Bayesian Estimation of the Normal Means under Model Perturbation

Dal-Ho Kim[1] · Seung-Cheol Han[2]

1. Introduction

Bayesian methods have become more attractive because of their practical approach in the theory and practice of statistics different from classical methods. Bayesian analysis is also more powerful since the analysis is based on a given data and prior information is used by experimenter, where it makes the efficiency of the result strengthen. In real data analysis, most social problems or experiments are para- meterized with several parameters rather than one parameter. This method is useful in many fields which one problem is related or dependent with others. However, multiple parametric models have had much difficulties in finding their estimators and inferring from them. Many statisticians made efforts to surmount these difficulties. Empirical and hierarchical Bayes methods are useful in statistics, especially in the context of simultaneous estimation of several parameters.

For example, agencies of the Federal Government have been involved in obtaining estimates of per capita income, unemployment rates, crop yields and so forth simultaneously for several state and local government areas. In such situations, quite often estimates of certain area means, or simultaneous estimates of several area means can be improved by incorporating information from similar neighboring areas. Examples of this type are especially suitable for empirical Bayes (EB) analysis. As described in Berger (1985), an EB scenario is one in which is known as relationships among the coordinates of a parameter vector, say $\Theta = (\Theta_1, \cdots, \Theta_n)^T$ allow use of the data to estimate some features of the prior distribution. For example, one may have reason to believe that the Θ_i's are iid from a prior $\pi_0(\lambda)$, where π_0 is structurally known except possibly for some unknown parameter λ. A parametric empirical Bayes (EB) procedure is one where λ is estimated from the marginal distribution of only the observations.

Closely related to the EB procedure is the hierarchical Bayes (HB) procedure which models the prior distribution in several stages. In the first stage, conditional on $\Lambda = \lambda$, Θ_i's are iid with a prior $\pi_0(\lambda)$. In the second stage, a prior distribution (often improper) is assigned to Λ. This is an example of a two stage

1) First Author : Associate Professor, Department of Statistics, Kyungpook National University
2) Graduate, Department of Statistics, Kyungpook National University

prior. The idea can be generalized to multistage priors, but that will not be pursued in this article.

It is apparent that both the EB and the HB procedures recognize the uncertainty in the prior information, but whereas the HB procedure models the uncertainty in the prior information by assigning a distribution (often noninformative or improper) to the prior parameters (usually called hyperparameters), the EB procedure attempts to estimate the unknown hyperparameters, typically by some classical methods like the method of moments(MME), method of maximum likelihood(MLE), etc., and use the resulting estimated priors for inferential purposes. It turns out that the two methods can quite often lead to comparable results, especially in the context of point estimation. However, when it comes to the question of measuring the standard errors associated with these two estimators, the HB method has a clear edge over a naive EB method. Whereas there are no clear cut measures of standard errors associated with EB point estimators, the same is not true with HB estimators. To be precise, if one estimates the parameter of interest by its posterior mean, then a very natural estimate of the risk associated with this estimator is its posterior variance. Estimates of the standard errors associated with EB point estimators usually need an ingenious approximations (see, e.g., Morris, 1981, 1983). However, the posterior variances, though often complicated, can be found exactly.

Deely and Lindley (1981) compared and contrasted the EB and the HB procedures much in the spirit of the discussion in the preceding paragraphs. However, unlike the present article, they did not emphasize simultaneous estimation problems, nor did they incorporate discussion of multivariate normal models.

The outline of the remaining sections is as follows. In Section 2 of this paper, we summarize the methods for finding the empirical and hierarchical Bayes estimators. Furthermore, we obtain the variance of HB estimator. As noted earlier, for the EB estimator, we have only the point estimator itself since it is not easy to find its standard error. In Section 3, we set up the model structure using the several different distribution of the errors for observing their effects of model perturbation for the error terms in obtaining the EB and HB estimators. In Section 4, we provide a numerical example. Based on a simulation study, we observe the performance of EB and HB estimators under model perturbation.

2. Bayesian Estimation of the Normal Means

We make a comparison of the EB and HB procedures for estimating the multivariate normal mean. We consider the following model.

