

Estimation of the trace of the scaled covariance matrix of a multivariate t -model using a known information[†]

A. H. Joarder¹, Sarjinder Singh²

¹Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia 31261 (e-mail: anwarj@kfupm.edu.sa)

²Department of Mathematics and Statistics, University of Southern Maine, Portland, Maine ME04104, USA (e-mail: sarjinder@yahoo.com)

Abstract. The trace of the scaled covariance matrix of the multivariate t -distribution is considered for estimation using a power transformation. The proposed estimator always dominates the usual maximum likelihood estimator in the sense of having smaller risk under a quadratic loss function. The dominance behaviour is proved analytically as well as computationally by using Monte-Carlo simulation.

Key words: Scale matrix; estimation of trace of scaled covariance matrix; multivariate t -distribution, Monte-Carlo simulation

1. Introduction

In this paper we consider the estimation of the trace of the scaled covariance matrix of the multivariate t -distribution. The trace of the scaled covariance matrix representing the total variation in the component variables is important in many statistical analyses e.g. in principal component analysis.

We assume Np -dimensional ($p \geq 2$) random vectors (not necessarily independent) $X_1, X_2, \dots, X_N$ have a joint p.d.f. (probability density function) given by

$$f(x_1, x_2, \dots, x_N) = \frac{|\Sigma|^{-N/2}}{C(v, Np)(2\pi)^{Np/2}} \left(1 + \sum_{j=1}^N (x_j - \mu)' (v\Sigma)^{-1} (x_j - \mu) \right)^{-(v+Np)/2} \quad (1.1)$$

[†] This paper is based on a Technical Report No. 234, Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

where the normalizing constant $C(v, Np)$ is given by

$$C(v, Np) = \frac{(v/2)^{Np/2} \Gamma(v/2)}{\Gamma((v + Np)/2)} \quad (1.2)$$

where $x_j = (x_{1j}, x_{2j}, \dots, x_{pj})'$, μ is an unknown $p \times 1$ vector of location parameters, Σ is a $p \times p$ unknown positive definite matrix of scale parameters and v (> 4) is assumed to be a known positive constant. Each p -dimensional random vector X_j ($j = 1, 2, \dots, N$) has a multivariate t -distribution with mean vector μ and covariance matrix $\Sigma/(1 - 2/v)$.

The joint p.d.f. in (1.1) represents the multivariate t -model; it has been considered, among others, by Zellner (1976) in the context of stock market problems. Lange, Little and Taylor (1989) applied the multivariate t -model to a variety of situations. For a discussion of the justification of the multivariate t -model the reader is referred to Kelejian and Prucha (1985) and Joarder and Ahmed (1998).

Olkin and Selliah (1977) considered the estimation of the trace of the covariance matrix of the multivariate normal distribution under a weighted squared error loss function. Joarder (1995) and Joarder and Beg (1999) considers estimation of the trace of the scale matrix of the multivariate t -distribution under a squared error loss function following Dey (1988). Joarder and Singh (1997) considered the estimation of the trace of the scale matrix of a multivariate t -model using a regression type of estimator. In this paper we consider the estimation of the trace of the scale matrix of the multivariate t -model using a power transformation. It may be remarked that we are estimating the trace of the scale matrix Σ (which is hereinafter referred to as scaled covariance matrix) instead of the trace of the covariance matrix since the scale matrix Σ determines the covariance matrix upto a known constant $v/(v - 2)$.

The estimation of the trace of the covariance matrix of the multivariate normal distribution by the maximum likelihood method is meaningful because the large sample properties of maximum likelihood estimators follow from independently and identically distributed observations. This is not necessarily true for the model considered in (1.1) unless $v \rightarrow \infty$. Hence we use MSE (Mean Square Error) Criterion to estimate the trace of scaled covariance matrix.

Let $\delta = \text{tr}(\Sigma)$ be the trace of the scaled covariance matrix Σ to be estimated. The usual estimator of δ is given by $\hat{\delta} = c_0 \text{tr}(A)$ where

$$A = \sum_{j=1}^N (X_j - \bar{X})(X_j - \bar{X})'$$

with $\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)'$ and $\bar{X}_i = \sum_{j=1}^N X_{ij}/N$, ($i = 1, 2, \dots, p$), is the sample sum of product matrix (Wishart matrix) based on the multivariate t -model. In this paper, we propose an improved estimator of $\delta = \text{tr} \Sigma$ based on the consideration of reducing the risk by incorporating a well known inequality of the Wishart matrix.

