Restricted Risk Bayes Estimation for the Mean of the Multivariate Normal Distribution

by

Shun-Yu Chen

Technical Report #83-33

Department of Statistics
Purdue University

August 1983

Research supported by the National Science Foundation under Grant MCS-8101670 and the David Ross Grant XR-0671.

Restricted Risk Bayes Estimation
for the Mean of the Multivariate
Normal Distribution

by

Shun-Yu Chen
Purdue University

Abstract

Let $X = (X_1,\ldots,X_p)^t$ to be an observation from a p-variate normal distribution with unknown mean vector $\theta = (\theta_1,\ldots,\theta_p)^t$ and known covariance matrix Σ. It is desired to estimate θ under the quadratic loss $L(\theta,\delta) = (\theta-\delta)^t\, Q\, (\theta-\delta)$. Suppose prior beliefs concerning θ can be approximately modeled by a conjugate prior distribution π which is $N_p(\mu,A)$, where μ, A are known. We find estimators of θ which have small Bayes risk and which also satisfy the constraint $R(\theta,\delta) \le tr(Q\Sigma) + c$, $R(\theta,\delta)$ being the frequentist risk of δ. Such estimators are good from both the frequentist and Bayesian perspectives.

Research supported by the National Science Foundation under Grant MCS-8101670 and David Ross Grant XR-0671.

AMS 1980 Subject Classification: 62C10, 62F10.

Key Words: Risk; Bayes risk; Relative saving risk.

1. Introduction

Let $X = (X_1,\ldots,X_p)^t$ be a p-variate normal distribution with known covariance matrix Σ and unknown mean vector $\theta = (\theta_1,\ldots,\theta_p)$. It is desired to estimate θ under quadratic loss

$$L(\theta,\delta) = (\theta-\delta)^t\, Q\, (\theta-\delta),$$

where Q is a given positive definite matrix. An estimator will be evaluated by its risk function

$$R(\theta,\delta) = E_\theta[L(\theta,\delta(x))].$$

When one has complete knowledge of the prior distribution π of θ, one would use a Bayes estimator δ^π which is an estimator minimizing the Bayes risk, i.e., $r(\pi,\delta^\pi) = \min_\delta r(\pi,\delta)$, where

$$r(\pi,\delta) = E^\pi[R(\theta,\delta)].$$

Unfortunately, the determination of π is often very inexact. In a finite amount of time, only subjective approximations to π can be constructed and the Bayes estimator can be sensitive to uncertain parts of the prior specification. (See Berger (1980c, 1983) for references.)

On the other hand, the minimax principle tries to protect against the worst possible state of nature. Although a minimax estimator is the "most robuts" Bayesian decision rule (in the sense of minimizing $\sup_\pi r(\pi,\delta)$), it often behaves poorly when looked at from the viewpoint of average risk, and can often be inadmissible (c.f. Berger (1980a, c, 1983)).

The situation occurring in practice frequently lies between these two extreme cases. Often, we have some idea about the prior distribution of θ but we are not willing to entirely rely on any specific prior. Hence it is important to consider the robustness (with respect to the specification of π) of the estimator selected. Thus, a reasonable goal is to develop a procedure which can incorporate prior knowledge, but is safe with respect to error in the specification of prior knowledge.

One way to do this problem is as follows: First, we restrict the risk; for example

$$(1.1) \qquad R(\theta,\delta) - R(\theta,\delta^0) \leq c$$

for all $\theta \in R^p$, where $\delta^0(x) = x$ is the MLE and minimax estimator of θ and c is a given nonnegative constant. (Note that $R(\theta,\delta^0) = tr(Q\ddagger)$.) And then, in the class of estimators satisfying the restricted condition (1.1) find an estimator which has minimum Bayes risk. Thus, we impose a constraint on the deviation of the risk of our estimator from the minimax estimator and then, subject to this constraint, we minimize the Bayes risk.

This problem was first proposed by Hodges and Lehmann (1952) and has been considered for various situations in Efron and Morris (1971), Shapiro (1972, 1975), Bickel (1980), Marazzi (1982), and Berger (1982a). It is known that exact mathematical solution of this problem is very messy and, even in the case of a spherically symmetric normal distribution, numerical solution is very difficult (c.f. Berger (1982a), Marazzi (1982)). For this reason, we will consider various simplifications of the problem.

By using the identity (essentially a variant of Stein's "unbiased estimator of risk")

$$R(\theta,\delta) - R(\theta,\delta^0) = E_\theta[\mathcal{D}\gamma(x)],$$

where $\gamma(x) = \ddagger^{-1}(\delta(x)-x)$ and $\mathcal{D}\gamma(x)$ is an expression involving partial derivatives of $\gamma_i(x)$, the condition (1.1) will clearly be satisfied if

$$(1.2) \qquad \mathcal{D}\gamma(x) \leq c.$$

Hence we can formulate the following approximate restricted risk Bayes problem:

> Select the estimator δ which minimizes $r(\pi,\delta)$ subject to (1.2).

When $\ddagger Q \ddagger$ is a diagonal matrix (always achievable by a linear transformation), Chen (1983) generalizes an idea of Berger (1982a) to show that the optimal estimator in the approximate restricted risk Bayes problem must be a smooth blending of the Bayes estimator δ^π and estimator arising from the differential equation $\mathcal{D}\gamma(x) = c$. Unfortunately, in the nonsymmetric multi-dimensional case it is generally impossible to solve the differential equation $\mathcal{D}\gamma(x) = c$ in closed form. Numerical solution would be a possibility, but in the interest of providing reasonably accessible and understandable estimators we further specialize the problem by considering a particular class of estimators for which solution of the differential equality is possible. We also restrict consideration to the case where the "approximate" prior that is specified is $N_p(\mu,A)$, μ being the prior mean vector and A the positive definite prior covariance matrix.

A natural class of estimators to consider is the class of "compromise" estimators

$$(1.3) \qquad \delta_\rho(x) = \rho(r)\ \delta^\pi(x) + (1-\rho(r))\ \delta^0(x)$$
$$= x - \rho(r)\ \Sigma\ (\Sigma+A)^{-1}\ (x-\mu),$$

where $r = (x-\mu)^t\ (\Sigma+A)^{-1}\ (x-\mu)$, and ρ is a continuous and piecewise differentiable function taking values in [0,1]. Small r support the validity of the prior information (if $x \sim N_p(\theta,\Sigma)$ and $\theta \sim N_p(\mu,A)$, then $r \sim \chi^2_p$), and hence suggest use of $\delta^\pi(x) = x - \Sigma(\Sigma+A)^{-1}(x-\mu)$ (so $\rho(r)$ should then be near one), while large r suggests that prior information is implausible and that $\delta^0(x) = x$ should be used (i.e., $\rho(r)$ should then be near zero). A "robust Bayes" estimator in this class was proposed by Berger (1980b). We will find the optimal choice of ρ in Section 2 (for the approximate restricted risk problem), a choice which cah offer substantial improvement over the robust Bayes estimator of Berger (1980b).

