HIERARCHICAL AND EMPIRICAL BAYES MULTIVARIATE ESTIMATION

Malay Ghosh*, Department of Statistics, University of Florida

Abstract

This article reviews and unifies the hierarchical and empirical Bayes approach for estimating the multivariate normal mean. Both the ANOVA and the regression models are considered.

Introduction

Empirical and hierarchical Bayes methods are becoming increasingly popular 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 known relationships among the coordinates of a parameter vector, say $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$ 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 the observations.

Closely related to the EB procedure is the hierarchical Bayes (HB) procedure which models the prior distribution in 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 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 method like the method of moments, method of maximum likelihood etc., and use the resulting estimated priors for inferential purposes. It turns out that the two methods can quite often

*This paper is dedicated to Professor D. Basu on the occasion of his 65th birthday. The research is partially supported by NSF Grant Numbers DMS 8701814 and DMS 8901334.

lead to comparable results, especially in the context of point estimation. This will be revealed in some of the examples appearing in the later sections. However, when it comes to the question of measuring the standard errors associated with these 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 approximation* (see, e.g., Morris, 1981, 1983), whereas the posterior variances, though often complicated, can be found exactly.

The above ideas will be made more concrete in the subsequent sections with the aid of examples. Ours is an expository article which compares and contrasts the EB and the HB methods for multivariate normal linear models. The outline of the remaining sections is as follows. In the next section, we address the problem of estimating the multivariate normal mean. EB procedures for such problems are discussed quite adequately in Efron and Morris (1973), Morris (1981, 1983) and Casella (1985). However, the interrelationship between the EB and the HB procedures for such problems is not discussed in these papers. Lindley and Smith (1972) introduced and provided a detailed discussion of the HB approach for estimating the multivariate normal mean. However, there is no mention of the EB approach in their paper.

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.

In the third section, we consider the regression problem. The EB and the HB methods are contrasted both for the balanced and unbalanced linear models. This section is largely a review of the work of Lindley and Smith (1972) as well as Morris (1981, 1983). However, for the unbalanced case, our calculations go beyond those of Lindley and Smith (1972). It is our belief that the present calculations will shed more light on some of the EB approximations of Morris (1983). For the balanced case, the reader is also referred to Berger (1985).

Extensive development of the EB methodology began with Robbins (1951, 1955), who called problems of the above type *compound decision problems.* In Robbins's terminology, an EB procedure is one where $X_1,\ldots,X_p$ are the *past* data about $\theta_1,\ldots,\theta_p$. The past data should be used together with the current data to infer on a current θ_i. The terminological distinction between the EB and compound decision problems will be ignored in this article, and the term *empirical Bayes* will be used to cover problems of both types. Also, Robbins's procedure is a nonparametric EB procedure in contrast to the parametric EB approach taken in this paper.

The term *hierarchical Bayes* was first used by Good (1965). Lindley and Smith (1972) called such priors *multistage priors.* As noted earlier, the latter used the idea very effectively for estimating the vector of normal means, as well as the vector of regression coefficients.

Estimation of the Multivariate Normal Mean

This section is devoted to a comparison of the EB and the HB procedures for estimating the multivariate normal mean. We begin with a simple example.

I. Conditional on $\theta_1,\ldots,\theta_p$, let $X_1,\ldots,X_p$ be independent with $X_i \sim N(\theta_i, \sigma^2)$, $i = 1,\ldots,p$, σ^2 (> 0) being known. Without loss of generality, assume $\sigma^2 = 1$.

II. The θ_i's have independent $N(\mu_i, A)$, $i = 1,\ldots,p$ priors. Write $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$, $\underset{\sim}{X} = (X_1,\ldots,X_p)^T$ and $\underset{\sim}{x} = (x_1,\ldots,x_p)^T$.

The posterior distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X} = \underset{\sim}{x}$ is then $N((1-B)\underset{\sim}{x} + B\underset{\sim}{\mu}, (1-B)\underset{\sim}{I}_p)$, where $B = (A+1)^{-1}$. Accordingly, the posterior mean (the usual Bayes estimate) of $\underset{\sim}{\theta}$ is given by

$$E(\underset{\sim}{\theta}|\underset{\sim}{X} = \underset{\sim}{x}) = (1-B)\underset{\sim}{x} + B\underset{\sim}{\mu}. \tag{1}$$

In an EB or a HB scenario, some or all of the prior parameters are unknown. In an EB set up, these parameters are estimated from the marginal distribution of $\underset{\sim}{X}$ which in this case is $N(\underset{\sim}{\mu}, B^{-1}\underset{\sim}{I}_p)$. 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 three cases.

Case I. Let $\mu_1 = \ldots = \mu_p = \mu$ (say), where μ (real) is unknown, but A (> 0) is known. Based on the marginal distribution of $\underset{\sim}{X}$, $\bar{X}$ is the UMVUE, MLE and the best equivariant estimator of μ. Accordingly, from (1), an EB estimator of $\underset{\sim}{\theta}$ is given by

$$\underset{\sim}{\theta}_{EB}^{(1)} = (1-B)\underset{\sim}{X} + B\bar{X}\underset{\sim}{1}_p. \tag{2}$$

The estimator given in (2) was proposed by Lindley and Smith (1972). They used a HB approach to arrive at the estimator given in (2). The procedure is described below.

Consider the HB model under which (i) conditional on $\underset{\sim}{\theta}$ and μ, $\underset{\sim}{X} \sim N(\underset{\sim}{\theta}, \underset{\sim}{I}_p)$; (ii) conditional on μ, $\underset{\sim}{\theta} \sim N(\mu\underset{\sim}{1}_p, A\underset{\sim}{I}_p)$; (iii) μ is uniform on $(-\infty, \infty)$. Then the joint (improper) pdf of $\underset{\sim}{X}$, $\underset{\sim}{\theta}$ and μ is given by

$$f(\underset{\sim}{x}, \underset{\sim}{\theta}, \mu) \propto exp\left[-\frac{1}{2}\|\underset{\sim}{x} - \underset{\sim}{\theta}\|^2\right] A^{-\frac{1}{2}p}\, exp\left[-\frac{1}{2A}\|\underset{\sim}{\theta} - \mu\underset{\sim}{1}_p\|^2\right]. \tag{3}$$

The factor $A^{-\frac{1}{2}p}$ could have been left out in (3), but will be needed for later calculations.

Integrating with respect to μ in (3), it follows that the joint (improper) pdf of $\underset{\sim}{X}$ and $\underset{\sim}{\theta}$ is

$$f(\underset{\sim}{x}, \underset{\sim}{\theta}) \propto exp\left[-\frac{1}{2}(\underset{\sim}{\theta}^T \underset{\sim}{D}\underset{\sim}{\theta} - 2\underset{\sim}{\theta}^T\underset{\sim}{x} + \underset{\sim}{x}^T\underset{\sim}{x})\right], \tag{4}$$

where $\underset{\sim}{D} = A^{-1}[(A+1)\underset{\sim}{I}_p - p^{-1}\underset{\sim}{J}_p]$ with $\underset{\sim}{J}_p = \underset{\sim}{1}_p\underset{\sim}{1}_p^T$. Recall that $B = (A+1)^{-1}$. It follows from (4) that the posterior distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X} = \underset{\sim}{x}$ is $N(\underset{\sim}{D}^{-1}\underset{\sim}{x}, \underset{\sim}{D}^{-1})$. Since $\underset{\sim}{D}^{-1} = (1-B)\underset{\sim}{I}_p + Bp^{-1}\underset{\sim}{J}_p$, one gets

$$E(\underset{\sim}{\theta}|\underset{\sim}{X} = \underset{\sim}{x}) = (1-B)\underset{\sim}{x} + B\bar{x}\underset{\sim}{1}_p; \tag{5}$$

$$V(\underset{\sim}{\theta}|\underset{\sim}{X} = \underset{\sim}{x}) = (1-B)\underset{\sim}{I}_p + Bp^{-1}\underset{\sim}{J}_p. \tag{6}$$

A naive EB approach as noted earlier uses the estimated posterior distribution $N((1-B)\underset{\sim}{x} + B\bar{x}\underset{\sim}{1}_p, (1-B)\underset{\sim}{I}_p)$ to infer about $\underset{\sim}{\theta}$. A comparison of (2) and (5) reveals that the EB and the HB approaches yield the same point estimate of $\underset{\sim}{\theta}$, but the naive EB approach estimates the posterior variance by $(1-B)\underset{\sim}{I}_p$, which is an underestimate when compared to (6). This point is discussed more fully below.

Based on (3), the posterior distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{x}$ and μ is $N((1-B)\underset{\sim}{x} + B\mu\underset{\sim}{1}_p, (1-B)\underset{\sim}{I}_p)$. Also, integrating with respect to $\underset{\sim}{\theta}$ in (3), it follows that the joint (improper) distribution of $\underset{\sim}{x}$ and μ is given by

$$f(\underset{\sim}{x}, \mu) \propto B^{\frac{1}{2}p} exp\left[-\frac{1}{2} B\|\underset{\sim}{x} - \mu\underset{\sim}{1}_p\|^2\right]. \tag{7}$$

It follows from (7) that the posterior distribution of μ given $\underset{\sim}{X} = \underset{\sim}{x}$ is $N(\bar{x}, (Bp)^{-1})$. Hence, one may note that

$$(1 - B)\underset{\sim}{I}_p = E[V(\underset{\sim}{\theta}|\underset{\sim}{X}, \mu)|\underset{\sim}{X}]; \tag{8}$$

$$\begin{aligned} Bp^{-1}\underset{\sim}{J}_p &= V[B\mu\underset{\sim}{1}_p|\underset{\sim}{X}] = V[(1 - B)\underset{\sim}{X} + B\mu\underset{\sim}{1}_p|\underset{\sim}{X}] \\ &= V[E(\underset{\sim}{\theta}|\underset{\sim}{X}, \mu)|\underset{\sim}{X}]. \end{aligned} \tag{9}$$

Thus a naive EB procedure ignores estimating $V[E(\underset{\sim}{\theta}|\underset{\sim}{X}, \mu)|\underset{\sim}{X}]$ which amounts to ignoring the uncertainty involved in estimating the prior parameters when estimating the posterior variance.

It is shown in Lindley and Smith (1972) that the risk of $\hat{\underset{\sim}{\theta}}_{EB}^{(1)}$ is *not* uniformly smaller than that of $\underset{\sim}{X}$ under the squared error loss $L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \|\underset{\sim}{\theta} - \underset{\sim}{a}\|^2$.

However, there is a Bayes risk superiority of $\hat{\theta}_{EB}^{(1)}$ over X which is described below.

