The Annals of Statistics
1985, Vol. 13, No. 4, 1509–1522

TWO-STAGE SEQUENTIAL ESTIMATION OF A MULTIVARIATE NORMAL MEAN UNDER QUADRATIC LOSS[1]

BY JAYALAKSHMI NATARAJAN AND WILLIAM E. STRAWDERMAN

Fairleigh Dickinson University and Rutgers University

In estimating a multivariate normal mean under quadratic loss, this paper looks into the existence of two-stage sequential estimators that are better both in risk (mean square error) and sample size than the usual estimator of a given fixed sample size. In other words, given any sample size n, we are looking for two-stage sequential estimators truncated at n, with a positive probability of stopping earlier and risk lower than that of the sample mean based on n observations. Sequential versions of James–Stein estimators are used to produce two-stage sequential estimators better in risk and sample size than the usual estimator—the sample mean. A lower bound on the largest possible probability of stopping earlier without losing in the risk is also obtained.

1. Introduction. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution of dimension three or more was established by Stein (1955). Suppose X is a $p \times 1$ random vector having the multivariate normal distribution with mean θ and identity covariance matrix, where θ is to be estimated under quadratic loss given by $L(\theta, \delta) = \|\delta - \theta\|^2$. Stein [10] proved that the usual estimator $\delta^0(x) = x$ is inadmissible when $p \geq 3$. James and Stein (1960) showed that estimators of the form

$$\delta^*(x) = \left(1 - \frac{a}{\|x\|^2}\right)x, \quad \text{where} \quad 0 < a < 2(p-2)$$

have lower risks than x. Since then, a considerable amount of work in finding significant improvements upon $\delta^0(x) = x$ in more general settings and under various loss functions has been accomplished by a number of authors. (See References.)

In this study, we are interested in the existence of two-stage sequential estimators that are better both in sample size and risk (expected loss) than the usual estimator of a given fixed sample size. In other words, given any sample size n, we are looking for two-stage sequential estimators truncated at n, with a positive probability of stopping earlier and risk lower than that of the usual estimator based on n observations. To achieve this end, we consider sequential

Received October 1983; revised May 1985.

[1]This paper was delivered at a Statistical Research Conference dedicated to the memory of Jack Kiefer and Jacob Wolfowitz. The Conference was held at Cornell University, July 6–9, 1983, and was a special meeting of the Institute of Mathematical Statistics, co-sponsored by the American Statistical Association. Financial support was provided by the National Science Foundation, the Office of Naval Research, the Army Research Office-Durham, and units of Cornell University. Research was supported by NSF Grant #MCS81-18564-01.

AMS 1980 *subject classifications*. Primary 62F10; secondary 62C99 62H12.

Key words and phrases. Admissibility, James–Stein estimation, sequential estimation.

versions of James–Stein estimators. Similar sequential estimators have been produced for the cases of simultaneous estimation of several Poisson parameters and the estimation of the variance of a normal distribution in Natarajan and Strawderman (1983a, b).

Section 2 deals with the case of identity covariance matrix. Suppose $Y_1 \cdots Y_n$ are $p \times 1$ i.i.d. $N(\theta, I)$ where the unknown θ is to be estimated under quadratic loss given by $L(\theta, \delta) = \|\delta - \theta\|^2$. Let $\overline{X}_i = \sum_{j=1}^{i} Y_{j/i}$ for $i = 1, 2, \dots, n$. The usual estimator based on n observations is $\overline{X}_n$ and has a constant risk p/n. We consider stopping rules based on $\|\overline{X}_m\|$ where $1 \leq m \leq n-1$ and sequential estimators of the form

$$\delta_c^m(Y_1 \cdots Y_n) = \begin{cases} \overline{X}_m, & \text{if } \|\overline{X}_m\| < c, \\ \overline{X}_n - \dfrac{a}{n\|\overline{X}_n\|^2}\overline{X}_n, & \text{otherwise,} \end{cases} \tag{1.1}$$

where $0 < a < 2(p-2)$.

We show that for $p \geq 3$, $n \geq 2$, $1 \leq m \leq n-1$, and $0 < a < 2(p-2)$, there exists $c^m(a) > 0$ such that for all $0 < c \leq c^m(a)$, the corresponding estimators δ_c^m have lower risks than $\overline{X}_n$. Furthermore, we show that when $p \geq 4$, if

$$\frac{n(n-m)}{m^2} \cdot \frac{p^2}{(p-2)^2} < 1$$

and

$$a \in \left[(p-2) - (p-2)\sqrt{1 - \frac{n(n-m)}{m^2}\frac{p^2}{(p-2)^2}},\right.$$
$$\left.(p-2) + (p-2)\sqrt{1 - \frac{n(n-m)}{m^2}\frac{p^2}{(p-2)^2}}\right],$$

the maximum possible value for c, namely $c^m(a)$ is such that when $\theta = 0$, the probability of stopping earlier than n, $P_0(\|\overline{X}_m\| < c^m(a))$ is at least $\frac{1}{2}$. This result implies, for example, that when $p \geq 6$, the sample size can be reduced by at least 25% of the fixed sample size n, with a probability of more than $\frac{1}{2}$ for θ near zero, while still maintaining a lower risk than $\overline{X}_n$. Modifications of (1.1) replacing $\overline{X}_m$ by the corresponding James–Stein estimators are also considered. The case of unknown covariance matrix is dealt with in Remark 3.

