Journal of the Korean
Data & Information Science Society
2000, Vol. 11, No. 2, pp. 313 ~ 318

Robust Bayesian Models for Meta-Analysis

Dal Ho Kim [1] · **Gea Joo Park** [2]

Abstract

This article addresses aspects of combining information, with special attention to meta-analysis. In specific, we consider hierarchical Bayesian models for meta-analysis under priors which are scale mixtures of normal, and thus have tail heavier than that of the normal. Numerical methods of finding Bayes estimators under these heavy tailed prior are given, and are illustrated with an actual example.

Key Words and Phrases: Gibbs sampler, Heavy-tailed priors, Hierarchical Bayes, Random effects model.

1. Introduction

Meta-analysis is the science of combining data from related but statistically independent studies using statistical methods. Each of the n studies provides information about an effect parameter, which we denote θ_i here for the ith study, $i = 1, \cdots, n$ and which for convenience here will be one dimensional.

The motivation for conducting a meta-analysis is straightforward: to integrate results over studies and to summarize information about certain treatment effects. More recently it has been the subject of increasing attention in the medical research literature. The need for formal methods of combining such results is evident when one considers the vast scope of current scientific and biomedical research.

The hierachical model provides a natural way to do meta-analysis from the perspectives of modeling and of theoretical development. The hierarchical Bayesian approach is distinguished by the construction and use of a formal statistical model at two levels. First, a parametric model is set up for each of the individual studies, in which a likelihood function relates the distribution of the sample statistics to one

[1]Assistant Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

[2]Graduate, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea

or more unknown parameters characterizing that study. Next, a second parametric statistical model is constructed which relates the parameters from the separate studies to each other. Significant applications of hierarchical models already have been made to meta-analysis. Variance component models have been proposed and developed as random effects model; see e.g. DerSimonian and Laird(1986), Goodman(1989) and DuMouchel(1990).

In general, there are two methodological frameworks for synthesizing information from a group of studies: a fixed effects approach and a random effects approach. Many meta-analysis use a random effects model to account for heterogeneity among study results, beyond the variation associated with fixed effects. Random effects models focus on learning about the populations of true effects for several possible studies. A host of techniques for random effects models are well-known and are available for making inferences about the overall mean of data.

In this paper, we consider hierarchical Bayesian models in order to combine statistical summaries from several studies. We will focus attention on estimation of the average study effect.

Substantial evidence has been presented to the effect that priors with tails that are flatter than those of likelihood function tend to be fairly robust (e.g. Box and Tiao(1973), O'Hagan(1989) and West(1987)). Priors which are scale mixtures of normal have flatter tails than those of the normal automatically by construction. This class of priors includes the student t family, double exponential, logistic and the exponential power family of Box and Tiao(1973) among others. In practice, such distributions for the simulation and the analysis of outlier models have been used widely in robustness studies.

Meta-analysis via hierarchical Bayesian models are illustrated using aspirin data of Goodman(1989). This meta-analysis example shows how different specifications of the prior distribution can affect the results.

The price to be paid for utilization of such heavy-tailed priors is computational; closed form is no longer possible. Recently, however, Markov chain Monte Carlo integration scheme known as the Gibbs sampler has proved to be a simple yet powerful tool in hierarchical Bayesian models; see e.g. Carlin and Polson(1991) and Gelfand and Smith(1991).

2. Main Results

We consider the common random effects model $y_i = \theta_i + \epsilon_i, i = 1, \ldots, n$, where the θ_i are unknown mean structures ($\theta_i = \mu + \delta_i$; δ_i are unexplainable study-to-study variation), and ϵ_i are independent random errors having density with mean 0. Morris and Normand(1992) described such hierarchical Bayesian approaches for meta-analysis. These approaches are aimed principally at estimating the mean, μ, of the population of effects, with θ_i as a sample from that population.

First we assume that the prior distributions for θ_i are normal. We use standard distributions as the hyperpriors for (μ, τ^2). The hierarchical Bayesian model is as follows.

