〈Research Paper〉

Journal of the Korean Statistical Society
Vol. 15, No. 2, 1986.

An Empirical Bayes Estimation of Multivariate Normal Mean Vector†

Hea-Jung Kim*

ABSTRACT

Assume that $X_1, X_2, \cdots, X_N$ are iid p-dimensional normal random vectors ($p\geq 3$) with unknown covariance matrix. The problem of estimating multivariate normal mean vector in an empirical Bayes situation is considered. Empirical Bayes estimators, obtained by Bayes treatment of the covariance matrix, are presented. It is shown that the estimators are minimax, each of which dominates the maximum likelihood estimator (MLE), when the loss is nonsingular quadratic loss. We also derive approximate credibility region for the mean vector that takes advantage of the fact that the MLE is not the best estimator.

1. Introduction and Summary

James and Stein (1961) showed that if $X_1, X_2, \cdots, X_N$ are independent p-dimensional random vectors ($p\geq 3$) following nonsingular normal distribution $N_p(\theta, \Sigma)$, then

$$\hat{\theta}_{JS}=\left[1-\frac{(p-2)}{(N-p+2)N\bar{X}'V^{-1}\bar{X}}\right]\bar{X} \tag{1·1}$$

will dominate $\bar{X}$ *w.r.t.* a quadratic loss

$$L[\hat{\theta}\,;\,(\theta, \Sigma)]=(\theta-\hat{\theta})'\Sigma^{-1}(\theta-\hat{\theta}), \tag{1·2}$$

where

$$\bar{X}=\frac{1}{N}\sum_{j=1}^{N}X_j, \text{ and } V=\sum_{j=1}^{N}(X_j-\bar{X})(X_j-\bar{X})'.$$

Stein (1962) also suggested the so-called positive part estimator to improve upon (1·1). Ever since James and Stein's result, statisticians have searched for uniformly better

* Department of Statistics, Dongguk University, Seoul 100, Korea.

† This work was supported by a grant from the Ministry of Education.

(minimax) estimators. Lin and Tsai(1973) have obtained, extending the Baranchik's (1970) result, conditions for a family of minimax estimators with unknown Σ, and Efron and Morris(1976) have obtained the class of estimation rules invariant under orthogonal transformations whose unbiased estimate of risk is bounded uniformly by the risk of MLE. Efron and Morris(1976) also provided minimax estimators using a particular loss function for estimating the inverse of a covariance matrix. One might expect a rush of applications of this powerful new statistical weapon(minimax estimator), but such has not been the case. Resistance has formed along several lines stated in Efron and Morris(1975).

Clearly, it would be desirable to provide uniformly better estimators which eliminate the resistance. In this paper we shall give such desirable estimators under the empirical Bayes approach. This approach, which uses the data to estimate the hyperparameters of prior distribution, has merit of robustness against misspecification of prior distribution. Haff(1978) and Carter and Rolph(1974) pointed that estimating the hyperparameters, one generates estimators of the original mean vector which often have smaller risks than estimators not based on the hyperparameter structure.

Bayes estimators of θ and Σ under a natural conjugate prior are derived in Section 2 of present paper. In Section 3 of this paper, we give empirical Bayes estimators of θ under Bayes treatment of Σ. This approach eliminates one line of the resistance that mistrusts the statistical interpretation of the mathematical formulation leading to the estimators. In Section 4, we obtain unique unbiased estimator of the risk and show that the estimators are minimax which dominate the MLE. We also present a numerical evidence which supports the assertion that our estimators are minimax. Finally, we derive limiting distribution of credibility region for θ which can be used to make inference about θ. Incidentally, this eliminates another line of the resistance.

2. Bayes Estimator of θ under a Natural Conjugate Prior

Suppose $X_1, X_2, \cdots, X_N$ are independent and identically distributed as $N_p(\theta, \Sigma)$. Then

$$(\bar{X},\ V)=\left(\frac{1}{N}\sum_{j=1}^{N} X_j,\ \sum_{j=1}^{N}(X_j-\bar{X})(X_j-\bar{X})'\right) \tag{2·1}$$

is sufficient for (θ, Σ). The joint density of $\bar{X}$ and V is given by

$$p(\bar{X},\ V|\theta,\ \Sigma)\propto\frac{|V|^{(N-p-2)/2}\exp[-1/2\,\mathrm{tr}(\Sigma^{-1}V)]}{|\Sigma|^{(N-1)/2}}\cdot\frac{\exp[-N/2(\theta-\bar{X})'\Sigma^{-1}(\theta-\bar{X})]}{|\Sigma|^{1/2}}. \tag{2·2}$$

