Ann. Inst. Statist. Math.
38 (1986), Part A, 101-108

TRIMMED MINIMAX ESTIMATOR OF A COVARIANCE MATRIX

DIPAK K. DEY AND C. SRINIVASAN*

(Received May 4, 1984; revised July 26, 1985)

Summary

In the problem of estimating the covariance matrix of a multivariate normal population, James and Stein (*Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1, 361-380, Univ. of California Press) obtained a minimax estimator under a scale invariant loss. In this paper we propose an orthogonally invariant trimmed estimator by solving certain differential inequality involving the eigenvalues of the sample covariance matrix. The estimator obtained, truncates the extreme eigenvalues first and then shrinks the larger and expands the smaller sample eigenvalues. Adaptive version of the trimmed estimator is also discussed. Finally some numerical studies are performed using Monte Carlo simulation method and it is observed that the trimmed estimate shows a substantial improvement over the minimax estimator.

1. Introduction

In this paper we consider the problem of estimating the normal covariance matrix under a scale invariant loss function introduced in James and Stein [12]. The usual estimator is S/k, where S is distributed according to the Wishart distribution $W_p(\Sigma, k)$. Although S/k is unbiased it is known that the sample eigenvalues of S tend to be more spread out than the population eigenvalues of Σ. This fact suggests that, Stein's phenomenon should be observed in estimation of Σ. The improved estimator (over S/k) are obtained in Stein [15], [16], Efron and Morris [5], Haff [7]-[11] and a host of others.

Since S/k itself is not minimax, it is more meaningful to consider the minimax estimators. James and Stein [12], obtained minimax estimator by considering the best invariant estimator with respect to the triangular group G_T^+ (the group consisting of lower triangular matrices

* The second author's research was supported by NSF Grant Number MCS 82-12968.
Key words and phrases: Covariance matrix, Wishart distribution, Stein's loss, minimax estimator, Stein's truncation, adaptive estimator.

with positive diagonal elements). This estimator, of course, depends on the coordinate system. Later on, Takemura [18] obtained an orthogonally invariant minimax estimator by averaging Stein's minimax estimator over the $p\times p$ orthogonal matrices with respect to Haar measure. However for higher dimensions ($p\geqq 3$), his estimator does not have a simple form. Recently Dey and Srinivasan [4] obtained an orthogonally equivariant minimax estimator of Σ which is expressable in closed form for any dimension.

In this paper our objective is to find an orthogonally invariant minimax estimator of Σ, having a simple form, which will truncate the extreme eigenvalues. This estimator will be very useful for practical application. We will consider the scale invariant loss function, considered in James and Stein [12], which is given as

$$L(\hat{\Sigma}, \Sigma)=\operatorname{tr}(\hat{\Sigma}\Sigma^{-1})-\log\det(\hat{\Sigma}\Sigma^{-1})-p\ . \tag{1.1}$$

The loss given in (1.1) was introduced and motivated in Section 5 of James and Stein [12], and will be referred to as Stein's loss. To aid to one's intuition about L, note that $L(\hat{\Sigma}_0, \Sigma_0)$ is the likelihood ratio statistic for testing $H_0:\Sigma=\Sigma_0$ against all alternatives. It is also interesting to observe that under the loss (1.1), the best invariant estimator coincides with the best unbiased estimator.

Suppose $S=RLR^t$ in which $RR^t=R^tR=I$ and $L=\operatorname{diag}(l_1, l_2, \cdots, l_p)$ is the diagonal matrix of eigenvalues, with $l_1\geqq l_2\geqq\cdots\geqq l_p$. The class of estimators which are considered is

$$\hat{\Sigma}=R\phi(L)R^t \tag{1.2}$$

where $\phi(L)=\operatorname{diag}(\phi_1(L), \phi_2(L), \cdots, \phi_p(L))$. Clearly it follows that $\phi^{(0)}(L)=(1/k)L$ determines the best invariant and unbiased estimator $\hat{\Sigma}_0=S/k$. The performance of the estimator is evaluated by considering the risk function given as $R(\hat{\Sigma}, \Sigma)=E(L(\hat{\Sigma}, \Sigma)\,|\,\Sigma)$.

In Section 2, using Stein [15] and Haff [11], we will express the risk difference of two estimators as a differential inequality involving sample eigenvalues. The solution of the differential inequality is obtained, which is motivated by Dey and Srinivasan [4] and Stein's [17] truncation method. The truncation method, proposed in Stein [17], gives rise to a minimax estimator of the multinormal mean, which is extremely effective in the presence of outlier. This is further studied in Dey and Berger [2] and extended in estimation of parameters from continuous and discrete exponential families by Ghosh and Dey [6].

