The Korean Communications
in Statistics Vol. 6, No. 1, 1999
pp. 181-191

Bayesian Methods for Combining Results from Different Experiments[1]

In Suk Lee[2], Dal Ho Kim[3] and Keun Baik Lee[4]

Abstract

We consider Bayesian models which allow multiple grouping of parameters for the normal means estimation problem. In particular, we consider a typical Bayesian hierarchical approach based on the partial exchangeability where the components within a subgroup are exchangeable, but the different subgroups are not. We discuss implementation of such Bayesian procedures via Gibbs sampling. We illustrate the proposed methods with numerical examples.

1. Introduction

Recently there has been an increased interest in methods for combining results from several experiments or observational studies. Given data from several experiments or observational studies initially believed to be similar, it is desired to estimate the means corresponding to one or more experiments of particular interest. A class of prior distributions for the means is specified to reflect the belief that there are subsets of means such that the means within each subset are similar, but the composition of such subsets is uncertain. One reason for this idea is the ability to make reliable inference for each experiment (or study) by "borrowing strength" from other selected experiments (or studies).

The related method is the meta-analysis, which combines inferential summaries from several studies into a single analysis. DuMouchel and Harris (1983) proposed Bayesian methods for combining the results of cancer studies in humans and other species. DuMouchel (1990) described and illustrated the use of Bayesian hierarchical models for meta-analysis. Hedges and Olkin (1985) surveyed the frequentist statistical methods of meta-analysis.

Often the population means are clustered into two or more subgroups. In this case, a typical Bayesian hierarchical approach, based on the exchangeability assumption, would have the weakness of Bayesian models. A useful substitute for exchangeability in this situation is

1) This paper was supported by Kyungpook National University Research Fund, 1997.
2) Professor, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.
3) Assistant Professor, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.
4) Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.

partial exchangeability, where the components within a subgroup are exchangeable, but the different subgroups are not. This idea has been suggested by Malec and Sedransk (1992) and implemented with normal data. Later Consonni and Veronese (1995) considered combining results from several binomial experiments. Recently Efron (1996) described empirical Bayes methods for combining likelihoods to get an interval estimate for any one of the log-odds ratio for a 2×2 table relating to the kth population in a series of medical experiments.

In this paper, we consider Bayesian models which allow multiple grouping of parameters for the normal means estimation problem. Such models allow different shrinkage within each group. This kind of behavior is termed "multiple shrinkage" by George (1986). We discuss implementation of such Bayesian procedures via Gibbs sampling. In Section 2, we review the idea of partially exchangeable model based on Malec and Sedransk (1992). In Section 3, we study a Bayesian analysis with partially exchangeable priors based on Gibbs sampling which is recently popularized by Gelfand and Smith (1990). This is Gibbs sampling version of Malec and Sedransk (1992). In Section 4, we illustrate the proposed methods using two numerical examples with no covariates.

2. Partially Exchangeable Models

Consider a collection of I independent normal experiments with experiment i having mean μ_i and variance σ_i^2. That is, the y_{ij} are independent with $y_{ij} \sim N(\mu_i, \sigma_i^2)$ for $i=1,\cdots,I,\ j=1,\cdots,n_i$. A typical Bayesian hierarchical approach would assume the μ_i's to be exchangeable. Exchangeability at times may appear to be too strong an assumption. For example, assume that, $\mu_i \mid \nu, \delta^2 \sim N(\nu, \delta^2)$ independently for each i and ν has an uniform prior on $(-\infty, \infty)$ where σ_i^2 and δ^2 are assumed to be known. Then the posterior expected values of μ_i is

$$E(\mu_i \mid \mathbf{y}, \delta^2) = \lambda_i\, \hat{\mu}_i + \left\{ (1-\lambda_i) \sum_{j=1}^{I} \lambda_j\, \hat{\mu}_j \Big/ \sum_{j=1}^{I} \lambda_j \right\} \tag{2.1}$$

and

$$Var(\mu_i \mid \mathbf{y}, \delta^2) = \frac{(\sigma_i^2/n_i)\delta^2}{\sigma_i^2/n_i + \delta^2} + \frac{\{(\sigma_i^2/n_i)/(\sigma_i^2/n_i + \delta^2)\}^2}{\sum_{i=1}^{I}(\sigma_i^2/n_i + \delta^2)}, \tag{2.2}$$

where $\mathbf{y}$ denotes the observed data and $\lambda_i = \delta^2 / \{\delta^2 + (\sigma_i^2/n_i)\}$.