The general phenomenon we are studying is the possibility of trading off some of the potential savings in risk of an inadmissible estimator for savings in sample size. Our papers investigating the possibility of similar trade-offs in the classical problems of estimating several Poisson parameters or of estimating a normal variance indicate that such trade-offs are possible in some generality. Our procedures would most likely be useful in a setting where there was a reasonably good prior guess as to the value of $\boldsymbol{\theta}$ and where observations are costly.

2. Results for the case of identity covariance matrix. We consider estimators given by (1.1). The following theorem proves, for all possible values of p, n, m, and a, the existence of an interval of values of c for which δ_c^m has a lower risk than $\overline{X}_n$. This insures us of the existence of sequential estimators we are looking for.

THEOREM 2.1. *For* $p \geq 3$, $n \geq 2$, $1 \leq m \leq n-1$, *and* $0 < a < 2(p-2)$, *there exists* $c^m(a) > 0$ *such that for all* $0 < c < c^m(a)$, δ_c^m *has a lower risk than* $\overline{X}_n$.

PROOF. If δ_n^{JS} denotes the corresponding James–Stein estimator based n observations, it is easy to note that

$$\begin{aligned} R(\theta, \delta_c^m) &\leq P_\theta(\|\overline{X}_m\| < c)(\|\theta\| + c)^2 + R(\theta, \delta_n^{\mathrm{JS}}) \\ &\leq \frac{p}{n} + P_\theta(\|\overline{X}_m\| < c)(\|\theta\| + c)^2 - \frac{2a(p-2) - a^2}{n(p-2+n\|\theta\|^2)}. \end{aligned} \tag{2.1}$$

Since $(n\|\theta\|^2 + p - 2)(\|\theta\| + c)^2 P_\theta(\|\overline{X}_m\| < c)$ tends to zero as $\|\theta\|$ tends to infinity, it is possible to select a $\|\theta\|_0$, such that for all c less than or equal to a preassigned number M and for all θ with $\|\theta\| > \|\theta\|_0$,

$$P_\theta(\|\overline{X}_m\| < c)(\|\theta\| + c)^2 - \frac{2a(p-2) - a^2}{n(p-2+n\|\theta\|^2)} < 0. \tag{2.2}$$

Since $P_\theta(\|\overline{X}_m\| < c)(\|\theta\| + c)^2$ tends to zero as c tends to zero, we can select $c^*(a)$ such that for all θ with $\|\theta\| \leq \|\theta\|_0$ and for $0 < c \leq c^*(a)$, we have

$$\begin{aligned} &P_\theta(\|\overline{X}_m\| < c)(\|\theta\| + c)^2 - \frac{2a(p-2) - a^2}{n(p-2+n\|\theta\|^2)} \\ &\quad \leq P_0(\|\overline{X}_m\| < c)(\|\theta\|_0 + c)^2 - \frac{2a(p-2) - a^2}{n(p-2+n\|\theta\|_0^2)} \\ &\quad < 0. \end{aligned} \tag{2.3}$$

Let $c^m(a) = \min(M, c^*(a))$. Combining (2.1), (2.2), and (2.3) we get that for all $0 < c \leq c^m(a)$, δ_c^m has a lower risk than $\overline{X}_n$. □

Thus, Theorem 2.1 assures us of the existence of a class of two-stage sequential estimators better both in risk and sample size than $\overline{X}_n$. The most interesting among the class of estimators (1.1) are the ones that not only have a lower risk than $\overline{X}_n$ but also give a high probability of estimating θ with as small a number of observations as possible, thus permitting the highest possible reduction in the expected number of observations and hence in the expected cost of taking observations. Of course, the reduction in the sample size is better measured by the ratio of m to n rather than the absolute value of m. In selecting a subclass of the estimators (1.1), two points of view can be taken. First, we can decide on a probability of stopping with m observations and see how small a ratio m/n is possible, while maintaining a risk lower than p/n. Since the probability of stopping earlier than n observations is maximum when $\|\theta\| = 0$ and tends to zero as $\|\theta\|$ tends to infinity, we are more interested in the probability of stopping with m observations only for small values of $\|\theta\|$. It seems reasonable to expect to stop approximately 50% of the time for values of $\|\theta\|$ near zero. This situation is dealt with in Theorem 2.4. On the other hand, we can fix the ratio m/n and find out the maximum possible probability of stopping such that δ_c^m has a risk lower than p/n. One of the interesting situations will be when $m/n = \frac{1}{2}$. Theorem 2.5 looks into this situation.

Before we proceed any further, we need the following two lemmas. The first, Lemma 2.1, expresses the risk of δ_c^m as the expectation of a function of a Poisson variable with parameter $n\|\theta\|^2/2$. Lemma 2.2 obtains a uniform bound for the coefficients of $(n\|\theta\|^2/2)^j/j!(p+2j-2)$ for all $j \geq 0$ in the Poisson expansion of the risk of δ_c^m. The proofs of both lemmas are technical in nature and lengthy. We outline the proofs in the Appendix. Detailed proofs can be found in Natarajan (1983).

LEMMA 2.1. *The risk of* δ_c^m *can be written as*

$$R(\theta, \delta_c^m) = \frac{p}{n} + \exp\left[-\frac{1}{2}n\|\theta\|^2\right] \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}n\|\theta\|^2\right)^j}{j!(p+2j-2)} B_j^m(h), \tag{2.4}$$

where $h = mc^2$ *and* $B_j^m(h)$ *is a function of* j, h, m, *and* n *such that* $nB_j^m(h)$ *depends on* m *and* n *only through* m/n. *The exact expression for* $B_j^m(h)$ *is very long and hence given in Appendix* (A.1).

