Journal of the Korean
Statistical Society
Vol. 23, No. 2, 1994

Bayesian Analysis of Randomized Response Models: A Gibbs Sampling Approach†

ManSuk Oh[1]

ABSTRACT

In Bayesian analysis of randomized response models, the likelihood function does not combine tractably with typical priors for the parameters of interest, causing computational difficulties in posterior analysis of the parameters of interest. In this article, the difficulties are solved by introducing appropriate latent variables to the model and using the Gibbs sampling algorithm.

KEYWORDS: Sampling survey, Monte Carlo, Posterior inference, Latent variables.

1. INTRODUCTION

In sampling survey, respondents may not give honest answers to sensitive questions like 'Do you have an experience of abortion?'. Randomized response models are designed to protect respondents on sensitive questions for more

†This research was supported by Ewha Womans University Research Fund.
[1]Deptpartment of Statistics, Ewha Womans University, Sodaemun-Gu, Seoul 120-750, Korea.

honest responses. It was first introduced by Warner (1965). In Warner's model each interviewee answers to either the sensitive question of interest or the complementary of it depending on the outcome of a random device such as tossing a coin. The outcome of the random device is unknown to the interviewer so that privacy of respondents can be protected. This would draw more honest answers and reduce non-honest-answer bias. But the price is that data incorporates outcomes of the random device as well as answers to the question and hence there is some loss of information from data. Also the information from data is not given in a convenient form for inference on the parameters of interest.

Horvitz, Shah, and Simons (1967) introduced the unrelated question scheme, in which each respondent answers to the sensitive question or the unrelated nonsensitive question depending on the outcome of a random device. This unrelated question scheme can handle polychotomous response data by giving the same number of categories in the unrelated question. Also it can handle qualitative questions.

A modification of the above schemes to reduce the loss of information while keeping the protection capability is the two stage model by Mangat and Singh (1990). Two stage models are more complex but they can be made more efficient than simple models by choosing appropriate probabilities in random devices.

As one can imagine there are situations where each interviewee answers a question with $t(t \geq 2)$ categories and at least one and at most $t-1$ of the categories are stigmatizing. Various randomized response models for such polychotomous response data are described in Chaudri and Mukerjee (1988). Some are direct extensions of the randomized response models for binary data described above and some are different types of extensions.

Bayesian analyses of randomized response models are given in Winkler and Franklin (1979), Pitz (1980), and O'Hagan (1987). Bayesian methods are particularly attractive in randomized response models because they incorporate useful prior information where only partial sample information is available

(Winkler and Franklin (1979)). However, because of the random device used for the collection of honest responses, the likelihood function of the parameters of interest, usually the proportion of each sensitive category in the question, has a complicated form, being discordant with useful prior distributions such as beta or Dirichlet priors. This causes computational difficulties in deriving posterior distributions of the parameters of interest. To get around this Winkler and Franklin (1979) used an approximation of the posterior distributions and O'Hagan(1987) found simple but restrictive estimators for the parameters of interest.

In this paper, Bayesian analysis of randomized response models is developed using the idea of data augmentation, specifically using the Gibbs sampling algorithm. In each model, appropriate latent variables are introduced to the model for direct implementation of the Gibbs sampling algorithm. With random samples from the Gibbs sampling algorithm, any posterior quantities including marginal posterior density functions of components can be easily obtained and hence complete posterior analysis of the parameters of interest is possible.

This paper is organized as follows. In Section 2, the Gibbs sampling algorithm is briefly described. In Section 3, Bayesian analysis of various randomized binary response models using the Gibbs sampling algorithm is described. In section 4, randomized response models for binary data are extended to handle polychotomous data and Bayesian analysis of these models are described. An example is given in Section 5 and summary in Section 6.

2. GIBBS SAMPLING

In this section the Gibbs sampling algorithm by Gelfand and Smith (1990) is briefly described.

One is interested in simulation from a posterior distribution of d-variate random variable $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Though it is difficult to generate samples from the joint distribution of $\boldsymbol{\theta}$, suppose that it is easy to simulate from the full conditional distributions $f(\theta_k|\theta_j, j \neq k, j = 1,\ldots,d)$ for $k = 1,\ldots,d$. Under this assumption, the Gibbs sampling algorithm proceeds as follows: First, start with an initial guess $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)},\ldots,\theta_d^{(0)})$ of $\boldsymbol{\theta}$. Then generate in turn

$$\begin{aligned} &\theta_1^{(1)} \quad \text{from} \quad f(\theta_1|\theta_j^{(0)}, j = 2,\ldots,d) \\ &\theta_2^{(1)} \quad \text{from} \quad f(\theta_2|\theta_1^{(1)}, \theta_j^{(0)}, j = 3,\ldots,d) \\ &\theta_3^{(1)} \quad \text{from} \quad f(\theta_3|\theta_1^{(1)}, \theta_2^{(1)}, \theta_j^{(0)}, j = 4,\ldots,d) \\ &\qquad\quad \vdots \\ &\theta_d^{(1)} \quad \text{from} \quad f(\theta_d|\theta_j^{(1)}, j = 1,\ldots,d-1). \end{aligned}$$

