Ann. Inst. Statist. Math.
Vol. 43, No. 4, 761–771 (1991)

ESTIMATION OF A COMMON MULTIVARIATE NORMAL MEAN VECTOR

K. KRISHNAMOORTHY

Department of Statistics, Temple University, Philadelphia, PA 19122, U.S.A.

(Received December 4, 1989; revised June 6, 1990)

Abstract. Let $X_1, \ldots, X_N$ be independent observations from $N_p(\mu, \Sigma_1)$ and $Y_1, \ldots, Y_N$ be independent observations from $N_p(\mu, \Sigma_2)$. Assume that X_i's and Y_i's are independent. An unbiased estimator of μ which dominates the sample mean $\bar{X}$ for $p \geq 1$ under the loss function $L(\mu, \hat{\mu}) = (\hat{\mu} - \mu)'\Sigma_1^{-1}(\hat{\mu} - \mu)$ is suggested. The exact risk (under L) of the new estimator is also evaluated.

Key words and phrases: Common mean vector, unbiased estimator, Wishart and noncentral Wishart distributions.

1. Introduction

Let $X_1, \ldots, X_N$ be independent observations from a p-variate normal distribution with mean vector μ and positive definite covariance matrix $\Sigma_1, N_p(\mu, \Sigma_1)$. Let $Y_1, \ldots, Y_N$ be independent observations from $N_p(\mu, \Sigma_2)$. Assuming that X_i's and Y_i's are independent, we consider the problem of estimating the common mean vector μ when Σ_1 and Σ_2 are unknown.

This problem, when $p = 1$, has been considered by many authors, among them, Graybill and Deal (1959), Brown and Cohen (1974), Cohen and Sackrowitz (1974) and Khatri and Shah (1974). These authors suggest unbiased estimators of μ with smaller variances than $\text{var}(\bar{X})$, sometimes smaller than both $\text{var}(\bar{X})$ and $\text{var}(\bar{Y})$ with certain restrictions on the sample size N, where $\bar{X} = (1/N)\sum_{i=1}^N X_i$ and $\bar{Y} = (1/N)\sum_{i=1}^N Y_i$.

For $p > 1$ case, Chiou and Cohen (1985) considered this problem with respect to the covariance criterion. Under this criterion, between the unbiased estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ of μ, $\hat{\mu}_1$ is preferable to $\hat{\mu}_2$ if $\text{cov}(\hat{\mu}_2) - \text{cov}(\hat{\mu}_1)$ is a positive semidefinite matrix, where $\text{cov}(\cdot)$ denotes the covariance matrix. They showed that none of the estimators, that are multivariate analogous to the ones considered in the papers cited in the previous paragraph, dominates either $\bar{X}$ or $\bar{Y}$ under the covariance criterion. Loh (1988) suggested some combined unbiased estimators for μ under the loss function $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\mu - \hat{\mu})$. None of these estimators is known to dominate either $\bar{X}$ or $\bar{Y}$ analytically. Recently, Kubokawa (1989) treated this problem under the loss $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})'P(\mu - \hat{\mu})$, where P is a known positive definite matrix, and proposed a combined estimator that dominates both $\bar{X}$ and $\bar{Y}$.

In this paper, we evaluate the merit of an estimator through the loss function

$$L(\mu, \hat{\mu}) = (\mu - \hat{\mu})'\Sigma_1^{-1}(\mu - \hat{\mu}). \tag{1.1}$$

This loss function is chosen for mathematical convenience. Under this loss (1.1), there are several estimators (for example, James and Stein's estimator (1961)), based on only one sample observations $X_1, \ldots, X_N$, that dominate $\bar{X}$. However, note that such estimators are not unbiased and also they necessitate (to dominate $\bar{X}$) that $p \geq 3$. So we hope that if we use both sample observations X_i's and Y_i's, we can find an unbiased estimator that beats $\bar{X}$ under the loss (1.1) for any $p \geq 1$.

In Section 2 of this paper, we suggest an unbiased estimator that is similar to the one given in Krishnamoorthy and Rohatgi (1988) for $p = 1$ case. We compute the exact risk of the new estimator and show that it is less than the risk of $\bar{X}$, for any $p \geq 1$ and $N \geq p + 5$, under the loss function (1.1).

We discuss the features of the new estimator and a relation between the loss (1.1) and the covariance criterion in Section 3. Numerical study (Table 1) indicates that the relative improvement of the new estimator over $\bar{X}$ is quite significant when $\mathrm{tr}(\Sigma_1(\Sigma_1 + \Sigma_2)^{-1})$ is moderately large.

In the Appendix, we derive expectations of some mixtures of Wishart and inverted Wishart random matrices (using some "matrix derivatives" results) that are needed in Section 2. These expectations are in a more general form and, if one is interested, moments of mixtures of elements of Wishart and inverted Wishart random matrices can be obtained from them.