LEMMA 2.2. *Following the notations in Lemma* 2.1, *for* $p \geq 4$ *and for all* $j \geq 0$,

$$B_j^m(h) \leq t_m(h), \tag{2.5}$$

where

$$t_m(h) = \frac{1}{n}\left[\frac{n(n-m)}{m^2} phE1_{(\chi_p^2 < h)} - \left(2a(p-2) - a^2\right)\left(1 - E1_{(\chi_p^2 < h)}\right)\right]. \tag{2.6}$$

Before we proceed with results on the probabilities of stopping earlier, we give the following theorem which compares the risks of δ_c^m for different pairs of values of (n, m) where n/m is a constant. To avoid the confusion caused in the notation, we denote (in the following theorem only) the estimators given in (1.1) by $\delta_c^{m,n}$ rather than just δ_c^m.

THEOREM 2.2. *Let* (n_1, m_1) *and* (n_2, m_2) *be two sets of positive integers such that* $n_1/m_1 = n_2/m_2$. *For any* $h > 0$, *define* c_1 *and* c_2 *by* $h = m_1c_1^2 = m_2c_2^2$.

Then for $p \geq 3$, $\delta_{c_1}^{m_1, n_1}$ has a risk lower than p/n_1 if and only if $\delta_{c_2}^{m_2, n_2}$ has a risk lower than p/n_2.

PROOF. Let l be such that $n_2 = n_1 l$ and $m_2 = m_1 l$. For $\theta = (\|\theta\|, 0, \dots, 0)$, define $\theta_l = (\|\theta_l\|, 0, \dots, 0)$ by $\|\theta_l\|^2 = l\|\theta\|^2$. By Lemma 2.1, $nB_j^{m,n}(h)$ depends only on n/m. Hence we have that $n_1 B_j^{m_1, n_1}(h) = n_2 B_j^{m_2, n_2}(h)$. From (2.4),

$$\begin{aligned} R\left(\theta, \delta_{c_2}^{m_2, n_2}\right) - \frac{p}{n_2} &= \frac{\exp\left[-\frac{1}{2} n_2 \|\theta\|^2\right]}{n_2} \sum_0^\infty \frac{\left(n_2 \|\theta\|^2/2\right)^j}{j!(p + 2j - 2)} n_2 B_j^{m_2, n_2}(h) \\ &= \frac{\exp\left[-\frac{1}{2} n_1 l \|\theta\|^2\right]}{n_1 l} \sum_0^\infty \frac{\left(n_1 l \|\theta\|^2/2\right)^j}{j!(p + 2j - 2)} n_1 B_j^{m_1, n_1}(h) \\ &= \frac{\exp\left[-\frac{1}{2} n_1 \|\theta_l\|^2\right]}{n_1 l} \sum_0^\infty \frac{\left(n_1 \|\theta_l\|^2/2\right)^j}{j!(p + 2j - 2)} n_1 B_j^{m_1, n_1}(h) \\ &= \frac{1}{l}\left[R\left(\theta_l, \delta_{c_1}^{m_1, n_1}\right) - \frac{p}{n_1}\right] \end{aligned}$$

and the result follows. □

The above theorem implies that the largest possible probability of stopping earlier without losing in the risk is the same for all values of (m, n) such that n/m is a constant.

The following theorem gives a sufficient condition for the risk of δ_c^m to be lower than p/n. This sufficient condition is used in subsequent theorems that involve the largest possible probability of stopping earlier.

THEOREM 2.3. *Let $t_m(h)$ be defined as in (2.6). Then $t_m(h_0) \leq 0$ for some $h_0 > 0$ implies that δ_c^m has a lower risk than $\overline{X}_n$ for all c such that $mc^2 \leq h_0$.*

PROOF. $t_m(h)$ is a continuous increasing function of h with $t_m(0) \leq 0$. Hence $t_m(h_0) \leq 0$ for some $h_0 > 0$ implies, by way of Lemma 2.2, that $B_j^m(h) \leq t_m(h) \leq t_m(h_0)$ for all $j \geq 0$ and all $h \leq h_0$ and Eq. (2.4) helps us conclude that $R(\theta, \delta_c^m) - R(\theta, \overline{X}_n) = R(\theta, \delta_c^m) - p/n \leq 0$ for all c such that $mc^2 \leq h_0$. □

Now we go into theorems involving the probability of stopping earlier. Given p and n, for any $1 \leq m \leq n - 1$ and $0 < a < 2(p - 2)$, let

$$\begin{aligned} h_a^m &= h_a^m(p, n) \\ &= \max\left\{h = mc^2 \text{ such that } \delta_c^m \text{ has a lower risk than } \overline{X}_n\right\}. \end{aligned} \tag{2.7}$$

h_a^m corresponds to the estimator of the form (1.1) with the largest possible probability of stopping earlier than n, without losing in the risk. Suppose we are interested in stopping with m observations in at least 50% of the times, when

$\theta = 0$. Given p and n, we are interested in m and a such that the corresponding value of h_a^m gives for $\theta = 0$ at least 50% probability of stopping with m observations, that is $P(\chi_p^2 < h_a^m) \geq \frac{1}{2}$. Theorem 2.4 gives for $p \geq 4$, a sufficient condition on the ratio n/m and an interval of values for a in order to achieve $P(\chi_p^2 < h_a^m) \geq \frac{1}{2}$. This theorem implies that when $p \geq 6$, the number of observations can be reduced by at least 25% with a probability of greater than or equal to one-half for $\theta = 0$.