This makes the first cycle and yields a sample $\boldsymbol{\theta}^{(1)}$. Now with $\boldsymbol{\theta}^{(1)}$ as a new initial value of $\boldsymbol{\theta}$, run the second cycle by the same way, augmenting a component in each step of the cycle. The t-th iteration of the cycle yields $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)},\ldots,\theta_d^{(t)})$. As t gets large, the joint distribution of $\boldsymbol{\theta}^{(t)}$ converges to the joint distribution of $\boldsymbol{\theta}$. Thus, for a sufficiently large t, one may consider $\boldsymbol{\theta}^{(t)}$ as a sample from the joint distribution of $\boldsymbol{\theta}$.

Determining the stopping time t and an appropriate sample size is not completely solved issue in the Gibbs sampling algorithm. But recently there are many useful suggestions on these issues. For the stopping time, a typical way is to plot marginal density functions of components of $\boldsymbol{\theta}$ periodically and stop when stability seems to be achieved. Also some quantitative measures are suggested by Gelman and Rubin (1992) and Geyer (1992), and an histogram teachnique by Ritter and Tanner(1993). For obtaining a sufficiently large number of observations for the computation of expectations and marginal density functions, there are two suggestions. One is to replicate the Gibbs sampling process m times with different starting values of $\boldsymbol{\theta}$ and consider the samples $\boldsymbol{\theta}_1^{(t)},\ldots,\boldsymbol{\theta}_m^{(t)}$ obtained as m independent samples of $\boldsymbol{\theta}$ from the joint distribution (Gelman and Rubin (1992)). The other is to perform one 'long-run' Gibbs sampling and take $\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+s)} \ldots, \boldsymbol{\theta}^{(t+(m-1)s)}$ as a sample of size m (Geyer

(1992)). Both schemes have advantages and disadvantages, so it is hard to tell which method is better (see comments following Gelman and Rubin (1992) and Geyer (1992)). A rule of thumb is that the one long-run strategy is preferred unless the domain of $\boldsymbol{\theta}$ is well separated so that at least one starting point from each separated region is necessary. In particular, it is not necessary to have independent samples to compute the marginal posterior estimates of components of $\boldsymbol{\theta}$ (Albert and Chib (1993)).

When a sample of size m, $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m$, is obtained, an approximate marginal posterior density function of θ_k, for $k = 1, \ldots, d$, is given by

$$\hat{f}(\theta_k) = \frac{1}{m} \sum_{i=1}^{m} f\Big(\theta_k | \theta_j^i, j \neq k, j = 1, \ldots, d\Big) \tag{2.1}$$

and the posterior expectation of $h(\boldsymbol{\theta})$ for a measurable function h is estimated by

$$\hat{E}h(\boldsymbol{\theta}) = \frac{1}{m} \sum_{i=1}^{m} h(\boldsymbol{\theta}^i).$$

3. ANALYSIS OF RANDOMIZED BINARY RESPONSE DATA

In this section various randomized response models for binary response data are described. In each model appropriate latent variables are introduced to the model so that all necessary full conditional distributions are given in convenient forms. Once conditional distributions of all random variables, including latent variables, in the model are given in convenient forms, generation of random samples by the Gibbs sampling algorithm and inference on the parameters of interest based on the sample are direct as shown in the previous section. Therefore, in the following subsections derivation of all necessary full conditional distributions will suffice.

3.1 Warner's Model.

In Warner's model, an interviewee answers 'Yes' or 'No' to either the sensitive question or the negation depending on the outcome of a random device such as tossing a coin. Simple random sampling with replacement is assumed for the collection of sample. A typical parameter of interest is π , the proportion of 'Yes' to the sensitive question. But the joint density function of data $(y_1, \ldots, y_n)$, $y_i = 0$ or 1, is proportional to $\lambda^{\sum y_i}(1-\lambda)^{n-\sum y_i}$, where λ is the probability of a 'Yes' response regardless of which question is answered. If we let p be the probability of the outcome of the random device leading to the sensitive question, for example, the probability of a head in tossing a coin, then $\lambda = p\pi + (1-p)(1-\pi)$. Thus, when a typical *Beta* prior is used for π, the likelihood function does not combine tractably with the prior and it prohibits full posterior analysis of the parameter of interest.