In Section 3, we propose an adaptive trimmed estimate of Σ and use Monte Carlo simulation method to study the risk behavior of the estimate. The percentage improvements are computed with respect to the risk of best invariant estimator and the minimax risk.

2. Derivation of the trimmed minimax estimator

Assume the loss function and the class of estimators are of the form (1.1) and (1.2). We first observe that

$$L(\hat{\Sigma}, \Sigma)=\operatorname{tr}(\hat{\Sigma}\Sigma^{-1})-\sum_{i=1}^{p}\log\phi_i(L)+\log\det(\Sigma)-p\,. \tag{2.1}$$

Clearly the last two terms are constant with respect to $\hat{\Sigma}$. Thus we define

$$R^*(\hat{\Sigma}, \Sigma)=E\left[\operatorname{tr}(\hat{\Sigma}\Sigma^{-1})-\sum_{i=1}^{p}\log\phi_i(L)\,|\,\Sigma\right]. \tag{2.2}$$

The following lemma gives the unbiased estimate of $R^*(\hat{\Sigma}, \Sigma)$. The proof is given in Stein [16] and Haff [11].

LEMMA 2.1. *The unbiased estimator of $R^*(\hat{\Sigma}, \Sigma)$ is given as*

$$\begin{aligned}\hat{R}^*(\hat{\Sigma}, \Sigma)=2\sum_{i=1}^{p}\sum_{t>i}(\phi_i-\phi_t)/(l_i-l_t)+2\sum_{i=1}^{p}\partial\phi_i/\partial l_i\\ +(k-p-1)\sum_{i=1}^{p}\phi_i/l_i-\sum_{i=1}^{p}\log\phi_i\,.\end{aligned} \tag{2.3}$$

We now need the following lemma to obtain an upper bound of the risk difference. The proof of which is given in Dey and Srinivasan [4].

LEMMA 2.2. *For $|x|\leqq u<1$,*

$$\log(1+x)\geqq x-\frac{3-u}{6(1-u)}x^2\,. \tag{2.4}$$

Now following James and Stein [12], the minimax estimator is given as $\hat{\Sigma}^S=TDT^t$, where $D=\operatorname{diag}(d_1, d_2, \cdots, d_p)$ and $T\in G_T^+$ with $TT^t=S$. By choosing $d_i=1/(k+p+1-2i)$, the minimax risk is given as

$$R(\hat{\Sigma}^S, \Sigma)=-\sum_{i=1}^{p}\log d_i-p\log 2-\sum_{i=1}^{p}\psi\left(\frac{k+1-i}{2}\right), \tag{2.5}$$

where $\psi(a)=\Gamma'(a)/\Gamma(a)$ is the digamma function.

The following theorem gives a minimax estimator of Σ. The proof is given in Dey and Srinivasan [4].

THEOREM 2.1. *Consider the estimator $\hat{\Sigma}^M=R\phi^M(L)R^t$ where $\phi^M(L)$ is given componentwise as*

(2.6) $$\phi_i^M(L)=l_i d_i\,,\qquad i=1,2,\cdots,p\,.$$

Then $R(\hat{\Sigma}^M,\Sigma)\leq R(\hat{\Sigma}^S,\Sigma)$ *and hence* $\hat{\Sigma}^M$ *is minimax.*

Now we will develop the trimmed minimax estimator of Σ. First we need the following notations. Define,

$$Y_i=\log l_i\,,\qquad i=1,2,\cdots,p$$

and

$$|Y|_{(1)}<\cdots<|Y|_{(m)}<\cdots<|Y|_{(p)}$$

be the ordered $|Y_i|$'s, where $m=[yp]$ and $0<y<1$. Also define

$$Z_i=(\operatorname{sgn} Y_i)\{|Y_i|\wedge|Y|_{(m)}\}\,,\qquad i=1,2,\cdots,p\,,$$

where $a\wedge b=\min(a,b)$.

