The Korean Communications in Statistics **Vol. 14** No. 1, 2007, pp. 81–92

Bayesian Inference for Multinomial Group Testing

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

Abstract

This paper consider trinomial group testing concerned with classification of N given units into one of k disjoint categories. In this paper, we propose Bayesian inference for estimating individual category proportions using the trinomial group testing model proposed by Bar-Lev *et al.* (2005). We compared a relative efficience (RE) based on the mean squared error (MSE) of MLE and Bayes estimators with various prior information. The impact of different prior specifications on the estimates is also investigated using selected prior distribution. The impact of different priors on the Bayes estimates is modest when the sample size and group size are large.

Keywords: Group testing; trinomial distribution; interval estimator; Bayesian; Dirichlet distribution.

1. Introduction

Group testing problem was originally suggested by Dorfman (1943) and studied in screening large populations for diseases. Group testing is a cost-efficient way to obtain a lot of information in a short amount of time and efforts. If the number of individuals to be tested is quite large, we expect that the cost of testing will also be large. Group testing is the best way to reduce the number of test needed to screen everyone and thereby reduce the costs. Instead of testing each item, observations are made on groups of item polled together with group size $s > 1$.

In most applications, the group response is binary, classified as either non-infected or infected. Although we often encounters dichotomous items, item

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

responses may also be categorical. Kumar (1970a, 1970b, 1972) formulated the multinomial group testing problem, which is a generalization of the binomial. Under the assumption that a probability distribution on the number of defective items exists, multinomial probability group testing models have been considered by Hughes-Oliver and Rosenberger (2000), Xie *et al.* (2001), Zhu *et al.* (2001) and Pfeifier *et al.* (2002). However, Maximum likelihood often becomes prohibitive for small sample sizes due to unreasonable results.

To overcome MLE problem, we suggest a Bayesian framework for multinomial group testing model. In the Bayesian approach, the parameter to be estimated is considered a random variable that follows prior distribution. By taking into account the prior information for the unknown parameter in the estimation process, Bayesian estimation is expected to improve the accuracy of the estimated value, especially when the sample size is small, a case in which the maximum likelihood estimation procedure usually does not work well (Tebbs *et al.*, 2003). The prior distribution is typically specified by subjective way based on personal belief either from experience with the parameter or from the statistical properties of the parameter to be estimated. Although the Bayesian approach has often been criticized because of the subjective nature of its prior selection, Chaubey and Li (1995) used a two-parameter Beta prior distribution for p and derive the Bayes estimator using a squared error loss function. Chick (1996) also used a two-parameter Beta prior for p and considered the use of unequal group sizes. Kwon (2004) recently suggest Bayes estimators in binomial group testing with beta and uniform priors. However, neither of them addresses Bayesian estimation of the parameter in the context of multinomial group testing design.

Bar-Lev *et al.* (2005) recently proposed multinomial probability group testing which assumes that all pooled items has none or some of k attributes, one of them causing contamination. Estimation of proportions in the Bar-Lev *et al.* (2005) model is derived by MLE approach. In group testing problem, ML estimates may lie outside the boundary of the parameter space or are typically more extreme (or can be zero) than the Bayes estimates for rare traits (Tebbs *et al.*, 2003).

In this paper, we propose a Bayesian inference methods for the parameters of a multinomial group testing model based on Bar-Lev *et al.* (2005) model. The rest of the article is arranged as follows. Section 2 gives a short review of Bar-Lev *et al.* (2005) multinomial group testing. In Section 3, we derive a joint posterior distribution for Bayesian trinomial group testing model and its credible interval. A small simulation described to compare MLE and Bayes estimator in terms of relative efficiency (RE) in Section 4. Conclusions in Section 5.

2. Preliminaries

Recently Bar-Lev *et al.* (2005) proposed multinomial group testing model which deals with more than two category responses. Bar-Lev *et al.* (2005) focus on choosing an optimal group size for pooled screening so as to collect pre-specified numbers of items of the various types with minimum testing expenditures and derived exact results for the underlying distributions of the stopping times. They provided the probabilities of the group types for multinomial group testing. Let us review briefly Bar-Lev *et al.* (2005) approach.

