Journal of Modern Applied Statistical Methods

Volume 6 | Issue 2 Article 20

11-1-2007

Bayesian Subset Selection of Binomial Parameters Using Possibly Misclassified Data

James D. Stamey
Baylor University

Thomas L. Bratcher
Baylor University

Dean M. Young
Baylor University

Journal of Modern Applied Statistical Methods
November, 2007, Vol. 6, No. 2, 551-560

1538 – 9472/07/$95.00

Bayesian Subset Selection of Binomial Parameters Using Possibly Misclassified Data

James D. Stamey Thomas L. Bratcher Dean M. Young
Baylor University

Three Bayesian approaches are considered for the selection of binomial proportion parameters when data is subject to misclassification. The cases where the misclassification is non-differential and differential were considered, thus extending previous work which considered only non-differential misclassification. In this article, various selection criteria are applied to a simulated data set and a real data set.

Key words: Bayes, posterior approximation, Gibbs Sampler, binomial parameter subset selection

Introduction

A decision maker is often interested in selecting the population from among several populations that will produce the largest or smallest parameter value. For example, an experimenter might be interested in determining which production technique gives the lowest percentage of defects; a crime analyst might consider which reporting district has the highest rate of violent crimes; a baseball fan might inquire about the best home run hitter of the twentieth century. In each case a selection of a population parameter must be made from a set of parameters using data from the populations of interest. This process is known as the subset-selection problem. Of course various procedures exist for selecting a subset that contains the best (largest or smallest) parameter. Here, the concern is with the Bayesian subset selection paradigm.

James Stamey is an Assistant Professor in the Department of Statistical Science at Baylor University. His research interests are in discrete data with misclassification errors and Bayesian analysis. Email: James_Stamey@baylor.edu. Tom Bratcher is a Professor of Statistics. His research interests are selection and ranking. Dean Young is Professor of Statistics. His research interests are multivariate analysis and linear models.

The concept of subset selection essentially began with an article by Gupta and Sobel (1957), who described a statistic that can be used in parameter ranking and selection for multiple populations. Early work on Bayesian subset selection was initiated by Bratcher & Bhalla (1974), who have used a constant loss function to derive a Bayesian subset selection procedure, and Govindarajulu & Harvey (1974). For other Bayesian subset-selection approaches and related topics, see Goel & Rubin (1977), Gupta & Hsu (1977), Berger (1979, 1980), Gupta & Yang (1985), Gupta & Liang (1987), Berger & Deely (1988), Dixon & Simon (1991, 1994), Schulter, Deely, & Nicholson (1997) and Deely & Smith (1998).

Examples abound where interest might be in selecting a subset of binomial proportion parameters using correctly classified and misclassified data. For example, Hanson, Johnson, & Gardner (2003) have considered the prevalence of the disease bovine brucellosis in cattle herds in twenty regions of Mexico. This application can be thought of as a type of quality control in which one wishes to determine a set of herds deemed most likely to develop bovine brucellosis or, conversely, perhaps a set of herds that could be considered least likely to have the disease. A second application of a subset-selection method for binomial proportion parameters using possibly misclassified data is auditing. For instance, Raats & Moors (2004) have estimated the proportion of errors in social security payments in the Netherlands combining

fallible and validation data. One could also compare or select a subset of the proportion parameters of errors in auditing across geographical regions, industries, or some other variable of interest.

In both of the above examples, one cannot reasonably assume the observed counts are infallible. Most diagnostic tests are well known to be fallible. That is, most diagnostic tests can indicate that subjects have a disease when they do not or that they are disease free when they are actually infected. An appropriate statistical model will adjust for the error rates of the fallible test. Joseph, Gyorkos, & Coupal (1995) and Dendukuri & Joseph (2001) considered the case of estimating the prevalence of one population with fallible data. Hanson et al. (2003) have extended this work to multiple populations. Hanson et al. (2003) assumed that the properties of the diagnostic tests remain constant across populations. This assumption is referred to as non-differential misclassification.

