The Annals of Statistics
1976, Vol. 4, No. 3, 629-638

AN IMPROVED ESTIMATOR OF THE GENERALIZED VARIANCE

BY R. W. SHORROCK AND J. V. ZIDEK

Université de Montréal and University of British Columbia

A multivariate extension is made of Stein's result (1964) on the estimation of the normal variance. Here the generalized variance $|\Sigma|$ is being estimated from a Wishart random matrix $S: p \times p \sim W(n, \Sigma)$ and an independent normal random matrix $X: p \times k \sim N(\xi, \Sigma \otimes 1_k)$ with ξ unknown. The main result is that the minimax, best affine equivariant estimator $((n+2-p)!/(n+2)!)|S|$ is dominated by $\min\{((n+2-p)!/(n+2)!)|S|, ((n+k+2-p)!/(n+k+2)!)|S+XX'|\}$. It is obtained by a variant of Stein's method which exploits zonal polynomials.

1. Introduction and summary. Consider a multivariate normal linear model in canonical form. A minimal sufficient statistic is (X, S) where X is a normally distributed $p \times k$ matrix with independent columns $X_i \sim N(\xi_i, \Sigma)$, S is a $p \times p$ Wishart matrix with n degrees of freedom with $ES = n\Sigma$, and X and S are independent. We assume that Σ is known to be positive definite (so $|\Sigma| > 0$).

Consider the problem of estimating the determinant $|\Sigma|$ of Σ with the quadratic loss function

$$L\{\phi(X, S); \Sigma, \xi\} = |\Sigma|^{-2}\{\phi(X, S) - |\Sigma|\}^2 . \tag{1.1}$$

Similar results will obtain for the fairly large class of "bowl-shaped" loss functions introduced in Brown (1968).

The problem is invariant under the transformations:

$$\begin{aligned} X &\to AX + B, & S &\to ASA', \\ \xi &\to A\xi + B, & \xi &\to A\Sigma A', \end{aligned} \tag{1.2}$$

where A is any nonsingular $p \times p$ matrix and B is any $p \times k$ matrix. Estimators equivariant under this affine group satisfy

$$\phi(AX + B, ASA') = |A|^2\phi(X, S)$$

and have the form $\phi(S) = c|S|$ where c is a constant. Such estimators have constant risk (expected loss) which is minimized by taking $c = (n-p+2)!/(n+2)!$. We thus obtain the best affine equivariant estimator, $c|S|$, which Selliah (1964) shows is minimax relative to the class of all estimators that depend on S alone. We do not know whether $c|S|$ is admissible relative to this class but Stein (1964) shows, for the case $p = 1$ ($|\Sigma| = \sigma^2$), that $c|S|$ is not admissible relative to the class of all estimators based on the sufficient statistic (X, S) and he exhibits a better estimator that uses X as well as S.

Received October 1974; revised November 1975.

AMS 1970 *subject classifications.* Primary 62F10; Secondary 62H99.

Key words and phrases. Equivariant, multivariate normal matrix, noncentral Wishart matrix, zonal polynomials.

In Section 3 the counterpart of Stein's result for general p is obtained by showing that the estimator

$$\phi(X, S) = \min\left\{\frac{(n+2-p)!}{(n+2)!}|S|, \frac{(n+k+2-p)!}{(n+k+2)!}|S+XX'|\right\} \tag{1.3}$$

has uniformly smaller risk than $c|S|$. The case $k = 1$ is presented in Section 2 because the argument is much simpler.

2. A superior alternative to $c|S|$—one unknown mean. For the case $k = 1$ of the problem stated in Section 1, $|\Sigma|$ is to be estimated (with loss given by equation (1.1)) from (X, S), where X and S are independent, $X \simeq N(\xi, \Sigma)$, and $S \simeq W_p(n, \Sigma)$. Among estimators equivariant for the action of the affine group displayed in equation (1.2), Selliah's solution $c|S|$ was seen to be best.

It is shown below that a superior alternative to $c|S|$ can be found by searching in a larger class than the affine-equivariant estimators. In creating a larger class in which to search, we are guided by the concluding remarks in Stein (1964) and, in fact, the class we create will be a p-variate analogue of that of Stein for $p = 1$. Given a problem invariant under a large group G for which a best equivariant estimator exists, Stein suggests seeking a better estimator by looking among estimators equivariant under a nonnormal subgroup H.

