The Annals of Statistics
1989, Vol. 17, No. 2, 890–906

ESTIMATION OF NORMAL MEANS: FREQUENTIST ESTIMATION OF LOSS[1]

BY K. L. LU AND JAMES O. BERGER

Purdue University and Cornell University, and Purdue University

In estimation of a p-variate normal mean with identity covariance matrix, Stein-type estimators can offer significant gains over the mle in terms of risk with respect to sum of squares error loss. Their maximum risk is still equal to p, however, which will typically be their "reported loss." In this paper we consider use of data-dependent "loss estimators." Two conditions that are attractive for such a loss estimator are that it be an improved loss estimator under some scoring rule and that it have a type of frequentist validity. Loss estimators with these properties are found for several of the most important Stein-type estimators. One such estimator is a generalized Bayes estimator, and the corresponding loss estimator is its posterior expected loss. Thus Bayesians and frequentists can potentially agree on the analysis of this problem.

1. Introduction. For the problem of estimating a multivariate normal mean, we consider a frequentist approach to conditioning, based on use of loss estimators. In classical decision theory, the following nonconditional approach is standard. Select a procedure $\delta(x)$, define a criterion (or loss) $L(\theta, \delta)$ and report $(\delta, R(\theta, \delta))$, where $R(\theta, \delta) = E_\theta L(\theta, \delta(X))$ is the risk function and gives the long run performance of δ for each θ. Because θ is unknown, however, $R(\theta, \delta)$ is also unknown. Thus it is common [and argued in Berger (1985a) to be necessary for a valid nonconditional frequentist interpretation] to report

$$\overline{R}_\delta = \sup_\theta R(\theta, \delta) \tag{1}$$

as the operational measure of accuracy.

The approach based on loss estimators recognizes that the ideal measure of performance of δ would be $L(\theta, \delta(x))$ itself, were it obtainable, and recommends reporting a data-dependent estimator $\hat{L}_\delta(x)$ of the loss, instead of the constant $\overline{R}_\delta$. To evaluate the success with which $L(\theta, \delta(x))$ is estimated, some criterion [or communication loss; see Berger (1985b)] needs to be considered. We will consider the simple quadratic communication loss

$$L^*(L, \hat{L}) = (\hat{L} - L)^2. \tag{2}$$

Since we are working in the frequentist domain, the performance of $\hat{L}$ will be

Received December 1986; revised April 1988.
[1]Research supported by NSF Grant DMS-84-01996, by a David Ross fellowship at Purdue University and by the U.S. Army Research Office through the Mathematical Sciences Institute at Cornell University.
AMS 1980 *subject classifications*. Primary 62J07, 62C10; secondary 62C15.
Key words and phrases. Estimated loss, communication loss, communication risk, Stein estimation, generalized Bayes estimator, posterior variance.

measured by the "communication risk"

$$R_L^*(\theta, \hat{L}_\delta) = E_\theta L^*(L(\theta, \delta(X)), \hat{L}_\delta(X)). \tag{3}$$

Usual notions of admissibility, etc., can be applied to loss estimators $\hat{L}$ and their corresponding risks R_L^*. We will, henceforth, do so without further comment.

We will be concerned with two criteria for a loss estimator $\hat{L}$. The first is that it improve on $\overline{R}_\delta$, the usual constant frequentist report. Thus we seek, for $\hat{L}_\delta$:

CONDITION 1. Improved loss estimation:

$$R_L^*(\theta, \hat{L}_\delta) \le R_L^*(\theta, \overline{R}_\delta) \quad \text{for all } \theta. \tag{4}$$

A second criterion of interest is that the average long run reported loss never be less than the average long run actual loss. [A related notion is the *guaranteed conditional confidence* of Brown (1978).] Formally, we will consider:

CONDITION 2. Frequentist validity:

$$E_\theta \hat{L}_\delta(X) \ge R(\theta, \delta) \quad \text{for all } \theta. \tag{5}$$

Satisfaction of Condition 1 is arguably necessary before a frequentist would have clear cause to adopt $\hat{L}_\delta$, instead of $\overline{R}_\delta$, as the reported loss. Satisfaction of Condition 2 is a less obvious requirement; a traditional frequentist might feel that this is necessary, arguing [as in Berger (1985a)] that violation of (5) means that the "average reported loss" could be less than the "average actual loss" (and hence be nonconservative), but others might disagree [cf. Robinson (1979a, b), Johnstone (1987) and Rukhin (1987)]. Inequality in (5) is typically acceptable as being "conservative" and necessary to obtain sensible $\hat{L}_\delta$.

It is possible to combine the decision problems involving δ and $\hat{L}$ [as in Rukhin (1987)], but we will not do so. The losses involved in estimating θ and L are typically very different and hard to construct on a common utility scale. In this paper we will, in fact, consider only certain previously proposed estimators δ, treating them as fixed in discussing the estimation of L. Thus our orientation is that of taking a procedure advocated for the original decision problem and studying how best to estimate its actual loss.

The specific problem to be addressed in this paper is that of estimating a p-variate normal mean $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^t$, based on $\mathbf{X} = (X_1, \dots, X_p)^t \sim N_p(\boldsymbol{\theta}, I)$, where I is the $p \times p$ identity matrix and $p \ge 3$. General classes of Stein-type estimators (see Section 2) and generalized Bayes estimators (see Section 3) will be studied. These include the following three familiar estimators, that will be used for illustrative purposes.

1. Positive part James–Stein estimator:

$$\delta(\mathbf{x}) = \left(1 - \frac{a}{\|\mathbf{x}\|^2}\right)^+ \mathbf{x}. \tag{6}$$

Here $+$ denotes the positive part, $\|\mathbf{x}\|^2 = \sum_{i=1}^p x_i^2$ and $0 \le a \le p - 2$, with $a = p - 2$ being the usual constant considered.

2. Lindley–Efron–Morris–Stein estimator [Lindley (1962) and Efron and Morris (1973)]:

$$\delta(\mathbf{x}) = \left(1 - \frac{a}{\|\mathbf{x} - \bar{x}\mathbf{1}\|^2}\right)^+ (\mathbf{x} - \bar{x}\mathbf{1}) + \bar{x}\mathbf{1}, \tag{7}$$

where $\bar{x} = \sum_{i=1}^{p} x_i/p$, $\mathbf{1} = (1,\dots,1)^t$ and $0 \le a \le p - 3$, with $a = p - 3$ being the usual constant considered.

