Bayes Empirical Bayes Estimation of a Proportion under Nonignorable Nonresponse

Balgobin Nandram, Worcester Polytechnic Institute, Worcester, MA
Jai Won Choi, National Center for Health Statistics
Jai Won Choi, Room 915, NCHS, 6525 Belcrest Road, Hyattsville, MD 20782

Key Words: Beta-Binomial model; Health survey; EM algorithm; Importance sampling; Latent variable; Selection approach.

Abstract

We analyze binary nonresponse data on doctor visits at a household level by state in the in the National Health Interview Survey (NHIS). We dichotomize the number of doctor visits to zero or at least one doctor visit by a household in the fifty states and the district of Columbia. There is substantial nonresponse in NHIS.

The proportion of households with doctor visits might be misleading when the nonrespondents are ignored. The main issue we address here is that nonresponse should not be treated the same way as response because respondents and nonrespondents can differ. Thus, we begin by assuming that nonresponse is nonignorable, but in our model we also accommodate the case in which the nonresponse may be considered ignorable.

We consider a nonignorable nonresponse model that expresses uncertainty about ignorability. We use a hierarchical Bayesian model in which the hyperparameters are assumed fixed but unknown, and these hyperparameters are estimated by the EM algorithm. Thus, our method is Bayes empirical Bayes, but we still have to fit the rest of our model using a small-scale sample based method. Our main result is that for some of the states the nonresponse mechanism is ignorable, and that 95% credible intervals of the probability of a household doctor visit and the probability that a household responds shed important light on the NHIS.

1. Introduction

There has been much activity in estimating survey nonresponse. For many of the health surveys the response is binary, and this method is easily applicable. For example, the National Health Interview Survey (NHIS) estimates the proportion of households with at least one doctor visit during the past year.

The NHIS executes the surveys on chronic and acute condition, doctor visit, hospital discharge, and medical care and utilization, disability, and other health topics. This survey finding helps to formulate improved health care policies. We study doctor visit as an indicator of health. We dichotomize the number of doctor visits to zero or at least one doctor visit by a household in the fifty states and the district of Columbia.

The NHIS questionnaire is divided into two sections: core and supplemental questions. The core part includes the basic health questions on condition, doctor visit, hospital discharge, and personal information, while the supplemental part includes different topics solicited annually from the general public, and it encompasses a wide range of topics such as prescription medicine, hypertension, diabetis, high blood pressure, HIV, and others. There are many possible biases such as those associated with questionnaire, interviewer, proxy respondent, recall and sample, we focuss only on the problem of nonresponse bias.

The Bayesian approach appears to be particularly suitable for such nonresponse problems (e.g., Little and Rubin 1987). A full Bayesian analysis is difficult. We consider the Bayes empirical Bayes model (Deely and Lindley 1981) to study nonignorable nonresponse. We estimate the hyperparameters using the EM algorithm, and afterward we assume that these estimates are known. The analysis is still complicated, even modal estimates using the EM algorithm are difficult to obtain, but there is some simplification and indeed an underestimation of variability. Although the Bayesian method is appropriate for the analysis of nonignorable nonresponse problem, the main difficulty is to model the relationship between the respondents and nonrespondents.

Stasny (1991) used a hierarchical Bayes model to study victimization in the National Crime Survey. Assuming that the hyperparameters are fixed but unknown, she used the selection approach developed primarily to study sample selection problems. Our approach is an extension of hers.

We use a nonignorable model that is centered on ignorable model with centering parameter γ. Here γ is the odds of a doctor visit among responding households relative to the odds of that among all households. Like Stasny (1991) we use small area estimation techniques where the states are assumed to follow a common stochastic process. We note that a model with a centering parameter was described by Forster and Smith (1998) who used a Bayesian graphical nonresponse model within the pattern mixture approach to analyze data from the British general election polling. However, the problem described is not within the small area context.

We use an empirical hierarchical Bayese approach to study the proportion possessing a characteristic for nonignorable nonresponse when there is uncertainty about ignorability. In Section 2, the National Health Interview Survey is briefly described. In Section 3 we fit the NHIS nonignorable nonresponse to an expansion model with the parameter γ. Section 4 includes an analysis of NHIS data. Section 5 has a summary.