Consider a joint natural conjugate prior distribution of θ and Σ with *pdf* given by

$$p(\theta, \Sigma) \propto \frac{\exp[-1/2 \cdot \theta'(c\Sigma)^{-1}\theta]}{|\Sigma|^{1/2}} \cdot \frac{\exp[-1/2 \cdot \mathrm{tr}(\Sigma^{-1}\Omega)]}{|\Sigma|^{m/2}} \tag{2·3}$$

where $c>0$, Ω, and $m>2p$ are parameters of the prior distribution. Here we used the Bayesian assumption that $\theta|c, \Sigma \sim N_p(0, c\Sigma)$ for the identification of hyperparameters. This is the same assumption as Efron and Morris(1973, 1976). The corresponding joint posterior *pdf* of θ and Σ becomes

$$p(\theta, \Sigma|\bar{X}, V) \propto \frac{\exp\{-1/2 \cdot \mathrm{tr}\Sigma^{-1}[V+\Omega+N(\theta-\bar{X})(\theta-\bar{X})'+\theta\theta'/c]\}}{|\Sigma|^{(N+m+1)/2}}. \tag{2·4}$$

The marginal posterior distributions of θ and Σ are

$$p(\theta|\bar{X}, V) \propto \frac{1}{[1+(N+1/c)(\theta-a)'K^{-1}(\theta-a)]^{(N+m-p)/2}} \tag{2·5}$$

which is a multivariate t-distribution, and

$$p(\Sigma|\bar{X}, V) \propto \frac{\exp[-1/2 \cdot \mathrm{tr}(\Sigma^{-1}K)]}{|\Sigma|^{(N+m)/2}}, \tag{2·6}$$

which is an inverted Wishart $W^{-1}(K, p, N+m)$ where

$$a=N\bar{X}/(N+1/c), \text{ and } K=V+\Omega+\frac{(N/c)\bar{X}\bar{X}'}{[N+(1/c)]}.$$

Hence with respect to the quadratic loss in (1·2), the Bayes estimator for θ is

$$\hat{\theta}_{\mathrm{Bayes}}=(1-B)\bar{X}, \text{ where } B=1/(1+cN), \text{ and } N+m-2p\geq 1. \tag{2·7}$$

If the statistician does not know B, or equivalently c, he cannot use the Bayes rule $\hat{\theta}_{\mathrm{Bayes}}$. He can, however, attempt to estimate B from the data.

3. An Empirical Bayes Estimator for θ

Let us look again at the Bayes estimator for θ. When B is unknown, we cannot use the Bayes rule

$$\hat{\theta}_{\mathrm{Bayes}}=(1-B)\bar{X}, \ N+m-2p\geq 1, \text{ where } B=1/(1+cN).$$

By estimating B from the data we can find what we call the empirical Bayes estimator of θ. Under (2·1), the sample mean vector $\bar{X}$ has conditional mean vector θ and the covariance matrix Σ/N,

$$\bar{X}|\theta, \Sigma \sim N_p(\theta, \Sigma/N), \tag{3·1}$$

and the unscaled covariance matrix of the sample observation vectors has Wishart distribution,

$$V|\Sigma \sim W(\Sigma, p, N-1). \tag{3·2}$$

The joint natural conjugate prior (2·3) provides the unknown mean vector as an independent sample from conditional multivariate normal distribution with mean zero and covariance matrix $c\Sigma$,

$$\theta \mid c,\ \Sigma \sim N_p(0,\ c\Sigma), \tag{3·3}$$

and the unknown covariance matrix Σ as inverted Wishart distribution,

$$\Sigma \mid m,\ \Omega \sim W^{-1}(\Omega,\ p,\ m). \tag{3·4}$$

To use the prior information of Σ in estimating B, it is instructive to do the analysis conditional on Σ being known and then to turn to the actual situation with unknown Σ.