The problem mentioned above can be solved by using the Gibbs sampling algorithm. To implement the Gibbs sampler in Warner's model, introduce n latent variables, $z_1, \ldots, z_n$, where $z_i = 1$ if the outcome of tossing a coin is head, $z_i = 0$ otherwise. The probability, p, of getting a head is known and z_i has a density function

$$f(z_i) = p^{z_i}(1-p)^{1-z_i}. \tag{3.1}$$

Then the conditional density function of y_i, given z_i and π, is

$$\begin{aligned} f(y_i|z_i,\pi) &= \begin{cases} \pi^{y_i}(1-\pi)^{1-y_i}, & \text{if } z_i = 1 \\ \pi^{1-y_i}(1-\pi)^{y_i}, & \text{if } z_i = 0 \end{cases} \\ &= \pi^{y_i z_i + (1-y_i)(1-z_i)}(1-\pi)^{(1-y_i)z_i + y_i(1-z_i)} \end{aligned} \tag{3.2}$$

The likelihood function $l(\pi, \boldsymbol{z})$, where $\boldsymbol{z} = (z_1, \ldots, z_n)$, is the product of $f(z_i)$ and $f(y_i|z_i,\pi)$ given in (3.1) and (3.2), respectively, for $i = 1, \ldots, n$. With a $Beta(\alpha, \beta)$ prior for π (uniform when $\alpha = \beta = 1$), the posterior density function of π is proportional to $\pi^{y_i z_i + (1-y_i)(1-z_i)+\alpha-1}(1-\pi)^{(1-y_i)z_i + y_i(1-z_i)+\beta-1}p^{z_i}(1-p)^{1-z_i}$. It can be easily seen that the conditional distributions of z_i, given

$\boldsymbol{y} = (y_1, \ldots, y_n)$ and π, is $Bernoulli(p_i^w)$, where

$$p_i^w = \begin{cases} \pi p \Big/ \Big(\pi p + (1-\pi)(1-p)\Big), & \text{if } y_i = 1 \\ (1-\pi)p \Big/ \Big((1-\pi)p + \pi(1-p)\Big), & \text{if } y_i = 0, \end{cases}$$

independently for $i = 1, \ldots, n$, and the conditional distribution of π, given $\boldsymbol{y}$ and $\boldsymbol{z}$, is $Beta(\alpha^w, \beta^w)$, where

$$\begin{aligned} \alpha^w &= \sum_{i=1}^n y_i z_i + \sum_{i=1}^n (1-y_i)(1-z_i) + \alpha \\ &= \text{ the number of } \{y_i = z_i, i = 1, \ldots, n\} + \alpha \\ \beta^w &= \sum_{i=1}^n (1-y_i) z_i + \sum_{i=1}^n y_i(1-z_i) + \beta \\ &= \text{ the number of } \{y_i \neq z_i, i = 1, \ldots, n\} + \beta. \end{aligned}$$

All these conditional distributions are of very simple forms and generation of random samples from these distributions can be done very easily. With the sample generated, the marginal posterior density function of π, $f(\pi|\boldsymbol{y})$, can be obtained using (2.1).

3.2 The Unrelated Question Model of Horvitz, Shah, and Simons.

The model of Horvitz et. al. (1967) replaces the complementary of sensitive question in Warner's model by an unrelated nonsensitive question. This model is more flexible in that it allows not only polychotomous responses but also qualitative questions.

To implement the Gibbs sampling algorithm in this model, introduce latent variables z_i as in Warner's model. If $z_i = 1$ then let $y_i = 1$ or 0 according to 'Yes' or 'No' to the sensitive question A and if $z_i = 0$ then to nonsensitive question B. Let π and π_B be the probabilities of 'Yes' responses to questions A and B, respectively. Then the conditional density of y_i, given z_i, π, π_B, is given by

$$f(y_i|z_i, \pi, \pi_B) = \pi^{y_i z_i}(1-\pi)^{(1-y_i)z_i}\pi_B^{y_i(1-z_i)}(1-\pi_B)^{(1-y_i)(1-z_i)}.$$

The parameter of interest is, of course, π. As a prior for π, a $Beta(\alpha, \beta)$ distribution is assumed. The probability π_B of a 'Yes' response to the unrelated question B may be known or may not. We consider the two cases separately.

a) When π_B is known: From $f(y_i|z_i, \pi, \pi_B)$, $f(z_i)$ and the prior of π, the conditional distribution of z_i is $Bernoulli(p_i^u)$, where

$$p_i^u = \begin{cases} p\pi \Big/ \Big(p\pi + (1-p)\pi_B\Big), & \text{if } y_i = 1 \\ p(1-\pi) \Big/ \Big(p(1-\pi) + (1-p)(1-\pi_B)\Big), & \text{if } y_i = 0, \end{cases}$$

independently for $i = 1, \ldots, n$, and the conditional distribution of π is $Beta(\alpha^u, \beta^u)$, where $\alpha^u = \sum_{i=1}^n y_i z_i + \alpha$ and $\beta^u = \sum_{i=1}^n (1 - y_i) z_i + \beta$.

b) When π_B is unknown: When π_B is unknown, it is also considered as a parameter. A typical prior for π_B is also a $Beta(\alpha^B, \beta^B)$ distribution. In this case, the conditional distribution of z_i and the conditional distribution of π are same as in the previous case since π_B is treated as a constant. But the conditional distribution of π_B is $Beta(\alpha^{u*}, \beta^{u*})$, where $\alpha^{u*} = \sum_{i=1}^n y_i(1 - z_i) + \alpha^B$ and $\beta^{u*} = \sum_{i=1}^n (1 - y_i)(1 - z_i) + \beta^B$.