2. National Health Interview Survey

The National Health Interview Survey (NHIS) has been conducted every year since 1957 by the National Center for Health Statistics to measure one aspect of health status of the U.S. noninstitutionalized civilian population. Through this sample survey, NCHS monitors the nation's health by checking the chronic and acute conditions, doctor visits, hospital episodes, disability, other special aspects of health of U.S population. One of the variables of interest in the NHIS is the number of doctor visits by an entire household in the past year. As an example in this research, we use the binary variable, *visit*, to be 0 if the number of doctor visits by all members of an household is 0, and 1 otherwise.

The NHIS nonrespondents are mainly refusals, non contacts, those households with language difficulties, or households not qualified. They may arise nonrandomly. For example, the language problem may be confined to recent immigrants, who are not representative households, and therefore, nonresponse from this source can be considered nonrandom nonresponses. We observed that the average NHIS nonresponse rate was about 2-3 percent until the 1980's and has been increasing annually and reached 8-12 percent in 1995.

The NHIS frame is basically a two stage sample survey design of probability proportional to population size. The first stage is the selection of primary sampling units, and the second stage is the selection of segments. On the average each segment includes about 4-12 households, and all the sample households in the segment are interviewed. Weighting in the NHIS is a multi-stage scheme, and one of the stages is ratio adjustment for nonresponse at the segment level. This ratio is the proportion of all sample persons to the respondents in the segment. This ratio estimator is adequate when respondents and nonrespondents are similar. However, this method can fail badly when these two groups differ according to important characteristics which an investigator wants to study.

We address the nonignorable nonresponse problem by expanding the method of random weighting, and the Bayesian method is introduced as a possible alternative to impute the NHIS nonresponses.

For our illustration we use the 50 states and the District of Columbia from the 1995 household survey. States with at least 8% nonrespondents are Colorado, Delaware, District of Columbia, Florida, Louisiana, Maryland, New York, South Carolina and West Virginia. Hawaii and Maine reported the highest proportions of doctor visits of 38% (see Table 1).

3. Nonignorable Nonresponse Model

In this section we describe a nonignorable model centered on an ignorable model, and we call it the expansion model. In a way we expand ignorable model to nonignorable one via an additional parameter γ.

Let the binary characteristic be

$$x_{ij} = \begin{cases} 1, & \text{household } j \text{ in state } i \text{ has doctor visits} \\ 0, & \text{otherwise} \end{cases}$$

and the response variable

$$y_{ij} = \begin{cases} 1, & \text{household } j \text{ in state } i \text{ responded} \\ 0, & \text{otherwise} \end{cases}$$

where $i = 1, \ldots, \ell$ is the number of states, and $j = 1, \ldots, n_i$ where n_i is the number of sample households.

We use a probabilistic structure to model x_{ij} and y_{ij}.

3.1 Expansion Model

The expansion model for nonignorable nonresponse is

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

$$y_{ij} \mid \pi_i, x_{ij} = 0 \overset{iid}{\sim} \text{Bernoulli } (\pi_i)$$

$$y_{ij} \mid \pi_i, \gamma_i, x_{ij} = 1 \overset{iid}{\sim} \text{Bernoulli } (\gamma_i \pi_i)$$

If $\gamma_i = 1$, the model becomes ignorable model and there is no difference between respondents and nonrespondents. The γ_i are the ratio of the odds of success among respondents to the odds of success among all individuals in the ith state. The γ_i show the extent of nonignorability of the nonrespondents and incorporate the uncertainty about ignorability into the model. The probability of responding in area i is $\delta_i = \pi_i(\gamma_i p_i + (1-p_i))$. Assuming all areas are similar, we take the parameters $(p_i, \delta_i, \gamma_i)$ to have a common distribution.