Two subset-selection criteria of Schluter et al. (1997) and a subset- selection criterion proposed by Stamey, Bratcher, & Young (2004) are applied here to the bovine brucellosis data found in Hanson et al. (2003). Also proposed is a method of extending the hierarchical model to allow for differential misclassification. Differential misclassification occurs when the false positive and false negative rates are different in each population. For this scenario it is assumed that an expensive error-free classifier is available for a small sample of units. A sample where both fallible and infallible observations are made is often called a validation sample. A simulated binomial parameter subset-selection problem with differential misclassification motivated by an auditing application in Raats & Moors (2004) is considered.

Methodology

A parametric hierarchical model for binomial data with misclassification analogous to Hanson et al. (2003) is provided and a Bayesian extension is proposed for the case of differential misclassification. For the non-differential misclassification model, consider the case where only a single classifier is utilized; however, the method is easily extended to allow for two or more classifiers. The hierarchical model is

$$Z_i \mid \pi_i, \eta, \theta \sim \text{binomial}(n_i, p_i)$$

with

$$p_i = \pi_i\eta + (1 - \pi_i)(1 - \theta),$$

where p_i is the population proportion of observable occurrences in population $i = 1, \ldots, m$. The parameter π_i is the true probability of a positive response for population i and is assumed to vary across populations. The parameter η = 1 – P(false negative) is the sensitivity, or probability that a true positive is observed and is assumed to be the same for all populations. The parameter θ = 1 – P(false positive) is the specificity, or probability that a true negative is labeled as a negative and is also assumed to be the same for all populations. The first-stage priors of the Bayesian hierarchical model are

$$\pi_i \sim \text{beta}(\alpha, \beta),$$

$$\eta \sim \text{beta}(\alpha_\eta, \beta_\eta),$$

and

$$\theta \sim \text{beta}(\alpha_\theta, \beta_\theta).$$

The beta prior is the usual first-stage prior for hierarchical binomial models and is consistent with the models of Hanson et al. (2003). One can elicit priors for the sensitivity and specificity by using the approaches of Chaloner (1996) and Kadane & Wolfson (1996).

To model the heterogeneity of the prevalences, the parametric prior of Hanson et al. (2003) is used for both its convenience and ease of interpretation. Here, $\alpha = \mu\gamma$ and $\beta = \gamma$, where the parameter μ is the grand mean of the population prevalences and γ controls the heterogeneity of the prevalences since the variance is $\frac{\mu(1-\mu)}{1+\gamma}$. Specifically, the larger the value of γ, the tighter the distribution of the prevalences. To finish the hierarchy, assume $\mu \sim \text{beta}(\alpha_\mu, \beta_\mu)$ and $\gamma \sim \text{gamma}(\alpha_\gamma, \beta_\gamma)$, where α_μ, β_μ, α_γ, and β_γ are hyperpriors specified by the

experimenter. The joint posterior of all parameters is proportional to

$$p(\pi_i,\theta,\eta,\mu,\gamma\,|\,\mathbf{d}) \propto \prod_{i=1}^{r}\pi_i^{\mu\gamma-1}(1-\pi_i)^{\gamma-1}\theta^{\alpha_\theta-1}(1-\theta)^{\beta_\theta-1}\eta^{\alpha_\eta-1}(1-\eta)^{\beta_\eta-1}p_i^{z_i}(1-p_i)^{n_i-z_i}.$$

Hanson et al. (2003) provided a method for eliciting values for the parameters of the priors. However, in the analyses diffuse non-informative priors are used. No apparent closed-form posterior distributions exist, but the parameters can be estimated using either Monte Carlo integration or Markov Chain Monte Carlo methods. The free software WinBugs is used to approximate the posterior densities that is used. These WinBugs software programs are available from the first author.

The assumption that the sensitivity and specificity do not vary across populations is quite strong and often fails in practice. Here the model of Hanson et al. (2003) is extended to the case where the sensitivity and specificity are not the same across populations. If it is believed that the misclassification parameters vary across populations, it is recommended to use one of the following approaches. If the number of populations is not large, an expert to elicit prior parameters for each specificity and sensitivity can be used, using methods detailed in Chaloner (1996) and Kadane & Wolfson (1996). This approach results in the following change in the hierarchical model: $\eta_i \sim \text{beta}(\alpha_{\eta i}, \beta_{\eta i})$ and $\theta_i \sim \text{beta}(\alpha_{\theta i}, \beta_{\theta i})$.

