Journal of the Korean
Data & Information Science Society
2001, Vol. 12, No.1, pp. 1 ~ 9

Sampling Based Approach for Combining Results from Binomial Experiments

Jang Sik Cho [1], Dal Ho Kim [2] and Sang Gil Kang [3]

Abstract

In this paper, the problem of information related to I binomial experiments, each having a distinct probability of success θ_i, $i = 1, 2, \cdots, I$, is considered. Instead of using a standard exchangeable prior for $\theta = (\theta_1, \theta_2, \cdots, \theta_I)$, we consider a partition of the experiments and take the θ_i's belonging to the same partition subset to be exchangeable and the θ_i's belonging to distinct subsets to be independent. And we perform Gibbs sampler approach for Bayesian inference on θ conditional on a partition. Also we illustrate the methodology with a real data.

Key Words and Phrases: Beta-binomial, Hierarchical prior, Partial exchangeability, Gibbs sampler, Adaptive rejection sampling, Borrowing strength;

1. Introduction

Consider a collection of I independent binomial experiments, with experiment i having size n_i and success provbability $\theta_i, i = 1, 2, \cdots, I$. A typical Bayesian hierarchical approach would assume the θ_i's to be *exchangeability*. Exchangeability at times may appear to be too strong an assumption. So this approach leads shrinkaged estimate for θ. Hence, more generally, when one suspects that the experiments may have various degrees of similarity in some respect, one may wish to adopt a more flexible approach, involving entertaining several *partial exchangeability* structures for the θ_i's and then combining the corresponding inferences.

O'Hagan(1988) proposed a Bayesian procedure based on long-tailed prior distributions that may be useful in solving problems of the type for normal data. Malec

[1]Assistant Professor, Department of Statistical Information Science, Kyungsung University, Pusan, 608-736, Korea
[2]Associate Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea
[3]Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea

and Sedransk(1992) suggested that Bayesian hierarchical approach would assume the θ_i's to be *partial exchangeability* and that implemented with normal data. Also Consonni and Veronese(1995) considered combining inference on θ under the θ_i's to be partial exchangeability for binomial data. But the weakness of their approach is to use approximating the beta-binomial likelihood to lead Bayesian inference on θ.

In this paper, to overcome the problem for approximating the beta-binomial likelihood, we obtain the inference on θ by Gibbs sampler approach for combining the conditional inferences according to the posterior distribution of each partition. Also we illustrate the mothodology with a real data, showing in particular how classifying factors might help choose the collection of partitions. It also presents an empirical comparison with alternative methodologies, such as parametric empirical Bayes and standard logistic regression.

2. Notations and Preliminaries

Densities are denoted generically by brackets, so joint, conditional, and marginal forms, for example, appear as $[X, Y], [X|Y]$, and $[X]$. Let g be a partition of $\{1, 2, \cdots, I\}$ comprising $d(g)$ subsets $S_k(g)$, $k = 1, 2, \cdots, d(g)$ and let G be the total number of partitions of the set $\{1, 2, \cdots, I\}$. For example, if $I = 5$, then two possible partitions are $g_1 = \{\{1,3\}, \{2,4,5\}\}$ and $g_2 = \{\{1,3\}, \{2\}, \{4,5\}\}$. Clearly, $d(g_1) = 2$ with $S_1(g_1) = \{1,3\}$ and $S_2(g_1) = \{2,4,5\}$. Similarly, $d(g_2) = 3$ with $S_1(g_2) = \{1,3\}, S_2(g_2) = \{2\}$ and $S_3(g_2) = \{4,5\}$.

Let X_i given θ_i be independently distributed as binomial (n_i, θ_i), $i = 1, 2, \cdots, I$. Typically, there will be several partitions g whose relative plausibility is described by a prior probability mass function $p(g)$. We consider a *partition* of the experimental set $\{1, 2, \cdots, I\}$, whose subsets are $S_1(g), S_2(g), \cdots, S_d(g)$.

The basic assumption is to regard as exchangeable only the θ_i's associated with experiments belonging to the same partition subset $S_k(g)$, whereas the θ_i's relative to experiments in distinct subsets are taken to be independent. We wish our prior distribution for θ to reflect the beliefs that (a) there are subsets of θ such that the θ_i within each subset are 'similar', and (b) there is uncertainty about the composition of such subsets of θ.