In Berger (1982b) a different type of estimator was proposed, one that incorporated the prior information μ and A and yet was guaranteed to be minimax. The estimator was based on a "decomposition to subproblems" technique proposed in Bhattacharya (1966). In Section 3, we modify this estimator by basing it on the optimal "compromise" estimators of Section 2. The resulting estimator appears to have substantially better performance than the optimal compromise estimator, in the sense of having smaller Bayes risk, while still satisfying the constraint (1.1).

If minimaxity is desired (i.e. $c = 0$ in (1.1)), the estimators discussed above have certain limitations (see Section 4) in very nonsymmetric situation. A new estimator, called a "weighted" minimax estimator is proposed in Section 4, and is shown to offer substantial improvement in Bayes risk (while preserving minimaxity) in certain cases.

Section 5 explicitly discusses the comparative performance of the "compromise", "subproblem", and "weighted minimax" estimators for the restricted risk Bayes problem. The overall conclusion is that last two of these estimators are superior, and guidelines are presented concerning the use of each of them.

Instead of using $\gamma(\pi,\delta)$ to measure the effectiveness of an estimator δ, it is convenient to use the linearly transformed relative savings risk of Efron and Morris (1972), defined by

$$\mathrm{RSR}(\pi,\delta) = \frac{\gamma(\pi,\delta) - \gamma(\pi,\delta^{\pi})}{\gamma(\pi,\delta^{0}) - \gamma(\pi,\delta^{\pi})} \quad .$$

This measures the proportion of the potential Bayesian improvement over δ^0 which is attained by the estimator δ. Small RSR is desirable from a Bayesian viewpoint.

2. The Optimal Compromise Estimator

Let X be a $N_p(\theta,\Sigma)$ and suppose θ has (approximately) the conjugate prior distribution π which is $N_p(\mu,A)$, where μ and A are known. Let $d_1 \geq d_2 \geq \cdots \geq d_p > 0$ be the characteristic roots of the matrix $D = (\Sigma+A)^{-1/2}\,\Sigma Q \Sigma(\Sigma+A)^{-1/2}$ and let $\tau = \sum_{i=1}^{p} d_i/d_1$. We will consider the class of compromise estimators δ_ρ defined in (1.3) with $E^r(\rho^2(r)) < \infty$, and $E^r(|r\rho'(r)|) < \infty$.

First, we need the following Stein identity. (c.f. Stein (1981), Hudson (1978), Berger (1982), and Chen (1983)). Let

$$(2.2) \qquad \delta(x) = x + \Sigma\gamma(x),$$

where γ_i is a continuous and piecewise differentiable function with $E_\theta^X[\gamma_i^2(x)] < \infty$ and $E_\theta^X[\gamma_i^{(i)}(x)| < \infty$, (where $\gamma_i^{(i)}(x) = \frac{\partial}{\partial x_i}\gamma_i(x)$ and so on). Then

$$(2.3) \qquad R(\theta,\delta) - R(\theta,\delta^0) = E_\theta[\mathcal{L}\gamma(x)],$$

where

$$(2.4) \qquad \mathcal{L}\gamma(x) = \gamma^t(x)\Sigma Q\Sigma\gamma(x) + 2\,\mathrm{tr}(V\Sigma Q\Sigma)$$

and $V = (V_{ij})$ with $V_{ij} = \frac{\partial\gamma_i}{\partial x_j}$. (here trA = trace of A).

Using this identity, we will give a sufficient condition for the compromise estimator δ_ρ to satisfy the constraint (1.1).

<u>Lemma 1</u>. If $\rho^2(r) - 4\rho'(r) \geq 0$ for all $r > 0$, then $R(\theta,\delta_\rho) - R(\theta,\delta^0) \leq C$ for all θ, provided that $\mathcal{D}\rho(r) \leq c/d_1$ for all $r > 0$, where

$$(2.5) \qquad \mathcal{D}\rho(r) = [\rho^2(r) - 4\rho'(r)]\, r - 2\tau\rho(r).$$

<u>Proof</u>. Given in the Appendix. ||

Let D_c be the class of compromise estimators satisfying the restricted condition $\mathcal{D}\rho(r) \leq c/d_1$ for all $r > 0$, where $\mathcal{D}\rho(r)$ is given in (2.5). Chen (1983) showed that the optimal estimator, δ_{ρ_c}, in the class D_c is

$$(2.6) \qquad \delta_{\rho_c}(x) = x - \min(1, \rho_{c,0}(r))\, \Sigma\, (\Sigma+A)^{-1}(x-\mu),$$

where

$$(2.7) \qquad \rho_{c,0}(r) = \begin{cases} \dfrac{2(\tau-2)^+}{r} & \text{if } c = 0 \text{ and } \tau > 2 \\[2ex] \dfrac{c}{2d_1 t}\,\dfrac{K_{v+1}(t)}{K_v(t)} & \text{if } c > 0 \text{ and } \tau \geq 2 \\[2ex] \dfrac{c}{2d_1 t}\,\dfrac{K_{v-1}(t)}{K_v(t)} & \text{if } c > 0 \text{ and } 1 \leq \tau < 2, \end{cases}$$

$v = |\tau-2|/2$, $t = \frac{1}{2}\sqrt{cr/d_1}$ and K_v is the second kind of modified Bessel function of order v.

The RSR of δ_{ρ_c} is given in the following theorem.

<u>Theorem 1</u>. For the estimator δ_{ρ_c} defined as in (2.6),

$$\mathrm{RSR}(\pi,\delta_{\rho_c}) = 1 - \frac{2b}{p} f_p(b) + \frac{c}{pd_1} (1-\psi_p(b)) - \psi_p(b)$$

$$- 2[1 - \sum_{i=1}^{p} \frac{d_i}{pd_1}] \int_b^\infty \rho_{c,0}(r) f_p(r) dr,$$

where b is the unique positive solution of the equation $\rho_{c,0}(r) = 1$, and f_p and ψ_p are the density function and cumulative distribution function of the chi-square distribution with p degrees of freedom, respectively.