Theorem 1

Consider the model $X|\theta \sim N(\theta, I_p)$ and the prior $\theta \sim N(\mu 1_p, AI_p)$. Let E denote expectation over the joint distribution of X and θ. Then, assuming the matrix loss $L_1(\theta, a) = (a-\theta)(a-\theta)^T$, and writing $\hat{\theta}_B$ as the Bayes estimator of θ under L_1,

$$EL_1(\theta, X) = I_p;$$

$$EL_1(\theta, \hat{\theta}_B) = (1-B)I_p;$$

$$EL_1(\theta, \hat{\theta}_{EB}^{(1)}) = (1-B)I_p + Bp^{-1}J_p. \tag{10}$$

Next assuming the quadratic loss $L_2(\theta, a) = (a-\theta)^T Q(a-\theta)$, where Q is a known non-negative definite (n.n.d.) weight matrix,

$$EL_2(\theta, X) = tr(Q),\ EL_2(\theta, \hat{\theta}_B) = (1-B)tr(Q); \tag{11}$$

$$EL_2(\theta, \hat{\theta}_{EB}^{(1)}) = (1-B)tr(Q) + B\ tr(Qp^{-1}J_p). \tag{12}$$

Proof. Note that $\hat{\theta}_B = (1-B)X + B\mu 1_p$. It is immediate that $EL_1(\theta, X) = E[(X-\theta)(X-\theta)^T] = E(I_p) = I_p$ and $EL_1(\theta, \hat{\theta}_B) = E[V(\theta|X)] = E[(1-B)I_p] = (1-B)I_p$. Also, since marginally $\bar{X} \sim N(\mu, (Bp)^{-1})$

$$\begin{aligned} EL_1(\theta, \hat{\theta}_{EB}^{(1)}) &= EL_1(\theta, \hat{\theta}_B) + E\,[(\hat{\theta}_B-\hat{\theta}_{EB}^{(1)})(\hat{\theta}_B-\hat{\theta}_{EB}^{(1)})^T] \\ &= (1-B)I_p + B^2 E\,[(\bar{X}-\mu)^2 1_p 1_p^T] \\ &= (1-B)I_p + Bp^{-1}J_p. \end{aligned}$$

This completes the proof of (10). To prove (11) and (12), write $L_2(\theta, a) = (\theta-a)^T Q(\theta-a) = tr(QL_1(\theta, a))$ and use (10).

Remark 1. It follows from (10)–(12) that $E[L_i(\theta, X) - L_i(\theta, \hat{\theta}_{EB}^{(1)})]$ is nonnegative definite for each $i = 1, 2$. Accordingly, $\hat{\theta}_{EB}^{(1)}$ has smaller Bayes risk than that of X both under the matrix loss L_1, and *a fortiori* the quadratic loss L_2. To our knowledge, this particular optimality of the Lindley-Smith estimator has not been pointed out before.

The perfect agreement between the EB and the HB point estimators of θ in Case I is an exception rather than the rule. We now consider cases II and III which reveal that the point estimators of $\underset{\sim}{\theta}$ can also differ under the two approaches.

Case II. Assume that $\underset{\sim}{\mu}$ is known, but its components need not be all equal. Moreover, this time A is unknown. The marginal distribution of $\underset{\sim}{X}$ is $N(\underset{\sim}{\mu}, B^{-1}\underset{\sim}{I}_p)$. Then $||\underset{\sim}{X}-\underset{\sim}{\mu}||^2$ is complete sufficient, and is distributed as $B^{-1}\chi_p^2$. Accordingly, for $p \geq 3$, the UMVUE of B is given by $(p-2)/||\underset{\sim}{X}-\underset{\sim}{\mu}||^2$. Substituting this estimator of B in (1), an EB estimator of $\underset{\sim}{\theta}$ is given by

$$\hat{\underset{\sim}{\theta}}_{EB}^{(2)} = \left(1 - \frac{p-2}{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}\right)\underset{\sim}{X} + \frac{p-2}{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}\underset{\sim}{\mu}$$

$$= \underset{\sim}{X} - \frac{p-2}{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}(\underset{\sim}{X}-\underset{\sim}{\mu}). \tag{13}$$

This is the celebrated James-Stein estimator (James and Stein, 1961). The EB interpretation of this estimator was given in a series of articles by Efron and Morris (1972, 1973, 1975). The most popular version of this estimator takes $\underset{\sim}{\mu} = \underset{\sim}{0}$.

It is shown in James and Stein that for $p \geq 3$, the risk of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ is smaller than that of $\underset{\sim}{X}$ under the squared error loss. However, if the loss is changed to the arbitrary quadratic loss L_2 of Theorem 1, then the risk dominance of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ over $\underset{\sim}{X}$ does not necessarily hold. Indeed, it is well-known that (see, e.g., Bock, 1975, or Berger, 1975) that under the loss L_2, $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ dominates $\underset{\sim}{X}$ if (i) $tr(\underset{\sim}{Q}) > 2ch_1(\underset{\sim}{Q})$ and (ii) $0 < p-2 < 2[tr(\underset{\sim}{Q})/ch_1(\underset{\sim}{Q}) - 2]$, where $ch_1(\underset{\sim}{Q})$ denotes the largest eigen-value of $\underset{\sim}{Q}$.

The Bayes risk of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ is, however, smaller than that of $\underset{\sim}{X}$ under the losses L_1 and L_2, the model given in Theorem 1, and the prior $N_p(\underset{\sim}{\mu}, A\underset{\sim}{I}_p)$. As before, let E denote expectation over the joint distribution of $\underset{\sim}{X}$ and $\underset{\sim}{\theta}$. The following theorem is proved.

Theorem 2

Let $\underset{\sim}{X}|\underset{\sim}{\theta} \sim N(\underset{\sim}{\theta}, \underset{\sim}{I}_p)$ and $\underset{\sim}{\theta} \sim N_p(\underset{\sim}{\mu}, A\underset{\sim}{I}_p)$. Then for $p \geq 3$,

$$E[L_1(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_{EB}^{(2)})] = \underset{\sim}{I}_p - B(p-2)p^{-1}\underset{\sim}{I}_p; \tag{14}$$

$$E[L_2(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_{EB}^{(2)})] = tr(\underset{\sim}{Q}) - B(p-2)p^{-1}tr(\underset{\sim}{Q}). \tag{15}$$

Proof. To prove (14), use the identity

$$E[L_1(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_{EB}^{(2)})] = E[L_1(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_B)] + E[(\hat{\underset{\sim}{\theta}}_B - \hat{\underset{\sim}{\theta}}_{EB}^{(2)})(\hat{\underset{\sim}{\theta}}_B - \hat{\underset{\sim}{\theta}}_{EB}^{(2)})^T]. \tag{16}$$

Next write

$$E[(\hat{\underset{\sim}{\theta}}_B - \hat{\underset{\sim}{\theta}}_{EB}^{(2)})(\hat{\underset{\sim}{\theta}}_B - \hat{\underset{\sim}{\theta}}_{EB}^{(2)})^T] = E\left[\left(B - \frac{p-2}{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}\right)^2 (\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T\right]. \tag{17}$$

Marginally, $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, B^{-1}\underset{\sim}{I}_p)$. Hence, $||\underset{\sim}{X} - \underset{\sim}{\mu}||^2$ is complete sufficient, while $(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^2 = B^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/(B^{-1}||\underset{\sim}{X} - \underset{\sim}{\mu}||^2)$ is ancillary. Hence, using Basu's Theorem (see Basu, 1955), $(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^2$ is distributed independently of $||\underset{\sim}{X} - \underset{\sim}{\mu}||^2$. Hence,

$$\begin{aligned} B^{-1}\underset{\sim}{I}_p &= E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T] \\ &= E[||\underset{\sim}{X} - \underset{\sim}{\mu}||^2\{(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^2\}] \\ &= E(||\underset{\sim}{X} - \underset{\sim}{\mu}||^2)E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^2]. \end{aligned}$$

Now using $E||\underset{\sim}{X} - \underset{\sim}{\mu}||^2 = B^{-1}p$, $E||\underset{\sim}{X} - \underset{\sim}{\mu}||^{-2} = B(p-2)^{-1}$ for $p \geq 3$, one gets

$$E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^2] = p^{-1}\underset{\sim}{I}_p; \tag{18}$$

$$\begin{aligned} &E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^4] \\ &\quad = E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T/||\underset{\sim}{X} - \underset{\sim}{\mu}||^2]E(||\underset{\sim}{X} - \underset{\sim}{\mu}||^{-2}) \\ &\quad = (p^{-1}\underset{\sim}{I}_p)B(p-2)^{-1}. \end{aligned} \tag{19}$$

It follows from (17)–(19) that

$$\begin{aligned} &E[(\hat{\underset{\sim}{\theta}}_B - \hat{\underset{\sim}{\theta}}_{EB}^{(2)})(\hat{\underset{\sim}{\theta}}_B - \hat{\underset{\sim}{\theta}}_{EB}^{(2)})^T] \\ &\quad = B^2(B^{-1}\underset{\sim}{I}_p) - 2B(p-2)p^{-1}\underset{\sim}{I}_p + B(p-2)p^{-1}\underset{\sim}{I}_p \\ &\quad = B\underset{\sim}{I}_p - B(p-2)p^{-1}\underset{\sim}{I}_p. \end{aligned} \tag{20}$$

Combining (10), (16) and (20), one gets (14). The proof of (15) is immediate from (14) by writing $L_2(\underset{\sim}{\theta}, \underset{\sim}{a}) = tr[\underset{\sim}{Q}L_1(\underset{\sim}{\theta}, \underset{\sim}{a})]$.

Remark 2. Taking $\underset{\sim}{Q}$ as a matrix with its $(i, i)^{th}$ element equal to 1 and the rest zeroes, it follows that the i^{th} component of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ dominates X_i when one compares their Bayes risks. This co-ordinatewise Bayes risk dominance of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ over $\underset{\sim}{X}$ appears in Efron and Morris (1973). One can derive (15) from their work by using an orthogonal transformation. The dominance of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ over $\underset{\sim}{X}$ under the matrix loss L_1 has not been pointed out before, but the approach appears in Reinsel (1985) for a more complex EB problem.