3.3 Mangat and Singh's Two Stage Model.

In Mangat and Singh (1990)'s model, each interviewee is given two random devices R_1 and R_2. According to the outcome of R_1 the respondent answers to the sensitive question (with probability T), or proceed with random device R_2. And then according to the outcome of R_2 the respondent answers to the sensitive question (with probability p) or the complementary just as in Warner's model. Hence the interviewee is to use R_2 only when directed by R_1. This two-stage randomized response model can be made more efficient than simple Warner's model by selecting R_1 with appropriate probabilities for any practical choice of R_2 (Mangat and Singh (1990)).

To implement the Gibbs sampling algorithm in this two stage model, introduce two sets of Bernoulli random variables z_{1i} and z_{2i}, $i = 1, \ldots, n$, having probabilities T and p of success, respectively. Then if $z_{1i} = 1$, or, $z_{1i} = 0$ and $z_{2i} = 1$, let y_i represent an answer to the sensitive question, and if $z_{1i} = 0$ and $z_{2i} = 0$, let y_i represent an answer to the complementary. Thus, $f(y_i|\pi, z_{1i}, z_{2i})$ is proportional to

$$\left[\pi^{y_i}(1-\pi)^{1-y_i}\right]^{z_{1i}}\left[\pi^{y_i}(1-\pi)^{1-y_i}\right]^{(1-z_{1i})z_{2i}}\left[(1-\pi)^{y_i}\pi^{1-y_i}\right]^{(1-z_{1i})(1-z_{2i})}.$$

With a $Beta(\alpha, \beta)$ prior for π, it can be seen that the conditional distribution of z_{1i} is $Bernoulli(p_{1i}^t)$, where

$$p_{1i}^t = \begin{cases} 1/2, & \text{if } z_{2i} = 1 \\ T\pi\Big/\Big[T\pi + (1-T)(1-\pi)\Big], & \text{if } y_i = 1, z_{2i} = 0 \\ T(1-\pi)\Big/\Big[T(1-\pi) + (1-T)\pi\Big], & \text{if } y_i = 0, z_{2i} = 0 \,, \end{cases}$$

independently for $i = 1, \ldots, n$, and the conditional distribution of z_{2i} is $Bernoulli(p_{2i}^t)$, where

$$p_{2i}^t = \begin{cases} p, & \text{if } z_{1i} = 1 \\ p\pi\Big/\Big[p\pi + (1-p)(1-\pi)\Big], & \text{if } y_i = 1, z_{1i} = 0 \\ p(1-\pi)\Big/\Big[p(1-\pi) + (1-p)\pi\Big], & \text{if } y_i = 0, z_{1i} = 0, \end{cases}$$

independently for $i = 1, \ldots, n$. Finally the conditional distribution of π is $Beta(\alpha^t, \beta^t)$, where

$$\alpha^t = \sum_{i=1}^{n}\Big[y_i z_{1i} + y_i(1-z_{1i})z_{2i} + (1-y_i)(1-z_{1i})(1-z_{2i})\Big] + \alpha$$

$$\beta^t = \sum_{i=1}^{n}\Big[(1-y_i)z_{1i} + (1-y_i)(1-z_{1i})z_{2i} + y_i(1-z_{1i})(1-z_{2i})\Big] + \beta.$$

4. ANALYSIS OF RANDOMIZED POLYCHOTOMOUS RESPONSE DATA

In many practical situations there are t $(t \geq 2)$ categories in the response and at least one and at most $t-1$ of the categories are stigmatizing. For example, in a survey to estimate the proportion of unmarried mothers in a region the response categories may be (1) being pregnant after marriage (2) being pregnant at the time of marriage (3) being unmarried at the time of childbirth.

There are various randomized response models for polychotomous response data. In general Bayesian analysis of these models faces even more serious computational difficulties compared with randomized response models for binary data, because the number of parameters increases and restrictions on the parameters are more complicated in models for polychotomous data. However, the Gibbs sampling algorithm handles the computational difficulties in the polychotomous case as easily as in the binary case because in the Gibbs sampling algorithm all random generations are done in univariate space.

4.1 Extension of Warner's Model.

A direct extension of Warner's model for polychotomous response can be made as follows. Let $P_1, \ldots, P_s$ be s different permutations of $(1, \ldots, t)$, for $1 \leq s \leq t!$. And let a respondent choose one of P_i with probability p_i, using some random mechanism. Since P_i's are permutations, we may write $P_i = (i(1), \ldots, i(t))$, where $i(j)$ is the number at the j-th position in P_i. Given that P_i is chosen by a respondent, let the respondent answers j if he/she truly belongs to category $i(j)$. For instance, when $P_i = (2,3,1)$ was chosen by a respondent, the respondent answers 2 if he/she truly belongs to category 3. Note that if $s = 1$ then the model is not randomized and if $s = t!$ then the model is fully randomized. Also note that this is a generalization of the vector response model in Chaudri and Mukerjee(1988, Ch. 3, p.42).