THEOREM 2.2. *Consider the estimator* $\hat{\Sigma}^m=R\phi^m(L)R^t$. *If* $\phi^m(L)$ *is given componentwise as*

(2.7) $$\phi_i^m(L)=\begin{cases}\phi_i^M(L)-\dfrac{l_i\tau(|Z|^2)}{b+|Z|^2}\log l_i\,, & \text{if } Y_i^2\leq|Y|_{(m)}^2\\[2ex] \phi_i^M(L)-\dfrac{l_i\tau(|Z|^2)}{b+|Z|^2}|Y|_{(m)}(\operatorname{sgn}\log l_i)\,, & \text{if } Y_i^2>|Y|_{(m)}^2\end{cases}$$

where $\phi_i^M(L)=l_i d_i$, $b>144(m-2)^2/25(k+p-1)^2$ *and* $\tau(|Z|^2)$ *is a function satisfying*

(2.8)
(i) $0<\tau(|Z|^2)<12(m-2)/5(k+p-1)^2$
and
(ii) $\tau(|Z|^2)$ *monotone nondecreasing in* $|Z|^2$ *and* $E[\tau'(|Z|^2)]<\infty$;

then $R(\hat{\Sigma}^m,\Sigma)\leq R(\hat{\Sigma}^M,\Sigma)$ *and hence* $\hat{\Sigma}^m$ *is minimax.*

PROOF. Define,

$$g(|Z|^2)=\frac{\tau(|Z|^2)}{b+|Z|^2}\,(>0)\,,\qquad\text{and}\qquad \eta_i=-g(|Z|^2)Z_i\,,\quad i=1,2,\cdots,p\,.$$

Then from (2.7), it follows that

(2.9) $$\phi_i^m(L)=l_i d_i+l_i\eta_i=l_i d_i+\gamma_i\ \text{(say)}\,,\qquad i=1,\cdots,p\,.$$

Now define $\alpha(L)$ to be the unbiased estimator of the risk difference of $\hat{\Sigma}^m$ with respect to $\hat{\Sigma}^M$, that is, $E[\alpha(L)]=R(\hat{\Sigma}^m,\Sigma)-R(\hat{\Sigma}^M,\Sigma)$. Using (2.9) and Lemma 2.1 it follows that

(2.10) $$\alpha(L)=2\sum_{i=1}^{p}\sum_{t>i}(l_i\eta_i-l_t\eta_t)/(l_i-l_t)+2\sum_{i=1}^{p}\{\eta_i+l_i(\partial\eta_i/\partial l_i)\}$$
$$+(k-p-1)\sum_{i=1}^{p}\eta_i-\sum_{i=1}^{p}\log(1+\eta_i/d_i)\ .$$

Now it is easy to observe that

$$|\eta_i/d_i|\leqq\frac{\tau(|Z|^2)}{d_i(b+|Z|^2)}|Z_i|\leqq\frac{(k+p+1-2i)\tau(|Z|^2)}{2\sqrt{b}}$$
$$\leqq\frac{12(k+p-1)(m-2)}{10\sqrt{b}\,(k+p-1)^2}=\frac{6(m-2)}{5\sqrt{b}\,(k+p-1)}\leqq 1/2\ .$$

Hence using Lemma 2.2, with $u=1/2$, one gets

$$\log(1+\eta_i/d_i)\geqq\eta_i/d_i-(5/6)\eta_i^2/d_i^2\ .$$

Thus from (2.9), it follows that

$$\alpha(L)\leqq 2\sum_{i=1}^{p}\sum_{t>i}(l_i\eta_i-l_t\eta_t)/(l_i-l_t)+(k-p+1)\sum_{i=1}^{p}\eta_i-\sum_{i=1}^{p}\eta_i/d_i$$
$$+2\sum_{i=1}^{p}l_i(\partial\eta_i/\partial l_i)+(5/6)\sum_{i=1}^{p}\eta_i^2/d_i^2\ .$$

Let us define,

$$\alpha_1(L)=2\sum_{i=1}^{p}\sum_{t>i}(l_i\eta_i-l_t\eta_t)/(l_i-l_t)+(k-p+1)\sum_{i=1}^{p}\eta_i-\sum_{i=1}^{p}\eta_i/d_i\ ,$$

and

$$\alpha_2(L)=2\sum_{i=1}^{p}l_i(\partial\eta_i/\partial l_i)+(5/6)\sum_{i=1}^{p}\eta_i^2/d_i^2\ .$$

We will show that $\alpha_1(L)<0$ and $\alpha_2(L)\leqq 0$.

Substituting the value of d_i, it is clear that,

(2.11) $$\alpha_1(L)=2\sum_{i=1}^{p}\sum_{t>i}\eta_i+2\sum_{i=1}^{p}\sum_{t>i}l_t(\eta_i-\eta_t)/(l_i-l_t)-2\sum_{i=1}^{p}(p-i)\eta_i$$
$$=2\sum_{i=1}^{p}\sum_{t>i}l_t(\eta_i-\eta_t)/(l_i-l_t)\ .$$

Now observe that, $\eta_i-\eta_t=g(|Z|^2)(Z_t-Z_i)$ and for $t>i$, $Z_t-Z_i<0$ and $l_i-l_t>0$. Thus from (2.11), it follows that $\alpha_1(L)<0$.