Lemma 1. If Σ is known, then an empirical Bayes estimator of the multivariate normal mean vector is

$$\hat{\theta}_{EB}(\Sigma^{-1}) = [1-(p-2)/(N\bar{X}'\Sigma^{-1}\bar{X})]\bar{X}. \tag{3·5}$$

Proof. It will be convenient to work with the problem in the canonical form. Define $Z=\sqrt{N}\Sigma^{-1/2}\bar{X}$, and $\gamma=\sqrt{N}\Sigma^{-1/2}\theta$. Then, from (3·1) and (3·3), $Z\mid\gamma \sim N_p(\gamma,\ I)$, and $\gamma \sim N_p(0,\ (cN)I)$. The Bayes estimator of γ is given by the posterior mean. Now marginally $Z \sim N_p(0,\ (1/B)I)$, and $\gamma\mid Z,\ \Sigma \sim N_p((1-B)Z,\ (1-B)I)$, where $B=1/(1+cN)$ so that the Bayes estimator of γ is

$$\hat{\gamma}_{\text{Bayes}}(\Sigma) = E[\gamma\mid Z,\ \Sigma] = (1-B)Z.$$

For the estimation of independent components, the empirical Bayes approach is well studied. Efron and Morris(1973) empirically estimates B as $(p-2)/(Z'Z)$, which gives the usual(James-Stein) empirical Bayes estimator. Rewriting in the original X's, we have the result. ▨

Now since the Bayes rule unconditional on Σ is the expected value of the Bayes rule conditional on Σ and since $\hat{\theta}_{EB}(\Sigma^{-1})$ depends only upon Σ^{-1}, we need only replace Σ^{-1} by $\hat{\Sigma}^{-1} \cong E[\Sigma^{-1}\mid\bar{X},\ V]$ to obtain an unconditional empirical Bayes estimator from (2·7). However, since (2·6) gives the marginal posterior distribution

$$\Sigma^{-1}\mid\bar{X},\ V \sim W(K^{-1},\ p,\ N+m-p-1) \tag{3·6}$$

to Σ^{-1}, we have the following theorem:

Theorem 2. If Σ is unknown, then an empirical Bayes estimator of (2·7) is

$$\hat{\theta}^*_{EB} = \left[1-\frac{(p-2)(N+m-2p-2)}{(N-1)(N+m-p-1)}\frac{1}{N\bar{X}'V^{-1}\bar{X}}\right]\bar{X}, \tag{3·7}$$

where $m>2p+2$.

Proof. Under the quadratic loss of the form $L(\Sigma^{-1},\ \hat{\Sigma}^{-1}) = \text{tr}(\Sigma^{-1}-\hat{\Sigma}^{-1})^2$, the Bayes estimator for Σ^{-1} is $E[\Sigma^{-1}|\overline{X}, V]$. This is found, from (3·6), to be

$$E[\Sigma^{-1}|\overline{X}, V]=(N+m-p-1)K^{-1}=(N+m-p-1)(V+\Omega+NB\overline{X}\overline{X}')^{-1}. \quad (3\cdot 8)$$

To get the empirical Bayes estimator of (2·7) with an estimated value of Σ^{-1}, we estimate the posterior expectation for Σ^{-1} from (3·8) by $(N+m-p-1)\hat{K}^{-1}$. We avoid iteration by defining $\hat{K}^{-1}$ as

$$\hat{K}^{-1}=[V+\Omega+V/(N-1)]^{-1}=[NV/(N-1)+\Omega]^{-1}.$$

From (3·2) and (3·4), $E[V]=E[E[V|\Sigma]]$ gives the moment estimator $\hat{\Omega}=(m-2p-2)\ V/(N-1)$. (Note the last term in K^{-1} in (3·8) has expectation Σ so we use $V/(N-1)$ to estimate it.) Rewriting in the original coordinates we have

$$\hat{\Sigma}^{-1}=[(N-1)(N+m-p-1)V^{-1}/(N+m-2p-2)]. \quad (3\cdot 9)$$

Finally, combining (3·5) and (3·9) gives the desired result. ■

Since it is possible for $N\overline{X}'V^{-1}\overline{X}$ to become smaller than $(p-2)(N+m-2p-2)/[(N-1)(N+m-p-1)]$ even if the components of θ are positive, it is not unreasonable to try to improve upon (3·7) by the "Positive part rule" which yields

$$\hat{\theta}^{**}_{EB}=\left[\mathrm{Max}\left(0,\ 1-\frac{(p-2)(N+m-2p-2)}{(N-1)(N+m-p-1)}\cdot\frac{1}{N\overline{X}'V^{-1}\overline{X}}\right)\right]\overline{X}. \quad (3\cdot 10)$$

See Efron and Morris(1976) for the "Positive part rule."