THEOREM 2.4. *For* $p \geq 4$, $n \geq 3$, $1 \leq m \leq n - 1$ *such that*

$$\frac{n}{m}\left(\frac{n}{m} - 1\right)\frac{p^2}{(p-2)^2} < 1 \tag{2.8}$$

and

$$a \in \left[p - 2 - (p-2)\sqrt{1 - \frac{n(n-m)}{m^2}\left(\frac{p}{p-2}\right)^2},\right. \\ \left. p - 2 + (p-2)\sqrt{1 - \frac{n(n-m)}{m^2}\left(\frac{p}{p-2}\right)^2}\right], \tag{2.9}$$

h_a^m *as defined by* (2.7) *is such that* $P(\chi_p^2 < h_a^m) \geq \frac{1}{2}$.

PROOF. Let h_0 be the median of χ_p^2 distribution. Then $h_0 \leq p$. From the definition of $t_m(h)$ given in (2.6),

$$t_m(h_0) = \frac{1}{2n}\left\{\frac{n(n-m)}{m^2}ph_0 - \left(2a(p-2) - a^2\right)\right\}$$
$$\leq \frac{1}{2n}\left\{\frac{n(n-m)}{m^2}p^2 - \left(2a(p-2) - a^2\right)\right\},$$

which is less than or equal to zero if a lies between the roots of the equation $a^2 - 2a(p-2) + (n(n-m)/m^2)p^2 = 0$, that is, if a belongs to the interval specified in (2.9), and this interval is nonempty if and only if (2.8) is satisfied. Hence if (2.8) and (2.9) are satisfied, we have that $t_m(h_0) \leq 0$, which in turn (using Theorem 2.3) implies that δ_c^m has a risk lower than p/n for all c such that $mc^2 \leq h_0$. The definition of h_a^m allows us to conclude that $h_a^m \geq h_0$ or $P(\chi_p^2 < h_a^m) \geq P(\chi_p^2 < h_0) = \frac{1}{2}$.

On the other hand, given the ratio n/m, the question of how large a probability of stopping is possible without making the risk larger than p/n is answered partially in the following theorem. We show that for large p, we can reduce the number of observations by 50% with a probability of at least 0.3 when $\theta = 0$. □

Before we state the theorem, we define $h_{\max}$ to be a real number such that

$$\frac{h_{\max} E1_{(\chi_p^2 < h_{\max})}}{1 - E1_{(\chi_p^2 < h_{\max})}} = \frac{(p-2)^2}{2p}. \tag{2.10}$$

Since $hE1_{(\chi_p^2 < h\max)}/(1 - E1_{(\chi_p^2 < h_{\max})})$ is a continuous, increasing function of h with

$$\frac{hE1_{(\chi_p^2 < h)}}{1 - E1_{(\chi_p^2 < h)}} \underset{h \to 0}{\longrightarrow} 0 \quad \text{and} \quad \frac{hE1_{(\chi_p^2 < h)}}{1 - E1_{(\chi_p^2 < h)}} \uparrow \infty \quad \text{as } h \uparrow \infty,$$

$h_{\max}$ is well defined and unique.

Theorem 2.5. *For $p \geq 4$, $n/m = 2$, and $a = p - 2$, we have $h_a^m \geq h_{\max}$ where $h_{\max}$ is defined by* (2.10). *Also, we have*

$$P\left(\chi_p^2 < h_{\max}\right) \geq \frac{(p-2)^2/2p^2}{1 + (p-2)^2/2p^2}.$$

Proof. When $p \geq 4$, $n/m = 2$, and $a = p - 2$, from (2.6), we have,

$$t_m(h) = \frac{1}{n}\left[2phE1_{(\chi_p^2 < h)} - (p-2)^2\left(1 - E1_{(\chi_p^2 < h)}\right)\right].$$

By definition of $h_{\max}$, we have that $t_m(h_{\max}) = 0$ and hence by Theorem 2.3, we can conclude that δ_c^m has a risk lower than p/n for all $0 < c < (h_{\max}/m)^{1/2}$, which implies that $h_a^m \geq h_{\max}$.

To prove the rest of the theorem, we observe that when $h = p$,

$$\begin{aligned}
\frac{2phE1_{(\chi_p^2 < h)}}{1 - E1_{(\chi_p^2 < h)}} &= \frac{2phE1_{(\chi_p^2 < p)}}{1 - E1_{(\chi_p^2 < p)}} \\
&> 2p^2 \\
&> (p-2)^2 \\
\Rightarrow \text{(by definition of } h_{\max}) \; p &> h_{\max} \\
\Rightarrow \frac{E1_{(\chi_p^2 < h_{\max})}}{1 - E1_{(\chi_p^2 < h_{\max})}} &= \frac{(p-2)^2}{2ph_{\max}} \\
&\geq (p-2)^2/2p^2 \\
\Rightarrow E1_{(\chi_p^2 < h_{\max})} &\geq \frac{(p-2)^2/2p^2}{1 + (p-2)^2/2p^2},
\end{aligned}$$

which concludes the theorem. □

In the following remarks, we consider various modifications of estimators (1.1).

