The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 215-223

A Study on the Posterior Density under the Bayes-empirical Bayes Models[1]

Joong K. Sohn[2], Heon Joo Kim[3]

Abstract

By using Tukey's generalized lambda distribution, approximate posterior density is derived under the Bayes-empirical Bayes model. The sensitivity of posterior distribution to the hyperprior distribution is examined by using Tukey's generalized lambda distribution which approximate many well-known distributions. Based upon Monte Carlo simulation studies it can be said that posterior distribution is sensitive to the variance of the prior distribution and to the symmetry of the hyperprior distribution. Also posterior distribution is approximately obtained by using the following methods : Lindley method, Laplace method and Gibbs sampler method.

1. Preliminaries

Under the Bayes approach, the prior distribution with known values of parameters s given in general for the parameters of interests. On the other hand, values f parameters of a prior distribution are estimated from the past data set under he empirical Bayes approach. However, Deely and Lindley(1981) claimed that the empirical Bayes method is usually non-Bayesian and suggested the Bayes-empirical Bayes approach combining Bayes and emprical Bayes methods.

There have been many literatures on the sensitivity of posterior distribution to the hyperprior distribution (for example, Berger(1990), Sivaganesan(1993)). On the other hand, for the Bayes-empirical Bayes approach there has been no study on the sensitivity of posterior distribution to the hyperprior distribution. Hence in this paper we study the effect of the hyperprior distribution to posterior distribution. To obtain posterior distributions, Lindley method, Laplace method and Gibbs sampling method are going to be used. Also as a hyperprior distribution, Tukey's generalized lambda distribution is given for Lindley method and Laplace method. But for Gibbs sampling method, Tukey's generalized lambda distribution is not appropriate due to the difficulty of obtaining full conditional distributions.

---

1) This research was supported by the Korean Science and Engineering Foundation 941-0100-034-1.
2) Professor, Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea.
3) Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea.

2. Approximation of the Posterior Distributions

Let $(x_1, \theta_1), \cdots, (x_n, \theta_n)$ be the past data set and (x_{n+1}, θ_{n+1}) be the present data set. Then the Bayes-empirical Bayes model is given as follows.

Step 1. Hyperparameter λ follows the density $h(\lambda)$.

Step 2. Given λ, parameters θ_i are independent and identically distributed with the density function $\pi(\theta_i|\lambda)$. Here $\pi(\theta_i|\lambda)$ is the usual prior distribution for θ_i.

Step 3. Given λ and θ_i's, x_i's are independent and have densities $f(x_i|\theta_i)$. Here the density $f(x_i|\theta_i)$ is independent from λ and different θ_j.

Step 4. Compute

$$\pi(\theta_{n+1}|x_1, \cdots, x_{n+1}) = \int \pi(\theta_{n+1}|x_1, \cdots, x_{n+1}, \lambda) h(\lambda|x_1, \cdots, x_{n+1}) d\lambda$$
$$= \frac{\int \pi(\theta_{n+1}|x_1, \cdots, x_{n+1}, \lambda) \prod_{i=1}^{n+1} \pi(x_i|\lambda) h(\lambda) d\lambda}{\int \prod_{i=1}^{n+1} \pi(x_i|\lambda) h(\lambda) d\lambda}$$

To compute the posterior distribution of θ_{n+1} based on the past data and the present data set, let $x^{(n+1)} = (x_1, \cdots, x_n, x_{n+1})$, where x_i is independent and identically distributed as a normal distribution with mean θ_i and variance one. As a prior distribution $\pi(\theta_i|\lambda)$ for θ_i a conjugate prior distribution with two-dimensional hyperparameter is assumed. Then one can easily see that a conjugate prior $\pi(\theta_i|\lambda)$ is a normal distribution with mean λ_1/λ_2 and variance $1/\lambda_2$.

