Ann. Inst. Statist. Math.
Vol. 42, No. 2, 331-343 (1990)

ESTIMATING THE COVARIANCE MATRIX AND THE GENERALIZED VARIANCE UNDER A SYMMETRIC LOSS

TATSUYA KUBOKAWA* AND YOSHIHIKO KONNO

Institute of Mathematics, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

(Received February 20, 1989; revised August 28, 1989)

Abstract. For estimating the power of a generalized variance under a multivariate normal distribution with unknown means, the inadmissibility of the best affine equivariant estimator relative to the symmetric loss is shown, and a class of improved estimators is given. The problem of estimating the covariance matrix is also discussed.

Key words and phrases: Covariance matrix, generalized variance, Wishart distribution, affine equivariant estimators, Stein's truncated estimator, inadmissibility.

1. Introduction

Suppose that X $(p \times r)$ has the normal distribution $N(\xi, \Sigma \otimes I_r)$ and that S $(p \times p)$ has the Wishart distribution $W_p(n, \Sigma)$, where $n \geq p$ and the matrix ξ of mean vectors is unknown. Assume that X and S are independent. The first problem we consider is to estimate the α-th power of the generalized variance $|\Sigma|^\alpha$ under a symmetric loss given by

$$L(d, |\Sigma|^\alpha) = d/|\Sigma|^\alpha + |\Sigma|^\alpha/d - 2\,, \tag{1.1}$$

where α may be positive or negative. Every estimator is evaluated by its risk function $R(\theta, d) = E_\theta[L(d, |\Sigma|^\alpha)]$ for $\theta = (\xi, \Sigma)$, unknown parameters.

The affine equivariant estimator under the transformation $(X, S) \to (AXH + B, ASA')$ for any nonsingular matrix A $(p \times p)$, any orthogonal matrix H $(r \times r)$ and any matrix B $(p \times r)$, is defined by $d(AXH + B, ASA') = |A|^{2\alpha} d(X, S)$, and it must be of the form $a|S|^\alpha$, $a > 0$. The inadmissibility of the best affine equivariant estimator has been shown by Shorrock and Zidek (1976), Sinha (1976) and Sugiura and Konno (1987) for squared error loss; by Sinha and Ghosh (1987), Sugiura and Konno (1988) and

*Now at Department of Mathematical Engineering and Information Physics, Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan.

Sugiura (1989) for entropy loss. Some of them are multivariate extensions of Stein (1964) and Strawderman (1974). Recently, Pal (1988) proposed the loss (1.1) as a combination of two entropy losses and called it convex entropy loss. He also pointed out that under (1.1), the loss of the estimator d for the generalized variance $|\Sigma|$ is equal to that of the estimator d^{-1} for the generalized precision $|\Sigma^{-1}|$, that is, $L(d, |\Sigma|) = L(d^{-1}, |\Sigma^{-1}|)$. Therefore, we shall call it a symmetric loss. The best affine equivariant estimator relative to the loss (1.1) can be shown to be

$$d_0 = a(\alpha)|S|^{\alpha} \tag{1.2}$$

where $n - 2|\alpha| - p + 1 > 0$ and

$$a(\alpha) = 2^{-\alpha p}\left[\Gamma_p\left(\frac{n}{2} - \alpha\right)\Big/\Gamma_p\left(\frac{n}{2} + \alpha\right)\right]^{1/2}$$

is a constant given by the p-variate Gamma function $\Gamma_p(x) = \pi^{p(p-1)/4} \cdot \prod_{i=1}^{p}\Gamma(x - (i-1)/2)$. It is noted that $\{a(\alpha)\}^{-1} = a(-\alpha)$. In Section 2, it is shown that the best affine equivariant estimator d_0 is dominated by

$$d_S = [\min\{a(|\alpha|)|S|^{|\alpha|}, b(|\alpha|)|S + XX'|^{|\alpha|}\}]^{\mathrm{sgn}(\alpha)} \tag{1.3}$$

where sgn $(\alpha) = 1$ for $\alpha > 0$; $= -1$ for $\alpha < 0$, and

$$b(\alpha) = 2^{-\alpha p}\left[\Gamma_p\left(\frac{n+r}{2} - \alpha\right)\Big/\Gamma_p\left(\frac{n+r}{2} + \alpha\right)\right]^{1/2},$$

with $\{b(\alpha)\}^{-1} = b(-\alpha)$. This is an extension of Pal (1988). From Lemma A.1 given in the Appendix, it should be noted that $a(|\alpha|) > b(|\alpha|)$. Section 3 provides a family of improved estimators.

Section 4 treats the problem of estimating the covariance matrix Σ under the loss

$$L(\delta, \Sigma) = \mathrm{tr}\,(\delta\Sigma^{-1} + \Sigma\delta^{-1}) - 2p\,. \tag{1.4}$$

The best affine equivariant estimator is of the form

$$\delta_0 = \{n(n-p-1)\}^{-1/2}S\,, \tag{1.5}$$

and it is demonstrated that δ_0 is dominated by Haff type estimators when $p \geq 2$ and $n - p - 1 > 0$.

