The Korean Communications in Statistics **Vol. 14** No. 1, 2007, pp. 121–131

A Closed-Form Bayesian Inferences for Multinomial Randomized Response Model

Tae-Young Heo[1] and Jong-Min Kim[2]

Abstract

In this paper, we examine the problem of estimating the sensitive characteristics and behaviors in a multinomial randomized response model using Bayesian approach. We derived a posterior distribution for parameter of interest for multinomial randomized response model. Based on the posterior distribution, we also calculated a credible intervals and mean squared error (MSE). We finally compare the maximum likelihood estimator and the Bayes estimator in terms of MSE.

Keywords: Randomized response; multinomial model; Bayesian inference; Dirichlet distribution; mean squared error.

1. Introduction

The frequency of socially undesirable, embarrassing, or prohibited acts or attitudes is usually underestimated in surveys. A randomized response (RR) technique is a procedure for collecting the information on sensitive characteristics without exposing the identity of the respondent. The RR technique was originally proposed by Warner (1965) as an alternative survey technique for socially undesirable or incriminating behavior questions. With the many benefits of Dirichlet prior in Bayesian framework, we propose a Bayesian multinomial approach to an extension of the binomial randomized response model suggested by Bar-Lev *et al.* (2005) and Kim *et al.* (2006).

1) Senior researcher, Electronics and Telecommunications Research Institute, 161 Gajeong-dong, Useong-gu, Daejeon 505-700, Korea.
Correspondence : taeyoung.heo@etri.re.kr

2) Associate professor, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN 56267, U.S.A.
E-mail : jongmink@morris.umn.edu

One of the most important problems in RR model is an estimation of the sensitive characteristic. To protect a respondent privacy, and increase response rates, RR models require the interviewee to give a Yes or No answer either to the sensitive question or to its negative depending on the outcome of a randomizing device not reported to the interviewer. However this technique has not been extended to Bayesian multinomial RR model, which is more applicable in practice. When the sample size is small and a number of cells contain only few observations or even no observations, maximum likelihood method is not appropriate.

In this paper we consider parameter estimation of sensitivity prevalence which usually has a rare trait proportion in a Bayesian framework. We start in Section 2 with a short review of Kim and Warde's (2005) multinomial RR model and a derivation of maximum likelihood estimator (MLE). In Section 3, we propose and demonstrate how to apply Bayesian approach to multinomial RR model and derive a posterior distribution and credible intervals, and compare MLE and Bayes estimator in terms of mean squared error (MSE).

2. Multinomial Randomized Response Model

With a discrete quantitative randomized response technique model using the Hopkins randomization device (Liu and Chow, 1976), Kim and Warde (2005) propose a multinomial RR model and derive estimators and their properties. We follow Kim and Warde's (2005) multinomial model set up which explicitly assumed a multinomial model for a single sensitive variable, denoted as A.

Suppose that there are two different colors of balls, red and green, in the device and that each of the green balls contains a discrete number $\{1, 2, \ldots, k\}$. All green balls represent a set of non-sensitive categories, $B = \{B_1, B_2, \ldots, B_k\}$ and all the values of A are also included.

We assume that each of the t individuals belongs to one of k mutually exclusive. The exhaustive categories $T = \{T_1, T_2, \ldots, T_k\}$ consisting of sensitive categories $A = \{A_1, A_2, \ldots, A_k\}$ and non-sensitive categories $B = \{B_1, B_2, \ldots, B_k\}$ and $T_i = A_i + B_i$, $i = 1, 2, \ldots, k$, is the sum of the number of sensitive and nonsensitive at i-th category.

Let t_i denote the random quantity in a category T_i so that $n = \sum_{i=1}^{k} t_i$. The random quantities a_i and b_i are defined similarly, so that $a = \sum_{i=1}^{k} a_i$ and $b = \sum_{i=1}^{k} b_i$, and $t_i = a_i + b_i$. Our goal, then, is to estimate $\pi_1, \pi_2, \ldots, \pi_k$, the proportions in the population associated with the sensitive categories $A = \{A_1, A_2, \ldots, A_k\}$. Based on the number of green balls in the device, $p_{b_i} = q_i/g$

is the proportion of green balls with number i for $i = 1, 2, \ldots, k$, where q_i is the number of green balls that contain number i, and $g = \sum_{i=1}^{k} q_i$; that is, the quantities p_{b_i} are known in advance.