Let a hyperpior distribution for $\Lambda \equiv (\lambda_1, \lambda_2)$ be $h(\lambda_1, \lambda_2)$. Then posterior distribution is a ratio of integrals of the form as follows :

$$\pi(\theta_{n+1}|x^{(n+1)}) = \frac{\int p(\Lambda) e^{L(\Lambda)} h(\Lambda) d\Lambda}{\int e^{L(\Lambda)} h(\Lambda) d\Lambda} \tag{2.1}$$

and

$$p(\Lambda) = \frac{\sqrt{\lambda_2+1}}{\sqrt{2\pi}} \exp\left(-\frac{\lambda_2+1}{2}\left(\theta_{n+1} - \frac{x_{n+1}+\lambda_1}{\lambda_2+1}\right)^2\right),$$

and L, a log-likelihood function based on $n+1$ data, is given by

$$L(\Lambda) = \log\left\{\prod_{i=1}^{n+1} \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_2+1}} \exp\left[-\frac{1}{2}\left(x_i^2 + \frac{\lambda_1^2}{\lambda_2} - \frac{(x_i+\lambda_1)^2}{\lambda_2+1}\right)\right]\right\} \tag{2.2}$$

Under the Bayes-empirical Bayes approach parameters λ_1 and λ_2 of hyperprior distribution are estimated from the past data. In practice it is frequently very hard to perform the integration in (2.1). This leads us to approximate such integrations by following methods.

Under the regularity condition Deely and Lindley(1981) showed that

$$\int h(\lambda)e^{L(\lambda)}d\lambda \approx e^{\hat{L}}\left[\frac{-2\pi}{\hat{L}_2}\right]\hat{h}\left\{1-\frac{\hat{h}_2}{2\hat{h}\,\hat{L}_2}+\frac{\hat{h}_1\,\hat{L}_3}{2\hat{h}\,\hat{L}_2^{2}}+R\right\}. \tag{2.3}$$

Here $\hat{h}\equiv h(\hat{\lambda})$, $\hat{h_i}\equiv h_i(\hat{\lambda})$, $\hat{L}\equiv L(\hat{\lambda})$, $\hat{L_i}\equiv L_i(\hat{\lambda})$, where h_i and L_i are the i^{th} derivatives and $\hat{\lambda}$ is the value which maximizes the log-likelihood function L of the equation(2.2).

With Laplace method, which Tierney and Kadane(1986) used to approximate moments and marginal density function, one can see that

$$\int e^{nM(\lambda)}d\lambda \approx \int \exp[nM(\hat{\lambda})-n(\lambda-\hat{\lambda})^2/2\sigma^2]d\lambda=\sqrt{2\pi}\sigma n^{-1/2}e^{nM(\hat{\lambda})}, \tag{2.4}$$

where $\hat{\lambda}$ maximizes M and $\sigma^2=-1/M''(\hat{\lambda})$.

To use Gibbs sampler introduced by Geman and Geman(1984) the full conditional distributions must be evaluated and is given by follows.

$$p(\theta_i|x^{(n+1)},\theta_{-i},\Lambda) = \frac{\prod_{j=1}^{n+1} f(x_j|\theta_j)\pi(\theta_j|\Lambda)h(\Lambda)}{\int \prod_{j=1}^{n+1} f(x_j|\theta_j)\pi(\theta_j|\Lambda)h(\Lambda)d\theta_i}$$

$$= \frac{f(x_i|\theta_i)\pi(\theta_i|\Lambda)}{\int f(x_i|\theta_i)\pi(\theta_i|\Lambda)d\theta_i}, \quad i=1,\cdots,n+1$$

and

$$p(\Lambda|x^{(n+1)},\theta^{(n+1)}) = \frac{\prod_{j=1}^{n+1} f(x_j|\theta_j)\pi(\theta_j|\Lambda)h(\Lambda)}{\int \prod_{j=1}^{n+1} f(x_j|\theta_j)\pi(\theta_j|\Lambda)h(\Lambda)d\Lambda}$$

$$= \frac{\prod_{j=1}^{n+1} \pi(\theta_j|\Lambda)h(\Lambda)}{\prod_{j=1}^{n+1}\int \pi(\theta_j|\Lambda)h(\Lambda)d\Lambda}$$

where $\theta_{-i}=(\theta_1,\cdots,\theta_{i-1},\theta_{i+1},\cdots,\theta_{n+1})$ and $\theta^{(n+1)}=(\theta_1,\cdots,\theta_{n+1})$. Also posterior distribution of θ_{n+1} is approximated as follows

$$\pi(\theta_{n+1}|x^{(n+1)}) \approx \frac{1}{m}\sum_{j=1}^{m}\pi(\theta_{n+1}|x^{(n+1)}, \theta_{-(n+1)j}^{(t)}, \Lambda_j^{(t)}), \tag{2.5}$$

where m is the number of path.

