The Annals of Statistics
1995, Vol. 23, No. 2, 593–597

A NOTE ON ADMISSIBILITY WHEN PRECISION IS UNBOUNDED

BY CHARLES ANDERSON AND NABENDU PAL

University of Southwestern Louisiana

The estimation of a common mean vector θ given two independent normal observations $X \sim N_p(\theta, \sigma_x^2 I)$ and $Y \sim N_p(\theta, \sigma_y^2 I)$ is reconsidered. It being known that the estimator $\eta X + (1 - \eta)Y$ is inadmissible when $\eta \in (0, 1)$, we show that when η is 0 or 1, then the opposite is true, that is, the estimator is admissible. The general situation is that an estimator X^* can be improved by shrinkage when there exists a statistic B which, in a certain sense, estimates a lower bound on the risk of X^*. On the other hand, an estimator is admissible under very general conditions if there is no reasonable way to detect that its risk is small.

We will prove, among other things, the following theorem:

THEOREM 1. *Assume independent normal observations $X \sim N_p(\theta, \sigma_x^2 I)$ and $Y \sim N_p(\theta, \sigma_y^2 I)$ are the only observables and suppose θ is to be estimated with squared Euclidean norm error. Then X is an admissible estimator.*

This result may appear surprising in view of the fact, proven in George (1991) and Krishnamoorthy (1992), that for any fixed $\eta \in (0,1)$, $\hat{\theta} = \eta X + (1 - \eta)Y$ is inadmissible. It may also be noted that under the conditions of Theorem 1, an unbiased estimator of σ_x^2 is given by

$$\hat{\sigma}_x^2 = p^{-1} X'(X - Y),$$

a fact which may lead one to doubt the validity of Theorem 1 since the results of George and Krishnamoorthy indicate that a shrinkage estimator which improves on X can sometimes be constructed using a variance estimator which is not independent of X. Hence a careful derivation of Theorem 1 will be given.

The reason X is admissible is basically that its precision cannot be bounded. To emphasize this, we note that a careful analysis of the main argument used by George in establishing the dominance of shrinkage estimators shows that what is required for such an argument is not an estimate of σ_x^2, but rather a well-behaved statistic representing a probable lower bound for this quantity. In Theorem 2, we attempt to clarify the situation by stating a simple pair of sufficient conditions for use of such a bound to produce a dominating shrinkage estimate. We may note immediately, however, that the

Received April 1993; revised July 1994.
AMS 1991 *subject classifications*. Primary 62C15; secondary 62H12.
Key words and phrases. Inadmissibility, shrinkage estimation, Stein's normal identity.

unbiased estimate mentioned above is not necessarily positive and hence does not satisfy the required conditions.

We begin with an easy general result. In the more general setup, X^* is an estimator of $\theta \in \mathbb{R}^p$ and T is an $\mathbb{R}^q$-valued statistic representing all other data. It is not assumed that T is independent of X^* or that X^* is normal.

LEMMA 1. *Suppose that for every $\theta_0 \in \mathbb{R}^p$ and for every open set V in $\mathbb{R}^q$ it is possible to vary all parameters so that $\theta \to \theta_0$, $E[\|X^* - \theta\|^2] \to 0$ and so some version of $P(T \in V | X = x)$ is uniformly bounded away from 0. Then no continuous function of X^* and T dominates X^*.*

PROOF. If $g(X^*, T)$ were such a function, then one could find $\varepsilon > 0$ and open sets U, V so that when $\theta \in U$, $X^* \in U$ and $T \in V$, then $\|g(X^*, T) - X^*\| > 2\varepsilon$ and $\|X^* - \theta\| < \varepsilon$. We could vary the parameters so the probability of all this happening at once was bounded below, causing $E\|g(X^*, t) - \theta\|^2$ to be bounded away from 0 while $E\|X^* - \theta\|^2 \to 0$, and this would contradict the assumption that $g(X^*, T)$ dominates X^*. □