2. Proposed estimation strategy

We propose a new estimator that incorporates the following well known inequality between arithmetic mean and geometric mean

$$\frac{\text{tr } A/p}{|A|^{1/p}} \geq 1 \quad (2.1)$$

as follows:

$$\hat{\delta} = c_0 \text{tr } A - c_0 c (\text{tr } A - p|A|^{1/p}). \quad (2.2)$$

where c is such that the estimator is nonnegative. The proposed estimator aims at reducing the Mean Square Error (MSE) by incorporating the information in (2.1).

In order to appraise the efficiency of the proposed estimator relative to the usual estimator given by

$$\tilde{\delta} = c_0 \text{tr } A \quad (2.3)$$

we use the MSE criterion. The estimators are based on the sample sum of product matrix A having probability density function

$$g(A) = \frac{|\Sigma|^{-n/2} |A|^{(n-p-1)/2}}{C(v, np) 2^{np/2} \Gamma_p(n/2)} (1 + \text{tr } \Sigma^{-1} A/v)^{-(v+np)/2} \quad (2.4)$$

where $A > 0$, $n = N - 1 \geq p$, and $\Gamma_p(\alpha)$ is the generalized gamma function defined by

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma((2\alpha - i + 1)/2)$$

(see e.g. Joarder, 1998).

3. The main results

The main results are presented in this section in the form of two theorems.

Theorem 3.1. *The MSE of the usual estimator $\tilde{\delta}$ is given by*

$$\begin{aligned} \text{MSE}(\tilde{\delta}) &= \left[\frac{nc_0}{1 - 2/v} \left(\frac{nc_0}{1 - 4/v} - 2 \right) + 1 \right] (\text{tr } \Sigma)^2 \\ &\quad + \frac{2nc_0^2}{(1 - 2/v)(1 - 4/v)} \text{tr } \Sigma^2. \end{aligned}$$

Proof. We have

$$\begin{aligned} \text{MSE}(\tilde{\delta}) &= E(\tilde{\delta} - \delta)^2 \\ &= c_0^2 E(\text{tr } A)^2 + \delta^2 - 2\delta c_0 E(\text{tr } A). \end{aligned}$$

The expected value of the Wishart matrix is given by $E(A) = n\Sigma/(1 - 2/v)$. Then the proof is immediate from Joarder (1998).

Theorem 3.2. *The MSE of the proposed estimator $\hat{\delta}$ is given by*

$$\text{MSE}(\hat{\delta}) = \text{MSE}(\tilde{\delta}) - c\beta_1 + c^2\beta_2 \quad (3.1)$$

where

$$\beta_1 = 2E[(c_0 \text{tr } A - \delta)(p|A|^{1/p} - \text{tr } A)] \quad (3.2)$$

$$\text{and } \beta_2 = c_0^2 E[p|A|^{1/p} - \text{tr } A]^2. \quad (3.3)$$

Proof: The MSE of the proposed estimator is given by

$$\begin{aligned} \text{MSE}(\hat{\delta}) &= E[c_0 \text{tr } A - cc_0(\text{tr } A - p|A|^{1/p}) - \delta]^2 \\ &= \text{MSE}(\tilde{\delta}) - c\beta_1 + c^2\beta_2 \end{aligned}$$

where β_1 and β_2 are defined in the theorem.

Using the results in Joarder (1998), it follows that

$$\begin{aligned} \beta_1 &= \frac{2}{1-2/v} \left[2 \left(\frac{2c_0(n/2 + 1/p)}{1-4/v} - p \right) \frac{\Gamma_p(n/2 + 1/p)}{\Gamma_p(n/2)} |\Sigma|^{1/p} \text{tr } \Sigma \right. \\ &\quad \left. - n \left(\frac{nc_0}{1-4/v} - 1 \right) (\text{tr } \Sigma)^2 - \frac{4nc_0}{1-4/v} \text{tr } \Sigma^2 \right] \end{aligned} \quad (3.4)$$

and

$$\begin{aligned} \beta_2 &= \frac{c_0^2}{(1-2/v)(1-4/v)} \left[\frac{4p|\Sigma|^{1/p}}{\Gamma_p(n/2)} \{p\Gamma_p(n/2 + 2/p)|\Sigma|^{1/p} \right. \\ &\quad \left. - 2\Gamma_p(n/2 + 1/p) \text{tr } \Sigma\} + n\{n(\text{tr } \Sigma)^2 + 2 \text{tr } \Sigma^2\} \right]. \end{aligned} \quad (3.5)$$

The $\text{MSE}(\hat{\delta})$ minimizes at $c = \beta_1/(2\beta_2)$ and the minimum value is given by

$$\text{Min. MSE}(\hat{\delta}) = \text{MSE}(\tilde{\delta}) - \beta_1^2/(4\beta_2). \quad (3.6)$$