To specify the prior distribution for θ, first condition on g, we may represent the desired similarity of the θ_i within a subset by assuming the following prior distribution: (a) there is independence from one subset another, and (b) within $S_k(g)$, and conditional on $\mu_k(g)$, the θ_i are independent with

$$[\theta_i|\mu_k(g)] \sim \text{Beta}\left(q_k\mu_k(g), q_k(1-\mu_k(g))\right), \tag{1}$$

where q_k is a known positive constant, and

$$[\mu_k(g)] \sim \text{Beta}\left(r_k m_k, r_k(1-m_k)\right), \tag{2}$$

where $0 < m_k < 1$ and $r_k > 0$ both known.

Finally, our prior beliefs about the set of specifications in (1) and (2) for $g = 1, 2, \cdots, G$ are denoted by $p(g)$ with $\sum_{i=1}^{G} p(g) = 1$; that is,

$$P(\text{the elements of } \mu \text{ are arranged according to partition } g). \tag{3}$$

In practice, many of the $p(g)$ would be assigned equal mass, $p(g) = 1/G$.

To compute the posterior distribution of θ given the data $\mathbf{x} = (x_1, \cdots, x_I)$, $p(\theta|\mathbf{x})$, one typically first derives $p(\theta|\mu, \mathbf{x})$, where $\mu = (\mu_1, \cdots, \mu_d)$, and then integrates it with respect to $p(\mu|\mathbf{x})$. It is immediate to verify that for $i \in S_k(g)$

$$[\theta_i|\mathbf{x}, \mu_k(g)] \sim \text{Beta}(q_k\mu_k(g) + x_i,\ q_k(1 - \mu_k(g)) + n_i - x_i). \tag{4}$$

On the other hand, computation of $p(\mu|\mathbf{x})$ becomes analytically intractable, because the marginal distribution of $\mathbf{X}$ given μ is a product of beta-binomials, with general term

$$p(x_i|\mu_k(g)) = \binom{n_i}{x_i} \frac{B(s_i + q_k\mu_k(g),\ n_i - s_i + q_k(1 - \mu_k(g)))}{B(q_k\mu_k(g),\ q_k(1 - \mu_k(g)))}, \tag{5}$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$. Several approximations of the beta-binomial distribution have been suggested by some authors. But they do not lead to closed-form posteriors for μ. So Consonni and Veronese (1995) obtained the approximated the likelihood function for $\mu_k(g)$(See Consonni and Veronese(1995)).

3. Sampling Based Approach

In section 2, the weakness of Consonni and Veronese(1995)'s approach is to use approximating beta-binomial likelihood. To overcome the problems for approximating, we use Gibbs sampling by Gelfand and Smith(1990, 1991). Also Gelman and Rubin(1992) introduced iterative simulation using multiple sequences. In this section, we use Gelman and Rubin's method.

For convenience, we let $\mu_k(g) = \mu_k$. The Gibbs sampling analysis is based on the following posterior distributions:

(I) $[\theta_i|\mathbf{x}, \mu_k, g] \sim \text{Beta}\ (x_i + q_k\mu_k, n_i - x_i + q_k(1 - \mu_k))$, for $i \in S_k(g)$.

(II) For each $k = 1, 2, \cdots, d(g)$, $[\mu_k|\mathbf{x}, \theta, g]$ has pdf:

$$p(\mu_k|\mathbf{x}, \theta, g) \ \propto \prod_{i \in S_k(g)} \frac{\theta_i^{q_k\mu_k}(1-\theta_i)^{q_k(1-\mu_k)}\mu_k^{r_k m_k - 1}(1-\mu_k)^{r_k(1-m_k)-1}}{\Gamma(q_k\mu_k)\Gamma(q_k(1-\mu_k))}$$

(III) $[g|\mathbf{x}, \theta, \mu]$ has pdf:

$$p(g|\mathbf{x}, \theta, \mu) \propto \prod_{k=1}^{d(g)} \prod_{i \in S_k(g)} \frac{\Gamma(q_k)\Gamma(r_k)\theta_i^{q_k\mu_k}(1-\theta_i)^{q_k(1-\mu_k)}\mu_k^{r_k m_k - 1}(1-\mu_k)^{r_k(1-m_k)-1}}{\Gamma(q_k m_k)\Gamma(q_k(1-\mu_k))\Gamma(r_k m_k)\Gamma(r_k(1-m_k))} \quad (6)$$