Now one can find that the value of $\hat{\Lambda}$ which maximizes the log-likelihood function L of the equation(2.2) is $\hat{\lambda_1}=\hat{\lambda_2}\,\bar{x}_{n+1}$, where $\bar{x}_{n+1}$ is the usual sample mean. Thus it is only necessary to decide the value of either λ_1 or λ_2. But still it is not easy to derive posterior distribution as long as λ_1 and λ_2 are unknown. Thus we assume that $\lambda_2=K$, $\lambda_1=\lambda$, where K is known and $h(\lambda)$ follows a Tukey's generalized lambda distribution with known parameters μ, β, γ_1, γ_2. Tukey's generalized lambda distribution, suggested by Tukey(1960) and generalized by Ramberg et al.(1979) is as follows : Let H and h be the cumulative density function and probability density function of λ, respectively. Also let H^{-1} be its inverse. Then

$$\lambda \equiv H^{-1}(p)=\mu+[p^{\gamma_1}-(1-p)^{\gamma_2}]/\beta, \qquad 0<p<1,$$

where μ is a location parameter, β is a scale parameter and γ_1 and γ_2 are shape parameters. The distribution H is symmetric when $\gamma_1=\gamma_2$ and $(\mu, \beta, \gamma_1, \gamma_2)$ is given from the first central moment to fourth central moment of the its distribution.

Then one can get the following relations.

$$L(\lambda) = \log\left\{\prod_{i=1}^{n+1}\frac{\sqrt{K}}{\sqrt{K+1}}\exp\left[-\frac{1}{2}\left(x_i^2+\frac{\lambda^2}{K}-\frac{(x_i+\lambda)^2}{K+1}\right)\right]\right\}$$

$$p(\lambda) = \frac{\sqrt{K+1}}{\sqrt{2\pi}}\exp\left(-\frac{K+1}{2}\left(\theta_{n+1}-\frac{x_{n+1}+\lambda}{K+1}\right)^2\right),$$

$$h(\lambda) = \beta[\gamma_1 p^{\gamma_1-1}+\gamma_2(1-p)^{\gamma_2-1}]^{-1}.$$

Thus posterior distribution by using Lindley methods is

$$\begin{aligned}\pi(\theta_{n+1}|x^{(n+1)}) \approx\ & \frac{\sqrt{K+1}}{\sqrt{2\pi}}\exp\left\{-\frac{K+1}{2}\left(\theta_{n+1}-\frac{x_{n+1}+\lambda}{K+1}\right)^2\right\} \\ & \times\left[1+\frac{K(K+1)}{2(n+1)}\left\{\theta_{n+1}^2-2\theta_{n+1}\left(\frac{x_{n+1}+K\bar{x}_{n+1}}{K+1}-H_1\right)\right.\right. \\ & \left.\left.+\left(\frac{x_{n+1}+K\bar{x}_{n+1}}{K+1}\right)^2-2\frac{x_{n+1}+K\bar{x}_{n+1}}{K+1}H_1-\frac{1}{K+1}\right\}\right],\end{aligned} \tag{2.6}$$

where

$$H_1 = \frac{\beta[\gamma_2(\gamma_2-1)(1-p)^{\gamma_2-2}-\gamma_1(\gamma_1-1)p^{\gamma_1-2}]}{(\gamma_1 p^{\gamma_1-1}+\gamma_2(1-p)^{\gamma_2-1})^2}.$$

Also posterior distribution obtained by using Laplace method is

$$\pi(\theta_{n+1}|x^{(n+1)}) \approx (\sigma^*/\sigma)\exp(n+1)(M^*(\hat{\lambda}^*)-M(\hat{\lambda})), \qquad (2.7)$$

where

$$M(\lambda) = \frac{1}{n+1}(L(\lambda)+\log h(\lambda))$$
$$M^*(\lambda) = \frac{1}{n+1}(L(\lambda)+\log h(\lambda)+\log p(\lambda))$$

and $\hat{\lambda}$ and $\hat{\lambda}^*$ maximize $M(\lambda)$ and $M^*(\lambda)$, respectively.