Let $p_{t_1}, p_{t_2}, \ldots, p_{t_k}$ denote the proportions in the population who are in categories $T = \{T_1, T_2, \ldots, T_k\}$. With n different interviewees using the Hopkins' device, b, the total number of people who are in the non-sensitive categories $B = \{B_1, B_2, \ldots, B_k\}$, is a random quantity with expected value $E[b] = ng/(r+g)$ where r denotes the number of red balls in the device. As $b_1, b_2, \ldots, b_k$ are also random quantities with expected value $E[b_i] = nq_i/(r+g)$; thus, it follows that $b_k = b - (b_1 + b_2 + \cdots + b_{k-1})$. We assume the distributions of T, A, and B are as follows:

$$T = \{T_1, T_2, \ldots, T_{k-1}\} \sim \text{Multinomial}(n, p_{t_1}, p_{t_2}, \ldots, p_{t_{k-1}}) = \frac{n!}{\prod_{i=1}^{k} t_i!} \prod_{i=1}^{k} p_{t_i}^{t_i},$$

$$A = \{A_1, A_2, \ldots, A_{k-1}\} \sim \text{Multinomial}(a, \pi_1, \pi_2, \ldots, \pi_{k-1}) = \frac{a!}{\prod_{i=1}^{k} a_i!} \prod_{i=1}^{k} p_{a_i}^{a_i},$$

$$B = \{B_1, B_2, \ldots, B_{k-1}\} \sim \text{Multinomial}(b, p_{b_1}, p_{b_2}, \ldots, p_{b_{k-1}}) = \frac{b!}{\prod_{i=1}^{k} b_i!} \prod_{i=1}^{k} p_{b_i}^{b_i}.$$

Suppose that $T = A + B$ is fixed and that respondents give truthful answers to both the sensitive and non-sensitive questions. Then, for random quantities a and b, we derive π_i as follows:

$$\pi_i = \frac{(r+g)p_{t_i} - q_i}{r},$$

where $p_{t_i} = t_i/n$.

If a random sample of size n is drawn and n_i is the number of respondents answering i and let $\widehat{p}_{t_i} = n_i/n$ denote the proportion of respondents answering i. Let $\widehat{\pi}_{\mathrm{M}_i}$ denote the estimates of π_i, it follows that

$$\widehat{\pi}_{\mathrm{M}_i} = \frac{(r+g)\widehat{p}_{t_i} - q_i}{r}. \tag{2.1}$$

By invariance property of MLE, the MLE of sensitive characteristic, $\widehat{\pi}_{\mathrm{M}_i}$, is turned out to be maximum likelihood estimator (MLE) of π_i. The estimates of the variance and covariance, respectively, are given by

$$\widehat{v}(\widehat{\pi}_{\mathrm{M}_i}) = \left(\frac{r+g}{r}\right)^2 \times \frac{\widehat{p}_{t_i}(1 - \widehat{p}_{t_i})}{n}$$

and

$$\widehat{\mathrm{Cov}}(\widehat{\pi}_{\mathrm{M}_i}, \widehat{\pi}_{\mathrm{M}_j}) = -\left(\frac{r+g}{r}\right)^2 \times \frac{\widehat{p}_{t_i}\widehat{p}_{t_j}}{n},$$

for $i \neq j$.

In Equation (2.1), if $\widehat{\pi}_{\mathrm{M}_i} < q_i/(r+g)$, then $\widehat{\pi}_{\mathrm{M}_i}$ can be negative. When we find the maximum likelihood estimator (MLE) of π_i, we should be careful of putting the number of red balls, r, and green balls, $g = \sum_{i=1}^{k} q_i$, in the random device. In this paper, we define the restricted MLE to be

$$\widehat{\pi}^*_{\mathrm{M}_i} = \frac{(r+g)\widehat{p}^*_{ti} - q_i}{r},$$

where

$$\widehat{p}^*_{t_i} = \begin{cases} \frac{q_i}{r+g}, & \frac{n_i}{n} \leq \frac{q_i}{r+g} \\ \widehat{p}_{t_i}, & \frac{q_i}{r+g} < \frac{n_i}{n} \leq 1. \end{cases}$$