2. Inadmissibility of the best affine equivariant estimator for powers of the generalized variance

We shall prove that the best affine equivariant estimator d_0, given by (1.2), is dominated by d_S, given by (1.3), under the loss (1.1). From symmetry of the loss, it is sufficient to show the case of $\alpha > 0$. Put

$$T = \begin{cases} (S + XX')^{-1/2} S (S + XX')^{-1/2} & \text{when } r \geq p, \\ I_r - X'(S + XX')^{-1} X & \text{when } r < p. \end{cases} \tag{2.1}$$

Note that $|T| = |S| / |S + XX'|$ in either case. Also put $V = \Sigma^{-1/2}(S + XX') \cdot \Sigma^{-1/2}$. Then the estimators $d_0 / |\Sigma|^\alpha$ and $d_S / |\Sigma|^\alpha$ are expressed as

$$\begin{aligned} d_0 / |\Sigma|^\alpha &= a(\alpha) |T|^\alpha |V|^\alpha, \\ d_S / |\Sigma|^\alpha &= \min\{a(\alpha)|T|^\alpha |V|^\alpha, b(\alpha)|V|^\alpha\}. \end{aligned} \tag{2.2}$$

Let $f(|T|) = 1$ for $|T| \geq \{b(\alpha)/a(\alpha)\}^{1/\alpha}$; $= 0$ otherwise. Using the mixture representation for the distributions of V (noncentral Wishart) and T (noncentral Beta) by Shorrock and Zidek (1976) or by Lemma 2.1 in Sugiura and Konno (1987), we can write the risk difference as

$$\begin{aligned} \Delta_1 &= R(\theta, d_0) - R(\theta, d_S) \\ &= E\left[\left\{a(\alpha)|T|^\alpha |V|^\alpha + \frac{1}{a(\alpha)|T|^\alpha |V|^\alpha} - b(\alpha)|V|^\alpha - \frac{1}{b(\alpha)|V|^\alpha}\right\} f(|T|)\right] \\ &= E_\Lambda^\kappa\left[a(\alpha) \frac{E_0[|V|^\alpha C_\kappa(V)|\kappa]}{E_0[C_\kappa(V)|\kappa]} \frac{E_0[f(|T|)|T|^\alpha C_\kappa(I-T)|\kappa]}{E_0[C_\kappa(I-T)|\kappa]}\right. \\ &\quad + a(\alpha)^{-1} \frac{E_0[|V|^{-\alpha} C_\kappa(V)|\kappa]}{E_0[C_\kappa(V)|\kappa]} \frac{E_0[f(|T|)|T|^{-\alpha} C_\kappa(I-T)|\kappa]}{E_0[C_\kappa(I-T)|\kappa]} \\ &\quad - b(\alpha) \frac{E_0[|V|^\alpha C_\kappa(V)|\kappa]}{E_0[C_\kappa(V)|\kappa]} \frac{E_0[f(|T|) C_\kappa(I-T)|\kappa]}{E_0[C_\kappa(I-T)|\kappa]} \\ &\quad \left. - b(\alpha)^{-1} \frac{E_0[|V|^{-\alpha} C_\kappa(V)|\kappa]}{E_0[C_\kappa(V)|\kappa]} \frac{E_0[f(|T|) C_\kappa(I-T)|\kappa]}{E_0[C_\kappa(I-T)|\kappa]}\right], \end{aligned} \tag{2.3}$$

where $E_0[\cdot|\kappa]$ stands for the conditional expectation for given κ at $\Lambda = \xi'\Sigma^{-1}\xi = 0$ and $E_\Lambda^\kappa[\cdot]$ stands for the expectation with respect to the random partition $\kappa = \{\kappa_1, \ldots, \kappa_p\}$ ranging over ordered partitions of nonnegative integer $k = \|\kappa\| = \kappa_1 + \cdots + \kappa_p$ with $\kappa_1 \geq \cdots \geq \kappa_p \geq 0$ having probability mass $\mathrm{etr}(-\Delta/2)C_\kappa(\Delta/2)/k!$. Here $C_\kappa(Z)$ denotes the normalized zonal polynomials of the positive definite matrix Z of order p corresponding to partition $\kappa = \{\kappa_1, \ldots, \kappa_p\}$ so that for all $k = 0, 1, 2, \ldots$,

$$(\mathrm{tr}\ Z)^k = \sum_{\{\kappa;\|\kappa\|=k\}} C_\kappa(Z) .$$

The precise definition and the properties of the zonal polynomials may be found in James (1964) and Muirhead (1982). For given κ, V has the $W_p(n+r, I)$ distribution and T has $\mathrm{Beta}_p(n/2, r/2)$, the p-variate Beta distribution with parameter $(n/2, r/2)$ if $r \geq p$, and has $\mathrm{Beta}_r((n+r-p)/2, p/2)$ if $r < p$. They are conditionally independent for given κ. The constant $A(\alpha)$ is given by