The weakness of this approach is that the amount of pooling in (2.1) is specified by the prior distribution. This may lead to unsatisfactory inferences when, for example, it is found that $\hat{\mu}_1, \cdots, \hat{\mu}_k$ are each close to $\hat{\mu}_{(1)}$ while $\hat{\mu}_{k+1}, \cdots, \hat{\mu}_I$ are each close to $\hat{\mu}_{(2)}$ and $\hat{\mu}_{(1)} \ll \hat{\mu}_{(2)}$. Here, estimation of μ_1 using (2.1) would include, perhaps inappropriately, a large contribution from $\hat{\mu}_{k+1}, \cdots, \hat{\mu}_I$. The difficulty is that the prior distribution for $\boldsymbol{\mu}$ is not sufficiently flexible. To overcome these problems, one may wish to adopt a more flexible approach, involving several partial exchangeability structures for the μ_i and then combining the corresponding inferences.

Let G denote the total number of partitions of the set $B=\{1, \cdots, I\}$. Denote a particular partition by g $(g=1, \cdots, G)$, and $d(g)$ denotes the number of subsets of B in the gth partition $(1 \le d(g) \le I)$. Also, let $S_k(g)$ denote the set of experiment labels in subset k for $k=1, \cdots, d(g)$. For example, for $I=3$ the partitions are $g_1=\{1;2;3\}$, $g_2=\{1,2;3\}$ $g_3=\{1,3;2\}$, $g_4=\{2,3;1\}$ and $g_5=\{1,2,3\}$, and $G=5$. Clearly $d(g_2)=2$ with $S_1(g_2)=\{1,2\}$ and $S_2(g_2)=\{3\}$.

The basic assumption is to regard as exchangeable only the μ_i's associated with experiments belonging to the same partition subset $S_k(g)$, whereas the μ_i's relative to experiments in distinct subsets are taken to be independent.

Specifically, we consider the prior distribution for $\boldsymbol{\mu}$ by first conditioning on g. Then for a given partition g, there is independence from one subset to another, and within $S_k(g)$, and conditional on $\nu_k(g)$, the expected value of the experiment means in $S_k(g)$, the μ_i are independent with

$$\mu_i | \nu_k(g) \sim N\{\nu_k(g), \delta_k^2(g)\}, \quad (i \in S_k(g)) \tag{2.3}$$

and

$$\nu_k(g) \sim N(\theta_k(g), \gamma_k^2(g)). \tag{2.4}$$

where $\delta_k^2(g)$, $\theta_k(g)$ and $\gamma_k^2(g)$ are assumed to be known. Our prior beliefs about the set of specifications in (2.3) and (2.4) for $i=1, \cdots, G$ are denoted by $\{p(g)\}$ with $\sum_{g=1}^{G} p(g)=1$; that is,

$$P(\text{ the elements of } \boldsymbol{\mu} \text{ are arranged according to partition } g\ ) = p(g).$$

In particular, $p(g)$ takes equal mass. That is, $p(g)=1/G$.

Conditional on partition g, it can be shown that,

$$E(\mu_i | \boldsymbol{y}, g) = \lambda_i(g)\ \hat{\mu}_i + \{1-\lambda_i(g)\}\phi_k(g)\ \hat{\mu}_k(g) + \{1-\lambda_i(g)\}\{1-\phi_k(g)\}\theta_k(g) \qquad (i \in S_k(g)) \tag{2.5}$$

and

$$Cov(\mu_i, \mu_t | \boldsymbol{y}, g) = \begin{cases} \delta_k^2(g)\{1-\lambda_i(g)\} + \{1-\lambda_i(g)\}^2\gamma_k^2(g)\{1-\phi_k(g)\} & (i=t;\ i, t \in S_k(g)), \\ \{1-\lambda_i(g)\}\{1-\lambda_t(g)\}\gamma_k^2(g)\{1-\phi_k(g)\} & (i \neq t;\ i, t \in S_k(g)), \\ 0 & (i \in S_{k_1}, t \in S_{k_2}, k_1 \neq k_2), \end{cases} \tag{2.6}$$

where

$$\lambda_i(g) = \delta_k^2(g)\{\delta_k^2(g) + (\sigma_i^2/n_i)\}^{-1},$$

$$\phi_k(g) = \delta_k^2(g)\left[\delta_k^2(g) + \{\delta_k^2(g) \Big/ \sum_{t \in S_k(g)} \lambda_t(g)\}\right]^{-1},$$

$$\hat{\mu}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$$

and

$$\hat{\mu}_k(g) = \sum_{t \in S_k(g)} \{\lambda_t(g)\}\ \hat{\mu}_t \Big/ \sum_{t \in S_k(g)} \lambda_t(g).$$