Ⅰ. Conditional on $\Theta_1, \cdots, \Theta_m$, let $X_1, \cdots, X_m$ be independent with

$X_i \sim N(\Theta_i, \sigma^2)$, $i = 1, \cdots, m$, $\sigma^2 (> 0)$ being known.

II. The Θ_i's have independent $N(\mu, A)$, $i = 1, \cdots, m$, priors.

Let write $\Theta = (\Theta_1, \cdots, \Theta_m)^T$, $\boldsymbol{X} = (X_1, \cdots, X_m)^T$ and $\boldsymbol{x} = (x_1, \cdots, \cdot x_m)^T$.

The posterior distribution of Θ given $\boldsymbol{X} = \boldsymbol{x}$ is then $N\big((1-B)\boldsymbol{x} + B\mu, (1-B)\boldsymbol{I}_m\big)$, where $B = \sigma^2/(\sigma^2 + A)$. Accordingly, the Bayes estimator of Θ is given by

$$\widehat{\Theta}^B = E(\Theta \mid \boldsymbol{X} = \boldsymbol{x}) = (1-B)\boldsymbol{x} + B\mu \tag{2.1}$$

In an EB or a HB scenario, some or all of the prior parameters are unknown. In an EB setup, these parameters are estimated from the marginal distribution of $\boldsymbol{X}$ which in this case is $N(\mu, B^{-1}\boldsymbol{I}_m)$. A HB procedure, on the other hand, models the uncertainty of the unknown prior parameters by assigning distributions to them. Such distributions are often called hyperpriors. We shall consider the following case.

We assume both μ (real) and A to be unknown. Recall that marginally $\boldsymbol{X} \sim N(\mu 1_m, B^{-1}\boldsymbol{I}_m)$, where $B = \sigma^2/(\sigma^2 + A)$. Hence, $\left(\overline{X}, \sum_{i=1}^m (X_i - \overline{X})^2\right)$ is complete sufficient, so that the UMVUE's of μ and B are given respectively by $\overline{X}$ and $\sigma^2(m-3) \Big/ \sum_{i=1}^m (X_i - \overline{X})^2$. Substituting these estimators of μ and B in (2.1), the EB estimator of Θ is given by

$$\widehat{\Theta}^{EB} = \boldsymbol{X} - \frac{\sigma^2(m-3)}{\sum_{i=1}^m (X_i - \overline{X})^2}(\boldsymbol{X} - \overline{X}1_m) \tag{2.2}$$

This modification of the James-Stein estimator was proposed by Lindley(1962). Whereas, the original James-Stein estimator shrinks $\boldsymbol{X}$ towards a specified point, the modified estimator given (2.2) shrinks $\boldsymbol{X}$ towards a hyperplane spanned by 1_m. Additionally, the estimator $\widehat{\Theta}^{EB}$ is known to dominate $\boldsymbol{X}$ for $m \geq 4$.

We now proceed to find the HB estimator of Θ. Consider the model where (i) conditional on Θ, μ and A, $\boldsymbol{X} \sim N(\Theta, \sigma^2\boldsymbol{I}_m)$; (ii) conditional on μ and A,

$\Theta_i \sim N(\mu, A)$, $i = 1, \cdots, m$; (iii) marginally μ and A are independently distributed with μ uniform on $(-\infty, \infty)$, and A has uniform improper pdf on $(0, \infty)$. Customarily, such a prior on μ is widely accepted as a reasonable objective prior. Then the joint (improper) pdf of X, Θ, μ and A is given by

$$f(x, \Theta, \mu, A) \propto \exp\left[-\frac{1}{2\sigma^2}\|x - \Theta\|^2\right] \times A^{-\frac{1}{2}m} \exp\left[-\frac{1}{2A}\|\Theta - \mu 1_m\|^2\right] \tag{2.3}$$

Now integrating with respect to μ, it follows from (2.3) that the joint (improper) pdf of X, Θ, and A is

$$f(x, \Theta, A) \propto A^{-\frac{m-1}{2}} \times \exp\left[-\frac{1}{2\sigma^2}\left(\Theta - \frac{1}{\sigma^2}E^{-1}x\right)^T\left(\Theta - \frac{1}{\sigma^2}E^{-1}x\right) + \frac{1}{\sigma^2}x^T x - \frac{1}{\sigma^4}x^T E^{-1}x\right]$$

where $E^{-1} = \sigma^2(1-B)I_m + \sigma^2 B m^{-1} J_m$. Hence, conditional on x and A,

$$\Theta \propto N\left[(1-B)x + B\bar{x}1_m,\ \sigma^2\left\{(1-B)I_m + \frac{B}{m}J_m\right\}\right] \tag{2.4}$$