$$(2.4) \qquad \begin{aligned} A(\alpha) &= E_0[|V|^\alpha C_\kappa(V)|\kappa]/E_0[C_\kappa(V)|\kappa] \\ &= 2^{\alpha p}\Gamma_p\left(\frac{n+r}{2}+\alpha\right)\left(\frac{n+r}{2}+\alpha\right)_\kappa \Big/ \left\{\Gamma_p\left(\frac{n+r}{2}\right)\left(\frac{n+r}{2}\right)_\kappa\right\}, \end{aligned}$$

where the multivariate hypergeometric coefficient is defined by

$$(a)_\kappa = \prod_{i=1}^{p}(a-(i-1)/2)(a+1-(i-1)/2)\cdots(a+\kappa_i-1-(i-1)/2) ,$$

$$(a)_0 = 1 .$$

Then from (2.3), the risk difference Δ_1 is rewritten by

$$\Delta_1 = E_\Delta^\kappa\left[E_0\left[\frac{C_\kappa(I-T)f(|T|)}{E_0[C_\kappa(I-T)|\kappa]}\{a(\alpha)|T|^\alpha - b(\alpha)\}\right.\right. \cdot\left.\left.\left\{A(\alpha) - \frac{1}{a(\alpha)b(\alpha)}|T|^{-\alpha}A(-\alpha)\right\}\right|\kappa\right]\right] .$$

Here we observe that

$$\begin{aligned} \frac{A(-\alpha)}{A(\alpha)} &= 2^{-2\alpha p}\Gamma_p\left(\frac{n+r}{2}-\alpha\right)\left(\frac{n+r}{2}-\alpha\right)_\kappa\Big/\left\{\Gamma_p\left(\frac{n+r}{2}+\alpha\right)\left(\frac{n+r}{2}+\alpha\right)_\kappa\right\} \\ &< 2^{-2\alpha p}\Gamma_p\left(\frac{n+r}{2}-\alpha\right)\Big/\Gamma_p\left(\frac{n+r}{2}+\alpha\right) = \{b(\alpha)\}^2 , \end{aligned}$$

so that $A(\alpha) - \{a(\alpha)b(\alpha)|T|^\alpha\}^{-1}A(-\alpha) \geq A(\alpha)|T|^{-\alpha}\{|T|^\alpha - b(\alpha)/a(\alpha)\}$, which implies that $\Delta_1 \geq 0$. Hence we get

THEOREM 2.1. *The best affine equivariant estimator d_0 is dominated by the estimator d_S given by* (1.3) *relative to the loss* (1.1) *when* $n-2|\alpha|-p+1>0$.

3. A class of improved estimators

Consider a class of estimators for $|\Sigma|^{\alpha}$ of the form

$$d_{\phi}=d_0\{1-\phi(|T|)|T|^{\varepsilon}\}^{\alpha}, \tag{3.1}$$

where the random matrix T is defined by (2.1). Strawderman (1974) and Sugiura (1988, 1989) treated another type of estimators whose forms are different in two cases $\alpha>0$ or $\alpha<0$. Here $\varepsilon\geq 2|\alpha|$ and $\phi(|T|)$ is a nondecreasing function of $|T|$ satisfying $0<\phi(|T|)\leq D$ for some constant $D>0$. Let

$$\lambda=\begin{cases}\varepsilon & \text{for}\quad 0<|\alpha|\leq 1\,,\\ \dfrac{D(1-D)^{|\alpha|-1}}{1-(1-D)^{|\alpha|}}|\alpha|\varepsilon & \text{for}\quad |\alpha|>1\,,\end{cases} \tag{3.2}$$

and define $C_p(\lambda,\varepsilon,\alpha,n,r)$ and $K_p(\varepsilon,\alpha,n,r)$ by

$$\begin{aligned}C_p(\lambda,\varepsilon,\alpha,n,r)&=\frac{1}{B_p(n/2+\lambda+\varepsilon-\alpha,r/2)}\\ &\quad\cdot\left\{\frac{B_p(n/2-\alpha,r/2)}{B_p(n/2+\alpha,r/2)}B_p(n/2+\lambda+\alpha,r/2)\right.\\ &\qquad\left.-B_p(n/2+\lambda-\alpha,r/2)\right\},\\ K_p(\varepsilon,\alpha,n,r)&=(2\alpha/\varepsilon)B_p(n/2-\alpha,r/2)/B_p(n/2+\alpha,r/2)\,,\end{aligned} \tag{3.3}$$

where $B_p(\alpha,\beta)=\Gamma_p(\alpha)\Gamma_p(\beta)/\Gamma_p(\alpha+\beta)$. By Lemma A.2 in the Appendix, $C_p(\lambda,\varepsilon,\alpha,n,r)$ is positive.