In their work, they used the k attributes $A_1, \ldots, A_k$ and fix the group size s. The random vector $(Z_{i1}, \ldots, Z_{ik})$ consists of ones and zeroes as follows, for a given group,

$$Z_{ij} = \begin{cases} 1, & \text{if the } i^{th} \text{ item in the group possesses attribute } A_j, \quad j = 1, 2, \ldots, k, \\ 0, & \text{if it does not.} \end{cases}$$

Let B_0 be the event that none of the attributes is shown in the group and $B_{x_1,\ldots,x_h}$ be the event that the attributes $A_{x_1}, \ldots, A_{x_h}$ are shown in at least one item of the given group while the other attributes are not shown in all items for $1 \geq h \geq k$ and distinct indices $1 \leq s_1 < \cdots < s_h \leq k$. They denote the probabilities of B_0 and $B_{x_1,\ldots,x_h}$ as follows

$$\begin{aligned} P(B_0) &= P(Z_{ij}) = 0 \;\; \forall i, j, \\ P(B_{x_1,\ldots,x_h}) &= P\Bigg(\sum_{i=1}^{s} Z_{ix_j} \geq 1 \;\; \text{for } j = 1, \ldots, h, \\ &\qquad \sum_{i=1}^{s} Z_{ix} = 0 \;\; \text{for } x \notin \{x_1, \ldots, x_h\}\Bigg). \end{aligned}$$

Since the presence of attribute x_k contaminates a group, they can combine all types containing x_k into one, hence, they distinguish between the $l = 2^{k-1} + 1$ types as follows

$$\begin{cases} B_0, \\ B_{x_1,\ldots,x_h}, \quad 1 \leq h \leq k-1, \quad 1 \leq x_1 < \cdots < x_h \leq k-1, \\ B_l = \bigcup_{h=0}^{k-1} \;\; \bigcup_{1 \leq x_1 < \cdots < x_h \leq k-1} B_{x_1,\ldots,x_h,k}, \end{cases} \tag{2.1}$$

where B_0 in (2.1) is purly clean which means of the attributes is present in gorup. Second types in (2.1) are the clean ones containing the at least one of the attributes $x_1, \ldots, x_{k-1}$, and third type in (2.1) is the contaminated one.

Every item can have any combination of attributes independently of the other items. Usually, the population prevalence rate of contaminated items is much smaller than those of the clean types taken together. The occurrence of the k attributes $A_1, \ldots, A_k$ in an item can be assumed to be independent. In this paper, we denote by p_i the probability that an item possesses attribute i.

They provided the probabilities of the $B_0, B_{x_1,\ldots,x_h}, B_l$ as follows, by independent assumption,

$$\begin{cases} P(B_0) = P(Z_{ij}=0)^s = \prod_{j=1}^{k} p_j^s, \\ P(B_{x_1,\ldots,x_h}) = \prod_{j\notin\{x_1,\ldots,x_h\}} p_j^s \prod_{m=1}^{k}(1-p_{x_m})^s, \\ P(B_l) = 1 - p_k^s. \end{cases}$$

Under the multinomial group testing of Bar-Lev *et al.* (2005), π_1 is the case when there is no element for attribute present, π_2 is the case when there is an element for attribute 1 but not for attribute 2 present, π_3 is the case when there is an element for attribute 2 present for trinomial model. In this paper, we will focus on a trinomial group testing model which is concerned with classification of N given units into one of three disjoint categories. Assume that the attributes are independently distributed in the population. For a group test of size s, we obtain the probabilities of the three different categories for trinomial distribution as follows:

$$\begin{aligned} \pi_1 &= (1-p_1)^s(1-p_2)^s, \\ \pi_2 &= (1-p_2)^s[1-(1-p_1)^s], \\ \pi_3 &= 1-(1-p_2)^s, \end{aligned}$$

where $\pi_1+\pi_2+\pi_3=1$. By using invariant property of MLE, the ML estimates of p_0, p_1, p_2 are given by