3. Generalized Bayes estimator [Strawderman (1971) and Berger (1980)]:

$$\delta(\mathbf{x}) = \left(1 - \frac{r_n(\|\mathbf{x}\|^2)}{\|\mathbf{x}\|^2}\right)\mathbf{x}, \tag{8}$$

where, letting $v = \|\mathbf{x}\|^2$,

$$r_n(v) = \frac{v\int_0^1 \lambda^n \exp(-\lambda v/2)\, d\lambda}{\int_0^1 \lambda^{(n-1)} \exp(-\lambda v/2)\, d\lambda}. \tag{9}$$

[As usual, the estimators in (6) or (8) could "shrink" toward any given point $\boldsymbol{\mu}$. We simply set $\boldsymbol{\mu} = \mathbf{0}$ for simplicity.] These estimators dominate the usual estimator $\delta_0(\mathbf{x}) = \mathbf{x}$ under sum of squares error loss, providing $p \ge 3$ for (6) and (8) and $p \ge 4$ for (7). They also have a number of other desirable qualities. For all such estimators, however, $\overline{R}_\delta = p$ (which is also the risk of δ_0) and therefore does not exhibit the gain obtained through use of these estimators.

Stein (1981) provides a loss estimator $\hat{L}_\delta(\mathbf{x})$ for the problem, which he terms the "unbiased estimator of risk." This loss estimator satisfies the frequentist validity criterion; indeed, it is the unique loss estimator for which equality holds in (5). It is not clear, however, if Condition 1 also holds. Furthermore, the unbiased estimator of risk can have obviously undesirable properties as a loss estimator. For instance, the unbiased estimator of risk for (6) is negative if $\|\mathbf{x}\|^2 < a < p$, which would be a silly reported loss. (The unbiased estimator of risk can be modified to eliminate such undesirable features; see Section 2.)

We will consider alternative loss estimators, corresponding to the δ in (6), (7) and (8), which satisfy both (4) and (5) and which are intuitively reasonable. For (8), the loss estimator we shall consider is posterior expected loss, a report with many attractive features. Indeed, in this problem we demonstrate the perhaps surprising fact that the posterior mean and variance (for natural Strawderman-type priors) are improved frequentist procedures for the original *and* loss estimation decision problems, respectively. In that these priors are based on subjective input (at a minimum, the choice of where to shrink), this provides the first illustration of which we are aware in which there is complete agreement between frequentist and *subjective* Bayesian analyses, "complete" in the sense that both the estimator of $\boldsymbol{\theta}$ and the reported accuracy satisfy relevant frequentist and Bayesian criteria.

Our loss estimators also satisfy the condition

$$0 \le \hat{L}(\mathbf{x}) < \overline{R}_\delta = p \quad \text{for all } \mathbf{x}. \tag{10}$$

Of note here is that one will always estimate the loss to be less than p, giving immediate and visible evidence of the expected improvement obtained through use of (6), (7) or (8).

We note that related problems have recently been considered by Johnstone (1987), although he does not require Condition 2. His efforts center around determination of the admissibility of Stein's unbiased estimator of risk for the mle $\mathbf{x}$ and the nonpositive part James–Stein estimator. He proves the surprising result that the unbiased estimator of risk for $\mathbf{x}$ (which turns out to be just p) is admissible for $p \leq 4$, but inadmissible for $p \geq 5$. For the usual James–Stein estimator, he establishes inadmissibility of the unbiased estimator of risk. Results for confidence procedures that are analogous to the results in this paper are given in Lu and Berger (1988). Other work, related to the idea of estimating the loss, includes Lehmann (1959), Sandved (1968), Kiefer (1977) and Robinson (1979a, b).

2. Loss estimation for positive part estimators. The estimators in (6) and (7) are of the form

$$\delta_a(\mathbf{x}) = \mathbf{x} - r(\mathbf{x})(\mathbf{x} - \boldsymbol{\mu}(\mathbf{x})), \tag{11}$$

where

$$\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\alpha} + B(\mathbf{x} - \boldsymbol{\alpha}), \tag{12}$$

$$r(\mathbf{x}) = 1 - \left(1 - \frac{a}{v(\mathbf{x})}\right)^+, \tag{13}$$

$$v(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\alpha})^t(I - B)(\mathbf{x} - \boldsymbol{\alpha}), \tag{14}$$

$\boldsymbol{\alpha}$ is a specified vector and B is an idempotent matrix with rank

$$\gamma = \operatorname{rank}(B). \tag{15}$$

For instance, (6) is of this form with $\boldsymbol{\alpha} = \mathbf{0}$ and $B = 0$, and (7) is of this form with $\boldsymbol{\alpha} = \mathbf{0}$ and $B = \mathbf{1}\mathbf{1}^t/p$. It is well known [cf. Berger (1976)] that δ_a in (11) is minimax if $0 \leq a \leq 2(p - 2 - \gamma)$. The usual choice of a is $a = p - 2 - \gamma$, for a variety of admissibility and empirical Bayes reasons.

We shall consider loss estimators, corresponding to δ_a, of the form

$$\hat{L}_t(\mathbf{x}) = \begin{cases} g(v(\mathbf{x})), & \text{if } v(\mathbf{x}) \leq a, \\ p - t/v(\mathbf{x}), & \text{if } v(\mathbf{x}) > a, \end{cases} \tag{16}$$

where $0 < t \leq b^2$,

$$b = \left[2(p - 2 - \gamma)a - a^2\right]^{1/2} \tag{17}$$

(equaling $p - 2 - \gamma$ if $a = p - 2 - \gamma$) and $g(v)$ is a nondecreasing convex function on $(0, a]$ such that $g(v) \geq 0$, $g(a) = p - t/a$ and the left derivative of g exists at a and is no larger than t/a^2.

A particularly attractive special case is

$$\hat{L}^*(\mathbf{x}) = \begin{cases} \underline{L}, & \text{if } v(\mathbf{x}) \le \underline{L} - p + 2a, \\ v(\mathbf{x}) + p - 2a, & \text{if } \underline{L} - p + 2a \le v(\mathbf{x}) \le a, \\ p - a^2/v(\mathbf{x}), & \text{if } v(\mathbf{x}) > a, \end{cases} \tag{18}$$

where $\underline{L} \le p - a$ is a specified "smallest loss" that will be reported. $\underline{L}$ can be selected according to prior information if available (e.g., $\underline{L}$ could be chosen to be the posterior variance at zero). A possible default value is $\underline{L} = p - a$ (although this may often be unreasonably small), in which case $\hat{L}^*$ becomes

$$\hat{L}^{**}(\mathbf{x}) = \begin{cases} p - a, & \text{if } v(\mathbf{x}) \le a, \\ p - a^2/v(\mathbf{x}), & \text{if } v(\mathbf{x}) > a. \end{cases} \tag{19}$$

Note that a wide range of t and g could be chosen in (16). Indeed, there is nothing within the frequentist theory considered here to prevent one from choosing some g which is, say, zero over an interval. This could result in a reported loss of zero, however, which is obviously silly. It will be seen later [see (31)] that $t = b^2$ corresponds to modification of the unbiased estimator of risk. The reasons for considering other t will be discussed at the end of the section.