However, if expert opinion is not available for each of the sensitivities and specificities, another method is needed. One possibility is to use validation data for each population. For instance, Raats & Moors (2004) have assumed that a large sample of accounts is audited by a fallible auditor, and then a small random sample of these accounts is double checked by an infallible expert. Suppose in each population r_i units are classified by both the fallible and infallible procedure. The validation data adds the following binomial likelihoods to the experiment likelihood:

$$T_i\,|\,\pi_i \sim \text{binomial}(r_i, \pi_i),$$

$$X_i\,|\,t_i, \eta_i \sim \text{binomial}(t_i, \eta_i),$$

and

$$Y_i\,|\,t_i, \theta_i \sim \text{binomial}(r_i - t_i, \theta_i).$$

Here, T_i is the number of positive responses determined by the infallible classifier, X_i is the number of true positive responses correctly labeled as positive by the fallible classifier, and Y_i is the number of true negative responses labeled as negative by the fallible classifier. Then, a hierarchical structure for the sensitivity and specificity parameters similar to that used on the prevalences is used. That is, $\eta_i \sim \text{beta}(\alpha_\eta, \beta_\eta)$ and $\theta_i \sim \text{beta}(\alpha_\theta, \beta_\theta)$ and define $\alpha_\eta = \mu_\eta\gamma_\eta$, $\beta_\eta = \gamma_\eta$, $\alpha_\theta = \mu_\theta\gamma_\theta$, and $\beta_\theta = \gamma_\theta$. The hierarchy is completed with the priors

$$\mu_\eta \sim \text{beta}(\alpha_{\mu\eta}, \beta_{\mu\eta}),$$
$$\gamma_\eta \sim \text{gamma}(\alpha_{\gamma\eta}, \beta_{\gamma\eta}),$$
$$\mu_\theta \sim \text{beta}(\alpha_{\mu\theta}, \beta_{\mu\theta}),$$

and

$$\gamma_\theta \sim \text{gamma}(\alpha_{\gamma\theta}, \beta_{\gamma\theta}).$$

The WinBugs computer programs used to approximate the posterior distributions are available from the first author.

Three Subset Selection Procedures

Reviewed next are two subset selection criteria from Schluter et al. (1997) and a decision theoretic subset selection criterion from Stamey, Bratcher, & Young (2004) and extend them to apply to the binomial parameter case using possibly misclassified data.

A Posterior Probabilities Approach (Schluter et al. (1997))

The first subset-selection procedure that is considered uses the posterior probability that a site has the largest prevalence or is largest by a multiple of, say, ν. That is,

$$p_i(\nu) = P(\pi_i > \nu\pi_j, \forall\, j \neq i\,|\,\mathbf{z}) \qquad (1)$$

where $\mathbf{z}$ represents the vector of observed data. The probability (1) does not have a closed form; however, MCMC methods make (1) trivial to calculate. Suppose that after an initial burn-in, the Gibbs sampler is run B iterations. One can

approximate the posterior probability (1) by counting the number of times

$$\pi_{ik} = \max(\nu\pi_{1k}, ..., \nu\pi_{i-1,k}, \pi_{ik}, \nu\pi_{i+1,k}, ..., \nu\pi_{mk}),$$

where $k = 1, \ldots, B$. Specifically, probability (1) is approximated as

$$p_i(\nu) \approx \frac{\#(\pi_{ik} = \max(\nu\pi_{1k}, ..., \nu\pi_{i-1,k}, \pi_{ik}, \nu\pi_{i+1,k}, ..., \nu\pi_{mk}))}{B}$$

where $\#(\cdot)$ denotes the number of elements in a set. In this case count the number of Gibbs sampler iterations such that the prevalence of interest is the largest. Schluter et al. (1997) have remarked that if $\nu = 1$, then (1) simply becomes the probability that π_i is the largest prevalence. The populations can be ranked via

i) the use of the posterior probability (1),
ii) the use of some probability threshold chosen such that the groups selected are the smallest subset where the sum of the $p_i(\nu)$ probabilities exceed the threshold, or
iii) the choice of $r < m$ largest probabilities to be included in the superior set.