In the above model, response Y_k of the kth individual follows a multinomial distribution with parameters 1 and $\{\sum_{i=1}^{t} p_i \pi_{i(j)}, j = 1, \ldots, t\}$. Thus as in the binary case, posterior distribution of $\{\pi_i, i = 1, \ldots, t\}$ is analytically intractable. But it can be handled by the Gibbs sampling algorithm as follows. Let a latent variable z_k represent the outcome of the random mechanism used by the kth individual to select a permutation, for $k = 1, \ldots, n$. In other words, if we define $\mathbf{1}_j^d$ the d-dimensional vector with the jth element equal to 1 and all other elements equal to 0, then let $P(z_k = \mathbf{1}_j^t) = p_j$. Then z_k follows a multinomial distribution with parameters 1 and $\{p_i,\ i = 1, \ldots, t\}$. And the joint probability density function of $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{z} = (z_1, \ldots, z_n)$ is given by

$$f(\boldsymbol{y}, \boldsymbol{z}|\pi) \propto \prod_{k=1}^{n} \prod_{i=1}^{s} \left(\prod_{j=1}^{t} \pi_{i(j)}^{I(y_k = \mathbf{1}_j^t)} p_i \right)^{I(z_k = \mathbf{1}_i^s)}.$$

Therefore, when a $Dirichlet(\alpha_1, \ldots, \alpha_t)$ prior is used for $\pi = (\pi_1, \ldots, \pi_t)$, the conditional distributions required for the Gibbs sampling algorithm are given by

$$z_k | y_k = \mathbf{1}_j^t, \pi \sim Multinomial\Big(1, \Big\{p_i \pi_{i(j)} \Big/ \sum_{j=1}^{t} p_i \pi_{i(j)}, i = 1, \ldots, t\Big\}\Big),$$

for $1 \leq j \leq t,\ k = 1, \ldots, n$ and

$$\pi | \boldsymbol{y}, \boldsymbol{z} \sim Dirichlet\Big(\Big\{\sum_{k=1}^{n} \sum_{i=1}^{s} \sum_{j=1}^{t} I\Big(i(j) = l, y_k = \mathbf{1}_j^t, z_k = \mathbf{1}_i^t\Big) + \alpha_l,\ l = 1, \ldots, t\Big\}\Big).$$

4.2 Extension of The Unrelated Question Model.

The unrelated question model can also be directly extended for polychotomous response. Let a respondent toss a coin with probability p of coming up a head. If the outcome is head, then the respondent answers honestly to the question A of interest. Otherwise the respondent answers to an unrelated question B which has the same number of categories as question A.

For an easy implementation of the Gibbs sampling algorithm in this model, let a variable z_k represent the outcome of tossing a coin by the kth individual, for $k = 1, \ldots, n$. Then, from similar calculations as in the previous sections, it can be easily seen that, when a $Dirichlet(\{\alpha_i, i = 1, \ldots, t\})$ prior is used for π, the conditional distribution of z_k, given $y_k = \mathbf{1}_i^t$ and π, is $Bernoulli(p\pi_i/(p\pi_i + (1-p)p_i))$, where p_i is the probability of the ith category in the unrelated question B, for $1 \le i \le t$, and

$$\pi|\boldsymbol{y}, \boldsymbol{z} \sim Dirichlet\Big(\Big\{\sum_{k=1}^{n} I\Big(y_k = \mathbf{1}_i^t, z_k = 1\Big) + \alpha_i, i = 1, \ldots, t\Big\}\Big).$$

4.3 Dichotomization of Polychotomous Response Data.

Many randomized strategies for polychotomous responses described in Chaudri and Mukerjee (1988) transform a polychotomous response question into several binary response questions. We describe some of the models and show that these models can also be handled by using the Gibbs sampling algorithm with appropriate latent variables.

A Dichotomized Warner's Model for Polychotomous Response.

Eriksson(1973) suggests the following randomized response model (Chaudri and Mukerjee (1988)). By simple random sampling, $t-1$ classes of sizes $n_1, \ldots, n_{t-1}$ are drawn independently from the population. Each interviewee in the ith class tosses a coin with probability p_i of coming up a head and announces membership in the i-th category of the question of interest if the outcome of tossing a coin is head and denounces membership in the ith category if the outcome of tossing a coin is tail. So individuals in the ith class only answers about the ith category and the probability of a 'Yes' response in the ith class is $p_i\pi_i + (1-p_i)(1-\pi_i)$, for $i = 1, \ldots, t-1$.