Remark 1. Estimators (1.1) can be modified by replacing $\overline{X}_m$ by James–Stein estimators based on m observations. For $p \geq 3$, $n \geq 2$, $1 \leq m \leq n - 1$, we consider estimators of the form

$$(2.11)\qquad \delta_c^{bm}(Y_1 \cdots Y_n) = \begin{cases} \overline{X}_m - \dfrac{b}{m\|\overline{X}_m\|^2}\overline{X}_m, & \text{if } \|\overline{X}_m\| < c, \\ \overline{X}_n - \dfrac{b}{n\|\overline{X}_n\|^2}\overline{X}_n, & \text{otherwise,} \end{cases}$$

where $0 < a < 2(p-2)$ and $0 \le b \le 2(p-2)$. □

In a proof analogous to that of Theorem 2.1, it can be shown that for all possible values of p, n, m, a, and b, there exists an interval of values for c such that δ_c^{bm} is better than $\overline{X}_n$.

REMARK 2. The estimators δ_c^m of (1.1) are considerably better than $\overline{X}_n$ in both risk and the probability of stopping earlier than n, only for small values of $\|\theta\|$. If we have reason to believe that the true parameter value is near θ^*, we could consider the following translations of the original estimators, namely,

$$(2.12)\qquad \delta_{c,\theta^*}^m = \begin{cases} \overline{X}_m, & \text{if } \|\overline{X}_m - \theta^*\| < c, \\ \overline{X}_n - \dfrac{a(\overline{X}_n - \theta^*)}{n\|\overline{X}_n - \theta^*\|^2}, & \text{otherwise.} \end{cases}$$

All the results of this section, with suitable modifications, work just as well for these estimators.

REMARK 3 (unknown covariance matrix Σ).

(a) *When* $\Sigma = \sigma^2 I$ *where* σ^2 *is unknown*. Suppose the loss in estimating θ is given by

$$L((\theta, \sigma^2)) = \|\delta - \theta\|^2/\sigma^2.$$

For $i = 1, 2, \ldots, n$, as before, let $\overline{X}_i$ denote the sample mean based on the first i observations. For each $= 1, 2, \ldots, n$, we also have a random variable S_i^2 distributed as $\sigma^2\chi^2_{p(i-1)}$, independent of $\overline{X}_i$. For $p \ge 3$, $n \ge 3$ we consider estimators of the form

$$(2.13)\qquad \psi_c^m(Y_1 \cdots Y_n) = \begin{cases} \overline{X}_m, & \text{if } \|\overline{X}_m\|^2/S_m^2 < c^2, \\ \overline{X}_n - \dfrac{a}{n\|\overline{X}_n\|^2/S_n^2}\overline{X}_n, & \text{otherwise,} \end{cases}$$

where $2 \le m \le n-1$ and $0 < a < 2(p-2)/(p(n-1)+2)$.

The proof of Theorem 2.1 can be modified slightly to show the existence of an interval of values for c such that ψ_c^m has a lower risk than $\overline{X}_n$.

(b) *General, unknown* Σ. Suppose the loss involved in estimating θ by δ is given by

$$L((\theta, \Sigma), \delta) = (\delta - \theta)'\Sigma^{-1}(\delta - \theta).$$

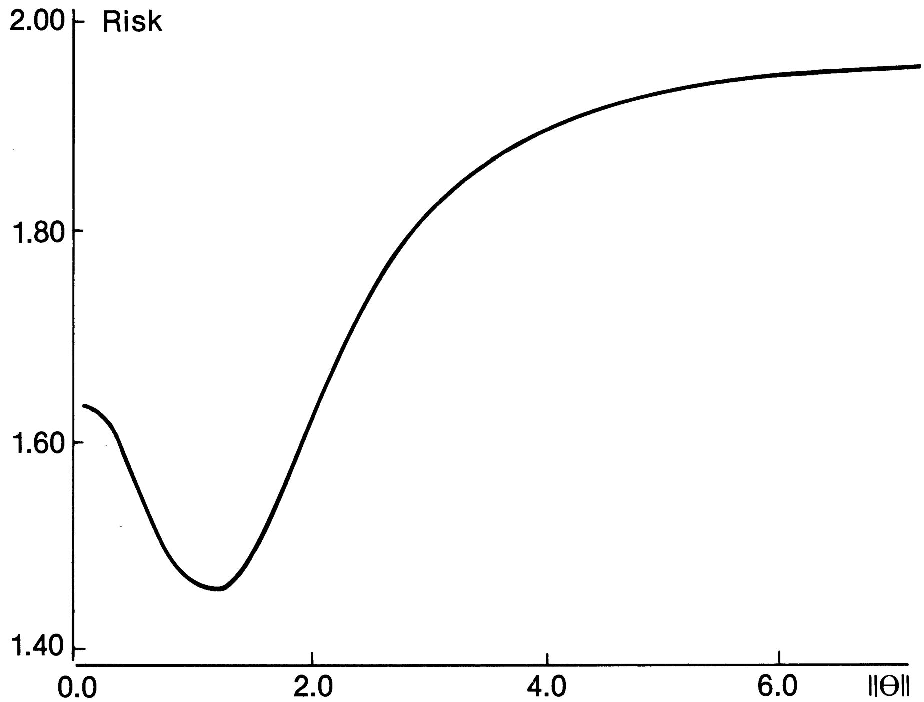

FIG. 1. *Risk of* δ_c^m *when* $p = 6$, $n = 3$, $m = 2$, $a = 4$, $h = p$.