A Predictive Probabilities Approach (Schluter, et al., 1997)

A second criterion is based on the predictive number of future occurrences in a future sample. The criterion is based on the probability that a future number of true positives, say W_i, exceeds some experimenter-chosen quantity, say w^*, or

$$pd_i(w^*) = P(W_i > w^* \mid \mathbf{z}, n_0) \qquad (2)$$

where n_0 represents the future sample size. To compute probability (2) with the Gibbs sampler, add the variables $W_i \mid \pi_i \sim$ binomial(n_0, π_i) for $i = 1, 2, \ldots, m$, to the likelihood. The approximation

$$pd_i(w_i) \approx \frac{\#(W_i \geq w^*)}{B}$$

is then straightforward to calculate. One can rank the populations via probability (2) and then either include the top r of them in a superior set or select all populations whose predictive probability (2) is greater than some user-specified value P_0. Difficulties with this criterion include determining a meaningful future sample size n_0 and defining a meaningful comparison number w^*.

A Decision Theoretic Approach (Bratcher & Bhalla (1974))

Stamey et al. (2004) used a constant loss function for Poisson parameters with misclassified data. Here a similar loss function for the binomial data case is utilized,

$$L(\pi) = \begin{cases} c_1 \#(S) + c_2 & \text{if } \pi_{\max} \notin S \\ c_1 \left[\#(S) - 1\right] & \text{if } \pi_{\max} \in S \end{cases},$$

where S denotes the superior set, $\#(S)$ denotes the number of parameters in the superior set, and $\pi_{\max} \in S$ represents placing the actual maximum proportion in the superior set. The corresponding risk is a linear combination of the expected size of the superior set and the probability of correct selection. Formally, the risk is

$$R(\pi) = c_1 E[\#(S)] + (c_1 + c_2)(1 - P(CS)) - c_1$$

where $P(CS)$ denotes probability of correct selection, i.e., $\pi_{\max}$ is selected. The Bayes threshold for inclusion is

$$p(\pi_i = \pi_{\max}) \geq 1/(c+1), \qquad (3)$$

where $c = c_2/c_1$. This loss ratio represents the relative seriousness of the two types of mistakes: leaving the largest parameter out of the superior set and putting a parameter in the superior set that is not the largest. Additionally, $c + 1$ may be considered the rate of change in $E[\#(S)]$ with respect to $P(CS)$. To guarantee that at least one parameter is placed in the superior set S, it is required that $c \geq m - 1$. The left side of (3) is approximated identically to (1) when $\nu = 1$. The estimated probabilities are then compared to $1/(c + 1)$, and the parameter π_k is placed in the superior set S when

$$p_i(\nu) \approx \frac{\#(\pi_{ik} = \max(\nu\pi_{1k}, ..., \nu\pi_{i-1,k}, \pi_{ik}, \nu\pi_{i+1,k}, ..., \nu\pi_{mk}))}{B}$$

$$> 1/(c + 1).$$

Results

The methods discussed are now applied to real data originally found in Hanson et al. (2003). Twenty cow herds in an area of Mexico where the disease is known to occur are sampled and tested with the buffered acidified plate agglutination (BAPA) serologic test. The BAPA is known to be imperfect, and its properties are discussed in Stemshorn et al. (1985). Point estimates of the sensitivity and specificity are .75 and .97, respectively. As in Hanson et al. (2003), this article used an equivalent sample size of 20 for the beta priors based on the prior means of .75 and .97, respectively. That is, seek beta priors with means of .75 and .97 where the sum of the parameters is 20; thus, $\eta \sim$ beta(15,5) and $\theta \sim$ beta(19.4, .6) was used. Had an equivalent sample size of 40 been used, it would have been assumed that $\eta \sim$ beta(30, 10) and $\theta \sim$ beta(38.8, 1.2). Interestingly, virtually identical inferences resulted from the two sets of priors with only a slight decrease in posterior variation. For this example the non-informative priors $\mu \sim$ beta(1, 1) and $\gamma \sim$ gamma(.001, .001) were used for the hierarchical parameters in the model for the prevalences.

WinBugs was used to approximate the posterior distributions. We show a plot of the approximate posterior densities for the bovine brucellosis prevalences in Figure 1. For this data one can visually see that clear differences exist among the posterior densities. The posterior distributions for the prevalences π_{15} and π_7 are centered at considerably larger values than the posterior distributions of the other prevalences.