<u>Proof</u>. Given in the Appendix. ||

When c = 0, the RSR of δ_{ρ_0} can be expressed explicitly as follows.

<u>Corollary 1</u>. When c = 0 and $\tau > 2$,

$$\mathrm{RSR}(\pi,\delta_{\rho_0}) = \frac{1}{p} \Big\{ [p-4(\tau-2) + \frac{4(\tau-2)^2}{p-2}][1-\psi_p(2(\tau-2))]$$

$$+ 4(\tau-2)[1 - \frac{2(\tau-2)}{p-2}] f_p(2(\tau-2)) \Big\} .$$

When $\tau = p$ (i.e., $(\Sigma+A)^{-1/2}(\Sigma Q \Sigma)(\Sigma+A)^{-1/2} = I_p$), then

$$\mathrm{RSR}(\pi,\delta_{\rho_0}) = 1 - \psi_p(2(p-2)) - \frac{4(p-2)}{p} f_p(2(p-2)).$$

Proof. It follows easily from Theorem 1. ||

Remark 1. $T^0_{p,p} \le RSR(\pi,\delta_{\rho_0}) \le 1 - \frac{\tau-2}{p-2}(1-T^0_{p,p})$,

where

$$T^0_{p,p} = \frac{1}{p}\int_{2(p-2)}^{\infty} r\left(1-\frac{2(p-2)}{r}\right)^2 f_p(r)dr.$$

(see Table 4 for values of $T^0_{p,p}$.)

Remark 2. For fixed p, $RSR(\pi,\delta_{\rho_0})$ is a decreasing function of τ and has minimum value $T^0_{p,p}$ when $\tau = p$.

3. The Modified Subproblem Estimator

In this section, we use the decomposition to subproblems approach of Bhattacharya (1966) and Berger (1979) to develop an alternative estimator, $\delta^{MB,c}$, satisfying the constraint (1.1) and yet performance better than δ_{ρ_c}. For simplicity of notation in the remainder of the paper, we will assume that Q, $\ddagger$, and A are diagonal matrix, with diagonal elements q_i, σ_i^2, and a_i, respectively. (For similar results in the nondiagonal case see Chen (1983).)

Suppose (without loss of generality) that x_i are index so that $d_1 \ge d_2 \ge \dots \ge d_p$, where $d_i = q_i\sigma_i^4/(\sigma_i^2+a_i)$ and let $d_{p+1} = 0$. Then define the i-th component of $\delta^{MB,c}$ as

$$\delta_i^{MB,c}(x) = x_i - \frac{\sigma_i^2}{\sigma_i^2+a_i}(x_i-\mu_i)\sum_{j=i}^{p}\frac{d_j-d_{j+1}}{d_i}\min(1,\rho_{c,0}^{(j)}(r_j)),$$

where

$$\text{(3.1)} \qquad \rho_{c,0}^{(j)}(r_j) = \begin{cases} \dfrac{2(j-2)^+}{r_j} & \text{when } c = 0 \\[2ex] \dfrac{c}{2d_1 t_j} \dfrac{K_{v_j+1}(t_j)}{K_{v_j}(t_j)} & \text{when } c > 0 \text{ and } j = 1 \\[2ex] \dfrac{c}{2d_1 t_j} \dfrac{K_{v_j-1}(t_j)}{K_{v_j}(t_j)} & \text{when } c > 0 \text{ and } j \geq 2 \end{cases}$$

$$r_j = \sum_{i=1}^{j} \frac{(x_i-\mu_i)^2}{\sigma_i^2+a_i} \;, \quad t_j = \frac{1}{2}\sqrt{\frac{cr_j}{d_1}} \;, \text{ and } v_j = \frac{1}{2}\,|j-2| .$$

Note that this estimator is indeed based on the optimal compromise estimator found in Section 2. Motivation for estimators of this form can be found in Berger (1982b). Note, at least, that if all $\rho_{c,0}^{(j)}(r_j) \geq 1$, then $\delta^{MB,c}$ is the conjugate prior Bayes rule.

Using Berger's (1979) decomposition theorem, it is easy to show that $R(\theta,\delta^{MB,c}) - R(\theta,\delta^0) \leq c$ for all θ. The RSR for $\delta^{MB,c}$ is given in the following theorem.

Theorem 2.

$$RSR(\pi,\delta^{MB,c}) = \frac{1}{\sum_{i=1}^{p} d_i} \sum_{i=1}^{p} \sum_{j=i}^{p} \sum_{k=i}^{p} \frac{(d_j-d_{j+1})(d_k-d_{k+1})}{d_i} \, T_{j,k}^{c*} \,,$$

where

$$T^{c*}_{j,k} = E^{Z}\{z_1^2[1-\min(1, \rho^{(j)}_{c,0}\ (r_j))][1-\min(1,\rho^{(k)}_{c,0}(r_k))]\} ,$$

$z_1,\ldots,z_p$ are i.i.d. $N(0,1)$, $r_j = \sum_{n=1}^{j} z_n^2$, $c* = c/d_1$, and $\rho^{(j)}_{c,0}\ (r_j)$ is defined in (3.1).

Proof. Given in the Appendix. ||

For $1 \leq j, K \leq 10$, the values of $T^{c*}_{j,k}$ for some c* are given in Table 4.

Remark 1. When c = 0, $\delta^{MB,0}$ is the same as δ^{MB} defined in Berger (1982b).

Remark 2. All numerical results indicate that the RSR of $\delta^{MB,c}$ is smaller than the RSR of δ_{ρ_c} (c.f. Tables 1, 2, and 3), but we were able to prove this analytically.

Remark 3. In Dey and Berger (1983) (See also Efron and Morris (1973b)) the possibility of estimating, separately, groups of coordinates was considered. Dey and Berger (1983) discussed a somewhat inferior version of the estimator $\delta^{MB,0}$, namely $\delta^{MB*} = (\delta_1^{MB*},\ldots,\delta_p^{MB*})$, where

$$\delta_i^{MB*}(x) = x_i - \frac{\sigma_i^2}{\sigma_i^2 + a_i}(x_i-\mu_i)\sum_{j=i}^{p}\frac{d_j-d_{j+1}}{d_k}\ \frac{(j-2)^+}{\sum_{n=1}^{j}\frac{(x_n-\mu_n)^2}{\sigma_n^2+a_n}} ,$$

showing that the "combined" estimator is always better than the "separate" estimator (in the sense of small RSR), where the "combined" estimator is δ^{MB*} for the entire parameter vector and the "separate" estimator is δ^{MB*} for two groups of coordinates separately, one group consisting of coordinates with larger d_i and the other of coordinates with small d_i. When considering $\delta^{MB,0}$, instead of δ^{MB*}, the "combined" estimator can still be shown to be better than the "separate" estimator. For details see Chen (1983).

