Ann. Inst. Statist. Math.
Vol. 50, No. 3, 523–530 (1998)

ESTIMATION OF THE SCALE MATRIX AND ITS EIGENVALUES IN THE WISHART AND THE MULTIVARIATE F DISTRIBUTIONS

PUI LAM LEUNG AND WAI YIN CHAN

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong

(Received June 12, 1996; revised April 25, 1997)

Abstract. In this paper, the problem of estimating the scale matrix and their eigenvalues in a Wishart distribution and in a multivariate F distribution (which arise naturally from a two-sample setting) are considered. A new class of estimators which shrink the eigenvalues towards their arithmetic mean are proposed. It is shown that the new estimator which dominates the usual unbiased estimator under the squared error loss function. A simulation study was carried out to study the performance of these estimators.

Key words and phrases: Covariance matrix, orthogonally invariant estimator, decision-theoretic estimation, shrinkage estimator.

1. Introduction and summary

There has been considerable research on the problem of estimating the covariance matrix Σ and their eigenvalues $\omega_1, \dots, \omega_m$ ($\omega_1 \geq \dots \omega_m \geq 0$) in a multivariate normal distribution using a decision-theoretic approach. Excellent reviews on this topic can be found in Muirhead (1987) and Pal (1993). It is shown that substantial improvement (reduction in risks) over the usual unbiased estimator of Σ can be obtained, essentially by focusing attention on the problem of estimating the eigenvalues of Σ by functions of all the eigenvalues of Σ. Works along this direction can be found in Stein (1975), Haff (1980), Dey and Srinivasan (1985), Lin and Perlman (1985) and Dey (1988).

Suppose that a random $m \times m$ positive definite matrix S has a nonsingular Wishart distribution with unknown scale matrix Σ and n degrees of freedom, i.e., $S \sim W_m(n, \Sigma)$. Stein (1975) considered the class of orthogonally invariant estimators of Σ of the form

$$\hat{\Sigma} = H\Phi(L)H' \tag{1.1}$$

where $S = HLH'$ with H the matrix of normalized eigenvectors ($HH' = H'H = I_m$), $L = \mathrm{diag}(l_1, \dots, l_m)$ is the diagonal matrix of eigenvalues of S with $l_1 > \dots > l_m > 0$ and $\Phi(L) = \mathrm{diag}(\phi_1(L), \dots, \phi_m(L))$, $\phi_i(L) \geq 0$ is a real valued function, $i = 1, \dots, m$.

In this paper, we consider the problem of estimating Σ using the loss function

$$L(\Sigma, \hat{\Sigma}) = \mathrm{tr}(\hat{\Sigma} - \Sigma)^2 \tag{1.2}$$

which is the natural multivariate extension of the squared error loss function. Although there are several plausible loss functions available, we only considered the loss function (1.2) in this paper for simplicity and convenience. We proposed a new estimator of the form

$$\hat{\Sigma}_\alpha = \frac{\alpha}{n} S + \frac{1-\alpha}{mn}(\mathrm{tr}\, S) I_m \tag{1.3}$$

where $0 \le \alpha \le 1$. Note that for $\alpha = 1$, $\hat{\Sigma}_\alpha$ corresponds to the usual unbiased estimator $\hat{\Sigma}_U = (1/n)S$ and $\hat{\Sigma}_\alpha$ is in the class of orthogonally invariant estimators defined in (1.1) with eigenvalues

$$\phi_i(L) = \frac{\alpha}{n} l_i + \frac{1-\alpha}{n}\bar{l}, \qquad (i = 1, \ldots, m) \tag{1.4}$$

where $\bar{l}$ is the average of $l_1, \ldots, l_m$, the eigenvalues of S. This estimator is motivated by the fact that the sample eigenvalues usually tend to be much more dispersed than the population eigenvalues (see Muirhead (1987)), i.e., $(1/n)l_1$ tends to over-estimate ω_1 and $(1/n)l_m$ tends to under-estimate ω_m. Intuitively, the unbiased estimator can be improved by shrinking the sample eigenvalues towards some central value. From equation (1.4), it is easy to see that the eigenvalue $\phi_i(L)$ is a linear combination of l_i and $\bar{l}$. α is the shrinkage parameter ranging from 0 to 1 representing various degrees of shrinkage. This particular form of shrinkage is also appeared in equation (18) in Friedman (1989) although he used this shrinkage in a different context, namely the discriminant analysis which he coined the name 'regularized discriminant analysis'. In Section 2, we stated the sufficient condition on α such that the unbiased estimator $\hat{\Sigma}_U$ is dominated by $\hat{\Sigma}_\alpha$.