To handle this model by the Gibbs sampling algorithm, let z_{ik} be 1 or 0 according to the outcome of tossing a coin for the kth interviewee in the ith class, for $k = 1, \ldots, n_i$ and $i = 1, \ldots, t-1$. Then the joint density function of $\boldsymbol{y} = \{y_{ik}, k = 1, \ldots, n_i,\ i = 1, \ldots, t-1\}$ and $\boldsymbol{z} = \{z_{ik}, k = 1, \ldots, n_i,\ i = 1, \ldots, t-1\}$, given $\pi_1, \ldots, \pi_{t-1}$, is proportional to

$$\prod_{i=1}^{t-1}\prod_{k=1}^{n_i}\Big(p_i\pi_i^{y_{ik}}(1-\pi_i)^{1-y_{ik}}\Big)^{z_{ik}}\Big((1-p_i)(1-\pi_i)^{y_{ik}}\pi_i^{1-y_{ik}}\Big)^{(1-z_{ik})}.$$

Thus, a natural-conjugate prior for $\pi_1, \ldots, \pi_{t-1}$ is that each π_i independently follows a $Beta(\alpha_i, \beta_i)$ distribution subject to the restriction

$$0 < \sum_{i=1}^{t-1} \pi_i < 1. \tag{4.1}$$

(Uniform prior on the region (4.1) if $\alpha_i = \beta_i = 1$, for all $i = 1, \ldots, t-1$.) With this natural-conjugate prior, it can be seen that the conditional distribution of z_{ik} is $Bernoulli(p^{pw})$, where

$$p^{pw} = \begin{cases} p_i\pi_i \Big/ \Big(p_i\pi_i + (1-p_i)(1-\pi_i)\Big), & \text{if } y_{ik} = 1 \\ p_i(1-\pi_i) \Big/ \Big(p_i(1-\pi_i) + (1-p_i)\pi_i\Big), & \text{if } y_{ik} = 0, \end{cases}$$

independently for $k = 1, \ldots, n_i$, $i = 1, \ldots, t-1$, and the conditional distribution of π_i is $Beta(\alpha^{pw}, \beta^{pw})$, where

$$\begin{aligned} \alpha^{pw} &= \sum_{i=1}^{t-1}\sum_{k=1}^{n_i}\Big[y_{ik}z_{ik} + (1-y_{ik})(1-z_{ik})\Big] + \alpha_i \\ \beta^{pw} &= \sum_{i=1}^{t-1}\sum_{k=1}^{n_i}\Big[y_{ik}(1-z_{ik}) + (1-y_{ik})z_{ik}\Big] + \beta_i, \end{aligned}$$

independently for $i = 1, \ldots, t-1$, subject to the restriction (4.1). All the conditional distributions are of convenient forms and the restriction can be handled easily in the Gibbs sampling algorithm (Gelfand and Smith (1992)).

A Dichotomized Unrelated Question Model for Polychotomous Response.

The unrelated question model for polychotomous response by Greenberg et. al. (1969) can be considered as follows. Draw t classes with sizes $n_1, \ldots, n_t$ from the population by simple random sampling with replacement. An individual in the ith class chooses the jth category with probability p_{ij}, for $j = 1, \ldots, t$. Given that the jth category is chosen, for $1 \le j \le t-1$, the interviewee announces membership in the jth category. If the tth category is chosen, the interviewee announces membership in a binary unrelated question with probability π_y of a 'Yes' response. Thus, this model also transforms a polychotomous response question into several binary response questions. And the probability of a 'Yes' response in the ith class is $\sum_{j=1}^{t-1} p_{ij}\pi_j + p_{it}\pi_y$.

In this model, if we let a variable z_{ik} represent the category chosen, i.e., let $z_{ik} = \mathbf{1}_j^t$ if the jth category is chosen, by the kth interviewee in the ith class, then clearly z_{ik} follows a multinomial distribution with parameters 1 and $\{p_{ij}, j = 1, \ldots, t\}$, and the joint density function of $\boldsymbol{y}=(y_{ik}, k = 1, \ldots, n_i, i = 1, \ldots, t)$ and $\boldsymbol{z}=(z_{ik}, k = 1, \ldots, n_i, i = 1, \ldots, t)$, given $\pi_1, \ldots, \pi_{t-1}$, is proportional to

$$\prod_{i=1}^{t}\prod_{k=1}^{n_i}\left[\prod_{j=1}^{t-1}\left(p_{ij}\pi_j^{y_{ik}}(1-\pi_j)^{1-y_{ik}}\right)^{I(z_{ik}=\mathbf{1}_j^t)}\right]\left(p_{it}\pi_y^{y_{ik}}(1-\pi_y)^{(1-y_{ik})}\right)^{I(z_{ik}=\mathbf{1}_t^t)},$$

subject to the restriction (4.1). Again, natural conjugate priors for $\pi_1, \ldots, \pi_{t-1}$ are independent *Beta* distributions with respective parameters (α_i, β_i), $i = 1, \ldots, t-1$, subject to the restriction (4.1). With this conjugate prior, necessary conditional distributions for the Gibbs sampling algorithm can be derived as

