King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 249

December 1999

Improved Estimators for the Scaled Convariance Matrix of the Scale Mixture of Multivariate Normal Distributions Under Entropy Loss

Anwar H. Joarder

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

IMPROVED ESTIMATORS FOR THE SCALED COVARIANCE MATRIX OF THE SCALE MIXTURE OF MULTIVARIATE NORMAL DISTRIBUTIONS UNDER ENTROPY LOSS

Anwar H. Joarder
Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Email: anwarj@kfupm.edu.sa

ABSTRACT

The covariance matrix of the scale mixture of multivariate normal distributions with unknown location vector is considered for estimation under entropy loss function to improve upon the usual estimators based on the sample sum of product matrix. The well-known results of the estimation of the covariance matrix of the multivariate normal model under the assumption of entropy loss function have been generalized to the proposed model.

Key Words and Phrases. Elliptical distributions; scale mixture of multivariate normal distributions; estimation of covariance matrix; entropy loss.

AMS 1991 Subject Classification: Primary 62H05, Secondary 62H12

1. INTRODUCTION

The scale mixture of multivariate normal distributions includes fat-tailed as well as thin-tailed distributions and has been used in modelling real world data. See for example Osiewalski and Steel (1993). The model is a broader class of distributions which accomodates multivariate normal and the t-distribution as special cases. On the one hand, it is a viable alternative to the usual multivariate normal distribution and on the other hand the results obtained under normality can be tested for its robustness using this model.

The estimation of the covariance matrix or its characteristics is important in multivariate analysis. However, the usual method of estimation of the covariance matrix is not meaningful when the observations follow a scale mixture of multivariate normal distributions.In fact maximum likelihood estimator is preferred to others because of its asymptotic normality which follows from the independence of the sample observations which is not necessarily true for our sample. We consider the estimation of the scaled covariance matrix of scale mixture of multivariate normal distributions under the entropy loss function.

The scale mixture of multivariate normal distributions considered in this paper is given by

$$f(\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N) = \int_0^\infty \frac{|\tau^2\mathbf{\Sigma}|^{-Np/2}}{(2\pi)^{Np/2}} exp\left(\frac{-1}{2}\sum_{j=1}^{N}(\mathbf{x}_j - \theta)'(\tau^2\mathbf{\Sigma})^{-1}(\mathbf{x}_j - \theta)\right) h(\tau)\, d\tau \tag{1.1}$$

where $\mathbf{X}_j = (\mathbf{X}_{1j}, \mathbf{X}_{2j}, \ldots, \mathbf{X}_{pj})', j = 1, 2, \ldots, N$. The p-dimensional column vector θ is the location parameter and the $p \times p$ positive definite matrix $\mathbf{\Sigma}$ is the scale matrix. The parameter τ is a nondiscrete positive random variable. The model in (1.1) is called the scale mixture of normal distributions as $\mathbf{X}_j|\tau \sim \mathcal{N}_p(\theta, \tau^2\mathbf{\Sigma}),\ j = 1, 2, \ldots, N$. The observations are not independent unless τ is degenerate with positive support. The model in (1.1) is a special case of broader class of elliptical distributions considered by Anderson, Fang and Hsu (1986).

Since the scale matrix $\mathbf{\Sigma}$ determines the covariance matrix up to a known constant $E(\tau^2) > 0$, we consider the estimation of $\mathbf{\Sigma}$ and hereinafter call it the scaled covariance matrix or simply the covariance matrix. In this paper we develop estimators for the scaled covariance matrix $\mathbf{\Sigma}$, the mean vector θ being unknown, under the entropy loss function

$$L(u(\mathbf{A}),\ \mathbf{\Sigma}) = tr(\mathbf{\Sigma}^{-1}u(\mathbf{A})) - ln|\mathbf{\Sigma}^{-1}u(\mathbf{A})| - p \tag{1.2}$$

where $u(\mathbf{A})$ is any estimator of $\mathbf{\Sigma}$ based on the sample sum of product matrix $\mathbf{A} = \sum_{j=1}^{N}\ (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'$ with $\bar{\mathbf{X}} = (\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \ldots, \bar{\mathbf{X}}_p)'$, $\bar{\mathbf{X}}_i = N^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij},\quad i = 1, 2, \ldots, p$. The above loss function is based on the form of the likelihood function and considered among others by James and Stein (1961).