We implement Stein's suggestion by considering estimators which are equivariant with respect to the subgroup H (of the affine group) whose action is described by

$$\begin{aligned} X &\to AX, & S &\to ASA' \\ \xi &\to AX, & \Sigma &\to A\Sigma A' \end{aligned} \tag{2.1}$$

where A is any nonsingular $p \times p$ real matrix. The equivariant estimators are defined as those functions ϕ satisfying

$$\phi(AX, ASA') = |A|^2\phi(X, S), \tag{2.2}$$

for all $p \times p$ nonsingular real matrices A. A standard argument shows that $\phi(X, S)$ is equivariant $\Leftrightarrow \phi(X, S) = \psi(X'S^{-1}X)|S + XX'|$ for some function ψ.

To facilitate the derivation of the required distribution theory write

$$\begin{aligned} X &= AU, & \xi &= A\xi_0 \\ S &= AWA', & \Sigma &= A1A' \end{aligned} \tag{2.3}$$

where $U \simeq N(\xi_0, 1)$ and $W \simeq W_p(n, 1)$, and $\xi_0' = (\lambda^{\frac{1}{2}}, 0, \cdots, 0)$. Then the risk function of any equivariant estimator $\phi(X, S) = \psi(X'S^{-1}X)|S + XX'|$ takes the form

$$R_\psi(\lambda) = E^\lambda\{\psi(T)|W + UU'| - 1\}^2 \tag{2.4}$$

where $T = U'W^{-1}U$. Let 0 be a random orthogonal matrix such that $\tilde{U} = 0U = ||\tilde{U}||(1, 0, \cdots, 0)'$, and set $\tilde{W} = 0W0'$. Then $||\tilde{U}||$ and $\tilde{W}$ are independent with $||\tilde{U}||^2 \simeq X_p'^2(\lambda)$ and $\tilde{W} \simeq W_p(n, 1)$ and $T = U'W^{-1}U = \tilde{U}'\tilde{W}^{-1}\tilde{U}$, i.e.

$$T = ||\tilde{U}||^2/(\tilde{W}_{11} - \tilde{W}_{12}\tilde{W}_{22}^{-1}\tilde{W}_{21}), \tag{2.5}$$

and $|W + UU'| = |\tilde{W} + \tilde{U}\tilde{U}'| = (1 + T)|\tilde{W}| = (1 + T)(\tilde{W}_{11} - \tilde{W}_{12}\tilde{W}_{22}^{-1}\tilde{W}_{21})|\tilde{W}_{22}|$, where $\tilde{W}$ is partitioned as

$$\tilde{W} = \begin{bmatrix} \tilde{W}_{11} & \tilde{W}_{12} \\ \tilde{W}_{21} & \tilde{W}_{22} \end{bmatrix} \tag{2.6}$$

and $\tilde{W}_{11}$ is a 1×1 matrix. Then standard distribution theory says

$$\begin{aligned} u &= \|\tilde{U}\|^2 \simeq \chi^2_{p+2\kappa} \\ v &= \tilde{W}_{11} - \tilde{W}_{12}\tilde{W}_{22}^{-1}\tilde{W}_{21} \simeq \chi^2_{n-p+1}, \\ w &= |\tilde{W}_{22}| \simeq \textstyle\prod_{i=1}^{p-1} \chi^2_{n-i+1} \end{aligned} \qquad \text{and} \tag{2.7}$$

where κ follows a Poisson law with $E\kappa = \lambda/2$ and the χ^2 random variables above are all conditionally independent given κ. Writing

$$R_\psi(\lambda) = E^\lambda E^{(u/v,\kappa)}\{\psi(u/v)(u + v)w - 1\}^2 \tag{2.8}$$

it may be seen that the inner conditional expectation is minimized by taking $\psi = \psi_\kappa$ where

$$\begin{aligned} \psi_\kappa(u/v) &= (n - p + 3)!/[(n + 2)!\,(n + 3 + 2\kappa)] \\ &\leqq (n - p + 3)!/(n + 3)!\,. \end{aligned} \tag{2.9}$$

It follows that for any estimator of the form $\phi(X, S) = \psi(X'S^{-1}X)|S + XX'|$, the estimator $\phi^*(X, S) = \min\{\phi(X, S), (n - p + 3)!\,|S + XX'|/(n + 3)!\}$ is as good as ϕ and strictly better than ϕ unless $\phi = \phi^*$ with probability one. For Selliah's estimator $c|S|$ a strictly better estimator is thus

$$\phi^*(X, S) = \min\left\{\frac{(n - p + 2)!}{(n + 2)!}|S|, \frac{(n - p + 3)!}{(n + 3)!}|S + XX'|\right\}. \tag{2.10}$$

3. A superior alternative to $c|S|$—the general case. We proceed as in Section 2 and first derive a class of equivariant estimators in which to search for an alternative to $c|S|$.