The estimator of μ_i in (2.5) is a convex combination of (a) the sample mean, $\hat{\mu}_i$, from experiment i, (b) an estimator, $\hat{\mu}_k(g)$, using data from all experimental units in $S_k(g)$, and (c) the mean, $\theta_k(g)$, from the prior distribution (2.4). In particular, $\hat{\mu}_k(g)$ exhibits the "gaining of strength" for estimation of the mean μ_i corresponding to experiment i.

Let $f(\boldsymbol{y}|g)$ denote the given normal likelihood. Then the posterior probability of g given $\boldsymbol{y}$ is given as

$$\begin{aligned} p(g|\boldsymbol{y}) &= f(\boldsymbol{y}|g)p(g) \Big/ \sum_{g=1}^{G} \{f(\boldsymbol{y}|g)p(g)\} \\ &\propto p(g) \prod_{k=1}^{d(g)} \{1-\phi_k(g)\}^{\frac{1}{2}} \prod_{i \in S_{k(g)}} \{1-\lambda_i(g)\}^{1/2} \\ &\quad \times \exp\left(-\frac{1}{2}\left[\sum_{k=1}^{d(g)} \sum_{i \in S_k(g)} \lambda_i(g) \frac{\{\hat{\mu}_i - \hat{\mu}_k(g)\}^2}{\delta_k^2(g)} + \sum_{k=1}^{d(g)} \phi_k(g) \frac{\{\hat{\mu}_k(g) - \theta_k(g)\}^2}{\gamma_k^2(g)}\right]\right). \end{aligned} \tag{2.7}$$

Then we can obtain the moments of posterior distribution as follows.

$$E(\mu_i|\, \mathbf{y}) = \sum_g p(g|\, \mathbf{y}) E(\mu_i|\, \mathbf{y}, g) \tag{2.8}$$

$$Var(\mu_i|\, \mathbf{y}) = E[\, Var(\mu_i|\, \mathbf{y}, g) \,|\, \mathbf{y}] + Var[E(\mu_i|\, \mathbf{y}, g) \,|\, \mathbf{y}] \tag{2.9}$$

and

$$Cov(\mu_i, \mu_j|\, \mathbf{y}) = E[\, Cov(\mu_i, \mu_j|\, \mathbf{y}, g) \,|\, \mathbf{y}] + Cov[E(\mu_i|\, \mathbf{y}, g), E(\mu_j|\, \mathbf{y}, g) \,|\, \mathbf{y}] \tag{2.10}$$

Note that the moments (2.8)-(2.10) are based on "model averaging" in Draper (1995) and Maleck and Sedransk (1992).

3. Sampling-Based Approach

In this section, we obtain the Bayes estimator of $\boldsymbol{\theta}$ under the partially exchangeable model using sampling-based approach. We consider the following hierarchical Bayesian model.

$$y_{ij} \overset{\text{ind}}{\sim} N(\mu_i, \sigma_i^2) \quad (i = 1, \cdots, I, \ \ j = 1, \cdots, n_i), \tag{3.1}$$

$$\mu_i | \nu_k(g) \ \sim \ N\{\nu_k(g), \delta_k^2(g)\} \quad (i \in S_k(g), \ \ k = 1, \cdots, d(g)) \tag{3.2}$$

$$\nu_k(g) \,|\, g \ \sim \ \text{uniform}(-\infty, \infty) \tag{3.3}$$

and

$$g \sim p(g) \quad (g = 1, \cdots, G) \tag{3.4}$$

For implementation of our Bayesian procedures, we use Gibbs sampler using the multiple sequences which is suggested by Gelman and Rubin (1992). This method is to use several independent sequences, with starting points sampled from an overdispersed distribution. Specifically, we run $c\ (\geq 2)$ parallel chains, each for $2q$ iterations. To diminish the effects of the starting distributions, the first q iterations of each chain are discarded. After q iterations, all the subsequent iterates are retained for finding the desired posterior distributions, posterior mean and variance, as well as for monitoring the convergence of the Gibbs sampler.