4. The Weighted Minimax Estimator

When d_1 (and maybe d_2) is much larger than the other d_i (i.e., $\tau = \sum_{i=1}^{p} d_i/d_1$ is less than or only slightly larger than two), then the RSR for the minimax estimators δ_{ρ_0} and $\delta^{MB,0}$ are large. In the following, we will propose a new minimax estimator, $\delta^{W,1}$, which has smaller RSR than the previous minimax estimators for this situation.

Let

$$(4.1) \qquad W = \frac{1}{p-2}\,[(I + (p-2)\,y_0\,D^{-1})^{1/2} - I],$$

where y_0 is the positive solution of the equation

$$(4.2) \qquad \frac{1}{p-2} \sum_{i=1}^{p} d_i \left[\sqrt{1 + \frac{(p-2)y}{d_i}} - 1\right] = y.$$

(Lemma 3, given in the Appendix, shows that equation (4.2) has a unique positive solution.) Then, define

$$\delta^{W,1}(x) = x - \min\left(1, \frac{2(p-2)^+}{r}\right) W \ddagger (\ddagger + A)^{-1}(x-\mu),$$

where W is given in (4.1) and $r = (x-\mu)^t(\ddagger+A)^{-1}(x-\mu)$.

As partial motivation for considering such an estimator, note that the optimal minimax estimator in the class D_c (see Section 2) of compromise estimators is $\delta_{\rho_0}(x) = x - \min\left(1, \frac{2(\tau-2)^+}{r}\right) \ddagger (\ddagger+A)^{-1}(x-\mu)$, which has the somewhat unappealing property of giving the same weight to each component of $\ddagger(\ddagger+A)^{-1}(x-\mu)$. An obvious modification is to give different weights to different components (usually, the bigger d_i, the less the weight). This will be seen to work quite well.

<u>Theorem 3</u>. $\delta^{W,1}$ is a minimax estimator and

$$RSR(\pi, \delta^{W,1}) = 1 - \frac{tr(WD)}{trD}(1 - T^0_{p,p}).$$

<u>Proof</u>. Given in the Appendix. ||

<u>Remark 1</u>. It can be shown that $\frac{tr(WD)}{trD} \geq \frac{\tau-2}{p-2}$, so that

$$RSR(\pi, \delta^{W,1}) \leq 1 - \frac{\tau-2}{p-2}(1 - T^0_{p,p}).$$

5. Comparisons and Conclusions

Analytic comparisons among δ_{ρ_c}, $\delta^{MB,c}$ and $\delta^{W,1}$ are difficult. Indeed the following result was the only explicit analytic result obtained.

Theorem 4. When (i) $d_1 > d_2 = \ldots = d_p$ or (ii) $d_1 = d_2 > d_3 = \ldots = d_p$, then

$$RSR(\pi, \delta^{W,1}) \leq RSR(\pi, \delta^{MB,0}) .$$

Proof. Given in the Appendix. ||

Of course, Theorem 1 (and Corollary 1), Theorem 2, and Theorem 3, together with Table 4, allow explicit computation of the RSR of any of the three estimators and hence selection of the best one for the given situation. To give a general feeling for the performance of the estimators, however, we give three examples of application, two in which $c = 0$ (i.e., minimaxity is desired) so that δ_{ρ_0}, $\delta^{MB,0}$, and $\delta^{W,1}$ are compared, and one in which $c > 0$ so that only δ_{ρ_c} and $\delta^{MB,c}$ are compared.

Example 1. (Randomized Block Design) Assume that

$$Y_{ij} = \alpha_i + \beta_j + e_{ij},\ i = 1,\ldots,p,\ j = 1,\ldots,n,$$

where the β's are i.i.d. $N(0,\tau^2)$ and the e_{ij}'s are i.i.d. $N_p(0,\sigma^2)$, independent of the β's. Also assume that σ^2 and τ^2 are known. Under the sum of squares error loss function, we want to estimate $\alpha = (\alpha_1,\ldots,\alpha_p)^t$. Letting $Y_j = (Y_{1j},\ldots,Y_{pj})^t$, $j = 1,\ldots,n$, it is clear that the Y_j are i.i.d. $N_p(\alpha,\Sigma)$, where

$$\Sigma = \sigma^2 I_p + \tau^2 1_p 1_p^t ,$$

and $1_p = (1,1,\ldots,1)^t$. Therefore, the UMVU estimator of α is

$$\overline{Y} = n^{-1} \sum_{j=1}^{n} Y_j ,$$

which is $N_p(\alpha, n^{-1}\Sigma)$. Suppose α has (approximately) the conjugate prior distribution $N_p(0,aI)$. First, we will transform to the diagonal case. Let H be a $(p\times p)$ orthogonal matrix with first row $\sqrt{1/p}(1,1,\ldots,1)$, and let $Z_j = HY_j$, $j = 1,\ldots,n$. Then $\overline{Z} = n^{-1} \sum_{j=1}^{n} Z_j$ is $N_p(\alpha^*,\Sigma^*)$, where $\alpha^* = H\alpha$ and $\Sigma^* = n^{-1}H\Sigma H^t$ = $\mathrm{diag}\,((\sigma^2+p\tau^2)/n,\ \sigma^2/n,\ldots,\sigma^2/n)$. Also, $\alpha^* \sim N_p(0,A^*)$, where $A^* = H(aI)H^t = aI$. Let $\delta^*(z) = H\delta(y)$. Then

$$\begin{aligned} L(\alpha,\delta) &= (\alpha-\delta)^t(\alpha-\delta) \\ &= (\alpha^*-\delta^*)^t(\alpha^*-\delta^*) \\ &= L(\alpha^*,\delta^*). \end{aligned}$$

Thus, $Q^* = I$, which implies that

$$\begin{aligned} D &= (\Sigma^*+A^*)^{-1/2}\Sigma^*Q^*\Sigma^*(\Sigma^*+A^*)^{-1/2} \\ &= n^{-2}\ \mathrm{diag}\ \left(\frac{(\sigma^2+p\tau^2)^2}{a+(\sigma^2+p\tau^2)/n},\ \frac{\sigma^4}{a+\sigma^2/n},\ldots,\frac{\sigma^4}{a+\sigma^2/n}\right) . \end{aligned}$$

For $p = 6$, $n = 3$, and $a = \tau^2 = 1$, the RSR for the estimators δ_{ρ_0}, $\delta^{MB,0}$, $\delta^{W,1}$

and for some different σ are given in the following table:

Table 1. RSR in a randomized block design.