The following theorems show that $\hat{L}_t(\mathbf{x})$ in (16) has frequentist validity and dominates $\overline{R}_\delta = p$, under suitable conditions. For use in the theorem, define

$$\begin{aligned} a_p(\gamma) &= \min\left\{p - 2 - \gamma, p - 4 - \gamma + \sqrt{(p - 3 - \gamma)^2 - 1}\right\}, \\ t_0 &= 2(p - 4 - \gamma)(1 + a) - a^2. \end{aligned} \tag{20}$$

THEOREM 2.1. *For $p \ge 5 + \gamma$, $0 < a \le a_p(\gamma)$ and $0 < t \le 2t_0$, $\hat{L}_t$ satisfies*

$$E_\theta\left[\hat{L}_t(\mathbf{X}) - \|\delta_a(\mathbf{X}) - \theta\|^2\right]^2 < E_\theta\left[p - \|\delta_a(\mathbf{X}) - \theta\|^2\right]^2 \quad \textit{for all } \theta. \tag{21}$$

NOTE. If $\gamma = 0$, i.e., we are considering the usual positive part James–Stein estimator, relevant values of $a_p(0)$ are $a_5(0) = 1 + \sqrt{3}$ and $a_p(0) = p - 2$ for $p \ge 6$. Also, $a_p(\gamma) = a_{p-\gamma}(0)$ for $\gamma \ge 1$.

COROLLARY 2.1. *For $p \ge 5 + \gamma$ and $0 \le a \le a_p^*(\gamma)$, $\hat{L}^*$ (and $\hat{L}^{**}$) satisfy (21). Here*

$$a_p^*(\gamma) = \min\left\{p - 2 - \gamma, p - 6 - \gamma + \sqrt{(p - 4 - \gamma)^2 + 4}\right\}. \tag{22}$$

NOTE. If $\gamma = 0$, relevant values of $a_p^*(0)$ are $a_5^*(0) = 1.2361$, $a_6^*(0) = 2.8284$, $a_7^*(0) = 4.6056$ and $a_p^*(0) = p - 2$ for $p \ge 8$; also, $a_p^*(\gamma) = a_{p-\gamma}^*(0)$ for $\gamma \ge 1$.

PROOF OF COROLLARY 2.1. It can be checked that, if $0 \le a \le a_p^*(\gamma)$, then

$$t = a^2 \le 2t_0. \tag{23}$$

Hence Theorem 2.1 yields the result. □

PROOF OF THEOREM 2.1. Under the given conditions on $g(v)$, it is possible to find a sequence $g_m(v)$ which converges to $g(v)$ pointwise and for which each g_m is [on $(0, a]$] nondecreasing, convex, everywhere twice differentiable with piecewise continuous second derivative and satisfies $g_m(v) > 0$, $g_m(a) = p - t/a$ and $g_m'(a) = t/a^2$. It thus suffices to establish the theorem for all such g_m. (We, henceforth, drop the subscript m.)

Define the auxiliary function

$$S_t(\mathbf{x}, \boldsymbol{\theta}) = \left(p - \|\delta_a(\mathbf{x}) - \boldsymbol{\theta}\|^2\right)^2 - \left(\hat{L}_t(\mathbf{x}) - \|\delta_a(\mathbf{x}) - \boldsymbol{\theta}\|^2\right)^2. \tag{24}$$

Using Lemma A.1 in the Appendix, calculation gives

$$E_{\boldsymbol{\theta}} S_t(\mathbf{X}, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \hat{S}_t(\mathbf{X}), \tag{25}$$

where $\hat{S}_t(\mathbf{x})$ is given by

$$\begin{aligned} \hat{S}_t(\mathbf{x}) = &- \left(p - \hat{L}_t(\mathbf{x})\right)\left(p - \hat{L}_t(\mathbf{x}) + 2r(\mathbf{x})^2 v(\mathbf{x})\right) \\ &+4 \sum_{i=1}^{p} \frac{\partial}{\partial x_i}\left[r(\mathbf{x})\left(p - \hat{L}_t(\mathbf{x})\right)(x_i - \mu_i)\right] - 2 \sum_{i=1}^{p} \frac{\partial^2}{\partial x_i^2}\left(p - \hat{L}_t(\mathbf{x})\right). \end{aligned} \tag{26}$$

Calculation yields [writing $v = v(\mathbf{x})$ for convenience], for $v \le a$,

$$\begin{aligned} \hat{S}_t(\mathbf{x}) = &\left(p - g(v)\right)\left[4(p - 2 - \gamma) - \left(p - g(v) + 2v\right)\right] \\ &+8\left(p - g(v) - g'(v)v\right) + 8g''(v)v + 4g'(v)\left(p - g(v)\right) \end{aligned} \tag{27}$$

and, for $v > a$,

$$\hat{S}_t(\mathbf{x}) = t(2t_0 - t)/v^2. \tag{28}$$

When $v \le a$, $\hat{S}_t(\mathbf{x}) > 0$ since $g'(v) \le 1$ and the function $p - g(v) + 2v$ is increasing in v with a maximum of $3b$. When $v > a$ and $0 < t \le 2t_0$, it is clear that $\hat{S}_t(\mathbf{x}) > 0$. It follows that $E_\theta \hat{S}_t(\mathbf{X}) > 0$, completing the proof. □

THEOREM 2.2. *For $p \ge 2 + \gamma$, $0 < a \le p - 2 - \gamma$ and $0 < t \le b^2$, $\hat{L}_t$ satisfies*

$$E_{\boldsymbol{\theta}} \hat{L}_t(\mathbf{X}) > E_{\boldsymbol{\theta}} \|\delta_a(\mathbf{X}) - \boldsymbol{\theta}\|^2. \tag{29}$$

PROOF. Using Lemma A.1 in the Appendix, calculation gives

$$\begin{aligned} E_{\boldsymbol{\theta}}\|\delta_a(\mathbf{X}) - \boldsymbol{\theta}\|^2 &= E_{\boldsymbol{\theta}}\|\mathbf{X} - \boldsymbol{\theta} - r(\mathbf{x})(\mathbf{x} - \boldsymbol{\mu}(\mathbf{x}))\|^2 \\ &= E_{\boldsymbol{\theta}}\left[p + r(\mathbf{X})^2 v(\mathbf{X}) - 2\sum_{i=1}^{p} \frac{\partial}{\partial X_i} r(\mathbf{X})\left(X_i - \mu_i(\mathbf{X})\right)\right] \\ &= E_{\boldsymbol{\theta}} \hat{L}^u(\mathbf{X}), \end{aligned} \tag{30}$$

where $\hat{L}^u$ is the unbiased estimator of risk given by

$$\hat{L}^u(\mathbf{x}) = \begin{cases} v(\mathbf{x}) - p + 2\gamma, & \text{if } v(\mathbf{x}) \le a, \\ p - b^2/v(\mathbf{x}), & \text{if } v(\mathbf{x}) > a. \end{cases} \tag{31}$$