$$\begin{aligned} \widehat{p}_0 &= 1-\widehat{p}_1-\widehat{p}_2 = \Big(\frac{n_1+n_2}{n_1+n_2+n_3}\Big)^{\frac{1}{s}} + \Big(\frac{n_1}{n_1+n_2}\Big)^{\frac{1}{s}} - 1, \\ \widehat{p}_1 &= 1-\Big(1-\frac{\widehat{\pi}_2}{1-\widehat{\pi}_3}\Big)^{\frac{1}{s}} = 1-\Big(\frac{n_1}{n_1+n_2}\Big)^{\frac{1}{s}}, \\ \widehat{p}_2 &= 1-(1-\widehat{\pi}_3)^{\frac{1}{s}} = 1-\Big(\frac{n_1+n_2}{n_1+n_2+n_3}\Big)^{\frac{1}{s}}, \end{aligned}$$

where $\widehat{\pi}_1 = n_1/n$, $\widehat{\pi}_2 = n_2/n$, and $\widehat{\pi}_3 = n_3/n$. For example, assuming no testing errors, N_3, has a binomial distribution with parameter n and $1-(1-p_2)^s$. For interval estimation, the confidence intervals to the group testing model is given by $\widehat{p}_2 \pm z_{\alpha/2}\sqrt{\widehat{Var}(\widehat{p}_2)/n}$, where $\widehat{\mathrm{Var}}(\widehat{p}_2) = \{1-(1-\widehat{p}_2)^s\}(1-\widehat{p}_2)^{2-s}/s^2$ and $z_{\alpha/2}$ denotes the upper $\alpha/2$ percentile of the standard normal distribution.

3. Bayesain Inference for Trinomial Group Testing

Bayesian methods have not been extensively used in group testing problem. The ultimate goal of this study is to improve procedure for estimating the parameters in group testing model using Bar-Lev *et al.* (2005) model. We now introduce the Bayesian trinomial group testing estimation problem which is more applicable in practice.

In this study, we will formally introduce the Bayesian method in a general framework but we will focus on a trinomial group testing model which is concerned with classification each of N given units into one of three disjoint categories. Note that in many cases, it is not easy to calculate the desired joint posterior distribution using analytical method. In our work, we analytically derived the joint and marginal posterior distribution for multinomial group testing model. Bayesian estimators typically provide more accurate results with less bias and smaller mean squared errors (MSE) between the true and the estimated values than the maximum likelihood estimators with proper prior information.

We denote by p_j the probability that an item possesses attribute i and $\sum_{j=1}^{3} p_j = 1$. We assume that the attributes are fixed and *iid* Bernoulli(p_j) random variables, $0 < p_j < 1$ and a common group size s. We assume that the response vector $(Y_{i1}, Y_{i2}, \ldots, Y_{i3})$ for item i with $i = 1, 2, \ldots, n$ and for attribute j, $j = 1, 2, 3$ consisting of zero and one, in addition to that we consider vectors having at least one non zero entry. To formally introduce group testing problem, we need to fix the group size s. Therefore $\tilde{N} = (N_1, N_2, N_3)$ has a trinomial distribution with parameters n and $\Pi = (\pi_1, \pi_2, \pi_3)$ in section 2. Therefore, the likelihood function of $\tilde{N}$ given that $\Pi = (\pi_1, \pi_2, \pi_3)$ is

$$f_{\tilde{N}|\tilde{\Pi}}(\tilde{n}|\tilde{\pi}) = \frac{n!}{\prod_{i=1}^{3} n_i} \prod_{i=1}^{3} \pi_i^{n_i}.$$

The number of defective groups, $\tilde{N} = (N_1, N_2, N_3)$, has a multinomial distribution with parameters $\tilde{n} = (n_1, n_2, n_3)$ and $\tilde{\pi} = (\pi_1, \pi_2, \pi_3)$. Thus, the likelihood for trinomial distribution of group testing problem can be expressed as

$$f_{\tilde{N}|\tilde{\Pi}}(\tilde{n}|\tilde{\pi}) = \frac{n!}{n_1! n_2! (n - n_1 - n_2)!} \pi_1^{n_1} \pi_2^{n_2} (1 - \pi_1 - \pi_2)^{n - n_1 - n_2},$$

where $\pi_1 + \pi_2 + \pi_3 = 1$ and $n = n_1 + n_2 + n_3$.

