The Annals of Statistics
1974, Vol. 2, No. 2, 367-373

COMPARISON OF LINEAR NORMAL EXPERIMENTS

BY OLE HÅVARD HANSEN AND ERIK N. TORGERSEN

Mφre and Romsdale Regional College, Molde and University of Oslo

Consider independent and normally distributed random variables $X_1, \cdots, X_n$ such that $0 < \text{Var } X_i = \sigma^2$; $i = 1, \cdots, n$ and $E(X_1, \cdots, X_n)' = A'\beta$ where A' is a known $n \times k$ matrix and $\beta = (\beta_1, \cdots, \beta_k)'$ is an unknown column matrix. (The prime denotes transposition.) The cases of known and totally unknown σ^2 are considered simultaneously. Denote the experiment obtained by observing $X_1, \cdots, X_n$ by $\mathscr{E}_A$. Let A and B be matrices of, respectively, dimensions $n_A \times k$ and $n_B \times k$. Then, if σ^2 is known, (if σ^2 is unknown) $\mathscr{E}_A$ is more informative than $\mathscr{E}_B$ if and only if $AA' - BB'$ is nonnegative definite (and $n_A \geqq n_B + \text{rank}(AA' - BB')$).

1. Introduction, notations and basic facts. A notion of "being more informative" for experiments was introduced by Bohnenblust, Shapley and Sherman and may be found in Blackwell [1]. This was generalized by LeCam in [4] (see also, Heyer [3]) to the notion of ε-deficiency.

An experiment will here be defined as a pair $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta : \theta \in \Theta))$ where $(\chi, \mathscr{A})$ is a measurable space and $(P_\theta : \theta \in \Theta)$ is a family of probability measures on $(\chi, \mathscr{A})$. $(\chi, \mathscr{A})$ is the sample space of $\mathscr{E}$ and Θ is the parameter set of $\mathscr{E}$.

DEFINITION. Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta : \theta \in \Theta))$ and $\mathscr{F} = ((\mathscr{Y}, \mathscr{B}), (Q_\theta : \theta \in \Theta))$ be two experiments with the same parameter set Θ.

Then we shall say that $\mathscr{E}$ is more informative than $\mathscr{F}$, if to each decision space $(D, \mathscr{T})$ (i.e. a measurable space) where $\mathscr{T}$ is finite, every bounded loss function $(\theta, d) \rightsquigarrow W_\theta(d)$ on $\Theta \times D$ (W_θ is assumed measurable for each θ) and every risk function r obtainable in $\mathscr{F}$ there is a risk function r' obtainable in $\mathscr{E}$ so that

$$r'(\theta) \leqq r(\theta), \qquad \theta \in \Theta.$$

If $\mathscr{E}$ is more informative than $\mathscr{F}$ then we will write this $\mathscr{E} \geqq \mathscr{F}$.

Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta \in \Theta))$ and $\mathscr{F} = ((\mathscr{Y}, \mathscr{B}), (Q_\theta \in \Theta))$ be two experiments such that $(P_\theta : \theta \in \Theta)$ is dominated, $\mathscr{Y}$ is a Borel subset of a complete separable metric space and $\mathscr{B}$ is the class of Borel subsets of $\mathscr{Y}$. Then it follows from Theorem 3 in LeCam's paper [4] (see Section 2 in [5]) that $\mathscr{E} \geqq \mathscr{F}$ if and only if there is a Markov kernel M from $(\chi, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$ so that

$$P_\theta M = Q_\theta : \qquad \theta \in \Theta.$$

Suppose $P_\theta M = Q_\theta$; $\theta \in \Theta$ for some Markov kernel M and consider a real-valued function g on Θ. Let X and Y be UMVU estimators of g in, respectively,

Received February 1973; revised April 1973.

AMS 1970 *subject classifications.* Primary 62B15; Secondary 62K99.

Key words and phrases. Informational inequality, invariant kernels, normal models.