In estimating $\mathbf{\Sigma}$ by $u(\mathbf{A})$, we consider the risk function $R(u(\mathbf{A}), \mathbf{\Sigma}) = E[L(u(\mathbf{A}), \mathbf{\Sigma})]$. An estimator $u_2(\mathbf{A})$ of $\mathbf{\Sigma}$ will be said to dominate another estimator $u_1(\mathbf{A})$ of $\mathbf{\Sigma}$ if for all $\mathbf{\Sigma}$ belonging to the class of positive definite matrices the inequality $R(u_2(\mathbf{A}), \mathbf{\Sigma}) \leq R(u_1(\mathbf{A}), \mathbf{\Sigma})$ holds, and the inequality $R(u_2(\mathbf{A}), \mathbf{\Sigma}) < R(u_1(\mathbf{A}), \mathbf{\Sigma})$ holds for at least one $\mathbf{\Sigma}$.

In Section 2, we consider estimators which are multiples of the sample sum of product matrix i.e. estimators of the type $cA,\ c > 0$, where $\mathbf{A}$ is the sample sum of product matrix based on the multivariate elliptical model. In Section 3, we consider estimators based on the lower triangular decomposition of the sample sum of product matrix i.e. estimators of the type $\mathbf{T\Delta T}'$ where $\mathbf{T}$ is a lower triangular matrix such that $\mathbf{A} = \mathbf{TT}'$ and $\mathbf{\Delta}$ is an arbitrary positive definite diagonal matrix with diagonal elements δ_i $(i = 1, 2 \ldots, p)$, in analogy with the work, in the context of the multivariate normal distribution, by James and Stein (1961).

In Section 4, we consider estimators based on the spectral decomposition of the sample sum of product matrix i.e. estimators of the form $\mathbf{R\Delta MR}'$ where $\mathbf{A}$ has the spectral decomposition $\mathbf{A} = \mathbf{RMR}'$, with

$$\mathbf{M} = diag(m_1, m_2, \ldots, m_p),\ (m_1 \geq m_2 \ldots \geq m_p),$$

once again in analogy with the works of Stein (1975) and Dey and Srinivasan (1985).

Dey and Srinivasan (1985) developed improved as well as minimax estimators for the scale matrix $\mathbf{\Sigma}$ of the multivariate normal distribution along the line of James and Stein (1961) and Stein (1975). Joarder (1992) extendeded some of these results for multivariate t-population. It may be mentioned that Theorem 7.1 and part of Theorem 7.2 of Joarder (1992) are special cases of Theorem 4.4.4 and Theorem 4.4.5 of Fang and Zhang (1990) respectively. In this paper contribution of Joarder (1992) developed in the spirit of Dey and Srinivasan (1985) has been isolated in Thorem 3.2 and Thorem 4.1.

It is well-known that if $\mathbf{X}_j|\tau \sim \mathcal{N}_p(\theta,\ \tau^2\mathbf{\Sigma}), j = 1, 2, \ldots, N$, then

$$\mathbf{A}|\tau = \tau^2\mathbf{W} \sim \mathcal{W}_p(n, \tau^2\mathbf{\Sigma}),\ n = N - 1 \tag{1.3}$$

and that

$$E(\mathbf{A}) = E[E(\mathbf{A})|\tau] = E[n(\tau^2\mathbf{\Sigma})] = n\mathbf{\Sigma}E(\tau^2). \tag{1.4}$$

In order to avoid future digressions of a trivial nature we outline some lemmas that will be needed in the sequel. The proofs are omitted as they easily follow from Joarder and Ali (1997).