3. Bayesian Inference for Multinomial Randomized Response Model

Assuming that a researcher wants to assess a sensitive characteristic and that he uses a question to which the answers are more than two categories. Suppose that in each of k categories, individuals are independently classified into one of T_i $(i = 1, \ldots, k)$ categories. Therefore $T = (T_1, \ldots, T_k)$ has a multinomial distribution with parameters n and $\tilde{p}_t = (p_{t_1}, \ldots, p_{t_k})$. So it follows that based on the observed values $\tilde{t} = (t_1, \ldots, t_k)$, the likelihood function of T given that $\tilde{p}_t = (p_{t_1}, \ldots, p_{t_k})$ is

$$f_{T|P_t}(\tilde{t}|\tilde{p}_t) = \frac{n!}{\prod_{i=1}^{k} t_i} \prod_{i=1}^{k} p_{t_i}^{t_i},$$

where $p_{t_i} \in \Theta_{r,g,q_i} = \{q_i/(r+g), (r+q_i)/(r+g)\}$. Based on $p_{t_i} = (r\pi_i + q_i)/(r+g)$ in Section 2, we derive the likelihood function of T given that $\tilde{\pi} = (\pi_1, \ldots, \pi_k)$ as follows:

$$f_{T|\Pi}(\tilde{t}|n, r, g, \tilde{q}, \tilde{\pi}) = \frac{n!}{\prod_{i=1}^{k} t_i} \prod_{i=1}^{k} \left(\frac{r}{r+g}\pi_i + \frac{q_i}{r+g}\right)^{t_i},$$

where $0 < \pi_i < 1$ and $\tilde{q} = (q_1, \ldots, q_k)$.

The proper conjugate prior for $\tilde{p}_t = (p_{t_1}, \ldots, p_{t_k})$ is the Dirichlet distribution denoted by $\text{Dirichlet}_1(\alpha_1, \ldots, \alpha_k)$ with density

$$f_{P_t}(\tilde{p}_t|\tilde{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}p_{t_i}^{\alpha_i-1},$$

where $\alpha_i > 0$ for $i = 1, \ldots, k$, $p_{t_i} \in \Theta_{r,g,q_i}$, and $\Gamma(\cdot)$ is the gamma distribution. By the Jacobian transformation of $p_{t_i} = (r\pi_i + q_i)/(r+g)$, a prior density for $\tilde{\pi} = (\pi_1, \ldots, \pi_k)$ is the Dirichlet distribution with density

$$f_{\Pi}(\tilde{\pi}|\tilde{\alpha}, r, g, \tilde{q}) = \left(\frac{r}{r+g}\right)^{k-1}\frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}\left(\frac{r}{r+g}\pi_i + \frac{q_i}{r+g}\right)^{\alpha_i-1},$$

where $\alpha_i > 0$ for $i = 1, \ldots, k$, and $\pi_i \in (0,1)$.

Therefore, the posterior distribution of T and P_t is as follows:

$$\begin{aligned}
f_{P_t,T}(\tilde{t}, \tilde{p}_t|\tilde{\alpha}) &= f_{T|P_t}(\tilde{t}|\tilde{p}_t) \times f_{P_t}(\tilde{p}_t|\tilde{\alpha}) \\
&= \frac{n!}{\prod_{i=1}^{k}t_i}\prod_{i=1}^{k}p_{t_i}^{t_i} \times \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}p_{t_i}^{\alpha_i-1} \\
&= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\prod_{i=1}^{k}p_{t_i}^{t_i+\alpha_i-1},
\end{aligned}$$

for $p_{t_i} \in \Theta_{r,g,q_i}$. The marginal distribution of T is given by

$$\begin{aligned}
f_T(\tilde{t}|n, \tilde{\alpha}) &= \int_{\Theta_{r,g,q_i}} f_{P_t,T}(\tilde{t}, \tilde{p}_t|\tilde{\alpha})d\tilde{p}_t \\
&= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\prod_{i=1}^{k}\int_{\Theta_{r,g,q_i}}p_{t_i}^{t_i+\alpha_i-1}dp_{t_i} \\
&= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\frac{\prod_{i=1}^{k}\Gamma(t_i+\alpha_i)}{\Gamma\left(\sum_{i=1}^{k}(t_i+\alpha_i)\right)}.
\end{aligned}$$