For p_i, we take

$$p_i \mid \mu_1, \tau_1 \overset{iid}{\sim} \text{Beta } (\mu_1\tau_1, (1-\mu_1)\tau_1).$$

The parameters (π_i, γ_i) are jointly independent with

$$\pi_i \mid \mu_2, \tau_2 \overset{iid}{\sim} \text{Beta } (\mu_2\tau_2, (1-\mu_2)\tau_2) \ \ \textit{and}$$

$$\gamma_i \mid \nu \overset{iid}{\sim} \Gamma(\nu, \nu) \ 0 < \gamma_i < 1/\pi_i \ \textit{and} \ 0 < \pi_i < 1.$$

Let $B(u, v) = \Gamma(u)\Gamma(v)/\Gamma(u+v)$ be the beta function. Then, the joint prior density for (π_i, γ_i) is given by $p(\pi_i, \gamma_i \mid \mu_2, \tau_2, \nu)$:

$$\nu\gamma_i^{\nu-1}\exp(-\nu\gamma_i)\frac{\pi_i^{\mu_2\tau_2-1}\ (1-\pi_i)^{(1-\mu_2)\tau_2-1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)\ I_i(\mu_2, \tau_2, \nu)}$$

$$\textit{where} \ \ I_i(\mu_2, \tau_2, \nu) \ \ \textit{is} \tag{1}$$

$$\int_0^1\int_0^1 \left\{\pi_i^{-1} exp(-\phi_i/\pi_i)\right\}^{\nu} f_1(\pi_i, \phi_i \mid \nu) d\pi_i \ d\phi_i$$

and, for $0 < \pi_i, \ \phi_i < 1$,

$$f_1(\pi_i, \phi_i \mid \nu) = \nu\phi_i^{\nu-1}\frac{\pi_i^{\mu_2\tau_2-1}\ (1-\pi_i)^{(1-\mu_2)\tau_2-1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}.$$

The joint prior distribution for $(\pi_i, \phi_i\, \gamma_i)$ is the product of $p(p_i \mid \mu_1, \tau_1)$ and $p(\pi_i, \gamma_i \mid \mu_2, \tau_2, \nu)$.

The ignorable model is a special case of the expansion model with $\gamma_i = 1$. Note that, if γ_i follow $\Gamma(\nu, \nu)$, $E(\gamma_i \mid \nu) = 1$ and $Var(\gamma_i \mid \nu) = 1/\nu$. That is, we have centered the expansion model on the ignorable model and that γ_i fluctuates about unity with a standard deviation $1/\sqrt{\nu}$ a priori.

We take uniform prior on μ_1 and μ_2, and the proper priors $p(\nu) = 1/(\nu+1)^2$ for $\nu \geq 0$, $p(\tau_r) = 1/(\tau_r+1)^2$ for $\tau_r \geq 0$ and $r = 1,\ 2$. However, we use the EM algorithm to estimate the hyperparameters directly. The EM algorithm is a general approach to iterate the computation of maximum likelihood estimation when the observation can be viewed as incomplete data.

Let $r_i = \sum_{j=1}^{n_i} y_{ij}$ be the number of responding households in the ith state and $x_i = \sum_{j=1}^{n_i} x_{ij}$ the number of households with at least one doctor visit in the ith state, and $n_i - r_i$ is the number of nonrespondents. Since the number of visits among the nonrespondents is unknown, we denote it by the latent variable z_i, and hence, the number of non-visits among them is $n_i - r_i - z_i$.

Then it is easy to show that the likelihood function is proportional to $f(\mathbf{x}, \mathbf{r}, \mathbf{z} \mid \mathbf{p}, \gamma, \pi)$, which is $\prod_{i=1}^{\ell}\sum_{z_i=0}^{n_i-r_i} f(x_i, r_i, z_i \mid p_i, \gamma_i, \pi_i)$, where $f(x_i, r_i, z_i \mid p_i, \gamma_i, \pi_i)$ is

$$\binom{n_i}{r_i}\binom{r_i}{x_i}\binom{n_i-r_i}{z_i}(\gamma_i\pi_i p_i)^{x_i}(\pi_i(1-p_i))^{r_i-x_i}$$

$$\times((1-\gamma_i\pi_i)p_i)^{z_i} \ \ ((1-\pi_i)(1-p_i))^{n_i-r_i-z_i}.$$

By Bayes' theorem the joint posterior density follows, but it is convenient to make the transformation $\phi_i = \gamma_i\pi_i$.