In Section 1 $c|S|$ is derived as the best affine-equivariant estimator. However, in order to follow Stein's suggestion in the general case we have found it convenient to choose a larger group, with respect to which $c|S|$ is also best equivariant, and then choose a subgroup of that larger group. To be precise, the problem outlined in Section 1, where

$$\begin{aligned} &X \sim N(\xi, \Sigma \otimes 1)\,, \qquad S \sim W_p(n, \Sigma)\,, \\ &\qquad X \text{ and } S \text{ are independent}, \\ &\qquad L(\phi; \Sigma, \xi) = |\Sigma|^{-2}(\phi - |\Sigma|)^2\,, \quad n \geqq p\,, \end{aligned} \tag{3.1}$$

is invariant under the direct product $\mathscr{G}$ of an affine group on the sample space of X and the orthogonal group on R^k which acts on the problem as follows:

$$\begin{aligned} X &\to (AX + B)0'\,, \qquad S \to ASA'\,, \\ \xi &\to (A\xi + B)0'\,, \qquad \Sigma \to A\Sigma A'\,, \end{aligned} \tag{3.2}$$

where A is nonsingular $p \times p$, 0 is orthogonal $k \times k$ and B is any $p \times k$ matrix. The subgroup $\mathscr{H}$ of $\mathscr{G}$ obtained by setting $B = 0$ in (3.2) is a nonnormal subgroup of $\mathscr{G}$.

The groups $\mathscr{G}$ and $\mathscr{H}$ are p-variate analogues of those considered by Stein (1964) and by analogy we consider estimators $\phi(X, S)$ of $|\Sigma|$ that satisfy

$$\phi(AX0', ASA') = |A|^2\phi(X, S) \tag{3.3}$$

for all $X, S, A, 0$. Setting $A = 1$ we see that $\phi(X0', S) = \phi(X, S)$ for all orthogonal 0 so that $\phi(X, S) = \psi(XX', S)$ depends on X only through XX'. Also,

$$\phi(AX, ASA') = |A|^2\phi(X, S) \Leftrightarrow \psi(AXX'A', ASA') = |A|^2\psi(XX', S)\,,$$

and by choosing A such that $A(S + XX')A' = 1$ and $AXX'A' = T$ where T is a diagonal matrix with nonnegative diagonal elements $t_1 \geqq t_2 \geqq \cdots \geqq t_p$ we conclude that

$$\begin{aligned}\phi(X, S) &= \psi(XX', S)\\ &= \psi(T, 1 - T)|S + XX'|\\ &= \psi(T)|S + XX'|\,;\end{aligned}$$

conversely, every estimator of this form is $\mathscr{H}$-equivariant as is easily seen.

We remark that the diagonal elements of T are the roots of the determinantal equation $|XX' - t(S + XX')| = 0$ and the preceding argument essentially shows that T is a maximal invariant for the operation of $\mathscr{H}$ on the space of (X, S). Also, since the ratio of any two nonzero functions ϕ and ϕ_0 satisfying (3.3) is constant on orbits of $\mathscr{H}$ and so is a function of T alone, we always have $\phi(X, S) = \psi_0(T)\phi_0(X, S)$ for some function ψ_0. The choice $\phi_0(X, S) = |S + XX'|$ to represent the class of $\mathscr{H}$-equivariant estimators seems most convenient for our analysis.