2. Main result

Consider the transformation $U_i = X_i$ and $V_i = X_i - Y_i$, $i = 1, 2, \ldots, N$. Define

$$\bar{U} = (1/N)\sum_{i=1}^{N} U_i \quad \text{and} \quad S_U = \sum_{i=1}^{N}(U_i - \bar{U})(U_i - \bar{U})'.$$

Let $\bar{V}$ and S_V be defined similarly.

When Σ_1 and Σ_2 are known, it can be easily seen that $\hat{\mu} = \bar{U} - \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}\bar{V}$ is the best unbiased estimator and hence better than both $\bar{X}$ and $\bar{Y}$ under the loss (1.1). When Σ_1 and Σ_2 are unknown, replacing $\Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$ by $aS_U S_V^{-1}$, where a is a positive constant, leads to the estimator

$$\hat{\mu}_a = \bar{U} - aS_U S_V^{-1}\bar{V}. \tag{2.1}$$

The rational for considering $\hat{\mu}_a$ is that $E(S_U) = (N-1)\Sigma_1$ and $E(S_V^{-1}) = (N-p-2)^{-1}(\Sigma_1 + \Sigma_2)^{-1}$, and a is chosen to minimize the risk. Since $E(\bar{U}) = \mu, E(\bar{V}) = 0$ and (S_U, S_V) is independent of $(\bar{U}, \bar{V}), \hat{\mu}_a$ is an unbiased estimator of μ.

We need the following lemma to derive the risk of $\hat{\mu}_a$ under the loss (1.1).

LEMMA 2.1. (i) *The conditional distribution of* S_U *given* S_V *is noncentral Wishart with* $n \equiv N-1$ *degrees of freedom, covariance matrix* $\Sigma_{11.2} = \Sigma_1 - \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}\Sigma_1$ *and noncentrality parameter* $\Sigma_{11.2}^{-1} A S_V A'$, *where* $A = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$. *In the standard notation* $S_U \mid S_V \sim W_p(n, \Sigma_{11.2}, \Sigma_{11.2}^{-1} A S_V A')$.

(ii) $E(S_U \mid S_V) = n\Sigma_{11.2} + A S_V A'$

(iii)
$$\begin{aligned}
&E(S_U C S_U \mid S_V)\\
&\quad = n(n+1)(\Sigma_{11.2} C \Sigma_{11.2})\\
&\qquad + (n+1)(\Sigma_{11.2} C A S_V A' + A S_V A' C \Sigma_{11.2})\\
&\qquad + \operatorname{tr}(C\Sigma_{11.2})(n\Sigma_{11.2} + A S_V A') + (\operatorname{tr} C A S_V A')\Sigma_{11.2}\\
&\qquad + A S_V A' C A S_V A'
\end{aligned}$$

where C *is a matrix of constants.*

PROOF. (i) Let $\boldsymbol{U} = (U_1, \ldots, U_N)$ be a $p \times N$ matrix. Noting that,

$$\boldsymbol{U} \mid (V_1, \ldots, V_N) \sim N(M_{p\times N}, I_N \otimes \Sigma_{11.2})$$

where $M = (\mu + AV_1, \ldots, \mu + AV_N)$ and $A = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$, we prove (i).

(ii) follows from (i) (for example, see Muirhead ((1982), p. 442)).

(iii) Let $W \sim W_p(n, \Theta, \Theta^{-1}\eta\eta')$. Note that $W \overset{d}{=} ZZ'$, where $\boldsymbol{Z}$ is a $p \times n$ matrix whose columns Z_i's are independently distributed as $N_p(\eta_i, \Theta)$ and $(\eta_1, \ldots, \eta_p) = \eta$ is a $p \times n$ matrix. So,

$$\begin{aligned}
E(WCW) &= E(ZZ'CZZ')\\
&= E\Big(\sum_{i=1}^{n} Z_i Z_i' C Z_i Z_i'\Big) + E\Big(\sum_{i\neq j} Z_i Z_i' C Z_j Z_j'\Big).
\end{aligned} \tag{2.2}$$

Now

$$E\Big(\sum_{i\neq j} Z_i Z_i' C Z_j Z_j'\Big) = \sum_{i\neq j} (\Theta + \eta_i\eta_i') C (\Theta + \eta_j\eta_j'). \tag{2.3}$$

Next, using the relation $Z_i \overset{d}{=} \xi_i + \eta_i$, where ξ_i's are i.i.d. as $N_p(0, \Theta)$, we get

$$\begin{aligned}
&E\Big(\sum_{i=1}^{n} Z_i Z_i' C Z_i Z_i'\Big)\\
&\quad = E\Big[\sum_{i=1}^{n} (\xi_i + \eta_i)(\xi_i + \eta_i)' C (\xi_i + \eta_i)(\xi_i + \eta_i)'\Big]\\
&\quad = 2n(\Theta C \Theta) + \operatorname{tr}(C\Theta)(n\Theta + \eta\eta') + 2\Theta C \eta\eta'\\
&\qquad + 2\eta\eta' C \Theta + (\operatorname{tr} C\eta\eta')\Theta + \sum_{i=1}^{n} \eta_i\eta_i' C \eta_i \eta_i'.
\end{aligned} \tag{2.4}$$

Combining (2.2), (2.3) and (2.4), replacing Θ by $\Sigma_{11.2}$ and $\eta\eta'$ by AS_VA', and after some simplification we get (iii).