Lemma 1.1 Consider the triangular decomposition $\mathbf{A} = \mathbf{TT}'$ where $\mathbf{A}$ is the sum of product matrix based on the scale mixture of the multivariate normal model given by (1.1) and $\mathbf{T}$ is a lower triangular matrix. Then the following identities hold:

$$E[tr(\mathbf{\Sigma}^{-1}\mathbf{T\Delta T}')] = E[tr(\mathbf{\Sigma}^{-1}\mathbf{U\Delta U}')]E(\tau^2), \tag{1.5}$$

$$E[ln(|\mathbf{\Sigma}^{-1}\mathbf{T\Delta T}'|)] = E[ln(|\mathbf{\Sigma}^{-1}\mathbf{U\Delta U}'|)] + 2pE[ln(\tau)], \tag{1.6}$$

provided $E(\tau^2)$ and $E(ln\tau)$ exist, where $\mathbf{U}$ is a lower triangular matrix such that $\mathbf{W} = \mathbf{UU}'$.

Lemma 1.2 Consider the spectral decomposition of $\mathbf{A} = \mathbf{RMR}'$ where $\mathbf{A}$ is the sample sum of product matrix based on the scale mixture of the multivariate normal model given by (1.1). Then the following results hold:

$$E[ln(|\mathbf{\Sigma}^{-1}\mathbf{RD}^*\mathbf{MR}'|)] = \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] + \sum_{i=1}^{p} ln(d_i^*) + 2pE(ln\tau), \tag{1.7}$$

$$E[tr(\mathbf{\Sigma}^{-1}\mathbf{RD}^*\mathbf{MR}' = 2E\left[\sum_{i=1}^{p}\sum_{t=i+1}^{p}\frac{d_il_i - d_tl_t}{l_i - l_t}\right] + (n-p+1)\sum_{i=1}^{p} d_i . \tag{1.8}$$

provided $E(ln\tau)$ exists, where l_i's and m_i's $(i = 1, 2, \ldots, p)$ are the characteristic roots of $\mathbf{W}$ and $\mathbf{A}$ respectively, d_i^*, $(i = 1, 2, \ldots, p)$ and $d_i, (i = 1, 2, \ldots, p)$ are the diagonal elements of the diagonal matrices $\mathbf{D}^*$ and $\mathbf{D}$ respectively where

$$d_i^* = (n + p + 1 - 2i)^{-1}/E(\tau^2) = d_i/E(\tau^2), \quad (i = 1, 2, \ldots, p). \tag{1.9}$$

2. ESTIMATORS BASED ON THE MULTIPLES OF THE SAMPLE SUM OF PRODUCT MATRIX

The covariance matrix $\mathbf{\Sigma}$ of the multivariate normal distribution is usually estimated by $c\mathbf{W}$ where $c > 0$ and $\mathbf{W}$ is the usual Wishart matrix. It is well known (see e.g. Muirhead, 1982, p 129) that under the entropy loss function, the best estimator (smallest risk) of the covariance matrix of the multivariate normal distribution, of the form $c\mathbf{W}$, is given by $\mathbf{W}/n$.

In this section we consider estimators of the form $c\mathbf{A}$, where $c > 0$, for the scaled covariance matrix $\mathbf{\Sigma}$ of the multivariate elliptical model and find optimum value of c for which the risk function of the estimator under the entropy loss function is minimized. The result is proved by Fang and Zhang (1990) for a broader class of elliptical distribution. A direct proof along Joarder and Ali (1997)is presented below.