THEOREM 3.1. *Assume that $\phi(|T|)$ is nondecreasing in $|T|$ with $\phi(|T|)\leq D$ and*

$$\frac{1}{(1-D)^{|\alpha|}}-1 \tag{3.4}$$

$$\leq \begin{cases} \min\{C_p(\lambda, \varepsilon, |\alpha|, n, r),\ K_p(\varepsilon, |\alpha|, n, r)\} & \\ & \textit{for} \quad r \geq p\,, \\ \min\{C_r(\lambda, \varepsilon, |\alpha|, n+r-p, p),\ K_r(\varepsilon, |\alpha|, n+r-p, p)\} & \\ & \textit{for} \quad r < p\,. \end{cases}$$

Then the estimator d_ϕ *given by* (3.1) *dominates the best affine equivariant estimator* d_0 *under the loss* (1.1).

PROOF. From symmetry of the loss (1.1), it suffices to consider the case $\alpha > 0$. We first observe that $\{1 - \phi(|T|)|T|^{\varepsilon}\}^{-\alpha} \leq (1 - D|T|^{\varepsilon})^{-\alpha} \leq K|T|^{\varepsilon} + 1$ for $K = (1 - D)^{-\alpha} - 1$. From (2.2) and (2.3), the risk difference is written as

$$\begin{aligned} (3.5)\quad \Delta_2 &= R(\theta, d_0) - R(\theta, d_\phi) \\ &= E[\{1 - (1 - \phi(|T|)|T|^{\varepsilon})^{\alpha}\}\{a(\alpha)|T|^{\alpha}|V|^{\alpha} \\ &\qquad - a(-\alpha)(|T||V|)^{-\alpha}(1 - \phi(|T|)|T|^{\varepsilon})^{-\alpha}\}] \\ &\geq E[\{1 - (1 - \phi(|T|)|T|^{\varepsilon})^{\alpha}\}\{a(\alpha)|T|^{\alpha}|V|^{\alpha} \\ &\qquad - a(-\alpha)(|T||V|)^{-\alpha}(K|T|^{\varepsilon} + 1)\}] \\ &= E_{\Lambda}^{\kappa}\left[\frac{A(-\alpha)a(-\alpha)}{E_0[C_\kappa(I - T)|\kappa]} \cdot E_0[k(|T|)|T|^{\lambda-\alpha}g(|T|)C_\kappa(I - T)|\kappa]\right], \end{aligned}$$

where $A(\alpha)$ is defined by (2.4), and for $0 < t < 1$,

$$(3.6)\qquad \begin{aligned} k(t) &= \{1 - (1 - \phi(t)t^{\varepsilon})^{\alpha}\}/t^{\lambda}\,, \\ g(t) &= \{a(\alpha)\}^2\{A(\alpha)/A(-\alpha)\}t^{2\alpha} - Kt^{\varepsilon} - 1\,. \end{aligned}$$

Here we shall demonstrate that $k(t)$ is nondecreasing in t when $0 < t < 1$. Observe that $(d/dt)\{\log k(t)\} \geq 0$ if $\lambda \leq h(\phi(t)t^{\varepsilon})$ for $h(s) = \alpha\varepsilon s(1 - s)^{\alpha-1}/\{1 - (1 - s)^{\alpha}\}$, $0 < s \leq D$. Further, it can be seen that $h(s)$ is nondecreasing in s when $0 < \alpha \leq 1$ and $h(s)$ is nonincreasing when $\alpha > 1$. Hence we have that

$$\inf_{0 < s \leq D} h(s) = \begin{cases} \lim\limits_{s \to 0} h(s) = \varepsilon & \text{for} \quad 0 < \alpha \leq 1\,, \\ \lim\limits_{s \to D} h(s) = \dfrac{D(1 - D)^{\alpha-1}}{1 - (1 - D)^{\alpha}}\,\alpha\varepsilon & \text{for} \quad \alpha > 1\,, \end{cases}$$

which, by the definition of λ, gives that $k(t)$ is nondecreasing. Also it follows that the derivative of $g(t)$ is positive under the conditions $\varepsilon \geq 2|\alpha|$

and (3.4), that is, $g(t)$ is nondecreasing. Since $g(t)$ changes sign at most once from negative, we have

$$(3.7)\qquad \Delta_2 \geq E_A^{\kappa}\left[\frac{A(-\alpha)a(-\alpha)}{E_0[C_\kappa(I-T)|\kappa]}\cdot k(t^*)E_0[|T|^{\lambda-\alpha}g(|T|)C_\kappa(I-T)|\kappa]\right],$$

where $t^* = \sup\{t|0<t<1, g(t)<0\}$. When $r \geq p$, we put

$$\begin{aligned} B(\alpha) &= E_0[|T|^\alpha C_k(I-T)|\kappa]/E_0[C_\kappa(I-T)|\kappa] \\ &= B_p\left(\frac{n}{2}+\alpha, \frac{r}{2}\right)\left(\frac{n+r}{2}\right)_\kappa \Big/ \left\{B_p\left(\frac{n}{2}, \frac{r}{2}\right)\left(\frac{n+r}{2}+\alpha\right)_\kappa\right\}. \end{aligned}$$