The conditional distribution of P_t given T is

$$f_{P_t|T}(\tilde{p}_t|\tilde{t},\tilde{\alpha},n) = \frac{f_{\tilde{\pi},T}(\tilde{t},\tilde{\pi}|\tilde{\alpha},n,r,g,\tilde{q})}{f_T(\tilde{t}|\tilde{\alpha},n,r,g,\tilde{q})} = \frac{\Gamma\left(\sum_{i=1}^{k}(t_i+\alpha_i)\right)}{\prod_{i=1}^{k}\Gamma(t_i+\alpha_i)}\prod_{i=1}^{k}p_{t_i}^{t_i+\alpha_i-1},$$

which means that $P_t|T$ has a Dirichlet$_1(t_1+\alpha_1,\ldots,t_k+\alpha_k)$.

Similarly, the posterior distribution of T and Π is as follows:

$$\begin{aligned} f_{\Pi,T}(\tilde{t},\tilde{\pi}|n,r,g,\tilde{q},\tilde{\alpha}) &= f_{T|\Pi}(\tilde{t}|n,r,g,\tilde{q},\tilde{\pi}) \times f_{\Pi}(\tilde{\pi}|\tilde{\alpha},r,g,\tilde{q}) \\ &= \frac{n!}{\prod_{i=1}^{k}t_i}\prod_{i=1}^{k}\left(\frac{r}{r+g}\pi_i+\frac{q_i}{r+g}\right)^{t_i} \\ &\quad\times\frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\left(\frac{r}{r+g}\right)^{k-1}\prod_{i=1}^{k}\left(\frac{r}{r+g}\pi_i+\frac{q_i}{r+g}\right)^{\alpha_i-1} \\ &= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\left(\frac{r}{r+g}\right)^{k-1} \\ &\quad\times\prod_{i=1}^{k}\left(\frac{r}{r+g}\pi_i+\frac{q_i}{r+g}\right)^{t_i+\alpha_i-1} \end{aligned}$$

for $0<\pi_i<1$. The marginal distribution of T is given by

$$\begin{aligned} f_T(\tilde{t}|n,r,g,\tilde{q},\tilde{\alpha}) &= \int_{(0,1)^k} f_{\Pi,T}(\tilde{t},\tilde{\pi}|n,r,g,\tilde{q},\tilde{\alpha})d\tilde{\pi} \\ &= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\left(\frac{r}{r+g}\right)^{k-1} \\ &\quad\times\prod_{i=1}^{k}\int_0^1\left(\frac{r}{r+g}\pi_i+\frac{q_i}{r+g}\right)^{t_i+\alpha_i-1}d\pi_i \end{aligned}$$

$$= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\prod_{i=1}^{k}\int_{\Theta_{r,g,q_i}} p_{t_i}^{t_i+\alpha_i-1}dp_{t_i}$$

$$= \frac{n!\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\left(\prod_{i=1}^{k}t_i\right)\left(\prod_{i=1}^{k}\Gamma(\alpha_i)\right)}\frac{\prod_{i=1}^{k}\Gamma(t_i+\alpha_i)}{\Gamma\left(\sum_{i=1}^{k}t_i+\alpha_i\right)}.$$

Thus, the conditional distribution of $\Pi|T$ is

$$f_{\Pi|T}(\tilde{\pi}|\tilde{t},\tilde{\alpha},n,r,g,\tilde{q}) = \frac{f_{\Pi,T}(\tilde{t},\tilde{\pi}|\tilde{\alpha},n,r,g,\tilde{q})}{f_T(\tilde{t}|\tilde{\alpha},n,r,g,\tilde{q})}$$

$$= \left(\frac{r}{r+g}\right)^{k-1}\frac{\Gamma\left(\sum_{i=1}^{k}(t_i+\alpha_i)\right)}{\prod_{i=1}^{k}\Gamma(t_i+\alpha_i)} \times\prod_{i=1}^{k}\left(\frac{r}{r+g}\pi_i+\frac{q_i}{r+g}\right)^{t_i+\alpha_i-1}.$$

