

# Statistical Proofs of Some Matrix Theorems\*

C. Radhakrishna Rao

*Statistics Department, Joab Thomas Building, Pennsylvania State University, University Park, PA  
16802, USA*

## Summary

Books on linear models and multivariate analysis generally include a chapter on matrix algebra, quite rightly so, as matrix results are used in the discussion of statistical methods in these areas. During recent years a number of papers have appeared where statistical results derived without the use of matrix theorems have been used to prove some matrix results which are used to generate other statistical results. This may have some pedagogical value. It is not, however, suggested that prior knowledge of matrix theory is not necessary for studying statistics. It is intended to show that a judicious use of statistical and matrix results might be of help in providing elegant proofs of problems both in statistics and matrix algebra and make the study of both the subjects somewhat interesting. Some basic notions of vector spaces and matrices are, however, necessary and these are outlined in the introduction to this paper.

*Key words:* Cauchy–Schwarz inequality; Fisher information; Kronecker product; Milne’s inequality; Parallel sum of matrices; Schur product.

## 1 Introduction

Matrix algebra is extensively used in the study of linear models and multivariate analysis in statistics [see for instance Rao (1973), Rao & Toutenburg (1999) and Rao & Rao (1998)]. It is generally stated that a knowledge of matrix theory is a prerequisite for the study of statistics. Recently a number of papers have appeared where statistical results are used to prove some matrix theorems. [Refs: Dey, Hande & Tiku (1994), Kagan & Landsman (1997), Rao (2000), Kagan & Smith (2001), Kagan & Rao (2003, 2005), Kagan, Landsman & Rao (2006)]. The object of this paper is to review some of these results and give some new results. It may be noted that the statistical results used to prove matrix theorems are derivable without using matrix theory. We consider only matrices and vectors with real numbers.

### 1.1 Some Basic Notions of Vector Spaces and Matrices

A symmetric  $p \times p$  matrix  $A$  is said to be nnd (nonnegative definite) if  $a' A a \geq 0$  for any  $p$ -vector  $a$ , which is indicated by  $A \geq 0$ .  $A$  is said to be pd (positive definite) if  $a' A a > 0$  for all non-null  $a$ , which is indicated by  $A > 0$ . We write  $A \geq B$  if  $A - B$  is nnd and  $A > B$  if  $A - B$  is pd. The notation  $C : p \times k$  is used to denote a matrix  $C$  with  $p$  rows and  $k$  columns. The transpose of  $C$  obtained by interchanging columns and rows is denoted by  $C' : k \times p$ . The rank of any matrix  $X$ , denoted by  $\rho(X)$ , is the number of independent columns (equivalently of independent rows) in  $X$ .

\*The paper is based on a keynote presentation at the 14th International Workshop on Matrices and Statistics meeting in Auckland, New Zealand, March 29–April 1, 2005.

We denote by  $\mu(X)$  the vector space generated by the column vectors of  $X : p \times k$ ,

$$\mu(X) = \{y : y = Xa, a \in R^k\}$$

where  $R^k$  the vector space of all  $k$ -vectors.

Given  $X : p \times q, \rho(X) = r$ , there exists a matrix  $Y : p \times s, s = \rho(Y) = p - r$  such that with respect to a given pd matrix  $V$

$$X'VY = 0$$

where  $0 : q \times s$  is a matrix with all elements zero. The vector spaces  $\mu(X)$  and  $\mu(Y)$  are said to be mutually  $V$ -orthogonal. [If  $V = I$ , the identity matrix,  $\mu(X)$  and  $\mu(Y)$  are simply said to be orthogonal]. The space  $R^p$  of all  $p$ -vectors is generated by

$$\mu(X) \oplus \mu(Y) = \{Xa + Yb, a \in R^q, b \in R^s\}.$$

If  $z \in R^p$ , then

$$z = Xa + Yb.$$

Multiplying by  $X'V$ , we have

$$\begin{aligned} X'Vz &= X'VXa \Rightarrow a = (X'VX)^- X'Vz \\ Xa &= X(X'VX)^- X'Vz \end{aligned}$$

so that the part of  $z$  lying in  $\mu(X)$  can be expressed as  $P_X z$  where

$$P_X = X(X'VX)^- X'V \quad (1.1)$$

is called the projection operator on the space  $\mu(X)$ . [The explicit expression (1.1) for a projection operator in terms of a generalized inverse defined in (1.3) was first reported in Rao (1967). See also Rao (1973, p. 47)].

Note that we can choose the orthogonal complement of  $\mu(X)$  as  $\mu(Y)$ , where

$$Y = I - X(X'VX)^- X'V. \quad (1.2)$$

Given a matrix  $X : m \times n$ , there exists a matrix  $X^- : n \times m$ , called the *g-inverse* (generalized inverse) of  $X$ , such that

$$XX^-X = X. \quad (1.3)$$

$X^-$  need not be unique unless  $X$  is a square matrix of full rank. If  $C$  is any matrix such that  $\mu(C) \subset \mu(X)$ , then  $XX^-C = C$ , so that  $XX^-$  behaves like an identity matrix in such situations. [See Rao (1973), Rao & Mitra (1971, Section 16.5), and Rao & Rao (1998, Chapter 8)].

## 2 Some Basic Results

Let  $x = (x_1, \dots, x_p)'$  be a  $p$ -vector rv (random variable) and  $y = (y_1, \dots, y_q)$  be a  $q$ -vector rv all with zero mean values. We denote the covariance matrix of  $x$  and  $y$  by

$$C(x, y) = E(xy') = \begin{pmatrix} E(x_1y_1) & E(x_1y_2) & \dots & E(x_1y_q) \\ \cdot & \cdot & \dots & \cdot \\ E(x_py_1) & E(x_py_2) & \dots & E(x_py_q) \end{pmatrix} = C(y, x)'$$

which is a  $p \times q$  matrix. The following results hold.

THEOREM 2.1. If  $A : p \times p$  is the covariance matrix of a  $p$ -vector rv  $x$ , then:

$$(a) \quad A \text{ is nnd.} \quad (2.1)$$

$$(b) \quad \text{If } \rho(A) = s < p, \text{ then there exists a matrix} \\ D : p \times (p - s) \text{ such that } y = D'x = 0 \text{ a.s.} \quad (2.2)$$

Conversely, if  $A : p \times p$  is an nnd matrix, then there exists a rv  $x$  such that

$$C(x, x) = A. \quad (2.3)$$

To prove (2.1) consider the scalar rv  $a'x$ ,  $a \in R^p$ . Its variance is

$$V(a'x) = E(a'xx'a) = a'Aa \geq 0 \Rightarrow A \text{ is nnd.}$$

The result (2.2) follows by choosing  $D$  as the orthogonal complement of  $A$ . Then

$$C(D'x, D'x) = \text{Cov}(D'x) = D'AD = 0 \Rightarrow D'x = 0 \text{ a.s.}$$

*Note:* The result (2.3) follows if we assume the existence of an nnd matrix  $A^{1/2}$  such that  $(A^{1/2})(A^{1/2}) = A$  (or a matrix  $B$  such that  $A = BB'$ ). Then we need only choose an rv  $z$  with  $I$  as its covariance matrix and define  $x = A^{1/2}z$  (or  $x = Bz$ ). However, we prove the result without using the matrix result on the existence of  $A^{1/2}$  (or the factorization  $A = BB'$ ). Later we show that the factorization  $A = BB'$  is a consequence of a statistical result.