Hence from (3.7), the risk difference Δ_2 is nonnegative if

$$(3.8)\qquad \{a(\alpha)\}^2\{A(\alpha)/A(-\alpha)\}B(\lambda+\alpha) - KB(\lambda+\varepsilon-\alpha) - B(\lambda-\alpha) \geq 0,$$

namely, $K \leq C(\kappa)$ where

$$C(\kappa) = [\{a(\alpha)\}^2\{A(\alpha)/A(-\alpha)\}B(\lambda+\alpha) - B(\lambda-\alpha)]/B(\lambda+\varepsilon-\alpha).$$

It can be easily shown that $C(\kappa) \geq C(0) = C_p(\lambda, \varepsilon, \alpha, n, r)$, which implies that the inequality (3.8) is guaranteed by the condition (3.4). When $r<p$, replace $B(\alpha)$ in (3.8) with

$$\begin{aligned} B^*(\alpha) &= E_0[|T|^\alpha C_\kappa(I-T)|\kappa]/E_0[C_\kappa(I-T)|\kappa] \\ &= B_r\left(\frac{n+r-p}{2}+\alpha, \frac{p}{2}\right)\left(\frac{n+r}{2}\right)_\kappa \Big/ \left\{B_r\left(\frac{n+r-p}{2}, \frac{p}{2}\right)\left(\frac{n+r}{2}+\alpha\right)_\kappa\right\}, \end{aligned}$$

and note that

$$\begin{aligned} &B_p\left(\frac{n}{2}-\alpha, \frac{r}{2}\right)\Big/ B_p\left(\frac{n}{2}+\alpha, \frac{r}{2}\right) \\ &\qquad = B_r\left(\frac{n+r-p}{2}-\alpha, \frac{p}{2}\right)\Big/ B_r\left(\frac{n+r-p}{2}+\alpha, \frac{p}{2}\right), \end{aligned}$$

which can be derived by the relation

$$(3.9)\qquad \Gamma_p\left(\frac{n}{2}+x\right)\Big/\Gamma_p\left(\frac{n+r}{2}+x\right) = \Gamma_r\left(\frac{n+r-p}{2}+x\right)\Big/\Gamma_r\left(\frac{n+r}{2}+x\right).$$

Then we can get the desired result and the proof of Theorem 3.1 is complete.

Example 3.1. In estimation of the generalized variance $|\Sigma|$ ($\alpha = 1$), we shall find the Stein type truncated estimator belonging to our class. Setting $\phi_c(t) = \max\{0, t^{-\varepsilon}(1 - c/t)\}$, $0 < c < 1$, in (3.1) yields the estimator

$$d_{\phi_c} = \min\{a(1)|S|, ca(1)|S + XX'|\}.$$

Let $\varepsilon = c/(1-c)$. Then we can see that $\phi_c(t)$ is nondecreasing for $0 < t < 1$, and that $\phi_c(t) \le \lim_{t \to 1} \phi_c(t) = 1 - c$. In the case of $r \ge p$, the condition (3.4) is written by

$$\varepsilon^{-1} \le \min\{C_p(\varepsilon, \varepsilon, 1, n, r), K_p(\varepsilon, 1, n, r)\}.$$

Since $B_p(n/2 - 1, r/2)/B_p(n/2 + 1, r/2) > 1$, it is clear that $\varepsilon^{-1} \le K_p(\varepsilon, 1, n, r)$. Also observe that $\lim_{\varepsilon \to \infty} C_p(\varepsilon, \varepsilon, 1, n, r) = B_p(n/2 - 1, r/2)/B_p(n/2 + 1, r/2) - 1 > 0$, which shows that

$$\varepsilon^{-1} \le C_p(\varepsilon, \varepsilon, 1, n, r) \tag{3.10}$$

holds for large ε, or c close to one. Hence for the constant c (close to one) satisfying (3.10), our improved class can include the Stein type truncated estimator d_{ϕ_c}. *The case of* $r < p$ is similarly shown.

4. Inadmissibility of the best affine equivariant estimator for the covariance matrix

Now we consider the problem of estimating the covariance matrix Σ. By developing the minimax estimators, James and Stein (1961) and Olkin and Selliah (1977) have shown the inadmissibility of the best affine equivariant estimator for entropy loss and quadratic loss, respectively. Haff (1980) has obtained other improved estimators under both losses as the empirical Bayes procedure. Of interest is to investigate the inadmissibility under the symmetric loss (1.4). We can easily get the best affine equivariant estimator

$$\delta_0 = cS \quad \text{for} \quad c = \{n(n-p-1)\}^{-1/2}. \tag{4.1}$$

For improving on δ_0, the Haff type estimator we consider is

$$\delta(t) = c\left(S + \frac{t}{\operatorname{tr} S^{-1}} I_p\right), \tag{4.2}$$