TABLE 1
The proportional decrease in communication risk for $\hat{L}^{**}$

$\|\theta\|$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
0.00	0.55	0.80	0.87	0.89	0.93	0.96	0.97	0.98	0.98
0.50	0.53	0.80	0.88	0.90	0.93	0.96	0.97	0.98	0.98
1.00	0.48	0.78	0.88	0.90	0.94	0.96	0.97	0.98	0.98
1.50	0.36	0.70	0.84	0.88	0.93	0.96	0.97	0.98	0.98
2.00	0.22	0.52	0.68	0.79	0.88	0.94	0.96	0.97	0.98
3.00	0.06	0.12	0.18	0.31	0.55	0.82	0.90	0.94	0.95
4.00	0.02	0.04	0.02	0.03	0.19	0.56	0.77	0.86	0.90
5.00	0.01	0.02	0.02	0.02	0.12	0.37	0.62	0.77	0.83
6.00	0.00	0.01	0.01	0.01	0.07	0.28	0.53	0.69	0.77
7.00	0.00	0.01	0.00	0.00	0.05	0.21	0.44	0.60	0.72
8.00	0.00	0.00	0.00	0.00	0.03	0.15	0.35	0.51	0.65
9.00	0.00	0.00	0.00	0.00	0.02	0.11	0.28	0.42	0.57
10.00	0.00	0.00	0.00	0.00	0.02	0.08	0.22	0.34	0.50

Thus

$$(32) \qquad \hat{L}_t(\mathbf{x}) - \hat{L}^u(\mathbf{x}) = \begin{cases} g(v(\mathbf{x})) - v(\mathbf{x}) + p - 2\gamma, & \text{if } v(\mathbf{x}) \le a, \\ (b^2 - t)/v(\mathbf{x}), & \text{if } v(\mathbf{x}) > a. \end{cases}$$

The above function is decreasing in $v(\mathbf{x})$ when $0 < v(\mathbf{x}) < a$. Hence, for all $\mathbf{x}$,

$$(33) \qquad \hat{L}_t(\mathbf{x}) \ge \hat{L}^u(\mathbf{x}).$$

Taking expectations completes the proof. □

To indicate the amount of improvement that can be expected in using the loss estimator $\hat{L}^{**}$ in (19), instead of $\overline{R}_{\delta_a} = p$, the proportional decrease in communication risk (i.e., $[R_L^*(\theta, p) - R_L^*(\theta, \hat{L}^{**})]/R_L^*(\theta, p))$ is given in Table 1 for $\gamma = 0$, $\boldsymbol{\alpha} = \mathbf{0}$, $a = a_p^*(\gamma)$ and various values of $\|\theta\|$. Clearly, substantial improvement is available for small $\|\theta\|$. Note that one should be careful in making comparisons across dimensions in Table 1. A reasonable rule of thumb in making such comparisons is to first standardize $\|\theta\|$ by dividing by $\sqrt{p}$. (The risks in Table 1 were calculated via Monte Carlo simulation and the standard errors of the entries in Table 1 are 0.01.)

Figure 1, for the case $\gamma = 0$, $\boldsymbol{\alpha} = \mathbf{0}$, $p = 8$ and $a = p - 2$, graphs the risk $R(\theta, \delta_a)$, $E_\theta \hat{L}^{**}(\mathbf{X})$, $E_\theta(p) = p$ and the actual loss estimator $\hat{L}^{**}$. The first three are graphed as functions of $\|\theta\|$, while $\hat{L}^{**}$ is graphed as a function of $\|\mathbf{x}\|$. Of course (5) is satisfied by $\hat{L}^{**}$. Note that one could obtain a closer correspondence between $E_\theta \hat{L}_{\delta_a}(\mathbf{X})$ and $R(\theta, \delta_a)$ if $\hat{L}_{\delta_a}$ were chosen as in (18) with smaller $\underline{L}$ than the default value $\underline{L} = p - a$ used in (19), but we feel that unbiased estimation of $R(\theta, \delta_a)$ is essentially irrelevant. If $\underline{L}$ is too small, a silly estimate of the actual loss could result.

Of considerable interest are admissibility questions concerning loss estimators. Admissibility of δ_a itself is well understood; indeed it is easy to show that δ_a, for $a < p - 2 - \gamma$, can be dominated by $\delta_{(p-2-\gamma)}$. Since the unbiased estimator of risk corresponding to $\delta_{(p-2-\gamma)}$ is given by (31), with $a = b = p - 2 - \gamma$, it might

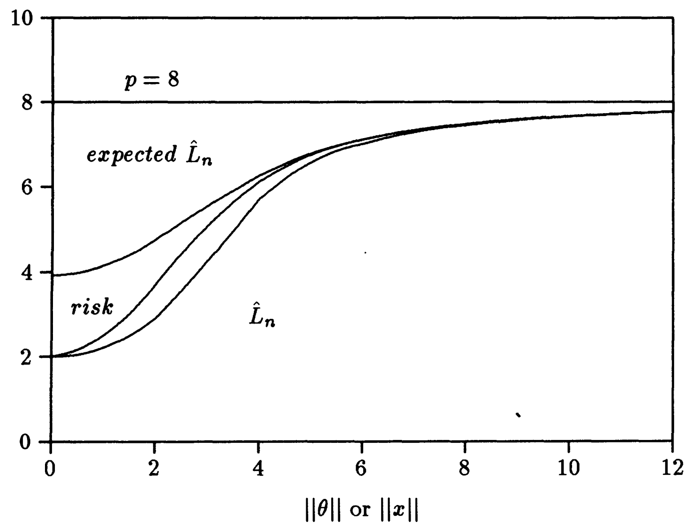

FIG. 1. *Graphs of $\hat{L}^{**}$, its expected value and the risk function.*

seem that only the choice $t = b^2$ in (16) would be of interest. This is not the case, though the following corollary does rule out certain choices of t.

COROLLARY 2.2. *For $t < t_0$,*

$$E_{\boldsymbol{\theta}}\left[\hat{L}_t(\mathbf{X}) - \|\delta_a(\mathbf{X}) - \boldsymbol{\theta}\|^2\right]^2 > E_{\boldsymbol{\theta}}\left[\hat{L}_{t_0}(\mathbf{X}) - \|\delta_a(\mathbf{X}) - \boldsymbol{\theta}\|^2\right]^2 \quad \textit{for all } \boldsymbol{\theta}. \tag{34}$$

PROOF. Since (27) is decreasing in g and (28) has a maximum at $t = t_0$, it follows that

$$\hat{S}_t(\mathbf{x}) < \hat{S}_{t_0}(\mathbf{x}) \quad \text{for all } \mathbf{x}. \tag{35}$$

The result is immediate. □

Comparing the $\hat{L}_t$ for $t_0 \le t \le b^2$ turns out to be more difficult. Indeed, none seem to be dominated by others in the class (those with smaller t doing better for large $\|\boldsymbol{\theta}\|$ and those with large t doing better for small $\|\boldsymbol{\theta}\|$). This is the interesting point and one could conjecture that the $\hat{L}_t$, for $t_0 \le t \le b^2$, are actually all admissible (or nearly so) for suitable choices of g. Furthermore, the choice $t = t_0$ is that which does best "at infinity" (i.e., for large $\|\boldsymbol{\theta}\|$), not the choice $t = b^2$ (which corresponds to the unbiased estimator of risk). While we still feel that $t = b^2$ is a good default choice of t, $t = t_0$ clearly deserves consideration.