Let $A_i = x_i + z_i + \mu_1\tau_1$, $B_i = n_i - x_i - z_i + (1-\mu_1)\tau_1)$, $C_i = r_i + x_i + \mu_2\tau_2$, and $D_i = n_i - r_i - z_i + (1-\mu_2)\tau_2)$. The joint posterior density of all the parameters $(\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\phi})$ for given data $(\mathbf{x}, \mathbf{r})$ is $f(\mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\pi}, \mid \mathbf{x}, \mathbf{r}, \mathbf{z})$ which is porportional to: $p(\nu)p(\mu_1)p(\mu_2)p(\tau_1)p(\tau_2)\ \times$

$$\prod_{i=1}^{\ell}\Bigg\{\binom{n_i-r_i}{z_i}\ \frac{p_i^{A_i-1}\ (1-p_i)^{B_i-1}}{B(\mu_1\tau_1,\ (1-\mu_1)\tau_1)}$$

$$\times\ \phi_i^{x_i+\nu-1}(1-\phi_i)^{z_i}\ \ \frac{\pi_i^{C_i-1}\ (1-\pi_i)^{D_i-1}}{B(\mu_2\tau_2,\ (1-\mu_2)\tau_2)}$$

$$\times\ \ \left\{\pi_i^{-1}exp(-\phi_i/\pi_i)\right\}^{\nu}\ \nu\ I_i^{-1}(\mu_2, \tau_2, \nu)\Bigg\}. \tag{2}$$

3.2 Computations

Because the posterior density is not accessible directly, we use a sampling based method to obtain samples from the posterior density to permit an inference. We marginalize out the parameters (p_i, π_i, ϕ_i) from the joint posterior density to obtain the posterior density $f(\mu_1, \tau_1, \mu_2, \tau_2, \nu, \mathbf{z} \mid \mathbf{x}, \mathbf{r})$.

Because it is difficult to obtain samples from this posterior density, we obtain the posterior mode for the hyperparameters μ_1, τ_1, μ_2, τ_2, and ν. For simplicity, we use the EM algorithm to obtain the estimates of the hyperparameters (Dempster, Laird, and Rubin, 1977). Then we assume that these estimated values are the true values. (This is the Bayes empirical Bayes procedure, Deely and Lindley 1981.)

The EM algorithm is initiated with the values of $\mathbf{z}$, μ_1, τ_1, μ_2, τ_2, and ν obtained by the method of moments.

Let $f_a(\mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{x}, \mathbf{r}, \mathbf{z})$ be proportional to $p(\nu)p(\tau_1)p(\tau_2)) \times \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} \right.$

$$\times \frac{B(A_i, B_i)}{B(\mu_1\tau_1,\ (1-\mu_1)\tau_1)}$$

$$\times \frac{B(C_i, D_i)}{B(\mu_2\tau_2,\ (1-\mu_2)\tau_2)}$$

$$\left. \times\ \nu\ B(x_i + \nu, z_i + 1) \right\}.$$

The posterior density $f(\mu_1, \tau_1, \mu_2, \tau_2, \nu, \mathbf{z} \mid \mathbf{x}, \mathbf{r})$ is proportional to

$$f_a(\mu_1, \tau_1, \mu_2, \tau_2, \nu, \mathbf{z} \mid \mathbf{x}, \mathbf{r}) \prod_{i=1}^{\ell} \{R_{z_i}(\mu_2, \tau_2, \nu)\}$$

where $R_{z_i}(\mu_2, \tau_2, \nu)$ is

$$\frac{\int_0^1 \int_0^1 \{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu} f_2(\pi_i, \phi_i \mid \nu, z_i, x_i, r_i) d\pi_i\ d\phi_i}{I_i(\mu_2, \tau_2, \nu)}$$

and $f_2(\pi_i, \phi_i \mid \nu, z_i, x_i, r_i)$ is

$$\frac{\phi_i^{x_i+\nu-1}(1-\phi_i)^{z_i}}{B(x_i+\nu, z_i+1)} \frac{\pi_i^{C_i-1}\ (1-\pi_i)^{D_i-1}}{B(C_i, D_i)}.$$