Remark 3. Efron and Morris (1973) found it convenient to define the concept of relative savings loss (RSL). Denote the given prior by ξ and the Bayes risk of an estimator $\underset{\sim}{e}$ of $\underset{\sim}{\theta}$ under the prior ξ and the loss L_2 by $r(\xi, \underset{\sim}{e})$. The RSL of $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ with respect to $\underset{\sim}{X}$ is defined by

$$
\begin{aligned}
RSL(\hat{\underset{\sim}{\theta}}_{EB}^{(2)}; \underset{\sim}{X}) &= [r(\xi, \hat{\underset{\sim}{\theta}}_{EB}^{(2)}) - r(\xi, \hat{\underset{\sim}{\theta}}_B)]/[r(\xi, \underset{\sim}{X}) - r(\xi, \hat{\underset{\sim}{\theta}}_B)] \\
&= 1 - [r(\xi, \underset{\sim}{X}) - r(\xi, \hat{\underset{\sim}{\theta}}_{EB}^{(2)})]/[r(\xi, \underset{\sim}{X}) - r(\xi, \hat{\underset{\sim}{\theta}}_B)]. \qquad (21)
\end{aligned}
$$

This is the proportion of the possible Bayes risk improvement over $\underset{\sim}{X}$ that is sacrificed by using $\hat{\underset{\sim}{\theta}}_{EB}^{(2)}$ rather than the ideal estimator $\hat{\underset{\sim}{\theta}}_B$ under the prior ξ. It follows from (11), (15) and (16) that $RSL(\hat{\underset{\sim}{\theta}}_{EB}^{(2)}; \underset{\sim}{X}) = 2/p$ for an arbitrary n.n.d. non-null matrix $\underset{\sim}{Q}$. Efron and Morris (1973) proved the result when $\underset{\sim}{Q} = \underset{\sim}{I}_p$ as well as when the $(i, i)^{th}$ element of $\underset{\sim}{Q}$ is 1 and the rest zeroes ($i = 1,\ldots,p$). For the matrix loss L_1, the RSL concept of Efron and Morris (1973) can be generalized to get

$$
\begin{aligned}
\text{RSL}(\hat{\underset{\sim}{\theta}}_{EB}^{(2)}; \underset{\sim}{X}) &= [\underset{\sim}{r}(\xi, \underset{\sim}{X}) - \underset{\sim}{r}(\xi, \hat{\underset{\sim}{\theta}}_B)]^{-1}[\underset{\sim}{r}(\xi, \hat{\underset{\sim}{\theta}}_{EB}^{(2)}) - \underset{\sim}{r}(\xi, \hat{\underset{\sim}{\theta}}_B)] \\
&= (B\underset{\sim}{I}_p)^{-1}(B(2/p))\underset{\sim}{I}_p = (2/p)\underset{\sim}{I}_p. \qquad (22)
\end{aligned}
$$

Suppose now we consider a HB approach in this case, where conditional on $\underset{\sim}{\theta}$ and A, $\underset{\sim}{X} \sim N(\underset{\sim}{\theta}, \underset{\sim}{I}_p)$, and conditional on A, $\underset{\sim}{\theta} \sim N(\underset{\sim}{\mu}, A\underset{\sim}{I}_p)$. Also, let A have marginal pdf $g_0(A)$. Then, the joint pdf of $\underset{\sim}{X}$, $\underset{\sim}{\theta}$ and A is

$$
f(\underset{\sim}{x}, \underset{\sim}{\theta}, A) \propto exp\left[-\tfrac{1}{2}||\underset{\sim}{x} - \underset{\sim}{\theta}||^2\right] A^{-\frac{1}{2}p}\, exp\left[-\tfrac{1}{2} A^{-1}||\underset{\sim}{\theta} - \underset{\sim}{\mu}||^2\right] g_0(A). \qquad (23)
$$

As before, the conditional distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{x}$ and A is $N((1-B)\underset{\sim}{x} + B\underset{\sim}{\mu}, (1-B)\underset{\sim}{I}_p)$, where $B = (A+1)^{-1}$. But integrating with respect to $\underset{\sim}{\theta}$, the joint pdf of $\underset{\sim}{X}$ and A is

$$
f(\underset{\sim}{x}, A) \propto (A+1)^{-\frac{1}{2}p}\, exp\left[-\frac{1}{2(A+1)}||\underset{\sim}{x} - \underset{\sim}{\mu}||^2\right] g_0(A). \qquad (24)
$$

Since $B = (A+1)^{-1}$, the joint pdf of $\underset{\sim}{X}$ and B is of the form

$$f(\underset{\sim}{x}, B) \propto B^{\frac{1}{2}p} \, exp\left[-\frac{1}{2} B \,||\underset{\sim}{x} - \underset{\sim}{\mu}||^2\right] g(B). \tag{25}$$

The HB approach of the above type was first proposed by Strawderman (1971), and was later generalized by Faith (1978). Assuming the Type II Beta density for A, namely $g_0(A) \propto A^{m-1}(1+A)^{-(m+n)}$, where m (> 0) and n (> 0), it is easy to see that

$$f(\underset{\sim}{x}, B) \propto B^{\frac{1}{2}p+n-1}(1-B)^{m-1} \, exp\left[-\frac{1}{2} B \,||\underset{\sim}{x} - \underset{\sim}{\mu}||^2\right]. \tag{26}$$

Now, using the iterated formula for conditional expectations,

$$\underset{\sim}{\hat{\theta}}{}^{(2)}_{HB} \equiv E(\underset{\sim}{\theta}|\underset{\sim}{x}) = E(E[\underset{\sim}{\theta}|B, \underset{\sim}{x}] \,|\, \underset{\sim}{x}) = (1 - \hat{\hat{B}})\underset{\sim}{x} + \hat{\hat{B}}\underset{\sim}{\mu}, \tag{27}$$

where

$$\hat{\hat{B}} = E(B|\underset{\sim}{x}) = \int_0^1 B^{\frac{1}{2}p+n}(1-B)^{m-1} \, exp\left[-\frac{1}{2} B||\underset{\sim}{x} - \underset{\sim}{\mu}||^2\right] dB$$

$$\div \int_0^1 B^{\frac{1}{2}p+n-1}(1-B)^{m-1} \, exp\left[-\frac{1}{2} B||\underset{\sim}{x} - \underset{\sim}{\mu}||^2\right] dB. \tag{28}$$

Strawderman (1971) considered the case $m = 1$, and found sufficient conditions on n under which the risk of $\underset{\sim}{\hat{\theta}}{}^{(2)}_{HB}$ is smaller than that of $\underset{\sim}{X}$. His results were generalized to a certain extent by Faith (1978).

We consider also the case $m = 1$, and interpreting (26) as the posterior pdf of B given $\underset{\sim}{x}$, find the posterior mode of B as

$$\hat{B}_{MO} = min((p+2n-2)/||\underset{\sim}{x} - \underset{\sim}{\mu}||^2, 1). \tag{29}$$

Substituting this estimator of B in (1), one gets the estimator

$$\underset{\sim}{\hat{\theta}}{}^{(3)}_{HB} = (1 - \hat{B}_{MO})\underset{\sim}{X} + \hat{B}_{MO}\underset{\sim}{\mu} = \underset{\sim}{X} - \hat{B}_{MO}(\underset{\sim}{X} - \underset{\sim}{\mu}) \tag{30}$$

of $\underset{\sim}{\theta}$. The special choice $n = 0$ leads to the positive part James-Stein estimator which is known to dominate the usual James-Stein estimator (see Lehmann, 1983, p. 302). This is intuitively very clear since the usual James-Stein estimator substitutes the UMVUE of B in (1), and this UMVUE can take values exceeding 1 with positive probability while $0 < B < 1$. This deficiency is rectified by $\hat{B}_{MO}$.

Case III. The model is similar to the one in Case I, except that now μ (real) and A (> 0) are both unknown. Recall that marginally $\underset{\sim}{X} \sim N(\mu \underset{\sim}{1}_p, B^{-1}\underset{\sim}{I}_p)$ where $B = (A+1)^{-1}$. Hence, $(\bar{X}, \sum_1^p (X_i-\bar{X})^2)$ is complete sufficient, so that the UMVUE's of μ and B are given respectively by $\bar{X}$ and $(p-3)/\sum_1^p (X_i-\bar{X})^2$. Substituting these estimators of μ and B in (1), the EB estimator of $\underset{\sim}{\theta}$ is given by

$$\hat{\underset{\sim}{\theta}}_{EB}^{(3)} = \left(1 - \frac{p-3}{\sum_1^p (X_i-\bar{X})^2}\right)\underset{\sim}{X} + \frac{p-3}{\sum_1^p (X_i-\bar{X})^2}\bar{X}\underset{\sim}{1}_p$$

$$= \underset{\sim}{X} - \frac{p-3}{\sum_1^p (X_i-\bar{X})^2}(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}_p). \tag{31}$$

This modification of the James-Stein estimator was proposed by Lindley (1962). Whereas, the original James-Stein estimator shrinks $\underset{\sim}{X}$ towards a specified point, the modified estimator given in (31) shrinks $\underset{\sim}{X}$ towards a hyperplane spanned by $\underset{\sim}{1}_p$.

The estimator $\hat{\underset{\sim}{\theta}}_{EB}^{(3)}$ is known to dominate $\underset{\sim}{X}$ for $p \geq 4$. Its Bayes risk under the L_1 and L_2 losses are not known however. We now prove a theorem to this effect quite in the spirit of Theorems 1 and 2.

Theorem 3

Assume the model and the prior given in Theorem 1. Then, for $p \geq 4$,

$$E[L_1(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_{EB}^{(3)})] = \underset{\sim}{I}_p - B(p-3)(p-1)^{-1}(\underset{\sim}{I}_p - p^{-1}\underset{\sim}{J}_p); \tag{32}$$

$$E[L_2(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_{EB}^{(3)})] = tr(\underset{\sim}{Q}) - B(p-3)(p-1)^{-1}tr[\underset{\sim}{Q}(\underset{\sim}{I}_p - p^{-1}\underset{\sim}{J}_p)]. \tag{33}$$

Proof. First write

$$E[L_1(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_{EB}^{(3)})] = E[L_1(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_B)] + E(\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B)(\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B)^T. \tag{34}$$

We write

$$\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B = \left(B - \frac{p-3}{\sum_1^p (X_i-\bar{X})^2}\right)(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}_p) + B(\bar{X} - \mu)\underset{\sim}{1}_p. \tag{35}$$

Now using the independence of $\underset{\sim}{X} - \bar{X}\underset{\sim}{1}_p$ and $\bar{X}$, and using the fact that $\bar{X} \sim N(\mu, (Bp)^{-1})$, one gets from (35),

$$E\left[\left(\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B\right)\left(\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B\right)^T\right]$$

$$= E\left[\left(B - \frac{p-3}{\sum(X_i-\bar{X})^2}\right)^2 (\underset{\sim}{X}-\bar{X}\underset{\sim}{1}_p)(\underset{\sim}{X}-\bar{X}\underset{\sim}{1}_p)^T\right] + B^2(Bp)^{-1}\underset{\sim}{J}_p. \tag{36}$$

Next using the independence of $(\underset{\sim}{X}-\bar{X}\underset{\sim}{1}_p)(\underset{\sim}{X}-\bar{X}\underset{\sim}{1}_p)^T \big/ \sum_1^p (X_i-\bar{X})^2$ with $\sum_1^p (X_i-\bar{X})^2$ (again by applying Basu's Theorem) and the facts that $E[(\underset{\sim}{X}-\bar{X}\underset{\sim}{1}_p)(\underset{\sim}{X}-\bar{X}\underset{\sim}{1}_p)^T] = B^{-1}(\underset{\sim}{I}_p-p^{-1}\underset{\sim}{J}_p)$, while $\sum_1^p (X_i-\bar{X})^2 \sim B^{-1}\chi^2_{p-1}$, it follows from (36) that for $p \geq 4$,

$$E\left[\left(\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B\right)\left(\hat{\underset{\sim}{\theta}}_{EB}^{(3)} - \hat{\underset{\sim}{\theta}}_B\right)^T\right]$$

$$= B^2B^{-1}\left(\underset{\sim}{I}_p-p^{-1}\underset{\sim}{J}_p\right) - 2B(p-3)(p-1)^{-1}\left(\underset{\sim}{I}_p-p^{-1}\underset{\sim}{J}_p\right)$$

$$+ (p-3)^2B(p-3)^{-1}(p-1)^{-1}\left(\underset{\sim}{I}_p-p^{-1}\underset{\sim}{J}_p\right) + Bp^{-1}\underset{\sim}{J}_p$$

$$= B\underset{\sim}{I}_p - B(p-3)(p-1)^{-1}\left(\underset{\sim}{I}_p-p^{-1}\underset{\sim}{J}_p\right). \tag{37}$$

Combining (10), (34) and (37), one gets (32). The proof of (33) is immediate from (32).