To estimate the posterior moments, we use Rao-Blackwellized estimates as in Gelfand and Smith(1991). Using above step (I), the posterior mean and the posterior variance are given by

$$E[\theta_i|\mathbf{x}, g] = E[E(\theta_i|\mu_k, g, \mathbf{x})|\mathbf{x}] = E\left[\frac{x_i + q_k\mu_k}{n_i + q_k}|\mathbf{x}\right], \quad (7)$$

$$\begin{aligned} Var[\theta_i|\mathbf{x}, g] &= E[Var(\theta_i|\mathbf{x}, \mu_k, g)|\mathbf{x}] + Var[E(\theta_i|\mathbf{x}, \mu_k, g)|\mathbf{x}] \\ &= E\left[\frac{(x_i + q_k\mu_k)(n_i - x_i + q_k(1-\mu_k))}{(n_i + q_k)^2(n_i + q_k + 1)}|\mathbf{x}\right] \\ &\quad + Var\left[\frac{x_i + q_k\mu_k}{n_i + q_k}|\mathbf{x}\right]. \end{aligned} \quad (8)$$

Thus we can calculate the $E(\theta_i|\mathbf{x})$ and $Var(\theta_i|\mathbf{x})$, respectively as follow,

$$E(\theta_i|\mathbf{x}) = \sum_g P(g|\mathbf{x})E(\theta_i|\mathbf{x}, g), \quad (9)$$

$$\begin{aligned} V(\theta_i|\mathbf{x}) &= E[V(\theta_i|\mathbf{x}, \beta_k(g))] + V[E(\theta_i|\mathbf{x}, \beta_k(g))] \\ &= \sum_{g=1}^{G} V(\theta_i|\mathbf{x}, g)P(g|\mathbf{x}) + \sum_{g=1}^{G} E(\theta_i|\mathbf{x}, g)^2 P(g|\mathbf{x}) \\ &\quad - \left(\sum_{g=1}^{G} E(\theta_i|\mathbf{x}, g)P(g|\mathbf{x})\right)^2. \end{aligned} \quad (10)$$

Using (7) and (8), $E(\theta_i|\mathbf{x}, g)$ and $Var(\theta_i|\mathbf{x}, g)$ are approximated by

$$E(\theta_i|\mathbf{x}, g) \approx (nm)^{-1} \sum_{j=1}^{m} \sum_{l=n+1}^{2n} \left[(x_i + q_k\mu_{kj}^{(l)})(n_i + q_k)^{-1}\right], \quad (11)$$

$$Var(\theta_i|\mathbf{x}, g) \approx (nm)^{-1} \sum_{j=1}^{m} \sum_{l=n+1}^{2n} \left(\frac{(x_i + q_k\mu_{kj}^{(l)})(n_i - x_i + q_k(1 - \mu_{kj}^{(l)}))}{(n_i + q_k)^2(n_i + q_k + 1)}\right)$$

$$+(nm)^{-1}\sum_{j=1}^{m}\sum_{l=n+1}^{2n}\left(\frac{x_i+q_k\mu_{kj}^{(l)}}{n_i+q_k}\right)^2$$
$$-\left((nm)^{-1}\sum_{j=1}^{m}\sum_{l=n+1}^{2n}\left(\frac{x_i+q_k\mu_{kj}^{(l)}}{n_i+q_k}\right)\right)^2, \quad (12)$$

where $\mu_k(g)_j^{(l)}$ denotes the generated value in lth iteration of jth chain.

In (6), $p(g|\mathbf{x})$ is approximated by $p(g|\mathbf{x})/\sum_{g=1}^{G}p(g|\mathbf{x})$, where

$$p(g|\mathbf{x}) \approx \frac{A}{mn}\sum_{j=1}^{m}\sum_{l=n+1}^{2n}\left(\prod_{k=1}^{d(g)}\prod_{i\in S_k(g)}\frac{\Gamma(q_k)\Gamma(r_k)\theta_i^{q_k\mu_{kj}^{(l)}}(1-\theta_i)^{q_k(1-\mu_{kl}^{(l)})}}{\Gamma(r_km_k)\Gamma(r_k(1-m_k))}\right)\cdot$$
$$\left(\prod_{k=1}^{d(g)}\prod_{i\in S_k(g)}\frac{\mu_{kj}^{(l)\,r_km_k-1}(1-\mu_{kj}^{(l)})^{r_k(1-m_k)-1}}{\Gamma(q_km_k)\Gamma(q_k(1-\mu_{kj}^{(l)}))}\right), \quad (13)$$

where A is the norming constant. Therefore, $E(\theta_i|\mathbf{x})$ and $Var(\theta_i|\mathbf{x})$ are approximated by (9)-(11).