By using Gibbs sampler we can obtain the posterior distribution of θ_{n+1} as follows :

$$\pi(\theta_{n+1}|x^{(n+1)}) \approx \frac{1}{m}\sum_{j=1}^{m} p(\theta_{n+1}|x^{(n+1)}, \theta^{(t)}_{-(n+1)j}, \lambda_j^{(t)}) \qquad (2.8)$$

where $\theta^{(t)}_{-(n+1)j}$ and $\lambda_j^{(t)}$, $j=1,2,\cdots,m$, are m random variates of $\theta_{-(n+1)}$ and λ after t iterations and $p(\theta_{n+1}|x^{(n+1)}, \theta_{-(n+1)}, \lambda)$ is normal distributed with mean $(x_{n+1}\lambda)/(K+1)$ and variance $1/(K+1)$. In particular $p(\lambda|x^{(n+1)}, \theta^{(n+1)})$ must be evaluated but is different types for another $h(\lambda)$. If $h(\lambda)$ is $N(\delta,\tau)$ then the full conditional distribution of λ is also normal with mean $\dfrac{K\tau\left(\sum_{i=1}^{n+1}\theta_i+\delta/\tau\right)}{K+\tau(n+1)}$ and variance $\dfrac{K\tau}{K+\tau(n+1)}$.

In this paper we will use only the case that hyperpior distribution is normal because the full conditional distribution is not closed form for another hyperprior distributions which may be another topics of Gibbs sampler.

3. Monte Carlo Simulation Studies

In this section we examine the sensitivity of posterior distribution to a hyperprior distribution via Monte Carlo simulation and also compare between Lindley method, Laplace method and Gibbs sampling method. To see the sensitivity of a posterior distribution we used the following scheme.

Simulation data were obtained from the subroutine GGNML of IMSL(International Mathematical and Statistical Libraries) with sample size $n=5, 10, 15, 30, 50, 100$ and known constant $K=0.1, 0.5, 1.0, 5.0, 10.0$.

Lindley method and Laplace method were used when a hyperprior distribution is Tukey's generalized lambda distribution with various values of parameters which approximate a uniform(0,1), a standard normal, a double exponential, a Cauchy and an exponential distribution with mean 1. The values of parameters of Tukey's generalized lambda distribution were based on the result of Ramberg et al(1979).

To compare between Lindley, Laplace and Gibbs sampling methods in terms of easiness and precision, normal distributions with combinations of means, 0, 0.5, 1, 5 and variances, 1, 4,9,25,100 are given as hyperprior distributions.

All computations were carried out and the rest results of not mentioned in this paper are available based on request. In Figure 1 and Figure 2 posterior distributions are estimated by the Simpson's composite rule. From figures one can see that posterior distribution of θ_{n+1} varies with variance of a hyperprior distribution more than sample size. For the case $n=5$ and $K=10.0$ posterior distribution of θ_{n+1} is plotted with various hyper distributions in Figure 3. From this figure one can see that posterior distribution is affected by the symmetricity of a hyperprior distribution. For the case that a hyperprior distribution is symmetric, uniform, standard normal, Laplace, double exponential and Cauchy distributions, posterior distributions are similar. But the case that the hyperpior distribution is asymmetric, that is, exponential distribution posterior distribution is different from another distributions. This fact is analogous to various combination of sample size n and the variance K. That is, posterior distribution is affected by symmetricity of a hyperprior distribution and variance of hyperparameter. In Figure 4, for the case that $n=15$, $K=0.1$ and a hyperprior distribution is a standard normal distribution, posterior distributions are estimated by various approximation methods containing Simpson's composite rule, as a yardstick for comparison. From figure one can see that Gibbs sampler method, the distributions by Laplace approximation method and Simpson's approximation provide nearly same shape to the true distribution. But that by Deely and Lindley's approximation method is different from true distribution and another distribution, because the error rate of Laplace approximation is $O(n^{-2})$ and that of Deely and Lindley's approximation is $O(n^{-1})$. In particular, Gibbs sampler method is almost not affected by sample size and variances of a hyper distribution.