We assume that there exists a  $(p - 1)$ -vector rv with any given  $(p - 1) \times (p - 1)$  nnd matrix as its covariance matrix and show that there exists a  $p$ -vector rv associated with any given  $p \times p$  nnd matrix. Since an rv exists when  $p = 1$ , there exists one when  $p = 2$  and so on.

Let us consider the nnd matrix  $A : p \times p$  in the partitioned form

$$A = \begin{pmatrix} A_{11} & \alpha \\ \alpha' & \beta \end{pmatrix}$$

where  $A_{11} : (p - 1) \times (p - 1)$ . Then, since  $A$  is nnd, the quadratic form

$$a'A_{11}a + 2c(\alpha'a) + c^2\beta \geq 0 \quad \text{for all } a \text{ and } c. \quad (2.4)$$

If  $\beta = 0$ , then  $\alpha'a = 0$  for all  $a$  which implies that  $\alpha = 0$ . If  $x_1$  is an rv associated with the nnd matrix  $A_{11}$ , then  $(x_1', 0)'$  is an rv associated with  $A$ .

If  $\beta \neq 0$ , then, choosing  $c = -(\alpha'a)/\beta$ , (2.4) reduces to

$$a'(A_{11} - \beta^{-1}\alpha\alpha')a \geq 0 \quad \text{for all } a$$

i.e.,  $A_{11} - \beta^{-1}\alpha\alpha'$  is nnd. Let  $x_1$  be an rv associated with  $(A_{11} - \beta^{-1}\alpha\alpha') : (p - 1) \times (p - 1)$  and  $z$  be a univariate rv with variance  $\beta$  and independent of  $x_1$ . Then it is easy to verify that the rv

$$u = \begin{pmatrix} x_1 + \alpha\beta^{-1}z \\ z \end{pmatrix} \quad (2.5)$$

has  $A$  as its covariance matrix.

THEOREM 2.2. Let  $A : p \times p$  be a symmetric matrix partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Then  $A$  is nnd if and only if

- (a)  $A_{11}$  is nnd,  
 (b)  $A_{22} - A_{21}A_{11}^{-1}A_{12} = A_{2.1}$  (Schur complement of  $A_{11}$ ) is nnd,  
 (c)  $\mu(A_{12}) \subset \mu(A_{11})$ .

The theorem was proved by Albert (1969) using purely matrix methods. Statistical proofs were given by Dey *et al.* (1994) and Rao (2000).

First we prove sufficiency using Theorem 2.1. (a) implies that there exists a random variable  $x$  with mean zero and covariance matrix  $A_{11}$ . (b) implies that there exists a random variable  $z$  with mean zero and covariance matrix  $A_{2.1}$ . We can choose  $z$  to be independent of  $x$ . Now consider the random variable  $w = (x', y')'$  where  $y = z + A_{21}A_{11}^{-1}x$ . Verify using (c) that  $C(w, w) = A$ , which implies that  $A$  is nnd.

To prove necessity, observe that there exists a random variable  $(x', y')'$  with the covariance matrix  $A$ . (a) follows since  $C(x, x) = A_{11}$ . If  $A_{11}$  is not of full rank, let  $a$  be a nonnull vector such that  $a'A_{11} = 0$  which implies that  $a'x = 0$  a.s. and  $C(a'x, b'y) = a'A_{12}b = 0 \forall b$ , i.e.,  $a'A_{12} = 0$ . Hence  $\mu(A_{12}) \subset \mu(A_{11})$  which proves (c).

Now consider the random variable  $z = y - A_{21}A_{11}^{-1}x$  whose covariance matrix is

$$C(z, z) = A_{22} - A_{21}A_{11}^{-1}A_{12}$$

which implies (b).

**THEOREM 2.3.** Let  $A$  and  $B$  be nnd matrices partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

If  $\rho(A + B) = \rho(A_{11} + B_{11})$ , then  $\rho(A) = \rho(A_{11})$  and  $\rho(B) = \rho(B_{11})$ .

The statistical proof of this theorem given by Dey *et al.* (1994) and Mitra (1973) is reproduced here.

Let  $(U_1' : U_2')' \sim N(0, A)$  and  $(V_1' : V_2')' \sim N(0, B)$  be independent. Then  $Z = (U_1' : U_2')' + (V_1' : V_2')' = (Z_1' : Z_2')' \sim N(0, A + B)$ . Since  $\rho(A + B) = \rho(A_{11} + B_{11})$ , there exists a matrix  $G$  such that

$$(A_{21} + B_{21}) = G(A_{11} + B_{11}), \quad (A_{22} + B_{22}) = G(A_{12} + B_{12})$$

which implies

$$\text{Cov}(Z_2 - GU_1) = 0 \Rightarrow U_2 + V_2 = GU_1 + GV_1.$$

Then  $\text{Cov}(U_2 - GU_1) = 0$  and  $\text{Cov}(V_2 - GV_1) = 0$  which implies  $U_2 = GU_1$ ,  $\rho(A) = \rho(A_{11})$  and  $V_2 = GV_1$ ,  $\rho(B) = \rho(B_{11})$ .

**THEOREM 2.4.** Let  $A$  be an nnd matrix. Then it can be factorized as

$$A = BB'.$$

Let  $(x_1, \dots, x_p) = x'$  be rv with the nnd matrix  $A : p \times p$  as its covariance matrix. Then we can recursively construct random variables

$$z_i = x_i - a_{i,i-1}x_{i-1} - \dots - a_{i1}x_1, \quad i = 1, \dots, p \quad (2.6)$$

such that  $z_1, \dots, z_p$  are mutually uncorrelated. It may be noted that

$$a_{i1}x_1 + \dots + a_{i,i-1}x_{i-1}$$

is the regression of  $x_i$  on  $x_1, \dots, x_{i-1}$ . Denote by  $b_i^2$ , the variance of  $z_i$ . We can write the transformation (2.6) as

$$Z = Fx$$

where  $F$  is a triangular matrix of rank  $p$ . Then

$$\Delta^2 = C(z, z) = FC(x, x)F' = FAF'$$

where  $\Delta^2$  is a diagonal matrix with diagonal elements  $b_1^2, \dots, b_p^2$ , some of which may be zeros, and

$$A = F^{-1}\Delta^2(F^{-1})' = (F^{-1}\Delta)(F^{-1}\Delta)' = BB'$$

where  $\Delta$  is the diagonal matrix with diagonal elements  $b_1, \dots, b_p$ .

