Stochastic Inequalities
IMS Lecture Notes – Monograph Series
Volume 22 (1993)

A GROUP ACTION ON COVARIANCES WITH APPLICATIONS TO THE COMPARISON OF LINEAR NORMAL EXPERIMENTS[1]

By MORRIS L. EATON

University of Minnesota

Consider a linear normal experiment with a fixed regression subspace and a known covariance matrix Σ. A classical method for comparing such experiments involves the covariance matrix of the Gauss–Markov estimator of the regression coefficients, say $V(\Sigma)$. We introduce a group action on covariance matrices and show that a maximal invariant is $V(\Sigma)$. The concavity of $V(\Sigma)$ in the Loewner ordering shows that $V(\Sigma)$ is monotone in the natural group induced ordering on covariances. In addition, the group structure is used to provide an easy proof of a main theorem in the comparison of linear normal experiments. A related problem concerns the behavior of $V(\Sigma)$ as a function of the elements of Σ. Some results related to positive dependence ideas are presented via examples.

1. Introduction

In simple linear model problems, the covariance matrix of the Gauss–Markov estimator of the vector of regression coefficients is often used to choose between competing linear models with the same regression coefficients. Given an $n \times k$ design matrix X of rank k with $1 \leq k < n$ and a known non–singular covariance matrix Σ, let $\mathcal{E}(X, \Sigma)$ denote the experiment with an observation vector Y whose distribution is multivariate normal $N(X\beta, \Sigma)$ where β is the k–vector of regression coefficients. The reason for the assumption that $k < n$ is explained at the end of Section 4.

Now, the covariance matrix of $\widehat{\beta}$, the Gauss Markov estimator of β, is

$$\text{Cov}(\widehat{\beta}) = (X'\Sigma^{-1}X)^{-1}. \tag{1.1}$$

For two experiments with the same $\beta \in R^k$, say $\mathcal{E}(X_i, \Sigma_i)$, $i = 1, 2$, it is well known that experiment $\mathcal{E}(X_1, \Sigma_1)$ is sufficient for $\mathcal{E}(X_2, \Sigma_2)$ iff

$$(X_1'\Sigma_1^{-1}X_1)^{-1} \leq (X_2'\Sigma_2^{-1}X_2)^{-1} \tag{1.2}$$

[1]Work supported in part by National Science Foundation Grant DMS–89–22607.

AMS 1991 *subject classifications.* Primary 62B15; Secondary 62J05.

Key words and phrases. Comparison of experiments, positive correlation, linear normal experiments, group induced orderings.

where "$\leq$" is the standard Loewner ordering on symmetric matrices ($A \leq B$ means $B - A$ is non–negative definite). [Here, we are using sufficiency in the sense discussed in Blackwell (1951, 1953) — that is, an experiment $\mathcal{E}_1$ is sufficient for $\mathcal{E}_2$ if for every decision problem and prior distribution, the Bayes risk from $\mathcal{E}_1$ is not greater than that from $\mathcal{E}_2$.] A proof of this and related results can be found in Hansen and Torgersen (1974). A few other relevant references include Ehrenfeld (1955), Torgersen (1972), Goel and De Groot (1979), Stepniak, Wang and Wu (1984), Torgersen (1984), Shaked and Tong (1990), and Torgersen (1991).

In the simple case when $k = 1$ and X is the vector of ones in $\mathbb{R}^n$, say $X = e$, replace β by θ. Thus the data vector Y is $N(\theta e, \Sigma)$ and

$$\text{var}(\widehat{\theta}) = (e'\Sigma^{-1}e)^{-1}. \tag{1.3}$$

With the further assumption that Σ is a correlation matrix, say $\Sigma = R$, (1.3) becomes

$$\phi(R) = \text{var}(\widehat{\theta}) = (e'R^{-1}e)^{-1}. \tag{1.4}$$

When R is an intraclass correlation matrix with off diagonal elements equal to $\rho \in (-(n-1)^{-1}, 1)$ then $\phi(R) = n^{-1}[(n-1)\rho + 1]$ which is increasing in ρ. For this case, then, experiments can be ordered by sufficiency in terms of ρ. In a recent paper, Shaked and Tong (1990) studied this and other problems by comparing experiments with i.i.d. observations to those with exchangeable variables. They showed that under certain conditions, positive dependence (corresponding to $\rho > 0$ in the example above) tends to decrease information (increase $\phi(R)$). This raises the rather natural question of how $\phi(R)$ behaves as a function of R when all the elements of R are non–negative. For example, when is it true that $\phi(R) \geq \phi(I_n)$ so that an experiment with i.i.d. observations is preferred to one with correlation matrix R?

In Section 2 of this paper, we present a number of examples — all of which concern the behavior of $\phi(R)$ when the elements of R are non–negative. For $n = 3$, the examples show that for some R's with non–negative elements, perfect estimation of θ is possible — that is, $\widehat{\theta}$ has variance zero. In other cases, $\phi(R)$ first increases and then decreases as certain elements of R increase. However, when R has some special structure, such as the circular correlation structure of Olkin and Press (1969) or the special correlation structure described in Tong (1990, p. 129), $\phi(R)$ increases as certain elements of R increase. These examples show that our rather vague intuitive feeling that "positive correlation tends to decrease information content in an experiment" is very far from the truth, even for rather simple normal experiments with three observations. But, when R has some special structure, our intuition may be correct.