$$z_{ik}|y_{ik}, \pi \sim Multinomial\Big(1, \{p_j^*,\ j = 1, \ldots, t\}\Big),$$

where

$$p_j^* = \frac{p_{ij}\pi_j^{y_{ik}}(1-\pi_j)^{1-y_{ik}}}{\sum_{j=1}^{t-1} p_{ij}\pi_j^{y_{ik}}(1-\pi_j)^{1-y_{ik}} + p_{it}\pi_y^{y_{ik}}(1-\pi_y)^{1-y_{ik}}},$$

for $j = 1, \ldots, t-1$ and

$$p_t^* = \frac{p_{it}\pi_y^{y_{ik}}(1-\pi_y)^{1-y_{ik}}}{\sum_{j=1}^{t-1} p_{ij}\pi_j^{y_{ik}}(1-\pi_j)^{1-y_{ik}} + p_{it}\pi_y^{y_{ik}}(1-\pi_y)^{1-y_{ik}}},$$

independently for $k = 1, \ldots, n_i,\ i = 1, \ldots, t$. And

$$\pi_j|\boldsymbol{y}, \boldsymbol{z}, \pi_l, l \neq j \sim Beta\Big(\sum_{i=1}^{t-1}\sum_{k=1}^{n_i} y_{ik}I(z_{ik} = \mathbf{1}_j^t)+\alpha_j, \sum_{i=1}^{t-1}\sum_{k=1}^{n_i}(1-y_{ik})I(z_{ik} = \mathbf{1}_j^t)+\beta_j\Big),$$

independently for $j = 1, \ldots, t-1$, subject to the restriction (4.1).

When π_y is also unknown, it is treated as a parameter as in the binary case. With a $Beta(\alpha_y, \beta_y)$ prior for π_y, the conditional distribution of π_y can be easily derived as

$$Beta\Big(\sum_{i=1}^{t-1}\sum_{k=1}^{n_i} y_{ik} I(z_{ik} = \mathbf{1}_t^t) + \alpha_y,\ \sum_{i=1}^{t-1}\sum_{k=1}^{n_i}(1 - y_{ik}) I(z_{ik} = \mathbf{1}_t^t) + \beta_y\Big),$$

and the generation of π_y is added to the Gibbs sampling cycle as in the binary case.

5. EXAMPLE

The following example is given in Chaudri and Mukerjee (1988, Example 3.2). Three classes of sizes n_1, n_2, n_3 are drawn independently by simple random sampling with replacement. There are three decks of 80 cards. On each card, one of the following questions, "Do you support candidate 1?", "Do you support candidate 2?", and "Do you like tea to coffee?", is written. The first deck has 30 cards of the first question, 30 of the second, and 20 of the third. The second deck has 30, 20, and 30 cards of the three questions, respectively. And the third deck has 20, 30, and 30 cards of the three questions, respectively. So there are three categories and the third category is unrelated to the first two categories of interest. Each interviewee in the ith class randomly draws a card from the i-th deck and responds to the question on the card, for $i = 1, 2, 3$. This is the dichotomized unrelated question model for polychotomous response described in Subsection 4.3.2.

In the original example in Chaudri and Mukerjee(1988), $n_1 = 100, n_2 = 150$, and $n_3 = 200$ and there were 37, 60, and 72 'Yes' responses in three classes, respectively. When the Gibbs sampling algorithm was applied as described in this article to the data with uniform priors for unknown probabilities of 'Yes' responses to the three categories, it showed the same results with those from a normal approximation with parameters given in Chaudri and Mukerjee(1988). This implies that the normal approximation is good in this example and the Gibbs sampling algorithm verifies it. And the normal approximation seems to be good because the sample sizes are large enough.

But there are situations where large samples are not affordable. In those situations the normal approximation would not be good and numerical schemes would be necessary for accurate posterior inferences. To illustrate the usefulness of the Gibbs sampling algorithm as a numerical scheme when sample sizes are not large, the example is slightly modified as follows. The sizes of the three classes are reduced to 14, 20, and 50, respectively. And we suppose that the numbers of 'Yes' responses in the three classes are 5, 8, and 18, respectively. So the proportions of 'Yes' responses are about the same but the sample sizes are reduced significantly.

For a fair comparison between Bayesian analysis and the normal approximation in Chaudri and Mukerjee(1988), uniform priors were used for all unknown probabilities in the model. The Gibbs sampling scheme was applied to the new artificial data, generating random samples from the conditional distributions given in Subsection 4.3.2. The one long-run strategy was adopted and the first 1,000 random samples were dropped as a warm-up time. Random samples were taken in every 5 cycles, until sufficiently large number, 6,000, of samples were obtained to ensure stability. So, total of 31,000 Gibbs sampling cycles were required to obtain a good sample of size 6,000.