Also, integrating with respect to Θ in (2.4), one gets the joint pdf of x and A given by

$$f(x, A) \propto (\sigma^2 + A)^{-\frac{m-1}{2}} \exp\left[-\frac{1}{2(\sigma^2 + A)}\sum_{i=1}^m (x_i - \bar{x})^2\right] \tag{2.5}$$

Since $B = \dfrac{\sigma^2}{\sigma^2 + A}$, it follows from (2.5) that the joint pdf of X and B is given by

$$f(x, B) \propto B^{\frac{m-5}{2}} \exp\left[-\frac{B}{2\sigma^2}\sum_{i=1}^m (x_i - \bar{x})^2\right] \tag{2.6}$$

where this HB approach was first proposed by Strawderman (1971).

It follows from (2.6) that

$$
\begin{aligned}
E(B|\boldsymbol{x}) = & \int_0^1 B^{\frac{m-3}{2}} \exp\left[-\frac{B}{2\sigma^2}\sum_{i=1}^m (x_i - \bar{x})^2\right] dB \\
& \div \int_0^1 B^{\frac{m-5}{2}} \exp\left[-\frac{B}{2\sigma^2}\sum_{i=1}^m (x_i - \bar{x})^2\right] dB
\end{aligned} \tag{2.7}
$$

$$
\begin{aligned}
E(B^2|\boldsymbol{x}) = & \int_0^1 B^{\frac{m-1}{2}} \exp\left[-\frac{B}{2\sigma^2}\sum_{i=1}^m (x_i - \bar{x})^2\right] dB \\
& \div \int_0^1 B^{\frac{m-5}{2}} \exp\left[-\frac{B}{2\sigma^2}\sum_{i=1}^m (x_i - \bar{x})^2\right] dB
\end{aligned} \tag{2.8}
$$

One can obtain $V(B|\boldsymbol{x})$ from (2.7) and (2.8), and use to obtain the HB estimator $E(\Theta|\boldsymbol{x})$ and its variance $V(\Theta|\boldsymbol{x})$

$$\hat{\Theta}^{HB} = E(\Theta|\boldsymbol{x}) = \boldsymbol{x} - E(B|\boldsymbol{x})(\boldsymbol{x} - \bar{x}\mathbf{1}_m) \tag{2.9}$$

$$
\begin{aligned}
V(\Theta|\boldsymbol{x}) = & V(B|\boldsymbol{x})(\boldsymbol{x} - \bar{x}\mathbf{1}_m)(\boldsymbol{x} - \bar{x}\mathbf{1}_m)^T \\
& + \sigma^2 \boldsymbol{I}_m - \sigma^2 E(B|\boldsymbol{x})\left(\boldsymbol{I}_m - \frac{1}{n}\boldsymbol{J}_m\right)
\end{aligned} \tag{2.10}
$$

3. Model Perturbation

Our interest is to find the EB and HB procedures for estimating the multivariate normal mean. We consider the following hierarchical model:

Ⅰ. $Y_i|\Theta_i \overset{iid}{\sim} N(\Theta_i, \sigma^2)$, $i = 1, \cdots, m$

Ⅱ. $\Theta_i|A \overset{iid}{\sim} N(0, A)$, $i = 1, \cdots, m$

In other words, our model can be rewritten as

$$Y_i = \Theta_i + e_i, \quad i = 1, \cdots, m \tag{3.1}$$

where $e_i \overset{iid}{\sim} N(0, \sigma^2)$.