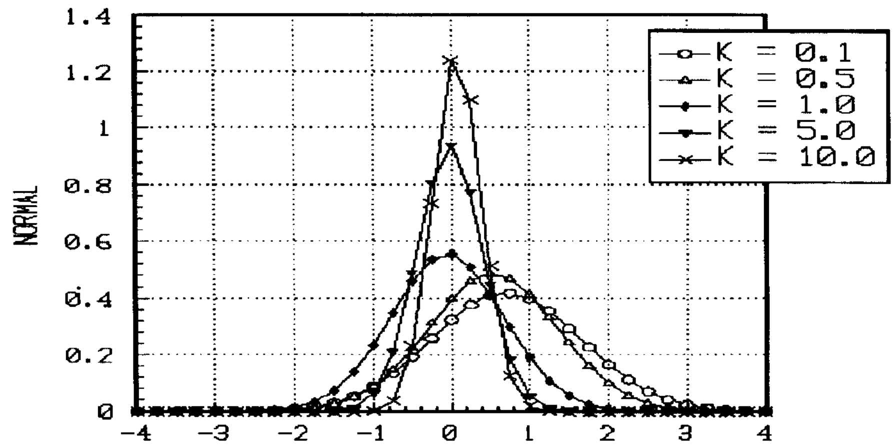

Fig 1. Posterior Distributions with a Standard Normal Prior with $n=30$

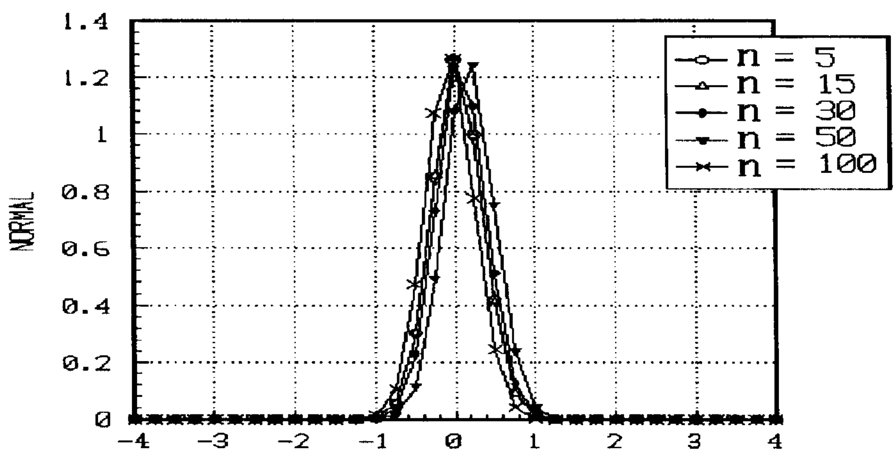

Fig 2. Posterior Distributions with a Standard Normal Prior with $K=10.0$

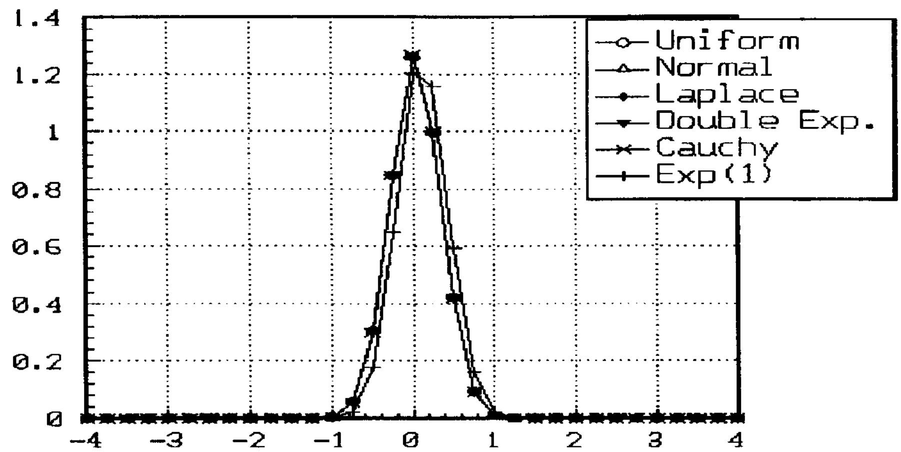

Fig 3. Posterior Distributions with $n=5$ **and** $K=10.0$

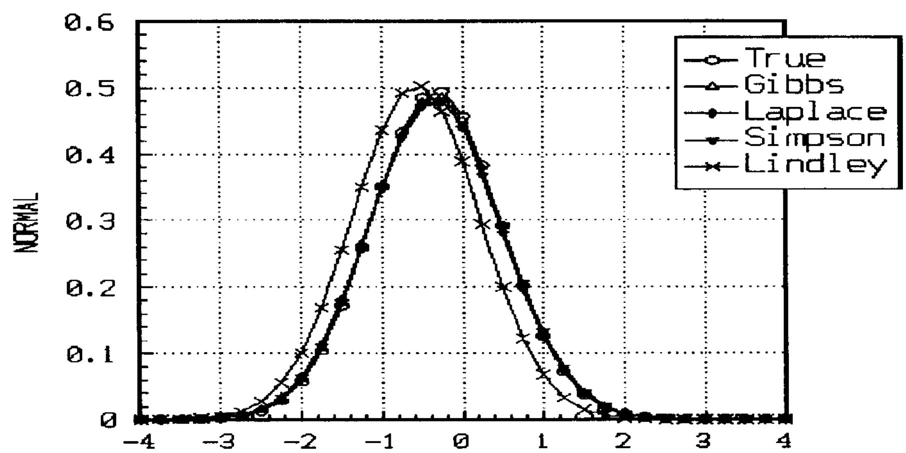

Fig 4. Approximated Posterior Distributions with $n=15$ **and** $K=0.5$

References

[1] Berger, J.O. (1990). Robust Bayesian Analysis: sensitivity to the prior, *Journal of Statistical Planning and Inference*, Vol. 25, 303-328.

[2] Besag (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society* B, Vol. 36, 192-326.

[3] Deely, J.J. and Lindley, D.V. (1981). Bayes empirical Bayes, *Journal of the American Statistical Association*, Vol. 76, 833-841.

[4] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, Vol. 85, No. 410, 398-409.

[5] Gelfand, A. E: and Smith, A. F. M. (1991). Gibbs Sampling for Marginal Posterior Expectations, *Communication in Statistics - Theory and Methods*, 20, 1747-1766.

[6] Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling, *Journal of the American Statistical Association*, Vol. 85, No. 412, 972-985.

[7] Gelfand, A. E., Smith, A. F. M. and Lee, T. (1992). Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, *Journal of the American Statistical Association*, Vol. 87, No. 418, 523-532.

[8] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol, Pami-6, No. 6, 721-741.

[9] Lindley, D.V. (1980). *Approximate Bayesian Methods*, Bayesian Statistics (eds.Bernado, J.M., DeGroot, M.H., Lindley, D.V. and Smith, A.F.M.), Valencia, Spain: University Press, 223-245.

[10] Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R. and Mykytka, E.F. (1979). A Probability Distribution and Its Uses in Fitting Data, *Technometircs*, 21, 201-214.

[11] Sivaganesan, S. (1993). Robust Bayesian Diagnostics, *Journal of Statistical Planning and Inference*, 35, 171-188.

[12] Tierney L. and Kadane, J.B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistical Association*, 81, 82-86.

[13] Tukey, J.W. (1960). The Practical Relationship between the Common Transformation of Percentages of Counts and of Amounts, Technical Report 36, Statistical Techniques Research Group, Princeton University, U.S.A..