Since we are restricting our considerations to the class of $\mathscr{H}$-equivariant estimators, we need only compare the risk functions of such estimators and invariance allows us to reduce the problem as follows. Write

$$\begin{aligned} S &= AWA'\,, & \Sigma &= AA' \\ X &= AU0'\,, & \xi &= A\xi_0 0' \end{aligned} \tag{3.4}$$

where $\xi_0\xi_0' = \Lambda$ is a diagonal matrix with diagonal elements $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p =$ the roots of the determinantal equations

$$|\xi\xi' - \lambda\Sigma| = 0 \Leftrightarrow |\xi_0\xi_0' - \lambda 1| = 0\,.$$

Then Λ is a maximal invariant for the operation of $\mathscr{H}$ on the parameter space and for any $\mathscr{H}$-equivariant estimator $\phi(X, S) = \psi(T)|S + XX'|$ we have

$$\begin{aligned} R_\phi(\Sigma, \xi) &= E\{\phi(X, S) - |\Sigma|\}^2|\Sigma|^{-2} \\ &= E\{\psi(T)|W + UU'| - 1\}^2 = R_\psi(\Lambda)\,, \end{aligned} \tag{3.5}$$

where T is a diagonal matrix with diagonal elements $=$ the roots of the equation

$|UU' - t(UU' + W)| = 0$ and

$$W \sim W_p(n, 1_p), \qquad U \sim N(\xi_0, 1_p \bar{\otimes} 1_k), \qquad \text{and} \qquad \Lambda = \xi_0 \xi_0' . \tag{3.6}$$

The problem of comparing $\mathscr{H}$-equivariant estimators $\phi(X, S)$ of $|\Sigma|$ on the basis of their risk functions $R_\phi(\Sigma, \xi)$ is thus equivalent to comparing functions $\psi(T)$ on the basis of $R_\psi(\Lambda)$.

We shall show, by a multidimensional analogue of the device of conditioning on an auxiliary Poisson variable κ used in Section 2 for one unknown mean, that for any ψ, ψ^* defined by

$$\psi^*(T) = \min\{\psi(T), \psi_0(T)\} \tag{3.7}$$

satisfies $R_{\psi^*}(\Lambda) \leqq R_\psi(\Lambda)$ for all Λ, the inequality being strict provided $\psi^*(T) \neq \psi(T)$ with positive probability.

Let V be a diagonal matrix with diagonal elements = the latent roots of $W + UU'$. Then from (3.5) we have

$$R_\psi(\Lambda) = E\{\psi(T)|V| - 1\}^2$$

for any ψ. To express R_ψ in a convenient form we first derive a representation for the noncentral ($\Lambda \neq 0$) law of (T, V) as a mixture of central (i.e., parameter free) laws. Following James (1964) we let $\kappa = \{\kappa_1 \geqq \kappa_2 \geqq \cdots \geqq \kappa_p\}$ range over all ordered sequences of p nonnegative integers and consider the zonal polynomials $C_\kappa(\Omega)$ normalized so that for all $k = 0, 1, 2, \cdots$

$$(\operatorname{tr}\Omega)^k = \textstyle\sum_{\{\kappa : ||\kappa|| = k\}} C_\kappa(\Omega), \qquad ||\kappa|| = \sum_{i=1}^p \kappa_i, \quad \operatorname{tr} = \text{trace}. \tag{3.8}$$

The $C_\kappa(\Omega)$ are homogeneous symmetric polynomials in the latent roots of the $p \times p$ matrix Ω and have the property, essential for our main result, of being nonnegative on the space of nonnegative definite matrices [see James (1968)].

Let $L = L(T, V)$ be any (measurable) function of $T = T(U, W)$, $V = V(U, W)$, and let E^Λ denote expectation assuming (3.6) with E^0 in the special case $\Lambda = 0 \Leftrightarrow \xi_0 = 0$. Then L, as a function of (U, W) is invariant under the transformations

$$U \to 0U\hat{0}, \qquad W \to 0W0'$$

where 0 is $p \times p$ orthogonal and $\hat{0}$ is $k \times k$ orthogonal, and we have

$$\begin{aligned} E^\Lambda L &= \exp(-\tfrac{1}{2}\operatorname{tr}\Lambda) E^0\{L \exp(\operatorname{tr}\xi_0' U)\} \\ &= \exp(-\tfrac{1}{2}\operatorname{tr}\Lambda) \textstyle\sum_\kappa E^0\{L C_\kappa(\Lambda \cdot UU')\}\alpha_\kappa \\ &= \exp(-\tfrac{1}{2}\operatorname{tr}\Lambda) \textstyle\sum_\kappa \dfrac{C_\kappa(\frac{1}{2}\Lambda)}{||\kappa||!} \dfrac{E^0\{L C_\kappa(UU')\}}{E^0\{C_\kappa(UU')\}} \end{aligned} \tag{3.9}$$

where α_κ is a coefficient whose precise form may be found in James (1964). The first equality follows at once from the forms of the central ($\xi_0 = 0$) and noncentral ($\xi_0 \neq 0$) densities of U. The second follows from the invariance of L by letting $U \to U\hat{0}$ and integrating with respect to the invariant probability measure $\hat{\lambda}$ on the group of $k \times k$ orthogonal matrices $\hat{0}$, using the formula