We now proceed to find the HB estimator of $\underset{\sim}{\theta}$. Consider the model where (i) conditional on $\underset{\sim}{\theta}$, μ and A, $\underset{\sim}{X} \sim N(\underset{\sim}{\theta}, \underset{\sim}{I}_p)$; (ii) conditional on μ and A, $\underset{\sim}{\theta} \sim N(\mu\underset{\sim}{1}_p, A\underset{\sim}{I}_p)$; (iii) marginally μ and A are independently distributed with μ uniform on $(-\infty, \infty)$, and A has uniform improper pdf on $(0, \infty)$. Then the joint (improper) pdf of $\underset{\sim}{X}$, $\underset{\sim}{\theta}$, μ and A is given by

$$f(\underset{\sim}{x}, \underset{\sim}{\theta}, \mu, A) \propto exp\left[-\frac{1}{2}\|\underset{\sim}{x} - \underset{\sim}{\theta}\|^2\right] A^{-\frac{1}{2}p}\, exp\left[-\frac{1}{2A}\|\underset{\sim}{\theta} - \mu\underset{\sim}{1}_p\|^2\right]. \tag{38}$$

Now integrating with respect to μ, it follows from (38) that the joint (improper) pdf of $\underset{\sim}{X}$, $\underset{\sim}{\theta}$ and A is

$$f(\underset{\sim}{x}, \underset{\sim}{\theta}, A) \propto A^{-\frac{1}{2}(p-1)} exp\left[-\frac{1}{2}(\underset{\sim}{\theta}-\underset{\sim}{D}^{-1}\underset{\sim}{x})^T\underset{\sim}{D}(\underset{\sim}{\theta}-\underset{\sim}{D}^{-1}\underset{\sim}{x}) - \frac{1}{2(A+1)}\sum_1^p (x_i-\bar{x})^2\right], \tag{39}$$

where $\underset{\sim}{D}$ is defined after (6). Recall $\underset{\sim}{D}^{-1} = (1-B)\underset{\sim}{I}_p + Bp^{-1}\underset{\sim}{J}_p$. Hence, conditional on $\underset{\sim}{x}$ and A, $\underset{\sim}{\theta} \sim N[(1-B)\underset{\sim}{x} + B\bar{x}\underset{\sim}{1}_p, (1-B)\underset{\sim}{I}_p + Bp^{-1}\underset{\sim}{J}_p]$. Also, integrating with respect to $\underset{\sim}{\theta}$ in (39), one gets the joint pdf of $\underset{\sim}{X}$ and A given by

$$f(\underset{\sim}{x}, A) \propto (A+1)^{-\frac{1}{2}(p-1)} exp\left[-\frac{1}{2(A+1)} \sum_1^p (x_i-\bar{x})^2\right]. \tag{40}$$

Since $B = (A+1)^{-1}$, it follows from (40) that the joint pdf of $\underset{\sim}{X}$ and B is given by

$$f(\underset{\sim}{x}, B) \propto B^{\frac{1}{2}(p-1)} exp\left[-\frac{1}{2} B \sum_1^p (x_i-\bar{x})^2\right] B^{-2}$$

$$= B^{\frac{1}{2}(p-5)} exp\left[-\frac{1}{2} B \sum_1^p (x_i-\bar{x})^2\right]. \tag{41}$$

It follows from (41) that

$$E(B|\underset{\sim}{x}) = \int_0^1 B^{\frac{1}{2}(p-3)} exp\left[-\frac{1}{2} B\sum_1^p (x_i-\bar{x})^2\right] dB$$

$$\div \int_0^1 B^{\frac{1}{2}(p-5)} exp\left[-\frac{1}{2} B\sum_1^p (x_i-\bar{x})^2\right] dB; \tag{42}$$

$$E(B^2|\underset{\sim}{x}) = \int_0^1 B^{\frac{1}{2}(p-1)} exp\left[-\frac{1}{2} B\sum_1^p (x_i-\bar{x})^2\right] dB$$

$$\div \int_0^1 B^{\frac{1}{2}(p-5)} exp\left[-\frac{1}{2}B\sum_1^p (x_i-\bar{x})^2\right] dB. \tag{43}$$

One can obtain $V(B|\underset{\sim}{x})$ from (42) and (43), and use these to obtain

$$E(\underset{\sim}{\theta}|\underset{\sim}{x}) = \underset{\sim}{x} - E(B|\underset{\sim}{x})(\underset{\sim}{x}-\bar{x}\underset{\sim}{1}_p); \tag{44}$$

$$V(\underset{\sim}{\theta}|\underset{\sim}{x}) = V[E(\underset{\sim}{\theta}|B, \underset{\sim}{x})|\ \underset{\sim}{x}] + E[V(\underset{\sim}{\theta}|B, \underset{\sim}{x})|\ \underset{\sim}{x}]$$

$$= V[\underset{\sim}{x} - B(\underset{\sim}{x}-\bar{x}\underset{\sim}{1}_p)|\ \underset{\sim}{x}] + E[(1-B)\underset{\sim}{I}_p + Bp^{-1}\underset{\sim}{J}_p|\ \underset{\sim}{x}]$$

$$= V(B|\ \underset{\sim}{x})(\underset{\sim}{x}-\bar{x}\underset{\sim}{1}_p)(\underset{\sim}{x}-\bar{x}\underset{\sim}{1}_p)^T + \underset{\sim}{I}_p - E(B|\underset{\sim}{x})(\underset{\sim}{I}_p - p^{-1}\underset{\sim}{J}_p). \tag{45}$$

Also, one can obtain a positive-part version of Lindley's estimator by substituting the posterior mode of B namely $min\left((p-5)\Big/\sum_1^p (X_i-\bar{X})^2,\ 1\right)$ in (1). Morris (1981) suggests approximations to $E(B|\underset{\sim}{x})$ and $E(B^2|\underset{\sim}{x})$ involving replacement of $\int_0^1$ by $\int_0^\infty$ both in the numerator as well in the denominator of (42) and (43). The resulting approximations turn out to be $E(B|\underset{\sim}{x}) \doteq (p-3)\Big/\sum_1^p (x_i-\bar{x})^2$ and, $E(B^2|\underset{\sim}{x}) \doteq (p-1)(p-3)\Big/\left\{\sum_1^p (x_i-\bar{x})^2\right\}^2$, so that $V(B|\underset{\sim}{x}) \doteq 2(p-3)\Big/\left\{\sum_1^p (x_i-\bar{x})^2\right\}^2$. Morris (1981) points out that the above approximations amount to putting a uniform prior to A on $(-1, \infty)$ rather than on $(0, \infty)$. Note that with Morris's approximations

$$E(\underset{\sim}{\theta}|\underset{\sim}{X}) \doteq \underset{\sim}{X} - \frac{p-3}{\sum_1^p (X_i-\bar{X})^2}(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}_p) = \hat{\theta}_{EB}^{(3)}, \tag{46}$$

which is Lindley's modification of the James-Stein estimator, while

$$V(\underset{\sim}{\theta}|\underset{\sim}{X}) \doteq \frac{2(p-3)}{\left(\sum_1^p (X_i-\bar{X})^2\right)^2}(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}_p)(\underset{\sim}{X} - \bar{X}\underset{\sim}{1}_p)^T$$

$$+ \underset{\sim}{I}_p - \frac{p-3}{\sum_1^p (X_i-\bar{X})^2}(\underset{\sim}{I}_p - p^{-1}\underset{\sim}{J}_p). \tag{47}$$

Morris (1981) considered a slightly more general version of the model where conditional on $\underset{\sim}{\theta}$, μ and A, $\underset{\sim}{X} \sim N(\underset{\sim}{\theta}, \sigma^2\underset{\sim}{I}_p)$, while the distributions of $\underset{\sim}{\theta}$, μ and A remain the same. If one redefines $B = \sigma^2/(\sigma^2+A)$, the only change that is needed in the calculations is that conditional on $\underset{\sim}{x}$ and A, $\underset{\sim}{\theta} \sim N((1-B)\underset{\sim}{x} + B\bar{x}\underset{\sim}{1}_p,\ \sigma^2[(1-B)\underset{\sim}{I}_p + Bp^{-1}\underset{\sim}{J}_p])$, while the conditional pdf of B given $\underset{\sim}{x}$, and accordingly $E(B|\underset{\sim}{x})$ and $V(B|\underset{\sim}{x})$ are modified by putting B/σ^2 in place of B in the exponents.

We now revisit the famous baseball data of Efron and Morris (1975). They considered the batting averages of 18 baseball players in 1970 after each had batted 45 times. Based on these batting averages, they estimated (in fact, predicted) the players' batting averages for the remainder of the season. We used formulas (42) and (43) with B/σ^2 replacing B in the exponents to get the exact expressions for $E(\theta_i|\underset{\sim}{x})$ and $V(\theta_i|\underset{\sim}{x})$. Also, we used Morris's approximations which are obtained by modifying (46) and (47). The results are given in Table 1. In

TABLE 1. The True Values (θ_i), the Maximum Likelihood Estimates (Y_i), the Hierarchical Bayes Estimates ($\hat{\theta}_{i,HB}$), the Hierarchical Bayes S.D.'s ($s_{i,HB}$), Morris's Approximate Estimates ($\hat{\theta}_{i,M}$), and Morris's Approximate S.D.'s ($s_{i,M}$).