It should be noted that a similar result can be obtained for the nonpositive part estimator

$$\delta(\mathbf{x}) = \boldsymbol{\mu}(\mathbf{x}) + \left(1 - \frac{a}{v(\mathbf{x})}\right)(\mathbf{x} - \boldsymbol{\mu}(\mathbf{x})) \tag{36}$$

and associated loss estimator

$$\hat{L}_\delta(\mathbf{x}) = p - t/v(\mathbf{x}), \tag{37}$$

or its positive part. Indeed, (37) for $t = t_0$ can be shown to minimize $R_L^*(\theta, \hat{L}_\delta)$ among all $0 \le t \le b^2$. Johnstone (1987) proved this for the basic James–Stein estimator corresponding to $a = p - 2$.

It would be interesting to determine if any of the $\hat{L}_\delta$ considered here are themselves admissible. We were unable to carry out any admissibility proofs, in part because of the additional difficulties imposed by working subject to the frequentist validity constraint. Also, we have not seriously addressed the issue of selecting from among the $\hat{L}_t$ in (16). In shrinkage estimation, the only method of selecting from among improved procedures that we are comfortable with is to use Bayesian criteria. A partly Bayesian approach is thus considered next.

3. Loss estimation for a generalized Bayes estimator. The estimator in (8) is of the form

$$\delta^n(\mathbf{x}) = \mathbf{x} - \frac{r_n(v(\mathbf{x}))}{\tau^2 v(\mathbf{x})}(\mathbf{x} - \mu(\mathbf{x})), \tag{38}$$

where

$$\mu(\mathbf{x}) = \alpha + B(\mathbf{x} - \alpha), \tag{39}$$

$$v(\mathbf{x}) = (\mathbf{x} - \alpha)^t (I - B)(\mathbf{x} - \alpha)/\tau^2, \tag{40}$$

α is a specified vector, B is an idempotent matrix with rank γ and $\tau \ge 1$ is a constant. [Usually $\tau^2 = 1$ will be chosen, but other values are sometimes desirable. Also, $n = (p - 2 - \gamma)/2$ is the usual choice of n, yielding estimators similar to $\delta_{(p-2-\gamma)}$ in (11).] This estimator is similar to those considered in Strawderman (1971) and is a special case of the general estimator in Berger (1980). The estimator is generalized Bayes with respect to the generalized prior

$$g_n(\theta) = \int_0^1 [\det\{D(\lambda)\}]^{1/2} \exp\{-(\theta - \alpha)^t D(\lambda)(\theta - \alpha)/2\} \times \lambda^{(n-1-p/2)}\, d\lambda, \tag{41}$$

where

$$D(\lambda) = [I - \lambda(I - B)/\tau^2]^{-1} - I. \tag{42}$$

This prior has several interesting interpretations. For a frequentist, this prior can be considered simply as a device to generate a desirable estimator, one which dominates $\delta_0(\mathbf{x}) = \mathbf{x}$ and is admissible [such as when $\gamma = 0$ and $n = (p - 2)/2$]. For a Bayesian, this prior would correspond to a belief that $(\theta - \alpha)^t (I - B)(\theta - \alpha)$ is small to moderate. This belief cannot be quantified exactly in Bayesian terms, since typically used g_n are improper, though Berger (1980) provides some insight into such an interpretation. In any case, it is very natural to utilize this prior to calculate the posterior expected loss for use as a loss estimator. Rather surprisingly, this loss estimator satisfies both Conditions 1 and 2.

Since the original decision loss is sum of squares error, the posterior expected loss of δ^n is simply the sum of the posterior variances. In Berger (1980) it is shown that this is given by [writing $v = v(\mathbf{x})$ for convenience]

$$\hat{L}_n(v) = p - \left[2n + (p - 2 - \gamma - 2n - v)u(v) + vu(v)^2\right]/\tau^2, \tag{43}$$

where $u(v) = r_n(v)/v$. It can be shown (using Lemma A.2 of the Appendix) that $\hat{L}_n(v)$ is strictly positive and monotone with limit p as v goes to infinity.

The following lemma is given in Berger (1980); a version of this lemma was first obtained by Morris (1977).

LEMMA 3.1. *$\hat{L}_n$ satisfies the frequentist validity criterion (Condition 2) when $n \le (p - 2 - \gamma)/2$.*

Thus it remains only to show that $\hat{L}_n$ satisfies (4) and is hence a better report than is $\overline{R}_\delta = p$. The following theorem gives conditions under which this is so. Its proof is long and technical and is relegated to the Appendix.

THEOREM 3.1. *For $p \ge 3$ and for $n = \frac{1}{2}$ or $1 \le n \le (p - 2 - \gamma)/2$, $\hat{L}_n$ satisfies*

$$E_\theta\left[\hat{L}_n(v(\mathbf{X})) - L(\theta, \delta^n(\mathbf{X}))\right]^2 < E_\theta\left[p - L(\theta, \delta^n(\mathbf{X}))\right]^2 \quad \textit{for all } \theta. \tag{44}$$

The gap $\frac{1}{2} < n < 1$ in Theorem 3.1 exists only because of technical difficulties. The theorem is undoubtedly still true in this case.

To indicate the amount of improvement that can be expected in use of the loss estimator $\hat{L}_n(v)$ in (43), instead of $\overline{R}_{\delta^n} = p$, the proportional decrease in communication risk (i.e., $[R_L^*(\theta, p) - R_L^*(\theta, \hat{L}_n)]/R_L^*(\theta, p)$) is given in Table 2 for $\tau = 1$, $\gamma = 0$, $\boldsymbol{\alpha} = \mathbf{0}$, $n = (p - 2)/2$ and various values of $\|\theta\|$. Clearly,

TABLE 2
The proportional decrease in communication risk for $\hat{L}_n$

$\|\theta\|$	$p = 3$	$p = 4$	$p = 5$	$p = 8$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
0.00	0.29	0.58	0.73	0.87	0.91	0.94	0.95	0.96	0.96
0.50	0.30	0.59	0.70	0.85	0.90	0.94	0.95	0.96	0.96
1.00	0.19	0.47	0.69	0.87	0.88	0.94	0.95	0.96	0.97
1.50	0.16	0.33	0.63	0.85	0.90	0.94	0.95	0.96	0.97
2.00	0.05	0.23	0.36	0.76	0.86	0.93	0.96	0.97	0.97
3.00	0.00	0.01	0.22	0.53	0.68	0.88	0.93	0.97	0.97
4.00	0.01	0.04	0.03	0.20	0.31	0.75	0.86	0.92	0.95
5.00	0.01	0.00	0.00	0.01	0.21	0.52	0.77	0.83	0.90
6.00	0.00	0.00	0.02	0.01	0.16	0.27	0.51	0.72	0.81
7.00	0.00	0.01	0.01	0.00	0.05	0.25	0.39	0.54	0.71
8.00	0.00	0.01	0.00	0.01	0.04	0.12	0.35	0.51	0.62
9.00	0.00	0.00	0.00	0.00	0.01	0.10	0.17	0.36	0.57
10.00	0.00	0.00	0.00	0.01	0.02	0.06	0.18	0.30	0.43