It is very easy to verify that the hypotheses outlined in Theorem 1 imply the hypotheses of Lemma 1 with $X^* = X$ and $T = Y$. More difficult is the task of showing that, under the conditions of Lemma 1, admissibility in the class of continuous functions of X and Y, the conclusion of Lemma 1, is equivalent to admissibility (in the class of measurable functions of X and Y). Since we are dealing now with an exponential family of distributions for (X, Y), we might be able to adapt the results of Brown (1986), where it is shown that generalized Bayes estimators of the canonical parameter are continuous and, in the Appendix, that the family of generalized Bayes estimators is essentially complete. Another possibility is to generalize the proof, suggested in a remark in Stein (1955), of the essential completeness of continuous functions of X when X is assumed to have the distribution specified in Theorem 1 and when only X is observed. We will adopt this latter approach.

PROOF OF THEOREM 1. Suppose $g(X, Y)$ is an estimator of θ which dominates X, with g being measurable but possibly not continuous. Since the problem is invariant under translations and orthogonal linear transformations acting simultaneously on X and Y, it is clear that for fixed $A \in \mathbb{R}^p$ and $Q \in O(p)$ (the set of $p \times p$ orthogonal matrices), X is also dominated by

$$g^*_{A,Q}(X, Y) = Q^{-1}[g(QX + A, QY + A) - A].$$

Let A and Q then be chosen randomly, independently of X and Y and with a smooth joint distribution. Then

$$h(X, Y) = E[g^*_{A,Q}(X, Y) | X, Y]$$

will be a smooth function of X and Y and will also dominate X. Then Lemma 1 applies. □

ALTERNATIVE PROOF OF THEOREM 1. We can also prove Theorem 1 without appealing to the symmetries of the problem by making use of Lebesgue's metric density theorem [see Section III of Hobson (1927) or Section 12 of Dunford and Schwartz (1958), for example]. Suppose $g(X, Y)$ dominates X, g being some Borel measurable function. Then there exists $\varepsilon > 0$ such that $\|g(x, y) - x\| > \varepsilon$ for all (x, y) in a Borel set M having positive Lebesgue measure $\lambda(M)$. Without loss of generality, let M be bounded and assume there is a positive lower bound for the Lebesgue measure of each cross section

$$M_x = \{y \in \mathbb{R}^p : (x, y) \in M\}$$

with x in

$$N = \{x \in \mathbb{R}^p : M_x \neq \varnothing\}.$$

Then

$$b = \inf\{P(Y \in M_x) : \theta \in N,\ x \in N,\ \sigma_y^2 = 1\}$$

is also positive. Let $\theta \in N$ be a point at which the metric density of N is 1, meaning that if K is a cube of vanishingly small diameter centered at θ, then $\lambda(K \cap N) \sim \lambda(K)$. It follows that as $\sigma_x^2 \to 0$, while θ and $\sigma_y^2 = 1$ remain fixed, we have $P(X \in N) \to 1$ and $\liminf P((X, Y) \in M) \geq b$. Thus the risk of $g(X, Y)$ is eventually greater than $b\varepsilon^2$ while the risk of X approaches 0, contradicting our assumptions. □

Now we will try to clarify the difference between the case $\eta \in \{0, 1\}$ and the case $\eta \in (0, 1)$ when considering the estimator $\hat{\theta} = \eta X + (1 - \eta)Y$. Abstracting from the arguments used by George (1991), we find that the key to making a successful improvement by shrinkage is obtaining a probable lower bound on the variance of the original estimator, in a sense spelled out technically in the following theorem, which therefore represents a partial converse to Theorem 1.

THEOREM 2. *Suppose $\theta \in \mathbb{R}^p$ is to be estimated by some measurable function of X^* and T, the loss of $\hat{\theta}$ being $\|\hat{\theta} - \theta\|^2$. Assume that X^* has the multivariate normal distribution centered at θ with unknown scalar nonzero covariance matrix $\sigma^2 I$, assume $p \geq 3$ and assume, finally, that B is a smooth nonnegative function of X^* and T satisfying*

(1) $E[B^2/\|X^*\|^2] \leq K\sigma^2 E[B/\|X^*\|^2]$ *for some known constant* K,

(2) $E[X^{*\prime}W/\|X^*\|^2] \geq 0$ *where* $W = \nabla E(B|X^*)$.