### 3 Cauchy–Schwarz (CS) and Related Inequalities

A simple (statistical) proof of the usual CS inequality is to use the fact that the second raw moment of an rv is nonnegative. Consider a bivariate rv,  $(x, y)$  such that

$$0 < b_{11} = E(x^2), \quad b_{12} = E(xy), \quad b_{22} = E(y^2).$$

Then

$$E\left(y - \frac{b_{12}}{b_{11}}x\right)^2 = b_{22} - \frac{b_{12}^2}{b_{11}} \geq 0 \tag{3.1}$$

which written in the form

$$b_{12}^2 \leq b_{11}b_{22} \tag{3.2}$$

is the usual CS inequality. If  $b_{11} = 0, x = 0$  a.s. and  $b_{12} = 0$  and the result (3.2) is true.

In particular, if  $(x, y)$  is a discrete bivariate distribution with  $(x_i, y_i)$  having probability  $p_i, i = 1, \dots, n$ , then

$$(\sum p_i x_i y_i)^2 \leq (\sum p_i x_i^2)(\sum p_i y_i^2) \tag{3.3}$$

since  $b_{12} = \sum p_i x_i y_i, b_{11} = \sum p_i x_i^2$  and  $b_{22} = \sum p_i y_i^2$ .

The matrix version of CS inequality is obtained by considering rv's  $x$  and  $y$  such that

$$\Sigma_{11} = C(x, x), \quad \Sigma_{12} = C(x, y), \quad \Sigma_{22} = C(y, y),$$

and noting that

$$C(y - \Sigma_{21}\Sigma_{11}^{-1}x, y - \Sigma_{21}\Sigma_{11}^{-1}x) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \geq 0. \tag{3.4}$$

A special case of (3.4) is obtained by considering rv's  $x = Az$  and  $y = Bz$  where  $z$  is an  $s$ -vector rv with uncorrelated components, in which case

$$\sum_{11} = AA', \quad \sum_{12} = AB', \quad \sum_{22} = BB'$$

and (3.4) reduces to

$$BB' \geq BA'(AA')^{-1}AB'. \tag{3.5}$$

The statistical proofs of the results in the following Lemmas are given in Dey *et al.* (1994).

LEMMA 3.1. *Let  $A$  and  $B$  be nnd matrices. Then*

$$[a'A^{-1}a][a'B^{-1}a] \geq [a'(A+B)^{-1}a][a'A^{-1}a + a'B^{-1}a]$$

*provided  $a \in \mu(A) \cap \mu(B)$ , with equality when either  $A$  or  $B$  has rank unity.*

Let  $A = X_1'X_1$  and  $B = X_2'X_2$ . Consider the linear models

$$Y_1 = X_1\beta, \quad Y_2 = X_2\beta, \quad Y_3 = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\beta = X_3\beta.$$

The best linear estimates of  $a'\beta$  under the three models are

$$T_i = a'\hat{\beta}_i, \quad \hat{\beta}_i = (X_i'X_i)^{-1}X_i'Y_i, \quad i = 1, 2, 3, \\ V(T_1) = a'A^{-1}a = w_1^{-1}, \quad V(T_2) = a'B^{-1}a = w_2^{-1}.$$

Now

$$a'(A+B)^{-1}a = V(T_3) \leq (w_1 + w_2)^{-2}V(w_1T_1 + w_2T_2) \\ = (w_1 + w_2)^{-2}[w_1^2V(T_1) + w_2^2V(T_2)] \\ = \frac{[a'A^{-1}a][a'B^{-1}a]}{[a'A^{-1}a] + [a'B^{-1}a]}.$$

If either  $\rho(A)$  or  $\rho(B)$  is unity,  $T_3$  is same as  $(w_1T_1 + w_2T_2)/(w_1 + w_2)$ . This completes the proof.

By using the same argument, the result of Lemma 3.1 can be generalized as follows.

LEMMA 3.2. If  $a \in \mu(A+B)$ , then

$$a'(A+B)^{-1}a = \min_{a_1 \in \mu(A), a_2 \in \mu(B)} [a_1'A^{-1}a_1 + a_2'B^{-1}a_2]$$

where  $a = a_1 + a_2$ .

LEMMA 3.3. Let  $A_i, i = 1, \dots, k$  be nnd matrices and  $a \in \cap \mu(A_i)$ . Then

$$a' \left( \sum_{i=1}^k A_i \right)^{-1} a \leq \sum_{i=1}^k \frac{1}{(a'A_i^{-1}a)^2} \div \left( \sum_{i=1}^k \frac{1}{a'A_i^{-1}a} \right)^2.$$

LEMMA 3.4. Let  $A$  be nnd matrix. Then for any choice of  $g$ -inverse

$$a'(A+aa')^{-1}a = a'A^{-1}a/(1+a'A^{-1}a) \text{ if } a \in \mu(A).$$

#### 4 Schur and Kronecker Product of Matrices

The following theorems are well known.

THEOREM 4.1. Let  $A$  and  $B$  be nnd matrices of order  $p \times p$ . Then the Schur product  $A \circ B$  is nnd. [ $A \circ B = (a_{ij}b_{ij})$ , where  $A = (a_{ij})$  and  $B = (b_{ij})$ ].

*Proof.* Let  $X$  be a  $p$ -vector rv such that  $C(X, X) = A$  and  $Y$  be an independent  $p$ -vector rv such that  $C(Y, Y) = B$ . Then it is seen that

$$C(X \circ Y, X \circ Y) = A \circ B$$

so that  $A \circ B$  is the covariance matrix of a rv, which proves the result.  $\square$

THEOREM 4.2. Let  $A : p \times p$  be an nnd matrix and  $B : q \times q$  be an nnd matrix. Then the Kronecker product  $A \otimes B$  is nnd. [ $A \otimes B = (a_{ij}B)$ , where  $A = (a_{ij})$ ].

*Proof.* Let  $X$  be a  $p$ -vector rv such that  $C(X, X) = A$  and  $Y$  be an independent  $q$ -vector rv such that  $C(Y, Y) = B$ . Then

$$C(X \otimes Y, X \otimes Y) = A \otimes B$$

so that  $A \otimes B$  is the covariance matrix of a rv, which proves the desired result.  $\square$

Note that Theorem 4.1 is a consequence of Theorem 4.2 as  $A \circ B$  is a principal submatrix of  $A \otimes B$ .