Theorem 2.1 Consider the scale mixture of the multivariate normal model given by (1.1). Then under the entropy loss function the unbiased estimator $\tilde{\mathbf{\Sigma}} = \mathbf{A}/(nE(\tau^2))$ of $\mathbf{\Sigma}$ has the smallest risk among the class of estimators $c\mathbf{A}$, for $c > 0$ and the corresponding minimum risk is given by

$$R(\tilde{\mathbf{\Sigma}},\ \mathbf{\Sigma}) = p\ ln(n) - \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] + p\ ln\left[E(\tau^2)\right] - 2p\ E(ln\ \tau) \qquad (2.1)$$

provided $E(\tau^2)$ and $E(ln\tau)$ exist.

Proof. The risk function of the estimator $c\mathbf{A}$ is given by

$$\begin{aligned} R(c\mathbf{A},\ \mathbf{\Sigma}) =& E[tr(\mathbf{\Sigma}^{-1}c\mathbf{A}) - ln(|\mathbf{\Sigma}^{-1}c\mathbf{A}|) - p] \\ =& c\ tr[\mathbf{\Sigma}^{-1}E(\mathbf{A})] - p\ ln(c) - E[ln\left(|\mathbf{\Sigma}^{-1}\mathbf{A}|\right)] - p. \end{aligned}$$

Then it follows from (1.3) and $|\mathbf{\Sigma}^{-1}\mathbf{W}| \sim \prod_{i=1}^{p} \chi^2_{n+1-i}$ (see e.g. Muirhead, 1982, pp 85, 100) that

$$\begin{aligned} R(c\mathbf{A},\ \mathbf{\Sigma}) =& c\ tr\left[\mathbf{\Sigma}^{-1}n\mathbf{\Sigma}\ E(\tau^2)\right] - p\ ln(c) \\ & - \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] - 2p\ E(ln\tau) - p \\ =& npc\ E(\tau^2) - \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] - p\ ln(c) - p - 2p\ E(ln\ \tau) \qquad (2.2) \end{aligned}$$

provided $E[log\tau]$ exists.

It is easily verified that $E[ln(\chi_{n+1-i})]$, $i = 1, 2, \ldots, p$ is finite. The risk function given by (2.2) is minimized at $c = [nE(\tau^2)]^{-1}$ and hence the corresponding estimator is given by $\tilde{\mathbf{\Sigma}} = \mathbf{A}/(nE(\tau^2))$. It then follows from (2.2), by putting $c = [nE(\tau^2)]^{-1}$ that the risk function of the estimator $\tilde{\mathbf{\Sigma}}$ is given by (2.1).

3. ESTIMATORS BASED ON A TRIANGULAR DECOMPOSITION OF THE SAMPLE SUM OF PRODUCT MATRIX

Following James and Stein (1961), we propose estimators of the form $\mathbf{T\Delta T}'$ where $\mathbf{T}$ is a lower triangular matrix such that the sample sum of product matrix $\mathbf{A}$ has the decomposition $\mathbf{A} = \mathbf{TT}'$ and $\mathbf{\Delta}$ an arbitrary positive definite diagonal matrix. We find the optimum value of $\mathbf{\Delta}$ for which the risk function of the estimator $\mathbf{T\Delta T}'$ under the entropy loss function is minimized and denote it by $\mathbf{D}^*$. The resulting estimator $\mathbf{\Sigma}^* = \mathbf{TD}^*\mathbf{T}$ dominates the unbiased estimator $\tilde{\mathbf{\Sigma}}$. The dominance behaviour is presented in Theorem 3.2. Note that Theorem 3.1 is a special case of Fang and Zhang (1990).

Theorem 3.1 Under the entropy loss function, the estimator $\mathbf{\Sigma}^* = \mathbf{TD}^*T'$ where $\mathbf{T}$ is a lower triangular matrix such that $\mathbf{A} = \mathbf{TT}'$ and $\mathbf{D}^* = diag(d_1^*, d_2^*, \ldots, d_p^*)$ with d_i^* $(i = 1, 2, \ldots, p)$ defined by (1.9), has the smallest risk among the class of estimators $\mathbf{T\Delta T}'$ where $\mathbf{\Delta}$ belongs to the class of all positive definite diagonal matrices.