i	θ_i	Y_i	$\hat{\theta}_{i,HB}$	$s_{i,HB}$	$[\hat{\theta}_{i,HB}-2s_{i,HB}, \hat{\theta}_{i,HB}+2s_{i,HB}]$	$\hat{\theta}_{i,M}$	$s_{i,M}$	$[\hat{\theta}_{i,M}-2s_{i,M}, \hat{\theta}_{i,M}+2s_{i,M}]$
1	0.346	0.395	0.308	0.046	[0.216,0.400]	0.293	0.073	[0.147,0.439]
2	0.300	0.375	0.301	0.044	[0.213,0.389]	0.288	0.071	[0.142,0.430]
3	0.279	0.355	0.295	0.043	[0.209,0.381]	0.284	0.069	[0.146,0.422]
4	0.223	0.334	0.288	0.042	[0.204,0.372]	0.280	0.067	[0.146,0.414]
5	0.276	0.313	0.281	0.041	[0.199,0.363]	0.275	0.066	[0.143,0.407]
6	0.273	0.291	0.281	0.041	[0.199,0.363]	0.275	0.066	[0.143,0.407]
7	0.266	0.269	0.274	0.040	[0.194,0.354]	0.271	0.066	[0.139,0.405]
8	0.211	0.247	0.267	0.040	[0.187,0.347]	0.266	0.066	[0.134,0.398]
9	0.271	0.247	0.260	0.040	[0.180,0.340]	0.262	0.067	[0.128,0.396]
10	0.232	0.247	0.260	0.040	[0.180,0.340]	0.262	0.067	[0.128,0.396]
11	0.266	0.224	0.252	0.040	[0.172,0.332]	0.257	0.068	[0.121,0.393]
12	0.258	0.224	0.252	0.040	[0.172,0.332]	0.257	0.068	[0.121,0.393]
13	0.306	0.224	0.252	0.040	[0.172,0.332]	0.257	0.068	[0.121,0.393]
14	0.267	0.224	0.252	0.040	[0.172,0.332]	0.257	0.068	[0.121,0.393]
15	0.228	0.224	0.252	0.040	[0.172,0.332]	0.257	0.068	[0.121,0.393]
16	0.288	0.200	0.244	0.041	[0.162,0.326]	0.252	0.070	[0.112,0.392]
17	0.318	0.175	0.236	0.043	[0.150,0.322]	0.247	0.073	[0.101,0.393]
18	0.200	0.148	0.227	0.045	[0.137,0.317]	0.241	0.077	[0.087,0.395]

what follows the true values θ_i's refer to the baseball players' actual batting averages for the remainder of the season. Also, $\hat{\theta}_{i,HB}$ and $\hat{\theta}_{i,M}$ denote respectively the HB estimate of θ_i and Morris's approximate estimate of θ_i. The standard errors associated with $\hat{\theta}_{i,HB}$ and $\hat{\theta}_{i,M}$ are denoted respectively by $s_{i,HB}$ and $s_{i,M}$. It turns out that

$$(18\sigma^2)^{-1}\sum_{i=1}^{18}(X_i-\theta_i)^2 = 0.976,$$

$$(18\sigma^2)^{-1}\sum_{i=1}^{18}(\hat{\theta}_{i,HB}-\theta_i)^2 = 0.299,$$

and

$$(18\sigma^2)^{-1}\sum_{i=1}^{18}(\hat{\theta}_{i,M}-\theta_i)^2 = 0.286$$

so that Morris's approximations serve well as point estimates. However, Morris's (1981) approximations to the s.d.'s are consistently larger than the actual ones, leading thereby to wider confidence intervals. It appears that Morris (1981) has reported that $\hat{\theta}_{i,HB}$'s and $s_{i,HB}$'s in his Table 1, p. 31, but his notations seem to suggest that these are $\hat{\theta}_{i,M}$'s and $s_{i,M}$'s.

So far we have considered only the case when the sampling variance σ^2 is known. In a more realistic set up, σ^2 is unknown. In such instances, one approach is to first find the Bayes estimator of $\underset{\sim}{\theta}$ assuming σ^2 to be known. Next find an estimator of σ^2, and substitute this estimator in the Bayes estimator found earlier. Berger (1985) discusses this approach. A slightly different classical EB approach can be found in Ghosh and Meeden (1986) or Ghosh and Lahiri (1987). These methods do not take into account the uncertainty involved in estimating σ^2. This deficiency can be rectified by putting a prior distribution (often non-informative) on σ^2 as well.

One important example is the unbalanced one-way ANOVA model. We propose a HB analysis with an unknown σ^2 as well as unknown parameters involved in the prior distribution of $\underset{\sim}{\theta}$. We find it convenient to reparametrize into $\sigma^2 = r^{-1}$ and $A = (\lambda r)^{-1}$. The remainder of this section is an adaptation of the arguments of Ghosh and Lahiri (1988).

Assume that

(a) conditional on $\underset{\sim}{\theta}$, m, λ and r, the random variables $X_1,\ldots,X_p$ and U are mutually independent with $X_i \sim N\left(\theta_i, (rn_i)^{-1}\right)$ $(i = 1,\ldots,p)$, while $U \sim r^{-1}\chi^2_{N-p}$ $(N = \sum_{i=1}^{p} n_i)$;

(b) conditional on m, λ and r, $\underset{\sim}{\theta} \sim N\left(m\underset{\sim}{1}_p, (\lambda r)^{-1}\underset{\sim}{I}_p\right)$;

(c) marginally, M, Λ and R are independently distributed with $M \sim$ *uniform*$(-\infty, \infty)$, R has pdf $g(r) \propto r^{-2}$, while Λ has pdf $h(\lambda) \propto \lambda^{-2}$.

Remark 3. Note that we have changed the notation from μ to m. If one assigns the noninformative prior $g(A, \sigma^2) \propto (\sigma^2)^{-1}$, then noting that $r = (\sigma^2)^{-1}$ and $\lambda r = A^{-1}$, one gets the prior on R and Λ as given in (c). It is possible to assign gamma priors (informative or noninformative) on R and ΛR as in Ghosh and Lahiri (1988), but we have decided to sacrifice that generality.

To identify the above model with an unbalanced one-way random effects ANOVA model, write $Y_{ij} = m + \tau_i + e_{ij}$ ($j = 1,\ldots,n_i$; $i = 1,\ldots,p$). Here, τ_i's and e_{ij}'s are mutually independent with τ_i's iid $N(0, (\lambda r)^{-1})$ and e_{ij}'s iid $N(0, r^{-1})$. Write $\theta_i = m + \tau_i$, $X_i = \bar{Y}_i = n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$ $(i = 1,\ldots,p)$ and $U = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2$. Clearly, $(X_1,\ldots,X_p, U)$ is minimal sufficient with joint distribution given in (a).

Under the above model, the joint pdf of $X_1,\ldots,X_p$, U, $\underset{\sim}{\theta}$, M, R and Λ is given by

$$f(\underset{\sim}{x}, u, \underset{\sim}{\theta}, m, r, \lambda) \propto r^{\frac{1}{2}p} exp\left[-\frac{1}{2} r\sum_{i=1}^{p} n_i(x_i - \theta_i)^2\right] r^{\frac{1}{2}(N-p)}$$

$$\times\ exp\left(-\frac{1}{2}ru\right)u^{\frac{1}{2}(N-p)-1}$$

$$\times\ (\lambda r)^{\frac{1}{2}p} exp\left[-\frac{1}{2}\lambda r\sum_{i=1}^{p}(\theta_i - m)^2\right](\lambda r)^{-2}. \tag{48}$$

Integrating with respect to m in (48), one gets the joint pdf of $\underset{\sim}{X}$, U, $\underset{\sim}{\theta}$, R and Λ given by

$$f(\underset{\sim}{x}, \underset{\sim}{u}, \underset{\sim}{\theta}, r, \lambda) \propto r^{\frac{1}{2}(N+p-1)-2} exp\left[-\frac{1}{2}r(\underset{\sim}{\theta}^T\underset{\sim}{D}\underset{\sim}{\theta} - 2\underset{\sim}{\theta}^T\underset{\sim}{G}\underset{\sim}{x} + \underset{\sim}{x}^T\underset{\sim}{G}\underset{\sim}{x} + u)\right]$$

$$\times\ u^{\frac{1}{2}(N-p)-1}\ \lambda^{\frac{1}{2}(p-1)-2}, \tag{49}$$

where $\underset{\sim}{G} = Diag(n_1,\ldots,n_p)$, $\underset{\sim}{D} = \underset{\sim}{G} + \lambda(\underset{\sim}{I}_p - p^{-1}\underset{\sim}{J}_p)$. Next integrating with respect to r in (49), it follows that the joint pdf of $\underset{\sim}{X}$, U, $\underset{\sim}{\theta}$ and $\underset{\sim}{\Lambda}$ is

$$f(\underset{\sim}{x}, u, \underset{\sim}{\theta}, \lambda)$$

$$\propto\ u^{\frac{1}{2}(N-p)-1}\lambda^{\frac{1}{2}(p-1)-2}(\underset{\sim}{\theta}^T\underset{\sim}{D}\underset{\sim}{\theta} - 2\underset{\sim}{\theta}^T\underset{\sim}{G}\underset{\sim}{x} + \underset{\sim}{x}^T\underset{\sim}{G}\underset{\sim}{x} + u)^{-\frac{1}{2}(N+p-1)-1}$$

$$= u^{\frac{1}{2}(N-p)-1}\lambda^{\frac{1}{2}(p-1)-2}$$

$$\times \left[(\underset{\sim}{\theta} - \underset{\sim}{D}^{-1}\underset{\sim}{G}\underset{\sim}{x})^T\underset{\sim}{D}(\underset{\sim}{\theta} - \underset{\sim}{D}^{-1}\underset{\sim}{G}\underset{\sim}{x}) + \underset{\sim}{x}^T(\underset{\sim}{G} - \underset{\sim}{G}\underset{\sim}{D}^{-1}\underset{\sim}{G})\underset{\sim}{x} + u\right]^{-\frac{1}{2}(N+p-3)}. \tag{50}$$

It is clear from (50) that the conditional pdf of $\underset{\sim}{\theta}$ given $\underset{\sim}{x}$, u and λ is multivariate-t with location parameter $\underset{\sim}{D}^{-1}\underset{\sim}{G}\underset{\sim}{x}$, scale parameter $(N\text{-}1)^{-1}[\underset{\sim}{x}^T(\underset{\sim}{G} - \underset{\sim}{G}\underset{\sim}{D}^{-1}\underset{\sim}{G})\underset{\sim}{x} + u]\underset{\sim}{D}^{-1}$ and degrees of freedom N-1. On simplification, one gets

$$\underset{\sim}{D}^{-1} = \underset{\sim}{K} + \lambda\left(\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}\right)^{-1}\underset{\sim}{K}\underset{\sim}{J}_p\underset{\sim}{K}$$

$$\times \left(\underset{\sim}{K} = Diag[(n_1+\lambda)^{-1},\ldots,(n_p+\lambda)^{-1}]\right); \tag{51}$$

$$\underset{\sim}{D}^{-1}\underset{\sim}{G} = Diag\left(n_1(n_1+\lambda)^{-1},\ldots,n_p(n_p+\lambda)^{-1}\right)$$

$$+ \lambda\left(\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}\right)^{-1}\begin{bmatrix}(n_1+\lambda)^{-1}\\ \vdots \\ (n_p+\lambda)^{-1}\end{bmatrix}$$

$$\times [n_1(n_1+\lambda)^{-1},\ldots,n_p(n_p+\lambda)^{-1}]; \tag{52}$$

$$\underset{\sim}{D}^{-1}\underset{\sim}{G}\underset{\sim}{x} = [n_1(n_1+\lambda)^{-1}x_1 + \lambda(n_1+\lambda)^{-1}\bar{x}.,\ldots,n_p(n_p+\lambda)^{-1}x_p + \lambda(n_p+\lambda)^{-1}\bar{x}.]^{\mathrm{T}}, \tag{53}$$

where $\bar{x}. = \left(\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}\right)^{-1}\left(\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}x_i\right)$. Further after much simplifications, one can write

$$\underset{\sim}{x}^T(\underset{\sim}{G} - \underset{\sim}{G}\underset{\sim}{D}^{-1}\underset{\sim}{G})\underset{\sim}{x}$$

$$= \lambda\left\{\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}x_i^2 - \left(\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}\right)^{-1}\left(\sum_{i=1}^{p} n_i(n_i+\lambda)^{-1}x_i\right)^2\right\}$$

$$= Q_\lambda(\underset{\sim}{x}) \text{ (say)}. \tag{54}$$

Integrating with respect to $\underset{\sim}{\theta}$ in (50), one finds the joint pdf of $\underset{\sim}{X}$, U and Λ given by

$$f(\underset{\sim}{x}, u, \underset{\sim}{\lambda}) \propto \lambda^{\frac{1}{2}(p-1)-2} u^{\frac{1}{2}(N-p)-1} \times \left(Q_\lambda(\underset{\sim}{x}) + u\right)^{-\frac{1}{2}(N+p-3)} \left|\left(Q_\lambda(\underset{\sim}{x}) + u\right)\underset{\sim}{D}^{-1}\right|^{\frac{1}{2}}$$

$$= \lambda^{\frac{1}{2}(p-1)-2} u^{\frac{1}{2}(N-p)-1} \left(Q_\lambda(\underset{\sim}{x}) + u\right)^{-\frac{1}{2}(N-3)} |\underset{\sim}{D}|^{-\frac{1}{2}}. \tag{55}$$