Another closely related problem is the estimation of the scale matrix Δ in a multivariate F distribution. This problem has been considered by various authors, namely, Muirhead and Verathaworn (1985), Leung and Muirhead (1988), Dey (1989), Gupta and Krishnamoorthy (1990), Konno (1991) and Leung (1992). This problem arise naturally from a two-sample setting (see Muirhead and Verathaworn (1985) and Leung and Muirhead (1988) for details).

Suppose that a random $m \times m$ positive definite matrix F has a multivariate F distribution with degrees of freedom n_1 and n_2 and scale matrix Δ, i.e., $F \sim F_m(n_1, n_2; \Delta)$. An unbiased estimator of Δ is $\hat{\Delta}_U = cF$, where $c = (n_2 - m - 1)/n_1$. Using a similar approach as in the Wishart situation, we proposed a new class of orthogonally invariant estimator of the form

$$\hat{\Delta}_\alpha = \alpha c F + \frac{(1-\alpha)c}{m}(\mathrm{tr}\, F) I_m \tag{1.5}$$

where $0 \le \alpha \le 1$. In Section 3, we proved that the unbiased estimator $\hat{\Delta}_U$ is dominated by $\hat{\Delta}_\alpha$ for some values of α using the squared error loss function

$$L(\Delta, \hat{\Delta}) = \mathrm{tr}(\hat{\Delta} - \Delta)^2. \tag{1.6}$$

Although the dominance result in Sections 2 and 3 are focused on estimating the unknown matrix Σ or Δ, we believe that the estimation of the unknown eigenvalues of Σ or Δ could also be improved by using the eigenvalues of $\hat{\Sigma}_\alpha$ or $\hat{\Delta}_\alpha$ respectively. Works on simultaneous estimation of eigenvalues directly can be found in Dey (1988) and Jin (1993). The loss function they used (which is analogous to (1.2)) is the sum of squared errors loss function

$$L(\delta, \hat{\delta}) = \sum_{i=1}^{m} (\hat{\delta}_i - \delta_i)^2 \tag{1.7}$$

where $\hat{\delta}_i$ are the eigenvalues of $\hat{\Sigma}_\alpha$ (or $\hat{\Delta}_\alpha$) and δ_i are the eigenvalues of Σ (or Δ). It is easy to see that if the unknown matrix Σ (or Δ) is of the form kI_m (k is constant), then the loss incurred by (1.2) (or by (1.6)) is equal to the loss incurred by (1.7).

In Section 4, a Monte Carlo simulation is carried out to study the performance of the proposed estimators in Section 2 and Section 3 using loss functions (1.2), (1.6) and (1.7).

2. Improved estimation of Σ

Suppose that a random $m \times m$ positive definite matrix S has a nonsingular Wishart distribution with unknown scale matrix Σ and n degrees of freedom, i.e. $S \sim W_m(n, \Sigma)$. Let $\hat{\Sigma}_U = (1/n)S$ be the usual unbiased estimate of Σ. The main result in this section is to provide a sufficient condition on α such that $\hat{\Sigma}_U$ is dominated by $\hat{\Sigma}_\alpha$ defined in (1.3).

THEOREM 2.1. *Using the loss function* (1.2), *the unbiased estimate* $\hat{\Sigma}_U$ *is dominated by* $\hat{\Sigma}_\alpha$ *defined in* (1.3) *for all nonnegative definite matrix* Σ *provided that* $m > 1$ *and* $(n-2)/(n+2) \le \alpha \le 1$.

This theorem can be considered as a limiting case of Theorem 3.2, the proof is deferred to Remark 3 after the proof of Theorem 3.2 in Section 3. Furthermore, although we can choose any α value between $(n-2)/(n+2)$ and 1, we suggest using $\alpha_1 = n/(n+2)$ and the corresponding regularized estimator of Σ is

$$\hat{\Sigma}_R = \frac{1}{n+2} S + \frac{2}{mn(n+2)} (\operatorname{tr} S) I_m \tag{2.1}$$

with eigenvalues

$$\phi_i(L) = \frac{1}{n+2} l_i + \frac{2}{n(n+2)} \bar{l} \quad (i = 1, \dots, m). \tag{2.2}$$

Again the justification of using α_1 is deferred to the next section.