I. $y_i|\ \theta_i \sim N(\theta_i, V_i)$ $(i = 1, \cdots, n;\ V_i$ is known);

II. $\theta_i|\ \mu, \tau^2 \sim N(\mu, \tau^2)$ $(i = 1, \cdots, n)$,

III. Marginally μ and τ^2 are independent with

$$\mu \sim Uniform(-\infty, \infty) \quad \text{and} \quad \tau^2 \sim IG(a/2, b/2).$$

$IG(\cdot, \cdot)$ represent a inverse gamma distribution. (A random variable X is said to have a $IG(\alpha, \beta)$ distribution if it has a pdf of the form $f(x) \propto x^{-(\alpha+1)} \exp(-\beta x^{-1})$ $I_{(0,\infty)}(x)$, where I denotes the usual indicator function). Notice that y_i is the observed study effect, θ_i is the true study effect, V_i is the within-study variance, μ is the average study effect, and τ^2 is the between-study variance. Because $\sqrt{V_i}$ is usually the standard error of an estimate, for simplicity we treat it as fixed.

So far we have relied primarily upon the hierarchical model of normal prior distribution, for modeling data and parameters. The use of a limited class of distributions results, however, in a limited and potentially inappropriate class of inferences. Models based on the normal distribution are notoriously 'nonrobust' to outliers, in the sense that a single aberrant data point can strongly affect the inference for all the parameters in the model, even those with little substantive connection to the outlying observation.

Now, we consider a refinement based on heavy tailed priors on θ_i's using scale mixtures of normals. It was introduced by West(1985) in the context of Bayesian modeling for robustness.

I. $y_i|\ \theta_i \sim N(\theta_i, V_i)$ $(i = 1, \cdots, n; V_i$ is known);

II. $\theta_i|\ \mu, \tau^2 \sim p(\theta_i|\ \mu, \tau^2)$ $(i = 1, \cdots, n)$,

$$\Longleftrightarrow \quad \text{IIa)} \quad \theta_i|\ \mu, \tau^2 \sim N(\mu, \lambda_i \tau^2)$$

$$\text{IIb)} \quad \lambda_i \sim g(\lambda_i) \ \text{ where } \int_0^\infty g(x)dx = 1$$

where λ_i is the latent variable.

III. Marginally μ and τ^2 are independent with

$$\mu \sim Uniform(-\infty, \infty) \quad \text{and} \quad \tau^2 \sim IG(a/2, b/2).$$

Note that the following list identifies the necessary functional forms for $g(\lambda_i)$ to obtain the heavy tailed prior distribution of θ_i.

t-priors : if $k/\lambda_i \sim \chi^2_k$ then θ_i is Student t with k degrees of freedom, location parameter μ and scale parameter τ.

double exponential priors : if $\lambda_i \sim \exp(1/2)$ then θ_i is double exponential with location parameter μ and scale parameter τ.

exponential power family priors : if λ_i has positive stable distribution with index $\alpha/2$ then θ_i has exponential power distribution with location parameter μ and scale parameter τ.

logistic priors : if $1/\sqrt{\lambda_i}$ has the asymptotic Kolmogorov distance distribution then θ_i is logistic with location parameter μ and scale parameter τ. (A random variable Z is said to have an asymptotic Kolmogorov distance distribution if it has a pdf of the form $f(z) = 8z\sum_{j=1}^{\infty}(-1)^{j-1}j^2\exp(-2j^2z^2)I_{(0,\infty)}(z)$).