$\mathscr{E}$ and $\mathscr{F}$. Then it follows from the Rao-Blackwell theorem that

$$\operatorname{Var}_\theta X \leqq \operatorname{Var}_\theta Y\,; \qquad \theta \in \Theta\,.$$

Let G be a group acting on χ, $\mathscr{Y}$ and Θ so that

$$P_\theta g^{-1} = P_{g(\theta)}\,; \qquad \theta \in \Theta$$

and

$$Q_\theta g^{-1} = Q_{g(\theta)}\,; \qquad \theta \in \Theta\,.$$

Call a Markov kernel M from $(\chi, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$ invariant if:

$$M(g[B]\,|\,g(x)) = M(B\,|\,x)\,; \qquad B \in \mathscr{B},\ x \in \chi\,.$$

Then ([2] and [4]) there is an invariant Markov kernel so that $P_\theta M = Q_\theta$; $\theta \in \Theta$ provided there is at least one Markov kernel with this property and provided certain regularity conditions are satisfied. An exposition of sufficient generality for the applications to be made here may be found in Section 2 in [6].

Consider now independent and normally distributed random variables $X_1, \cdots, X_n$ such that $0 < \operatorname{Var} X_i = \sigma^2$; $i = 1, \cdots, n$ and $E(X_1, \cdots, X_n)' = A'\beta$ where A' is a known $n \times k$ matrix and $\beta = (\beta_1, \cdots, \beta_k)'$ is an unknown column matrix. (The prime denotes transposition.) We shall simultaneously treat the cases of known and totally unknown σ^2. The experiment obtained by observing X will be denoted by $\mathscr{E}_A$.

The purpose of this paper is to present a simple criterion for the informational inequality, $\mathscr{E}_A \geqq \mathscr{E}_B$, when A and B are matrices with the same number of rows.

Our point of departure was the following result of C. Boll [2]:

Let $j = 1, 2, \cdots$ and $c \geqq 0$ be given constants and consider the experiment $\mathscr{F}_{c,j}$ of observing independent random variables Z and W such that Z is $N(\zeta, (1 + c)\sigma^2)$ distributed and W/σ^2 is χ_j^2 distributed. It is assumed that ζ and $\sigma^2 > 0$ are totally unknown. Then Boll proved that $\mathscr{F}_{c,j} \geqq \mathscr{F}_{0,k}$ if and only if $c > 0$ and $j \geqq k + 1$, or, $c = 0$ and $j \geqq k$.

2. Comparison of "reduced" linear normal models. In this section we shall treat simultaneously the case of known and unknown variance σ^2. In the first case our parameter set Θ is $]-\infty, \infty[^k$ for some positive integer k. In the last case $\Theta = .]-\infty, \infty[^k \times]0, \infty[$ for some nonnegative integer k.

Consider two experiments $\mathscr{E}$ and $\mathscr{F}$ defined as follows:

$\mathscr{E}$ is the experiment obtained by observing $k + p$ independent normally distributed random variables $X_1, \cdots, X_{k+p}$ such that $\operatorname{Var} X_i = \sigma^2$; $i = 1, \cdots, k + p$. $EX_i = \beta_i$; $i = 1, \cdots, k$ and $EX_i = 0$; $i = k + 1, \cdots, k + p$.

$\mathscr{F}$ is the experiment obtained by observing $l + q$ independent normally distributed random variables $Y_1, \cdots, Y_{l+q}$ such that $\operatorname{Var} Y_i = \sigma^2$; $i = 1, \cdots, l + q$, $EY_i = c_i\beta_i$; $i = 1, \cdots, l$ and $EY_i = 0$; $i = l + 1, \cdots, l + q$.