We now derive the risk of $\hat{\mu}_a$.

THEOREM 2.1. (i) *The risk of* $\hat{\mu}_a = \bar{U} - aS_US_V^{-1}\bar{V}$ *under the loss* (1.1) *is given by*

$$
\begin{aligned}
(2.5)\quad & R(\mu, \hat{\mu}_a) \\
&= R(\mu, \bar{X}) \\
&\quad + (ac_1/N)\{[an(n-1)(n+p+1) - 2n(n-p)(n-p-3)]\operatorname{tr} D \\
&\qquad + [2(n-p)(n-p-3) \\
&\qquad\quad + a(2n+p+2)(p-2n) - 2a(2n-p-2)](\operatorname{tr} D^2) \\
&\qquad + 4a(3n-2p-1)(\operatorname{tr} D^3) + a(9n-6p-1)(\operatorname{tr} D)(\operatorname{tr} D^2) \\
&\qquad + [2(n-p)(n-p-3) - a(4n^2-(p+2)^2+2)](\operatorname{tr} D)^2 \\
&\qquad + a(3n-2p-3)(\operatorname{tr} D)^3\}
\end{aligned}
$$

where $c_1 = [(n-p)(n-p-1)(n-p-3)]^{-1}$ *and* $D = (\Sigma_1+\Sigma_2)^{-1/2}\Sigma_1(\Sigma_1+\Sigma_2)^{-1/2}$.

(ii) *Moreover, for* $n \geq (p+4)$ *and* $a_o = (n-p)(n-p-3)/((n-1)(n+p+1))$

$$R(\mu, \hat{\mu}_{a_o}) < R(\mu, \bar{X})$$

for all positive definite matrices Σ_1 *and* Σ_2.

PROOF. (i)

$$
\begin{aligned}
(2.6)\qquad R(\mu, \hat{\mu}_a) &= E(\hat{\mu}_a - \mu)'\Sigma_1^{-1}(\hat{\mu}_a - \mu) \\
&= R(\mu, \bar{X}) - 2aN^{-1}E\operatorname{tr}(S_US_V^{-1}) \\
&\quad + a^2N^{-1}E\operatorname{tr}(S_V^{-1}S_U\Sigma_1^{-1}S_US_V^{-1}(\Sigma_1+\Sigma_2)).
\end{aligned}
$$

To get (2.6) we used the conditional expectation $E(\bar{U} - \mu \mid \bar{V}) = \Sigma_1(\Sigma_1+\Sigma_2)^{-1}\bar{V}$ and $NE(\bar{V}\bar{V}') = (\Sigma_1 + \Sigma_2)$. Using Lemma 2.1(ii), we first compute

$$
\begin{aligned}
(2.7)\quad & E\operatorname{tr}(S_US_V^{-1}) \\
&= E[\operatorname{tr} E(S_US_V^{-1} \mid S_V)] \\
&= n(n-p-1)^{-1}\operatorname{tr}(\Sigma_{11.2}(\Sigma_1+\Sigma_2)^{-1}) + E\operatorname{tr}(AS_VA'S_V^{-1}).
\end{aligned}
$$

Let $Q = (\Sigma_1+\Sigma_2)^{-1/2}S_V(\Sigma_1+\Sigma_2)^{-1/2} \sim W_p(n, I)$. Then $E\operatorname{tr}(AS_VA'S_V^{-1}) = E\operatorname{tr}(DQDQ^{-1})$ and using Corollary A.1(i), it follows from (2.7) that

$$(2.8)\qquad E\operatorname{tr}(S_US_V^{-1}) = (n-p-1)^{-1}[n\operatorname{tr} D - \operatorname{tr} D^2 - (\operatorname{tr} D)^2].$$

Next, from Lemma 2.1(iii), we have

$$(2.9)\qquad E\operatorname{tr}(S_V^{-1}S_U\Sigma_1^{-1}S_US_V^{-1}(\Sigma_1+\Sigma_2))$$