[James (1964)]

$$\int \exp\{\operatorname{tr} \xi' U \hat{0}\} \, d\hat{\lambda}(\hat{0}) = \textstyle\sum_\kappa \alpha_\kappa C_\kappa(\xi\xi' UU') . \tag{3.10}$$

The third equality follows by letting $W \to 0W0'$ and $U \to 0U$ and using the formula [James (1964)]

$$\int C_\kappa(\Omega_1 0 \Omega_2 0') \, d\lambda(0) = C_\kappa(\Omega_1) C_\kappa(\Omega_2) / C_\kappa(1_p) , \tag{3.11}$$

where λ is the invariant probability measure on the group of $p \times p$ orthogonal matrices, together with the relation

$$E^0 C_\kappa(UU') = C_\kappa(\tfrac{1}{2} 1_p) / (\alpha_\kappa ||\kappa||!) \tag{3.12}$$

which follows by comparing equations (24) and (27) in James (1964).

Equation (3.9) is a representation of the noncentral law of (T, V) as a mixture but is inconvenient because of the presence of UU'. To obtain a representation directly in terms of (T, V) we first observe that if we make the transformation

$$\begin{aligned} \tilde{U} &= (W + UU')^{-\frac{1}{2}} U , & U &= \tilde{W}^{\frac{1}{2}} \tilde{U} \\ \tilde{W} &= W + UU' & W &= \tilde{W}^{\frac{1}{2}} (1 - \tilde{U}\tilde{U}')^{\frac{1}{2}} \tilde{W}^{\frac{1}{2}} \end{aligned} \tag{3.13}$$

then $\tilde{U}$ and $\tilde{W}$ are independent when $\xi_0 = 0$, and $\tilde{W} \sim W_p(n + k, 1_p)$. This can be seen by checking that the joint density of $(\tilde{U}, \tilde{W})$ factors into the $W_p(n+k, 1_p)$ density over the space $\{\tilde{W}: \tilde{W} \geqq 0\}$ and a density of the form $c|1 - \tilde{U}\tilde{U}'|^{\frac{1}{2}(n-p-1)}$ with respect to Lebesgue measure on the space $\{\tilde{U}: 0 \leqq \tilde{U}\tilde{U}' \leqq 1_p\}$; the Jacobian $\partial(U, W)/\partial(\tilde{U}, \tilde{V})$ is seen to be $|\tilde{W}|^{\frac{1}{2}k}$ by considering the sequence of mappings

$$\begin{bmatrix} U \\ W \end{bmatrix} \to \begin{bmatrix} U \\ W + UU' \end{bmatrix} \to \begin{bmatrix} (W + UU')^{-\frac{1}{2}} U \\ W + UU' \end{bmatrix} , \tag{3.14}$$

the first of which has Jacobian $= 1$.

Since the diagonal elements of T are the latent roots of the matrix $\tilde{U}\tilde{U}'$ and the diagonal elements of V are the roots of $\tilde{W}$, we see that T and V are also independent when $\xi_0 = 0$ and that $L = L(T, V)$ can also be thought of as a function of $(\tilde{U}, \tilde{W})$ that is invariant under maps $\tilde{W} \to 0\tilde{W}0'$ where 0 is $p \times p$ orthogonal. Such maps also leave the law of $\tilde{W}$ invariant, so we can write, using equation (3.11)

$$\begin{aligned} E^0\{LC_\kappa(UU')\} &= E^0\{LC_\kappa(\tilde{W}^{\frac{1}{2}} \tilde{U}\tilde{U}' \tilde{W}^{\frac{1}{2}})\} \\ &= E^0\{LC_\kappa(\tilde{U}\tilde{U}'\tilde{W})\} = E^0\{LC_\kappa(\tilde{U}\tilde{U}' 0 \tilde{W} 0')\} \\ &= E^0\{LC_\kappa(\tilde{U}\tilde{U}') C_\kappa(\tilde{W})\} / C_\kappa(1_p) \\ &= E^0\{LC_\kappa(T) C_\kappa(V)\} / C_\kappa(1_p) , \end{aligned} \tag{3.15}$$