In order to complete the model specification from a Bayesian framework, we specify a joint prior distribution for all parameter of the model. In this work, we adopted Dirichlet distribution as a prior information. In order to fully specify the Bayesian model, Dirichlet priors are assigned to the $\pi_j, j = 1,2,3$, *i.e.*, $\Pi \sim$ Dirichlet$(\alpha_1, \alpha_2, \alpha_3)$, where α_i, $i = 1,2,3$ are the hyper-parameter of the prior distribution. The prior distribution should express our knowledge of the model parameters, before the data is taken into account. In many situations we do not have any prior information, or we do not want to use it. Many approaches have been suggested to construct these noninformative prior distributions. The first and most obvious way to define a noninformative prior distribution of parameters is by the Dirichlet distribution on the parameter space. The Dirichlet prior corresponds well with our intuition about a noninformative prior. There are several advantages of incorporating the Dirichlet distribution in trinomial set up. First, Dirichlet distribution is a conjugate family of trinomial distribution. Second, Dirichlet prior $\tilde{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ is appropriate for small p, since for large value of α, the majority of the probability distribution of the random variable is closed to zero. Third, estimates derived using Dirichlet prior are consistent, and can be computed efficiently by conjugate property.

The Dirichelt prior is given by

$$f_{\tilde{\Pi}}(\tilde{\pi}|\tilde{\alpha}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha - \alpha_1 - \alpha_2)} \pi_1^{\alpha_1 - 1} \pi_2^{\alpha_2 - 1} (\pi - \pi_1 - \pi_2)^{\alpha - \alpha_1 - \alpha_2 - 1},$$

for values of $\alpha = \sum_{i=1}^{3} \alpha_i$. The joint distribution of $\tilde{N}$ and $\tilde{\pi}$, conditioned on $\tilde{\alpha}$, is given by

$$f_{\tilde{N},\tilde{P}}(\tilde{n}, \tilde{\pi}|\tilde{\alpha}) = f_{\tilde{N}|\tilde{\pi}}(\tilde{n}|\tilde{\pi}) \cdot f_{\tilde{\pi}}(\tilde{\pi}|\tilde{\alpha}).$$

Using transformation with $\pi_1 = (1-p_1)^s(1-p_2)^s$, $\pi_2 = (1-p_2)^s[1-(1-p_1)^s]$, and $\pi_3 = [1-(1-p_2)^s]$ so that $\Pi = (\pi_1, \pi_2, \pi_3)$, the joint distribution of $\tilde{N}$ and $\tilde{P}$, conditioned on $\tilde{\alpha}$, is given by

$$f_{\tilde{N},\tilde{P}}(\tilde{n}, \tilde{p}|\tilde{\alpha}) = f_{\tilde{N}|\tilde{p}}(\tilde{n}|\tilde{p}) \cdot f_{\tilde{p}}(\tilde{p}|\tilde{\alpha}) \cdot |J|,$$

for $0 < p_1$, $p_2 < 1$ and the Jacobian is $|J| = \left|s^2(1-p_1)^{s-1}(1-p_2)^{2s-1}\right|$. The marginal distribution of $\tilde{N}$ can be represented by the product of gamma function (not shown in paper).