Using (1), the posterior probabilities that each prevalence was the largest were calculated. Table 1 gives results for the posterior probabilities approach of selecting the largest prevalence for values of ν of 1, 1.1, and 1.25. In the table there are sites and corresponding posterior probabilities where $P(\pi_i > \nu\pi_j, \forall\, j \neq i \,|\, \mathbf{z})$ exceed 0.01 when $\nu = 1$. If one use criterion ii) with a threshold of .9 in conjunction with the posterior probability criterion, one can see from Table 1 that the two prevalences π_{15} and π_7 were the only elements contained in the superior set S using the posterior probability criterion. If the threshold had been increased to .99, then the prevalences π_{14} and π_{19} would be added to the superior set S. If one increases ν to 1.1 and 1.25, then it becomes evident from Table 1 that π_{15} is the sole choice for the largest prevalence.

Next, the predictino approach criterion is applied to the bovine brucellosis data. It was assumed a future sample size of $n_0 = 10$ and provided the probabilities for various values of w^*. Figure 2 is a plot of the results for values of w^* ranging from 0 to 5. For illustrative purposes supposed $w^* = 3$ and $P_0 = .8$, then placed a rectangle or box in the area of Figure 2 where the prediction criterion holds. All curves that fall inside the box, which in this case corresponded to the prevalences π_7, π_{14}, and π_{15}, satisfied the prediction criterion. The graph could easily be changed to allow for different values of P_0 and w^*.

Consider the decision theoretic approach to selecting herds with the largest bovine brucellosis prevalence. Only the prevalences π_{15}, π_7, and π_{14} would be selected at the boundary for the rate of change, $(c + 1) = 20$, which gave a critical probability of $1/(c + 1) = .05$. Thus, for this example we assumed it is 19 times more serious to leave the largest prevalence out of the superior set than to include a prevalence in the superior set S that is not the largest. If it were to be considered to be 99 more times serious to leave the largest prevalence out of the superior set than to include a prevalence in the superior set S that is not the largest, the critical probability would decrease to .01, and the prevalences π_{15}, π_7, π_{14}, and π_9 would be included in the superior set.

Auditing Application

As a second example, data were simulated similar to that found in Raats & Moors (2004). Suppose we wish to compare 15 locations in terms of the proportion of errors in accounts. As in Raats & Moors (2004), we assumed that the initial audit is fallible, that is, some accounts that are in error could be missed

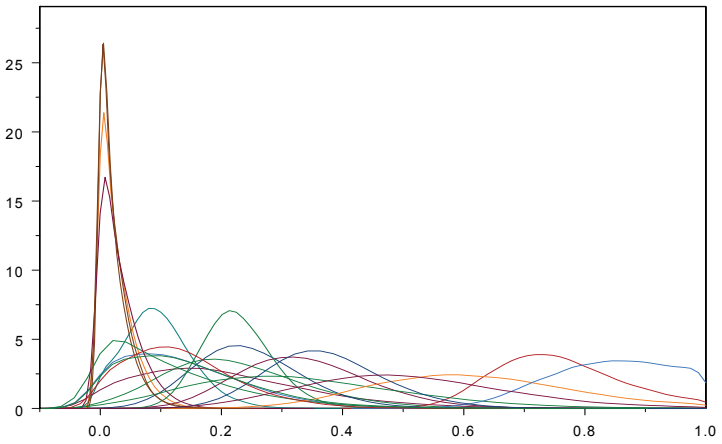

Figure 1. Posterior densities of prevalences for bovine data

Table 1. Posterior probabilities of having the largest prevalence

ν	Herd 15	Herd 7	Herd 14	Herd 19	Others
1	.773	.158	.052	.013	.004
1.1	.469	.032	.000	.000	.000
1.25	.123	.000	.000	.000	.000