3. Improved estimation of Δ

Closely related to the problem in Section 2 is the problem of estimating the unknown scale matrix Δ of a multivariate F distribution. Let $A \sim W_m(n_1, \Delta)$ and independent of $B \sim W_m(n_2, I)$. Then the random $m \times m$ positive definite matrix $F = A^{1/2}B^{-1}A^{1/2}$ has a multivariate F distribution with degrees of freedom n_1 and n_2 and scale matrix Δ, i.e., $F \sim F_m(n_1, n_2; \Delta)$. Throughout this section, we assume that $n_1 > m+1$ and $n_2 > m+3$. It is shown in Muirhead and Verathaworn (1985) that an unbiased estimator of Δ is $\hat{\Delta}_U = cF$, where $c = (n_2 - m - 1)/n_1$. Similar to Section 2, the main result in this section is to provide a sufficient condition on α such that $\hat{\Delta}_U$ is dominated by $\hat{\Delta}_\alpha$ defined in (1.5). Before we state and prove this result, we need the following lemma.

LEMMA 3.1. *Let $F \sim F_m(n_1, n_2; \Delta)$. Then*

$$\text{(i)} \qquad E[\operatorname{tr}(F^2)] = \frac{n_1}{k_0 k_1 k_3}\{[k_1(n_1+1)+2]\operatorname{tr}(\Delta^2) + (n_1+k_1)(\operatorname{tr}\Delta)^2\},$$

$$\text{(ii)} \qquad E[(\operatorname{tr} F)^2] = \frac{n_1}{k_0 k_1 k_3}\{2(n_1+k_1)(\operatorname{tr}\Delta^2) + (k_2 n_1 + 2)(\operatorname{tr}\Delta)^2\}$$

where $k_i = n_2 - m - i$.

PROOF. The proof of (i) given in Konno (1988), Corollary 2.4. (ii) can be proved using the Wishart identity (see Haff (1979)) or by using similar technique as in Konno (1988). Both methods are reasonably straight forward and hence are omitted.

THEOREM 3.2. *Using the loss function* (1.6), *the unbiased estimate $\hat{\Delta}_U$ is dominated by $\hat{\Delta}_\alpha$ defined in* (1.5) *for all nonnegative definite matrix Δ provided that $m > 1$ and*

$$\max\left\{\frac{n_1-2}{n_1+2} - \frac{4n_1}{(n_1+2)(n_2-m-1)}, 0\right\} \le \alpha \le 1.$$

PROOF. Using Lemma 3.1, it can be shown that the risk difference between $\hat{\Delta}_U$ and $\hat{\Delta}_\alpha$ is

$$\begin{aligned} H(\Delta) &= E[L(\Delta, \hat{\Delta}_U)] - E[L(\Delta, \hat{\Delta}_\alpha)] \\ &= \frac{c^2(1-\alpha^2)n_1}{mk_0k_1k_3}[a_1\operatorname{tr}(\Delta^2) + a_2(\operatorname{tr}\Delta)^2] \end{aligned}$$

where $c = (n_2 - m - 1)/n_1$,

$$\begin{aligned} a_1 &= m[(n_1+1)k_1+2] - 2(n_1+k_1) - \frac{2mn_1k_0k_3}{(1+\alpha)k_1}, \\ a_2 &= m(n_1+k_1) - n_1k_2 - 2 + \frac{2n_1k_0k_3}{(1+\alpha)k_1}. \end{aligned}$$

Using the fact that $(\operatorname{tr}\Delta)^2 \geq \operatorname{tr}(\Delta^2)$ and $a_2 > 0$, we obtain a lower bound for $H(\Sigma)$ as follow:

$$H(\Delta) \geq \frac{(m-1)(1-\alpha^2)k_1}{mn_1k_3}\left[n_1 + 2 - \frac{2n_1k_3}{(1+\alpha)k_1}\right]\operatorname{tr}(\Delta^2). \tag{3.1}$$

A sufficient condition for $H(\Delta) \geq 0$ is that $0 \leq \alpha \leq 1$ and the term in the square bracket is nonnegative. This completes the proof.