where t is a nonnegative constant. Then we prove

THEOREM 4.1. *Assume that* $n-p-1>0$ *and* $p \geq 2$. *Then the estimator* $\delta(t)$ *given by* (4.2) *dominates the best affine equivariant estimator* δ_0 *if*

$$0 < t \leq \frac{(p-1)(2n-p)}{pn(n-p)} . \tag{4.3}$$

PROOF. Letting $\mathrm{diag}\,(l_1,\ldots,l_p) = \mathrm{diag}\,(l_i) = H'SH$ for some orthogonal matrix H, we observe that

$$\begin{aligned} H'\{S + (t/\mathrm{tr}\,S^{-1})I\}^{-1}H &= \mathrm{diag}\,(\{l_i + t/\mathrm{tr}\,S^{-1}\}^{-1}) \\ &\leq \mathrm{diag}\left(\frac{1}{l_i}\left\{1 - \frac{t}{l_i\,\mathrm{tr}\,S^{-1}} + \left(\frac{t}{l_i\,\mathrm{tr}\,S^{-1}}\right)^2\right\}\right) \\ &= H'\left\{S^{-1} - \frac{t}{\mathrm{tr}\,S^{-1}}\,S^{-2} + \left(\frac{t}{\mathrm{tr}\,S^{-1}}\right)^2 S^{-3}\right\}H , \end{aligned}$$

which yields that

$$\mathrm{tr}\left(S + \frac{t}{\mathrm{tr}\,S^{-1}}\,I_p\right)^{-1}\Sigma \leq \mathrm{tr}\left\{S^{-1} - \frac{t}{\mathrm{tr}\,S^{-1}}\,S^{-2} + \left(\frac{t}{\mathrm{tr}\,S^{-1}}\right)^2 S^{-3}\right\}\Sigma .$$

Hence the risk difference is written as

$$\begin{aligned} \Delta_3 &= R(\theta,\delta_0) - R(\theta,\delta(t)) \\ &= E\left[c\,\mathrm{tr}\,S\Sigma^{-1} + c^{-1}\,\mathrm{tr}\,S^{-1}\Sigma - c\,\mathrm{tr}\left(S + \frac{t}{\mathrm{tr}\,S^{-1}}\,I_p\right)\Sigma^{-1} \right. \\ &\qquad \left. - c^{-1}\,\mathrm{tr}\left(S + \frac{t}{\mathrm{tr}\,S^{-1}}\,I_p\right)^{-1}\Sigma\right] \\ &\geq E\left[c^{-1}t\,\frac{\mathrm{tr}\,S^{-2}\Sigma}{\mathrm{tr}\,S^{-1}} - ct\,\frac{\mathrm{tr}\,\Sigma^{-1}}{\mathrm{tr}\,S^{-1}} - c^{-1}t^2\,\frac{\mathrm{tr}\,S^{-3}\Sigma}{(\mathrm{tr}\,S^{-1})^2}\right] \\ &\geq E\left[c^{-1}t\,\frac{\mathrm{tr}\,S^{-2}\Sigma}{\mathrm{tr}\,S^{-1}} - ct\,\frac{\mathrm{tr}\,\Sigma^{-1}}{\mathrm{tr}\,S^{-1}} - c^{-1}t^2\,\frac{p}{n-p-1}\right], \end{aligned} \tag{4.4}$$

since $E[\mathrm{tr}\,S^{-3}\Sigma/(\mathrm{tr}S^{-1})^2] \leq E[\mathrm{tr}\,S^{-2}\,\mathrm{tr}\,S^{-1}\Sigma/(\mathrm{tr}\,S^{-1})^2] \leq E[\mathrm{tr}\,S^{-1}\Sigma] = p/(n-p-1)$. Here, if the inequalities

$$E\left[\frac{\mathrm{tr}\,\Sigma^{-1}}{\mathrm{tr}\,S^{-1}}\right] \leq n-p+1 , \tag{4.5}$$

$$E\left[\frac{\operatorname{tr} S^{-2}\Sigma}{\operatorname{tr} S^{-1}}\right] \geq \frac{n-1}{(n-p-1)(n-p)}, \tag{4.6}$$

are valid, then combining (4.4), (4.5) and (4.6) gives that

$$\Delta_3 \geq c^{-1}t(n-1)/\{(n-p-1)(n-p)\} - ct(n-p-1) - c^{-1}t^2p/(n-p-1),$$

which is nonnegative under the condition (4.3). Hence, in order to complete the proof, we only need to show (4.5) and (4.6).