Now consider $\alpha_2(L)$. Letting I_A denote the usual indicator function of set A, it is clear that

$$\frac{\partial Z_j}{\partial Y_i}=\begin{cases} I_{\{|Y_i|\leq|Y|_{(m)}\}} & \text{if } j=i\\ (\operatorname{sgn} Y_i)(\operatorname{sgn} Y_j)I_{\{|Y_j|>|Y_i|=|Y|_{(m)}\}} & \text{if } j\neq i\end{cases}$$

and hence

$$-\frac{\partial \eta_i(Y)}{\partial Y_i}=g(|Z|^2)I_{\{|Y_i|\leq |Y|_{(m)}\}}+2g'(|Z|^2)Z_i\{Z_iI_{\{|Y_i|\leq|Y|_{(m)}\}}$$
$$+\sum_{j\neq i} Z_j(\operatorname{sgn} Y_i)(\operatorname{sgn} Y_j)I_{\{|Y_j|>|Y_i|=|Y|_{(m)}\}}\} .$$

Now observing that,

$$|Z|^2=\sum_{i=1}^{m} Z_{(i)}^2+(p-m)Z_{(m)}^2=\sum_{i\ni Y_i^2\leq|Y|_{(m)}^2} Y_i^2+(p-m)|Y|_{(m)}^2$$

it clearly follows

$$\sum_{i=1}^{p} l_i\frac{\partial}{\partial l_i}[g(|Z|^2)Z_i]=[g(|Z|^2)m+2g'(|Z|^2)|Z|^2] .$$

Therefore,

$$\begin{aligned}\alpha_2(L)&=-\frac{2m\tau(|Z|^2)}{b+|Z|^2}-4|Z|^2\frac{\tau'(|Z|^2)(b+|Z|^2)-\tau(|Z|^2)}{(b+|Z|^2)^2}\\&\quad+(5/6)\frac{\tau^2(|Z|^2)}{(b+|Z|^2)^2}\sum_{i=1}^{p}(k+p+1-2i)^2Z_i^2\\&\leqq-\frac{2m\tau(|Z|^2)}{b+|Z|^2}-\frac{4|Z|^2\tau'(|Z|^2)}{b+|Z|^2}+\frac{4|Z|^2\tau(|Z|^2)}{(b+|Z|^2)^2}\\&\quad+(5/6)\frac{\tau^2(|Z|^2)}{b+|Z|^2}(k+p-1)^2|Z|^2\\&\leqq-\frac{2(m-2)\tau(|Z|^2)}{b+|Z|^2}+(5/6)\frac{(k+p-1)^2\tau^2(|Z|^2)}{b+|Z|^2}\\&=-\frac{\tau(|Z|^2)}{b+|Z|^2}\left\{2(m-2)-\frac{5(k+p-1)^2}{6}\tau(|Z|^2)\right\}\leqq 0\\&\qquad\qquad\text{(by (2.8))}.\end{aligned}$$

This completes the proof of the theorem.

Remark 2.1. The estimator given in (2.7) is a very simple minimax estimator, however ϕ_i^m's are not order preserving. Following Barlow et al. [1] one can get rid of this problem by performing an isotonic regression technique over ϕ_i^m's. See Lin [13] for a complete description of this modification.

3. An adaptive choice of trimming point

The estimator given in (2.7) has a very elegant form. However in the expression (2.7), the quantity m is not known. An appealing possibility is to let the data select m, the trimming point. Since the optimum choice of m should minimize the risk, therefore the obvious method of selection is to choose that $m\geqq 3$ (say m^*) which minimizes

$R(\hat{\Sigma}^m, \Sigma)$.

Theoretical analysis of this estimator is immensely difficult, due to the complicated dependence of m^* on the sample eigenvalues.

For $p=5$, 6, 10, and for several values of k, we generated 100 independent and identically distributed observations from a $W_p(I, k)$ distribution. We considered $\Sigma=I$, to obtain maximum percentage improvement in risks.

