k}p_{j'k}$, and $q_{jk}^{(4)} = \pi_{4jk}p_{jk} / \sum_{j'=1}^{r} \sum_{k'=1}^{c} \pi_{4j'k'}p_{j'k'}$.

For the ignorable nonresponse model we only need $\mu_s = \frac{1}{4}$, $s = 1, 2, 3, 4$, $\tau = 4$ for above $\underset{\sim}{p}, \underset{\sim}{\pi}_{jk}$, and $\underset{\sim}{y}_s$, $s = 2, 3, 4$.

The joint conditional posterior density $p(\underset{\sim}{\mu}, \ \tau \mid \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c)$ of $\underset{\sim}{\mu}$, τ is

$$p(\underset{\sim}{\mu}, \ \tau \mid \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c) \propto \frac{\prod_{s=1}^{4} \delta_s^{\mu_s \tau}}{\{D(\underset{\sim}{\mu}\tau)\}^{rc}} \tau^{\alpha_0 - 1} e^{-\beta_0 \tau},$$

$\sum_{s=1}^{4} \mu_s = 1$, $\mu_s \geq 0$, $s = 1, 2, 3, 4$, $\tau > 0$ where $\delta_s = \prod_{j=1}^{r} \prod_{k=1}^{c} \pi_{sjk}$ and $D(\underset{\sim}{\mu}\tau) = \{\prod_{s=1}^{4} \Gamma(\mu_s \tau)\} / \Gamma(\tau)$ is the Dirichlet function.

We do not need to get a sample directly from $p(\underset{\sim}{\mu} \mid \tau, \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c)$. But, letting $\underset{\sim}{\mu}_{(s)}$ denote the vector of all components of $\underset{\sim}{\mu}$ except μ_s, we have $p(\mu_s \mid \underset{\sim}{\mu}_{(s)}, \tau, \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c) \propto \delta_s^{\mu_s \tau} / \{\Gamma(\mu_s \tau)\}^{rc} \quad \delta_4^{(1-\mu_1-\mu_2-\mu_3)\tau} / \{\Gamma((1 - \mu_1 - \mu_2 - \mu_3)\tau)\}^{rc}$, where $0 \leq \mu_s \leq 1 - \sum_{s'=1, \ s' \neq s}^{3} \mu_{s'}$, $s = 1, 2, 3$.

We use a grid to draw a sample from $p(\mu_s \mid \underset{\sim}{\mu}_{(s)}, \tau, \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c)$. We have used 50 grids (i.e., we have divided the range of μ_s, $(0, \ 1 - \sum_{s'=1, \ s' \neq s}^{3} \mu_{s'})$, into 50 intervals of equal widths) to form the probability mass function of μ_s, $s = 1, 2, 3$. To draw a random deviate, we first draw a random variable from this probability mass function, and this indicates which of the 50 intervals is selected. Then, we obtain the random deviate for μ_s by drawing a uniform deviate in this interval. This procedure is efficient because μ_s is bounded, it does not lie close to 0 or 1, the intervals are very narrow, and it is very "cheap" to construct the discrete probability mass function for each μ_s, $s = 1, 2, 3$. Finally, μ_4 is obtained from its conditional posterior density by taking $\mu_4 = \sum_{s=1}^{3} \mu_s$.

The conditional posterior density of τ is $p(\tau \mid \underset{\sim}{\mu}, \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c)$

$$\propto \left[\prod_{s=1}^{4} \frac{\delta_s^{\mu_s \tau}}{\{\Gamma(\mu_s \tau)\}^{rc}}\right] \tau^{\alpha_0 - 1} e^{-\beta_0 \tau}, \quad \tau > 0.$$

To draw a random deviate from above, we proceed in the same manner as for μ except that we transform τ from the positive half of the real line to $(0, \ 1)$. It is more convenient to perform a grid in a bounded interval. Thus, letting $\tau = \phi/(1 - \phi)$ in (??), we have $p(\phi \mid \underset{\sim}{\mu}, \underset{\sim}{\pi}_{jk}, j = 1, \ldots, r, \ k = 1, \ldots, c) \propto \frac{1}{(1-\phi)^2} \left\{ \left[\prod_{s=1}^{4} \frac{\delta_s^{\mu_s \tau}}{\{\Gamma(\mu_s \tau)\}^{rc}}\right] \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \right\}_{\tau = \frac{\phi}{1-\phi}}$, $0 < \phi < 1$. We use 50 intervals of equal width to draw ϕ, and the random deviate for τ is $\phi/(1 - \phi)$.

The Gibbs sampler is executed by drawing a random deviate from each of $\underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{y}, \underset{\sim}{mu}$, and$\tau$ iterating the entire procedure until convergence.