## 5 Results Based on Fisher Information

### 5.1 Properties of Information Matrix

Let  $x$  be a  $p$ -vector rv with density function  $f(x, \theta)$ , where  $\theta' = (\theta_1, \dots, \theta_q)$  is a  $q$ -vector parameter. Define the Fisher score vector

$$J_x = \left( \frac{\partial \log f}{\partial \theta_1}, \dots, \frac{\partial \log f}{\partial \theta_q} \right)'$$

and information matrix of order  $q \times q$

$$I_x(\theta) = E(J_x J_x') = \text{Cov}(J_x, J_x)$$

which is nnd.

*Monotonicity of information:* Let  $T(x)$  be a function of  $x$ . Then under some regularity conditions (given in Rao (1973, pp. 329–332))

$$I_T(\theta) \leq I_x(\theta) \quad (5.1)$$

where  $I_T(\theta)$  is the information matrix derived from the distribution of  $T$ .

*Additivity of information:* If  $x$  and  $y$  are independent random variable whose densities involve the same parameters, then

$$I_{x,y}(\theta) = I_x(\theta) + I_y(\theta). \quad (5.2)$$

*Information when  $x$  has a multivariate normal distribution:* Let  $x \sim N_p(B\mu, A)$ , i.e., distributed as  $p$ -variate normal with mean  $B\mu$  and a pd covariance matrix  $A$ . Then the information in  $x$  on  $\mu$  is

$$I_x(\mu) = B'A^{-1}B. \quad (5.3)$$

If  $y = Gx$ , then

$$I_y(\mu) = B'G'(GAG')^{-1}GB \quad (5.4)$$

and if  $G$  is invertible,  $I_x(\mu) = I_y(\mu)$ . Note that if  $G$  is of full rank,  $I_y(\mu) = I_x(\mu)$ , i.e., there is no loss of information in the transformation  $y = Gx$ .

Suppose that the parameter  $\theta$  is partitioned into two subvectors,  $\theta' = (\theta^{(1)'}, \theta^{(2)'})$  where  $\theta^{(1)'}$  =  $(\theta_1, \dots, \theta_p)$  and  $\theta^{(2)'}$  =  $(\theta_{p+1}, \dots, \theta_s)$ . The corresponding partition of the score vector is

$$J_x' = (J_{1x}', J_{2x}') \quad (5.5)$$

and the information matrix is

$$I_x(\theta) = \begin{pmatrix} E(J_{1x}J_{1x}') & E(J_{1x}J_{2x}') \\ E(J_{2x}J_{1x}') & E(J_{2x}J_{2x}') \end{pmatrix} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}. \quad (5.6)$$

We define the conditional score vector of  $\theta^{(1)}$ , when  $\theta^{(2)}$  is not of interest and considered as a nuisance parameter, by

$$\tilde{J}_{1x} = J_{1x} - I_{12}I_{22}^{-1}J_{2x} \quad (5.7)$$

which is the residual of  $J_{1x}$  subtracting the regression on  $J_{2x}$ , in which case

$$\tilde{I}_{1x}(\theta) = E_{\theta}(\tilde{J}_{1x}\tilde{J}_{1x}') = I_{11} - I_{12}I_{22}^{-1}I_{21}. \quad (5.8)$$

As one sees from the definition  $\tilde{I}_{1x} \leq I_{11}$ , the equality sign holding if and only if the components of  $J_{1x}$  and  $J_{2x}$  are uncorrelated. If

$$I_x^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}$$

then

$$\tilde{I}_{1x}(\theta) = I_{11} - I_{12}I_{22}^{-1}I_{21} = (I_x^{11})^{-1} \quad (5.9)$$

which can be proved as follows.

Consider a random variable  $p + q$  vector r.v.

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N_{p+q} \left( \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \right).$$

Then

$$I_{x,y}(\mu) = I^{11}$$

using (5.3) with  $B = (1, 1, \dots, 1, 0, 0, \dots, 0)$ . Further

$$z = x - I_{12}I_{22}^{-1}y \text{ and } y$$

are independent. Hence

$$\begin{aligned} I^{11} &= I_{x,y}(\mu) = I_{z,y}(\mu) \\ &= I_z(\mu) + I_y(\mu) = I_z(\mu) = (I_{11} - I_{12}I_{22}^{-1}I_{21})^{-1} \end{aligned}$$

since  $C(z, z) = I_{11} - I_{12}I_{22}^{-1}I_{21}$ .

The monotonicity property remains the same for conditional information. Thus if  $T = T(x)$  is a statistic and  $I_T(\theta)$  is pd, then

$$\tilde{I}_{1T}(\theta) \leq \tilde{I}_{1x}(\theta). \quad (5.10)$$

But the additivity property may not hold. If  $x$  and  $y$  are independent random variables, we can only say

$$\tilde{I}_{1(x,y)}(\theta) \geq \tilde{I}_{1x}(\theta) + \tilde{I}_{1y}(\theta) \quad (5.11)$$

which is described as superadditivity, unless  $x$  and  $y$  are identically distributed in which case equality holds. For a proof, reference may be made to Kagan & Rao (2003).

## 5.2 Some Applications

**THEOREM 5.1.** *If  $A$  and  $B$  are pd matrices, then*

$$A \geq B \Rightarrow A^{-1} \leq B^{-1}. \quad (5.12)$$

*Proof.* Consider independent rv's

$$x \sim N(\mu, B) \text{ and } y \sim N(0, A - B + \delta^2 I)$$

where  $\delta^2 I$  is added to  $A - B$  to make  $C(y, y)$  positive definite. Note that  $\delta$  can be chosen as small

as we like. Then

$$(x + y) \sim N(\mu, A + \delta^2 I)$$

and

$$I_x(\mu) = B^{-1}, I_y(\mu) = 0, I_{x+y} = (A + \delta^2 I)^{-1}.$$

Using monotonicity (5.1) and additivity (5.2) properties of Fisher information, we have

$$(A + \delta^2 I)^{-1} = I_{x+y} \leq I_{x,y} = I_x + I_y = B^{-1}. \quad (5.13)$$

Since (5.13) is true for all  $\delta$ , taking the limit as  $\delta \rightarrow 0$ , we have the result (5.12).

The equality given in the following theorem is useful in the distribution theory of quadratic forms (Rao, 1973, p.77). We give a statistical proof of the equality.  $\square$

**THEOREM 5.2.** *Let  $\Sigma : p \times p$  be a pd matrix and  $B : k \times p$  of rank  $k$ ,  $C : p - k \times p$  of rank  $p - k$ , be matrices such that  $BC' = 0$ . Then*

$$C'(C\Sigma C')^{-1}C + \Sigma^{-1}B'(B\Sigma^{-1}B')B\Sigma^{-1} = \Sigma^{-1}. \quad (5.14)$$

*Proof.* Consider the random variable  $X \sim N_p(\mu, \Sigma)$  and the transformation

$$Y_1 = CX \sim N_{p-k}(C\mu, C\Sigma C'),$$

$$Y_2 = B\Sigma^{-1}X \sim N_k(B\Sigma^{-1}\mu, B\Sigma^{-1}B').$$

Here  $Y_1$  and  $Y_2$  are independent, and using additivity (5.1) of Fisher information we have

$$\Sigma^{-1} = I_X(\mu) = I_{Y_1, Y_2}(\mu) = I_{Y_1}(\mu) + I_{Y_2}(\mu).$$

Note that  $I_{Y_1}(\mu)$  is the first term and  $I_{Y_2}(\mu)$  is the second term on the left hand side of (5.14). See (Rao, 1973, p.77) for an algebraic proof.  $\square$

## 6 Milne's Inequality

Milne (1925) proved the following result.