In Table 1, we consider our improved minimax estimator as given in (2.7) with $b=5.8(m-2)^2/(k+p-1)^2$ and $\tau(|Z|^2)$ is treated as a constant c with $c=6(m-2)/5(k+p-1)^2$. The numerical studies show that such a choice is optimum in terms of minimum risk criterion. We then compute the minimax risk $(R(\hat{\Sigma}^S))$ and the risk of the best invariant estimator $(R(\hat{\Sigma}_0))$ for different p and k. Then using Monte Carlo simulation method, we compute the risk of our trimmed minimax estimator $R(\hat{\Sigma}^m, \Sigma)$ for $m=3,\cdots, p$. Then we find m^* which minimizes $R(\hat{\Sigma}^m, \Sigma)$. The percentage improvements in risk of our adaptive trimmed estimator is then computed over $R(\hat{\Sigma}_0)$ and $R(\hat{\Sigma}^S)$. The average standard error is about .05. The numbers in parentheses are respectively $\{R(\hat{\Sigma}_0)-R(\hat{\Sigma}^{m^*})\}\times 100/R(\hat{\Sigma}_0)$ and $\{R(\hat{\Sigma}^S)-R(\hat{\Sigma}^{m^*})\}\times 100/R(\hat{\Sigma}^S)$.

Table 1 indicates that the percentage improvements are significant. Table 1 also indicates that our estimator uniformly dominates Takemura's [18] estimator in terms of risk.

Table 1. Risks, Optimum trimming point (m^*) and Percentage Improvements of $\hat{\Sigma}^{m^*}$ over $\hat{\Sigma}_0$ and $\hat{\Sigma}^S$.

k	$R(\hat{\Sigma}_0)$	$R(\hat{\Sigma}^S)$	m^*	$R(\hat{\Sigma}^{m^*})$
		$p=5$		
5	5.96	4.76	3	3.04(48.95, 36.13)
6	3.98	3.28	3	1.90(52.27, 42.00)
10	1.87	1.65	4	.88(52.80, 46.67)
15	1.14	1.05	4	.43(62.32, 59.04)
		$p=6$		
6	7.05	5.54	3	3.14(55.36, 43.27)
10	2.78	2.38	4	.97(64.94, 59.18)
15	1.65	1.49	6	.84(49.05, 43.47)
		$p=10$		
10	11.29	8.57	4	5.30(53.03, 38.09)
15	5.04	4.19	6	2.12(58.00, 49.43)

Acknowledgement

The authors wish to thank Mr. Subhendu Lahiri for the help in numerical computation. The authors are also grateful to the referee for suggestions which improved the presentation of this paper.

UNIVERSITY OF CONNECTICUT
UNIVERSITY OF KENTUCKY

REFERENCES

[1] Barlow, R. E. et al. (1972). *Statistical Inference Under Order Restrictions*, John Wiley and Sons, New York.

[2] Dey, D. K. and Berger, J. (1983). On truncation of shrinkage estimators in simultaneous estimation of normal means, *J. Amer. Statist. Ass.*, **78**, 865-869.

[3] Dey, D. K., Ghosh, M. and Srinivasan, C. (1986). Simultaneous estimation of parameters under entropy loss, to appear in *J. Statist. Plann. Inf.*

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

[5] Efron, B. and Morris, C. (1976). Multivariate empirical Bayes estimation of covariance matrices, *Ann. Statist.*, **4**, 22-32.

[6] Ghosh, M. and Dey, D. K. (1984). Trimmed estimates in simultaneous estimation of parameters in exponential families, *J. Multivariate Anal.*, **15**, 183-200.

[7] Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix, *J. Multivariate Anal.*, **7**, 374-385.

[8] Haff, L. R. (1979a). Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity, *Ann. Statist.*, **7**, 1264-1276.

[9] Haff, L. R. (1979b). An identity for the Wishart distribution with applications, *J. Multivariate Anal.*, **9**, 531-542.

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

[11] Haff, L. R. (1982). Solutions of the Euler-Lagrange equations for certain multivariate normal estimation problems, unpublished manuscript.

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

[13] Lin, S. P. (1977). Improved procedures for estimating a correlation matrix, unpublished Ph.D. thesis, Department of Statistics, University of Chicago.

[14] Lin, S. P. and Perlman, M. D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix, *Multivariate Analysis*, (ed. P. R. Krishnaiah), **6**, 411-429, North Holland, Amsterdam.

[15] Stein, C. (1975). Rietz lecture, *38th Annual Meeting IMS*, Atlanta, Georgia.

[16] Stein, C. (1977). Unpublished notes on estimating the covariance matrix.

[17] Stein, C. (1981). Estimation of the mean of multivariate normal distribution, *Ann. Statist.*, **9**, 1135-1151.

[18] Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population, *Tsukuba J. Math.*, **8**, 367-376.