Using $|\underset{\sim}{D}| \propto \left\{\prod_1^p (n_i + \lambda)\right\}\left(\sum_{i=1}^p n_i(n_i + \lambda)^{-1}\right)$, it follows from (55) that the conditional pdf of $\underset{\sim}{\Lambda}$ given $\underset{\sim}{x}$ and u is

$$f(\lambda|\underset{\sim}{x}, u) \propto \lambda^{\frac{1}{2}(p-3)-1} \left\{\prod_1^p (n_i + \lambda)^{-\frac{1}{2}}\right\} \times \left\{\sum_{i=1}^p n_i(n_i + \lambda)^{-1}\right\}^{-\frac{1}{2}} [Q_\lambda(\underset{\sim}{x}) + u]^{-\frac{1}{2}(N-3)}. \tag{56}$$

From the properties of multivariate-t, it follows that $E(\underset{\sim}{\theta}|\underset{\sim}{x}, u, \lambda) = \underset{\sim}{D}^{-1}\underset{\sim}{G}\underset{\sim}{x}$, given in (53), and $V(\underset{\sim}{\theta}|\underset{\sim}{x}, u, \lambda) = (N-3)^{-1}[Q_\lambda(\underset{\sim}{x}) + u]\underset{\sim}{D}^{-1}$. One obtains now $E(\underset{\sim}{\theta}|\underset{\sim}{x}, u)$ and $V(\underset{\sim}{\theta}|\underset{\sim}{x}, u)$ by using (56) and the formulas

$$\begin{aligned} E(\underset{\sim}{\theta}|\underset{\sim}{x}, u) &= E[E(\underset{\sim}{\theta}|\underset{\sim}{x}, u, \Lambda)|\underset{\sim}{x}, u]; \\ V(\underset{\sim}{\theta}|\underset{\sim}{x}, u) &= V[E(\underset{\sim}{\theta}|\underset{\sim}{x}, u, \Lambda)|\underset{\sim}{x}, u] + E[V(\underset{\sim}{\theta}|\underset{\sim}{x}, u, \Lambda)|\underset{\sim}{x}, u]. \end{aligned} \tag{57}$$

As noted already, the posterior mean of $\underset{\sim}{\theta}$ is given by (53) for known λ. Ghosh and Meeden (1986) used a classical EB procedure to estimate λ and used this estimator of λ in (53) to obtain an estimator of $\underset{\sim}{\theta}$. Although, the resulting estimator of $\underset{\sim}{\theta}$ was quite satisfactory for point estimation purposes (see Ghosh and Lahiri, 1987), the method suffered from the earlier criticism of not modelling the uncertainty in λ. The Ghosh-Meeden procedure was not particularly suitable for the construction of credible intervals or sets.

Shrinking Towards Regression Surfaces

In the preceding section, the sample mean was either shrunk towards a specified point or a subspace spanned by the vector $\underset{\sim}{1}_p$. The present section generalizes the ideas of the preceding section by shrinking the sample mean towards an arbitrary regression surface. This can be achieved by using either an EB or a HB approach. The HB approach is discussed in detail in Lindley and Smith (1972) with known variance components. Morris (1983) provides a thorough discussion of the EB procedure. We attempt a synthesis between the two, and argue that Morris's EB procedure is indeed an attempt to approximate a bonafide HB procedure, and is clearly superior to a naive EB procedure.

We begin with Morris's set up, except that we assign distributions on the unknown hyperparameters, rather than estimate them on the basis of the marginal distributions of the observations. The following model is proposed.

(A) Conditional on $\underset{\sim}{\theta}$, $\underset{\sim}{b}$ and a, let $X_1,\ldots,X_p$ be independently distributed with $X_i \sim N(\theta_i, V_i)$, $i = 1,\ldots,p$, where the V_i's are known positive constants;

(B) Conditional on $\underset{\sim}{b}$ and a, $\Theta_1,\ldots,\Theta_p$ are in- dependently distributed with $\Theta_i \sim N(\underset{\sim}{z}_i^T\underset{\sim}{b},\ a)$ $(i = 1,\ldots,p)$, where $\underset{\sim}{z}_1,\ldots,\underset{\sim}{z}_p$ are known regression vectors of dimension r and $\underset{\sim}{b}$ is $r\times 1$.

(C) $\underset{\sim}{B}$ and A are marginally independent with $\underset{\sim}{B} \sim uniform(R^r)$ and $A \sim uniform(0,\infty)$. We assume that $p \geq r+3$. Also, we write $\underset{\sim}{Z}^T = (\underset{\sim}{z}_1,\ldots,\underset{\sim}{z}_p)$; $\underset{\sim}{G} = Diag(V_1,\ldots,V_p)$ and assume rank $(\underset{\sim}{Z}) = r$.

Now the joint (improper) pdf of $\underset{\sim}{X} = (X_1,\ldots,X_p)^T$, $\underset{\sim}{\Theta} = (\Theta_1,\ldots,\Theta_p)^T$, $\underset{\sim}{B}$ and A is given by

$$f(\underset{\sim}{x},\underset{\sim}{\theta},\underset{\sim}{b},a) \propto exp\left[-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\theta})^T\underset{\sim}{G}^{-1}(\underset{\sim}{x}-\underset{\sim}{\theta})\right] a^{-\frac{1}{2}p} exp\left[-\frac{1}{2a}\|\underset{\sim}{\theta} - \underset{\sim}{Z}\underset{\sim}{b}\|^2\right]. \qquad (58)$$

Integrating with respect to $\underset{\sim}{b}$ in (58), one finds the joint (improper) pdf of $\underset{\sim}{X}$, $\underset{\sim}{\Theta}$ and A given by

$$f(\underset{\sim}{x},\underset{\sim}{\theta},a)$$

$$\propto a^{-\frac{1}{2}(p-r)} exp\left(-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\theta})^T\underset{\sim}{G}^{-1}(\underset{\sim}{x}-\underset{\sim}{\theta}) - \frac{1}{2a}\underset{\sim}{\theta}^T[\underset{\sim}{I}_p - \underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T]\underset{\sim}{\theta}\right). \qquad (59)$$

Write $\underset{\sim}{E}^{-1} = \underset{\sim}{G}^{-1} + a^{-1}\left(\underset{\sim}{I}_p - \underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\right)$. Then, one can write

$$\begin{aligned}(\underset{\sim}{x}-\underset{\sim}{\theta})^T\underset{\sim}{G}^{-1}(\underset{\sim}{x}-\underset{\sim}{\theta}) &+ a^{-1}\underset{\sim}{\theta}^T\left(\underset{\sim}{I}_p - \underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\right)\underset{\sim}{\theta}\\ &= \underset{\sim}{\theta}^T\underset{\sim}{E}^{-1}\underset{\sim}{\theta} - 2\underset{\sim}{\theta}^T\underset{\sim}{G}^{-1}\underset{\sim}{x} + \underset{\sim}{x}^T\underset{\sim}{G}^{-1}\underset{\sim}{x}\\ &= (\underset{\sim}{\theta} - \underset{\sim}{E}\underset{\sim}{G}^{-1}\underset{\sim}{x})^T\underset{\sim}{E}^{-1}(\underset{\sim}{\theta} - \underset{\sim}{E}\underset{\sim}{G}^{-1}\underset{\sim}{x}) + \underset{\sim}{x}^T(\underset{\sim}{G}^{-1} - \underset{\sim}{G}^{-1}\underset{\sim}{E}\underset{\sim}{G}^{-1})\underset{\sim}{x}. \qquad (60)\end{aligned}$$

From (59) and (60) it follows that

$$E(\underset{\sim}{\Theta}|\underset{\sim}{x},a) = \underset{\sim}{E}\underset{\sim}{G}^{-1}\underset{\sim}{x}\ ;\ V(\underset{\sim}{\Theta}|\underset{\sim}{x},a) = \underset{\sim}{E}. \qquad (61)$$

Write $u_i = V_i/(a+V_i)$ $(i = 1,\ldots,p)$, and $\underset{\sim}{D} = Diag(1-u_1,\ldots,1-u_p)$. Then, on simplification, it follows that

$$\underset{\sim}{E} = a(\underset{\sim}{I}_p - D) + (\underset{\sim}{I}_p - \underset{\sim}{D})\underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\Big(a(\underset{\sim}{I}_p - \underset{\sim}{D})\Big); \tag{62}$$

$$\underset{\sim}{E}\underset{\sim}{G}^{-1} = \underset{\sim}{D} + (\underset{\sim}{I}_p - \underset{\sim}{D})\underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\underset{\sim}{D}; \tag{63}$$

$$\underset{\sim}{E}\underset{\sim}{G}^{-1}\underset{\sim}{x} = [(1-u_1)x_1 + u_1\underset{\sim}{z}_1^T\hat{\underset{\sim}{b}},\ldots,(1-u_p)x_p + u_p\underset{\sim}{z}_p^T\hat{\underset{\sim}{b}}]^T, \tag{64}$$

where $\hat{\underset{\sim}{b}} = (\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{x})$. Then,

$$\underset{\sim}{G}^{-1} - \underset{\sim}{G}^{-1}\underset{\sim}{E}\underset{\sim}{G}^{-1} = a^{-1}[\underset{\sim}{D} - \underset{\sim}{D}\underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\underset{\sim}{D}]. \tag{65}$$

Hence,

$$\begin{aligned} &\underset{\sim}{x}^T(\underset{\sim}{G}^{-1} - \underset{\sim}{G}^{-1}\underset{\sim}{E}\underset{\sim}{G}^{-1})\underset{\sim}{x} \\ &= a^{-1}\left[\sum_{i=1}^{p}(1-u_i)x_i^2 - \Big(\sum_{i=1}^{p}(1-u_i)x_i\underset{\sim}{z}_i\Big)^T(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\Big(\sum_{i=1}^{p}(1-u_i)x_i\underset{\sim}{z}_i\Big)\right] \\ &= Q_a(\underset{\sim}{x}) \text{ (say).} \end{aligned} \tag{66}$$

Combining (59), (60) and (66), the joint pdf of $\underset{\sim}{X}$ and A is given by

$$f(\underset{\sim}{x}, a) \propto |\underset{\sim}{E}|^{\frac{1}{2}} a^{-\frac{1}{2}(p-r)} exp\Big[-\tfrac{1}{2}Q_a(\underset{\sim}{x})\Big]. \tag{67}$$