Observe that $R_{z_i}(\mu_2, \tau_2, \nu)$ is the ratio of the expectations of $\{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu}$ over $f_2(\pi_i, \phi_i \mid \nu, z_i, x_i, r_i)$ and $f_1(\pi_i, \phi_i \mid \nu)$ in the numerator and denominator, respectively, and $f_1(.)$ is given in (1). With the known hyperparameters, we can obtain samples from the joint posterior density $f(z_i, p_i, \pi_i, \phi_i \mid \mathbf{x},\ \mathbf{r})$ which is given by

$$g_1(p_i \mid z_i, x_i, r_i)\ g_2(\pi_i, \phi_i \mid z_i, \mathbf{x},\ \mathbf{r})\ g_3(z_i \mid \mathbf{x},\ \mathbf{r}).$$

The posterior of $p_i \mid z_i, x_i, r_i$ is

$$g_1(p_i \mid x_i, r_i, z_i) \overset{iid}{\sim} \text{Beta}\ (A_i, B_i).$$

from which the samples for p_i are obtained.

The joint posterior $\pi_i, \phi_i \mid x_i, r_i, z_i$ is $g_2(\pi_i, \phi_i \mid x_i, r_i, z_i)$ which is proportional to

$$\{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu}\ g_a(\pi_i, \phi_i \mid x_i, r_i, z_i, \nu)$$

where $g_a(\pi_i, \phi_i \mid z_i, x_i, r_i, \nu)$ is

$$\frac{\phi_i^{x_i+\nu-1}(1-\phi_i)^{z_i}}{B(x_i+\nu, z_i+1)} \frac{\pi_i^{C_i-1}\ (1-\pi_i)^{D_i-1}}{B(C_i,\ D_i)}.$$

The probability mass function of z_i is $g_3(z_i \mid x_i, r_i)$. Also observe that

$$g_3(z_i = t \mid \mathbf{x}, \mathbf{r}) = \frac{\omega_t}{\sum_{t=0}^{n_i - r_i} \omega_t} \quad t = 0, ..., n_i - r_i,$$

where ω_t is proportional to

$$\omega_t \binom{n_i - r_i}{z_i} B(A_i,\ B_i)$$

$$\times\ B(x_i + \nu,\ z_i + 1)\ \times\ B(C_i,\ D_i)\ \times\ I_{z_i}(\mu_2, \tau_2, \nu)$$

where

$$I_{z_i}(\mu_2, \tau_2, \nu) = \int_0^1 \int_0^1 \left\{\frac{exp(-\phi_i/\pi_i)}{\pi_i}\right\}^{\nu} \times$$

$$\frac{\phi_i^{x_i+\nu-1}(1-\phi_i)^{z_i}}{B(x_i+\nu,\ z_i+1)} \frac{\pi_i^{C_i-1}\ (1-\pi_i)^{D_i-1}}{B(C_i,\ D_i)}\ d\pi_i\ d\phi_i.$$

We first obtain z_i from equation $z_i \mid \mathbf{x}, \mathbf{r}$. Then for each z_i we fill in the (p_i, π_i, ϕ_i) with draws from $g_1(.)$ and $g_2(.)$. We draw 1000 values for z_i $i = 1,...,\ell$ from the equation $g_3(z_i \mid \mathbf{x}, \mathbf{r})$. To draw samples from $g_2(.)$, we use the Metropolis algorithm with proposal density $g_a(\pi_i, \phi_i \mid x_i, r_i, z_i, \nu)$. Assuming the chain is at the sth iterate, then the jumping probability to the $(s+1)st$ iterate is $A_{s,s+1} = min\{1,\ (\psi(\pi_i^{(s+1)}, \phi_i^{(s+1)})/\psi(\pi_i^{(s)}, \phi_i^{(s)}))\}$, where

$$\psi(\pi_i, \phi_i) = \{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu}.$$