In implementing the Gibbs sampler, one should be able to draw samples from the conditional densities given in (I)-(III). Simulation from the conditional densities (I) which is beta density can be done by standard methods. However, the posterior pdf of μ_k given $\mathbf{x},\theta$ and g is known only up to a multiplicative constant. In order to simulate from this density, one general approach is to use the Metropolis-Hasting accept-reject algorithm. Fortunately, the task becomes simpler for us because of the following result.

Lemma For each $k=1,2,\cdots,d(g)$, $logp(\mu_k|\mathbf{x},\theta,g)$ is a concave function of μ_k.

Proof Let $|S_k(g)|$ denotes the number of elements in $S_k(g)$ and consider $p(\mu_k|\mathbf{x},\theta,g) \propto \prod_{i\in S_k(g)}\frac{\theta_i^{q_k\mu_k}(1-\theta_i)^{q_k(1-\mu_k)}\mu_k^{r_km_k-1}(1-\mu_k)^{r_k(1-m_k)-1}}{\Gamma(q_k\mu_k)\Gamma(q_k(1-\mu_k))}$. Then

$$\begin{aligned}
&logp(\mu_k|\mathbf{x},\theta,g)\\
&=C+\sum_{i\in S_k(g)}[q_k\mu_klog(\theta_i)+q_k(1-\mu_k)log(1-\theta_i)]\\
&+|S_k(g)|\cdot[(r_km_k-1)log(\mu_k)+(r_k(1-m_k)-1)log(1-\mu_k)]\\
&-|S_k(g)|\cdot[log(\Gamma(q_k\mu_k))+log(\Gamma(q_k(1-\mu_k)))]
\end{aligned} \quad (14)$$

where C is the norming constant. Hence,

$$\frac{\partial}{\partial\mu_k}logp(\mu_k|\mathbf{x},\theta,g)$$

$$= \sum_{i \in S_k(g)} \left[q_k log\left(\frac{\theta_i}{1-\theta_i}\right)\right] + |S_k(g)| \cdot \left(\frac{r_k m_k - 1}{\mu_k} - \frac{r_k(1-m_k)-1}{1-\mu_k}\right)$$
$$-|S_k(g)| \cdot \left(\frac{\int_0^\infty e^{-z} z^{q_k\mu_k} q_k log(z)dz}{\Gamma(q_k\mu_k+1)} - \frac{1}{\mu_k}\right)$$
$$-|S_k(g)| \cdot \left(\frac{\int_0^\infty e^{-z} z^{q_k(1-\mu_k)}(-q_k log(z))}{\Gamma(q_k(1-\mu_k)+1}) + \frac{1}{1-\mu_k}\right). \quad (15)$$

Therefore,

$$\frac{\partial^2}{\partial \mu_k^2} logp(\mu_k|\mathbf{x}, \theta, g)$$
$$= -|S_k(g)| \cdot \left(\frac{r_k m_k}{\mu_k^2} + \frac{r_k(1-m_k)}{(1-\mu_k)^2}\right)$$
$$-|S_k(g)| \cdot (Var_I(q_k log(z)) + Var_{II}(-q_k log(z))) < 0, \quad (16)$$

where Var_I and Var_{II} are the variances for Gamma$(1, q_k\mu_k+1)$ and Gamma$(1, q_k(1-\mu_k)+1)$. The proof is completed.

Because of the log-concavity of this posterior density, we can use the adaptive rejection sampling algorithm of Gilks and Wild(1992) to simulate from this density(See Gilds and Wild(1992)).

4. An Example

In this section, we illustrate the results of section 3 by providing a complete Bayesian hierarchical analysis of the mortality of pine seedlings data previously analyzed by Fienberg(1980). For more details on this experiment and analysis of the data using a logistic regression model, see Fienberg(1980). Our object is to compare and contrast the present hierarchical Bayes method with Fienberg's logistic regression method and parametric empirical Bayes method.

Because the number of plants for each combination of Seedling Type (LS = Longleaf Seedling, SS = Slash Seedling) and Depth of Planting(DH = Depth too High, DL = Depth too Low) was fixed by design, we have four binomial experiments each of size $n_i = 100$, $i = 1, \cdots, 4$, corresponding to the four combinations(LS,DH),(LS, DL), (SS, DH), and (SS, DL). The data are reported in the first three columns of Table 1.