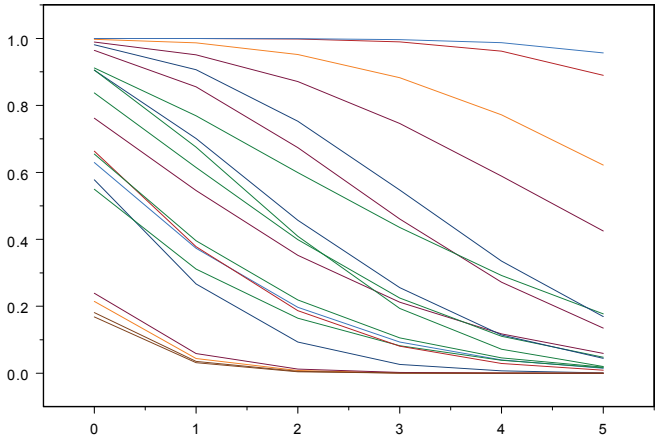

Figure 2. Predictive probabilities for bovine brucellosis data. The rectangle includes herds that satisfy a predictive probability of 3 or more events with probability greater than .8.

and some accounts that are correct could be labeled as in error. For each of the 15 locations, the parameters of the populations with the following distributions: $\pi_i \sim$ beta(2, 18), $\eta_i \sim$ beta(12, 8), and $\theta_i \sim$ beta(19, 1) were generated. These distributions are consistent with Raats & Moors (2004) in the sense that the overall proportion of errors is small with a mean of 10%, the sensitivity is moderate with a mean of 60%, and the specificity is high with a mean of 95%. For each site the following was generated $z_i \sim$ binomial(500, p_i), $t_i \sim$ binomial(60, π_i), $x_i \sim$ binomial(t_i, η_i), and $y_i \sim$ binomial(60 – t_i, θ_i), where $p_i = \pi_i\eta_i + (1 - \pi_i)(1 - \theta_i)$.

For the hierarchical model, allow for differential misclassification by using diffuse priors for all hyperprior distributions. Specifically, let $\mu \sim$ beta(1, 1), $\gamma \sim$ gamma(.001, .001), $\mu_\eta \sim$ beta(1, 1), $\gamma_\eta \sim$ gamma(.001, .001), $\mu_\theta \sim$ beta(1, 1), and $\gamma_\theta \sim$ gamma(.001, .001).

Two competing models were considered. The first was an independence-based model where each of the 15 sites was modeled independently and, thus, no information-sharing occurred among the sites. For the independence models beta(1, 1) priors were used for all parameters. Also considered was the hierarchical model of Hanson et al. (2003), previously used on the first example, where all the specificities and sensitivities were constant. For this non-differential misclassification model, the actual distributions from which the sensitivities and specificities were generated are used as the prior distributions. That is, the priors $\eta \sim$ beta(12, 8) and $\theta \sim$ beta(19, 1) were assumed. The generated proportions, posterior means of the validation data hierarchical model, and 95% intervals for all three models are provided in Table 3.

Note that the 95% intervals for the hierarchical model and the independence model both contained the true parameter values in all cases while the non-differential misclassification model missed two of the parameters. Also, the hierarchical model had the narrowest intervals, thus supporting the use of this model.

Table 4 gives the sites and corresponding posterior probabilities of having the largest prevalence for parameters where $P(\pi_i > \nu\pi_j, \forall\, j \neq i \mid \mathbf{z})$ exceed 0.01 when $\nu = 1$. Probabilities are provided for the case where $\nu =$ 1 and 1.1. Assuming criterion ii) with a probability threshold of .9, it was determined that the proportions π_8, π_1, and π_3, were included in the superior set because the sum of their probabilities is .923. In Table 4 are the three largest proportions used to generate the data in order from largest to smallest are π_8, π_3, and π_1. Thus, the posterior probability procedure included the three largest proportions in this example. If the threshold was increased to .99, then the proportions π_8, π_1, π_3, π_{10}, π_7, π_9, and π_2 would all be included in the superior set *S*.

If non-differential misclassification is incorrectly assumed, then one would have incorrectly concluded that π_{13} was the largest proportion with a corresponding posterior probability of .865 of being the largest proportion. Also, if the incorrect non-differential misclassification model were applied, one would have determined that the second largest proportion was π_{10} with a posterior probability of .106 of being the largest proportion. In this case the non-differential misclassification assumption leads to incorrect inferences because neither site 13 nor 10 was actually among the three largest proportions.