4. Risks and Minimax Property of Estimators

So far we have considered empirical Bayes estimators. In this section we give general expressions for the risks of the empirical Bayes estimators dealt in the previous sections. These calculations yield a way of comparing the empirical Bayes and maximum likelihood estimators.

Theorem3. Suppose $\overline{X}|\theta, \Sigma\sim N_p(\theta, \Sigma/N)\ (p\geq 3)$ where $\Sigma=DG$ with G and D being a known pxp matrix and an unknown constant, respectively. Suppose, furthermore, that $\hat{D}\sim DX_n^2/(n+2)$, independent of $\overline{X}$. With respect to the invariant loss in (1·2), the invariant estimator of θ has the form

$$\hat{\theta}_{IE}=[1-(p-2)\tau(F)/F]\overline{X},\quad F=N\overline{X}'G^{-1}\overline{X}/\hat{D} \quad (4\cdot 1)$$

with $\tau(F)$ any real valued function on $(0, \infty)$. If the risk $R_n(\lambda)$ of (4·1) is finite and if $\tau(F)$ is absolutely continuous, then a unique unbiased estimator of $R_n(\lambda)$ based on F exists and is

$$\hat{R}_n(F)=1/N\{p-(p-2)[(p-2)\tau(F)(2-\tau(F))/F+4\tau'(F)(1+C_n\tau(F))]\}, \quad (4\cdot2)$$

where $C_n=(p-2)/(n+2)$, and $\tau'(F)$ is derivative of $\tau(F)$. Note that $R_n(\lambda)$ is a function only of $\lambda=N\theta'G^{-1}\theta/2D$.

Proof. Ths assertion is easily proved by making the usual transformation $Z=\sqrt{N}G^{-1/2}\bar{X}$ to reduce to the situation of Efron and Morris(1976). ▨

Corollary 1. $\hat{\theta}_{EB}$ in (3·5) has unique unbiased estimator of risk

$$\hat{R}_n(F)=[p-(p-2)^2/(N\bar{X}'\Sigma^{-1}\bar{X})]/N. \quad (4\cdot3)$$

Proof. It is easy to see that (3·5) gives $F=NX'\Sigma^{-1}X$, $\tau(F)=1$, and $D=1$. Consequently, the right side of (4·2) gives (4·3). ▨

Corollary 2. $\hat{\theta}_{EB}{}^*$ in (3·7) has unique unbiased estimator of risk

$$\hat{R}_n(F)=\frac{1}{N}\left\{p-(p-2)\frac{p-2}{(N-p+2)N\bar{X}'V^{-1}\bar{X}}\left[\frac{(N+m-2p-2)(N-p+2)}{(N-1)(N+m-p-1)}\right]\right.$$
$$\left.\left[2-\frac{(N-p+2)(N+m-2p-2)}{(N-1)(N+m-p-1)}\right]\right\},\ \text{where } n=N-p. \quad (4\cdot4)$$

Proof. Under (3·1) and (3·2), define $Y=\sqrt{N}\bar{X}$, and $V=\sum_{\alpha=1}^{N-1}Z_\alpha Z_\alpha'$ so that $Y\sim N_p(\sqrt{N}\theta,\Sigma)$, and $Z_\alpha \underset{\sim}{\text{iid}} N_p(0,\Sigma)$, and make the transformation $Y^*=\Sigma^{-1/2}Y$, $V^*=\Sigma^{-1/2}V\Sigma^{-1/2}$, $\mu^*=\sqrt{N}\Sigma^{-1/2}\theta$, and $Z_\alpha{}^*=\Sigma^{-1/2}Z_\alpha$. Suppose that a pxp orthogonal matrix Γ is defined by $\gamma_{1i}=Y_i^*/\sqrt{Y^{*\prime}Y^*}$, $i=1,2,\cdots,p$ with the other $p-1$ arbitrary rows. Then

$$\bar{X}'\Sigma^{-1}X/(\bar{X}'V^{-1}\bar{X})=UU/(U'B^{-1}U)=1/b^{11}=b_{11\cdot2,3,\cdots,p},$$

where $B=\Gamma V^*\Gamma'=\begin{bmatrix}b_{11}B_{12}\\B_{21}B_{22}\end{bmatrix}$, and $U=\Gamma Y^*=(\sqrt{Y^{*\prime}Y^*},0,0,\cdots,0)'$.