The joint distribution of $(\mathbf{y}, \boldsymbol{\mu}, \nu, g)$ can be written as

$$f(\mathbf{y}, \boldsymbol{\mu}, \nu, g) \propto \left\{ \prod_{k=1}^{d(g)} \prod_{i \in S_k(g)} f(\mathbf{y}_i | \mu_i) \right\} p(g) \prod_{k=1}^{d(g)} \prod_{i \in S_k(g)} \left[\frac{1}{\delta_k(g)} \exp\left\{ - \frac{(\mu_i - \nu_k(g))^2}{2\delta_k^2(g)} \right\} \right],$$

where

$$\mathbf{y} = (\mathbf{y}_1, \cdots, \mathbf{y}_I),$$

$$\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i}) \quad (i = 1, \cdots, I)$$

and

$$\boldsymbol{\mu}=(\mu_1,\cdots,\mu_I).$$

From the joint distribution we can obtain all the full conditional distributions as follows.

$$f(\boldsymbol{\mu}|\,\boldsymbol{y},\nu_k(g),g)$$
$$\propto \prod_{k=1}^{d(g)}\prod_{i\in S_k(g)}\frac{1}{\sqrt{n_i/\sigma_i^2+1/\delta_k^2(g)}}\exp\left[-\frac{(\mu_i-((1-\lambda_i(g))\nu_k(g)+\lambda_i(g)\,\bar{y}_{i.}))^2}{2(n_i/\sigma_i^2+1/\delta_k^2(g))}\right] \tag{3.5}$$

where

$$\lambda_i(g)=\frac{\delta_k^2(g)}{\sigma_i^2/n_i+\delta_k^2(g)}\qquad(i=1,\cdots,I).$$

And

$$f(\nu_k(g)|\,\boldsymbol{y},\boldsymbol{\mu},g)$$
$$\propto\frac{\sqrt{m(g)}}{\delta_k(g)}\exp\left[-\frac{m(g)}{2\delta_k^2(g)}(\nu_k(g)-\sum_{i\in S_k(g)}\mu_i/m(g))^2\right] \tag{3.6}$$

where $m(g)$ is the size of $S_k(g)$ under the partition g. Finally,

$$p(g|\,\boldsymbol{y},\boldsymbol{\mu},\nu_k(g))$$
$$\propto\prod_{k=1}^{d(g)}\prod_{i\in S_k(g)}\frac{1}{\delta_k(g)}\exp\left[-\frac{1}{2\delta_k^2(g)}(\mu_i-\nu_k(g))^2\right]. \tag{3.7}$$

Thus we obtain the following full conditional distributions ;

(i) $\mu_i|\,\boldsymbol{y},\nu_k(g),g\sim N[(1-\lambda_i(g))\nu_k(g)+\lambda_i(g)\,\bar{y}_{i.},\ n_i/\sigma_i^2+1/\delta_k^2(g)]$ $(i\in S_k(g),\ k=1,\cdots,d(g))$;

(ii) $\nu_k(g)|\,\boldsymbol{y},\boldsymbol{\mu},g\sim N[\sum_{i\in S_k(g)}\mu_i/m(g),\ \delta_k^2(g)/m(g)]\quad(k=1,\cdots,d(g))$;

(iii)
$$p(g|\,\boldsymbol{y},\boldsymbol{\nu},\nu_k(g))=\left[\prod_{k=1}^{d(g)}\prod_{i\in S_k(g)}\frac{1}{\delta_k(g)}\exp\{-\frac{1}{2\delta_k^2(g)}(\mu_i-\nu_k(g))^2\}\right]$$
$$\div\left[\sum_{g=1}^{G}\prod_{k=1}^{d(g)}\prod_{i\in S_k(g)}\frac{1}{\delta_k(g)}\exp\{-\frac{1}{2\delta_k^2(g)}(\mu_i-\nu_k(g))^2\}\right].$$

To estimate the posterior moments, we use Rao-Blackwellized estimates as in Gelfand and Smith (1991). Note that

$$E[\mu_i | \mathbf{y}, \nu_k(g), g] = (1-\lambda_i(g))\nu_k(g) + \lambda_i(g)\ \bar{y}_{i.}, \tag{3.8}$$

where $i \in S_k(g)$, $k=1,\cdots,d(g)$. Using (3.8), we have

$$E(\mu_i | \mathbf{y}, g) \approx \frac{1}{cq} \sum_{t=1}^{c} \sum_{l=q+1}^{2q} [(1-\lambda_i(g))\nu_k^{(t,l)}(g) + \lambda_i(g)\ \bar{y}_{i.}]. \tag{3.9}$$