For a general linear normal experiment $\mathcal{E}(X, \Sigma)$, the quantity

$$\psi(\Sigma) = (X'\Sigma^{-1}X)^{-1} \tag{1.5}$$

is used to order experiments in terms of sufficiency. For X fixed, this induces a natural equivalence relationship on Σ's — namely, Σ_1 and Σ_2 are equivalent iff $\psi(\Sigma_1) = \psi(\Sigma_2)$. An alternative way to induce an equivalence relationship on Σ's is to consider $n \times n$ non-singular linear transformations A which fix the regression space (the column space of X), and then define a group action on the set $\mathcal{S}_n^+$ of all $n \times n$ positive definite Σ's. More precisely, suppose the observation vector Y satisfies

$$\mathcal{L}(Y) = N(X\beta, \Sigma), \tag{1.6}$$

where $\mathcal{L}(\bullet)$ denotes "the law of $\bullet$". Let $G(X)$ be the group of all $n \times n$ non-singular matrices which satisfy $AX = X$ (A fixes the regression subspace). In terms of sufficiency, Y and AY are equivalent since they are 1–1 transformations of each other. Further,

$$\begin{aligned} \mathcal{L}(Y) &= N(X\beta, \Sigma) \\ \mathcal{L}(AY) &= N(X\beta, A\Sigma A'). \end{aligned}$$

Thus, the group action $\Sigma \to A\Sigma A'$, $A \in G(X)$, also induces an equivalence relationship on $\mathcal{S}_n^+$ which is clearly relevant for the comparison of experiments. A main result in Section 3 shows that the two equivalence relations are the same. This is accomplished by showing that $\psi(\Sigma)$ is, in fact, a maximal invariant under the action of $G(X)$ on $\mathcal{S}_n^+$. A basic lemma which is used to prove that $\psi(\Sigma)$ is a maximal invariant is also used to give a very easy proof that (1.2) implies that $\mathcal{E}(X_1, \Sigma_1)$ is sufficient for $\mathcal{E}(X_2, \Sigma_2)$.

Finally, the action of $G(X)$ on $\mathcal{S}_n^+$ induces a natural partial ordering — namely $\Sigma_1 < \Sigma_2$ iff Σ_1 is in the convex hull of the $G(X)$ orbit of Σ_2. This type of ordering arises in a number of problems in probability and statistics — see Eaton and Perlman (1977) or Eaton (1987). However, unlike the case here, only compact groups have arisen naturally in the examples familiar to me. Because the function

$$\Sigma \to \psi(\Sigma)$$

is concave (in the Loewner ordering — see Ylvisaker (1964)), it follows immediately that ψ is decreasing in this $G(X)$-induced ordering on $\mathcal{S}_n^+$. A main result in Section 4 relates the ordering induced by $G(X)$ to the ordering induced by ψ.

2. Examples with Non-Negative Correlations

Throughout this section, R denotes an $n \times n$ correlation matrix with non-negative elements. With e denoting the vector of ones in R^n, the examples below concern the behavior of

$$\phi(R) = (e'R^{-1}e)^{-1} \tag{2.1}$$

when R is non–singular. Cases where R is singular are also of interest and will be discussed separately. When Y is $N_n(\theta e, R)$ and R is non–singular, then $\phi(R)$ is the variance of the Gauss–Markov estimator of θ. Marginally, each coordinate of Y is $N(\theta, 1)$ so it is natural to ask for conditions on R which imply that

$$\phi(R) \geq \phi(I_n). \tag{2.2}$$

In other words, when is the experiment $\mathcal{E}_1(e, I_n)$ consisting of i.i.d. observations sufficient for $\mathcal{E}_2(e, R)$? Clearly (2.2) holds for $n = 2$, since R has non–negative elements. However, the following example for $n = 3$ exhibits, what is to some, a rather counter–intuitive result.

EXAMPLE 2.1 For $n = 3$ consider

$$R = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & a \\ a & a & 1 \end{pmatrix}, \qquad 0 \leq a < 2^{-1/2}.$$

Given that a is non–negative, the condition on a is necessary and sufficient that R be positive definite. An easy calculation shows that

$$h(a) = (e'R^{-1}e)^{-1} \tag{2.3}$$

is given by

$$h(a) = \frac{1 - 2a^2}{3 - 4a}, \qquad 0 \leq a < 2^{-1/2}. \tag{2.4}$$

Differentiation shows that h is concave, increases in $[0,1/2)$ and decreases in $(1/2, 2^{-1/2})$. The maximum value of $1/2$ is for $a = 1/2$, and

$$h(2^{-1/2}) = 0. \tag{2.5}$$

In fact, when the correlation matrix is

$$R_0 = \begin{pmatrix} 1 & 0 & 2^{-1/2} \\ 0 & 1 & 2^{-1/2} \\ 2^{-1/2} & 2^{-1/2} & 1 \end{pmatrix}, \tag{2.6}$$

R_0 is singular. When

$$\mathcal{L}(Y) = N_3(\theta e, R_0), \tag{2.7}$$

let $c_3 = (1 - 2^{1/2})^{-1}$, $c_1 = c_2 = -2^{-1/2}c_3$, and let $c \in R^3$ have coordinates c_1, c_2, c_3. Then $c'Y$ is an unbiased estimator of θ which has variance zero. Thus perfect estimation of θ is possible when the covariance matrix is R_0.