For each $i = 1, 2, \dots, n$, let $\overline{X}_i$ denote the sample mean based on the first i observations. For each $i = 1, 2, \dots, n$, we also have a $p \times p$ Wishart matrix S_i independent of $\overline{X}_i$ with $E(S_i) = (i - 1)\Sigma$.

For $p \geq 3$, $n \geq p + 2$, and $p + 1 \leq m \leq n - 1$ we consider estimators of the form

$$\psi_c^{*m}(Y_1 \cdots Y_n) = \begin{cases} \overline{X}_m, & \text{if } \overline{X}_m' S_m^{-1} \overline{X}_m < c^2, \\ \overline{X}_n - \dfrac{a}{n\overline{X}_n' S_n^{-1} \overline{X}_n} \overline{X}_n, & \text{otherwise,} \end{cases} \tag{2.14}$$

where $0 < a < 2(p - 2)/(n - p + 2)$.

Existence of an interval of values for c such that ψ_c^{*m} has a lower risk than $\overline{X}_n$ can be shown in a manner similar to the proof of Theorem 2.1.

REMARK 4. The choice of a preliminary sample size m affects the probability of early stopping as well as the mean squared error. If such a procedure were to be used in practice it would seem desirable to choose c (for any given m) as large as possible, such that the mean squared error stays bounded by that of the

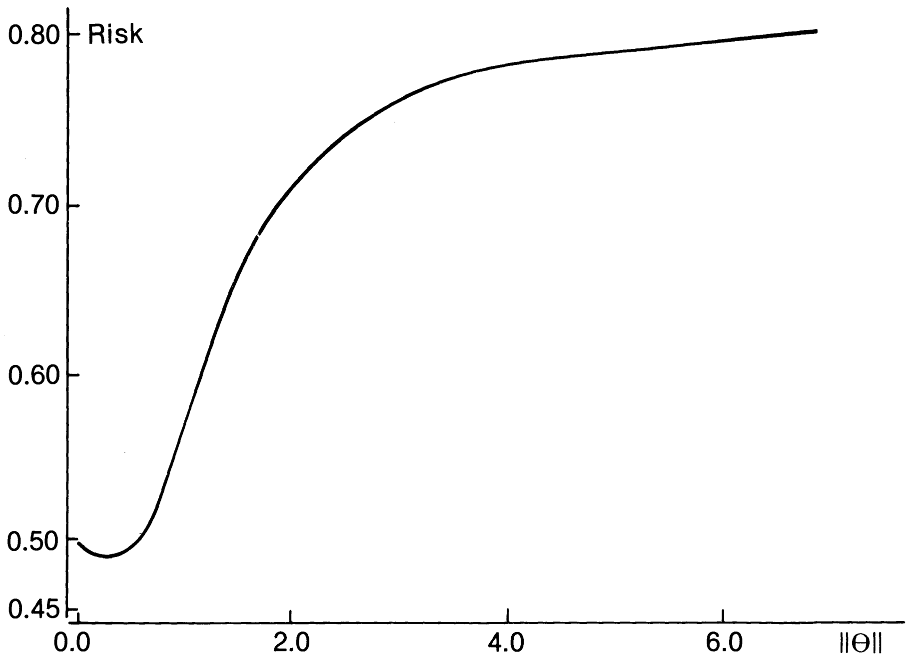

FIG. 2. *Risk of* δ_c^m *when* $p = 8$, $n = 10$, $m = 5$, $a = 6$, $h = p - 0.5\sqrt{2p} = 6$.

classical procedure. A reasonable way to choose m would then be to minimize the expected sample size for $\boldsymbol{\theta} = \mathbf{0}$.

3. Numerical results. Appearing in Figures 1 and 2 are the graphs of the risks of δ_c^m for two sets of values of p, n, m, a, and h not covered by the theorems of Section 2. By using Eq. (A.5) of the Appendix, the expression for $B_j^m(h)$ given by (A.1) can be reduced to expectations involving functions of binomial variables only. Given h, m, p, n, and a, $B_j^m(h)$ were computed for a number of values of j and then expression (2.4) for the risk was used to calculate the same for various values of $\|\theta\|$.

The graph in Figure 1 represents the risk of δ_c^m for various values of $\|\theta\|$ when $p = 6$, $n = 3$, $m = 2$, $h = p$, and $a = p - 2$. The graph indicates that the risk of δ_c^m stays below the value of p/n and hence, by Theorem 2.2, the result holds for all pairs (m, n) such that $m/n = \frac{2}{3}$. Note that for this set of values of p, n, and m, the condition (2.8) of Theorem 2.4 is not satisfied but yet the result of Theorem 2.4 holds.

Figure 2 deals with the situation when $m = n/2$. For $p = 8$ and $a = p - 2$, h is taken to be equal to $p - 0.5\sqrt{2p}$. The graph shows that the risk of δ_c^m when $n = 10$ and $m = 5$ stays below the value of p/n. Even though this case is covered by Theorem 2.5, the probability of stopping earlier when $h = p - 0.5\sqrt{2p}$ is much higher than the value given by the theorem.

APPENDIX

The outlines of the proofs of Lemmas 2.1 and 2.2 are given in this Appendix. Detailed proofs can be found in Natarajan (1983). Lemma 2.1 is restated here to include the expression for $B_j^m(h)$.