Finally, we describe how to specify α_0 and β_0. We have obtained iterates for the missing cell counts from the Gibbs sampler for the ignorable case (i.e., $\underset{\sim}{\pi}_{jk} = \underset{\sim}{\pi}$) which we denote by $n_{sjk}^{(h)}$, $h = 1, \ldots, M = 1000$. For each h we fit the model $\underset{\sim}{p} \sim$ Dirichlet($\underset{\sim}{1}$) and $\underset{\sim}{\pi}_{jk} \overset{iid}{\sim}$ Dirichlet($\underset{\sim}{\alpha}$) where $\alpha_s = \mu_s \tau$, $s = 1, \ 2, \ 3, \ 4$ to obtain the likelihood function $\prod_{j=1}^{r} \prod_{k=1}^{c} \left[\frac{\Gamma(\sum_{s=1}^{4} \alpha_s)}{\Gamma(\sum_{s=1}^{4} (\alpha_s + n_{sjk}^{(h)})} \prod_{s=1}^{4} \frac{\Gamma(\alpha_s + n_{sjk}^{(h)})}{\Gamma(\alpha_s)} \right]$, $\alpha_s > 0$, $s = 1, \ 2, \ 3, \ 4$. Using the Nelder-Mead algorithm to maximize the likelihood function over

$\alpha_s > 0,\ s = 1,\ 2,\ 3,\ 4$ at the h^{th} iterate, we obtain $\underset{\sim}{\alpha}^{(h)},\ h = 1, \ldots, M$. Now letting $\tau^{(h)} = \sum_{s=1}^{4} \alpha_s^{(h)}$, we fit the prior, Gamma(α_0, β_0), to $\tau^{(h)}, h = 1, \ldots, M$ taking $\alpha_0 = a^2/b$ and $\beta_0 = a/b$ where $a = M^{-1} \sum_{h=1}^{M} \tau^{(h)}$ and $b = (M-1)^{-1} \sum_{h=1}^{M} (\tau^{(h)} - a)^2$.

3. Bayes Factor: Tests of Association and Nonignorability

We construct two tests for (a) the presence of nonignorability and (b) association between BMD and FI. These tests apply to any $r \times c$ table. The test in (a) is an assessment of whether the ignorable model or the nonignorable model prevails, and the test in (b) is assessment of the assumption that $p_{jk} = q_{1j} q_{2k},\ j = 1, \ldots, r,\ k = 1, \ldots, c$, where $\sum_{j=1}^{r} q_{1j} = 1$ and $\sum_{k=1}^{c} q_{2k} = 1$. We use the Bayes factor, the ratio of the marginal likelihoods under two scenarios (e.g., association versus no association). Let $\underset{\sim}{y}_s = (y_{s11}, \ldots, y_{src})$, s=1,...,4, $\underset{\sim}{y} = (\underset{\sim}{y}_s, s = 1, 2, 3, 4) = (\underset{\sim}{y}_1, \underset{\sim}{y}_{(1)})$. Note that we observe $\underset{\sim}{y}_1$, but $\underset{\sim}{y}_{(1)}$ is a set of latent variables. So each marginal likelihood is simply the probability that $\underset{\sim}{y}_1$ is the observed value of $\underset{\sim}{Y}_1$, which we denote by $p(\underset{\sim}{y}_1)$. Let the marginal likelihood for the ignorable model be $p_{ig}(\underset{\sim}{y}_1) = a$ under association and $p_{ig}(\underset{\sim}{y}_1) = b$ under no association, and the marginal likelihood for the nonignorable model be $p_{nig}(\underset{\sim}{y}_1) = A$ under association and $p_{nig}(\underset{\sim}{y}_1) = B$ under no association. the marginal likelihoods, a and b for ignorable model and A and B for nonignorable model are obtained by Markov Chain Monte Carlo method.

The resemblance of the ignorable and noningnorable models depends on how close A or B to one under association and no association respectably. Bayes Factor for association is a/b under ignorable model and A/B under nonignorable model. If Bayes Factor is greater than 6, we may reject the idea that the two attributes of the 2-way table are not associated.

4. Data and Empirical Analysis

We apply our methodology to the data in the 3×3 categorical table (not included). We present results associated with the observed data. We do not include any table related to the analysis due to space limitation.

We compared the ignorable nonresponse model to the nonignorable nonresponse model. The numerical standard errors (NSE) are small indicating that the computations are repeatable. The posterior means (PM) are very similar for the two models. The posterior standard deviations (PSD) are larger for the nonignorable model, making the 95% credible intervals wider.

We also compare the estimation of π_s in the ignorable nonresponse model to π_{sjk} in the nonignorable nonresponse model in which we present the range of the values for the nine cells of each of $s = 1, 2, 3, 4$ of the posterior means (PM). The range indicates the extent of the nonignorability. The PM's of π_s are within the range of the π_{sjk} and as expected the PSD's are larger for the nonignorable model.

We have presented the Bayes factors for testing the goodness of fit of the ignorable model and the nonignorable model. We note that the estimates of A and B are about the same with a NSE of roughly 4 in each case. The differences are small indicating little preference. There are very little evidence for any preference between the ignorable and nonignorable nonresponse models. It is also true that there is very little evidence for association between BMD and FI.