This example has the following implication. Given Y_1 and Y_2 which are i.i.d. $N(\theta, 1)$, suppose we can further select a Y_3 which is marginally $N(\theta, 1)$, but which has correlation ρ with Y_1 and Y_2, $\rho \in [0, 2^{-1/2}]$. From a design point of view, some intuition suggests that $\rho = 0$ may be the preferred value of ρ for inferential problems concerning θ. However, the example shows that $\rho = 2^{-1/2}$ is the preferred value for all inferential problems since perfect estimation of θ is then possible. $\square$

REMARK 2.1 The above example is easily extended to correlation matrices of the form

$$R = \begin{pmatrix} 1 & \delta & a \\ \delta & 1 & a \\ a & a & 1 \end{pmatrix}$$

where $0 \leq \delta < 1$, $0 \leq a < 1$, and $1 + \delta - 2a^2 > 0$. In this case,

(1) for a fixed, $a \in [0, 2^{-1/2})$, the function $\delta \to (e'R^{-1}e)^{-1}$ is increasing in $\delta, \delta \in [0, 1)$

(2) for $\delta \in [0, 1)$ fixed, the function $a \to (e'Re)^{-1}$ is concave on $[0, ((1 + \delta)/2)^{1/2})$, has a maximum at a point strictly between the two endpoints, and converges to zero as a converges to the right endpoint. When $a = ((1 + \delta)/2)^{1/2}$, perfect estimation of θ is again possible. $\square$

REMARK 2.2 Extensions of Example 2.1 to higher dimensions is easy. For dimension n, let u be a fixed n–vector of length one with non–negative coordinates and let b be a real number in [0,1). Then the $(n + 1) \times (n + 1)$ correlation matrix

$$R = \begin{pmatrix} I_n & bu \\ bu' & 1 \end{pmatrix}$$

has non–negative elements and is non–singular. When $u'e \neq 1$, the behavior of

$$b \to (e'R^{-1}e)^{-1}$$

is similar to that of h defined in (1.3). When $u'e \neq 1$ and $b = 1$, then R is singular and again, perfect estimation is possible. $\square$

The following proposition gives a sufficient condition for (2.2) to hold.

PROPOSITION 2.1 *Let R be an $n \times n$ positive definite correlation matrix with non–negative entries such that $Re = ce$ for some real number c. Then*

$$(e'R^{-1}e)^{-1} \geq \frac{1}{n}. \tag{2.8}$$

PROOF Since $Re = ce$ and R has non-negative elements, $c \geq 1$. Thus,

$$(e'R^{-1}e)^{-1} = \left(\frac{e'e}{c}\right)^{-1} = \frac{c}{n} \geq \frac{1}{n}. \qquad \square$$

Examples of R's which satisfy the assumptions of Proposition 2.1 include intraclass correlation matrices with non-negative entries and the circular stationary correlation matrices (see Olkin and Press (1969)) with non-negative entries.

The final example of this section deals with a class of correlation matrices which might be called intra–inter–class correlation matrices as described in Tong (1990, p. 129). For a positive integer n, let $n_1 \geq n_2 \geq \cdots \geq n_r \geq 1$ be a partition of n — that is, each n_i is a positive integer and

$$\sum_1^r n_i = n. \tag{2.9}$$

Partition a correlation matrix R into $n_i \times n_j$ blocks, $1 \leq i, j \leq r$ and assume

$$\begin{cases} \text{(i) } R_{ii} \text{ has intraclass correlation structure with} \\ \qquad \text{correlation coefficient } \rho_2 \in [0,1) \\ \text{(ii) for } i \neq j, \text{ the block } R_{ij} \text{ has all entries equal} \\ \qquad \text{to } \rho_1 \in [0, \rho_2]. \end{cases} \tag{2.10}$$

PROPOSITION 2.2 *If R has the structure given in* (2.10), *then*

$$e'R^{-1}e = \frac{\sum_1^r n_i(1-\rho_2+(\rho_2-\rho_1)n_i)^{-1}}{1+\rho_1\sum_1^r n_i(1-\rho_2+(\rho_2-\rho_1)n_i)^{-1}} \tag{2.11}$$

PROOF This is proved in Appendix I. $\square$

Using the expression (2.11) it is easy to show that

(i) for fixed $\rho_2, e'R^{-1}e$ is decreasing in ρ_1 for $\rho_1 \in [0, \rho_2]$

(ii) for fixed $\rho_1, e'R^{-1}e$ is decreasing in ρ_2 for $\rho_2 \in (\rho_1, 1)$.