Because $P_t|T$ has a Dirichlet$(t_1+\alpha_1,\ldots,t_k+\alpha_k)$, its marginal has a Beta density function as follows:

$$f_{P_{t_i}|T}(p_{t_i}|\tilde{t},\tilde{\alpha},n) = \frac{\Gamma\left(\sum_{i=1}^{k}(t_i+\alpha_i)\right)}{\Gamma(t_i+\alpha_i)\Gamma\left(\sum_{j=1}^{k}(t_j+\alpha_j)-(t_i+\alpha_i)\right)} \times p_{t_i}^{t_i+\alpha_i-1}(1-p_{t_i})^{(\sum_{j=1}^{k}(t_j+\alpha_j)-(t_i+\alpha_i)-1)},$$

for $i=1,\ldots,k$.

Considering the squared error loss function, $L(p_{t_i},a)=(p_{t_i}-a)^2$, we calculate the Bayesian estimate of p_{t_i} which is the mean of posterior $f_{P_{t_i}|T}(p_{t_i}|\tilde{t},\tilde{\alpha},n)$.

A simple closed-form expression for $\widehat{p}_{tB_i}$ is given by

$$\widehat{p}_{tB_i} = E\left[p_{t_i}|T\right] = \frac{\text{Beta}\left(t_i+\alpha_i+1,\sum_{j=1}^{k}(t_j+\alpha_j)-(t_i+\alpha_i)\right)}{\text{Beta}\left(t_i+\alpha_i,\sum_{j=1}^{k}(t_j+\alpha_j)-(t_i+\alpha_i)\right)},$$

where

$$\text{Beta}(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Since the facts that $n=\sum_{j=1}^{k}t_j$ and $\beta_i=(\sum_{j=1}^{k}\alpha_j)-\alpha_i$, we obtain a more simpler form for $\widehat{p}_{tB_i}$ as follows:

$$\widehat{p}_{tB_i}=\frac{\text{Beta}\left(t_i+\alpha_i+1,n-t_i+\beta_i\right)}{\text{Beta}\left(t_i+\alpha_i,n-t_i+\beta_i\right)}.$$

We finally obtain the estimate of π_{B_i} as follows

$$\widehat{\pi}_{\text{B}_i}=E\left[p_{t_i}|T\right]=\frac{(r+g)\widehat{p}_{t\text{B}_i}-q_i}{r}.$$

The variance of $\widehat{\pi}_{\text{B}_i}$ is given by

$$\begin{aligned}\text{Var}(\widehat{\pi}_{\text{B}_i})&=\left(\frac{r+g}{r}\right)^2\times\text{Var}(\widehat{p}_{tB_i})\\&=\left(\frac{r+g}{r}\right)^2\times\left[E(\widehat{p}^{\,2}_{tB_i})-E(\widehat{p}_{tB_i})^2\right]\\&=\left(\frac{r+g}{r}\right)^2\times\left[\frac{\text{Beta}\left(t_i+\alpha_i+2,n-t_i+\beta_i\right)}{\text{Beta}\left(t_i+\alpha_i,n-t_i+\beta_i\right)}\right.\\&\quad\left.-\left(\frac{\text{Beta}\left(t_i+\alpha_i+1,n-t_i+\beta_i\right)}{\text{Beta}\left(t_i+\alpha_i,n-t_i+\beta_i\right)}\right)^2\right].\end{aligned}$$

By constructing the posterior distribution, we calculate a $100(1-\alpha)\%$ credible intervals for p_{t_i} as follows

$$\int_{L_{p_{t_i}}}^{U_{p_{t_i}}}f_{P_{t_i}|T}(p_{t_i}|\tilde{t},\tilde{\alpha},n)dp_{t_i}=1-\alpha,$$

where $0<L_{p_{t_i}}<U_{p_{t_i}}<1,$ for $i=1,2,\ldots,k$.