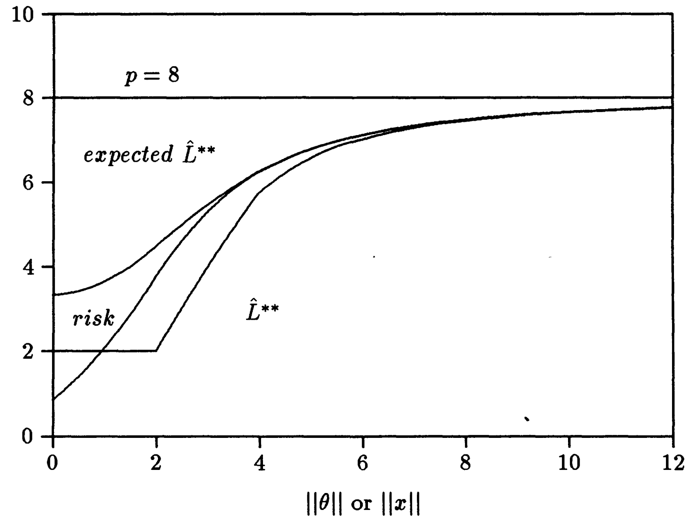

FIG. 2. *Graphs of $\hat{L}_n$, its expected value and the risk function.*

substantial improvement is available for small $||\theta||$. (The standard errors of the entries in Table 2 are 0.01.)

Figure 2, for the case $\tau = 1$, $\gamma = 0$, $\boldsymbol{\alpha} = \mathbf{0}$, $p = 8$ and $n = 3$, graphs the risk $R(\boldsymbol{\theta}, \delta^n)$, $E_{\boldsymbol{\theta}}\hat{L}_n(v(\mathbf{X}))$, $E_{\boldsymbol{\theta}}p = p$ and the actual loss estimator $\hat{L}_n$. The first three are graphed as functions of $||\boldsymbol{\theta}||$, while $\hat{L}_n$ is graphed as a function of $||\mathbf{x}||$. It is somewhat curious that the risk and $\hat{L}_n$ are rather similar, as functions.

4. Conclusions. The choice of $\hat{L}_\delta$ as in (18) or (19) is a bit ad hoc, and possibly conditionally inappropriate for small $||\mathbf{x}||$. The best method of assuring that a conditional error measure (or loss) is sensible is to use a posterior measure. For this reason, we would recommend utilization of δ^n and $\hat{L}_n$ from Section 3, rather than δ_a and $\hat{L}^*$ or $\hat{L}^{**}$ from Section 2. It is truly appealing to be able to report a posterior mean and variance sum, with complete assurance of improved frequentist performance. Note, from Berger (1980), that δ^n and $\hat{L}_n$ can be easily calculated using a recurrence relation.

A comparison of Tables 1 and 2 also seems to indicate superiority of $\hat{L}_n$ in terms of communication risk. For small p, $\hat{L}_n$ offers much greater improvement. Indeed, for $p = 3$ and 4, $\hat{L}^{**}$ is actually *worse* than p, while $\hat{L}_n$ still offers substantial improvement. [Of course, these loss estimators refer to different estimators δ, but δ^n in (38) and δ_a in (11) are actually very similar for $a = 2n = p - 2$, as considered here.]

Although the numerical calculations in Sections 2 and 3 indicate the theoretical advantage in use of these loss estimators, the main practical advantage should be stressed. This is simply that one can report a *sensible* loss $\hat{L}_\delta$ which is (often significantly) smaller than p, the only unconditional frequentist report possible. Thus the increased accuracy available through use of Stein-type estimation can be communicated within the frequentist paradigm.

Acknowledgments. We are grateful to the Associate Editor and referees for suggestions which substantially improved the paper.

APPENDIX

LEMMA A.1. *Suppose* $\mathbf{X}$ *has the* $N_p(\boldsymbol{\theta}, I)$ *distribution and* $\phi(\mathbf{x})$ *is a continuous function with piecewise continuous partial derivatives. Then, for all* $\boldsymbol{\theta}$ *and* $i = 1, \dots, p$,

$$E_{\boldsymbol{\theta}}\phi(\mathbf{X})(X_i - \theta_i) = E_{\boldsymbol{\theta}}\frac{\partial}{\partial X_i}\phi(\mathbf{X}), \tag{45}$$

providing the expectation on the right-hand side of (45) *exists.*

Furthermore, if $\partial\phi(\mathbf{x})/\partial x_i$, $i = 1, \dots, p$, *are continuous and have piecewise continuous partial derivatives and if the expectation on the right-hand side of* (46) *exists, then, for all* $\boldsymbol{\theta}$,

$$E_{\boldsymbol{\theta}}\phi(\mathbf{X})(X_i - \theta_i)^2 = E_{\boldsymbol{\theta}}\left(\frac{\partial^2\phi(\mathbf{X})}{\partial X_i^2} + \phi(\mathbf{X})\right). \tag{46}$$

PROOF. See Stein (1981) and Johnstone (1987). □

LEMMA A.2. *If* $n > 0$ *and* $v > 0$, *then* $r_n(v)$, *as defined in* (9), *satisfies*:

(a) $0 < r_n(v) < 2n$.
(b) $r_n(v)$ *is increasing in* v.
(c) $\lim_{n\to\infty} r_n(2nc)/[2n\min\{1, c\}] = 1$.
(d) $r_n(v)/v$ *is decreasing in* v.
(e) $r_n(v)/v \le n/(n+1)$.
(f) $\lim_{v\to 0} r_n(v)/v = n/(n+1)$.

PROOF. See Berger (1980). □

LEMMA A.3. *For* $u(v) = r_n(v)/v$,

$$u'(v) = \frac{1}{2}u^2(v) - \left(\frac{n+1}{v} + \frac{1}{2}\right)u(v) + \frac{n}{v} \tag{47}$$

and

$$\begin{aligned} u''(v) = {} & \frac{1}{2}u^3(v) - \left[\frac{3(n+1)}{2v} + \frac{3}{4}\right]u^2(v) \\ & + \left[\frac{(n+1)(n+2)}{v^2} + \frac{2n+1}{v} + \frac{1}{4}\right]u(v) - \left[\frac{n(n+2)}{v^2} + \frac{n}{2v}\right]. \end{aligned} \tag{48}$$

PROOF. Follows from formulas (3.6) and (4) of Lemma 3.1.1 in Berger (1980). □

The following lemma provides the basic representation needed to prove Theorem 3.1.