$$
\begin{aligned}
&= E[E\,\mathrm{tr}(S_V^{-1}S_U\Sigma_1^{-1}S_US_V^{-1}(\Sigma_1+\Sigma_2)\mid S_V)]\\
&= n(n+1)E\,\mathrm{tr}[S_V^{-1}\Sigma_{11.2}\Sigma_1^{-1}\Sigma_{11.2}S_V^{-1}(\Sigma_1+\Sigma_2)]\\
&\quad+(n+1)E\,\mathrm{tr}[S_V^{-1}\Sigma_{11.2}\Sigma_1^{-1}AS_VA'S_V^{-1}(\Sigma_1+\Sigma_2)]\\
&\quad+(n+1)E\,\mathrm{tr}[S_V^{-1}AS_VA'\Sigma_1^{-1}\Sigma_{11.2}S_V^{-1}(\Sigma_1+\Sigma_2)]\\
&\quad+\mathrm{tr}(\Sigma_1^{-1}\Sigma_{11.2})E\,\mathrm{tr}[S_V^{-1}AS_VA'S_V^{-1}(\Sigma_1+\Sigma_2)]\\
&\quad+n\,\mathrm{tr}(\Sigma_1^{-1}\Sigma_{11.2})E\,\mathrm{tr}[S_V^{-1}\Sigma_{11.2}S_V^{-1}(\Sigma_1+\Sigma_2)]\\
&\quad+E\,\mathrm{tr}(\Sigma_1^{-1}AS_VA')\,\mathrm{tr}[S_V^{-1}\Sigma_{11.2}S_V^{-1}(\Sigma_1+\Sigma_2)]\\
&\quad+E\,\mathrm{tr}[S_V^{-1}AS_VA'\Sigma_1^{-1}AS_VA'S_V^{-1}(\Sigma_1+\Sigma_2)].
\end{aligned}
$$

Using the relations that, for any matrices F and G, $\mathrm{tr}(FG)=\mathrm{tr}(GF)$ and for symmetric matrices F, G and H of same order $\mathrm{tr}(FGH)=\mathrm{tr}(GFH)$, (2.9) can be written as

$$
\begin{aligned}
(2.10)\qquad &E\,\mathrm{tr}(S_V^{-1}S_U\Sigma_1^{-1}S_US_V^{-1}(\Sigma_1+\Sigma_2))\\
&= n(n+1)E\,\mathrm{tr}[Q^{-1}(D-2D^2+D^3)Q^{-1}]\\
&\quad+2(n+1)E\,\mathrm{tr}[Q^{-2}(D-D^2)QD]\\
&\quad+(p-\mathrm{tr}\,D)E\,\mathrm{tr}(Q^{-2}DQD)\\
&\quad+n(p-\mathrm{tr}\,D)E\,\mathrm{tr}[Q^{-1}(D-D^2)Q^{-1}]\\
&\quad+E[\mathrm{tr}(DQ)\,\mathrm{tr}(Q^{-1}(D-D^2)Q^{-1})]\\
&\quad+E\,\mathrm{tr}[Q^{-2}DQDQD].
\end{aligned}
$$

All these expectations on the rhs of (2.10) are evaluated and given in the Appendix (Corollary A.1(ii), (iii), Theorem A.2(i), (ii) and the equation (A.3)). After substituting these expectations in (2.10) and then combining the resulting equation with (2.8) and (2.6), we get (2.5).

(ii) When $a=a_o$, (2.5) can be simplified as

$$
\begin{aligned}
(2.11)\quad &R(\mu,\bar X)-R(\mu,\hat\mu_{a_o})\\
&= a_o^2c_1N^{-1}[n(n-1)(n+p+1)\,\mathrm{tr}\,D\\
&\qquad+(2n^2+8n-2np-p^2-2p-2)(\mathrm{tr}\,D^2)\\
&\qquad-4(3n-2p-1)(\mathrm{tr}\,D^3)-(9n-6p-1)(\mathrm{tr}\,D)(\mathrm{tr}\,D^2)\\
&\qquad+(2n^2-2np-p^2-2p)(\mathrm{tr}\,D)^2\\
&\qquad-(3n-2p-3)(\mathrm{tr}\,D)^3].
\end{aligned}
$$

Since $D=(\Sigma_1+\Sigma_2)^{-1/2}\Sigma_1(\Sigma_1+\Sigma_2)^{-1/2}$ is a positive definite matrix with all eigen-values greater than zero and less than or equal to unity, $p\ge\mathrm{tr}\,D\ge\mathrm{tr}\,D^2\ge\mathrm{tr}\,D^3>0$. Applying these inequalities along with the relation $2n^2>2np$ to (2.11), it can be checked that $R(\mu,\bar X)-R(\mu,\hat\mu_{a_o})>0$ if $[n(n-1)(n+p+1)-(p^2+2p+2)-(4n-8p-4)-p(9n-6p-1)-(p^2+2p)p-(3n-2p-3)p^2]\,\mathrm{tr}\,D>0$ which is equivalent to

$$(2.12)\qquad n^3+n^2p-n(3p^2+10p+5)+(p^3+6p^2+7p+2)>0.$$

It can be easily shown that the lhs of (2.12) is positive when $n = p+4$ and a strictly increasing function of n for $n \geq p+4$ and so the inequality (2.12) is true for all $n \geq p+4$. Thus we prove (ii).