Thus, $\phi(R)$ in (2.1) is increasing in ρ_1 for $\rho_1 \in [0, \rho_2]$ and is increasing in ρ_2 for $\rho_2 \in [\rho_1, 1)$. Hence we have

COROLLARY 2.3 *A correlation matrix with the structure* (2.10) *satisfies* $\phi(R) \geq \phi(I_n)$. $\square$

REMARK 2.3 Using (2.11) it is possible to give an alternative proof of a recent result of Shaked and Tong (1992). To describe this result, fix ρ_1 and ρ_2 with $0 \leq \rho_1 < \rho_2 < 1$ and regard $\phi(R)$ as a function of the partition $\mathbf{n} =$

$(n_1, n_2, \ldots, n_r, 0, \ldots, 0)$ — an n–dimensional vector. Let $\mathbf{n}$ and $\mathbf{n}^*$ be two partitions such that $\mathbf{n}$ majorizes $\mathbf{n}^*$ (see Marshall and Olkin (1979) for the relevant definitions) and let R and R^* be correlation matrices corresponding to these two partitions (with the same ρ_1 and ρ_2). Then, the experiment based on $Y^* \sim N(\theta e, R^*)$ is sufficient for the experiment based on $Y \sim N(\theta e, R)$. The proof of this in Shaked and Tong (1992) is based on results in Torgersen (1984). However, using (2.11), a direct verification that the function ϕ is a Schur convex function of partitions $\mathbf{n}$ is not difficult. Thus if $\mathbf{n}$ majorizes $\mathbf{n}^*$, then $\phi(R^*) \leq \phi(R)$ so the Shaked and Tong result follows. □

REMARK 2.4 It is well known that for Σ for positive definite and x fixed, the function

$$\Sigma \to (x'\Sigma^{-1}x)^{-1}$$

is a concave function of Σ (see Section 4 for a more general result and further discussion). Thus, on any line segment contained in the set of positive definite matrices, the function $(x'\Sigma^{-1}x)^{-1}$ is concave on that line segment. In particular, $(x'\Sigma^{-1}x)^{-1}$ is concave in each element of Σ, as long as Σ remains positive definite. These remarks explain the concavity property in Example 2.1. □

3. A Group Action on Covariances

Consider an observation vector Y in R^n which has a $N(X\beta, \Sigma)$ distribution where the design matrix X is $n \times k$ of rank k, the known covariance matrix Σ is non–singular, and $\beta \in R^k$ is the vector of regression coefficients. As described in Section 1, such experiments can be compared via the function

$$\psi(\Sigma) = (X'\Sigma^{-1}X)^{-1} \tag{3.1}$$

which is the covariance matrix of the Gauss–Markov estimator $\widehat{\beta}$ of β. Let $G(X)$ be the group of $n \times n$ non–singular matrices A which satisfy $AX = X$. Such A's fix the elements of the regression subspace. Then $G(X)$ acts on the set $\mathcal{S}_n^+$ of $n \times n$ positive definite matrices via the group action

$$S \to ASA'. \tag{3.2}$$

Clearly, ψ in (3.1) is invariant under this group action. A main result in this section shows that ψ is a maximal invariant. To establish this, some preliminaries are needed. First, each $n \times k$ X of rank k can be written

$$X = \Gamma X_0 M \tag{3.3}$$

where Γ is an $n \times n$ orthogonal matrix, M is a $k \times k$ non-singular matrix, and X_0 is the special design matrix

$$X_0 = \begin{pmatrix} I_k \\ O \end{pmatrix}. \tag{3.4}$$

Also note that

$$X = \Gamma V X_0 \tag{3.5}$$

where

$$V = \begin{pmatrix} M & O \\ O & I_{n-k} \end{pmatrix} \tag{3.6}$$

is $n \times n$ and non-singular.

LEMMA 3.1 *Given $\Sigma \in \mathcal{S}_n^+$, there exists an $A \in G(X_0)$ such that*

$$A\Sigma A' = \begin{pmatrix} (X_0\Sigma^{-1}X_0')^{-1} & O \\ O & I_{n-k} \end{pmatrix}. \tag{3.7}$$

PROOF Partition Σ as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where Σ_{11} is $k \times k$ and Σ_{22} is $(n-k) \times (n-k)$. It is clear that $A \in G(X_0)$ iff A has the form

$$A = \begin{pmatrix} I_k & A_{12} \\ O & A_{22} \end{pmatrix}$$

where A_{22} is $(n-k) \times (n-k)$ and non-singular. Now, pick $A_{22} = \Sigma_{22}^{-1/2}$ and $A_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$. With this choice of A, some algebra yields

$$A\Sigma A' = \begin{pmatrix} \Sigma_{11\cdot 2} & O \\ O & I_{n-k} \end{pmatrix}$$

where $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. But it is well known that

$$\Sigma_{11\cdot 2} = (X_0'\Sigma^{-1}X_0)^{-1}.$$

□

COROLLARY 3.2 *The function*

$$\psi_0(\Sigma) = (X_0'\Sigma^{-1}X_0)^{-1}$$

is a maximal invariant under the action of $G(X_0)$.