Then, for any constant $c \in (0, 2(p-2)/K)$, the risk of $\hat{\theta}^0 = X^*$ is larger than the risk of $\hat{\theta}^c = X^*[1 - cB/\|X^*\|^2]$ for all values of θ and σ^2.

PROOF. Using Stein's normal identity and proceeding as in George (1991), the risk difference $R(\hat{\theta}^c, \theta) - R(\hat{\theta}^0, \theta)$ is seen to be

$$c^2 E(B^2/\|X^*\|^2) - 2cK\sigma^2(p-2)E(B/\|X^*\|^2) - 4c\sigma^2 E(X^{*\prime}W/\|X^*\|^2) \geq 0.$$

If B is a known positive constant, then condition (2) holds trivially but condition (1) holds only if B can be taken to provide a known bound on σ^2. More generally, if B is merely independent of X^*, then condition (2) is still trivial, because $W \equiv 0$, and condition (1) reduces to the condition

$$\sigma^2 \geq K^{-1}E(B^2)/E(B),$$

which says in a sense that B is underestimating $K\sigma^2$. □

Note that the conditions of the theorem are especially simple when B is independent of X^* and distributed as chi-squared times a constant. Examples covered by this case may be found in a host of papers, for example, in Baranchik (1970), Strawderman (1971) and Lin and Tsai (1973).

Our Theorem 2 merely extracts the main argument from George's Theorem 2.1. Thus if we take X^* and B to be, in his notation, δ_η and $S = \|X - Y\|^2$, then the conditions of the theorem are, in his notation, (1) $A_1/A_2 \leq K$ and (2) $A_3 \geq 0$. One notes that the difficult part of George's paper is the proof of our condition (1). The slightly stronger result of Krishnamoorthy (1991) does not follow from our Theorem 2.

Using Theorem 2, one can actually generalize George's result slightly by taking $B = SH$ where H is any nonnegative, nondecreasing function of $\|X^*\|^2$ which can be bounded away from 0 and ∞. Then condition (1) with $B = SH$ is equivalent to condition (1) with $B = S$, which was proved by George in his Lemma 3.1. Condition (2) can be established as follows: First note that the conditional expectation of S given X^* is of the form $\alpha + \beta Z'Z$, where α and β are nonnegative functions of the unknown parameters and $Z = X^* - \theta$. Then $\nabla E[B|X^*] = \beta(2HZ + Z'Z\,\nabla H)$. The inner product of $Z'Z\,\nabla H$ with $U = X^*/\|X^*\|^2$ is never negative while, by the normal identity, the inner product of HZ with U has the same expectation as the divergence of HU, which is easily seen to be nonnegative.

Acknowledgments. Our thanks to a referee and an Associate Editor for suggesting some useful changes; also to K. Krishnamoorthy, Calvin Berry, Bimal K. Sinha, and Bill Strawderman for some helpful conversations. The referee helped in an important way to make the thrust of the paper clearer.

REFERENCES

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

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory.* IMS, Hayward, CA.

DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators Part 1: General Theory.* Interscience, New York.

GEORGE, E. I. (1991). Shrinkage domination in a multivariate common mean problem. *Ann. Statist.* **19** 952–960.

HOBSON, E. W. (1927). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series.* Dover, New York.

KRISHNAMOORTHY, K. (1992). On a shrinkage estimator of a normal common mean vector. *J. Multivariate Anal.* **40** 109–114.

LIN, P. E. and TSAI, H. L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.* **1** 142–145.

STEIN, C. (1955). Inadmissibility of the usual estimator for the mean vector of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF SOUTHWESTERN LOUISIANA
LAFAYETTE, LOUISIANA 70504-1006