Remark 2.1. Interchanging the roles of X and Y, it can be easily seen that $\bar{Y}$ is inadmissible under the loss $L(\mu, \hat{\mu}) = (\hat{\mu} - \mu)'\Sigma_2^{-1}(\hat{\mu} - \mu)$ for any $p \geq 1$ and $N \geq p+5$.

Remark 2.2. Noting that $R(\mu, \bar{X}) = \operatorname{tr} D/(n+1)$, it follows from (2.11) that $\lim_{n\to\infty}(R(\mu, \bar{X}) - R(\mu, \hat{\mu}_{a_o}))/R(\mu, \bar{X}) = \operatorname{tr} D/p$. Therefore, the relative improvement of $\hat{\mu}_{a_o}$ over $\bar{X}$ is significant for large values of n and moderately large values of $\operatorname{tr} D$. This is also evident from Table 1.

Remark 2.3. It is to be noted that the estimator $\hat{\mu}_a$ in (2.1) changes under the permutations of the observations X_i's and Y_i's because the cross-product matrix S_V in $\hat{\mu}_a$ involves $\sum_{i=1}^n (X_i Y_i' + Y_i X_i')$. Thus, one can obtain $n!$ distinct estimators by permuting X_i's and Y_i's. These $n!$ estimators can also be obtained by just permuting Y_i's keeping X_i's fixed. We also note each of such estimators is unbiased and has the same risk function as that of $\hat{\mu}_a$. Therefore, the estimator $\hat{\mu}_a$ can be improved as follows: Let $\mathcal{P}$ denote the set of all permutations on the integers $1, 2, \ldots, n$ and $\kappa = \{i_1, \ldots, i_n\}$ be an element in $\mathcal{P}$. Also let $S_{V(\kappa)} = \sum_{j=1}^n (X_j - Y_{i_j} - \bar{V})(X_j - Y_{i_j} - \bar{V})'$ and $\hat{\mu}_{a(\kappa)} = \bar{U} - aS_U S_{V(\kappa)}^{-1}\bar{V}$. Then,

$$\hat{\mu}_a^* = \sum_{\kappa\in\mathcal{P}} \hat{\mu}_{a(\kappa)}/n! \tag{2.13}$$

is invariant under the permutations of the observations and an unbiased estimator of μ. Also as $L(\mu, \hat{\mu})$ in (1.1) is a strictly convex function of $\hat{\mu}$, Jensen's inequality implies that

$$L(\mu, \hat{\mu}_a^*) < \sum_{\kappa\in\mathcal{P}} L(\mu, \hat{\mu}_{a(\kappa)})/n!. \tag{2.14}$$

Thus, taking expectation on both sides of (2.14) and using the fact that $R(\mu, \hat{\mu}_{a(\kappa)})$ is the same for all $\kappa \in \mathcal{P}$, we prove $\hat{\mu}_a^*$ dominates $\hat{\mu}_{a(\kappa)}$ for each $\kappa \in \mathcal{P}$. In particular, it dominates $\hat{\mu}_a$ in (2.1).

3. Concluding remarks

The estimator $\hat{\mu}_a^*$ in (2.13) is itself inadmissible because it is not a function of minimal sufficient statistics. However, the numerical comparison in Table 1 shows that the percentage relative improvement of $\hat{\mu}_{a_0}$ over $\bar{X}$ is quite significant for moderately large values of $\operatorname{tr} D$ where $D = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$. Therefore, the use of $\hat{\mu}_{a_o}$ or $\hat{\mu}_{a_o}^*$ over $\bar{X}$ under the loss function (1.1) is certainly an advantage.

We next like to point out a relation between the covariance criterion and the loss (1.1). Consider the loss function $L_T(\mu, \hat{\mu}) = (\hat{\mu} - \mu)'T(\hat{\mu} - \mu)$ where T is an arbitrary positive definite matrix. It is not too difficult to show that an estimator

$\hat{a}$ dominates $\hat{b}$ for covariance ordering if and only if it dominates $\hat{b}$ under L_T for all positive definite matrices T. Since the loss (1.1) is a particular case of $L_T(\mu, \hat{\mu})$, it is a slightly weaker criterion than the covariance criterion. So we do not know whether or not $\hat{\mu}_{a_o}$ dominates $\bar{X}$ under the covariance criterion.

In Table 1, the numbers with "%" are the values of $100[R(\mu, \bar{X}) - R(\mu, \hat{\mu}_{a_o})]/R(\mu, \bar{X})$.

Table 1. Percentage relative improvement of $\hat{\mu}_{a_o}$ over $\bar{X}$.