PROOF The invariance of ψ_0 is obvious. If $\psi_0(\Sigma_1) = \psi_0(\Sigma_2)$, use Lemma 3.1 to find A_1 and A_2 so that

$$A_1\Sigma_1A_1' = \begin{pmatrix} \psi_0(\Sigma_1) & O \\ O & I_{n-k} \end{pmatrix} = \begin{pmatrix} \psi_0(\Sigma_2) & O \\ O & I_{n-k} \end{pmatrix} = A_2\Sigma_2A_2'.$$

Thus Σ_1 and Σ_2 are in the same orbit so ψ_0 is maximal. □

THEOREM 3.3 *A maximal invariant under the action of $G(X)$ on $\mathcal{S}_n^+$ is $\psi(\Sigma)$ in* (3.1).

PROOF The invariance of ψ is clear. If $\psi(\Sigma_1) = \psi(\Sigma_2)$, then writing X in the form (3.3) yields

$$X_0'(\Gamma'\Sigma_1\Gamma)^{-1}X_0 = X_0'(\Gamma'\Sigma_2\Gamma)^{-1}X_0. \tag{3.8}$$

From Corollary 3.2, there then exists an $A \in G(X_0)$ such that

$$A\Gamma'\Sigma_1\Gamma A' = \Gamma'\Sigma_2\Gamma. \tag{3.9}$$

Setting $B = \Gamma A\Gamma'$, B is in $G(X)$ and $B\Sigma_1 B' = \Sigma_2$. □

Using Lemma 3.1 it is fairly straightforward to give a proof from first principles that $\mathcal{E}(X_1, \Sigma_1)$ is sufficient for $\mathcal{E}(X_2, \Sigma_2)$ iff (1.2) holds. Here, we just sketch the details. Let Y_i be the data vector for $\mathcal{E}(X_i, \Sigma_i)$, $i = 1, 2$.

CLAIM 1 *Without loss of generality,* $X_1 = X_2 = X_0$. To see this, use (3.5) to find an $n \times n$ non-singular matrix C_i such that $C_iX_i = X_0$, $i = 1, 2$. Then

$$\mathcal{L}(C_iY_i) = N(X_0\beta, C_i\Sigma_iC_i'), \qquad i = 1, 2.$$

Because C_i is non-singular, the experiments $\mathcal{E}(X_i, \Sigma_i)$ and $\mathcal{E}(C_iX_i, C_i\Sigma_iC_i')$ are equivalent. Further, (1.2) holds iff

$$(X_0'(C_1\Sigma_1C_1')^{-1}X_0)^{-1} \leq (X_0'(C_2\Sigma_2C_2')^{-1}X_0)^{-1}.$$

This establishes Claim 1. □

Now take $X_1 = X_2 = X_0$ so we want to prove

THEOREM 3.4 *The experiment $\mathcal{E}(X_0, \Sigma_1)$ is sufficient for $\mathcal{E}(X_0, \Sigma_2)$ iff*

$$(X_0'\Sigma_1^{-1}X_0)^{-1} \leq (X_0'\Sigma_2^{-1}X_0)^{-1}. \tag{3.10}$$

PROOF Assume (3.10) holds. Using Lemma 3.1, find $A_i \in G(X_0)$ such that

$$\mathcal{L}(A_iY_i) = N(X_0\beta, A_i\Sigma_iA_i'), \qquad i = 1, 2$$

where

$$A_i\Sigma_iA_i' = \begin{pmatrix} (X_0'\Sigma_i^{-1}X_0)^{-1} & O \\ O & I \end{pmatrix}.$$

Since the A_i's are non-singular, the experiments $\mathcal{E}(X_0, \Sigma_i)$ and $\mathcal{E}(X_0, A_i\Sigma_iA_i)$ are equivalent. But, when (3.10) holds, we can then find a random vector Z which is $N(0, \Delta)$ and is independent of Y_1 such that for all β,

$$\mathcal{L}(A_1Y_1 + Z) = \mathcal{L}(A_2Y_2). \tag{3.11}$$

Of course,

$$\Delta = A_2\Sigma_2A_2' - A_1\Sigma_1A_1'$$

is non–negative definite. But (3.11) clearly implies that $\mathcal{E}(X_0, A_1\Sigma_1A_1')$ is sufficient for $\mathcal{E}(X_0, A_2\Sigma_2A_2')$ (see Lehmann (1959, p. 75)).

For the converse, let $T_i = A_i\Sigma_iA_i'$ so the experiment $\mathcal{E}(X_0, T_i)$ has data vector A_iY_i, $i = 1,2$. Consider the decision problem of estimating β with a loss function

$$L(a,\beta) = (a-\beta)'D(a-\beta)$$

where D is a fixed non–negative definite matrix. When β has a $N(0,\alpha I)$ prior distribution with $\alpha \in (0,\infty)$ standard calculations yield a Bayes risk for experiment $\mathcal{E}(X_0, T_i)$ of

$$\begin{aligned} r_i(\alpha) &= \alpha^2 \mathrm{tr} T_i(T_i+\alpha I)^{-1}D(T_i+\alpha I)^{-1} \\ &\quad + \alpha \mathrm{tr}[\alpha(T_i+\alpha I)^{-1} - I]D[\alpha(T_i+\alpha I)^{-1} - I]. \end{aligned}$$