LEMMA 2.1. *The risk of* δ_c^m *can be written in the form given by* (2.4) *where* $h = mc^2$ *and for* $j \geq 0$,

$$B_j^m(h) = (p + 2j - 2)\Bigg[\sum_{k=0}^{j} \binom{j}{k} {\frac{m}{n}}^{k} \left(\frac{n-m}{n}\right)^{k-k}$$

$$\cdot \left\{ \frac{p(n-m)}{nm} E1_{(\chi^2_{p+2k} < h)} + \left(\frac{n^2 - m^2}{n^2 m}\right)\right.$$

$$\text{(A.1)} \qquad \left. \cdot \frac{e^{-h/2}(h/2)^{p/2+k-1}}{\Gamma(p/2+k)}(2k - h) + \frac{2a}{n} E1_{(\chi^2_{p+2k} < h)} \right\}$$

$$- \left(\frac{a^2 + 4aj}{n}\right)\left(\frac{n-m}{n}\right)^{p/2+j-1} \frac{1}{p + 2j - 2}$$

$$\cdot \left\{ \sum_{k=0}^{\infty} \frac{\omega^m_{j-1,k}}{k!} E1_{(\chi^2_{p+2k} < hn/(n-m))} \right\}\Bigg] - \frac{a(2(p-2) - a)}{n}$$

with

(A.2)

$$w^m_{j-1,k} = \begin{cases} (p + 2j - 2)(p + 2j) \cdots (p + 2j + 2k - 4)\left(\dfrac{m}{2n}\right)^k, & \text{if } k \geq 1, \\ 1, & \text{if } k = 0 \end{cases}$$

and for any given h, $nB_j^m(h)$ *depends on* n *and* m *only through* n/m.

PROOF OF LEMMA 2.1. The expression (2.4) for the risk of δ_c^m can be obtained by first separating the terms involved into two groups, one with the terms involving $\overline{X}_m$ only and the other containing terms involving both $\overline{X}_m$ and $\overline{X}_n$. Terms involving $\overline{X}_m$ only can be handled using the fact that $\|\overline{X}_m\|^2$ is noncentral χ_p^2 with noncentrality parameter $m\|\theta\|^2/2$. The terms involving both $\overline{X}_m$ and $\overline{X}_n$ can be dealt with by considering first the conditional expectation of $\overline{X}_m$ given $\overline{X}_n$ and then using the fact that $\|\overline{X}_n\|^2$ is noncentral χ_p^2 with noncentrality parameter $n\|\theta\|^2/2$. A change of the order of summation yields the required expression (2.4).

Proof of Lemma 2.2. $P_\theta(\|\overline{X}_m\|^2 < h/m)$ can be expressed as an expectation of a function of a Poisson variable with parameter $n\|\theta\|^2/2$ in two different ways. First, using the fact that $\overline{X}_m|\overline{X}_n \sim N(\overline{X}_n, ((n-m)/nm)I)$, and $\overline{X}_n \sim N(\theta, (1/N)I)$, we get that

$$P_\theta\left(\|\overline{X}_m\|^2 < \frac{h}{m}\right) = E_\theta E\left\{1_{(\|\overline{X}_m\|^2 < h/m)} \mid \|\overline{X}_n\|\right\}$$

$$\text{(A.3)} \qquad = \exp\left[-\tfrac{1}{2}n\|\theta\|^2\right] \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}n\|\theta\|^2\right)^j}{j!} \left(\frac{n-m}{n}\right)^{p/2+j} \cdot \sum_{k=0}^{\infty} \frac{w_{jk}^m}{k!} E1_{(\chi^2_{p+2} < hn/(n-m))},$$

where w_{jk}^m is given by (A.2). On the other hand, $\overline{X}_m \sim N(\theta, (1/m)I)$ gives

$$P_\theta\left(\|\overline{X}_m\|^2 < h/m\right) = \exp\left[-\tfrac{1}{2}m\|\theta\|^2\right] \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}m\|\theta\|^2\right)^k}{k!} E1_{(\chi^2_{p+2k} < h)}$$

$$\text{(A.4)} \qquad = \exp\left[-\tfrac{1}{2}n\|\theta\|^2\right] \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}n\|\theta\|^2\right)^j}{j!} \cdot \sum_{k=0}^{j} \binom{j}{k} \left(\frac{n-m}{n}\right)^{j-k} E1_{(\chi^2_{p+2k} < h)}.$$

Comparing the corresponding coefficients in (A.3) and (A.4), we get

$$\text{(A.5)} \qquad \left(\frac{n-m}{n}\right)^{p/2+j} \sum_{k=0}^{\infty} \frac{w_{jk}^m}{k!} P\left(\chi^2_{p+2k} < \frac{hn}{n-m}\right) = \sum_{k=0}^{j} \binom{j}{k} \left(\frac{m}{n}\right)^k \left(\frac{n-m}{n}\right)^{j-k} P\left(\chi^2_{p+2k} < h\right).$$