$p = 3$

DN	10	20	30
(.99, .99, .99)	32%	74%	87%
(.9, .8, .7)	27	59	69
(.7, .6, .5)	21	43	50
(.1, .4, .9)	16	33	38
(.5, .5, .5)	17	35	42
(.1, .1, .9)	13	26	30
(.4, .3, .2)	10	21	24
(.25, .25, .25)	8	17	20
(.1, .1, .1)	3	7	8

$p = 5$

DN	10	20	30
(.99, .99, .99, .99, .99)	7%	56%	76%
(.9, .9, .8, .7, .7)	7	46	76
(.7, .6, .6, .5, .4)	5	32	41
(.1, .1, .7, .7, .9)	5	28	36
(.5, .5, .5, .5, .5)	5	28	36
(.1, .1, .1, .9, .9)	4	23	30
(.4, .3, .4, .3, .2)	3	18	22
(.25, .25, .25, .25, .25)	3	13	17
(.1, .1, .1, .1, .1)	1	5	7

$p = 10$

DN	20	30	50
(.99, .99, .99, .99, .99, .99, .99, .99, .99, .99)	19%	47%	73%
(.9, .9, .9, .9, .9, .8, .8, .8, .8, .8)	18	41	62
(.9, .9, .9, .7, .7, .7, .6, .6, .6, .5)	16	34	51
(.5, .5, .5, .5, .5, .5, .5, .5, .5, .5)	12	24	35
(.1, .1, .9, .9, .2, .2, .8, .8, .1, .9)	12	24	35
(.1, .1, .2, .2, .5, .5, .7, .7, .8, .9)	11	22	32
(.4, .4, .4, .3, .3, .3, .2, .2, .2, .2)	7	14	20
(.25, .25, .25, .25, .25, .25, .25, .25, .25, .25)	6	11	17
(.1, .1, .1, .1, .1, .1, .1, .1, .1, .1)	2	4	6

Acknowledgements

The author would like to thank Sanat K. Sarkar for helpful discussions during the preparation of this paper. He is grateful to the referees for their comments and suggestions.

Appendix

We first give "matrix derivatives" of some functions of matrices which are needed in evaluating the expectations given in Section 2. "Matrix derivative" is essentially a collection of partial derivatives arranged in orderly arrays. For more details and application the readers can refer to Roger ((1980), p. 82) and von Rosen (1988).

We now define some basic tools which are needed in the sequel. For a matrix $A : p \times q$, $\text{vec}(A)$ denotes the $pq \times 1$ vector obtained by arranging the columns of A one after another in a longer column. For matrices $A = (a_{ij})$ and $B = (b_{ij})$, $A \otimes B = (a_{ij}B)$ is the *Kronecker product.* Let $J_{\alpha\beta}$ denote the $p \times p$ matrix whose (α, β) element is unity and other elements are zeroes and let

$$I_{(p,p)} = \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} J_{\alpha\beta} \otimes J_{\beta\alpha}.$$

The derivatives given in the following lemma can be found in Roger (1980) and von Rosen (1988).

LEMMA A.1. *Let X be a $p \times p$ symmetric matrix whose elements are otherwise functionally independent. Let $Y : p \times p$ and $Z : p \times p$ be matrix valued functions of X, and y and z be real valued functions of X. Further, let A and B denote the matrices of constants. Then,*

$$\text{(i)} \qquad \frac{\partial Y}{\partial X} = \Sigma_\kappa \left(\frac{\partial y_{ij}}{\partial x_{kl}} J_{ij} \otimes \gamma_{kl} J_{kl} \right), \qquad \gamma_{kl} = \begin{cases} 1 & \text{if } k = l \\ \frac{1}{2} & \text{if } k \neq l \end{cases}$$

where $\kappa = \{(i, j, k, l) : 1 \leq i, j, k, l \leq p\}$.

$$\text{(ii)} \quad \frac{\partial X}{\partial X} = \frac{1}{2}[\text{vec}(I_p)\text{vec}(I_p)' + I_{(p,p)}]$$

$$\text{(iii)} \quad \frac{\partial\, \text{tr}(AX)}{\partial X} = (A + A')/2$$

$$\text{(iv)} \quad \frac{\partial |X|^q}{\partial X} = q|X|^q X^{-1} \qquad (\textit{provided } X \textit{ is nonsingular})$$

$$\text{(v)} \quad \frac{\partial (YZ)}{\partial X} = \frac{\partial Y}{\partial X}(Z \otimes I_p) + (Y \otimes I_p)\frac{\partial Z}{\partial X}$$

$$\text{(vi)} \quad \frac{\partial Zy}{\partial X} = \frac{\partial Z}{\partial X} y + \left(Z \otimes \frac{\partial y}{\partial X} \right)$$

$$\text{(vii)} \quad \frac{\partial (AYB)}{\partial X} = (A \otimes I_p) \frac{\partial Y}{\partial X} (B \otimes I_p)$$

$$\text{(viii)} \quad \frac{\partial Y^{-1}}{\partial X} = -(Y^{-1} \otimes I_p) \frac{\partial Y}{\partial X} (Y^{-1} \otimes I_p) \qquad (\textit{provided } Y \textit{ is nonsingular})$$

$$\text{(ix)} \quad \frac{\partial\, \text{tr}(Y)}{\partial X} = \frac{\partial (\sum_{i=1}^{p} e_i' Y e_i)}{\partial X} = \sum_{i=1}^{p} (e_i' \otimes I_p) \frac{\partial Y}{\partial X} (e_i \otimes I_p)$$

$$\text{(x)} \quad \frac{\partial y(z(X))}{\partial X} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial X}$$

where I_p denotes the identity matrix of order $p \times p$ and e_i denotes the $p \times 1$ vector whose i-th element is unity and others are zeroes.