Letting $\alpha \to \infty$ produces the limit

$$r_i(\infty) = \mathrm{tr}T_iD, \qquad i = 1,2.$$

When $\mathcal{E}(X_0, T_1)$ is sufficient for $\mathcal{E}(X_0, T_2)$, we then must have

$$\mathrm{tr}T_1D \le \mathrm{tr}T_2D \tag{3.12}$$

for all non–negative definite D. This is clearly equivalent to (3.10). □

REMARK 3.1 When comparing $\mathcal{E}(X_1, \Sigma_1)$ and $\mathcal{E}(X_2, \Sigma_2)$, the argument in Claim 1 above shows that it is sufficient to consider the case $X_1 = X_2 = X_0$ where X_0 is given in (3.4). For comparing $\mathcal{E}(X_0, \Sigma_1)$ and $\mathcal{E}(X_0, \Sigma_2)$, the group $G(X_0)$ is obviously relevant since $\psi(\Sigma) = (X_0\Sigma^{-1}X_0')^{-1}$ is maximal invariant and characterizes sufficiency. □

4. A Group Induced Ordering on Covariances

Again consider an experiment $\mathcal{E}(X, \Sigma)$ corresponding to a random vector Y with $\mathcal{L}(Y) = N(X\beta, \Sigma)$, $\beta \in \mathbb{R}^k$. The function

$$\psi(\Sigma) = (X'\Sigma^{-1}X)^{-1} \tag{4.1}$$

is a maximal invariant under the action of $G(X)$ on $\mathcal{S}_n^+$, and is clearly relevant for the comparison of experiments. An important property of ψ, established in Ylvisaker (1964), is that ψ is concave in the Loewner ordering on $\mathcal{S}_n^+$. That is, for $S_1, S_2 \in \mathcal{S}_n^+$ and $\alpha \in [0,1]$, the matrix

$$\psi(\alpha S_1 + (1-\alpha)S_2) - \alpha\psi(S_1) - (1-\alpha)\psi(S_2)$$

is non–negative definite. This we write as

$$\psi(\alpha S_1 + (1-\alpha)S_2) \geq \alpha\psi(S_1) + (1-\alpha)\psi(S_2). \tag{4.2}$$

To define an ordering on $\mathcal{S}_n^+$, let $C(S)$ denote the convex hull of the $G(X)$–orbit of $S \in \mathcal{S}_n^+$. In other words, $C(S)$ is the convex hull of

$$\{ASA' \mid A \in G(X)\}.$$

DEFINITION 4.1 Write $S_1 \leq_G S_2$ if $S_1 \in C(S_2)$.

Since $C(S_2)$ is the convex hull of the $G(X)$–orbit of S_2, each point in $C(S_2)$ is a finite convex combination of points in the orbit. Thus S_1 is in $C(S_2)$ iff S_1 can be represented as

$$S_1 = \sum_1^r \alpha_i A_i S_2 A_i'$$

for some integer $r \geq 1$ where $\alpha_i \geq 0$ and $\Sigma\alpha_i = 1$. In other words, $S_1 \leq_G S_2$ iff S_1 has the above representation. It is easily checked that $S_1 \leq_G S_2$ iff for any $A, B \in G(X)$,

$$AS_1A' \leq_G BS_2B'.$$

Using this, it follows that $S_1 \leq_G S_2$ iff $C(S_1) \subseteq C(S_2)$. Thus, $\leq_G$ is a pre–order in the sense described in Marshall and Olkin (1979, p. 13). Group induced orderings of this type have arisen in a number of contexts related to both probability inequalities and inequalities more generally. The classical majorization ordering is a group induced ordering, as is one version of the submajorization ordering. Some relevant references are Eaton and Perlman (1977), Marshall and Olkin (1979), Eaton (1982) and Eaton (1987). It should be noted that in all of the examples I know, except the current one, the underlying groups are compact.

The main result of this section shows that $S_1 \leq_G S_2$ iff $\psi(S_1) \geq \psi(S_2)$. Therefore, the $G(X)$ induced ordering on covariances is the same as the comparison of experiments ordering given by ψ in (4.1). The implication in one direction is easy.

THEOREM 4.1 *If $S_1 \leq_G S_2$, then $\psi(S_1) \geq \psi(S_2)$ (in the Loewner ordering).*

PROOF Since $S_1 \leq_G S_2$ we can find $A_1, \ldots, A_r$ in $G(X)$ and non–negative numbers $\alpha_1, \ldots, \alpha_r$ satisfying $\Sigma\alpha_i = 1$ such that

$$S_1 = \sum_1^r \alpha_i A_i S_2 A_i'. \tag{4.3}$$

Using the concavity of ψ in the Loewner ordering, we have

$$\psi(S_1) = \psi\left(\sum_1^r \alpha_i A_i S_2 A_i'\right) \geq \sum_1^r \alpha_i \psi(A_i S_2 A_i') = \psi(S_2).$$

The final equality follows from the $G(X)$–invariance of ψ. □

For the converse, we first establish a special case.