Thus we can obtain

$$E(\mu_i | \mathbf{y}) = \sum_g p(g| \mathbf{y}) E(\mu_i | \mathbf{y}, g)\ . \tag{3.10}$$

Next to estimate the posterior variances $Var(\mu_i | \mathbf{y})$, at first we consider

$$Var(\mu_i | \mathbf{y}, \nu_k(g), g) = n_i / \sigma_i^2 + 1/\delta_k^2(g) \tag{3.11}$$

where $i \in S_k(g)$, $i=1,\cdots,d(g)$. Then we can obtain

$$\begin{aligned} &Var(\mu_i | \mathbf{y}, g) \\ &= E[Var(\mu_i | \mathbf{y}, \nu_k(g), g)] + Var[E(\mu_i | \mathbf{y}, \nu_k(g), g)] \\ &\approx (n_i/\sigma_i^2 + 1/\delta_k^2(g)) + \frac{1}{cq} \sum_{t=1}^{c} \sum_{l=q+1}^{2q} [(1-\lambda_i(g))\nu_k^{(t,l)}(g) + \lambda_i(g)\ \bar{y}_{i.} - E_i]^2 \end{aligned} \tag{3.12}$$

where $E_i = \sum_{t=1}^{c} \sum_{l=q+1}^{2q} \{(1-\lambda_i(g))\nu_k^{(t,l)}(g) + \lambda_i(g)\ \bar{y}_{i.}\}/cq$. Finally we can have

$$\begin{aligned} &Var(\mu_i | \mathbf{y}) \\ &= E[Var(\mu_i | \mathbf{y}, g)] + Var[E(\mu_i | \mathbf{y}, g)] \\ &= \sum_{g=1}^{G} Var(\mu_i | \mathbf{y}, g) p(g| \mathbf{y}) + \sum_{g=1}^{G} E(\mu_i | \mathbf{y}, g)^2 p(g| \mathbf{y}) - \left(\sum_{g=1}^{G} E(\mu_i | \mathbf{y}, g) p(g| \mathbf{y}) \right)^{2}. \end{aligned}$$

This can be approximated easily using (3.9) and (3.12).

4. Numerical Examples

4.1 Example 1

In this subsection, we will illustrate the results obtained in Section 2 and 3 with sample data. Suppose we have four sample means, $\bar{y}_1 = 2.0$, $\bar{y}_2 = 5.1$, $\bar{y}_3 = 5.2$ and $\bar{y}_4 = 15.0$. So we have $I=4$. We consider the following partitions; $g_1 = \{1;2;3;4\}$, $g_2 = \{1;2,3;4\}$,

$g_3=\{1;2,4;3\}$, $g_4=\{1;2;3,4\}$, $g_5=\{1;2,3,4\}$ and $g_6=\{1,2,3,4\}$. Clearly $d(g_1)=4$, $d(g_2)=3$, $d(g_3)=3$, $d(g_4)=3$, $d(g_5)=2$ and $d(g_6)=1$. For the prior we choose $\sigma_i^2=1.0$ and $(\delta_1(1),\delta_2(1),\delta_3(1),\delta_4(1))=(1.0,1.5,2.0,2.5)$, $(\delta_1(2),\delta_2(2),\delta_3(2))=(1.0,1.5,2.0)$, $(\delta_1(3),\delta_2(3),\delta_3(3))=(1.0,1.5,2.0)$, $(\delta_1(4),\delta_2(4),\delta_3(4))=(1.0,1.5,2.0)$, $(\delta_1(5),\delta_2(5))=(1.0,1.5)$, and $\delta_1(6)=1.0$. We have 10 Gibbs chains which use starting points drawn by the overdispersed distribution. We use 2000 replications. Actually the first 1000 are used for burn-in samples so that the last 1000 samples are used for our computations.