The joint posterior distribution is given by

$$f_{\tilde{P}|\tilde{N}}(\tilde{p}|\tilde{n}, \tilde{\alpha}) = \frac{f_{\tilde{N},\tilde{P}}(\tilde{n}, \tilde{p}|\tilde{\alpha})}{f_N(\tilde{n}|\tilde{\alpha})}.$$

The joint posterior distribution is a product of the joint prior distribution and the likelihood function. We obtain analytically the marginal posterior distribution by integration over some parameters. This joint posterior distribution is all that is needed to make inference about the unknown parameters. The full conditional distributions are derived from the joint posterior distribution. The final full conditional distribution for proportion p_1 is given by

$$f_{P_1|\tilde{N}}(p_1|\tilde{n},\tilde{\alpha}) = \frac{s\Gamma(n_1+n_2+\alpha_1+\alpha_2)}{\Gamma(n_1+\alpha_1)\Gamma(n_2+\alpha_2)}(1-p_1)^{s(n_1+\alpha_1-\frac{1}{s})}[1-(1-p_1)^s]^{(n_2+\alpha_2-1)}$$

and, in the same manner, p_2 is given by

$$f_{P_2|\tilde{N}}(p_2|\tilde{n},\tilde{\alpha}) = \frac{s\Gamma(n+\alpha)}{\Gamma(n_1+n_2+\alpha_1+\alpha_2)\Gamma(n-n_1-n_2+\alpha-\alpha_1-\alpha_2)} \times(1-p_2)^{s(n_1+n_2+\alpha_1+\alpha_2-\frac{1}{s})}[1-(1-p_2)^s]^{(n-n_1-n_2+\alpha-\alpha_1-\alpha_2-1)}.$$

With $f_{P_i|\tilde{N}}(p_i|\tilde{n},\tilde{\alpha})$ and a given loss function, say, $L(p_i,a)$, (where a denotes the action taken), the Bayes estimate of p_i with respect to $L(p_i,a)$ is the value of a that minimizes

$$E[L(P_i,a)|\tilde{n},\tilde{\alpha}] = \int_0^1 L(p_i,a)f_{P_i|\tilde{N}}(p_i|\tilde{n},\tilde{\alpha})dp_i,$$

for $i=1,2$. For the remainder of this section, and for all comparisons in Section 4, only squared-error loss is considered; *i.e.*, $L(p_i,a)=(p_i-a)^2$, so that the Bayes estimate of p_i is the mean of posterior $f_{P_i|N}(p_i|\tilde{n},\tilde{\alpha})$. A closed-form expression for $\widehat{p}_{B_1}$, the mean of the posterior, is given by

$$\widehat{p}_{B_1} = 1 - \frac{\text{Beta}(n_1+\alpha_1+\frac{1}{s}, n_2+\alpha_2)}{\text{Beta}(n_1+\alpha_1, n_2+\alpha_2)},$$

where $\text{Beta}(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ with Γ denote gamma function. Similarly a closed-form expression for $\widehat{p}_{B_2}$, the mean of the posterior, is given by

$$\widehat{p}_{B_2} = 1 - \frac{\text{Beta}(n_1+n_2+\alpha_1+\alpha_2+\frac{1}{s}, n-n_1-n_2+\alpha-\alpha_1-\alpha_2)}{\text{Beta}(n_1+n_2+\alpha_1+\alpha_2, n-n_1-n_2+\alpha-\alpha_1-\alpha_2)}.$$

3.1. Credible Intervals

In the group testing literature, methods for confidence intervals construction have not been studied extensively. Thompson (1962) provide an approximate confidence intervals for the population proportion of viruliferous insects, based on the exact variance and Student-t approach. The predominant strategy is to use approximate Wald-type confidence intervals, based on the normal distribution, using the asymptotic variance of $\widehat{p}_{\mathrm{M}_i}$, for $i = 1, 2$. Straightforward calculations show this interval is given by

$$\widehat{p}_{\mathrm{M}_i} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\{1-(1-\widehat{p}_{\mathrm{M}_i})^s\}(1-\widehat{p}_{\mathrm{M}_i})^{2-s}}{ns^2}},$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$ percentile of the standard normal distribution.