LEMMA 4.2 *Consider the ordering* $\leq_G$ *induced on* $\mathcal{S}_n^+$ *by* $G(X_0)$. *If*

$$(X_0' S_2^{-1} X_0)^{-1} \leq (X_0' S_1^{-1} X_0)^{-1} \tag{4.4}$$

then there exists a discrete probability measure μ *on* $G(X_0)$ *such that*

$$S_1 = \int A S_2 A' \mu(dA). \tag{4.5}$$

PROOF Let

$$T_i = (X_0' S_i^{-1} X_0)^{-1}, \qquad i = 1, 2.$$

From Lemma 3.1, there exists $A_i \in G(X_0)$ such that

$$A_i S_i A_i' = \begin{pmatrix} T_i & O \\ O & I_{n-k} \end{pmatrix}, \qquad i = 1, 2.$$

By assumption, $\Delta = T_1 - T_2$ is non–negative definite. Write $\Delta = \Sigma_1^k v_i v_i'$ where $v_i, \ldots, v_k$ are vectors in R^k. Fix $u_0 \in R^{n-k}$ such that $u_0' u_0 = 1$. Let B be the random $k \times (n-k)$ matrix which takes on the values $\pm(2k)^{1/2} v_i u_0'$ with probabilities $1/2k$. Denote by μ_0 the distribution (on $G(X_0)$) of the random matrix

$$A = \begin{pmatrix} I_k & B \\ O & I_{n-k} \end{pmatrix} \in G(X_0).$$

Because $\mathcal{E}B = 0$ and $\mathcal{E}BB' = \Delta$, it is easy to verify that

$$\mathcal{E}_{\mu_0} A(A_2 S_2 A_2') A' = A_1 S_1 A_1'. \tag{4.6}$$

Setting $\tilde{A} = A_1^{-1} A A_2$, let μ denote the distribution of $\tilde{A}$ on $G(X_0)$. Then (4.6) can be written

$$\mathcal{E}_\mu \tilde{A} S_2 \tilde{A}' = S_1 \tag{4.7}$$

which is just (4.5). □

Of course, (4.5) is just the assertion that S_1 is in $C(S_2)$ — that is, $S_1 \leq_G S_2$ when $G = G(X_0)$. The general case is now easy.

THEOREM 4.3 *Consider the ordering* $\leq_G$ *induced on* $\mathcal{S}_n^+$ *by* $G(X)$. *If*

$$(X' S_2^{-1} X)^{-1} \leq (X' S_1^{-1} X)^{-1}, \tag{4.8}$$

then $S_1 \leq_G S_2$.

PROOF Using (3.3), write $X = \Gamma X_0 M$ so (4.8) is equivalent to

$$(X_0'(\Gamma' S_2 \Gamma)^{-1} X_0)^{-1} \leq (X_0'(\Gamma' S_1 \Gamma)^{-1} X_0)^{-1}. \tag{4.9}$$

From Lemma (4.2), $\Gamma' S_1 \Gamma$ is in the convex hull of the $G(X_0)$–orbit of $\Gamma' S_2 \Gamma$. Thus, we can write

$$\Gamma' S_1 \Gamma = \Sigma \alpha_i A_i (\Gamma' S_2 \Gamma) A_i'$$

where $A_i \in G(X_0)$, $\alpha_i \geq 0$ and $\Sigma \alpha_i = 1$. Therefore

$$S_1 = \Sigma \alpha_i (\Gamma A_i \Gamma') S_2 (\Gamma A_i \Gamma')'.$$

Since $A_i \in G(X_0)$, $\Gamma A_i \Gamma' \in G(X)$ so S_1 is in the $G(X)$–orbit of S_2. Thus $S_1 \leq_G S_2$. □

Now, consider experiments $\mathcal{E}(X_1, \Sigma_1)$ and $\mathcal{E}(X_2, \Sigma_2)$. To compare these experiments, we can take $X_1 = X_2 = X_0$ (see Remark 3.1). The results in this section show that $\mathcal{E}(X_0, \Sigma_1)$ is sufficient for $\mathcal{E}(X_0, \Sigma_2)$ iff $\Sigma_2 \leq_G \Sigma_1$ where $G = G(X_0)$. Note that this result is not correct when $k = n$ since in this case $X_0 = I_n$ and $G = \{I_n\}$; so the convex hull of the G–orbit of Σ_1 is just $\{\Sigma_1\}$. However, the characterization (1.2) of sufficiency does continue to hold when $X_1 = X_2 = X_0$ and $k = n$. Thus the assumption that $k < n$ is necessary.

Appendix I

In this appendix, we establish a result which verifies equation (2.11) of Proposition 2.2. Let $x_1, \ldots, x_r$ be vectors in R^n which satisfy

(i) $x_i' x_j = 0 \quad$ if $i \neq j$

(ii) $x_i' x_i = a_i^2 \quad$ with $a_i > 0$.

Also, let $x = \Sigma_1^r x_i$. For $\alpha \geq 0$ and $\beta \geq 0$, consider the matrix

$$A = I + \alpha x x' + \beta \sum_1^r x_i x_i'. \tag{A.1}$$

The following proposition gives a formula for $x' A^{-1} x$.