**THEOREM 6.1.** *For any constants  $w_i > 0$ ,  $i = 1, \dots, n$ , such that  $w_1 + \dots + w_n = 1$  and given  $\rho_i$ ,  $(|\rho_i| < 1)$ ,  $i = 1, \dots, n$ ,*

$$\left( \sum \frac{w_i}{1 - \rho_i^2} \right)^2 \geq \left( \sum \frac{w_i}{1 - \rho_i} \right) \left( \sum \frac{w_i}{1 + \rho_i} \right) \geq \sum \frac{w_i}{1 - \rho_i^2}. \quad (6.1)$$

*Proof.* We give a statistical proof of Milne's inequalities.

To prove the right hand side inequality, consider the bivariate discrete distribution of  $(X, Y)$ :

$$\begin{array}{lll} \text{probability} & w_1, & \dots, w_n \\ \text{value of } X & (1 - \rho_1)^{-1}, & \dots, (1 - \rho_n)^{-1} \\ \text{value of } Y & (1 + \rho_1)^{-1}, & \dots, (1 + \rho_n)^{-1} \end{array} \quad (6.2)$$

We use the well known formula

$$2C(X, Y) = E(X_1 - X_2)(Y_1 - Y_2) \quad (6.3)$$

where  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are independent samples from the distribution (6.2). A typical term on the right hand side of (6.3) is

$$\begin{aligned} & \left( \frac{1}{1-\rho_i} - \frac{1}{1-\rho_j} \right) \left( \frac{1}{1+\rho_i} - \frac{1}{1+\rho_j} \right) \\ &= -\frac{(\rho_i - \rho_j)^2}{(1-\rho_i^2)(1-\rho_j^2)} < 0. \end{aligned}$$

Hence  $0 \geq C(X, Y) = E(XY) - E(X)E(Y)$ . The right hand side inequality follows by observing that

$$E(XY) = \sum \frac{w_i}{1-\rho_i^2}, \quad E(X) = \sum \frac{w_i}{1-\rho_i} \quad \text{and} \quad E(Y) = \sum \frac{w_i}{1+\rho_i}.$$

□

COROLLARY 1. For any rv  $X$  with  $\Pr\{|X| < A\} = 1$ ,

$$E\left(\frac{1}{A-X}\right)E\left(\frac{1}{A+X}\right) \geq E\left(\frac{1}{A^2-X^2}\right) \quad (6.4)$$

We use the same arguments as in the main theorem exploiting (6.3), choosing  $A - X$  and  $A + X$  as the variables  $X$  and  $Y$ .

COROLLARY 2. (Matrix version of Milne's inequality). Let  $A, V_1, \dots, V_n$  be symmetric commuting matrices such that  $A > 0, -A < V_i < A, i = 1, \dots, n$ . Then

$$\left(\sum w_i(A - V_i)^{-1}\right)\left(\sum w_i(A + V_i)^{-1}\right) \geq \sum w_i(A^2 - V_i^2)^{-1} \quad (6.5)$$

where  $w_i > 0$  and  $w_1 + \dots + w_n = 1$ .

A purely statistical proof of (6.5) is not known. The proof given in Kagan & Rao (2003) depends on the result that an nnd matrix  $A$  has an nnd square root  $A = B^2$ . Is there a statistical proof of this result?

A different statistical proof of Milne's inequality (6.1) given in Kagan & Rao (2003) is as follows.

*Proof.* Let  $x_1, \dots, x_n$  be independent bivariate normal vectors such that

$$E(x_i) = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad C(x_i, x_i) = \frac{1}{v_i} \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}$$

in which case the elements of  $I_{x_i}$  are

$$\begin{aligned} I_{11x_i} &= I_{22x_i} = \frac{v_i}{1-\rho_i^2}, \\ I_{12x_i} &= I_{21x_i} = \frac{-\rho_i v_i}{1-\rho_i^2}, \end{aligned}$$

and the information based on  $x$ , the whole set of  $x_1, \dots, x_n$  is

$$\begin{aligned} I_{11x} &= I_{22x} = \sum \frac{v_i}{1-\rho_i^2}, \\ I_{12x} &= I_{21x} = -\sum \frac{\rho_i v_i}{1-\rho_i^2}, \end{aligned}$$

where

$$\tilde{I}_{1x} = \sum \frac{v_i}{1-\rho_i^2} - \left(\sum \frac{\rho_i v_i}{1-\rho_i^2}\right)^2 \left(\sum \frac{v_i}{1-\rho_i^2}\right)^{-1}. \quad (6.6)$$

Consider now the vector  $x^1$  of the first components of  $x_{11}, \dots, x_{1n}$  of  $x$ . The distribution of  $x^1$  depends only on  $\theta_1$  and

$$I_{x^1}(\theta_1) = \sum v_i$$

and trivially

$$I_{1x^1}(\theta_1) = I_{x^1}. \tag{6.7}$$

Since  $x^1$  is a statistic, by applying the super additivity result (5.11) to (6.6) and (6.7), we have

$$\sum \frac{v_i}{1 - \rho_i^2} - \left( \sum \frac{\rho_i v_i}{1 - \rho_i^2} \right)^2 \left( \sum \frac{v_i}{1 - \rho_i^2} \right)^{-1} \geq \sum v_i. \tag{6.8}$$

Dividing both sides of (6.8) by  $\sum v_i$  and setting  $w_i = v_i / \sum v_i$  leads after simple calculation to the right-hand side inequality of (6.1).

Note that the classical inequality between arithmetic and geometric means gives

$$\left( \sum \frac{w_i}{1 - \rho_i} \right) \left( \sum \frac{w_i}{1 + \rho_i} \right) \leq \left[ \frac{1}{2} \sum \left( \frac{w_i}{1 - \rho_i} + \frac{w_i}{1 + \rho_i} \right) \right]^2 = \left( \sum \frac{w_i}{1 - \rho_i^2} \right)^2$$

which establishes the left-hand side inequality in (6.1). □

### 7 Convexity of Some Matrix Functions

Let

$$A_i = \begin{pmatrix} A_{11i} & A_{12i} \\ A_{21i} & A_{22i} \end{pmatrix}, \quad i = 1, \dots, n$$

be nnd matrices. Further let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \sum_1^n w_i A_i, \quad w_i > 0 \text{ and } \sum_1^n w_i = 1,$$

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = A_1 \circ A_2 \circ \dots \circ A_n = \Pi_1^n \circ A_i.$$

Since  $A$  and  $B$  are nnd matrices, we have

$$A_{22} - A_{21} A_{11}^- A_{12} \geq 0, \quad A_{11} - A_{12} A_{22}^- A_{21} \geq 0, \tag{7.1}$$

$$B_{22} - B_{21} B_{11}^- B_{12} \geq 0, \quad B_{11} - B_{12} B_{22}^- B_{21} \geq 0. \tag{7.2}$$

As a consequence of (7.1) and (7.2), we have the following theorems.