Here $(\pi_i^{(s)}, \phi_i^{(s)})$ are obtained independently from

$$\phi_i^{(s)} \mid r_i, x_i, z_i, \nu \overset{ind}{\sim} \text{Beta}\ (x_i + \nu, z_i + 1)$$

and

$$\pi_i^{(s)} \mid r_i, x_i, z_i, \mu_2, \tau_2 \overset{iid}{\sim} \text{Beta}\ (C_i,\ D_i).$$

We ran the Metropolis step 100 times and we took the last one.

We finally obtain a sample $(\pi_i^{(h)}, \phi_i^{(h)}, \gamma_i^{(h)})$ by taking $\gamma_i^{(h)} = \pi_i^{(h)}\phi_i^{(h)}$, $h = 1, ..., M$. Inference can now be made in standard way.

4. Analysis of NHIS data

In this section we assess the fit of the expansion model to the NHIS data. Then, we discuss posterior inference about the parameters of the expansion model. We note that the EM algorithm converges within 10 steps and the estimates of the hyperparameters are $\mu_1 = 0.309$, $\tau_1 = 1237.408$, $\mu_2 = 0.864$, $\tau_2 = 2.846$, and $\nu = 526.223$.

4.1 Model Assessment

We assess our model by using a Bayesian cross validation analysis. Let $q_i = \gamma_i p_i/(p_i\gamma_i + (1-p_i))$. Using Bayes theorem, it is easy to show that for the respondents

$$x_i \mid p_i, \pi_i, r_i \overset{iid}{\sim} \text{Binomial}\,(r_i, q_i).$$

Let $\hat{p}_i = x_i/r_i$ and $x_{(i)}$ and $r_{(i)}$ be vectors of all x_i and r_i with the i-th state deleted. For the M iterates from the sampling based method we compute the weights

$$\omega_{ih} = \frac{\{q_{ih}^{x_i}(1-q_{ih})^{r_i-x_i}\}^{-1}}{\sum_{h=1}^{M}\{q_{ih}^{x_i}(1-q_{ih})^{r_i-x_i}\}^{-1}}$$

where

$$q_{ih} = \gamma_i^{(h)} p_i^{(h)}/\{(p_i^{(h)}\gamma_i^{(h)} + (1-p_i^{(h)})\}$$

and $(\gamma_i^{(h)}, p_i^{(h)})$ $h = 1, ..., M,$ $i = 1, ..., \ell$ are iterates from the sampling based method.

Then we compute

$$dres_i = \frac{\hat{p}_i - E(p_i \mid x_{(i)}, r_{(i)})}{Std(p_i \mid x_{(i)}, r_{(i)})}, \quad i = 1, \ldots, \ell$$

where $E(p_i \mid x_{(i)}, r_{(i)}) = \sum_{h=1}^{M} \omega_{ih} q_{ih} = \bar{q}_i$ and $Std(p_i \mid x_{(i)}, r_{(i)})$ is

$$\{\sum_{h=1}^{M} \omega_{ih}\{(q_{ih}-\bar{q}_i)^2 + \frac{q_{ih}(1-q_{ih})}{r_i}\}\}^{1/2}.$$

We use the $dres_i$ to assess the model fit. When we plotted $dres_i$ versus $E(p_i \mid x_{(i)}, r_{(i)})$, all the points are between -2.5 and 2.5, $dres_i$ is symmetric about the horizontal line at zero (i.e., of the 51 points, 25 points are above the line). Colorado is a state among the ones with the smallest proportions of doctor visits. Yet this state has the largest $E(p_i \mid x_{(i)}, r_{(i)})$. For further exploration we compute $\rho_i = Pr(\gamma_i \le 1 \mid \mathbf{x}, \mathbf{r})$. For Colarodo the ρ_i is 0.002 (i.e., the extent of nonignorability is extreme).

4.2 Posterior Inference

Our main results are presented in Table 1.