Theorem 3.2 Under the entropy loss function, the risk of the estimator $\mathbf{\Sigma}^*$ is given by

$$\begin{aligned} R(\mathbf{\Sigma}^*, \, \mathbf{\Sigma}) = &\sum_{i=1}^{p} ln(n+1+p-2i) - \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] \\ &+p \; ln \left[E(\tau^2)\right] - 2p \; E(ln\tau). \end{aligned} \tag{3.1}$$

Furthermore, $\mathbf{\Sigma}^*$ dominates the unbiased estimator $\tilde{\mathbf{\Sigma}} = \mathbf{A}/(nE(\tau^2))$.

Proof. It follows from Lemma 1.1 that the risk function of the estimator $\mathbf{T\Delta T}'$ is given by

$$R(\mathbf{T\Delta T}',\ \mathbf{\Sigma}) = E\left[tr(\mathbf{\Sigma}^{-1}\mathbf{T\Delta T}')\right] - E\left[ln(|\mathbf{\Sigma}^{-1}\mathbf{T\Delta T}'|)\right] - p$$
$$= E\left[tr(\mathbf{\Sigma}^{-1}\mathbf{U\Delta U}')\right] E(\tau^2) - E\left[ln(|\mathbf{\Sigma}^{-1}\mathbf{U\Delta U}'|)\right]$$
$$-2p\ E(ln\tau) - p \qquad (3.2)$$

Then following Muirhead (1982, pp 130-132), it can be proved that the risk function given by (3.2) does not depend on $\mathbf{\Sigma}$ and that

$$R(\mathbf{T\Delta T}',\ \mathbf{\Sigma}) = \sum_{i=1}^{p} \delta_i(n+1+p-2i)E(\tau^2) - 2p\ E(ln\tau)$$
$$- \left[\sum_{i=1}^{p} ln(\delta_i) + \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})]\right] - p. \qquad (3.3)$$

This attains its minimum value of

$$R(\mathbf{\Sigma}^*,\ \mathbf{\Sigma}) = -2p\ E(ln\tau) - \sum_{i=1}^{p} ln d_i^* - \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] \qquad (3.4)$$

when $\delta_i = d_i^*$, $(i = 1, 2, \ldots, p)$ defined by (1.9). Then by putting the value of d_i^* from (1.9) we get the risk function of Σ^* as given by (3.1). The risk function of the unbiased estimator $\tilde{\Sigma}$ has already been calculated in (2.1) so that

$$R(\mathbf{\Sigma}^*, \mathbf{\Sigma}) - R(\tilde{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \sum_{i=1}^{p} ln\left(\frac{n+1+p-2i}{n}\right).$$

The righthand side of the above equation is negative (See Theorem 3.1 of Joarder and Ali, 1997). Hence it follows that the estimator $\mathbf{\Sigma}^*$ dominates the unbiased estimator $\tilde{\mathbf{\Sigma}}$.

4. ESTIMATORS BASED ON THE SPECTRAL DECOMPOSITION OF THE SAMPLE SUM OF PRODUCT MATRIX

Let the sample sum of product matrix $\mathbf{A}$ have the spectral decomposition $\mathbf{A} = \mathbf{RMR}'$. We consider, following Stein (1975) and Dey and Srinivasan (1985), estimators of the form $\mathbf{R}\phi(\mathbf{M})\mathbf{R}'$ where $\phi(\mathbf{M}) = diag(\phi_1(M), \phi_2(M), \quad \ldots, \phi_p(\mathbf{M}))$, and $\phi_i(M) > 0$ $i = 1, 2, \ldots, p$ is a function of the characteristic roots $m_1, m_2, \ldots m_p$. The main result is presented in the form of the following theorem.