We find a nice closed-form expression for the equal-tail credible interval. The lower bound (LB) of p_{t_i} may find by solving following equation

$$\begin{aligned}\frac{\alpha}{2}&=\int_0^{p_{Lt_i}}f_{P_{t_i}|T}(p_{t_i}|\tilde{t},\tilde{\alpha},n)dp_{t_i}\\&=\int_0^{L_{p_{t_i}}}\frac{\Gamma(n+\alpha_i+\beta_i)}{\Gamma(t_i+\alpha_i)\Gamma(n-t_i+\beta_i)}p_{t_i}^{t_i+\alpha_i-1}(1-p_{t_i})^{n-t_i+\beta_i-1}dp_{t_i}.\end{aligned}$$

We obtain a LB as $L_{p_{t_i}}=\text{Beta}(\alpha/2;t_i+\alpha_i,n-t_i+\beta_i)$, where $\text{Beta}(\gamma;a,b)$ denotes the γ quantile of the Beta distribution. Similarly, the upper bound (UB) of p_{t_i}

is $U_{p_1} = 1 - \text{Beta}(1-\alpha/2; t_i + \alpha_i, n - t_i + \beta_i)$. Using $p_{t_i} = (r\pi_i + q_i)/(r+g)$, we finally derived the following LB and UB of π_i as follows:

$$L_{\pi_i} = \frac{(r+g) \times L_{p_{t_i}} - q_i}{r} = \frac{(r+g) \times \text{Beta}(\frac{\alpha}{2}; t_i + \alpha_i, n - t_i + \beta_i) - q_i}{r}.$$

Similarly, a upper bound of π_i is

$$U_{\pi_i} = \frac{(r+g) \times \text{Beta}(1 - \frac{\alpha}{2}; t_i + \alpha_i, n - t_i + \beta_i) - q_i}{r}.$$

We also calculate the mean squared error (MSE) for fixed as follows $\widehat{\pi}_{\text{B}_i}$ and $\widehat{\pi}_{\text{M}_i}$ are given by

$$\begin{aligned}\text{MSE}(\widehat{\pi}_{\text{B}_i}) &= E_{T|\Pi_i}\left[(\widehat{\pi}_{\text{B}_i} - \pi_i)^2\right] \\ &= \sum_{t_i=0}^{n} (\widehat{\pi}_{\text{B}_i} - \pi_i)^2 \\ &\quad \times \binom{n}{t_i} \left(\frac{r}{r+g}\pi_i + \frac{q_i}{r+g}\right)^{t_i} \left[1 - \left(\frac{r}{r+g}\pi_i + \frac{q_i}{r+g}\right)\right]^{n-t_i}.\end{aligned}$$

Now we are investigating two different estimators, Bayes estimators and maximum likelihood estimators, in order to assess the impact due to prior information in Bayesian analysis by small simulation. We demonstrate the effect of prior strength using various setting of values of $(\alpha_1, \alpha_2, \alpha_3)$ and sample size.

For the noninformative prior distribution, Beta 3 and Beta 4, following $\alpha_i = 1$ and $\alpha_i = 1/2$, respectively. The other cases for informative prior, Beta 1 and Beta 2, were used to see the impact of prior distribution.

To compare the performance of an estimator, we do with the MSE to evaluate the performance of two estimators. The purpose of the numerical computation in this paper is to confirm that the Bayesian approach multinomial RR model is more efficient in terms of MSE with proper prior information.

Using the informative priors, Beta 1 and Beta 2, the Bayes estimators outperform the maximum likelihood estimators when π_3 is small for $n = 100$. With noninformative prior, Beta 3 (Uniform prior) and Beta 4 (Jeffrey's prior), when π_3 is small, MLE is better than Bayes estimator, however, when π_3 increases, Bayes estimator is slightly better than MLE in terms of MSE. When $n = 250$, the reduction in MSE realized by using a Bayes procedure diminishes noticeably; however, Bayes estimators based on informative priors still continue to