LEMMA A.4. *For all* θ,

$$E_\theta[p - L(\theta, \delta^n)]^2 - E_\theta[\hat{L}_n(v) - L(\theta, \delta^n)]^2 = E_\theta \Delta_n^p(v)/\tau^4, \tag{49}$$

where, defining $q = p - \gamma$,

$$\begin{aligned} \Delta_n^p(v) = {} & \frac{4}{v^2}(q - 2n - 2)(q - 2n - 4)[(n+1)r_n - nv] \\ & + (2n - r_n)(4q - 4 - 2v + 7r_n - 10n) \\ & + \frac{r_n}{v}[(2n - r_n)(-2q - 8r_n + 20n) + 2q^2 - 8nq \\ & \qquad - 12q + 8n^2 + 48n + 16] \\ & + \frac{r_n^2}{v^2}[(2n - r_n)(3r_n - 8 - 10n) + q^2 + 4nq - 8n^2 - 20n - 12]. \end{aligned} \tag{50}$$

In the special case when $n = (p - 2 - \gamma)/2$, *we have* [*letting* $\Delta_n^*(v)$ *denote* $\Delta_n^p(v)$]

$$\begin{aligned} \Delta_n^*(v) = {} & (2n - r_n)[12n - 7(2n - r_n) + 4 - 2v] \\ & + \frac{r_n}{v}[(2n - r_n)(8(2n - r_n) - 4) + 24n] \\ & + \frac{r_n^2}{v^2}[(2n - r_n)(-8 - 4n - 3(2n - r_n)) + 4n^2 - 4n - 8]. \end{aligned} \tag{51}$$

PROOF. Using integration by parts and Lemma A.3, calculation yields

$$\Delta_n^p(v) = d_0 + d_1 u + d_2 u^2 + d_3 u^3 + d_4 u^4, \tag{52}$$

where $u = r_n/v$,

$$\begin{aligned} d_0 &= 8nq - 20n^2 - 8n - 4nv - \frac{4nk}{v}(q - 2n - 4), \\ d_1 &= 2q^2 - 12nq - 12q + 48n^2 + 48n + 16 - 4qv + 24nv + 4v + 2v^2 \\ & \qquad + \frac{4k}{v}(n+1)(q - 2n - 4), \\ d_2 &= q^2 + 4nq - 28n^2 - 36n - 12 + 2qv - 36nv - 7v^2, \\ d_3 &= 16nv + 8v + 8v^2, \\ d_4 &= -3v^2. \end{aligned} \tag{53}$$

Reorganization of (52) yields (50). When $n = (p - 2 - \gamma)/2$, (50) reduces to (51), completing the proof of Lemma A.4. □

PROOF OF THEOREM 3.1 when $n = (p - 2 - \gamma)/2$. Using Lemma A.4, it is clear that Theorem 3.1 will be established in this case if we show that $\Delta_n^*(v) > 0$. This will be established in a series of lemmas covering the various possible situations. For convenience define

$$a_n = \sup\left\{2 \le y < 6: 6 - \frac{7}{2n}(2n - r_n(ny)) > y\right\}, \quad \text{for } n \ge 2, \tag{54}$$

$$\alpha_n(y) = r_n(ny)/ny, \quad \text{for } y > 0, \tag{55}$$

$$A_n(y) = (2n - r_n(ny))/n, \quad \text{for } y > 0. \tag{56}$$

LEMMA A.5. *For a fixed $y > 0$, a_n is increasing in n, $\alpha_n(y)$ is increasing in n and $A_n(y)$ is decreasing in n. For all $a > 0$ and $v > an$, $2n - r_n(v) < A_n(a)n$.*

PROOF. By (4.6) and (4.7) in Berger (1980) we have

$$\alpha_n(y) = \frac{2}{y}\left\{1 - \left(\sum_{i=0}^{\infty} \frac{\Gamma(n+1)n^i}{\Gamma(n+1+i)}\left(\frac{y}{2}\right)^i\right)^{-1}\right\}, \tag{57}$$

which is increasing in n for a fixed y, since

$$\frac{\Gamma(n+1)n^i}{\Gamma(n+1+i)} = \left(\frac{n}{n+i}\right)\left(\frac{n}{n+i-1}\right)\cdots\left(\frac{n}{n+1}\right) \tag{58}$$

is increasing in n. Therefore,

$$6 - \frac{7}{2n}(2n - r_n(ny)) = 6 - 7\left(1 - \frac{y}{2}\alpha_n(y)\right) \tag{59}$$

is increasing in n, a_n is increasing in n and $A_n(y) = 2 - y\alpha_n(y)$ is decreasing in n. Therefore, for $v > an$,

$$2n - r_n(v) \le 2n - r_n(an) \le A_n(a)n. \tag{60}$$

Thus, the lemma is proved. □

LEMMA A.6. *For $n \ge \frac{1}{2}$ and for all $v > 0$, $\Delta_n^*(v) > 0$.*

PROOF. We consider the cases $\frac{1}{2} \le n < 2$, $2 \le n \le 6$ and $n > 6$ separately. For each case, we prove this lemma for $v < 2n$, $2n \le v \le a_n n$ and $v > a_n$ separately.

Suppose $2 \le n \le 6$. By calculating directly, we obtain $5.577 < a_2 < 5.578$. Taking $a = 5.577$ as a bound,

$$A_2(a) \le A_2(a_2) = \frac{2}{7}(6 - a_2) \le 0.1209. \tag{61}$$

By Lemma A.5, for any $n \ge 2$, and for $v > an$,

$$2n - r_n(v) \le A_n(a)n \le A_2(a). \tag{62}$$

From the definition of a_n and (2) of Lemma A.2, we have for $v \ge a_n n$,

$$12n - 7(2n - r_n(v)) \ge 12n - 7(2n - r_n(a_n n)) = 2a_n n. \tag{63}$$

From Lemma A.5, for $n \geq 2$,

$$a_n \geq a_2 \geq a. \tag{64}$$

By (63) and (64), we have for $v \geq an$,

$$12n - 7(2n - r_n(v)) \geq 2an. \tag{65}$$

From (51) and the facts that $r_n(v)/v < 1$, $4n^2 - 4n - 8 \geq 0$ and, for $v > an$,

$$2n - r_n(v) \leq A_n(a) \text{ and } \frac{r_n(v)}{v} \leq \frac{r_n(an)}{an} = \alpha_n(a) \leq \alpha_6(a) \leq \alpha_6(2), \tag{66}$$

it follows that, for $v \geq an$,

$$\begin{aligned} \Delta_n^*(v) &\geq 2(2n - r_n(v))(an + 2 - v) \\ &\quad + \frac{4nr_n(v)}{v}(6 - A_2(a) - (n + 2)\alpha_6(2)A_2(a)). \end{aligned} \tag{67}$$

Let $M = 6 - A_2(a) - 8\alpha_6(2)A_2(a)$. Then, by (61) and $\alpha_6(2) = 0.7102$ we obtain

$$M \geq 6 - 0.1209(1 + 8(0.7102)) = 5.1922. \tag{68}$$

Hence, if $an \leq v \leq an + 2$, then $\Delta_n^*(v) > 0$ for $2 \leq n \leq 6$. Now we only need to consider the case of $v > an + 2$. Since $r_n(v) > 2n - A_2(a)n$, we have for $v > an + 2$ and for $2 \leq n \leq 6$,

$$\Delta_n^*(v) \geq 2(2n - r_n(v))(an + 2 - v) + \frac{4}{v}Mn^2(2 - A_2(a)). \tag{69}$$