By Theorem 4.3.3(Anderson, 1958) the conditional distribution of $b_{11\cdot2,3,\cdots,p}$ does not depend on Γ, it is unconditionally distributed as $\chi_{(N-p)}{}^2$, that is, $\bar{X}'\Sigma^{-1}\bar{X}/(\bar{X}'V^{-1}\bar{X})\sim\chi^2{}_{(N-p)}$, being independent of $\bar{X}'\Sigma^{-1}\bar{X}$. Set $(N-p+2)\hat{D}=\bar{X}'\Sigma^{-1}\bar{X}/(\bar{X}'V^{-1}\bar{X})$ and $F=N\bar{X}'\Sigma^{-1}\bar{X}/\hat{D}$, then we have a situation identical to the Theorem 3 with $n=N-p$, $D=1$, $G=\Sigma$. Comparing (3·7) and (4·1), we get $\tau(F)=\dfrac{(N+m-2p-2)(N-p+2)}{(N-1)(N+m-p-1)}$, and $F=N(N-p+2)\bar{X}'V^{-1}\bar{X}$. Consequently the right side of (4·2) gives (4·4). ▨

Corollary 3. The unbiased estimator of the risk of $\hat{\theta}^{**}{}_{EB}$ in (3·10) is

$$\hat{R}_n(F)=\begin{cases}\left(\dfrac{N-p-2}{N-p+2}F-p\right)/N & \text{if } \hat{\theta}^{**}{}_{EB}=0\\ \hat{R}_n(F)\ \text{derived in (4·4)} & \text{if } \hat{\theta}^{**}{}_{EB}=\hat{\theta}^*{}_{EB},\end{cases} \quad (4\cdot5)$$

where $F=N(N-p+2)\bar{X}'V^{-1}\bar{X}$.

Proof. Here we need to prove only for the case of $\hat{\theta}^{**}_{EB}=0$. From (2·1) we know that $(N-p)N\bar{X}'V^{-1}\bar{X}/p\sim F_{p,N-p}(\lambda)$, where $\lambda=N\theta'\Sigma^{-1}\theta/2$. Hence $E\left[\left(\frac{N-p-2}{N-p+2}F-p\right)\Big/N\right]=\theta'\Sigma^{-1}\theta$ which is the risk of $\hat{\theta}_{EB}^{**}=0$ under (1·2). ▨

Lin and Tsai(1973) and Efron and Morris(1976) derived conditions for a family of minimax estimators, each of which dominates the "usual" one, for the problem of simultaneously estimating means of three or more dependent normal variables with unknown covariance matrix Σ. Using their conditions, we show how to compare the risks of our estimators to the corresponding quantity for $\bar{X}$.

Theorem 4. Empirical Bayes estimators, derived in (3·7) and (3·10), are minimax estimators which dominate the MLE.

Proof. According to general minimax conditions theorem, Lin and Tsai(1976) and Efron and Morris(1976), it suffices to show that $\tau(F)$ involved in (4·4) and (4·5) is absolutely continuous, and

$$0\leq\tau(F)\leq 2 \quad \text{for all } F\geq 0.$$

Since as is well known, relative to the loss function (1·2), $\bar{X}$ is minimax with constant risk p/N, and since $\hat{R}_n(F)$, defined in (4·4) and (4·5), is unique unbiased estimator of risk $R_n(\lambda)$ based on F,

$$E[\hat{R}_n(F)]=R_n(\lambda)\leq p/N \quad \text{if } \hat{R}_n(F)\leq p/N \quad (\equiv 0\leq\tau(F)\leq 2) \quad \text{for all } F.$$

Observe that right hand sides of (4·4) and (4·5) (or (3·7) and (3·10)) involve that $0<\tau(F)=\frac{(N+m-2p-2)(N-p+2)}{(N-1)(N+m-p-1)}<1$ except for the case $\frac{(p-2)(N+m-2p-2)}{(N-1)(N+m-p-1)}>N\bar{X}'V^{-1}\bar{X}$ in (3·10). In this case, set $0\leq\tau(F)=(N-p+2)N\bar{X}'V^{-1}\bar{X}/(p-2)\leq 1$ to get $\hat{\theta}^{**}_{EB}=0$. Then the estimators satisfy the minimax conditions and, hence, are minimax which dominate the MLE. ▨

Following example gives a numerical evidence that our estimator (3·7) dominates the MLE.