For this same data the prediction subset-selection criterion was applied. It was assumed a future sample size of 50. For the validation-data model with differential misclassification, the plot for all 15 sites for values of w^* from 0 to 6 is given in Figure 3. Included is a decision box for $w^* = 2$ and $P_0 = .6$. It was found that sites 1, 3, 7, 8, and 10 satisfied this particular configuration and, therefore, π_1, π_3, π_7, π_8, and π_{10} would be placed in the superior set. Recall that π_1, π_3, and π_8 were the largest three proportions so, again, this proposed prediction subset-selection criterion yielded very reasonable results.

For the decision theoretic approach, this article again considered *c*'s of 19 and 99 that yielded critical probabilities of .05 and .99. For a critical probability of .05, the proportions π_8, π_1, and π_3 were included in the superior set *S*.

Table 3. Posterior means and intervals for simulated auditing example
(Intervals that failed to cover the true parameter are bolded.)

Site	True value	Posterior mean differential hierarchical	95% Interval differential hierarchical	95% Interval independence	95% Interval non-differential hierarchical
1	.141	0.157	(.080, .244)	(.096, .290)	(.117, .465)
2	.076	0.102	(.051, .161)	(.049, .175)	(.000, .230)
3	.148	0.137	(.089, .202)	(.099, .258)	(.000, .162)
4	.017	0.047	(.012, .097)	(.004, .087)	(.000, 162)
5	.055	0.044	(.010, .093)	(.004, .083)	(.000, .174)
6	.131	0.100	(.055, .146)	(.054, .163)	(.000, .214)
7	.126	0.118	(.067, .180)	(.073, .229)	(.000, .205)
8	.201	0.190	(.128, .262)	(.145, .295)	(.122, .480)
9	.103	0.119	(.068, .175)	(.070, .183)	(.003, .267)
10	.101	0.120	(.063, .190)	(.067, .221)	**(.150, .542)**
11	.059	0.063	(.023, .108)	(.017, .117)	(.000, .219)
12	.092	0.090	(.046, .145)	(.037, .149)	(.000, .214)
13	.070	0.061	(.017, .115)	(.011, .113)	**(.200, .658)**
14	.119	0.079	(.037, .135)	(.029, .142)	(.072, .372)
15	.089	0.090	(.038, .153)	(.028, .163)	(.065, .361)

Table 4. Posterior probabilities of having the largest proportion of errors

ν	π_8	π_1	π_3	π_{10}	π_7	π_9
1	.618	.246	.059	.026	.021	.015
1.1	.452	.142	.023	.010	.007	.005

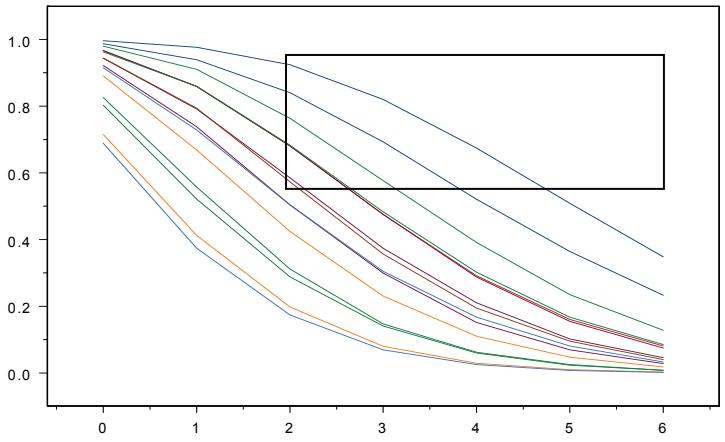

Figure 3. Predictive probabilities for auditing example. The rectangle includes populations that satisfied a predictive probability of 2 or more events with probability greater than .6.

For a critical probability of .01, the proportions π_{10}, π_7, and π_9 also entered the superior set. For the decision theoretic approach, this article again considered *c*'s of 19 and 99 that yielded critical probabilities of .05 and .99. For a critical probability of .05, the proportions π_8, π_1, and π_3 were included in the superior set *S*. For a critical probability of .01, the proportions π_{10}, π_7, and π_9 also entered the superior set.