Alternatively, by constructing the exact posterior distribution, we may obtain the interval estimates of parameter by locating the $\alpha/2$ and $1-\alpha/2$ quantiles of the relevant Beta distribution, or by the method of highest prior density (HPD). For parameter, we calculate a $100(1-\alpha)\%$ credible intervals for p_i as follows

$$\int_{L_{p_i}}^{U_{p_i}} f_{P_i|\tilde{N}}(p_i|\tilde{n}, \tilde{\alpha})dp_i = 1-\alpha,$$

where $0 < L_{p_i} < U_{p_i} < 1$ for $i = 1, 2$. We denote credible interval by (L_{p_i}, U_{p_i}), in practice, L_{p_i} and U_{p_i} may be determined using an equal-tail credible interval (95% equal-tail credible interval). In this study, we can find a nice closed-form expression for the equal-tail credible interval. The lower bound (LB) of p_1 may find by solving following equation

$$\frac{\alpha}{2} = \int_0^{L_{p_1}} f_{P_1|\tilde{N}}(p_1|\tilde{n}, \tilde{\alpha})dp_1$$

and obtain $L_{p_1} = 1 - \{1 - [\mathrm{Beta}(\alpha/2; n_2 + \alpha_2, n_1 + \alpha_1)]^{1/s}\}$. Similar to L_{p_1}, we obtain U_{p_1} as $U_{p_1} = 1 - \{1 - [\mathrm{Beta}\{1-(\alpha/2); n_2 + \alpha_2, n_1 + \alpha_1\}]^{1/s}\}$.

In the same manner, the lower bound for $\widehat{p}_{\mathrm{B}_2}$, $L_{p_2} = 1 - [\mathrm{Beta}(\alpha/2; n - n_1 - n_2 + \alpha - \alpha_1 - \alpha_2, n_1 + n_2 + \alpha_1 + \alpha_2)]^{1/s}$. Similar to L_{p_2}, we obtain U_{p_2} as $U_{p_2} = 1 - [\mathrm{Beta}\{1-(\alpha/2); n - n_1 - n_2 + \alpha - \alpha_1 - \alpha_2, n_1 + n_2 + \alpha_1 + \alpha_2\}]^{1/s}$.

4. Comparison of Two Estimators

In this Section, we are investigating several different Bayes estimators $\widehat{p}_{\mathrm{B}}$ based on different prior information and maximum likelihood estimators, $\widehat{p}_{\mathrm{M}}$, in order to assess the impact due to prior information by small simulation. The performance of an estimator may be evaluated by relative efficiency

$$\mathrm{RE}(\widehat{p}_{\mathrm{M}_2} \text{ to } \widehat{p}_{\mathrm{B}_2}) = \frac{\mathrm{MSE}(\widehat{p}_{\mathrm{M}_2})}{\mathrm{MSE}(\widehat{p}_{\mathrm{B}_2})}.$$

For p fixed, the exact mean squared error(MSE) of $\widehat{p}_{B_2}$, is given by

$$\begin{aligned}\mathrm{MSE}(\widehat{p}_{B_2}) &= E_{\tilde{N}|\tilde{P}}\left[(\widehat{p}_{B_2} - p_2)^2\right] \\ &= \sum_{t=0}^{n}(\widehat{p}_{B_2} - p_2)^2 \times \binom{n}{t}[1-(1-p)^s]^t(1-p)^{s(n-t)}.\end{aligned}$$

For various setting of values of $(\alpha_1, \alpha_2, \alpha_3)$ and sample size $n = 50, 200, 375$ considered in small simulation. If we do not know the values of α_i for $i = 1, 2, 3$ or some equivalent information, we may either put priors on the hyper-parameters, or one may adopt an empirical Bayes approach. In many cases it is desirable to select a so called noninformative prior in order to reduce the impact of the prior on the results of the analysis.

Kwon (2004) points that group testing is most useful when p is small and should be considered only in such cases, the distribution of p seen in group testing applications strongly skewed to the right, with p almost certainly < 0.3 and usually < 0.05 in binary group testing model. In trinomial case also similar facts that p_2 should be small to ensure the usefulness. To set the hyper parameters in Dirichlet prior to see the impact of prior information, we consider several possible sets of hyper-parameters, two cases all marginal priors are derived from a symmetric Dirichlet distribution (all parameters taking common value) for the model. For the first prior, $\alpha_i = 1$, and for the second prior $\alpha_i = 1/2$ (called Jeffreys' prior). The other cases for informative prior were used to see how change the relative efficiency (RE) as hyper-parameter changes in prior distribution.