**THEOREM 7.1.** *Let  $C_i, i = 1, \dots, n$  be symmetric matrices of the same order. Then*

$$\sum_1^n w_i C_i^2 \geq \left( \sum_1^n w_i C_i \right)^2, \tag{7.3}$$

$$\prod_1^n \circ C_i^2 \geq \left( \prod_1^n \circ C_i \right)^2. \tag{7.4}$$

*Proof.* Choose  $A_{11i} = C_i^2, A_{12i} = A_{21i} = C_i$  and  $A_{22i} = I$ . Then  $A_i$  is nnd by the sufficiency part of Theorem 2.2. Hence using (7.1) and (7.2), we have (7.3) and (7.4) respectively. □

THEOREM 7.2. Let  $C_i, i = 1, \dots, n$ , be pd matrices. Then

$$\sum_1^n w_i C_i^{-1} \geq \left( \sum_1^n w_i C_i \right)^{-1}, \quad (7.5)$$

$$\prod_1^n \circ C_i^{-1} \geq \left( \prod_1^n \circ C_i \right)^{-1}. \quad (7.6)$$

*Proof.* Choose  $A_{11i} = C_i^{-1}$ ,  $A_{12i} = A_{21i} = C_i$ ,  $A_{22i} = I$ . Applying the same argument as in Theorem 7.1, the results (7.5) and (7.6) follow from (7.3) and (7.4) respectively. The inequality (7.5) is proved by Olkin & Pratt (1958). See also Marshall & Olkin (1979, Chapter 16). In particular, if  $C_1$  and  $C_2 > 0$ , then  $(C_1 \circ C_2) \geq (C_1^{-1} \circ C_2^{-1})^{-1}$ .  $\square$

THEOREM 7.3. Let  $C_i : s_i \times s_i$ ,  $i = 1, \dots, n$  and  $B_i : s_i \times m$ ,  $i = 1, \dots, n$ , be matrices such that  $\sum B_i' B_i = I_m$ . Then

$$\sum B_i' C_i C_i' B_i \geq \left( \sum B_i' C_i B_i \right)^2. \quad (7.7)$$

In particular, when  $C_i$  are symmetric for all  $i$ , (7.9) reduces to

$$\sum B_i' C_i^2 B_i \geq \left( \sum B_i' C_i B_i \right)^2. \quad (7.8)$$

*Proof.* Since the matrix

$$\begin{pmatrix} B_i' C_i C_i' B_i & B_i' C_i B_i \\ B_i' C_i B_i & B_i' B_i \end{pmatrix}$$

is nnd, we get (7.8) and (7.9) by applying (7.3). Some of these results are proved in Kagan & Landsman (1997) and Kagan & Smith (2001) using Information Inequality (5.1)  $\square$

THEOREM 7.4. Let  $A$  and  $B$  be pd matrices of order  $n \times n$  and  $X$  and  $Y$  be  $m \times n$  matrices. Then

$$(X \circ Y)' (A \circ B)^{-1} (X \circ B) \leq (X' A^{-1} X) \circ (Y' B^{-1} Y). \quad (7.9)$$

*Proof.* The result follows by considering the Schur product of the nnd matrices

$$\begin{pmatrix} X' A^{-1} X & X' \\ X & A \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} Y' B^{-1} Y & Y' \\ Y & B \end{pmatrix}$$

and taking the Schur complement.  $\square$

THEOREM 7.5. (Fiedler, 1961). Let  $A$  be a pd matrix. Then

$$A \circ A^{-1} \geq I. \quad (7.10)$$

*Proof.* The Schur product

$$\begin{pmatrix} A & I \\ I & A^{-1} \end{pmatrix} \circ \begin{pmatrix} A^{-1} & I \\ I & A \end{pmatrix} = \begin{pmatrix} A \circ A^{-1} & I \\ I & A^{-1} \circ A \end{pmatrix} \quad (7.11)$$

is nnd, since each matrix on the left hand side of (7.11) is nnd. Multiplying the right hand side matrix of (7.13) by the vector  $b' = (a', -a')$  on the left and by  $b$  on the right, where  $a$  is an arbitrary vector, we have

$$a'(A \circ A^{-1})a - a'a \geq 0 \quad \forall a \Rightarrow (A \circ A^{-1}) \geq I.$$

Purely matrix proofs of some of the results in this section can be found in Ando (1979, 1998).  $\square$

**THEOREM 7.6.** *Let  $A_1, \dots, A_n$  be  $s \times s$  pd matrices and define the partitions*

$$A_i = \begin{pmatrix} A_{11i} & A_{12i} \\ A_{21i} & A_{22i} \end{pmatrix} \text{ and } A_i^{-1} = \begin{pmatrix} A_i^{11} & A_i^{12} \\ A_i^{21} & A_i^{22} \end{pmatrix},$$

where  $A_{11i}$  is of order  $p \times p$ . Then

$$\left( \sum_1^n w_i A_i^{11} \right)^{-1} \geq \sum_1^n w_i (A_i^{11})^{-1}. \quad (7.12)$$

*Proof.* Let

$$x_i \sim N(v_i \theta, A_i), i = 1, \dots, n$$

be independent  $s$ -dimensional normal variables, where  $\theta' = (\theta'_1, \theta'_2)$  is a  $p + (s - p) = s$  dimensional parameter and  $v_j$  are known constants such that  $v_1^2 + \dots + v_n^2 = 1$ . Set  $x' = (x_1, \dots, x_n)'$ ,  $v_i^2 = w_i$ . Then

$$I_{x_i} = w_i A_i^{-1}, I_x = \sum w_i A_i^{-1}.$$

From (5.8) and the result,  $(A_i^{11})^{-1} = A_{11i} - A_{12i} A_{22i}^{-1} A_{21i}$ ,

$$\tilde{I}_{1x_i} = w_i (A_i^{11})^{-1}, \tilde{I}_{1x} \leq \left( \sum_1^n w_i A_i^{11} \right)^{-1}, \quad (7.13)$$

Due to superadditivity (5.11), (7.15) leads to (7.14).  $\square$

## 8 Inequalities on Harmonic Mean and Parallel Sum of Matrices

The following theorem is a matrix version of the inequality on harmonic means given in Rao (1996).