σ	$\sqrt{1/2}$	1	$\sqrt{2}$	2	3
$RSR(\pi,\delta_{\rho_0})$	1.000	1.000	0.935	0.447	0.167
$RSR(\pi,\delta^{MB,0})$	0.831	0.600	0.332	0.156	0.073
$RSR(\pi,\delta^{W,1})$	0.687	0.419	0.193	0.087	0.051

Example 2. Suppose $d = (d_1, d_2,\ldots,d_p)$ is the diagonal elements of the matrix $D = (\Sigma+A)^{-1/2}\Sigma Q\Sigma(\Sigma+A)^{-1/2}$, where Σ, A, and Q are diagonal matrices. The RSR of the estimators δ_{ρ_0}, $\delta^{MB,0}$, and $\delta^{W,1}$, for various d, are given in Table 2.

Table 2. RSR for δ_{ρ_0}, $\delta^{MB,0}$, and $\delta^{W,1}$.

d	$RSR(\pi,\delta_{\rho_0})$	$RSR(\pi,\delta^{MB,0})$	$RSR(\pi,\delta^{W,1})$
(1.0,0.5,0.1)	1.000	.830	.535
(1.0,1.0,1.0,0.5)	.223	.190	.164
(1.0,1.0,0.5,0.5)	.368	.325	.179
(1.0,1.0,0.5,0.1)	.549	.460	.328
(1.0,1.0,0.1,0.1)	.818	.794	.465
(1.0,1.0,0.2,0.2)	.670	.632	.332
(1.0,0.8,0.8,0.8)	.247	.158	.139
(1.0,0.7,0.4,0.1)	.819	.472	.322
(1.0,1.0,1.0,0.001)	.368	.295	.537
(1.0,0.7,0.4,0.1,0.1)	.783	.433	.311
(1.0,1.0,1.0,0.5,0.1)	.256	.176	.222
(1.0,0.8,0.6,0.001,0.001)	.720	.412	.650
(1.0,1.0,0.9,0.8,0.4,0.1)	.192	.113	.167
(1.0,1.0,0.6,0.5,0.5,0.4)	.226	.162	.086

Remark 1. When d_p/d_2 or d_p/d_3 is close to one, $\delta^{W,1}$ appears to be better than $\delta^{MB,0}$ (see also Theorem 4).

Remark 2. When $p < 6$, $\delta^{W,1}$ appears to be superior to $\delta^{MB,0}$, except when d_{i+1}/d_i is very small (say, less than or equal to 0.1) for some $3 \leq i \leq p-1$.

Remark 3. When $p \geq 6$ and at least four of the d_j/d_1 are moderately close to one, then $\delta^{MB,0}$ is often better and rarely much worse than $\delta^{W,1}$. Hence $\delta^{MB,0}$ is reasonable to use in this situation.

Remark 4. We can use the fact that a convex combination of minimax estimators is also nimimax to develop an estimator improving on both $\delta^{MB,0}$ and $\delta^{W,1}$. (See Chen (1983) for details.) The improvement obtained did not seem to be substantial enough to justify the added complexity, however.

Example 3. Suppose $p = 6$, $c = 0.1$. The RSR of the estimators δ_{ρ_c} and $\delta^{MB,c}$ for various $(d_1,\ldots,d_p)$ are given in Table 3.

Table 3. RSR for δ_{ρ_c} and $\delta^{MB,c}$, when $p = 6$ and $c = 0.1$.

d	$RSR(\pi,\delta_{\rho_c})$	$RSR(\pi,\delta^{MB,c})$
(1.0,0.2,0.2,0.2,0.2,0.2)	.576	.208
(1.0,0.5,0.5,0.4,0.3,0.3)	.352	.071
(1.0,1.0,1.0,0.5,0.4,0.1)	.183	.049
(1.0,1.0,1.0,0.7,0.7,0.6)	.086	.032

Acknowledgment. The author wishes to thank Professor James Berger for his helpful suggestions and patient corrections.

Appendix

<u>Proof of Lemma 1</u>.

Using the identity (2.3), we have

$$\text{(A.1)} \qquad R(\theta,\delta_\rho) - R(\theta,\delta^0) = E_\theta[(\rho^2(r)-4\rho'(r))(x-\mu)^t(\ddagger+A)^{-1/2}$$
$$\cdot D(\ddagger+A)^{-1/2}(x-\mu)-2\rho(r)\,\mathrm{tr}D].$$

but

$$\text{(A.2)} \qquad (x-\mu)^t(\ddagger+A)^{-1/2}D(\ddagger+A)^{-1/2}(x-\mu) \le d_1 r.$$

Together with (A.1) and (A.2), it follows that

$$R(\theta,\delta_\rho) - R(\theta,\delta^0) \le d_1 E_\theta[(\rho^2(r)-4\rho'(r))r-2\tau\rho(r)]$$
$$< c. \quad ||$$

<u>Lemma 2</u>. Let $X \sim N_p(\mu,B)$, M be a $(p\times p)$ positive definite matrix and ρ be a function of r, where $r = (x-\mu)^t B^{-1}(x-\mu)$. If $E^r(\rho^2(r)) < \infty$, then

$$E^x(\rho(r)(x-\mu)^t B^{-1/2} M B^{-1/2}(x-\mu)) = (\mathrm{tr}M/p)\ E^r(r\rho(r)).$$

<u>Proof</u>. Let Λ be the $(p\times p)$ orthogonal matrix such that $H = \Lambda M \Lambda^{\ddagger}$ is the diagonal

matrix with diagonal elements $h_1 \geq h_2 \geq \cdots \geq h_p$, and let

$$z = \Lambda B^{-1/2}(x-\mu).$$

Then Z is $N_p(0,I)$ and

$$r = (x-\mu)^t B^{-1}(x-\mu) = z^t z$$

has a chi-square distribution with p degrees of freedom. But

$$[(x-\mu)^t B^{-1/2} M B^{-1/2}(x-\mu)]^2 \leq h_1^2 r^2,$$

which implies that

$$E^x[(x-\mu)^t B^{-1/2} M B^{-1/2}(x-\mu)] < \infty.$$

Therefore, by the Cauchy-Schwarz inequality

$$E^x[\rho(r)(x-\mu)^t B^{-1/2} M B^{-1/2}(x-\mu)]$$

exists, and

(A.3) $$E^x[\rho(r)(x-\mu)^t B^{-1/2} M B^{-1/2}(x-\mu)]$$

$$= E^x[\rho(r) z^t H z$$

$$= \sum_{i=1}^{p} h_i E^z(\rho(r) z_i^2).$$

Since $z_1, z_2, \ldots, z_p$ are i.i.d.,

$$\text{(A.4)} \qquad E^Z(\rho(r)Z_i^2) = \frac{1}{p} E^Z(r\rho(r)).$$

The desired result follows from (A.3), (A.4) and the fact that trH = trM.