Here, our interest is moved toward the issue that the normality of error terms is not satisfied and they are distributed from any other densities. The followings are provided as four cases:

Ⅰ. Skewed distribution $e_i = X_i - \beta$,
where $X_i \sim \text{Exp}(\Theta = 1/\beta)$

Ⅱ. Uniform distribution $e_i \sim U(-c, c)$

Ⅲ. Heavy tailed distribution $e_i \sim t_\nu$ (low value of ν)

Ⅳ. Bimodal (Mixture of normals)

$$e_i \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{with probability } \pi_1 \\ N(\mu_2, \sigma_2^2) & \text{with probability } \pi_2 = 1 - \pi_1 \end{cases}$$

with $\pi_1 \mu_1 + \pi_2 \mu_2 = 0$.

Additionally, without loss of generality, we may assume that

$$\pi_1 = \pi_2 = \frac{1}{2}, \ \mu_1 = -\mu_2, \text{ and } \sigma_1^2 = \sigma_2^2. \tag{3.2}$$

Note that it is necessary that density of e_i's for each case have mean 0 and equal variance. Especially, the variance of the mixture distribution of normals can be obtained as follows.

A random variable x has a normal mixture distribution if the data originate from a fixed number k of normal densities. A k-component normal mixture has pdf

$$f_k(x) = \sum_{j=1}^{k} \pi_j \phi(x; \mu_j, \sigma_j) \tag{3.3}$$

where $\phi(x; \mu_j, \sigma_j)$ is a normal pdf with mean μ_j and standard deviation σ_j and π_j are weights satisfying

$$\sum_{j=1}^{k} w_j = 1, \qquad w_j \geq 0.$$

Given the mixture parameters $\{\pi_j, \mu_j, \sigma_j\}$, $j = 1, \cdots, k$, the mean μ_{eq} and variance σ_{eq}^2 of the distribution are

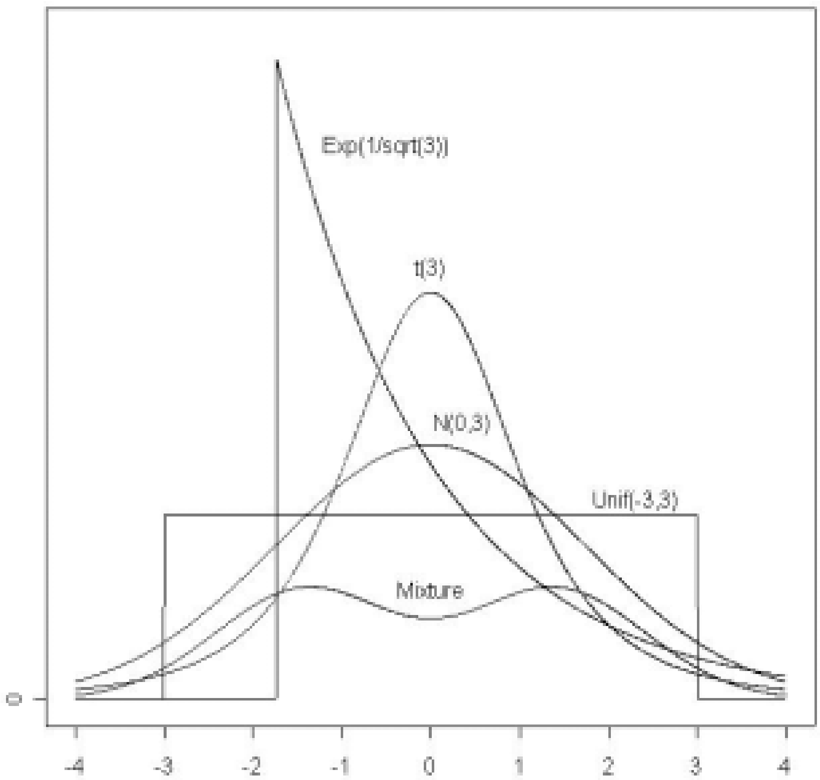
$$\mu_{eq} = \sum_{j=1}^{k} \pi_j \mu_j \tag{3.4}$$

$$\sigma_{eq}^2 = \sum_{j=1}^{k} \pi_j (\sigma_j^2 + \mu_j^2) - \mu_{eq}^2 \tag{3.5}$$

From (3.2), (3.4) and (3.5), using $k = 2$, variance of normal mixture is easily calculated.