Figures 1 and 2 show marginal density functions of proportions of supporters for candidates 1 and 2, respectively. The solid lines represent the marginal density functions from the Gibbs sampling algorithm and the dashed lines represent those from the normal approximation with parameters obtained as in Chaudri and Mukerjee(1988). The true marginal density functions were calculated using a simple quadrature method and they are shown as dotted lines in the figures. It can be clearly seen that the true density functions are invisibly different than those from the Gibbs sampling algorithm.

6. SUMMARY

Bayesian analyses of various randomized response models using the Gibbs sampling algorithm are described. A great advantage of using the Gibbs sampling algorithm in randomized response models is "the simplicity" in implementation. No expert skills is necessary here. Just introducing some latent variables gives all the full conditional distributions in very convenient forms for easy implementation of the Gibbs sampling algorithm. With samples obtained from the Gibbs sampling algorithm, one can do full analysis about posterior distributions of the parameters of interest.

The simplicity of the Gibbs sampling algorithm is especially attractive here because parameters are restricted. In most analytic approximations and numerical integration schemes, restricted parameters prohibit easy implementation of the schemes. However, because simulation is actually done in one dimensional space in each step of a cycle in the Gibbs sampling algorithm, restriction can be handled easily by the inverse cumulative distribution function technique.

ACKNOWLEDGEMENTS

The author thanks Professor B.S. So and two anonymous referees for helpful comments.

REFERENCES

(**1**) Albert, J. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data: A Gibbs Sampling Approach. *Journal of the American Statistical Association*, **88**, **669–679**.

(**2**) Chaudri, A. and Mukerjee, R. (1988). *Randomized Response: Theory and Techniques*, New York: Marcel Dekker.

(**3**) Eriksson, S. (1973). Randomized Interviews for Sensitive Questions. Ph. D. Thesis, Gothenberg University.

(4) Gelfand, A. and Smith, A.F.M. (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, **398–409**.

(5) Gelfand, A. and Smith, A.F.M. (1992). Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling. *Journal of the American Statistical Association*, **87**, **523–532**.

(6) Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**, **457–472**.

(7) Geyer, C. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, **7**, **473–482**.

(8) Greenberg, B.G., Abul-Ela, Abdel-Latif, A., Simmons, W.R., and Horvitz, B.G. (1969). The Unrelated Question Randomized Response Model. *Journal of the American Statistical Association*, **64**, **520–539**.

(9) Horvitz, D.G., Shah, B.V., and Simmons, W.R. (1967). The Unrelated Randomized Response Model. *Proceedings of the Social Statistics Section*, American Statistical Association, **65–72**.

(10) Mangat, N.S. and Singh, R. (1990). An Alternative Randomized Response Procedure. *Biometrika*, **77**, **439–442**.

(11) O'Hagan, A. (1987). Bayes Linear Estimators for Randomized Response Models. *Journal of the American Statistical Association*, **82**, **580–585**.

(12) Pitz, G.F. (1980). Bayesian Analysis of Random Response Models. *Psychological Bulletin*, **87**, **209–212**.

(13) Ritter, C. and Tanner, M.A. (1993). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddly Gibbs samper. *Journal of the American Statistical Association*.

(14) Warner, S.L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, **60**, **63–69**.

(15) Winkler, R.L. and Franklin, L.A. (1979). Warner's Randomized Response Model: A Bayesian Approach, *Journal of the American Statistical Association*, **74**, **207–214**.

Figure 1. Marginal Density Functions of π_1

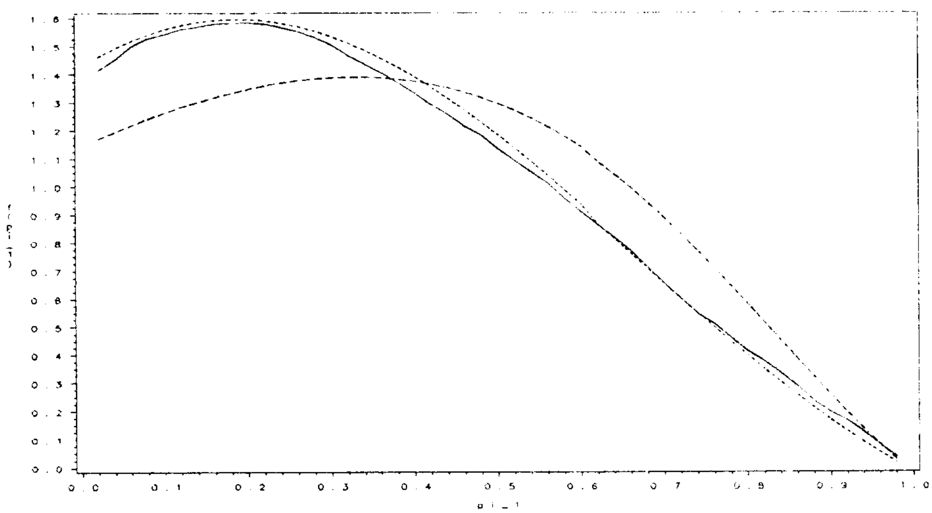

Figure 2. Marginal Density Functions of π_2