<u>Proof of Theorem 1.</u>

It is clear that

$$r(\pi,\delta_{\rho_c}) - r(\pi,\delta^\pi)$$

$$= E^X\{[1-\min(1,\rho_{c,0}(r))]^2(x-\mu)^t(\Sigma+A)^{-1}\Sigma Q\Sigma(\Sigma+A)^{-1}(x-\mu)\} \ .$$

By Lemma 2,

$$r(\pi,\delta_{\rho_c}) - r(\pi,\delta^\pi)$$

$$= E^r\{[\max(0,1-\rho_{c,0}(r))]^2 r\}$$

$$= \int_b^\infty (1-\rho_{c,0}(r))^2 r f_p(r)dr \ .$$

Using the fact that $r\rho_{c,0}(r) = c/d_1$ for $r \geq b$ and noting that

$$\int_b^\infty r f_p(r)dr = 2bf_p(b) + p(1-\psi_p(b))$$

and

$$r(\pi,\delta^0) - r(\pi,\delta^\pi) = \sum_{i=1}^{p} d_i ,$$

the desired result follows. ||

Proof of Theorem 2.

Let $z_i = (\sigma_i^2+a_i)^{-1/2}(x_i-\mu_i)$, and observe that the z_i are i.i.d. N(0,1) and that

$$r(\pi,\delta^{MB,c}) - r(\pi,\delta^\pi)$$

$$= E^X \left\{ \sum_{i=1}^{p} \left[\frac{q_i\sigma_i^2}{\sigma_i^2+a_i} x_i \left(1- \sum_{j=i}^{p} \frac{d_j-d_{j+1}}{d_i} \min\left(1,\rho_{c,0}^{(j)}(r_j)\right)\right)\right]^2 \right\}$$

$$= E^Z \left\{ \sum_{i=1}^{p} d_i z_i^2 \left[\sum_{j=i}^{p} \frac{d_j-d_{j+1}}{d_i} \left(1-\min\left(1,\rho_{c,0}^{(j)}(r_j)\right)\right)\right]^2 \right\}$$

$$= \sum_{i=1}^{p} \sum_{j=i}^{p} \sum_{k=i}^{p} \frac{(d_j-d_{j+1})(d_k-d_{k+1})}{d_i}$$

$$\cdot E^Z[z_i^2(1-\min(1,\rho_{c,0}^{(j)}(r_j)))(1-\min(1,\rho_{c,0}^{(k)}(r_k))].$$

The desired result follows.||

Lemma 3. When $p \geq 3$, the equation (4.1) has a unique positive solution.

Proof.

Let

$$h(y) = y - \frac{1}{p-2} \left[\sum_{i=1}^{p} d_i\left(\sqrt{1+ \frac{(p-2)y}{d_i}} -1\right)\right].$$

Clearly, $h(0) = 0$ and $h(y) > 0$ for sufficiently large y. On the other hand, for a sufficiently small positive mumber ε, $\sqrt{1+\varepsilon} \approx 1 + \frac{\varepsilon}{2}$. Thus we have

$$h(\varepsilon) \approx \varepsilon - \frac{1}{p-2} \left[\sum_{i=1}^{p} d_i \left(1+\frac{(p-2)\varepsilon}{2d_i} - 1\right) \right]$$

$$= (1 - \frac{p}{2})\varepsilon < 0.$$

Therefore, h has at least one positive root.

But

$$h'(y) = 1 - \frac{1}{2} \sum_{i=1}^{p} \left(1+\frac{(p-2)y}{d_i}\right)^{-1/2}$$

and

$$h''(y) = \frac{p-2}{4} \sum_{i=1}^{p} \frac{1}{d_i} \left[1+ \frac{(p-2)y}{d_i} \right]^{-3/2} > 0$$

for all $y > 0$. Thus, h is concave upward. Hence h has a unique positive root. This completes the proof. ||

Proof of Theorem 3.

Let

$$\gamma(x) = -\min\left(1, \frac{2(p-2)}{r}\right) W (\ddagger+A)^{-1}(x-\mu).$$

Using the identity (2.3), when $r \leq 2(p-2)$, we have

$$\mathcal{D}\gamma(x) = (x-\mu)^t(\ddagger+A)^{-1} W^2D(x-\mu) - 2tr(WD)$$

$$\leq r(ch_{max}(W^2D)) - 2tr(WD)$$

$$\leq 2(p-2)(ch_{max}(W^2D)) - 2tr(WD).$$

Since $ch_{max}(W^2D) \leq ch_{max}(W^2D+\frac{2}{p-2} WD) = \frac{1}{p-2} tr(WD)$, it follows that $\mathcal{D}\gamma(x) \leq 0$.

On the other hand, when $r > 2(p-2)$,

$$\mathcal{D}\gamma(x) = \left[\frac{2(p-2)}{r}\right]^2 (x-\mu)^t W^2D (\ddagger+A)^{-1}(x-\mu)$$

$$+ \left[\frac{4(p-2)}{r^2}\right] (x-\mu)^t WD(\ddagger+A)^{-1}(x-\mu) - \frac{4(p-2)}{r} tr(WD)$$

$$= \left[\frac{4(p-2)}{r^2}\right]\{(x-\mu)^t [(p-2)W^2D+2WD] (\ddagger+A)^{-1}(x-\mu)\}$$

$$- \frac{4(p-2)}{r} tr(WD)$$

$$\leq \frac{4(p-2)}{r} \{ch_{max} [(p-2)W^2D+2WD] - tr(WD)\}$$

$$\leq 0.$$

Thus δ is a minimax estimator of θ.