We have studied the sensitivity of inference about the p_{jk} with respect to the prior distribution of τ. That is, we have taken $\tau \sim$ Gamma$(\kappa\alpha_0,\ \beta_0)$, where κ is a sensitivity parameter that we have taken to be 1 in our analysis above. Making κ bigger than 1 induces no changes in the posterior (PM) and posterior standard deviation (PSD) of the p_{jk} because $\alpha_0 = 698$ and $\beta_0 = 1.07$. We calculate PM's and PSD's for $\kappa = .10,\ .25,\ .50,\ .75,\ 1.00$. The changes in the PM's are negligible (i.e., the posterior means of the p_{jk} are not sensitive to changes in κ). There is some changes in the PSD's: For values of $\kappa \geq .25$ the changes are not important. At $\kappa = .10$ the changes are substantial, but there is a huge difference between $\kappa = .10$ and $\kappa = 1.00$.

We have also studied the sensitivity of the Bayes factors at $\kappa = .10, .25, .50, .75$. For the nonignorable model the marginal log-likelihoods are $-75.58, -72.00, -70.86, -70.49$ for with association and $-67.11, -69.07, -71.75, -72.54$ without association. Thus, the marginal likelihood are not sensitive to the prior distribution of τ.

5. Concluding Remarks

We have shown how to analyze multinomial data from $r \times c$ categorical tables when there are both item nonresponse and unit nonresponse and the nonresponse mechanism may be nonignorable. We have also shown that by using the Bayes factor (ratio of the marginal likelihoods of two models) we can test for association between the two categories, and it may also be possible to assess nonignorability.

For the 3×3 categorical data of BMD and FI we used, we are able to estimate the cell probabilities very well. Also, while the chi-squared test shows strong evidence for association between BMD and FI, our Bayes factor shows "strong" evidence for no association under the ignorable model, but the evidence for no association under the nonignorable model is "not worth more than a bare mention." We have also shown that there is strong evidence that the ignorable model (simpler) is to be chosen over the nonignorable model. (Positive evidence is $2 \leq 2\log(\text{Bayes Factor}) \leq 6$.)

Further research can try to reduce the number of parameters in the nonignorable model. For example, one can fit a model in which the π_{sjk} can be taken to be $\pi_{sjk} = \pi_s \gamma_{jk}$ in which a prior distribution is put on the γ_{jk} so that the nonignorable model is centered on an ignorable one. It is also possible to consider representing the data in two categorical tables, one with the complete data and the other with the incomplete data, instead of three supplemental tables as we did.

References

Altham, P. M. (1976), "Discrete variable analysis for individuals grouped into families," *Biometrika*, 63, 263-269.

Brier, S. E. (1980), "Analysis of contingency tables under cluster sampling," *Biometrika*, 67, 591-596.

Chen, T. and Fienberg, S. E. (1974), "Two-dimensional contingency tables with both completely and partially cross-classified data," *Biometrics*, 30, 629-642.

Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270-281.

Choi, J. W. and McHugh, R. B. (1989), "A reduction factor in goodness-of-fit and independence for clustered and weighted observations," *Biometrics*, 45, 979-996.

Cohen, J. E. (1976), "The distribution of the chi-squared under cluster sampling from contingency tables," *Journal of the American Statistical Association*, 71, 591-596.

Draper, D. (1995), "Assessment and propagation of model uncertainty" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 57, 45-97.

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

Greene, M, A., Smith, L. E., Levenson, S .H., Hiser, S. and Mah, J. C. (2002), "Raking fire data," *U.S. Consumer Product Safety Commission*, 1-11.

Kass, R. and Raftery, A. (1996), "Bayes factors," *Journal of the American Statistical Association*, 90, 773-795.

Kalton, G. and Kasprzyk, D. (1986), "The treatment of missing survey data," *Survey Methodology*, 12, 1-16.

Little, R. J. (1985), "Nonresponse adjustments in longitudinal surveys: Models for categorical data," *Bulletin of the International Statistical Institute*, 15.1, 1-15.

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

Nandram, B. and Choi, J. W. (2002a), "Hierarchical Bayesian Nonresponse Models for Binary Data from small areas with Uncertainty about Ignorability," *Journal of the American Statistical Association*, 97, 381-388.

Nandram, B. and Choi, J. W. (2002b), "A Bayesian Analysis of a Proportion under Nonignorable Nonresponse," *Statistics in Medicine*, 21, 1189-1212.

Nandram, B., G. Han and Choi, J.W. (2002), "A Hierarchical Bayesian Nonignorable Nonresponse Model for Multinomial Data From Small Areas," *Survey Methodology*, 28, 145-156.

Rao, J. N. K. and Scott, A. J. (1981), "The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables," *Journal of the American Statistical Association*, 76, 221-230.

Rao, J. N. K. and Scott, A. J. (1984), "On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data," *The Annals of Statistics*, 12, 46-60.

Wang, H.,(2001), *Two-way contingency tables with marginally and conditionally imputed nonrespondents*, Ph.D. Dissertation, Department of Statistics, University of Wisconsin-Madison.