We shall use the notations $\boldsymbol{y} = (y_1, \cdots, y_n)$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)$. Then the joint pdf is given by

$$f(\boldsymbol{y}|\ \boldsymbol{\theta}, \mu, \boldsymbol{\lambda}, \tau^2) \propto \prod_{i=1}^{n}\Big[e^{-\frac{(y_i-\theta_i)^2}{2V_i}}(\lambda_i\tau^2)^{-\frac{1}{2}}e^{-\frac{(\theta_i-\mu)^2}{2\lambda_i\tau^2}}g(\lambda_i)\Big]\tau^{-2(\frac{a}{2}+1)}e^{-\frac{b}{2\tau^2}}.$$

To implement Gibbs sampler the full conditional distributions are derived as follows:

(i) $\theta_i|\ \boldsymbol{y}, \mu, \boldsymbol{\lambda}, \tau^2 \sim N\left(\frac{y_i\lambda_i\tau^2+V_i\mu}{\lambda_i\tau^2+V_i},\ \frac{V_i\lambda_i\tau^2}{\lambda_i\tau^2+V_i}\right) \qquad (i = 1, \cdots, n)$;

(ii) $\mu|\ \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \tau^2 \sim N\left(\frac{\sum_{i=1}^{n}\theta_i\lambda_i^{-1}\tau^{-2}}{\sum_{i=1}^{n}\lambda_i^{-1}\tau^{-2}},\ \frac{1}{\sum_{i=1}^{n}\lambda_i^{-1}\tau^{-2}}\right)$;

(iii) $f(\lambda_i|\ \boldsymbol{y}, \boldsymbol{\theta}, \mu, \tau^2) \propto \lambda_i^{-\frac{1}{2}}\exp\left[-\frac{(\theta_i-\mu)^2}{2\lambda_i\tau^2}\right]g(\lambda_i) \qquad (i = 1, \cdots, n)$;

(iv) $\tau^2|\ \boldsymbol{y}, \boldsymbol{\theta}, \mu, \lambda \sim IG\left(\frac{n+a}{2},\ \frac{1}{2}\left\{b + \sum_{i=1}^{n}\frac{(\theta_i-\mu)^2}{\lambda_i}\right\}\right)$.

There are two interesting special cases. The one is the t-priors and the other is the double exponential priors. For t-priors case, the above (iii) will reduce to as follows.

(iiia) $\lambda_i|\ \boldsymbol{y}, \boldsymbol{\theta}, \mu, \tau^2 \sim IG\left(\frac{k+1}{2},\ \frac{(\theta_i-\mu)^2+k\tau^2}{2\tau^2}\right) \qquad (i = 1, \cdots, n)$.

For the double exponential priors case, the above (iii) will be

(iiib) $\lambda_i|\ \boldsymbol{y}, \boldsymbol{\theta}, \mu, \tau^2 \sim GIG\left(\frac{1}{2},\ \sqrt{\frac{(\theta_i-\mu)^2}{\tau^2}},\ \frac{(\theta_i-\mu)^2}{\tau^2}\right) \qquad (i = 1, \cdots, n)$,

where GIG denotes the generalized inverse Gaussian distribution. Note that $GIG(\frac{1}{2}, \sqrt{(\theta_i-\mu)^2/\tau^2}, (\theta_i-\mu)^2/\tau^2)$ is the reciprocal of an inverse Gaussian $(\sqrt{(\theta_i-\mu)^2/\tau^2}, (\theta_i-\mu)^2/\tau^2)$.

3. Illustrations

Now we illustrate the robust Bayesian meta-analysis using real data set. Table 1 gives the difference in raw mortality y_i from all causes for aspirin versus placebo in post-myocardial infarction patients in six major randomized multicenter trials conducted in the U.S. and Europe. Also shown are effective sample sizes n_i (average of number taking aspirin and number taking placebo) and estimated standard errors $\sqrt{\hat{V}_i} = \sqrt{\hat{Var}(y_i)}$ for each study.