Remark 1. The lower bound of $H(\Delta)$ in the right hand side of (3.1) is maximized at $\alpha_2 = n_1(n_2 - m - 3)/[(n_1 + 2)(n_2 - m - 1)]$ and this α_2 always lies between 0 and 1. The corresponding regularized estimator of Δ is

$$\hat{\Delta}_R = \frac{n_2 - m - 3}{n_1 + 2}F + \frac{2(n_1 + n_2 - m - 1)}{mn_1(n_1 + 2)}(\operatorname{tr}F)I_m \tag{3.2}$$

with eigenvalues

$$\phi_i(L) = \frac{n_2 - m - 3}{n_1 + 2}l_i + \frac{2(n_1 + n_2 - m - 1)}{n_1(n_1 + 2)}\bar{l}, \qquad (i = 1, \ldots, m) \tag{3.3}$$

where $\bar{l}$ is the average of $l_1, \ldots, l_m$, the eigenvalues of F.

Remark 2. When $n_2 - m - 1 < 4n_1/(n_1 - 2)$, the sufficient condition in Theorem 3.2 becomes $0 \leq \alpha \leq 1$. In particular, $\hat{\Delta}_U$ is dominated by $[(n_2 - m - 1)/(n_1 m)](\operatorname{tr}F)I_m$. An intuitive justification is that when n_2 is small, the variance of the off-diagonal elements of F is too large to provide useful information about Δ. We will be better off to use only the diagonal elements of F, or $\operatorname{tr}F$, in estimating Δ.

Remark 3. Note that $F = A^{1/2}B^{-1}A^{1/2}$ where $A \sim W_m(n_1, \Delta)$ and independent of $B \sim W_m(n_2, I)$. B can be considered as the sum of squares and cross product matrix of n_2 independently and identically distributed $m \times 1$ standard normal random vectors. By strong law of large numbers, $n_2^{-1}B$ converges to I_m or n_2F converges to A almost surely, as n_2 tends to infinity. This becomes the problem of estimating the scale matrix in a Wishart distribution considered in Section 2. Furthermore, n_2F is uniformly integrable so exchange between $\lim_{n_2\to\infty}$ and expectation is possible. Therefore, the result in Theorem 2.1 can be obtained by passing the limit as $n_2 \to \infty$ in the results in Theorem 3.2. In particular, as n_2 tends to infinity, the sufficient condition in Theorem 3.2 becomes $(n_1 - 2)/(n_1 + 2) \leq \alpha \leq 1$ which is the same sufficient condition stated in Theorem 2.1 when we take $n_1 = n$ and $\Delta = \Sigma$. The value α_2 in Remark 1 becomes $n_1/(n_1 + 2)$ which is same as the α_1 defined in Section 2.

4. Simulation study

For estimating the scale matrix Σ in the Wishart distribution, a Monte Carlo simulation study was carried out to compare the risks of $\hat{\Sigma}_U$ and $\hat{\Sigma}_R$ defined in (2.1). For $m = 3$ and $n = 5, 10, 25$, a sample of 1000 Wishart $W_3(n, \Sigma)$ matrices were generated for three different choices of Σ. Then these 1000 matrices were used to construct $\hat{\Sigma}_U$ and $\hat{\Sigma}_R$ and from these average losses (with respect to the loss function (1.2)) were obtained. The design used here replicates the setting used in Dey (1988) in order to compare with his simulation results directly.

Table 1 summarizes these results. In this table the value given for each combination of Σ and n is the percentage reduction in average loss (PRIAL) for $\hat{\Sigma}_R$ compared with $\hat{\Sigma}_U$, i.e., it is the estimate of

$$\frac{E[L(\Sigma, \hat{\Sigma}_U) - L(\Sigma, \hat{\Sigma}_R)]}{E[L(\Sigma, \hat{\Sigma}_U)]} \times 100.$$

Table 2 shows the PRIAL for the eigenvalues of $\hat{\Sigma}_R$ given in equation (2.2) using the loss function (1.7), i.e., it is the sum of squared difference between $\phi_i(L)$ in (2.2) and the eigenvalues of Σ. The results show that the PRIAL in Tables 1 and 2 are very similar and in fact they are equal for the case $\Sigma = I_3$ as mentioned in Section 1. The PRIAL is large especially when n is small. The reduction in risk is larger than those given in Table 1 in Dey (1988).

Next, we consider the problem of estimating the scale matrix Δ in a multivariate F distribution. For $m = 4$ and $n_1 = n_2 = 10, 15, 20$, 1000 matrices of A's and B's are generated independently from $W_4(n_1, \Delta)$ and $W_4(n_2, I_4)$ respectively for three different choices of Δ. Again these Δ are taken from Dey (1988) for direct comparison. They are then transformed into

$$F = A^{1/2} B^{-1} A^{1/2}.$$

Table 1. PRIAL of $\hat{\Sigma}_R$ over $\hat{\Sigma}_U$ using loss function (1.2).