Example Suppose we take $m=2p+3$ to reflect minimal prior knowledge but to permit $E[V]$ to exist. Then the unbiased risk (4·4) is

$$\hat{R}_n(F)=\frac{1}{N}\left[p-(p-2)\frac{(p-2)}{F}t(2-t)\right], \quad \text{where } t=\frac{(N+1)(N-p+2)}{(N-1)(N+p+2)}.$$

By taking expectation on both sides, and using Jensen's inequality, we get the inequality

Table 1. Lower bounds for savings of risk
$[\underset{(\lambda, N, p)}{\text{Min}} R(\theta, X) - R(\theta, \hat{\theta}^*_{EB})]$

$\theta' \Sigma^{-1} \theta$ \ Sample size (N) Dimension (p)	$N=10,\ p=5$	$N=20,\ p=5$	$N=100,\ p=5$
0	.05810	.06245	.01716
5	.00582	.00297	.00016
10	.00276	.00152	.00007
15	.00187	.00102	.00005
20	.00141	.00077	.00003
30	.00095	.00051	.00002
40	.00071	.00038	.00001
50	.00057	.00031	.00001
60	.00048	.00026	.00001
70	.00041	.00022	.00001
80	.00036	.00019	.00001
90	.00032	.00017	.00001
100	.00028	.00015	.000009

$$\frac{p}{N} - R_n(\lambda) \geq \frac{(p-2)^2}{NE[F]} t(2-t), \text{ where } E[F] = \frac{(N-p+2)(N\theta' \Sigma^{-1}\theta + p)}{(N-p-2)}.$$

This inequality gives the lower bounds for the saving of risk. The lower bounds are tabulated in Table 1 for the various amounts of p, N and $\theta' \Sigma^{-1} \theta$. Table 1 shows that the empirical Bayes estimator dominates the MLE.

5. Approximate Cerdibility Region for Multivariate Normal Mean

The goal of this section is to show that one can inference about θ under (3·7). This is done via the limiting distribution of credibility region for θ.

Theorem 5. Let a, K and c be the same quantities defined in (2·6).
Then

$$\lim_{N \to \infty} \mathscr{L}\{K^{*-1/2}(\theta - a) \mid \bar{X},\ V\} = N_p(0, I), \tag{5·1}$$

where $K^* = K/\{q(N+1/c)\} > 0$.

Proof. Rewrite (2·5) n the standard form of multivariate Student t-density for $\theta \mid X$, V as

$$p(\theta \mid X,\ V) = M_p |K^*|^{-1/2} / [q + (\theta - a)' K^{*-1} (\theta - a)]^{(p+q)/2}, \tag{5·2}$$

where $\quad M_p = q^{q/2} \Gamma[(q+p)/2] / [\pi^{p/2} \Gamma(q/2)]$, and $q = N - 2p + m > N + 2$.

Using the transformation defined by $H = K^{*-1/2}(\theta - a)$, we get the density of H, which is

$$p(H|\bar{X}, V)=M_p|K^*|^{-1/2}[q+H'H]^{-(q+p)/2}J(\theta\to H)=M_p[q+H'H]^{-(q+p)/2}, \quad (5\cdot3)$$

where $J(\theta\to H)=|K^*|^{1/2}$.

Hence, as $q\to\infty$ or, equivalently, $N\to\infty$, (5·3) approaches multivariate normal density. Such that

$$\lim_{N\to\infty} p(H|\bar{X}, V)=\lim_{N\to\infty}\frac{q^{q/2}\Gamma[(q+p)/2]}{\pi^{p/2}\Gamma(q/2)}[q+H'H]^{-(q+p)/2}=(2\pi)^{-p/2}\exp[-1/2H'H],$$

which completes the proof. Note that $a=\hat{\theta}_{\text{Bayes}}$. ▨

Corollary 1. If N is large, it is approximately true that

$$\lim_{N\to\infty}\mathscr{L}\{\hat{K}^{*-1/2}(\theta-\hat{\theta}_{EB}^*)|\bar{X}, V\}=N_p(0, I) \quad (5\cdot4)$$