The effect of the prior is examined in terms of RE. Table 4.1 shows the values of parameter estimates and RE($\widehat{p}_{M_2}$ to $\widehat{p}_{B_2}$) and illustrates that the posterior distribution is relatively sensitive to the choice of prior distribution. The magnitude of the different between the MSE of $\widehat{p}_{B_2}$ and $\widehat{p}_{M_2}$ depends on the impact on the prior distribution.

The more informative the prior distribution, the more the parameter estimate tends to be pulled toward the center of its prior distribution, and the greater the impact of the prior on the $\widehat{p}_{B_2}$. The impact of prior distribution tended to become less intense as the sample size and group size became larger. These results demonstrate that a poorly specified prior can severely bias the $\widehat{p}_{B}$ by pulling them to the wrong center, particularly when the sample size is not large. Interestingly, when p_2 is small, RE is not good for noninformative and Jeffrey's prior, but good for Dirichlet(5, 1, 1). When p_2 is small, lower bounds for the MLE can be negative where is out of parameter space that makes the MLE better than Bayes estimator in some small p_2. Dirichlet(5, 1, 1) and Dirichlet(1, 5, 1) provide the same RE

Table 4.1: Comparison of $\widehat{p}_{B_2}$ and $\widehat{p}_{M_2}$ in terms of the relative efficiency

RE($\widehat{p}_{M_2}$ to $\widehat{p}_{B_2}$)	$n = 100$		$n = 375$		$n = 500$	
Prevalence	$s = 5$	$s = 10$	$s = 5$	$s = 10$	$s = 5$	$s = 10$
Dirichlet(1,1,1)						
$p_2 = 0.01$	0.888	0.967	0.966	0.990	0.974	0.992
$p_2 = 0.02$	0.971	1.015	0.991	1.004	0.994	1.003
$p_2 = 0.05$	1.030	1.054	1.008	1.016	1.007	1.012
$p_2 = 0.10$	1.056	1.088	1.015	1.024	1.011	1.018
$p_2 = 0.20$	1.086	1.267	1.022	1.057	1.016	1.042
Dirichlet(0.5,0.5,0.5)						
$p_2 = 0.01$	0.970	0.991	0.991	0.997	0.974	0.998
$p_2 = 0.02$	0.994	1.005	0.998	1.001	0.994	1.001
$p_2 = 0.05$	1.013	1.021	1.003	1.005	1.007	1.004
$p_2 = 0.10$	1.022	1.040	1.006	1.010	1.011	1.008
$p_2 = 0.20$	1.039	1.153	1.010	1.031	1.016	1.023
Dirichlet(5,1,1)						
$p_2 = 0.01$	1.021	1.107	1.005	1.028	1.004	1.021
$p_2 = 0.02$	1.110	1.141	1.029	1.036	1.021	1.027
$p_2 = 0.05$	1.131	1.068	1.034	1.018	1.026	1.013
$p_2 = 0.10$	1.065	0.907	1.017	0.967	1.013	0.974
$p_2 = 0.20$	0.894	0.620	0.963	0.797	0.972	0.834
Dirichlet(1,5,1)						
$p_2 = 0.01$	1.021	1.107	1.005	1.028	1.004	1.021
$p_2 = 0.02$	1.110	1.141	1.029	1.036	1.021	1.027
$p_2 = 0.05$	1.131	1.068	1.034	1.018	1.026	1.013
$p_2 = 0.10$	1.065	0.907	1.017	0.967	1.013	0.974
$p - 2 = 0.20$	0.894	0.620	0.963	0.797	0.972	0.834
Dirichlet(1,1,5)						
$p_2 = 0.01$	0.191	0.333	0.452	0.635	0.522	0.697
$p_2 = 0.02$	0.337	0.535	0.639	0.800	0.701	0.841
$p_2 = 0.05$	0.686	0.904	0.883	0.969	0.909	0.976
$p_2 = 0.10$	0.898	1.085	0.967	1.021	0.975	1.016
$p_2 = 0.20$	1.082	1.360	1.021	1.076	1.016	1.056
Dirichlet(5,5,5)						
$p_2 = 0.01$	0.256	0.498	0.517	0.753	0.584	0.800
$p_2 = 0.02$	0.504	0.912	0.758	0.965	0.804	0.973
$p_2 = 0.05$	1.201	1.267	1.052	1.070	1.039	1.052
$p_2 = 0.10$	1.281	0.781	1.073	0.905	1.054	0.925
$p_2 = 0.20$	0.791	0.350	0.912	0.522	0.931	0.578