To prove (4.5) and (4.6), Haff's identity is useful. For any $p \times p$ matrix $V = (v_{ij}(S))$ and $S = (s_{ij})$, define $V_{(c)} = (v'_{ij})$ where

$$v'_{ij} = \begin{cases} v_{ij} & \text{for } i = j, \\ cv_{ij} & \text{for } i \neq j, \end{cases}$$

and $D = (\partial/\partial s_{ij})_{(1/2)}$. For a real-valued function $h(S)$, Haff (1979) obtained the following identity:

$$E[h(S) \operatorname{tr} V\Sigma^{-1}] = 2E[h(S) \operatorname{tr}(DV)] + 2E\left[\operatorname{tr}\left\{\frac{\partial h(S)}{\partial S} \cdot V_{(1/2)}\right\}\right] + (n-p-1)E[h(S) \operatorname{tr} S^{-1}V]. \tag{4.7}$$

Putting $h(S) = 1/\operatorname{tr} S^{-1}$ and $V = I$ in (4.7) yields

$$\begin{aligned} E\left[\frac{\operatorname{tr}\Sigma^{-1}}{\operatorname{tr} S^{-1}}\right] &= 2E\left[\operatorname{tr}\left(\frac{\partial}{\partial S}\frac{1}{\operatorname{tr} S^{-1}}\right)\right] + n - p - 1 \\ &= 2E[\operatorname{tr} S^{-2}/(\operatorname{tr} S^{-1})^2] + n - p - 1 \\ &\leq n - p + 1, \end{aligned} \tag{4.8}$$

since

$$\frac{\partial}{\partial S}(\operatorname{tr} S^{-1})^{-1} = -(\operatorname{tr} S^{-1})^{-2}\frac{\partial}{\partial S}\operatorname{tr} S^{-1} = (\operatorname{tr} S^{-1})^{-2}S_{(2)}^{-2}, \tag{4.9}$$

which can be derived based on Lemma 3.2 in Haff (1979). Hence we get (4.5). For (4.6), put $h(S) = 1/\operatorname{tr} S^{-1}$ and $V = S^{-1}\Sigma$ in (4.7). Then,

$$1 = 2E\left[\frac{\operatorname{tr}(DS^{-1}\Sigma)}{\operatorname{tr} S^{-1}}\right] + 2E\left[\operatorname{tr}\left\{\left(\frac{\partial}{\partial S}\frac{1}{\operatorname{tr} S^{-1}}\right)(S^{-1}\Sigma)_{(1/2)}\right\}\right] + (n-p-1)E\left[\frac{\operatorname{tr} S^{-2}\Sigma}{\operatorname{tr} S^{-1}}\right]. \tag{4.10}$$

Here Lemma 3.1(iii) in Haff (1979) gives that tr $DS^{-1}\Sigma = -(1/2)$ tr $S^{-2}\Sigma - (1/2)(\text{tr } S^{-1})(\text{tr } S^{-1}\Sigma)$, so that from (4.9) and (4.10), we have

$$\begin{aligned} 1 &= (n-p-2)E\left[\frac{\text{tr } S^{-2}\Sigma}{\text{tr } S^{-1}}\right] + 2E\left[\frac{\text{tr } S^{-3}\Sigma}{(\text{tr } S^{-1})^2}\right] - E[\text{tr } S^{-1}\Sigma] \\ &\le (n-p)E\left[\frac{\text{tr } S^{-2}\Sigma}{\text{tr } S^{-1}}\right] - \frac{p}{n-p-1}, \end{aligned}$$

which implies (4.6). Therefore Theorem 3.1 is completely proved.

By symmetry of the loss (1.4), we get

COROLLARY 4.1. *The best affine equivariant estimator* δ_0^{-1} *for* Σ^{-1} *is improved on by* $\{\delta(t)\}^{-1}$ *relative to the loss* (1.4) *under the condition* (4.3).

Remark 4.1. Sinha and Ghosh (1987) established the domination by a Stein type truncated estimator for entropy loss. It is interesting if we could show the improvement of δ_0 by a Stein type estimator under the loss (1.4).

Remark 4.2. One may consider the class of orthogonally invariant estimators as in Dey and Srinivasan (1985), that is, of the form $R\phi(L)R'$ where $S = RLR'$, $L = \text{diag}\{l_1, \ldots, l_p\}$, $l_1 > \cdots > l_p$ are eigenvalues of S, R is the matrix of normalized eigenvectors, $\phi(L) = \text{diag}\{\phi_1(L), \ldots, \phi_p(L)\}$ and $\phi_i(L)$'s are functions from L to $(0, \infty)$. But we could not derive superior alternatives (e.g., an orthogonally invariant minimax estimator) to the best affine equivariant estimator (4.1) under the loss (1.4) since it is difficult to evaluate the expectation of the form tr $\Sigma\{R\phi(L)R'\}^{-1}$.

Acknowledgement

The authors are grateful to the referee for his helpful comments.