Performing Gibbs sampling approach, we choose the parameters of prior and hyperprior distributions as follows: In Bayesian hierarchical modeling it is customary to assign a noninformative prior to the last stage parameters (μ_k in our case). So we take into consideration is the uniform distribution on μ_k. This amounts to setting $r_k = 2$ and $m_k = 1/2$ in (2). For the values of $q_k(g), g = 1, 2, \cdots, 6$, we choose any value on the interval [1, 5]. In deriving the hierarchical Bayes estimates of the present section, we have considered Gibbs sampler with $m = 5$ and $2n = 10000$. A sample of size 10,000 is taken to obtain the Monte-Carlo estimates, as stability seems to be achieved with this sample size.

In Table 1, x_i denotes count of alive seedlings out of 100 in experiment i, $\hat{\theta}_i$ is observed survival rate(MLE), that is, $\hat{\theta}_i = x_i/n_i$, θ_i^{Gibbs} denotes hierarchical Bayes estimates using Gibbs sampler approach, θ_i^* denotes estimate of θ_i based on the logistic regression model by Fienberg(1980), and θ_i^{PEB} denotes parametric empirical Bayes estimate by Consonni and Veronese(1995).

Table 2 reports possible partitions and the posterior probability of a selected collection of partitions that are most supported by the data, besides those relative to g_1 for the independent model and g_6 for the exchangeable model.

It is immediately recognizable that experiment 1 is isolated, indicating a lack of exchageability relative to the other experiments. So partition $\{1, 2, 3, 4\}$ has the lowest posterior probability in Table 2. On the other hand, partition $\{\{1\}, \{2, 3, 4\}\}$ has the highest posterior probability, showing a substantial degree of similarity for experiments 2, 3, and 4. This is clearly reflected in the posterior expectation of θ_1, which is only very mildly affected by the results of experiments 2, 3, and 4. θ_1^* and θ_1^{PEB} lead shrinkaged estimate towards the results of experiments 2, 3, and 4. But θ_1^{Gibbs} turn out to be fairly close to the MLE, $\hat{\theta}_1$. On the other hand, the estimates of θ_4 are more strongly influenced by the data of experiments 2 and 3 due to a substantial borrowing strength effect, which does lower $\hat{\theta}_4$ as θ_4^{Gibbs} and θ_4^{PEB}. But

Table 1: Data, Estimates of θ_i, and (Standard Errors)

Experiment	i	x_i	$\hat{\theta}_i$	θ_i^{Gibbs}	θ_i^*	θ_i^{PEB}
LSDH	1	59	0.59	0.589(0.048)	0.606	0.610(0.047)
LSDL	2	89	0.89	0.885(0.031)	0.874	0.885(0.030)
SSDH	3	88	0.88	0.876(0.033)	0.864	0.875(0.031)
SSDL	4	95	0.95	0.945(0.022)	0.966	0.940(0.022)

Table 2: Posterior Probabilities for a Selected Collection of Partitions g

g	Partition	$P(g\|\mathbf{x})$
1	{ {1}, {2}, {3}, {4} }	0.180
2	{ {1}, {2,3}, {4}}	0.114
3	{{1}, {2,4}, {3} }	0.001
4	{{1}, {2}, {3,4}}	0.114
5	{{1}, {2,3,4}}	0.482
6	{1,2,3,4}	0.106

θ_4^* based on the logistic model is higher than $\hat{\theta}_4$ due to the additive positive effects related to Slash Seedling and Depth too Low.

In conclusion, our analysis clarifies that experiment 1 is clearly distinct from the other experiments, which can be regarded as essentially similar. In other words, the effect of planting is appreciable only in Depth too High in conjunction with Longleaf Seedlings.

References

(1) Consonni,G. and Veronese,P.(1995). A Bayesian Method for Combining Results From Several Binomial Experiments, *Journal of the American Statistical Association*, 90, 935-944.

(2) Fienberg, S.E.(1980). *The Analysis of Cross-Classified Categorical Data*(2nd ed.), Cambridge, MA: The MIT Press.

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

(4) Gelfand,A.E. and Smith,A.F.M.(1991). Gibbs Sampling for Marginal Posterior Expectations, *Communication in Statatistics-Theory and Methods*, 20, 1747-1766.

(5) Gelman,A. and Rubin,D.B.(1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Sciences*, 7, 457-511.

(6) Gilks, W.R. and Wild, P.(1992). Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics*, 41, 337-348.

(7) Malec,D. and Sedransk,J.(1992). Bayesian methodology for the results from different experiments when the specifications for pooling are uncertain, *Biometrika*, 79, 593-601.

(8) O'Hagan,A.(1988). Modelling with Heavy Tails, *Bayesian Ststistics*, 3, 345-359.

(9) Rao (1984). Variance components models for combining estimators. *W. G. Cochran's Impact on Statistics*, 203-221.