**THEOREM 8.1.** *Let  $A_1$  and  $A_2$  be random pd matrices of the same order, and  $E$  denote expectation. Then*

$$E \left[ (A_1^{-1} + A_2^{-1})^{-1} \right] \leq \left[ (E(A_1))^{-1} + (E(A_2))^{-1} \right]^{-1}. \quad (8.1)$$

*Proof.* Let  $C$  be any pd matrix. Then

$$\begin{pmatrix} A_1 C^{-1} A_1 & A_1 \\ A_1 & C \end{pmatrix} \quad (8.2)$$

is nnd by the sufficiency part of Theorem 2.2. Taking expectation of (8.2), and considering the Schur complement, we have

$$E(A_1 C^{-1} A_1) \geq E(A_1) (E(C))^{-1} E(A_1). \quad (8.3)$$

Now choosing  $C = A_1 + A_2$  and noting that

$$(A_1^{-1} + A_2^{-1})^{-1} = A_1 - A_1 (A_1 + A_2)^{-1} A_1$$

the result (8.1) follows by an application of (8.3). The same result was proved by Prakasa Rao (1998) using a different method.  $\square$

**THEOREM 8.2.** *Let  $A_1, \dots, A_p$  be random pd matrices and  $M_1, \dots, M_p$  be their expectations.*

Then

$$E(\tilde{A}) \leq \tilde{M} \tag{8.4}$$

where

$$\tilde{A} = (A_1^{-1} + \dots + A_p^{-1})^{-1}, \quad \tilde{M} = (M_1^{-1} + \dots + M_p^{-1})^{-1}$$

*Proof.* The result (8.4) is proved by repeated applications of (8.1). □

**THEOREM 8.3.** (Parallel Sum Inequality). *Let  $A_{ij}$ ,  $i = 1, \dots, p$  and  $j = 1, \dots, q$  be  $pd$  matrices. Denote*

$$\begin{aligned} \tilde{A}_{.i} &= (A_{1i}^{-1} + \dots + A_{pi}^{-1})^{-1}, \\ A_{.j} &= A_{j1} + \dots + A_{jq}, \quad \tilde{A}_{..} = (A_{1.}^{-1} + \dots + A_{p.}^{-1})^{-1}. \end{aligned}$$

Then

$$\sum \tilde{A}_{.i} \leq \tilde{A}_{..} \tag{8.5}$$

The result follows from (8.4) by considering  $A_{ij}$ ,  $i = 1, \dots, p$ , as possible values of  $A_1, \dots, A_p$ , giving uniform probabilities to index  $j$ , and taking expectation over  $j$ .

We can get a weighted version of the inequality (8.5) by attaching weights  $w_j$  to index  $j$ ,  $j = 1, \dots, q$ .

**THEOREM 8.4.** *Let  $A_1, \dots, A_n$  be  $nnd$  matrices of the same order  $p$  and  $B : p \times k$  of rank  $k$  and  $\mu(B) \subset \mu(A_i)$  for all  $i$ . Then*

$$\left(\sum (B' A_i^- B)^{-1}\right)^{-1} \geq B' \left(\sum A_i\right)^- B \tag{8.6}$$

where  $A_i^-$  denotes any  $g$ -inverse of  $A_i$ .

*Proof.* Let  $A_i = X_i' X_i$  and consider independent linear models

$$Y_i = X_i \theta + \epsilon_i, \quad E(\epsilon_i) = 0, \quad \text{and } C(\epsilon_i, \epsilon_i) = I, \quad i = 1, \dots, n. \tag{8.7}$$

The best estimate of  $B'\theta$  (in the sense of having minimum dispersion error matrix) from the entire model is

$$B' \left(\sum X_i' X_i\right)^- \sum X_i' Y_i$$

with the covariance matrix

$$B' (\Sigma A_i)^- B. \tag{8.8}$$

The estimate of  $B'\theta$  from the  $i$ -th part of the model,  $Y_i = X_i \theta + \epsilon_i$ , is

$$\hat{\psi}_i = B' (X_i' X_i)^- X_i' Y_i \tag{8.9}$$

with the covariance matrix

$$B' A_i^- B. \tag{8.10}$$

Combining the estimates (8.9) in an optimum way we have the estimate of  $B'\theta$

$$(\Sigma (B' A_i^- B)^{-1})^{-1} \Sigma (B' A_i^- B)^{-1} \hat{\psi}_i$$

with the covariance matrix

$$(\Sigma (B' A_i^- B)^{-1})^{-1}. \tag{8.11}$$

Obviously (8.11) is not smaller than (8.8) which yields the desired inequality. □

The results of Theorems 8.1 and 8.4 are similar to those of Dey *et al.* (1994).

### 9 Inequalities on the Elements of an Inverse Matrix

The following theorem is of interest in estimation theory.

THEOREM 9.1. *Let*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}$$

where  $\Sigma : (p + q) \times (p + q)$  is a pd matrix and  $\Sigma_{11}$  is of order  $p \times p$ . Then

$$\Sigma^{11} \geq (\Sigma_{11})^{-1} \tag{9.1}$$

*Proof.* Consider the rv

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q} \left( \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

where  $X$  is  $p$ -variate and  $Y$  is  $q$ -variate. Using the formulae (5.3) and (5.4) on information matrices, we have

$$I_{X,Y}(\mu) = \Sigma^{11}, \quad I_X(\mu) = \Sigma_{11}^{-1} \tag{9.2}$$

and the information inequality (5.1)

$$I_{X,Y}(\mu) \geq I_X(\mu)$$

yields the desired result. □

### 10 Carlen's Superadditivity of Fisher Information

Let  $f(u)$  be the probability density of a  $p + q$  vector variable  $u' = (u_1, \dots, u_{p+q})$  and define

$$J_f = (j_{rs}) = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

where

$$j_{rs} = E \left( \frac{\partial \log f}{\partial u_r} \cdot \frac{\partial \log f}{\partial u_s} \right)$$

and  $f_{11}$  and  $f_{22}$  are matrices of orders  $p$  and  $q$  respectively. Let  $g(u_1, \dots, u_p)$  and  $h(u_{p+1}, \dots, u_{p+q})$  be the marginal probability densities of  $u_1, \dots, u_p$  and  $u_{p+1}, \dots, u_{p+q}$  respectively with the corresponding  $J$  matrices,  $J_g$  of order  $p \times p$  and  $J_h$  of order  $q \times q$ . Then Carlen (1991) proved that

$$\text{Trace } J_f \geq \text{Trace } J_g + \text{Trace } J_h. \tag{10.1}$$

We prove the result (10.1), using Fisher information inequalities. Let us introduce a  $(p + q)$ -vector parameter

$$\theta' = (\theta'_1, \theta'_2) = (\theta_1, \dots, \theta_p, \theta_{p+1}, \dots, \theta_{p+q})$$

where  $\theta_1$  is a  $p$ -vector and  $\theta_2$  is a  $q$ -vector, and consider the probability densities

$$f(u_1 + \theta_1, \dots, u_{p+q} + \theta_{p+q}),$$

$$g(u_1 + \theta_1, \dots, u_p + \theta_p) \text{ and } h(u_{p+1} + \theta_{p+1}, \dots, u_{p+q} + \theta_{p+q}).$$

It is seen that in terms of information

$$\begin{aligned} I_f(\theta_1) &= f_{11} \geq I_g(\theta_1) = J_g \\ I_f(\theta_2) &= f_{22} \geq I_h(\theta_2) = J_h \end{aligned}$$

which shows that

$$f_{11} \geq J_g, \quad f_{22} \geq J_h.$$

Taking traces, we get the desired result (10.1). A similar proof is given by Kagan & Landsman (1997).