The second column of Table 1 contains the observed proportions $\hat{p}$ of households with at least one doctor visit. The third column contains the 95% credible for the population proportion p of households with at least one doctor visit. For fourteen of the states the credible intervals do not contain the observed proportions (the $\hat{p}_i$ are marked with a +). This implies that the observed values may be unreasonable estimates for the true proportions.

The fourth column of Table 1 shows that the 95% credible intervals of γ for each state. The fifth column contains ρ, the posterior probability that γ is less than one. Recall that when $\gamma = 1$, the expansion model is an ignorable model. For seven states the intervals do not include 1. These are Colorado, Georgia, Lousiana, Massachusetts, South Carolina, Virginia, and Washington. For these seven states, ρ_i are very small with values of 0.002, 0.007, 0.003, 0.025, 0.012, 0.006, and 0.021, respectively, which are extremes. Thus, the nonresponse mechansims for these states should be treated as nonignorable.

We have performed a sensitivity analysis to assess how inference is affected by the choice of the hyperparameters τ_1, τ_2 and ν. We kept μ_1 and μ_2 at the modal estimates, and set τ_1 at 500, 1000, 2000, τ_2 at 3, 6, 12, and ν at 250, 500, 1000. For each of the 27 combinations, we obtained the modal estimates and computed the 95% credible interval for p, γ, and δ. We found that inference is virtually unchanged for p and δ. There were some changes for γ as ν changes, but this is small.

Finally we observed the relation between ignorability and goodness of fit using dres described in Section 4.1. A cross tabulation of the standardized deleted residuals at the two levels ($\mid dres_i \mid \le 2$) and ($\mid dres_i \mid > 2$), and ρ_i at two levels (≤ 0.05) and (> 0.05) for ignorable and nonignorable states. Fisher's exact test from $SAS^{(R)}$ PROC FREQ gives a p-value of 0.978 (left-tail), 0.292 (right or two-level). Thus, the expansion model fits the states with ignorable and noningorable nonresponse equally well.

5. Summary

We have presented a Bayes empirical Bayes method to estimate the proportion of doctor visits and the probability that a household responds in the NHIS, incorporating a degree of uncertainty about the ignorability of the nonresponse mechanism. Our method assumes that the hyperparameters are fixed but unknown, and they are estimated using modal estimates from the EM algorithm. However, we have shown that nonreponse is nonignorable and as such it needs to be treated accordingly.

We found that moderate misspecifications of the parameters τ_1, τ_2 and ν have little consequence on inference about p_i and δ_i as well as γ_i.

Our method is potentially useful to incorporate uncertainty about ignorability of the nonresponse mechanism. We have shown that it is possible to decide for which states the nonresponse mechanism can be treated as ignorable. For these states it is possible to use the ratio method for nonresponse adjustment. For the other states one must be reluctant to use the ratio method. In either case our method provides adjusted estimates for p_i and δ_i based on the extent of ignorability.

It is possibly to do a full Bayesian approach by using the algorithm of Nandram (1998). One other important extension is to polychotomous data that are so prominent in many complex surveys. It is also possible to do Bayesian predictive inference for the finite population mean of each state.

Reference

Deely, J. J. and Lindley, D. V. (1981), "Bayes Empirical Bayes," *Journal of the American Statistical Association*, **76**, 833-841.

Dempster, A. P., Laird, N. M. and Rubin, D. B.(1977), "Maximum Likelihood From Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B, **39**, 1-22.

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

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

Nandram, B. (1998), "A Bayesian Analysis of the Three-Stage Hierarchical Multinomial Model," *Journal of Statistical Computation and Simulation*, 61, 97-126.

Stasny, E. A. (1991), "Hierarchical Models for the Probabilities of a Survey Classification and Nonresponse: An Example from the National Crime Survey," *Journal of the American Statistical Association*, **86**, 296-303.