Let $y = v/n$. Then, for $2 \leq n \leq 6$,

$$\Delta_n^*(v) > 2(2n - r_n(ny))\left(a + \frac{1}{3} - y\right)n + \frac{4}{y}Mn(2 - A_2(a)). \tag{70}$$

Note that

$$\begin{aligned} \frac{2n - r_n(ny)}{2} &= \left(\int_0^1 \lambda^{n-1} \exp\left(\frac{1 - \lambda}{2}ny\right) d\lambda\right)^{-1} \\ &\leq n\left(1 - \frac{\beta}{n}\right)^{-n} \exp\left(-\frac{\beta y}{2}\right), \end{aligned} \tag{71}$$

where $\beta = 2[1 - 2/(a + \frac{1}{3})]$. Thus, in order to prove that $\Delta_n^*(v) > 0$, it is sufficient to show that

$$y\left(y - a - \frac{1}{3}\right)\exp\left(-\frac{\beta y}{2}\right) < \frac{1}{6}\left(1 - \frac{\beta}{2}\right)^2(2 - A_2(a))M, \tag{72}$$

since $(1 - \beta/n)^{-n}$ is decreasing in n for fixed β. The function

$$y(y - a - \tfrac{1}{3})\exp(-\beta y/2)$$

has the maximum value of 0.0845847 at

$$y = \frac{1}{\beta}\left\{\frac{\beta}{2}\left(a + \frac{1}{3}\right) + 2 + \sqrt{\frac{\beta^2}{4}\left(a + \frac{1}{3}\right) + 4}\right\} = 7.7858997. \tag{73}$$

From (61) and (68), we have

$$0.186215 \le \frac{1}{6}\left(1 - \frac{\beta}{2}\right)^2 (2 - A_2(a))M. \tag{74}$$

Thus, for $2 \le n \le 6$ and for $v \ge an$, $\Delta_n^*(v) > 0$ holds.

In the other cases, the proof is similar. See Lu and Berger (1986) for details. □

Proof of Theorem 3.1 when $n < (p - 2 - \gamma)/2$. Using Lemma A.4, it is clear that Theorem 3.1 will be established if we show that $\Delta_n^p(v) > 0$. This is shown to be true in Lemma A.7; here $m = p - 2n - 2 - \gamma$. We freely use the previous lemmas showing that $\Delta_n^*(v) > 0$.

Lemma A.7. *For $m \ge 0$, for $n \ge \frac{1}{2}$ and for all $v > 0$, $\Delta_n^p(v) > 0$.*

Proof. We consider the cases $m \ge 2$ and $m < 2$ separately. When $m \ge 2$, from (52) and (51) calculation gives

$$\begin{aligned} \Delta_n^p(v) &= \Delta_n^*(v) + \frac{4}{v^2}m(m-2)(n+1)r_n(v) - \frac{4}{v}m(m-2)n \\ &\quad + 4m(2n - r_n(v)) + \frac{r_n(v)}{v}\left[-2m(2n - r_n(v)) + 2m(m-2)\right] \\ &\quad + \frac{r_n(v)^2}{v^2}(m^2 + 8nm + 4m) \\ &\ge \Delta_n^*(v) + \frac{2m(m-2)}{v}\left[\frac{2(n+1)r_n(v)}{v} + r_n(v) - 2n\right]. \end{aligned} \tag{75}$$

Since $\Delta_n^*(v) > 0$, in order to prove that $\Delta_n^p(v) > 0$, we only need to show that

$$\frac{2(n+1)r_n(v)}{v} + r_n(v) - 2n > 0. \tag{76}$$

This is equivalent to

$$G_n(v) > \frac{v}{2n(n+1)} + \frac{1}{n}, \tag{77}$$

where

$$G_n(v) = \frac{2}{2n - r_n(v)} = \int_0^1 \lambda^{n-1} \exp\left(\frac{1-\lambda}{2}v\right) d\lambda. \tag{78}$$

Since, for $k = 1, 2$,

$$\frac{\partial^k}{\partial v^k} G_n(v) = \int_0^1 \lambda^{n-1}\left(\frac{1-\lambda}{2}v\right)^k \exp\left(\frac{1-\lambda}{2}\right) d\lambda \tag{79}$$

and

$$G_n(0) = \frac{1}{n}, \qquad G_n'(0) = \frac{1}{2n(n+1)}, \tag{80}$$

(77) holds. In the other cases, the proof is similar. See Lu and Berger (1986) for details. This proves the lemma. □

REFERENCES

BERGER, J. (1976). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. *J. Multivariate Anal.* **6** 256–264.

BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761.

BERGER, J. (1984). A review of J. Kiefer's work on conditional frequentist statistics. In *The Collected Works of Jack Kiefer* (L. Brown, I. Olkin, J. Sacks and H. Wynn, eds.). Springer, New York.

BERGER, J. (1985a). The frequentist viewpoint and conditioning. In *Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **1** 15–44. Wadsworth, Monterey, Calif.

BERGER, J. (1985b). In defense of the likelihood principle: Axiomatics and coherency. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 33–65. North-Holland, Amsterdam.

BROWN, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Ann. Statist.* **6** 59–71.

BROWNIE, C. and KIEFER, J. (1977). The ideas of conditional confidence in the simplest setting. *Comm. Statist. A—Theory Methods* **6** 691–751.

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

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press.

JOHNSTONE, I. (1987). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.). Springer, New York.

KIEFER, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *J. Amer. Statist. Assoc.* **72** 789–827.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LINDLEY, D. V. (1962). Discussion of "Confidence sets for the mean of a multivariate normal distribution" by C. Stein. *J. Roy. Statist. Soc. Ser. B* **24** 285–287.

LU, K. L. and BERGER, J. O. (1986). Estimation of normal means: The estimated loss frequentist approach. Technical Report 87-16, Dept. Statistics, Purdue Univ.

LU, K. L. and BERGER, J. O. (1988). Estimated confidence procedures for multivariate normal means. *J. Statist. Plann. Inference*. To appear.

MORRIS, C. (1977). Interval estimation for empirical Bayes generalizations of Stein's estimator. In *Proc. 22nd Conf. on the Design of Experiments in Army Research Development and Testing*. Army Research Office Report 77-2.

ROBINSON, G. K. (1979a). Conditional properties of statistical procedures. *Ann. Statist.* **7** 742–755.

ROBINSON, G. K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist.* **7** 756–771.

RUKHIN, A. L. (1987). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.). Springer, New York.

SANDVED, E. (1968). Ancillary statistics and estimation of the loss in estimation problems. *Ann. Math. Statist.* **39** 1756–1758.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

STRAWDERMAN, W. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907