Also, for each $j \geq 0$,

$$\left\{\left(\frac{n-m}{n}\right)^{p/2+j-1} \frac{w_{j-1,k}^m}{k!}\right\}_{k=0,1,2,\dots}$$

can be considered as the probability densities of a discrete random variable since

$$\sum_{k=0}^{\infty} \left(\frac{n-m}{n}\right)^{p/2+j-1} \frac{w_{j-1,k}^m}{k!} = 1$$

and, with j as a parameter, this family is MLR. Hence we have for all $j \geq 0$,

$$\text{(A.6)}\qquad \begin{aligned} &\left(\frac{n-m}{n}\right)^{p/2+j-1} \sum_{k=0}^{\infty} \frac{w_{j-1,k}^m}{k!} P\left(\chi_{p+2k}^2 < \frac{hn}{n-m}\right) \\ &\geq \left(\frac{n-m}{n}\right)^{p/2+j} \sum_{k=0}^{\infty} \frac{w_j^m}{k!} P\left(\chi_{p+2k}^2 < \frac{hn}{n-m}\right). \end{aligned}$$

Given $h > 0$, let k_0 be the smallest integer such that $2k_0 - h \geq 0$. Let Δ denote $(n-m)p/nm$. Define a function $g(k)$ on the set of nonnegative integers as follows:

$$\text{(A.7)}\qquad g(k) = \begin{cases} \Delta E 1_{(\chi_{p+2}^2 < h)}, & \text{if } k = 0, \\ \Delta E 1_{(\chi_{p+2k}^2 < h)}, & \text{if } 1 \leq k < k_0, \\ \Delta E 1_{(\chi_{p+2k}^2 < h)} + (2k - h) & \\ \quad \cdot \dfrac{(n^2 - m^2)}{n^2 m} 2 f_{p+2k}(h), & \text{if } k \geq k_0. \end{cases}$$

where $f_{p+2k}(h)$ is the p.d.f. of a χ^2 distribution with $p + 2k$ d.f. Using (A.5), (A.6), and (A.7) in the expression (A.1) for $B_j^m(h)$ and denoting expectations with regard to binomial $(j, m/n)$ distribution by $E_j(\cdot)$, we have for all $j \geq 0$,

$$\text{(A.8)}\qquad \begin{aligned} B_j^m(h) &\leq (p + 2j - 2) E_j g(k) + \frac{2a(p-2) - a^2}{n}\left(E 1_{(\chi_p^2 < h)} - 1\right) \\ &\leq \frac{n}{m} \sum_{k=1}^{j+1} \binom{j+1}{k} \left(\frac{m}{n}\right)^k \left(\frac{n-m}{n}\right)^{j+1-k} (p - 4 + 2k) g(k-1) \\ &\quad + \frac{2a(p-2) - a^2}{n}\left(E 1_{(\chi_p^2 < h)} - 1\right). \end{aligned}$$

If we define $\Delta = (n-m)p/nm$ and $s(k)$ by

$$\text{(A.9)}\qquad s(k) = \begin{cases} \Delta E 1_{(\chi_p^2 < h)}, & \text{if } k = 1, \\ \Delta E 1_{(\chi_{p+2k-4}^2 < h)}, & \text{if } 2 \leq k \leq k_0 \\ \Delta E 1_{(\chi_{p+2k-4}^2 < h)} + (2(k-1) - h) & \\ \quad \cdot \left(\dfrac{n^2 - m^2}{n^2 m}\right) 2 f_{p+2k-4}(h), & \text{if } k \geq k_0 + 1, \end{cases}$$

it can be verified in a straightforward way that $s(k)$ is a decreasing function of k

and

$$(p + 2k - 4)g(k - 1) \le hs(k) \le hs(1), \quad \text{for all } k \ge 1. \tag{A.10}$$

Using (A.10) in (A.8), we get

$$\begin{aligned} B_j^m(h) &\le \frac{n}{m} hs(1) \sum_{k=1}^{j+1} \binom{j+1}{k} \left(\frac{m}{n}\right)^k \left(\frac{n-m}{n}\right)^{j+1-k} \\ &\quad + \frac{2a(p-2) - a^2}{n} \left(E1_{(\chi_p^2 < h)} - 1\right) \\ &\le \frac{n}{m} hs(1) - \frac{2a(p-2) - a^2}{n} \left(1 - E1_{(\chi_p^2 < h)}\right) \\ &= t_m(h) \end{aligned} \tag{A.11}$$

as defined in (2.6), completing the proof.

Acknowledgment. We wish to thank the referees for their comments and suggestions that improved this paper.

REFERENCES

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

BERGER, J., BOCK, M. E., BROWN, L. D., CASELLA, G. and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771.

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

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

GHOSH, M., SINHA, B. K. and MUKHOPADHYAY, N. (1976). Multivariate sequential point estimation. *J. Multivariate Anal.* **6** 281–294.

JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symposium Math. Stat. Prob.* **1** 361–379. Univ. California Press.

NATARAJAN, J. (1983). Sequential James–Stein estimation and related results. Ph.D. thesis, Rutgers University.

NATARAJAN, J. and STRAWDERMAN, W. E. (1983a). Simultaneous sequential estimation of several Poisson parameters. Unpublished manuscript.

NATARAJAN, J. and STRAWDERMAN, W. E. (1983b). Sequential estimation of the variance of a normal distribution. Unpublished manuscript.

STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symposium Math. Stat. Prob.* **1** 197–206. Univ. California Press.

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

STRAWDERMAN, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *Ann. Statist.* **1** 1189–1194.

COLLEGE OF BUSINESS ADMINISTRATION
FAIRLEIGH DICKINSON UNIVERSITY
MADISON, NJ 07940

STATISTICS DEPARTMENT
RUTGERS UNIVERSITY
NEW BRUNSWICK, NJ 08903