Here $c_1, \cdots, c_l$ are known constants and we shall assume that $k \geqq l$, p and q are given nonnegative integers. The unknown parameter is $(\beta_1, \cdots, \beta_k)$ when σ^2 is known and it is $(\beta_1, \cdots, \beta_k, \sigma^2)$ when σ^2 is unknown. If σ^2 is known then—

by sufficiency—$X_{k+1}, \cdots, X_{k+p}$ may be deleted from $\mathscr{E}$ and $Y_{l+1}, \cdots, Y_{l+q}$ may be deleted from $\mathscr{F}$. If σ^2 is unknown then—by sufficiency—$X_{k+1}, \cdots, X_{k+p}$ may be replaced by $S = X_{k+1}^2 + \cdots + X_{k+p}^2$ and $Y_{l+1}, \cdots, Y_{l+q}$ by $T = Y_{l+1}^2 + \cdots + Y_{l+q}^2$. In order to avoid trivialities we will assume that $l \geqq 1$ when σ^2 is known and assume that $k + p,\ l + q \geqq 1$ when σ^2 is unknown.

If $p \geqq 1$, $q \geqq 1$, $k = l = 1$ and σ^2 is unknown, then Boll [2] has shown that $\mathscr{E} \geqq \mathscr{F}$ if and only if either $p \geqq q$ and $|c_1| = 1$ or $p \geqq q + 1$ and $|c_1| < 1$. Boll's criterion generalizes as follows:

PROPOSITION 2.1. *If σ^2 is known then $\mathscr{E} \geqq \mathscr{F}$ if and only if $|c_i| \leqq 1$; $i = 1, \cdots, l$. If σ^2 is unknown then $\mathscr{E} \geqq \mathscr{F}$ if and only if $|c_i| \leqq 1$; $i = 1, \cdots, l$ and $p \geqq q + \#\{i : 1 \leqq i \leqq l, |c_i| < 1\}$.*

REMARK. The "invariance" part of the proof below is very similar to that of Boll and the proof might—as Boll's was—have been completed by considering Laplace transforms. We will here, however, replace the "Laplace transform" part of the proof by a comparison of unbiased estimators of σ^2.

PROOF OF THE PROPOSITION. Let $\tilde{\mathscr{E}}$ denote the experiment obtained from $\mathscr{E}$ by deleting $X_i : l < i \leqq k$. Clearly $\tilde{\mathscr{E}} \geqq \mathscr{F} \Rightarrow \mathscr{E} \geqq \mathscr{F}$, and that the converse holds may be seen by letting the additive group R^{k-l} act as follows:

$$g = (g_{l+1}, \cdots, g_k) : (\beta_1, \cdots, \beta_k, \sigma^2) \to (\beta_1, \cdots, \beta_l, \beta_{l+1} + g_{l+1}, \cdots, \beta_k + g_k, \sigma^2) ;$$

$$(X_1, \cdots, X_{k+p}) \to (X_1, \cdots, X_l, X_{l+1} + g_{l+1}, \cdots, X_k + g_k, X_{k+1}, \cdots, X_{k+p})$$

and $(Y_1, \cdots, Y_{l+q}) \to (Y_1, \cdots, Y_{l+q})$. Obviously $\mathscr{E}$ and $\mathscr{F}$ are both invariant under this group and any invariant kernel from $\mathscr{E}$ to $\mathscr{F}$ does not depend on $X_{l+1}, \cdots, X_k$. It follows that we may—without loss of generality—assume that $l = k$.

Consider now first the case of known variance σ^2. Let $c_i \neq 0$. Then the UMVU estimator of β_i in $\mathscr{E}$ is X_i while the UMVU estimator for β_i in $\mathscr{F}$ is Y_i/c_i. The variances of these estimators are, respectively, σ^2 and σ^2/c_i^2. Hence $\mathscr{E} \geqq \mathscr{F} \Rightarrow |c_i| \leqq 1$; $i = 1, \cdots, k$. Conversely, assume $|c_i| \leqq 1$; $i = 1, \cdots, k$. Let $Z_1, \cdots, Z_k$ be independently and normally distributed random variables such that: $(Z_1, \cdots, Z_k)$ is independent of $(X_1, \cdots, X_k)$, $EZ_i = 0$; $i = 1, \cdots, k$ and $\operatorname{Var} Z_i = (1 - c_i^2)\sigma^2$; $i = 1, \cdots, k$. Then $Z_i + c_i X_i$ has the same distribution as $Y_1, \cdots, Y_k$.