Σ	$n = 5$	$n = 10$	$n = 25$
diag(1, 1, 1)	40.939	25.460	11.900
diag(4, 2, 1)	36.828	22.512	11.143
diag(25, 1, 1)	22.977	13.376	7.437

Table 2. PRIAL of $\hat{\Sigma}_R$ over $\hat{\Sigma}_U$ using loss function (1.7).

Σ	$n = 5$	$n = 10$	$n = 25$
diag(1, 1, 1)	40.939	25.460	11.900
diag(4, 2, 1)	38.371	24.334	13.249
diag(25, 1, 1)	22.530	13.138	7.440

Table 3. PRIAL of $\hat{\Delta}_R$ over $\hat{\Delta}_U$ using loss function (1.6).

Δ	$n_1 = n_2 = 10$	$n_1 = n_2 = 15$	$n_1 = n_2 = 20$
$\text{diag}(1,1,1,1)$	63.129	42.829	33.065
$\text{diag}(8,4,2,1)$	55.484	37.482	28.484
$\text{diag}(25,1,1,1)$	40.877	27.515	20.133

Table 4. PRIAL of $\hat{\Delta}_R$ over $\hat{\Delta}_U$ using loss function (1.7).

Δ	$n_1 = n_2 = 10$	$n_1 = n_2 = 15$	$n_1 = n_2 = 20$
$\text{diag}(1,1,1,1)$	63.129	42.829	33.065
$\text{diag}(8,4,2,1)$	56.393	39.683	31.247
$\text{diag}(25,1,1,1)$	37.968	26.839	19.708

Table 3 shows the PRIAL for $\hat{\Delta}_R$ defined in (3.2) compared with the unbiased estimator $\hat{\Delta}_U$ using the loss function (1.6). Similarly, Table 4 shows the PRIAL for the eigenvalues of $\hat{\Delta}_R$ given in equation (3.3) using loss function (1.7). Again the results in Tables 3 and 4 are very similar although the PRIAL's are not as large as those given in Table 2 in Dey (1988).

References

Dey, D. K. (1988). Simultaneous estimation of eigenvalues, *Ann. Inst. Statist. Math.*, **40**, 137–147.

Dey, D. K. (1989). On estimation of the scale matrix of the multivariate F distribution, *Comm. Statist. Theory Methods*, **18**, 1373–1383.

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

Friedman, J. H. (1989). Regularized discriminant analysis, *J. Amer. Statist. Assoc.*, **84**, 165–175.

Gupta, A. K. and Krishnamoorthy, K. (1990). Improved estimators of eigenvalues of $\Sigma_1\Sigma_2^{-1}$, *Statist. Decisions*, **8**, 247–263.

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.

Jin, C. (1993). A note on simultaneous estimation of eigenvalues of a multivariate normal covariance matrix, *Statist. Probab. Lett.*, **16**, 197–203.

Konno, Y. (1988). Exact moments of the multivariate F and beta distributions, *J. Japan Statist. Soc.*, **18**, 123–130.

Konno, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F-distribution, *Ann. Inst. Statist. Math.*, **43**, 157–165.

Leung, P. L. (1992). Estimation of eigenvalues of the scale matrix of the multivariate F distribution, *Comm. Statist. Theory Methods*, **21**, 1845–1856.

Leung, P. L. and Muirhead, R. J. (1988). On estimating characteristic roots in a two-sample problem, *Statistical Decision Theory and Related Topic IV*, Vol. 1 (eds. S. S. Gupta and J. O. Berger), 397–402, Springer, New York.

Lin, S. P. and Perlman, M. D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix, *Multivariate Analysis VI* (ed. P. R. Krishnaiah), 411–429, North Holland, Amsterdam.

Muirhead, R. J. (1987). Developments in eigenvalue estimation, *Advances in Multivariate Statistical Analysis*, 277–288.

Muirhead, R. J. and Verathaworn, T. (1985). On estimating the latent roots of $\Sigma_1\Sigma_2^{-1}$, *Multivariate Analysis VI* (ed. P. R Krishnaiah), 431–447, North-Holland, Amsterdam.

Pal, N. (1993). Estimating the normal dispersion matrix and the precision matrix from a decision-theoretic point of view: a review, *Statistical Papers*, **34**, 1–26.

Stein, C. (1975). Estimation of a covariance matrix, *Rietz Lecture*, *39th Annual meeting IMS*, Atlanta, Georgia.