For the calculation of $RSR(\pi,\delta^{W,1})$, note first that

$$r(\pi,\delta^{W,1}) - r(\pi,\delta^{\pi})$$

$$= E\{(x-\mu)^t \left[I-\min(1, \frac{2(p-2)}{r}) W\right]^2 D(\ddagger+A)^{-1}(x-\mu)\} .$$

Lemma 2 implies that

$$r(\pi,\delta^{W,1}) - r(\pi,\delta^{\pi})$$

$$= p\, trD - 2\, tr(WD)\, E^r\left[\min(1, \frac{2(p-2)}{r})r\right]$$

$$+ tr(W^2D)\, E^r \left[(\min(1, \frac{2(p-2)}{r}))^2 r\right] .$$

Noting that

$$tr(W^2D) = tr(WD),$$

the desired result follows. ||

<u>Proof of Theorem 4</u>.

(i) Without loss of generality, we will assume that $d_1 = 1$. Let $d_2 = d_3 = \ldots = d_p = t$. Then, by Theorem 2,

$$r(\pi,\delta^{MB,0}) - r(\pi,\delta^{\pi}) = (1-t)^2 + 2t(1-t)\, T^0_{1,p} + t^2\, T^0_{p,p}.$$

Since $\sum_{i=1}^{p} d_i = 1 + (p-1)t$ and $T^0_{1,p} \geq T^0_{p,p}$,

$$\text{(A.5)} \qquad r(\pi,\delta^{MB,0}) - r(\pi,\delta^{\pi}) \geq \sum_{i=1}^{p} d_i - (pt + t\,(1-t))(1-T^0_{p,p}) \ .$$

On the other hand,

$$\text{(A.6)} \qquad r(\pi,\delta^{W,1}) - r(\pi,\delta^{\pi}) = \sum_{i=1}^{p} d_i - y\,(1-T_{p,p}).$$

Now, if we can show that

$$\text{(A.7)} \qquad y_0 \geq pt + t\,(1+t),$$

then the desired result follows from (A.5) and (A.6).

From the proof in Lemma 3, we know that showing that (A.7) holds is equivalent to showing that

$$\text{(A.8)} \qquad h(x) \leq 0,$$

where

$$h(x) = x + \frac{1}{p-2}\left[1 + (p-1)t - \sqrt{1+(p-2)x} - (p-1)t\sqrt{1 + \frac{(p-2)x}{t}}\right]$$

and $x = pt + t\,(1-t)$. After a little algebra, the inequality (A.8) is equivalent to

$$\text{(A.9)} \qquad \frac{-t(1-t)^2}{(p-2)^2}\,[(p-2)^3\, t\,(4-t) + (p-2)^2\, t\,(t^2-7t+14) + (p-2)\,(-2t^2+6t+3) + (1+t)] \le 0.$$

But it is clear that (A.9) holds for $0 < t \le 1$, completing the proof of part (i). The proof of part (ii) is similar to that of part (i). ||

Table 4. Values of $T^{c*}_{j,k}$

$c* = 0$

j \ k	1	2	3	4	5	6	7	8	9	10
1	1.0000									
2	1.0000	1.0000								
3	.4675	.4675	.2975							
4	.2707	.2702	.1830	.1353						
5	.1723	.1723	.1191	.0915	.0727					
6	.1160	.1160	.0810	.0632	.0516	.0427				
7	.0811	.0811	.0570	.0448	.0370	.0313	.0267			
8	.0583	.0583	.0410	.0324	.0270	.0230	.0200	.0174		
9	.0427	.0427	.0301	.0239	.0199	.0171	.0150	.0132	.0117	
10	.0318	.0318	.0225	.0179	.0150	.0129	.0113	.0101	.0090	.0080

$c* = .01$

j \ k	1	2	3	4	5	6	7	8	9	10
1	.8502									
2	.6462	.5282								
3	.4026	.3417	.2584							
4	.2477	.2133	.1684	.1297						
5	.1602	.1388	.1114	.0887	.0712					
6	.1084	.0943	.0755	.0615	.0507	.0422				
7	.0760	.0662	.0537	.0437	.0364	.0309	.0264			
8	.0547	.0476	.0388	.0317	.0266	.0228	.0198	.0172		
9	.0401	.0350	.0285	.0234	.0197	.0170	.0148	.0131	.0116	
10	.0299	.0261	.0213	.0175	.0148	.0128	.0112	.0100	.0089	.0079

$c* = .05$

j \ k	1	2	3	4	5	6	7	8	9	10
1	.6902									
2	.5082	.4109								
3	.3317	.2775	.2173							
4	.2136	.1812	.1471	.1169						
5	.1416	.1210	.0995	.0815	.0666					
6	.0971	.0832	.0690	.0572	.0478	.0402				
7	.0685	.0588	.0490	.0408	.0346	.0296	.0254			
8	.0495	.0425	.0355	.0297	.0253	.0219	.0191	.0166		
9	.0364	.0313	.0262	.0220	.0188	.0163	.0143	.0127	.0112	
10	.0272	.0234	.0196	.0165	.0141	.0123	.0109	.0097	.0087	.0077

Table 4 (Cont'd)

c* = .1

j \ k	1	2	3	4	5	6	7	8	9	10
1	.5870									
2	.4275	.3461								
3	.2850	.2381	.1899							
4	.1881	.1592	.1311	.1060						
5	.1268	.1080	.0901	.0749	.0620					
6	.0879	.0751	.0630	.0530	.0449	.0380				
7	.0624	.0534	.0450	.0381	.0326	.0281	.0242			
8	.0453	.0388	.0328	.0278	.0239	.0208	.0183	.0160		
9	.0334	.0287	.0242	.0206	.0178	.0156	.0138	.0122	.0108	
10	.0250	.0215	.0182	.0155	.0134	.0118	.0104	.0093	.0084	.0075

c* = .2

j \ k	1	2	3	4	5	6	7	8	9	10
1	.4628									
2	.3344	.2728								
3	.2278	.1911	.1558							
4	.1543	.1309	.1097	.0907						
5	.1062	.0907	.0768	.0650	.0547					
6	.0747	.0639	.0545	.0466	.0400	.0343				
7	.0536	.0459	.0393	.0338	.0293	.0255	.0222			
8	.0391	.0336	.0288	.0248	.0217	.0190	.0168	.0148		
9	.0290	.0249	.0214	.0185	.0162	.0143	.0127	.0114	.0101	
10	.0218	.0187	.0161	.0139	.0122	.0108	.0097	.0087	.0078	.0070

c* = .4

j \ k	1	2	3	4	5	6	7	8	9	10
1	.3248									
2	.2338	.1942								
3	.1627	.1384	.1159							
4	.1133	.0972	.0833	.0706						
5	.0798	.0688	.0595	.0515	.0442					
6	.0572	.0494	.0429	.0375	.0328	.0285				
7	.0416	.0360	.0313	.0275	.0242	.0214	.0189			
8	.0307	.0266	.0232	.0204	.0181	.0161	.0144	.0128		
9	.0229	.0199	.0174	.0153	.0136	.0122	.0110	.0099-	.0088	
10	.0173	.0150	.0131	.0116	.0103	.0093	.0084	.0076	.0069	.0062