Suppose next that σ^2 is unknown and that $\mathscr{E} \geqq \mathscr{F}$. By the result proved above: $|c_i| \leqq 1$; $i = 1, \cdots, k$. We will demonstrate that $p \geqq q + \#\{i : 1 \leqq i \leqq k, |c_i| < 1\}$.

Let G be the group of transformations of Θ of the form: $(\beta_1, \cdots, \beta_k, \sigma^2) \to (g_1 + g\beta_1, \cdots, g_k + g\beta_k, g^2\sigma^2)$ where $g_1, \cdots, g_k$ and g are constants. If $k = (g_1, \cdots, g_k, g) \in G$ then we let it move $(X_1, \cdots, X_k, S)$ to $(g_1 + gX_1, \cdots, g_k + gX_k, g^2S)$ and $(Y_1, \cdots, Y_k, T)$ to $(g_1 + gY_1, \cdots, g_k + gY_k, g^2T)$. It may then be checked that $\mathscr{E}$ and $\mathscr{F}$ are both invariant under G. Moreover, since G has an invariant mean we may restrict attention to invariant kernels (see Section 2 in [6]). It follows that we may assume that $(X_1, \cdots, X_k, S)$, $(Y_1, \cdots, Y_k, T)$ has a joint distribution where the conditional distribution, M, of $(Y_1, \cdots, Y_k, T)$ given $(X_1, \cdots, X_k, S)$ satisfies

$$(*) \quad M(B_1 \times \cdots \times B_k \times B|_{x_1,\cdots,x_k,s}) \\ = M((g_1 c_1 + gB_1) \times \cdots \times (g_k c_k + gB_k) \times g^2 B|_{g_1+gx_1,\cdots,g_k+gx_k,g^2 s}) \,.$$

Suppose first that $p = 0$. Then σ^2 is not estimable in $\mathscr{E}$. Hence $q = 0$. By inserting $g_i = X_i - gX_i$; $i = 1, \cdots, k$ in the identity (*) we get:

$$\begin{aligned} P(Y_1 \in B_1, \cdots, Y_k \in B_k|_{x_1,\cdots,x_k}) &= P(Y_1 \in g(B_1 - c_1 X_1) + c_1 X_1, \cdots, Y_k \in g(B_k - c_k X_k) + c_k X_k|_{x_1,\cdots,x_k}) \\ &= P\Big(\frac{1}{g}(Y_1 - c_1 X_1) + c_1 X_1 \in B_1, \cdots, \\ &\qquad \frac{1}{g}(Y_k - c_k X_k) + c_k X_k \in B_k|_{x_1,\cdots,x_k}\Big) \,. \end{aligned}$$

It follows—by letting $g \to \infty$—that the conditional distribution of $(Y_1, \cdots, Y_k)$ given $X_1, \cdots, X_k$ is the one point distribution at $(c_1 X_1, \cdots, c_k X_k)$, i.e. we may as well assume $Y_i = c_i X_i$; $i = 1, \cdots, k$. Hence $\sigma^2 = \operatorname{Var} Y_i = c_i^2 \operatorname{Var} X_i = c_i^2 \sigma^2$, so that $c_i^2 = 1$; $i = 1, \cdots, k$. This proves the desired inequality when $p = 0$.