Conclusion

In this article, three ranking criteria were applied to a hierarchical binomial model with misclassification first proposed in Hanson et al. (2003). These criteria are easy to use and understand and are computationally practical because of currently available statistical software. This has also extended the non-differential misclassification model of Hanson et al. (2003) to allow for differential misclassification. The example using simulated audit data with misclassified observations illustrates the importance of appropriately incorporating differential misclassification in the analysis. It is note that the Bayesian binomial parameter selection methods proposed here could also apply to psychology and medical subset-selection problems, where interest might lie in comparing various treatments when a fallible diagnostic test is used to assess presence of a particular psychological or medical condition. Finally, the computations in this article have been performed using WinBugs, which is a free statistical computing package available on the Internet.

References

Berger, R. L. (1980), Minimax subset selection for the multinomial distribution, *Journal of Statistical Planning and Inference 4,* 391-402.

Berger, R. L. (1979), Minimax subset selection for loss measured by subset size, *Annals of Statistics 7,* 1333-1338.

Berger, J. O., & Deely, J. (1988), A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology, *Journal of the American Statistical Association 83,* 364-373.

Bratcher, T.L., & Bhalla, P. (1974), On the properties of an optimal selection procedure, *Communications in Statistics - Theory and Methods 3,* 191-196.

Chaloner, K. (1996), Elicitation of prior distributions, *Bayesian Biostatistics*, 41-156.

Deely, J. J., & Smith, A. F. (1998), Quantitative refinements for comparison of institutional performance, *Journal of the Royal Statistical Society, A, 161*, 5-12.

Dendukuri, N., & Joseph, L. (2001), Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests, *Biometrics 57,* 208-217.

Dixon, D. O., & Simon, R. (1991), Bayesian subset analysis**,** *Biometrics 47*, 871-881.

Dixon, D. O., & Simon, R. (1994), Corrections: Bayesian subset analysis**,** *Biometrics 50*, 322.

Goel, P. K., & Rubin, H. (1977), On selecting a subset containing the best population-a Bayesian approach, *Annals of Statistics 5,* 969-983

Govindarajulu, Z., & Harvey, C. (1974), Bayesian procedures for ranking and selection problems, *Annals of the Institute of Statistical Mathematics 26*, 35-53.

Gupta, S. S., & Hsu, J. C. (1977), On the monotonicity of Bayes subset selection procedures, *Proceedings of the 41st Session of the International Statistical Institute 47*, New Delhi, India, 208–211.

Gupta, S. S., & Liang, T. (1987), On some Bayes and empirical Bayes selection procedures, *Probability and Bayesian Statistics* (Ed., R. Viertl), New York: Plenum Publishing Corporation, 233–246.

Gupta, S. S., & Sobel, M. (1957), On a statistic which arises in ranking and selection problems, *Annals of Mathematical Statistics 28*, 957-967.

Gupta, S. S., & Yang, H.-M. (1985), Bayes-P^{*} subset selection procedures, *Journal of Statistical Planning and Inference* 12, 213-233.

Hanson, T., Johnson, W., & Gardner, I. (2003), Hierarchical models for estimating herd prevalence and test accuracy in the absence of a gold standard, *Journal of Agricultural, Biological, and Environmental Statistics 8,* 223-239.

Joseph, L., Gyorkos, T. W., & Coupal, L. (1995), Bayesian Estimation of Disease Prevalence and Parameters of Diagnostic Tests in the Absence of a Gold Standard, *American Journal of Epidemiology 141,* 263-272.

Kadane, J. B., & Wolfson, L. J. (1996), Priors for the design and analysis of clinical trials, *Bayesian Biostatistics*, 157-184.

Raats, V. A., & Moors, J. J. A. (2004), Double-checking auditors: a Bayesian approach, *Journal of the Royal Statistical Society, D*, *52,* 351-366.

Schluter, D. C., Deely, J. J., & Nicholson, D.G. (1997), Ranking and Selecting Motor Vehicle Accident Sites Using a Hierarchical Bayesian Model, *The Statistician 46,* 293-316.

Stamey, J. D., Bratcher, T. L., & Young, D. M. (2004), Parameter subset selection and multiple comparisons of Poisson rate parameters with misclassified data, *Computational Statistics and Data Analysis 45,* 467479.

Stemshorn, B. W., Forbes, L. B., Eaglesome, M. D., Nielsen, K. H., Robertson, F. J., & Samagh, B. S. (1985), A comparison of standard serologic tests for the diagnosis of bovine brucellosis in Canada, *Canadian Journal of Comparative Medicine 49,* 391–394.