We have three separated groups. For our partially exchangeable normal priors, the posterior probabilities of g are presented in Table 4.1. Table 4.1 shows that $p(g=2|\boldsymbol{y})$ is the highest posterior probability. That is, the grouping $g_2=\{1;2,4;3\}$ is the best grouping and has the highest weight in calculating the posterior moments. Posterior means and posterior standard deviations are presented in Table 4.2. Table 4.2 shows that posterior means are closer to $\bar{y}_{i.}$ in the partially exchangeable case than in the exchangeable case. We observe that multiple shrinkages have been occurred in the calculation of the posterior means in the case of the partial exchangeability.

Table 4.1. Posterior probabilities for a selected collection of partitions g for 4 sample means

g	Partition	$p(g\|\boldsymbol{y})$
1	{1;2;3;4}	0.40791960
2	{1;2,3;4}	0.51283640
3	{1;2,4;3}	0.03133038
4	{1;2;3,4}	0.02368235
5	{1;2,3,4}	0.01336101
6	{1,2,3,4}	0.01087032

Table 4.2. Estimates of μ_i and standard errors (in parenthesis) for 4 sample means

i	Exchangeable model	Partially exchangeable model
1	4.4125	2.011535
	(0.728869)	(1.748060)
2	5.9625	5.187431
	(0.728869)	(1.500487)
3	6.0125	5.232480
	(0.728869)	(1.428161)
4	10.9125	14.817180
	(0.728869)	(1.513829)

4.2 Example 2

In this subsection we accomplish the same calculation process as in Subsection 4.1 with 3 sample means. We have $\bar{y}_1 = 1.1$, $\bar{y}_2 = 1.2$ and $\bar{y}_3 = 10.0$. We have $I=3$. The interesting partitions are $g_1 = \{1;2;3\}$, $g_2 = \{1,2;3\}$, $g_3 = \{1;2,3\}$ and $g_4 = \{1,2,3\}$. Clearly $d(g_1)=3$, $d(g_2)=2$, $d(g_3)=2$ and $d(g_4)=1$. For the prior we choose $\sigma_i^2 = 1.0$ and $(\delta_1(1), \delta_2(1), \delta_3(1)) = (1.0, 1.5, 2.0)$, $(\delta_1(2), \delta_2(2)) = (1.0, 1.5)$, $(\delta_1(3), \delta_2(3)) = (1.0, 1.5)$, and $\delta_1(4) = 1.0$. The estimation procedures using Gibbs sampling are the same as that of Subsection 4.1.

We have two separated groups. The posterior probabilities of partition g are presented in Table 4.3. As one might expect it, Table 4.3 indicates that $p(g=2 \mid \mathbf{y})$ has the highest posterior probability. That is, the grouping $g_2 = \{1,2;3\}$ is the best grouping and has the highest weight in calculating the posterior moments. Posterior means and the corresponding standard errors are presented in Table 4.4. Table 4.4 shows that the posterior means are much closer to $\bar{y}_{i\cdot}$ in the partially exchangeable case than in the exchangeable case. So we have the same observation as Subsection 4.1.

Table 4.3. Posterior probabilities for a selected collection of partitions g for 3 sample means

g	Partition	$p(g\mid y)$
1	{1;2;3}	0.35632670
2	{1,2;3}	0.42607390
3	{1;2,3}	0.09401340
4	{1,2,3}	0.12358590

Table 4.4. Estimates of μ_i and standard errors (in parenthesis) for 3 sample means

i	Exchangeable model	Partially exchangeable model
1	2.60 (0.541667)	1.272638 (1.6947900)
2	2.65 (0.541667)	1.556361 (1.6658700)
3	7.05 (0.541667)	9.199573 (2.1242790)

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] Draper, D. (1995). Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society*, Ser. B, 57, 45-97

[3] DuMouchel, W. (1990). Bayesian metaanalysis, in *Statistical Methodology in the Pharmaceutical Science*, 509-529.

[4] DuMouchel, W. and Harris, J. (1983). Bayesian methods for combining the results of cancer studies in human and other species. *Journal of the American Statistical Association*, 78, 293-307.

[5] Efron, B. (1996). Empirical Bayes methods for combining likelihoods. *Journal of the American Statistical Association*, 91, 538-565.

[6] 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.

[7] Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations, *Communication in Statistics - Theory & Methods*, 20, 1747-1766.

[8] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Sciences*, 7, 457-511.

[9] George, E. I. (1986). Combining minimax shrinkage estimators, *Journal of the American Statistical Association*, 81, 437-445.

[10] Hedges, J. V. and Olkin, I. (1985). *Statistical Methods for Meta-analysis,* Orlando: Academic Press.

[11] 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.