This proves that the proposed estimator is always better than the usual estimator in the sense of having smaller Mean Square Error. However the proposed estimator is not operational as it involves the unknown parameter c . The operational value of c is given by

$$\hat{c} = \frac{\hat{\beta}_1}{2\hat{\beta}_2} \quad (3.7)$$

which can be obtained by estimating β_1 and β_2 given by (3.4) and (3.5) respectively. The estimation of β_1 and β_2 can be done by replacing Σ by the

the usual estimator $c_0 A$. If $c_0 = 1/(n + 1)$, the estimators would be maximum likelihood estimators. Consequently the proposed estimator estimator of $\delta = \text{tr}(\Sigma)$ is

$$\hat{\delta} = c_0 \text{tr } A - c_0 \hat{c}(\text{tr } A - p|A|^{1/p}). \quad (3.8)$$

with $\hat{c}$ given by (3.7) where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimators of β_1 and β_2 respectively.

3. Empirical study

A Monte Carlo study has been carried out by generating 100 Wishart matrices from different Wishart distributions based on the multivariate t -model. We estimated the value of c for each matrix and observed that $\hat{c}$ lies between 0 and 1. Thus the proposed estimator

$$\hat{\delta} = c_0 \text{tr } A - c_0 c(\text{tr } A - p|A|^{1/p})$$

is always nonnegative. The Percent Relative Efficiency of the proposed estimator defined by

$$\frac{\text{MSE}(\hat{\delta}) - \text{MSE}(\tilde{\delta})}{\text{MSE}(\tilde{\delta})} = \frac{-c\beta_1 + c^2\beta_2}{\text{MSE}(\tilde{\delta})}$$

is presented in Table 1 for $n = 25$ and $p = 3$. The following three diagonal matrices

$$\Sigma = \text{Diag}(1, 1, 1), \quad \Sigma = \text{Diag}(4, 2, 1), \quad \Sigma = \text{Diag}(25, 1, 1)$$

considered in Table 1 are due to Dey (1988).

Table 1. Percent Relative Efficiency of the proposed estimator over the usual estimator

ν	Diag(1, 1, 1)	Diag(4, 2, 1)	Diag(25, 1, 1)
05	105.32	130.31	153.90
10	102.13	117.56	148.76
15	101.53	112.07	127.15

A general scale matrix of order 3×3 given by $\Sigma = ((\sigma_{ik}))$ where $\sigma_{11} = 25$, $\sigma_{12} = -2$, $\sigma_{13} = 4$, $\sigma_{22} = 4$, $\sigma_{23} = 1$, $\sigma_{33} = 9$ is also considered. In this case, the Percent Relative Efficiency of the proposed estimator over the usual one is given by 170.44, 165.57 and 158.11 for $\nu = 5, 10$ and 15 respectively.

Thus we conclude that the proposed estimator $\hat{\delta}^*$ in (3.8) remains always more efficient than the usual estimator $\tilde{\delta}$ given by (2.3) analytically as well as empirically.

Acknowledgement. The first author gratefully acknowledges the excellent research facilities available at King Fahd University of Petroleum and Minerals, Saudi Arabia.

References

- Dey DK (1988) Simultaneous estimation of eigenvalues. *Ann Inst Statist Math* 40:137–147
- Joarder AH (1995) Estimation of the trace of the scale matrix of a multivariate t -model. *Proceeding of the Econometrics Conference at Monash University, Monash University, Australia* 467–474
- Joarder AH (1998) Some useful Wishart expectations based on the multivariate t -model. *Statistical Papers* 39:223–229
- Joarder AH, Ahmed SE (1998) Estimation of the scale matrix of a class of elliptical distributions. *Metrika* 48:149–160
- Joarder AH, Beg GK (1999) Estimation of the trace of the scale matrix of a multivariate t -model. *Statistica* 59:181–191
- Joarder AH, Singh Sarjinder (1996) Estimation of the trace of the scale matrix of a multivariate t -model using regression estimator. *Statistics* 28:161–168
- Kelejian HH, Prucha IR (1985) Independent or uncorrelated disturbances in linear regression: An illustration of difference. *Economic Letters* 19:35–38
- Olkin I, Selliah JB (1977) Estimating covariance in a multivariate normal distribution. *Statistical Decision Theory and Related Topics II*, (eds. Gupta SS, Moore D), 313–326, Academic Press, New York
- Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modelling using the t distribution. *J Amer Statist Assoc* 84:881–896
- Zellner A (1976) Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error term. *J Amer Statist Assoc* 71:400–405