Appendix

LEMMA A.1. *For constants* $a(\alpha)$ *and* $b(\alpha)$ *given in Section* 1, *the inequality* $a(\alpha) > b(\alpha)$ *for* $\alpha > 0$ *holds.*

PROOF. Put $G_1(n, r) = \Gamma_p(n/2 - \alpha)\Gamma_p((n+r)/2 + \alpha)/\{\Gamma_p(n/2 + \alpha)\cdot \Gamma_p((n+r)/2 - \alpha)\}$. Then the inequality $a(\alpha) > b(\alpha)$ is equivalent to $G_1(n, r) > 1$. Letting $f_i = (n - i + 1)/2 - \alpha$, we can see that

$$G_1(n,r)=\prod_{i=1}^{p}\left[\Gamma(f_i)\Gamma\left(f_i+2\alpha+\frac{r}{2}\right)\Big/\left\{\Gamma(f_i+2\alpha)\Gamma\left(f_i+\frac{r}{2}\right)\right\}\right]$$
$$=\prod_{i=1}^{p}\left[E_i[Z^{2\alpha+r/2}]/\{E_i[Z^{2\alpha}]E_i[Z^{r/2}]\}\right],$$

where E_i stands for expectation with respect to the probability measure P_i given by $P_i(A)=\int_A z^{f_i-1}e^{-z}dz/\Gamma(f_i)$. Since both $Z^{2\alpha}$ and $Z^{r/2}$ are increasing, $E_i[Z^{2\alpha+r/2}]>E_i[Z^{2\alpha}]E_i[Z^{r/2}]$, which shows that $G_1(n,r)>1$.

LEMMA A.2. *The constant* $C_p(\lambda,\varepsilon,\alpha,n,r)$ *given by* (3.3) *is positive for* $r\geq p$. *Also* $C_r(\lambda,\varepsilon,\alpha,n+r-p,p)$ *is positive for* $r<p$.

PROOF. We shall consider the case of $r\geq p$. Put

$$G_2(n,r)$$
$$=B_p\left(\frac{n}{2}-\alpha,\frac{r}{2}\right)B_p\left(\frac{n}{2}+\lambda+\alpha,\frac{r}{2}\right)\Big/\left\{B_p\left(\frac{n}{2}+\alpha,\frac{r}{2}\right)B_p\left(\frac{n}{2}+\lambda-\alpha,\frac{r}{2}\right)\right\}.$$

Then $C_p(\lambda,\varepsilon,\alpha,n,\mathrm{r})>0$ if and only if $G_2(n,r)>1$. Denote $E_i^*[\cdot]$ is the expectation according to the probability measure P_i^* given by $P_i^*(A)=\int_A z^{f_i-1}(1-z)^{r/2-1}dz/B(f_i,r/2)$ for $f_i=(n-i+1)/2-\alpha$. $G_2(n,r)$ is represented as

$$G_2(n,r)=\prod_{i=1}^{p}\left[B\left(f_i,\frac{r}{2}\right)B\left(f_i+2\alpha+\lambda,\frac{r}{2}\right)\Big/\left\{B\left(f_i+2\alpha,\frac{r}{2}\right)B\left(f_i+\lambda,\frac{r}{2}\right)\right\}\right]$$
$$=\prod_{i=1}^{p}\left[E_i^*[Z^{2\alpha+\lambda}]/\{E_i^*[Z^{2\alpha}]E_i^*[Z^{\lambda}]\}\right],$$

which is greater than 1 by the same arguments as in the proof of Lemma A.1. Similarly, we can prove the case of $r<p$.

REFERENCES

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss, *Ann. Statist.*, **13**, 1581–1591.

Haff, L. R. (1979). An identity for the Wishart distribution with applications, *J. Multivariate Anal.*, **9**, 531–544.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix, *Ann. Statist.*, **8**, 586–597.

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

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 361–380, Univ. of California Press.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Olkin, I. and Selliah, J. B. (1977). Estimating covariances in a multivariate normal distribution, *Statistical Decision Theory and Related Topics*, II, (eds. S. S. Gupta and D. S. Moore), 313–326, Academic Press, New York.

Pal, N. (1988). Decision-theoretic estimation of generalized variance and generalized precision, *Comm. Statist. A—Theory Methods*, **17**, 4221–4230.

Shorrock, R. W. and Zidek, J. V. (1976). An improved estimator of the generalized variance, *Ann. Statist.*, **4**, 629–638.

Sinha, B. K. (1976). On improved estimators of the generalized variance, *J. Multivariate Anal.*, **6**, 617–625.

Sinha, B. K. and Ghosh, M. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss, *Statist. Decisions*, **5**, 201–227.

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.

Strawderman, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss, *Ann. Statist.*, **2**, 190–198.

Sugiura, N. (1988). A class of improved estimators of powers of the generalized variance and precision under squared loss, *Statistical Theory and Data Analysis II*, (ed. K. Matusita), North-Holland, Amsterdam.

Sugiura, N. (1989). Entropy loss and a class of improved estimators for powers of the generalized variance, (to appear in *Sankhyā*).

Sugiura, N. and Konno, Y. (1987). Risk of improved estimators for generalized variance and precision, *Advances in Multivariate Statistical Analysis,* (ed. A. K. Gupta), 353–371, Reidel, Dordrecht.

Sugiura, N. and Konno, Y. (1988). Entropy loss and risk of improved estimators for the generalized variance and precision, *Ann. Inst. Statist. Math.*, **40**, 329–341.