where 0 is a random orthogonal $p \times p$ matrix distributed according to the invariant probability measure λ of equation (3.11) and independent of $(\tilde{U}, \tilde{W})$. From equations (3.9) and (3.15) (and in particular, with (3.15) applied to the case $L \equiv 1$), and from the independence of T and V when $\xi_0 = 0$, we have

$$E^\Lambda\{L(T, V)\} = \textstyle\sum_\kappa \pi\{\kappa\} \dfrac{E^0\{L(T, V) C_\kappa(T) C_\kappa(V)\}}{E^0\{C_\kappa(T)\} E^0\{C_\kappa(V)\}} , \quad \text{where} \tag{3.16}$$

$$\pi\{\kappa\} = \exp(-\tfrac{1}{2} \operatorname{tr} \Lambda) C_\kappa(\tfrac{1}{2}\Lambda) / ||\kappa||! .$$

Using equation (3.8) it is easy to show that $\sum_\kappa \pi\{\kappa\} = 1$ where the summation is over all ordered sequences $\kappa = \{\kappa_1 \geqq \cdots \geqq \kappa_p\}$. This, together with the fact that $\Omega \geqq 0 \Rightarrow C_\kappa(\Omega) \geqq 0$, allows us to describe the noncentral law of (T, V) as follows: T and V are conditionally independent given κ with

$$(3.17)\qquad \begin{aligned} E^\kappa\{L(T)\} &= E^0\{L(T)C_\kappa(T)\}/E^0\{C_\kappa(T)\}\,, \\ E^\kappa\{L(V)\} &= E^0\{L(V)C_\kappa(V)\}/E^0\{C_\kappa(V)\}\,; \end{aligned}$$

here E^0 denotes expectation with respect to the corresponding central law when $\xi_0 = 0 \Leftrightarrow \Lambda = 0 \Leftrightarrow \kappa = 0$. The law π of κ can be thought of as a random partitioning of a Poisson variable $||\kappa||$ into p parts $\{\kappa_1 \geqq \cdots \geqq \kappa_p\}$; it follows easily from equation (3.8) that $||\kappa||$ follows a Poisson law with $E\{||\kappa||\} = \frac{1}{2}\operatorname{tr}\Lambda$.

We are now able to imitate the method of Section 2, using κ as the analogue of the Poisson variable used there. Writing

$$(3.18)\qquad \begin{aligned} R_\psi(\Lambda) &= E\{\psi(T)|V| - 1\}^2 \\ &= EE^{(\kappa,T)}\{\psi(T)|V| - 1\}^2\,, \end{aligned}$$

we see that the inner conditional expectation is minimized by taking $\psi = \psi_\kappa$ where

$$(3.19)\qquad \begin{aligned} \psi_\kappa &= E^{(\kappa,T)}\{|V|\}/E^{(\kappa,T)}\{|V|\}^2 \\ &= E^\kappa\{|V|\}/E^\kappa\{|V|^2\} \\ &= E^0\{|V|C_\kappa(V)\}/E^0\{|V|^2C_\kappa(V)\}\,. \end{aligned}$$

The second equality follows from conditional independence of V and T given κ, and the third from equation (3.17). Now we can express the expectations occurring in the last line of equation (3.19) in terms of expectations of the form $E\{C_\kappa(S)\}$ where S has central Wishart laws $W_p(n+k+2, 1_p)$ and $W_p(n+k+4, 1_p)$ for the numerator and denominator respectively. Then a simple calculation using equation (3.12) and the known form of the normalizing constants for the central Wishart densities yields

$$(3.20)\qquad \begin{aligned} \psi_\kappa &= \textstyle\prod_{i=1}^p (n + k + 3 - i + \kappa_i)^{-1} \\ &\leqq \textstyle\prod_{i=1}^p (n + k + 3 - i)^{-1} \\ &= \psi_0 = (n + k + 2 - p)!/(n + k + 2)!\,. \end{aligned}$$

We thus conclude that for any ψ, if we define $\psi^*(T) = \min\{\psi(T), \psi_0\}$, then $R_{\psi^*}(\Lambda) \leqq R_\psi(\Lambda)$ for all Λ and the inequality is strict provided $\psi^*(T) \neq \psi(T)$ with positive probability. In terms of the original problem this says that for any estimator $\phi(X, S) = \psi(T)|S + XX'|$, the estimator $\phi^*(X, S) = \psi^*(T)|S + XX'|$ satisfies

$$(3.21)\qquad E^{(\xi,\Sigma)}\{\phi^*(X, S) - |\Sigma|\}^2 \leqq E^{(\xi,\Sigma)}\{\phi(X, S) - |\Sigma|\}^2$$

for all (ξ, Σ) and the inequality is strict as long as $\phi^*(X, S) \neq \phi(X, S)$ with positive probability.