Proof. Note that all the unknown parameters $E[\Sigma^{-1}|\bar{X}, V]$, Ω and B, appeared in Lemma 1 and Theorem 2, are estimated by method of moments. Thus $\hat{\Sigma}^{-1}$, $\hat{\Omega}$ and $\hat{B}$ are consistent estimators of corresponding parameters, respectively. Using properties of consistent estimator, and the relation

$$\hat{\theta}_{\text{Bayes}}=E[\theta|\bar{X}, V]=E[E[\theta|\bar{X}, V, \Sigma]],$$

we have obtained empirical Bayes estimators which have property

$$\hat{\theta}^*_{EB}\xrightarrow{p}E_\Sigma[\hat{\theta}_{EB}(\Sigma^{-1})], \text{ and } E_\Sigma[\hat{\theta}_{EB}(\Sigma^{-1})]\xrightarrow{p}\hat{\theta}_{\text{Bayes}} \text{ as } N\to\infty.$$

This gives that

$$\hat{\theta}_{EB}^*\xrightarrow{p}\hat{\theta}_{\text{Bayes}}, \text{ for} \quad (5\cdot5)$$

$$P\{|\hat{\theta}_{\text{Bayes}}-\hat{\theta}_{EB}^*|>\varepsilon\}\le P\{|E_\Sigma\hat{\theta}_{EB}(\Sigma^{-1})-\hat{\theta}_{EB}^*|>\varepsilon/2\}+P\{|\hat{\theta}_{\text{Bayes}}-E_\Sigma[\hat{\theta}_{EB}(\Sigma^{-1})]|>\varepsilon/2\}.$$

The relations in (2·7), (3·8) and (5·5) give

$$\hat{K}^{*-1}\xrightarrow{p}K^{*-1} \quad (5\cdot6)$$

The limiting distribution of $K^{*-1/2}(\theta-\hat{\theta}_{\text{Bayes}})$ is $N_p(0, I)$, and since

$\hat{\theta}^*_{EB}\xrightarrow{p}\hat{\theta}_{\text{Bayes}}$, and $\hat{K}^{*-1}\xrightarrow{p}K^{*-1}$, $\hat{K}^{*-1/2}(\theta-\hat{\theta}^*_{EB})$ has the same distribution. ▨

Corollary 1 gives another aspect of our estimator (3·7) in a sense that we can inference about θ under (3·7). For example, substituting the consistent parameter values into the approximation gives $(1-\alpha)\times100\%$ approximate credibility region for θ as

$$(\theta-\hat{\theta}^*_{EB})'\hat{K}^{*-1}(\theta-\hat{\theta}^*_{EB})\le\chi^2_{(\alpha;p)}.$$

Acknowledgement

The detailed and constructive comments of referees were very useful in making the paper more readable and in correcting some errors. Their efforts are greatly appreciated.

References

(1) Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons.

(2) Branchik, A.J. (1970). A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution. *Ann. Math. Statist,* *41*, 642~645.

(3) Carter, G.M., and J.E. Rolph (1974). Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities. *J. Amer. Statist. Assoc.*, Vol. *69*, *No. 348*, 880~885.

(4) Efron, B., and C. Morris (1973). Stein's Estimation Rule and Its Competitors: An Empirical Bayes Approach. *J. Amer. Statist. Assoc.*, Vol. *68*, *No. 341*, 117~130.

(5) Efron, B., and C. Morris (1975). Data Analysis Using Stein's Estimator and Its Generalizations. *J. Amer. Statist. Assoc.*, Vol. *70*, 311~319.

(6) Efron, B., and C. Morris (1976). Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution. *Ann. of Statist.*, Vol. *4*, 11~21.

(7) Haff, L.R. (1978). The Multivariate Normal Mean with Intraclass Correlated Components: Estimation of Urban Fire Alarm Probabilities. *J. Amer. Statist. Assoc.*, Vol. *73*, *No.364*, 767~774.

(8) James, W., and C. Stein (1961). Estimation with Quadratic Loss. *Proceedings of the Fourth Berkeley Symposium*, Vol. *1*, 361~379.

(9) Lin, Pi-Erh, and Tsai, Hui-Liang (1973). Generalized Bayes Minimax Estimators of the Multivariate Normal Mean with Unknown Covariance Matrix. *Ann. of Statist.* Vol. 1, 142~145.

(10) Stein, C. (1962). Confidence Sets for the Mean of a Multivariate Normal Distribution. *J. of Royal Statist. Soc., Series B*, Vol. *24*, 265~285.