LEMMA A.2. *Let $S \sim W_p(n, \Sigma), n \geq p+1$. The pdf of S is given by*

$$f(S) = c|S|^{(n-p-1)/2}|\Lambda|^{n/2}e^{-\operatorname{tr}(\Lambda S)}, \quad \Lambda \equiv (1/2)\Sigma^{-1} \quad \text{and}$$
$$\frac{\partial f}{\partial \Lambda} = (n/2)\Lambda^{-1}f(S) - Sf(S).$$

PROOF. Use Lemma A.1(x), (ii) and (iii).

In the following theorem, we use these derivative results to compute the expections given in Section 2.

For $n \geq p+4$, let $c_1 = [(n-p)(n-p-1)(n-p-3)]^{-1}$ and $c_2 = (n-p-1)c_1$.

THEOREM A.1. *Let $S \sim W_p(n, \Sigma)$ and B be a $p \times p$ matrix of constants. Then,*

(i) $E(S^{-1}BS) = (n-p-1)^{-1}(n\Sigma^{-1}B\Sigma - (\operatorname{tr} B)I_p - B)$

(ii) $B(S^{-2}BS) = nc_1(\operatorname{tr}\Sigma^{-1})\Sigma^{-1}B\Sigma + nc_2\Sigma^{-2}B\Sigma - c_1(\operatorname{tr}\Sigma^{-1})(I_p \operatorname{tr} B + B) - 2(c_1+c_2)\Sigma^{-1}B - c_2[(\operatorname{tr}\Sigma^{-1}B)I_p + (\operatorname{tr} B)\Sigma^{-1}]$

(iii) $E(S^{-2}BSBS) = n^2[c_1(\operatorname{tr}\Sigma^{-1})\Sigma^{-1}B\Sigma B\Sigma + c_2\Sigma^{-2}B\Sigma B\Sigma] - nc_1(\operatorname{tr}\Sigma^{-1})\cdot[(\operatorname{tr} B)B\Sigma - \Sigma^{-1}B\Sigma^2 B + B^2\Sigma + B\Sigma B + \operatorname{tr}(B\Sigma B)I_p - \operatorname{tr}(\Sigma B)\Sigma^{-1}B\Sigma] - 2nc_1(\Sigma^{-1}B\Sigma^{-1}B) - nc_2[(\operatorname{tr}\Sigma^{-1}B)B\Sigma + (\operatorname{tr} B)\Sigma^{-1}B\Sigma + \operatorname{tr}(\Sigma^{-1}B\Sigma B)I_p + 2(\Sigma^{-1}B\Sigma B) + \operatorname{tr}(B\Sigma B)\Sigma^{-1} - \operatorname{tr}(\Sigma B)\Sigma^{-2}B\Sigma - \Sigma^{-2}B\Sigma^2 B] + (c_1+c_2)[4B^2 + 2(\operatorname{tr} B^2)I_p - 2n\Sigma^{-1}B^2\Sigma] + (2c_1+c_2)(\operatorname{tr} B)B + c_2(\operatorname{tr} B)^2 I_p$.

PROOF. (i) We know that

$$E(S^{-1}B) = \int_{S>0} S^{-1}Bf(S)dS = 2(n-p-1)^{-1}\Lambda B. \tag{A.1}$$

Differentiating both sides of (A.1) (using Lemma A.1(vii) and Lemma A.2) with respect to Λ, we get

$$\begin{aligned} E(S^{-1}B \otimes S) &= n(n-p-1)^{-1}(\Sigma^{-1}B \otimes \Sigma) \\ &\quad - (n-p-1)^{-1}[\operatorname{vec}(I_p)\operatorname{vec}(I_p)' + I_{(p,p)}](B \otimes I_p). \end{aligned} \tag{A.2}$$

Postmultiplying both sides of (A.2) by $\operatorname{vec}(I_p)$, we get (i).

(ii) It is known that (for example, see Haff (1979)),

$$E(S^{-2}B) = 4c_1(\operatorname{tr}\Lambda)\Lambda B + 4c_2\Lambda^2 B. \tag{A.3}$$

Use Lemma A.1(vi), (vii) and Lemma A.2 to differentiate both sides of (A.3) with respect to Λ. We get

$$\begin{aligned} &E(S^{-2}B \otimes S) \\ &\quad = n[c_1(\operatorname{tr}\Sigma^{-1})\Sigma^{-1}B + c_2\Sigma^{-2}B] \otimes \Sigma \\ &\qquad - c_1(\operatorname{tr}\Sigma^{-1})[\operatorname{vec}(I_p)\operatorname{vec}(I_p)' + I_{(p,p)}](B \otimes I_p) \\ &\qquad - 2c_1(\Sigma^{-1}B \otimes I_p) - c_2[\operatorname{vec}(I_p)\operatorname{vec}(I_p)' + I_{(p,p)}](\Sigma^{-1}B \otimes I_p) \\ &\qquad - c_2(\Sigma^{-1} \otimes I_p)[\operatorname{vec}(I_p)\operatorname{vec}(I_p)' + I_{(p,p)}](B \otimes I_p), \end{aligned} \tag{A.4}$$

and now (ii) follows from (A.4).