Plots of above four densities including the normal condition are shown in Figure 1 when $\sigma^2 = 3$. Thus parameter of each density is $\Theta = 1/\sqrt{3}$, $c = 3$, and $\nu = 3$. Also, using (3.2), we have $\mu_1 = \sqrt{2}$ and $\sigma_1^2 = \sigma_2^2 = 1$ to satisfy the equal variance between 5 densities.

Figure 1. Plots of densities for error terms



4. Simulation Study

Under the model setup in Section 3, given data $\boldsymbol{y} = (y_1, \cdots, y_m)^T$, we shall calculate the simulated Bayes risk differences for two EB and HB estimators under 4 model perturbation cases given by

$$\frac{1}{mR}\sum_{i=1}^{m}\sum_{r=1}^{R}(\hat{\theta}_{ir}^{EB(p)}-\hat{\theta}_{ir}^{EB(n)})^2 \tag{4.1}$$

$$\frac{1}{mR}\sum_{i=1}^{m}\sum_{r=1}^{R}(\hat{\theta}_{ir}^{HB(p)}-\hat{\theta}_{ir}^{HB(n)})^2 \tag{4.2}$$

We shall now conduct a simualtion study and then calculate the empirical and hierarchical Bayes estimators. We proceed our simulation in the following way:

Step 1. Start the values $A=1$ and $m=10$.

Step 2. Iterate the following procedure 10,000 times.

A) Generate θ_i, $i=1,\cdots,m$, from the normal density with mean 0 and variance A.

B) Choose one of the error model and generate e_i's from its density. Adding $y_i=\theta_i+e_i$, we have new samples $\boldsymbol{y}=(y_1,\cdots,y_m)^T$.

C) Use (2.2) and (2.9), respectively, to obtain the $\hat{\theta}_{ir}^{EB(p)}$ and $\hat{\theta}_{ir}^{HB(p)}$ estimators with given data $\boldsymbol{y}$.

D) Calculate $\hat{\theta}_{ir}^{EB(n)}$ and $\hat{\theta}_{ir}^{HB(n)}$ with given normal data.

E) After 10,000 iterations, calculate (4.1) and (4.2).

Step 3. Modifying $A(=1,2,3)$ and $m(=10,30,50,100)$, repeat the Step 1 and 2.

After 10,000 times iterations, these given quantities are provided in Table 1 and 2 when $\sigma^2=2$ and Table 3 and 3 when $\sigma^2=3$, respectively.

5. Conclusion

In this paper, we observed the effects for estimating the Bayes estimators when assumption that the error terms are independently and normally distributed is no longer satisfied. As for nonnormal cases about error such as exponential distribution, uniform distribution, heavy-tailed distribution (i.e., Student t-distribution with low degrees of freedom) and mixture distribution of normals, we computed the Bayes risk difference, which indicates difference between Bayes estimators whether normality of error is provided or not. Our simulations show that uniform case has the smallest quantity than any other cases. However, as m is large enough, (we use $m=100$), all Bayes risk differences among 4 cases are nearly same. This is followed by the Central Limit Theorem. Therefore, if we fall in the situation that the basic assumption for the error is not satisfied, finding more information for analysis will be an alternative solution.

Table 1. Simulated Bayes risk difference for the Empirical Bayes(EB) estimator when $\sigma^2 = 2$.

A	m	Simulated Bayes risk difference for EB			
		Exponential	Uniform	t	Mixture
1	10	1.857682	1.562864	2.020998	1.609981
	30	0.970262	0.823018	0.981484	0.845192
	50	0.760994	0.673663	0.771765	0.680867
	100	0.612954	0.559175	0.622611	0.565584
2	10	2.159192	1.952189	2.155329	1.990813
	30	1.413695	1.326626	1.436548	1.337724
	50	1.252127	1.196393	1.257925	1.202925
	100	1.132534	1.101349	1.145327	1.106238
3	10	2.428180	2.307849	2.429307	2.323360
	30	1.785523	1.725255	1.784403	1.736631
	50	1.657591	1.616681	1.665927	1.619311
	100	1.540722	1.520612	1.553463	1.517464

Table 2. Simulated Bayes risk difference for the Hierarchical Bayes(HB) estimator when $\sigma^2 = 2$.