results because of the feature of $\widehat{p}_{B_2}$. As the prior became further away from the distribution of true parameter in terms of the shape of the distribution density function, RE tends to become smaller (see, Dirichlet(1, 1, 5) and Dirichlet(5, 5, 5). From our small simulations, we obtain the similar results from Kwon (2004)'s Bayesian approach binomial group testing model.

5. Discussion

The aim of this paper is essentially on deriving of closed-form Bayes estimates of parameters based on Bar-Lev *et al.* (2005) multinomial group testing approach. Bayesian approaches we present give a more flexible and reliable estimates taking into prior information than MLE, but the impact of different prior specification on the Bayesian estimation largely depends on the sample size and group size. However, Bayesian approach is still recommendable to use for group testing model since, in doing so, we obtain a reliable estimates for rate trait proportion with appropriate prior information. We derive the credible intervals and compare the results to perform Bayesian inference with different prior distribution for parameter of interest for rate traits using MLE and Bayes estimator.

Acknowledgements

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

References

Bar-Lev, S. K., Stadje, W. and van der Duyn Schouten, F. A. (2005). Multinomial group testing models with incomplete identification. *Journal of Statistical Planning and Inference*, **135**, 384–401.

Chaubey, Y. P. and Li, W. (1995). Comparison between maximum likelihood and Bayes methods for estimation of binomial probability with sample compositing. *Journal of Official Statistics*, **11**, 379–390.

Chick, S. E. (1996). Bayesian models for limiting dilution assay and group test data. *Biometrics*, **52**, 1055–1062.

Dorfman, R. (1943). The detection of defective members of large populations. *The Annals of Mathematical Statistics*, **14**, 436–440.

Hughes-Oliver, J. M. and Rosenberger, W. F. (2000). Efficient estimation of the prevalence of multiple rare traits. *Biometrika*, **87**, 315–327.

Kumar, S. (1970a). Multinomial group-testing. *SIAM Journal of on Applied Mathematics*, **19**, 340–350.

Kumar, S. (1970b). Group-testing to classify all units in a trinomial sample. *Studia Scientiarum Mathematicarum Hungarica*, **5**, 229–247.

Kumar, S. (1972). Trinomial group–testing with an unknown proportion of units in the three categories. *Annals of the Institute of Statistical Mathematics*, **24**, 171–181.

Kwon, S. (2004). Bayes estimators in group testing. *The Korean Communication in Statistics*, **11**, 619–630.

Pfeiffer, R. M., Rutter, J. L., Gail, M., Struewing, J. and Gastwirth, J. L. (2002). Efficiency of DNA pooling to estimate joint allele frequencies and measure linkage disequilibrium. *Genetic Epidemiology*, **22**, 94–102.

Tebbs, J. M., Bilder, C. R. and Moser, B. K. (2003). An empirical Bayes group-testing approach to estimating small proportions. *Communications in Statistics: Theory and Methods*, **32**, 983–995.

Xie, M. (2001). Regression analysis of group testing samples. *Statistics in Medicine*, **20**, 1957–1969.

Zhu, L., Hughes-Oliver, J. M. and Young, S. S. (2001). Statistical decoding of potent pools based on chemical structure. *Biometrics*, **57**, 922–930.

[Received October 2006, Accepted December 2006]