Theorem 4.1 Let $\hat{\mathbf{\Sigma}} = \mathbf{R}\phi(\mathbf{M})\mathbf{R}$ be an estimator for $\mathbf{\Sigma}$ where $\mathbf{A}$ has the spectral decomposition $\mathbf{A} = \mathbf{RMR}'$, with $\phi(\mathbf{M}) = \mathbf{D}^*\mathbf{M}$. Then under the entropy loss function given by (1.2), the estimator $\hat{\mathbf{\Sigma}} = \mathbf{RD}^*\mathbf{MR}'$ dominates the estimator $\mathbf{\Sigma}^* = \mathbf{TD}^*T'$.

Proof. The risk function of the estimator $\hat{\mathbf{\Sigma}}$ is given by

$$\begin{aligned} R(\hat{\mathbf{\Sigma}},\ \mathbf{\Sigma}) =& E\left[tr(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}) - ln(|\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}|) - p\right] \\ =& E[tr(\mathbf{\Sigma}^{-1}\mathbf{R}\phi(\mathbf{M})\mathbf{R}')] - E[ln(|\mathbf{\Sigma}^{-1}\mathbf{R}\phi(\mathbf{M})\mathbf{R}'|)] - p. \end{aligned}$$

Then it follows from Lemma 1.2 that

$$\begin{aligned} R(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) =& 2E\left[\sum_{i=1}^{p}\sum_{t=i+1}^{p} \frac{d_i l_i - d_t l_t}{l_i - l_t}\right] + (n-p+1)\sum_{i=1}^{p} d_i \\ & - \sum_{i=1}^{p} ln(d_i^*) - \sum_{i=1}^{p} E[ln(\chi^2_{n+1-i})] - 2pE[ln(\tau)] - p, \end{aligned} \tag{4.1}$$

where l_i's are the characteristic roots of the Wishart matrix $\mathbf{W}$ and d_i's are given by (1.9). The rest of the proof is immediate from Joarder and Ali (1997).

We note that the smallest risk estimator in the class of estimators $c\mathbf{A}$ is an unbiased estimator given by $\tilde{\mathbf{\Sigma}} = \mathbf{A}/(nE(\tau^2))$. The estimator $\mathbf{\Sigma}^* = \mathbf{TD}^*\mathbf{T}'$ dominates $\tilde{\mathbf{\Sigma}}$. However, the estimator $\mathbf{\Sigma}^*$ is dominated by $\hat{\mathbf{\Sigma}} = \mathbf{RD}^*\mathbf{MR}'$.

It may be remarked that if τ is degenerate at $\tau = 1$, then the scale mixture of multivariate normal model given by (1.1) turns into the joint density of the product of N p-dimensional normal densities and the estimators coincide, as expected, with the usual estimators of the scaled covariance matrix $\mathbf{\Sigma}$ of the multivariate normal model. Finally we remark that it remains open to develop estimation strategy for the scaled covariance matrix for more general class of elliptical distributions considered by Anderson, Fang and Hsu (1986).

ACKNOWLEDGEMENT

The author acknowledges excellent research facilities provided by King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

REFERENCES

Anderson, T.W., Fang, K.T and Hsu, H. (1986). Maximum likelihood estimates and likelihood ratio criteria for multivariate elliptically contoured distributions. *Canad. J. Statist.,* **14**, 55–59.

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

Fang, K.T. and Zhang, Y.T (1990). *Generalized Multivariate Analysis.* Sprringer-Verlag

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.,* **1**, 361–379. University of California Press, Berkeley.

Joarder, A.H. (1992). Estimation of the Scale Matrix of a Multivariate T-model. Unpublished Ph.D. Thesis. The University of Western Ontario, Canada.

Joarder, A.H. and Ali, M.M. (1997). Estimation of the scale matrix of a multivariate t-model under entropy loss function. *Metrika*, 46, 21–32.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory* . Wiley, New York, Chapter 5.

Osiewalski, J. and Steel, M.F.J. (1993). Robust Bayesian inference in elliptical regression models. *J. Econometrics*, **57**, 345–363.

Stein, C. (1975). Reitz Lecture, 39th Annual Meeting IMS. Atlanta, Georgia.