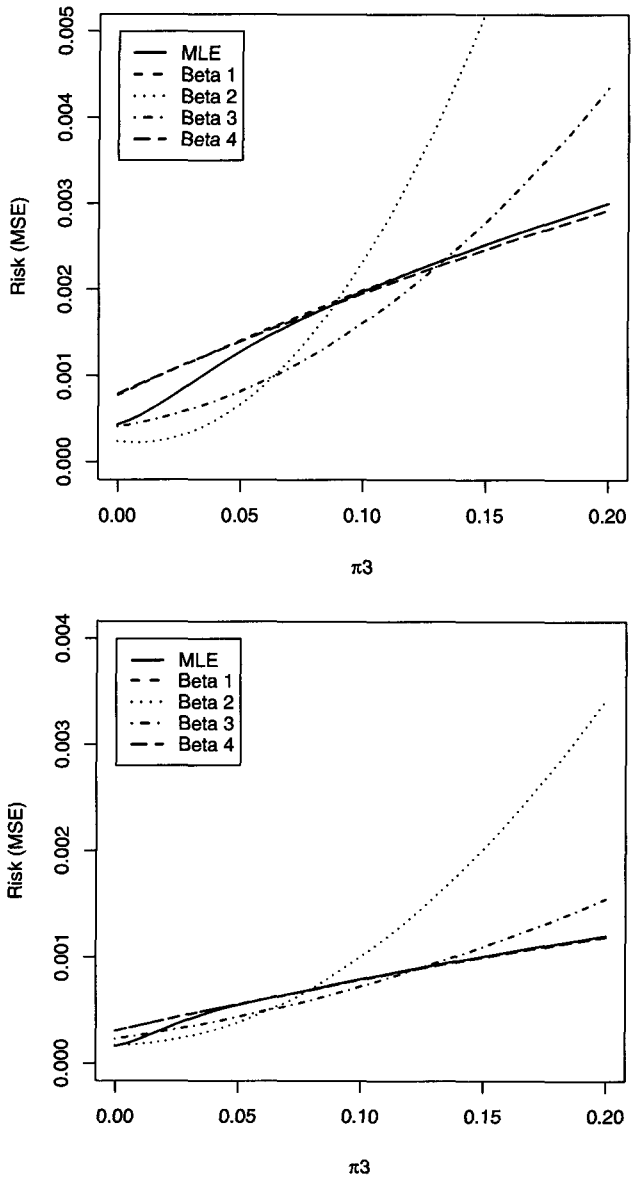

Figure 3.1: Top panel: MSE of $\widehat{\pi}_{M_3}$, $\widehat{\pi}_{B_3}$ when $n = 100, r = 70, g = 30, q_3 = 4$ and π_3 ranges from 0 to 0.2 Bottom panel: MSE of $\widehat{\pi}_{M_3}$, $\widehat{\pi}_{B_3}$ when $n = 250, r = 70, g = 30, q_3 = 4$ and π_3 ranges from 0 to 0.2; Beta 1 represents Beta(5, 95), Beta 2 represents Beta (2, 38), Beta 3 represents Beta (1, 2),Beta 4 represents Beta (0.5, 1)

have smaller MSE when π_3 is small. When examining Figure 3.1, one will note the overall large reduction in MSE for the Bayes estimator when compared to the MLE, especially when $\pi_3 \leq 0.13$. As we expected, when $\pi_3 = 0.05$, the reduction is greatest. In Figure 3.1, we also see that MLE and the Bayes estimator tend to be identical with increasing sample size. The Bayesian approach allows more flexibility in terms of how we may incorporate prior information into the parameter estimation process. The practical selection of a prior distribution may be delicate and subjective. However, for a large sample size, the relative weight of the prior information becomes negligible.

Acknowledgements

The authors are grateful to an anonymous referees for several valuable comments and suggestions.

References

Bar-Lev, S. K., Bobovich, E. and Boukai, B. (2003). A common conjugate prior structure for several randomized response models. *Test*, **12**, 101–113.

Kim, J.-M. and Warde, W. D. (2005). Some new results on the multinomial randomized response model. *Communications in Statistics: Theory and Methods*, **34**, 847–856.

Kim, J.-M., Tebbs J. M. and An, S.-W. (2006). Extensions of Mangat's randomized-response model. *Journal of Statistical Planning and Inference*, **136**, 1554–1567.

Liu, P. T. and Chow, L. P. (1976). New discrete quantitative randomized response model. *Journal of the American Statistical Association*, **71**, 72–73.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63–69.

[Received October 2006, Accepted December 2006]