## 11 Inequalities on Principal Submatrices of an nnd Matrix

A  $k \times k$  principal submatrix of an nnd matrix  $A : n \times n$  is a matrix  $[A] : k \times k$  obtained by retaining columns and rows indexed by the sequence  $(i_1, \dots, i_k)$  where  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ . It is clear that if  $A \geq 0$ , then  $[A] \geq 0$ .

**THEOREM 11.1.** *If  $A$  is nnd matrix, then (i)  $[A^2] \geq [A]^2$ , and if  $A$  is pd (ii)  $[A^{-1}] \geq [A]^{-1}$ , and (iii)  $[X'] [A]^{-1} [X] \leq [X' A^{-1} X]$  where  $X$  has the same number of rows as  $A$ .*

*Proof.* All the results are derived by observing that

$$\begin{pmatrix} A & X \\ X' & X' A^{-1} X \end{pmatrix} \geq 0 \quad (11.1)$$

which implies

$$\begin{pmatrix} [A] & [X] \\ [X'] & [X' A^{-1} X] \end{pmatrix} \geq 0 \quad (11.2)$$

and taking a Schur complement for special choices of  $X$ . For (i), choose  $A^2$  for  $A$  and  $X = A$ . For (ii), choose  $X = I$ . (iii) Follows directly from (11.2). Different proofs and additional results are given by Zhang (1998).  $\square$

## References

- Albert, A. (1969). Conditions for positive and nonnegative definiteness in terms of pseudo inverses. *SIAM J. Appl. Math.*, **17**, 434–440.
- Ando, T. (1979). Concavity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra and Appl.*, **26**, 203–241.
- Ando, T. (1998). *Operator-Theoretic Methods for Matrix Inequalities*. Hokusei Gakuen University.
- Carlen, E.A. (1991). Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, **101**, 194–211.
- Dey, A., Hande, S. & Tiku, M.L. (1994). Statistical proofs of some matrix results. *Linear and Multilinear Algebra*, **38**, 109–116.
- Fiedler, M. (1961). Über eine Ungleichung für positiv definite Matrizen. *Math. Nachr.*, **23**, 197–199.
- Kagan, A. & Landsman, Z. (1997). Statistical meaning of Carlen's superadditivity of Fisher information. *Statistics and Probability Letters*, **32**, 175–179.
- Kagan, A., Landsman, Z. & Rao, C.R. (2006). Sub- and superadditivity à la Carlen of matrices related to Fisher information. To appear in *J. Statistical Planning and Inference*.
- Kagan, A. & Rao, C.R. (2003). Some properties and applications of the efficient Fisher score. *J. Statist. Planning and Inference*, **116**, 343–352.
- Kagan, A. & Rao, C.R. (2005). On estimation of a location parameter in presence of an ancillary component. *Teoriya Veroyatnostej i ee Primenenie*, **50**, 171–176.
- Kagan, A. & Smith, P.J. (2001). Multivariate normal distributions, Fisher information and matrix inequalities. *Internat. J. Math. Education in Sci. Technology*, **32**, 91–96.
- Marshall, A.W. & Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*. New York: Academic Press.

- Milne, E.A. (1925). Note on Rosseland's integral for absorption coefficient. *Monthly Notices of Royal Astronomical Society*, **85**, 979–984.
- Mitra, S.K. (1973). Statistical proofs of some propositions on nonnegative definite matrices. *Bull. Inter. Statist. Inst.*, **XLV**, 206–211.
- Olkin, I. & Pratt, J. (1958). A multivariate Tchebycheff's inequality. *Ann. Math. Stat.*, **29**, 226–234.
- Prakasa Rao, B.L.S. (1998). An inequality for the expectation of harmonic mean of random matrices. *Tech. Report*, Indian Statistical Institute, Delhi.
- Rao, C.R. (1967). Calculus of generalized inverse of matrices: Part 1 – General theory. *Sankhyā A*, **19**, 317–350.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. Second edition. New York: Wiley.
- Rao, C.R. (1996). Seven inequalities in estimation theory. *Student*, **1**, 149–158.
- Rao, C.R. (2000). Statistical proofs of some matrix inequalities. *Linear Algebra and Applic.*, **321**, 307–320.
- Rao, C.R. & Mitra, S.K. (1971). *Generalized Inverse of Matrices and its Applications*. New York: John Wiley.
- Rao, C.R. & Rao, M.B. (1998). *Matrix Algebra and its Applications to Statistics and Econometrics*. Singapore: World Scientific.
- Rao, C.R. & Toutenburg, H. (1999). *Linear Models: Least Squares and Alternatives*, (Second Edition). Springer.
- Zhang, Fuzhen (1998). Some inequalities on principal submatrices of positive semidefinite matrices. *Tech. Report*, Nova South Eastern University, Fort Lauderdale.

## Résumé

Les ouvrages sur les modèles linéaires et l'analyse multivariée incluent un chapitre sur l'algèbre matricielle, tout à fait à juste titre, car les résultats matriciels sont utilisés dans la discussion des méthodes statistiques dans ces domaines. Dans les années récentes, sont apparus de nombreux articles dans lesquels les résultats statistiques obtenus sans avoir recours aux théorèmes matriciels ont été utilisés pour démontrer des résultats matriciels qui sont eux-mêmes utilisés pour générer d'autres résultats statistiques. Cela peut avoir une valeur pédagogique. On ne veut pas dire pour autant que la connaissance préalable de la théorie matricielle n'est pas nécessaire pour étudier les statistiques. On cherche à montrer qu'un usage judicieux de résultats statistiques et matriciels pourrait aider à fournir des démonstrations élégantes de problèmes autant en statistique qu'en algèbre matricielle et rendre l'étude des deux sujets plutôt intéressante. Des notions basiques des espaces vectoriels et des matrices sont cependant nécessaires et elles sont évoquées dans l'introduction de cet article.

[Received June 2005, accepted November 2005]