A	m	Simulated Bayes risk difference for HB			
		Exponential	Uniform	t	Mixture
1	10	1.231158	1.212605	1.268657	1.222339
	30	0.629980	0.631424	0.633770	0.632475
	50	0.504239	0.503871	0.504661	0.505580
	100	0.416289	0.415944	0.416219	0.416324
2	10	1.652823	1.613780	1.632313	1.617263
	30	1.178952	1.179666	1.184807	1.181775
	50	1.102116	1.104663	1.104613	1.103606
	100	1.054553	1.054176	1.055971	1.055721
3	10	2.113442	2.082222	2.094915	2.078273
	30	1.873554	1.875266	1.874860	1.877861
	50	1.837898	1.837979	1.838611	1.839089
	100	1.803275	1.802409	1.805150	1.802743

Table 3. Simulated Bayes risk difference for the Empirical Bayes(EB) estimator when $\sigma^2 = 3$.

A	m	Simulated Bayes risk difference for EB			
		Exponential	Uniform	t	Mixture
1	10	2.703815	2.095144	2.892363	2.121254
	30	1.216920	0.949108	1.503161	0.964257
	50	0.909195	0.721589	1.081143	0.732254
	100	0.643260	0.544182	0.866316	0.547590
2	10	2.973919	2.568638	3.218669	2.591406
	30	1.695215	1.504795	1.899276	1.511079
	50	1.385524	1.274057	1.593561	1.283963
	100	1.194081	1.117276	1.318456	1.116336
3	10	3.225439	2.979093	3.415805	3.008586
	30	2.132386	1.989260	2.488465	1.994696
	50	1.889301	1.794401	2.081812	1.800633
	100	1.688603	1.647797	1.840588	1.650419

Table 4. Simulated Bayes risk difference for the Hierarchical Bayes(HB) estimator when $\sigma^2 = 3$.

A	m	Simulated Bayes risk difference for HB			
		Exponential	Uniform	t	Mixture
1	10	1.680362	1.645251	1.690728	1.643587
	30	0.723506	0.718995	0.726910	0.720451
	50	0.533489	0.535871	0.542892	0.535366
	100	0.379559	0.381354	0.381060	0.381009
2	10	2.066405	2.027080	2.157352	2.032453
	30	1.195144	1.195040	1.204596	1.194297
	50	1.013015	1.015317	1.018567	1.015581
	100	0.898230	0.898681	0.897748	0.897851
3	10	2.454180	2.436912	2.487310	2.445730
	30	1.762546	1.759589	1.770420	1.759352
	50	1.655088	1.653654	1.656708	1.656593
	100	1.571932	1.571272	1.572184	1.571244

References

1. Cassella, G. (1985) : An introduction to empirical Bayes data analysis, *American Statistician* 39, 83-87
2. Efron, B., and Morris, C. (1972) : Limiting the risk of Bayes and empirical Bayes estimators-Part II : The empirical Bayes case, *J. Amer. Statist. Assoc.* 67, 13-139
3. Efron, B., and Morris, C. (1973) : Stein's estimation rule and its competitors - an empirical Bayes approach, *J. Amer. Statist. Assoc.* 68, 117-130
4. Kass, R.E. and Steffey, D. (1989) : Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Amer. Statist. Assoc.* 84, 717-726
5. Lehmann. E. L. (1983) : *Theory of Point Estimation*, Wiley, New York
6. Lindley, D., and Smith, A. F. M. (1972) : Bayes estimates for the linear model, *J. R. Statist. Soc.* B 34, 1-41
7. Morris, C.N. (1981) : Parametric empirical Bayes confidence intervals, *In Scientific Inference, Data Analysis, and Robustness*, eds. G.E.P.Box, T.Leonard and C.F. Jeff Wu, Academic Press, 25-30
8. Morris, C.N. (1981) : Parametric empirical Bayes inference and applications, *J. Amer. Statist. Assoc.* 78, 47-65
9. Rao, C.R. (1973) : Linear Statistical Inference and Its Application, Wiley, New York
10. Robbins, H. (1951) : Aymptotically subminimax solutions of compound statistical decision problems, *Proc, 2nd Berk. Symp. Math. Statist. Prob.* V1, University of California Press, Berkeley, 131-184
11. Strawderman, W.E. (1971) : Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* 42, 385-388