Writing $\underset{\sim}{F} = \underset{\sim}{G}^{-1} + a^{-1}\underset{\sim}{I}_p$, and using Exercise 2.4, p. 32 of Rao (1973), one gets

$$\begin{aligned} |\underset{\sim}{E}^{-1}| &= \begin{vmatrix} \underset{\sim}{F} & \underset{\sim}{Z} \\ \underset{\sim}{Z}^T & a(\underset{\sim}{Z}^T\underset{\sim}{Z}) \end{vmatrix} \div |a(\underset{\sim}{Z}^T\underset{\sim}{Z})| \\ &= |\underset{\sim}{F}|\,|a(\underset{\sim}{Z}^T\underset{\sim}{Z}) - \underset{\sim}{Z}^T\underset{\sim}{F}^{-1}\underset{\sim}{Z}| \div |a(\underset{\sim}{Z}^T\underset{\sim}{Z})| \\ &\propto a^{-p}\Big\{\prod_1^p(a + V_i)\Big\}|\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z}|. \end{aligned} \tag{68}$$

It is clear from (67) and (68) that

$$f(a|\underset{\sim}{x}) \propto a^{\frac{1}{2}r}\left\{\prod_{1}^{p}(a+V_i)^{-\frac{1}{2}}\right\}\left|\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z}\right|^{-\frac{1}{2}}exp\left[-\tfrac{1}{2}Q_a(\underset{\sim}{x})\right]. \quad (69)$$

Now writing $U_i = V_i/(A + V_i)$ $(i = 1,\ldots,p)$, using (69), and the iterated formulas for conditional expectations and variances, one gets

$$E[\Theta_i|\underset{\sim}{x}] = E[E(\Theta_i|\underset{\sim}{x}, A)|\underset{\sim}{x}] = E[(1-U_i)x_i + U_i\underset{\sim}{z}_i^T\underset{\sim}{b}|\underset{\sim}{x}]; \quad (70)$$

$$\begin{aligned} V[\Theta_i|\underset{\sim}{x}] &= V[E(\Theta_i|\underset{\sim}{x}, A)|\underset{\sim}{x}] + E[V(\Theta_i|\underset{\sim}{x}, A)|\underset{\sim}{x}] \\ &= V[(1-U_i)x_i + U_i\underset{\sim}{z}_i^T\underset{\sim}{b}|\underset{\sim}{x}] + E[AU_i + AU_i^2\underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i|\underset{\sim}{x}] \\ &= V[U_i(x_i - \underset{\sim}{z}_i^T\underset{\sim}{b})|\underset{\sim}{x}] + E[V_i(1-U_i) + V_iU_i(1-U_i)\underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i|\underset{\sim}{x}]; \quad (71) \end{aligned}$$

$$\begin{aligned} &Cov[\Theta_i, \Theta_j|x_i] \\ &= Cov[U_i(x_i - \underset{\sim}{z}_i^T\underset{\sim}{b}), U_j(x_j - \underset{\sim}{z}_j^T\underset{\sim}{b})|\underset{\sim}{x}] + E[AU_iU_j\underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z})^{-1}\underset{\sim}{z}_j|\underset{\sim}{x}]. \quad (72) \end{aligned}$$

Morris (1983) provides approximations for $E(\Theta_i|\underset{\sim}{x})$ and $V(\Theta_i|\underset{\sim}{x})$, $i = 1,\ldots,p$. He estimates the parameter a from the marginal distribution of $X_1,\ldots,X_p$ by employing some non-Bayesian method, and substitutes this estimate in the expressions for $E[\Theta_i|\underset{\sim}{x}, a]$ and $V[\Theta_i|\underset{\sim}{x}, a]$ instead of finding posterior expectations and variances of functions involving A. Thus, using Morris's method, $E[\Theta_i|\underset{\sim}{x}]$ is approximated by $(1-\hat{u}_i)x_i + \hat{u}_i\underset{\sim}{z}_i^T\hat{\underset{\sim}{\hat{b}}} = x_i - \hat{u}_i(x_i - \underset{\sim}{z}_i^T\hat{\hat{b}})$, while $V(\Theta_i|\underset{\sim}{x})$ is approximated by $v_i(x_i - \underset{\sim}{z}_i^T\hat{\hat{b}})^2 + V_i(1-\hat{u}_i)[1 + \hat{u}_i\underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\hat{D}\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i]$, $i = 1,\ldots,p$. In the above $v_i = [2/(p-r-2)]\hat{u}_i^2(\tilde{V}+\hat{a}) \div (V_i+\hat{a})$, $i = 1,\ldots,p$, $\tilde{V} = \sum_{i=1}^{p} V_i(V_i+\hat{a})^{-1} \div \sum_{i=1}^{p}(V_i+\hat{a})^{-1}$, $\hat{D} = Diag(1-\hat{u}_1,\ldots,1-\hat{u}_p)$, and $\hat{\hat{b}}$ is obtained from $\hat{b}$ by substituting the estimator of a. The v_i's are purported to estimate $V(U_i|\underset{\sim}{x})$'s. It is not clear whether such an approximation can be justified very rigorously since $\hat{\hat{b}}$ also involves the $\hat{u}_i$'s and $\hat{u}_i$ is *not* distributed independently of the $x_i - \underset{\sim}{z}_i^T\hat{\hat{b}}$.

We examine now how formulas (70) and (71) work in estimating the batting averages of Ty Cobb during 1905-1928. Morris (1983) took a similar undertaking except that his major emphasis was to examine whether Ty Cobb was "ever a true .400 hitter". To make our results comparable to those of Morris (1983), we fit a quadratic to Ty Cobb's batting averages, that is we take $\underset{\sim}{b} = (b_1, b_2, b_3)^T$, $\underset{\sim}{z}_i = (1, i, i^2)^T$, $i = 1,\ldots,24$. In the average year 1 refers to 1905, and year 24 refers to 1928. We provide in Table 2 the actual batting averages

TABLE 2. The Actual Batting Averages of Ty Cobb (Y_i), the Number of Times He Was at Bat (n_i), the HB Estimates ($\hat{\theta}_{i,HB}$), the Corresponding S.D.'s ($s_{i,HB}$), Morris's Approximate Estimates ($\hat{\theta}_{i,M}$), and the Corresponding S.D.'s ($s_{i,M}$).

i	n_i	Y_i	$\hat{\theta}_{i,HB}$	$s_{i,HB}$	$[\hat{\theta}_{i,HB}-2s_{i,HB}, \hat{\theta}_{i,HB}+2s_{i,HB}]$	$\hat{\theta}_{i,M}$	$s_{i,M}$	$[\hat{\theta}_{i,M}-2s_{i,M}, \hat{\theta}_{i,M}+2s_{i,M}]$
1	150	.240	.298	.020	[.258, .338]	.303	.026	[.251, .355]
2	350	.320	.325	.015	[.295, .355]	.327	.018	[.293, .363]
3	605	.350	.344	.013	[.318, .370]	.345	.015	[.315, .375]
4	581	.324	.337	.014	[.309, .365]	.339	.015	[.309, .369]
5	573	.377	.366	.014	[.338, .394]	.366	.015	[.336, .396]
6	509	.385	.373	.014	[.345, .401]	.373	.015	[.343, .403]
7	591	.420	.393	.016	[.361, .425]	.393	.017	[.359, .427]
8	553	.410	.391	.015	[.361, .421]	.391	.015	[.361, .421]
9	428	.390	.384	.015	[.354, .414]	.385	.015	[.355, .415]
10	345	.368	.379	.015	[.349, .409]	.379	.016	[.347, .411]
11	563	.369	.379	.014	[.351, .407]	.380	.014	[.352, .408]
12	542	.371	.381	.014	[.353, .409]	.381	.015	[.351, .411]
13	588	.383	.386	.013	[.350, .412]	.386	.014	[.358, .414]
14	421	.382	.385	.015	[.355, .415]	.385	.015	[.355, .415]
15	497	.384	.385	.014	[.357, .413]	.385	.014	[.357, .413]
16	428	.334	.364	.016	[.332, .396]	.365	.018	[.329, .400]
17	507	.389	.383	.014	[.355, .411]	.383	.014	[.355, .411]
18	526	.401	.385	.015	[.355, .415]	.384	.015	[.354, .414]
19	556	.340	.355	.014	[.327, .383]	.356	.015	[.326, .386]
20	625	.338	.350	.014	[.322, .378]	.351	.014	[.323, .379]
21	415	.378	.362	.015	[.332, .392]	.361	.016	[.329, .393]
22	233	.339	.342	.016	[.310, .374]	.342	.018	[.306, .378]
23	490	.357	.342	.015	[.312, .372]	.342	.017	[.308, .376]
24	353	.323	.322	.015	[.290, .352]	.322	.019	[.284, .360]

(Y_i) of Ty Cobb, the number of times he was at bat (n_i), our estimated batting averages ($\hat{\theta}_{i,HB}$), the corresponding standard errors ($s_{i,HB}$), Morris's approximations ($\hat{\theta}_{i,M}$) for these batting averages, and the corresponding approximate standard errors ($s_{i,M}$). Following Morris, we took $V_i = (.367)(.633)/n_i$, $i = 1,\ldots,24$.

It follows from Table 2 that $\sum_{i=1}^{24}(\hat{\theta}_{i,HB} - Y_i)^2 = .007377$ and $\sum_{i=1}^{24}(Y_i - \hat{\theta}_{i,M})^2 = .008244$. Thus, Morris's approximations lead to about a 11.0% increase in the overall mean squared error. Also, the $s_{i,M}$'s though mostly very close to $s_{i,HB}$'s can lead upto a 30% increase. More important, our two standard deviation confidence intervals around the posterior means are usually mugh tighter than the corresponding ones given in Morris (1983). However, as mentioned earlier, Morris's EB procedure is much superior to a naive EB procedure, since the latter can seriously underestimate the actual standard errors. This is evidenced in our actual calculations which are not reported here. We should also point out that both $[\hat{\theta}_{i,HB} \pm 2s_{i,HB}]$'s and $[\hat{\theta}_{i,M} \pm 2s_{i,M}]$'s cover the true Y_i's 23 out of 24 times which is approximately 95.8%. Also, $[\hat{\theta}_{i,HB} \pm s_{i,HB}]$'s and $[\hat{\theta}_{i,M} \pm s_{i,M}]$'s cover the true Y_i's 17 out of 24 times which is approximately 70.8%. Thus a normal approximation to the posterior distribution is not totally out of the way.

One of Cobb's greatest claims to fame is that he has the highest lifetime batting average of any baseball player in the modern era. Ty Cobb's actual overall batting average in 1905-1928 is .367. Also, $\hat{\theta}_{HB} = \sum_{i=1}^{24} n_i\hat{\theta}_{i,HB}/\sum_{i=1}^{24} n_i = .366$ and $\hat{\theta}_M = \sum_{i=1}^{24} n_i\hat{\theta}_{i,M}/\sum_{i=1}^{24} n_i = .366$. This shows that both the HB and EB estimates of the overall batting average of Ty Cobb essentially match the reality.