Table 4 (Cont'd)

c* = .6

j \ k	1	2	3	4	5	6	7	8	9	10
1	.2439									
2	.1755	.1483								
3	.1237	.1068	.0911							
4	.0875	.0760	.0662	.0571						
5	.0626	.0546	.0479	.0421	.0367					
6	.0453	.0396	.0349	.0309	.0274	.0241				
7	.0333	.0292	.0257	.0229	.0204	.0183	.0162			
8	.0248	.0217	.0192	.0171	.0153	.0138	.0124	.0111		
9	.0186	.0163	.0144	.0129	.0116	.0105	.0095	.0086	.0078	
10	.0142	.0124	.0110	.0098	.0088	.0080	.0073	.0067	.0061	.0055

c* = .8

j \ k	1	2	3	4	5	6	7	8	9	10
1	.1896									
2	.1366	.1174								
3	.0971	.0850	.0737							
4	.0694	.0611	.0540	.0472						
5	.0502	.0443	.0394	.0350	.0309					
6	.0367	.0325	.0289	.0259	.0232	.0206				
7	.0272	.0240	.0215	.0193	.0174	.0157	.0141			
8	.0203	.0180	.0160	.0145	.0131	.0119	.0108	.0090		
9	.0154	.0136	.0122	.0110	.0100	.0091	.0083	.0076	.0069	
10	.0117	.0104	.0093	.0084	.0076	.0070	.0064	.0059	.0054	.0049

c* = 1.0

j \ k	1	2	3	4	5	6	7	8	9	10
1	.1507									
2	.1087	.0949								
3	.0778	.0691	.0607							
4	.0561	.0500	.0447	.0395						
5	.0409	.0365	.0328	.0295	.0263					
6	.0301	.0269	.0243	.0219	.0198	.0178				
7	.0224	.0201	.0181	.0164	.0149	.0136	.0123			
8	.0169	.0151	.0136	.0124	.0113	.0103	.0095	.0086		
9	.0128	.0115	.0104	.0095	.0086	.0079	.0073	.0067	.0061	
10	.0090	.0088	.0079	.0072	.0066	.0061	.0056	.0052	.0048	.0044

Bibliography

[1] Berger, J. (1979). Multivariate estimation with nonsymmetric loss function. Optimizing Methods in Statistics. (J. S. Rustagi, ed.). Academic Press, New York.

[2] Berger, J. (1980a). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. Ann. Statist. 8, 545-571.

[3] Berger, J. (1980b). A robust generalized Bayes estimator and confidence region of a multivariate normal mean. Ann. Statist. 8, 716-761.

[4] Berger, J. (1980c). Statistical Decision Theory: Foundations, Concepts, and Methods. Springer-Verlag, New York.

[5] Berger, J. (1982a). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. Statistical Decision Theory and Related Topic III, eds. S. S. Gupta and J. Berger, New York: Academic Press.

[6] Berger, J. (1982b). Selecting a minimax estimator of a multivariate normal mean. Ann. Statist. 10, 81-92.

[7] Berger, J. (1983). Robust Bayesian Viewpoint. In Robustness in Bayesian Statistics, J. Kadane (ed.). North-Holland, Amsterdam.

[8] Bhattacharya, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. Ann. Math. Statist. 37, 1819-1824.

[9] Bickel, P. J. (1980). Minimax estimation of the mean of a normal distribution subject to doing well at a point. Technical Report. Department of Statistics, University of California, Berkeley.

[10] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. Ann. Math. Statist. 42, 855-903.

[11] Chen, S. Y. (1983). Restricted Risk Bayes Estimator. Ph.D. Thesis. Department of Statistics, Purdue University.

[12] Dey, D. and Berger, J. (1983) Combining coordinates in simultaneous estimation of normal mean. To appear in J. Statist. Planning and Inference.

[13] Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators - Part 1: the Bayes case. J. Amer. Statist. Assoc. 66, 807-815.

[14] Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators - Part 2: the empirical Bayes case. J. Amer. Statist. Assoc. 67, 130-139.

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

[16] Efron, B. and Morris, C. (1973b). Combining possibly related estimation problems. J. Roy. Statist. Soc. B, 35, 379-421.

[17] Hodges, J. L., Jr., and Lehmann, E. L. (1952). Some applications of the Cramer-Rao inequality. Proc. Second Berkeley Symp. Math. Statist. Prob. 13-22. Univer. of California Press.

[18] Huber, P. J. (1964). Robust estimation of a location parameter. Ann. Math Stat. 35, 73-101.

[19] Huber, P. J. (1977). Robust Covariances. Statistical Decision Theory and applications in multiparameter estimation. Statistical Decision Theory and Related Topics. II. S. S. Gupta and D. S. Moore. eds. Academic Press. New York.

[20] Huber, P. J. (1981). Robust Statistics. Wiley. New York.

[21] Hudson, H. M (1974). Empirical Bayes estimation. Technical Report #58, Stanford University.

[22] Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. Ann. Statist. 6, 473-484.

[23] James, W. and Stein, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Stat. Prob. 1, 361-379. University of California Press.

[24] Marazzi, A. (1980). Robust Bayesian estimation for the linear model. Research Report No. 27. Fachgruppe fuer Statistik, Eidgenoessische Technische Hochschule, Zurich.

[25] Marazzi, A. (1982). On constrained minimation of the Bayes risk for the linear model. Research Report No. 34, Fachgruppe Fure Statistik Eidgenoessische Technische Hochschule, Zurich.

[26] Shapiro, S. H. (1972). A compromise between Bayes and minimax approaches to estimation. Technical Report No. 31, Department of Statistics, Stanford University.

[27] Shapiro, S. H. (1975). Estimation of location and scale parameters - a compromise. Communications in Statist. 4(12), 1093-1108.

[28] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 197-206. University of California Press.

[29] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135-1151.