Table 1: **Credible Intervals and Probability**

state	$\hat{p}$	Inter(p)	Inter(γ)	ρ
Al	0.34	(0.298,0.341)	(0.965,1.079)	0.135
Ak	0.34+	(0.285,0.336)	(0.998,1.107)	0.036
Az	0.29	(0.280,0.324)	(0.981,1.093)	0.083
Ar	0.26+	(0.276,0.322)	(0.965,1.062)	0.264
Ca	0.31	(0.290,0.323)	(0.971,1.049)	0.170
Co*	0.27+	(0.271,0.315)	(1.035,1.156)	0.002
Ct	0.31	(0.283,0.327)	(0.971,1.086)	0.123
De*	0.37+	(0.288,0.338)	(0.968,1.149)	0.139
DC*	0.32	(0.285,0.334)	(0.980,1.160)	0.084
Fl*	0.31	(0.282,0.333)	(0.933,1.078)	0.239
Ga	0.30	(0.280,0.319)	(1.011,1.097)	0.007
Hi	0.38+	(0.294,0.345)	(0.966,1.111)	0.180
Id	0.29	(0.283,0.333)	(0.985,1.029)	0.496
Il	0.28	(0.268,0.325)	(0.872,1.028)	0.446
In	0.31	(0.287,0.328)	(0.987,1.077)	0.068
Ia	0.36+	(0.297,0.345)	(0.968,1.068)	0.230
Ks	0.30	(0.283,0.328)	(0.970,1.091)	0.133
Ky	0.34+	(0.294,0.338)	(0.987,1.105)	0.058
La*	0.32	(0.285,0.329)	(1.022,1.138)	0.003
Me	0.38+	(0.293,0.342)	(0.964,1.110)	0.184
Md*	0.33	(0.289,0.346)	(0.892,1.132)	0.078
Ma	0.29	(0.279,0.320)	(1.000,1.092)	0.025
Mi	0.34	(0.301,0.359)	(0.892,1.045)	0.368
Mn	0.34	(0.294,0.340)	(0.971,1.081)	0.116
Ms	0.29	(0.280,0.328)	(0.969,1.072)	0.213
Mo	0.30	(0.284,0.326)	(0.980,1.071)	0.102
Mt	0.28	(0.280,0.331)	(0.982,1.059)	0.114
Ne	0.33	(0.288,0.336)	(0.974,1.104)	0.144
Nv	0.35+	(0.287,0.338)	(0.969,1.124)	0.134
NH	0.31	(0.284,0.334)	(0.975,1.086)	0.147
NJ	0.31	(0.285,0.347)	(0.882,1.043)	0.371
NM	0.28+	(0.281,0.326)	(0.960,1.069)	0.216
NY*	0.30	(0.276,0.321)	(0.962,1.085)	0.134
NC	0.30	(0.283,0.335)	(0.901,1.088)	0.119
ND	0.32	(0.283,0.336)	(0.982,1.084)	0.095
Oh	0.34	(0.301,0.352)	(0.914,1.061)	0.167
Ok	0.31	(0.284,0.331)	(0.981,1.100)	0.080
Or	0.27+	(0.276,0.320)	(0.967,1.043)	0.340
Pa	0.34	(0.301,0.357)	(0.889,1.033)	0.410
RI	0.28+	(0.281,0.330)	(0.974,1.076)	0.234
SC*	0.33	(0.287,0.334)	(1.005,1.135)	0.012
SD	0.29	(0.281,0.331)	(0.990,1.063)	0.140
Tn	0.32	(0.289,0.332)	(0.987,1.088)	0.062
Tx	0.29	(0.276,0.318)	(0.926,1.035)	0.411
Ut	0.26+	(0.278,0.326)	(0.961,1.084)	0.231
Vt	0.32	(0.286,0.334)	(0.973,1.089)	0.216
Va	0.33	(0.293,0.334)	(1.016,1.108)	0.006
Wa	0.33	(0.294,0.336)	(1.002,1.099)	0.021
WV*	0.32	(0.286,0.336)	(0.951,1.136)	0.101
Wi	0.33	(0.295,0.338)	(0.956,1.055)	0.229
Wy	0.36+	(0.286,0.336)	(0.985,1.081)	0.188

NOTE. * States with 8% or more nonrespondents