It is instructive to look at the special case of *equal variances*, that is, when $V_1 = \ldots = V_p = V$. Then $u_1 = \ldots = u_p = V/(V+a) = u$ (say). In this case $\underset{\sim}{D} = (1-u)\underset{\sim}{I}_p$, $\underset{\sim}{Z}^T\underset{\sim}{D}\underset{\sim}{Z} = (1-u)\underset{\sim}{Z}^T\underset{\sim}{Z}$, $\hat{\hat{\underset{\sim}{b}}} = (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\underset{\sim}{x} = \hat{\underset{\sim}{b}}$, the usual least squares estimate of $\underset{\sim}{b}$. Moreover, $a + V = Vu^{-1}$ so that $a = V(1-u)/u$, $Q_a(\underset{\sim}{x}) = a^{-1}(1-u)SSE$, where $SSE = \sum_{i=1}^{p} x_i^2 - \left(\sum_{i=1}^{p} x_i\underset{\sim}{z}_i\right)^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\left(\sum_{i=1}^{p} x_i\underset{\sim}{z}_i\right)$, the usual error SS. Since $|da/du| = Vu^{-2}$, it follows from (69) that the conditional pdf of U given $\underset{\sim}{x}$ is

$$f(u|\underset{\sim}{x}) \propto \left((1-u)/u\right)^{\frac{1}{2}r} u^{\frac{1}{2}p-2}(1-u)^{-\frac{1}{2}r} exp\left(-\frac{u}{2V} SSE\right)$$

$$= u^{\frac{1}{2}(p-r-4)} exp\left(-\frac{u}{2V} SSE\right). \tag{73}$$

It follows from (70) and (71) that

$$E(\Theta_i|\underset{\sim}{x}) = x_i - E(U|\underset{\sim}{x})(x_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}}); \tag{74}$$

$$V(\Theta_i|\underset{\sim}{x}) = V(U|\underset{\sim}{x})(x_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2 + V - VE(U|\underset{\sim}{x})\left(1 - \underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i\right). \tag{75}$$

If one adopts Morris's approximations as in the second section, then one estimates $E(U|\underset{\sim}{x})$ by

$$\hat{\hat{U}} = \int_0^\infty u^{\frac{1}{2}(p-r-2)} exp\left(-\frac{u}{2V}SSE\right)du \Big/ \int_0^\infty u^{\frac{1}{2}(p-r-2)} exp\left(-\frac{u}{2V}SSE\right)du$$

$$= V(p-r-2)/SSE$$

and $E(U^2|\underset{\sim}{x})$ by

$$\int_0^\infty u^{\frac{1}{2}(p-r)} exp\left(-\frac{u}{2V}SSE\right)du \Big/ \int_0^\infty u^{\frac{1}{2}(p-r-4)} exp\left(-\frac{u}{2V}SSE\right)du$$

$$= V^2(p-r)(p-r-2)/(SSE)^2.$$

Accordingly, $V(U|x)$ is approximated by $2V^2(p-r-2) \div (SSE)^2 = [2/(p-r-2)]\hat{\hat{U}}^2$. These calculations suggest that $E(\Theta_i|\underset{\sim}{x})$ should be approximated by $x_i - \hat{\hat{U}}(x_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})$ and $V(\Theta_i|\underset{\sim}{x})$ should be approximated by

$$s_{i,G}^2 = \left(2/(p-r-2)\right)\hat{\hat{U}}^2(x_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2 + V\left[1 - \hat{\hat{U}}\left(1 - \underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i\right)\right]. \tag{76}$$

The expression $s_{i,G}^2$ *does not* agree with the expression s_i^2 given in (4.1) of Morris (1983) (with the obvious changes in his notations). It seems to us that Morris's (4.1) uses his (1.17) which involves a slight oversight. We shall discuss this point now.

Morris (1983) starts with an EB approach, where he assumes conditions (A) and (B) with $V_1 = \ldots = V_p = V$ (say). With this formula for known $\underset{\sim}{b}$ and a, the Bayes estimator of $\underset{\sim}{\theta}$ is given by

$$\hat{\underset{\sim}{\theta}}_B = (1-u)\underset{\sim}{X} + u\underset{\sim}{Z}\underset{\sim}{b},\ u = V/(V+a). \tag{77}$$

If $\underset{\sim}{b}$ and u are unknown, Morris (1983) estimates them by $\hat{\underset{\sim}{b}}$ and $\hat{a}$ respectively, where $\hat{\underset{\sim}{b}} = (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T\underset{\sim}{X}$, the least squares estimator of $\underset{\sim}{b}$ and $\hat{u} = (p-r-2)V/SSE$, $SSE = \sum_{i=1}^{p}(X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2$, the error SS. Note that $\hat{u}$ is the UMVUE of u since marginally $SSE \sim Vu^{-1}\chi^2_{p-r}$.

Morris (1983) proposes the EB estimator $\hat{\underset{\sim}{\theta}}_{EB} = (\hat{\theta}_{1,EB},\ldots,\hat{\theta}_{p,EB})^T$ of $\underset{\sim}{\theta}$, where

$$\hat{\theta}_{i,EB} = (1-\hat{u})X_i + \hat{u}\underset{\sim}{z}_i^T\hat{\underset{\sim}{b}} \quad (i = 1,\ldots,p). \tag{78}$$

Then,

$$\begin{aligned} E(\hat{\theta}_{i,EB} - \theta_i)^2 &= E(\theta_i - \hat{\theta}_{i,B})^2 + E(\hat{\theta}_{i,B} - \hat{\theta}_{i,EB})^2 \\ &= V(1-u) + E\left[(u-\hat{u})(X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}}) + u\underset{\sim}{z}_i^T(\hat{\underset{\sim}{b}} - \underset{\sim}{b})\right]^2. \end{aligned} \tag{79}$$

Using the marginal independence of $X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}}$ and $\underset{\sim}{z}_i^T\hat{\underset{\sim}{b}}$, and noting that $V(\hat{\underset{\sim}{b}}) = Vu^{-1}(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}$, it follows from (79) that

$$E(\hat{\theta}_{i,EB} - \theta_i)^2 = V(1-u) + E\left[(u-\hat{u})^2(X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2\right] + Vu\underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i. \tag{80}$$

Since $(\hat{\underset{\sim}{b}}, SSE)$ is complete sufficient for $(\underset{\sim}{b}, u)$ and $(X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2/SSE$ is ancillary, they are independently distributed by Basu's (1955) theorem. Now using $E(X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2 = Vu^{-1}\left(1 - \underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i\right)$ and $SSE \sim Vu^{-1}\chi^2_{p-r}$, it follows on simplification that

$$\begin{aligned} &E\left[(u-\hat{u})^2(X_i - \underset{\sim}{z}_i^T\hat{\underset{\sim}{b}})^2\right] \\ &= Vu\left(1 - \underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i\right) - Vu\frac{p-r-2}{p-r}\left(1 - \underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i\right). \end{aligned} \tag{81}$$

Combining (80) and (81), it follows that

$$E(\hat{\theta}_{i,EB} - \theta_i)^2 = V - V\frac{p-r-2}{p-r}\left(1 - \underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i\right)u. \tag{82}$$

In Morris's (1.17), $\underset{\sim}{z}_i^T(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{z}_i = r/p$ for every i which does not seem to be the case.

Acknowledgements

I am indebted to the referee and Mr. Gauri Sankar Datta for their constructive comments which led to a substantial improvement over the original version of the paper. I am also indebted to Dr. Li Chu Lee and Mr. Gauri Sankar Datta for carrying out the numerical computations, to Mr. Gwowen Shieh for catching a number of typos, and to Professor Carl Morris and Professor Ron Christensen for supplying me with the full batting average data of Ty Cobb.

References

Basu, D. (1955): On statistics independent of a complete sufficient statistic, *Sankhyā* 15, 377-380.

Berger, J. (1975): Minimax estimation of location vectors for a wide class of densities, *Ann. Statist.* 3, 1318-1328.

Berger, J. (1985): *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition, Springer-Verlag, New York.

Bock, M. E. (1975): Minimax estimators of the mean of a multivariate normal distribution, *Ann. Statist.* 3, 209-218.

Casella, G. (1985): An introduction to empirical Bayes data analysis, *American Statistician* 39, 83-87.

Deely, J. J., and Lindley, D. V. (1981): Bayes empirical Bayes, *J. Amer. Statist. Assoc.* 76, 833-841.

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, 130-139.

Efron, B., and Morris, C. (1973): Stein's estimation rule and its competitors – an empirical Bayes approach, *J. Amer. Statist. Assoc.* 68, 117-130.

Efron, B., and Morris, C. (1975): Data analysis using Stein's estimator and its generalizations, *J. Amer. Statist. Assoc.* 70, 311-319.

Faith, R. E. (1978): Minimax Bayes and point estimations of a multivariate normal mean, *J. Mult. Anal.* 8, 372-379.

Ghosh, M., and Lahiri, P. (1987): Robust empirical Bayes estimation of means from stratified samples, *J. Amer. Statist. Assoc.* 82, 1153-1162.

Ghosh, M., and Lahiri, P. (1988): A hierarchical Bayes approach to small area estimation with auxiliary information, presented at the Joint Indo. U.S. Workshop on Bayesian Analysis in Statistics and Econometrics held in Bangalore, India, December 19-23, 1988.

Ghosh, M., and Meeden, G. (1986): Empirical Bayes estimation in finite population sampling, *J. Amer. Statist. Assoc.* 81, 1058-1062.

Good, I.J. (1965): *The Estimation of Probabilities, An Essay on Modern Bayesian Methods*, M.I.T. Press, Cambridge, Massachusetts.

James, W., and Stein, C. (1961): Estimation with quadratic loss, *Proc. 4th Berk. Symp. Math. Statist. Prob.* 1, 361-380, Univ. California Press, Berkeley.

Lehmann, E. L. (1983): *Theory of Point Estimation*, Wiley, New York.

Lindley, D. (1962): Discussion of Professor Stein's paper, *J. R. Statist. Soc. B* 24, 265-296.

Lindley, D., and Smith, A. F. M. (1972): Bayes estimates for the linear model, *J. R. Statist. Soc. B* 34, 1-41.

Morris, C. (1981): Parametric empirical Bayes confidence intervals, *Scientific Inference, Data Analysis, and Robustness*, eds. G.E.P. Box, T. Leonard and C.F. Jeff Wu, Academic Press, 25-50.

Morris, C. (1983): Parametric empirical Bayes inference and applications, *J. Amer. Statist. Assoc.* 78, 47-65.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications*, 2nd Edition, Wiley, New York.

Reinsel, G. (1985): Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model, *J. Amer. Statist. Assoc.* 80, 642-651.

Robbins, H. (1951): Asymptotically subminimax solutions of compound statistical decision problems, *Proc. 2nd Berk. Symp. Math. Statist. Prob.* V1, University of California Press, Berkeley, 131-184.

Robbins, H. (1955): An empirical Bayes approach to statistics, *Proc. 3rd Berk. Symp. Math. Statist. Prob.* V1, 157-164.

Strawderman, W. E. (1971): Proper Bayes minimax estimators of the multivariate normal mean, *Ann. Math. Statist.* 42, 385-388.