Table 1. Aspirin data (Goodman, 1989)

Study level	UK-1	CDPA	GAMS	UK-2	PARIS	AMIS
effective sample size(n_i)	619.50	764.50	313.00	841.00	608.00	2262.00
mortality diff in %(y_i)	2.77	2.50	1.84	2.56	2.31	-1.15
estimated s.e. in %($\sqrt{\hat{V}_i}$)	1.65	1.31	2.34	1.67	1.98	0.90

The data show results from six studies of the protective effect of aspirin among patients following a herat attack. The measure of effect in each study is difference in percent mortality between the asprin and placebo groups, which was positive but not significant at conventional levels for five of the six studies, while the sixth and largest study showed a reversal in group contrast.

It would be worth investigating how assessment of the effectiveness of aspirin is altered by the use of a long-tailed prior for the effects θ_i, instead of the normal prior postulated for the hierarchical model. Our interest is to estimate the overall effect μ from such studies.

In deriving the robust Bayes estimates of parameters based on heavy-tailed prior distributions using scale mixtures of normal, we have considered Gibbs sampler with 10 independent sequences, each with a sample of size 2500 with a burn in sample of another 2500.

Table 2 provides the Bayes estimates of μ and τ^2 for the normal, double exponential and t priors with degrees of freedom 1, 3, 5, and 10 together with τ^2. We have used $a = b = 0.0005$ to ensure some form of diffuse gamma prior for the inverse of the variance component. Notice that the Bayes estimates of the overall mean mortality difference μ are not much sensitive to priors for θ_i.

Table 2. Estimates of μ and τ^2 for the aspirin data

	t(1)	t(3)	t(5)	t(10)	Normal	DE
μ	1.76802	1.51987	1.44795	1.42001	1.38258	1.52011
τ^2	0.82139	1.68588	2.04264	2.36358	2.59783	2.01356

To monitor the convergence of the Gibbs sampler, we use CODA which may be used to perform convergence diagnostics and statistical and graphical output analysis of the simulated values. If the potential scale reduction factors ($\hat{R}$) are near 1, it is reasonable to assume that the desired convergence is achieved in the Gibbs sampling algorithm (see Gelman and Rubin(1992) for complete discussion).

Table 3 provides the $\hat{R}$ values corresponding to the estimands using Cauchy and double exponential priors based on $10 \times 2500 = 25000$ simulated values.

Table 3. Potential scale reduction factors for the aspirin data

Priors	μ	τ^2	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
DE	1.00	1.01	1.00	1.01	1.00	1.00	1.00	1.01
Cauchy	1.01	1.02	1.01	1.02	1.01	1.01	1.01	1.02

REFERENCES

1. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, John Wiley &Sons, New York.
2. Carlin, B. P. and Polson, N. G. (1991). Inference for nonconjugate Bayesian models using the Gibbs sampler, *Canadian Journal of Statistics*, 19, 399-405.
3. DerSimonian, R. and Laired, N. (1986). Meta-analysis in clinical trials, *Controlled Clinical Trials*, 7, 177-188.
4. DuMouchel, W. H. (1990). Bayesian meta-analysis, *Statistical Methodology in the Pharmaceutical Sciences*, Ed. D. A. Berry, Marcel Dekker, New York, 509-529.
5. Goodman, S. N. (1989). Meta-analysis and evidence, *Controlled Clinical Trials*, 10, 188-204.
6. Gelfand, A. E. and Smith, A. F. M.(1991). Gibbs sampling for marginal posterior expectations, *Communication in Statistics - Theory and Methods*, 20, 1747-1766.
7. Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.
8. Morris, C. N. and Normand, S. L. (1992). Hierarchical models for combining information and for meta-analyses, *Bayesian Statistics 4*, Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford University Press, 321-344.
9. O'Hagan, A. (1989). Modeling with heavy tails, *Bayesian Statistics 3*, Eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindely and A.F.M. Smith, Oxford University Press, 349-359.
10. West, M. (1985). Generalized linear models: scale parameters, outlier accommodation and prior distributions, *Bayesian Statistics 2*, Amsterdam: North-Holland.
11. West, M. (1987). On scale mixtures of normal distributions, *Biometrika*, 74, 646-648.