For the special case $\phi(X, S) = c|S|$ we have $\psi(T) = c|S|/|S + XX'|$, and the improved estimator is

$$
\begin{aligned}
\phi^*(X, S) &= \min\left\{\frac{c|S|}{|S + XX'|}, \phi_0\right\} |S + XX'| \\
&= \min\{c|S|, \phi_0|S + XX'|\} \qquad (3.22)\\
&= \min\left\{\frac{(n+2-p)!}{(n+2)!}|S|, \frac{(n+k+2-p)!}{(n+k+2)!}|S + XX'|\right\}.
\end{aligned}
$$

In this case $\phi^*(X, S) \neq \phi(X, S) \Leftrightarrow c|S|/|S + XX'| > \phi_0 \Leftrightarrow |T| > \phi_0/c$ and this occurs with positive probability because $\phi_0/c < 1$. So the estimator in equation (3.22) has expected squared error strictly smaller than has $c|S|$ for all parameter values. It is the minimum of two Selliah estimators, $\phi_0|S + XX'|$ being the analogue of $c|S|$ when $\xi = EX = 0$, and it chooses $c|S|$ or $\phi_0|S + XX'|$ on the basis of a preliminary likelihood ratio test of the hypothesis $\xi = 0$, with $|T|$ as the test criterion. Intuitively we expect $\phi_0|S + XX'|$ to overestimate $|\Sigma|$ when $\xi \neq 0$ and so prefer $c|S|$, unless the "overestimator" actually gives a smaller value so does not appear to be overestimating in our sample.

4. Discussion. The work in Section 3 generalizes the result of Stein (1964) which treats the case $p = 1 \Leftrightarrow$ estimating the error variance σ^2 in a univariate linear model with normal errors. For this problem Brewster (1972) [see also Brewster and Zidek (1974)] has produced a smooth formal Bayes estimator of σ^2 that is, like Stein's, uniformly better than the usual estimator, and unlike Stein's, is for $k = 1$ admissible among all scale-invariant estimators. In Section 2 the case $k = 1$ (but general p) of estimating $|\Sigma|$ was reduced to a problem that is formally almost identical to that treated by Stein and by Brewster.

The works cited above and the present paper all use a technical device of conditioning which can be described as follows. The parametric statistical problem $\{P_\theta : \theta \in 0\}$, may be visualized as a Markov transition $\Theta \to \mathscr{X}$ from the parameter space Θ to the space of a sufficient statistic X. We then factor this transition through a third space $\mathscr{T}$, thereby obtaining the two stage Markov representation $\Theta \to \mathscr{T} \to \mathscr{X}$, $\theta \to \tau \to X$. Given a loss function L, the risk function of a nonrandomized decision rule ϕ can be written

$$R_\phi(\theta) = E^\theta E^\tau L(\phi(X), X, \theta)$$

and the symmetries of the problem may then allow us to find a function $\hat{L}(\phi(X), X, \tau)$ such that

$$R_\phi(\theta) = E^\theta E^\tau \hat{L}(\phi(X), X, \tau)$$

for all estimators ϕ under consideration. The original problem is thereby represented as a mixture of inference problems based on $\mathscr{T} \to \mathscr{X}$ with $\mathscr{T}$ as parameter space and loss function $\hat{L}$, and for these problems $(\mathscr{T} \to \mathscr{X}, \hat{L})$ it may be easier to see what is a good procedure. In particular it is seen that a given procedure is definitely unreasonable when viewed as a procedure for $(\mathscr{T} \to \mathscr{X}, \hat{L})$.

The work in Section 2 fits into this framework with $\mathscr{T} =$ the space of the hypothetical Poisson variable κ. The argument essentially says that the estimator $c|S|$, when viewed in terms of the dilated problem $\lambda \to \kappa \to X'S^{-1}X$, produces an estimator of κ which with positive probability lies outside the convex hull of the space $\{0, 1, 2, \cdots\}$ of κ. An analogous interpretation holds in the k-means case treated in Section 3.