(iii) Premultiplying both sides of (ii) by B and then proceeding as in the proofs of (i) and (ii) we get (iii).

COROLLARY A.1. *Let $S \sim W_p(n, I)$ and $n \geq p+4$.*

(i) $E\,\mathrm{tr}(S^{-1}BSB) = (n-p-1)^{-1}[(n-1)\,\mathrm{tr}\,B^2 - (\mathrm{tr}\,B)^2]$

(ii) $E\,\mathrm{tr}(S^{-2}BSB) = [c_1(np-p-2) + c_2(n-2)](\mathrm{tr}\,B^2) - (c_1p + 2c_2)(\mathrm{tr}\,B)^2$

(iii)
$$\begin{aligned} &E\,\mathrm{tr}(S^{-2}BSBSB) \\ &\quad = [c_1(np-4)(n-1) + c_2(n^2-3n+4)](\mathrm{tr}\,B^3) \\ &\qquad - [c_1(np-4) + 3c_2(n-1)](\mathrm{tr}\,B)(\mathrm{tr}\,B^2) + c_2(\mathrm{tr}\,B)^3. \end{aligned}$$

THEOREM A.2. *Let $S \sim W_p(n, I)$. Then, for a matrix of constants A,*

(i) $E(\mathrm{tr}\,AS)(\mathrm{tr}\,S^{-1}AS^{-1}) = c_1(n-2)(n+1)(\mathrm{tr}\,A)^2 - 2c_1(2n-p-2)\,\mathrm{tr}(A^2)$

(ii) $E\,\mathrm{tr}(AS)\,\mathrm{tr}(S^{-1}A^2S^{-1}) = c_1(n-2)(n+1)(\mathrm{tr}\,A)(\mathrm{tr}\,A^2) - 2c_1(2n-p-2)\,\mathrm{tr}(A^3)$.

PROOF. Let B be a matrix of constants and θ be a real variable. Then it is easy to check that

$$E(\mathrm{tr}\,AS)(\mathrm{tr}\,S^{-1}BS^{-1}) = -\left.\frac{\partial E\,\mathrm{tr}(S^{-1}BS^{-1})e^{-\theta(\mathrm{tr}\,AS)}}{\partial\theta}\right|_{\theta=0}. \tag{A.5}$$

Using the derivative $\partial|I + 2\theta A|^{n/2}/\partial\theta\,|_{\theta=0} = n\,\mathrm{tr}\,A$ in (A.5) it follows that

$$\begin{aligned} &E(\mathrm{tr}\,AS)(\mathrm{tr}\,S^{-1}BS^{-1}) \\ &\quad = c_1(n-2)(n+1)(\mathrm{tr}\,A)(\mathrm{tr}\,B) - 2c_1(2n-p-2)(\mathrm{tr}\,AB) \end{aligned} \tag{A.6}$$

Now (i) and (ii) follow from (A.6).

REFERENCES

Brown, L. D. and Cohen, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information, *Ann. Statist.*, **2**, 963–976.

Chiou, W. and Cohen, A. (1985). On estimating a common multivariate normal mean vector, *Ann. Inst. Statist. Math.*, **37**, 499–506.

Cohen, A. and Sackrowitz, H. B. (1974). On estimating the common mean of two normal distributions, *Ann. Statist.*, **2**, 1274–1282.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators, *Biometrics*, **15**, 543–550.

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

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

Khatri, C. G. and Shah, K. R. (1974). Estimation of location parameters from two linear models under normality, *Comm. Statist. Simulation Comput.*, **3**, 647–665.

Krishnamoorthy, K. and Rohatgi, V. K. (1988). Unbiased estimation of the common mean of two or more normal populations, Tech. Report, Department of Mathematics and Statistics, Bowling Green State University, Ohio.

Kubokawa, T. (1989). Estimating common parameters of growth curve models under a quadratic loss, *Comm. Statist. Theory Methods*, **18**, 3149–3155.

Loh, W. L. (1988). Estimating the common mean of two multivariate normal distributions, Tech. Report, No. 88-48, Purdue University, Indiana.

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

Roger, G. S. (1980). *Matrix Derivatives*, Dekker, New York.

von Rosen, D. (1988). Moments for the inverted Wishart distribution, *Scand. J. Statist.*, **15**, 97–109.