PROPOSITION A.1 *With x and A as defined above,*

$$x' A^{-1} x = \frac{\sum_1^r a_i^2 (1 + \beta a_i^2)^{-1}}{1 + \alpha \sum_1^r a_i^2 (1 + \beta a_i^2)^{-1}}. \tag{A.2}$$

PROOF Without loss of generality (just make an orthogonal transformation of coordinates), we can assume that $n = r$ and $x_i = a_i \epsilon_i$ where ϵ_i is the i^{th}

standard basis vector in R^r. Under this assumption, $\sum_1^r x_i x_i'$ is a diagonal matrix D with diagonal entries a_i^2, $i = 1, \ldots, r$. Further

$$x = \sum_1^r x_i = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} = a \in R^r$$

and

$$A = I + \alpha aa' + \beta D.$$

Therefore,

$$x'A^{-1}x = a'(I + \beta D + \alpha aa')^{-1}a = v'(I + \alpha vv')^{-1}v$$

where v is given by

$$v = (I + \beta D)^{-1/2}a.$$

But, it is easy to show that

$$v'(I + \alpha vv')^{-1}v = \frac{vv'}{1 + \alpha vv'}. \tag{A.3}$$

From this, we have

$$x'A^{-1}x = \frac{a'(I + \beta D)^{-1}a}{1 + \alpha a'(I + \beta)^{-1}a}$$

from which the result follows by noting that

$$a'(I + \beta D)^{-1}a = \sum_1^r a_i^2(1 + \beta a_i^2)^{-1}. \qquad \square$$

To apply this result to correlation matrices R satisfying (2.10), first let $n_1 \geq \cdots \geq n_r \geq 1$ be a partition of n. Then, let $e^{(i)} \in R^n$ be the vector whose first $n_1 + \cdots + n_{i-1}$ coordinates are zero, whose next n_i coordinates are one, and whose remaining coordinates are zero. Then it is clear that

$$\begin{cases} \text{(i)} \;\; e^{(i)\prime}e^{(j)} = 0 \text{ if } i \neq j \\ \text{(ii)} \;\; e^{(i)\prime}e^{(i)} = n_i \\ \text{(iii)} \;\; \sum_1^r e^{(i)} = e \end{cases} \tag{A.4}$$

where e is the vector of ones in R^n. An easy calculation shows that

$$R = (1 - \rho_2)I + \rho_1 ee' + (\rho_2 - \rho_1)\sum_1^r e^{(i)}e^{(i)\prime}.$$

A direct application of Proposition A.1 and a bit of algebra yields Proposition 2.2.

References

BLACKWELL, D. (1951). Comparison of experiments. In *Proc. Second Berkeley Symp. Math. Statist. Probab.*, J. Neyman, ed., University of California Press, Berkeley, CA. 93–102.

BLACKWELL, D. (1953). Equivalent comparison of experiments. *Ann. Math. Statist.* **24** 265–272.

EATON, M. L. (1982). A review of selected topics in probability inequalities. *Ann. Statist.* **10** 11–43.

EATON, M. L. (1987). *Lectures on Topics in Probability Inequalities.* Centrum voor Wiskunde en Informatica, Amsterdam.

EATON, M. L. AND PERLMAN, M. (1977). Reflection groups, generalized Schur functions and the geometry of majorization. *Ann. Probab.* **5** 829–860.

EHRENFELD, S. (1955). Complete class theorem in experimental design. In *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, J. Neyman, ed., University of California Press, Berkeley, CA. 69–75.

GOEL, P. K. AND DE GROOT, M. H. (1979). Comparison of experiments and information measures. *Ann. Statist.* **7** 1066–1077.

HANSEN, O. H. AND TORGERSEN, E. N. (1974). Comparison of linear normal experiments. *Ann. Statist.* **2** 367–373.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

MARSHALL, A. W. AND OLKIN, I. (1979). *Inequalities. Theory of Majorization and its Applications*, Academic Press, New York.

OLKIN, I. AND PRESS, S. J. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Statist.* **40** 1358–1373.

SHAKED, M. AND TONG, Y. L. (1990). Comparison of experiments for a class of positively dependent random variables. *Canadian J. Statist.* **18** 79–86.

SHAKED, M. AND TONG, Y. L. (1992). Comparison of experiments via dependence of normal variables with a common marginal distribution. *Ann. Statist.* **20** 614–618.

STEPNIAK, C., WANG, S.-G. AND WU, C. F. J. (1984). Comparison of linear experiments with known covariances. *Ann. Statist.* **12** 358–365.

TONG, Y. L. (1990). *The Multivariate Normal Distribution.* Springer–Verlag, New York.

TORGERSEN, E. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383–1399.

TORGERSEN, E. N. (1984). Orderings of linear models. *J. Statist. Plan. Inf.* **9** 1–17.

TORGERSEN, E. N. (1991). *Comparison of Experiments.* Cambridge University Press, Cambridge.

YLVISAKER, D. (1964). Lower bounds for minimum covariance matrices in time series regression problems. *Ann. Math. Statist.* **35** 362–368.

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455