Another problem that can be viewed in this way is that of estimating a multivariate normal mean with quadratic loss. Suppose $x \sim N_p(\xi, 1)$ where $p \geqq 3$ and we seek estimators of the form $\hat{\psi}(x) = \psi(||x||^2)x$ where ψ is a real-valued function, with loss function $L(\psi, \xi) = ||\hat{\psi} - \xi||^2$. Then, setting $\lambda = ||\xi||^2$ and taking $\kappa \simeq P_0(\lambda/2)$ so that $||x||^2 \simeq \chi^2_{p+2\kappa}$, we have

$$\begin{aligned} R_\psi(\lambda) &= E^\lambda||\psi(||x||^2)x - \xi||^2 \\ &= E^\lambda\{\psi^2(||x||^2)||x||^2 - 2\xi'x\psi(||x||^2) + ||\xi||^2\} \\ &= E^\lambda E^\kappa\{\psi^2(||x||^2)||x||^2 - 4\kappa\psi(||x||^2) + 2\kappa\} \\ &= E^\lambda E^\kappa \hat{L}(\psi(||x||^2), ||x||^2, \kappa)\,. \end{aligned}$$

The third equality follows from the relation

$$E\{\xi'x\psi(||x||^2)\} = E\{2\kappa\psi(||x||^2)\}\,,$$

which is proved in James and Stein (1960) and can also be seen from the formula

$$E\{2\kappa\}E\{\psi(||x||^2)x\} = E\{\kappa\psi(||x||^2)\}\xi\,,$$

which can be proved using essentially the method Baranchik (1973) used for the special case $\psi(||x||^2) = 1/||x||^2$. Define $\psi(t) = t^{-1}\phi(t)$ and

$$\tilde{L}(\phi, 2\kappa) = \frac{2(\phi - 2\kappa)^2}{p - 2 + 2\kappa} + \frac{2\kappa(p-2)}{p - 2 + 2\kappa}\,.$$

Then a straightforward calculation shows

$$E^\kappa\tilde{L}\{\phi(\tilde{T}), 2\kappa\} = E^\kappa\tilde{L}\{\psi(||x||^2), ||x||^2, \kappa\}$$

where $\tilde{T} \simeq \chi^2_{p-2+2\kappa}$ and $\kappa = 0, 1, 2, \cdots$. The original problem is thus equivalent to a mixture of problems of estimating 2κ from $\tilde{T} \simeq \chi^2_{p-2+2}$ with quadratic loss $\tilde{L}$. If we choose $\phi(\tilde{T}) = \tilde{T}$ we get $\hat{\psi}(x) = x$, the usual estimator, but if we choose $\phi(\tilde{T}) = \{\tilde{T} - (p-2)\}^+$, we get the James–Stein (1960) estimator

$$\hat{\psi}(x) = \left\{1 - \frac{p - 2}{||x||^2}\right\}^+ x\,.$$

The usual estimator $\hat{\psi}(x) = x \Leftrightarrow \phi(\tilde{T}) = \tilde{T}$ is evidently inadmissible because $\tilde{T}$ can substantially overestimate 2κ : $E\tilde{T} = p - 2 + 2\kappa$, and the James–Stein estimator seems to be a natural improvement.

REFERENCES

[1] BARANCHIK, A. J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. *Ann. Statist.* **1** 312–321.

[2] BREWSTER, J. F. (1972). On the admissibility of scale and quantile estimators. Ph.D. dissertation, Univ. of British Columbia.

[3] BREWSTER, J. F. and ZIDEK, J. V. (1974). Improving on equivariant estimators. *Ann. Statist.* **2** 21–38.

[4] BROWN, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.* **39** 29–48.

[5] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501.

[6] JAMES, A. T. (1968). Calculation of zonal polynomial coefficients by the use of the Laplace Beltrami operator. *Ann. Math. Statist.* **39** 1711–1718.

[7] JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.

[8] SELLIAH, J. B. (1964). Estimation and testing problems in a Wishart distribution. Stanford Univ. Technical Report No. 10, Dept. of Statist.

[9] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.

10] STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

CENTRE DE RECHERCHES MATHÉMATIQUES
UNIVERSITÉ DE MONTRÉAL
MONTRÉAL, PQ
CANADA

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
2075 WESTBROOK PLACE
VANCOUVER, B. C.
CANADA V6T 1W5