Suppose next that $p \geqq 1$ and put $U_i = (Y_i - c_i X_i)/S^{\frac{1}{2}}$; $i = 1, \cdots, k$ and $U = T/S$. It follows from (*) that $(U_1, \cdots, U_k, U)$ is independent of $(X_1, \cdots, X_k, S)$. Writing $Y_i = c_i X_i + S^{\frac{1}{2}} U_i$ we see that $S^{\frac{1}{2}} U_i$ is $N(0, (1 - c_i^2)\sigma^2)$ distributed. Furthermore:

$$\begin{aligned} E \exp[\textstyle\sum_{j=1}^k it_j c_j X_j][\prod_{j=1}^k E \exp(it_j S^{\frac{1}{2}} U_j)] Ee^{itSU} &= [\textstyle\prod_{j=1}^k E \exp(it_j c_j X_j) E \exp(it_j S^{\frac{1}{2}} U_j)] \cdot Ee^{itSU} \\ &= [\textstyle\prod_{j=1}^k E \exp(it_j Y_j)] \cdot Ee^{itT} \\ &= E[\textstyle\prod_{j=1}^k \exp(it_j Y_j) \cdot e^{itT}] \\ &= E[\textstyle\prod_{j=1}^k \exp(it_j c_j X_j)] E[\prod_{j=1}^k \exp(it_j S^{\frac{1}{2}} U_j)] Ee^{itSU} \,. \end{aligned}$$

It follows that $S^{\frac{1}{2}} U_1, \cdots S^{\frac{1}{2}} U, SU$ are independent. Hence:

$$\sigma^{-2}[\textstyle\sum_{i:|c_i|<1} (SU_i^2/(1 - c_i^2)) + SU]$$

has a χ^2 distribution with $q + \#\{i : |c_i| < 1\}$ degrees of freedom. This yields an unbiased randomized estimator of σ^2 based on S with variance $[q + \#\{i : |c_i| < 1\}]^{-1} 2\sigma^4$. The UMVU estimator based on S has variance $p^{-1} 2\sigma^4$. Hence $p \geqq q + \#\{i : |c_i| < 1\}$.

Finally, suppose σ^2 is unknown, that $|c_i| \leqq 1$; $i = 1, \cdots, k$ and that $p \geqq q + \#\{i : |c_i| < 1\}$. Write $\{i : |c_i| < 1\} = \{i_1, \cdots, i_m\}$. Then $\mathscr{F}$ may be constructed on the basis of $X_1, \cdots, X_{k+p}$ by putting:

$$\begin{aligned} Y_i &= c_i X_i \qquad \text{if} \quad |c_i| = 1 \,, \\ Y_{i_r} &= c_{i_r} X_{i_r} + (1 - c_{i_r}^2)^{\frac{1}{2}} X_{k+r} \,, \qquad r = 1, \cdots, m \end{aligned}$$

and

$$Y_{k+j} = X_{k+m+j}\,; \qquad j = 1, \cdots, q\,.$$

Hence $\mathscr{E} \geqq \mathscr{F}$.

3. Comparison of linear normal models. For each given $n_A \times k$ matrix A' let $\mathscr{E}_A$ be the experiment obtained by observing n_A independent and normally distributed random variables such that: $\operatorname{Var} X_i = \sigma^2$; $i = 1, \cdots, k$ and $E(X_1, \cdots, X_k)' = A'(\beta_1, \cdots, \beta_k)'$. The parameter set Θ is $]-\infty, \infty[^k$ if σ^2 is known and it is $]-\infty, \infty[^k \times]0, \infty[$ if σ^2 is unknown. The basic criterion for "being more informative" within this class of experiments is:

THEOREM 3.1. *Let A' and B' be given matrices of, respectively, dimension $n_A \times k$ and $n_B \times k$.*

If σ^2 is known then:[1]

$$\mathscr{E}_A \geqq \mathscr{E}_B \Leftrightarrow AA' \geqq BB'\,.$$

If σ^2 is unknown then:

$$\mathscr{E}_A \geqq \mathscr{E}_B \Leftrightarrow AA' \geqq BB' \quad \textit{and} \quad n_A \geqq n_B + \operatorname{rank}(AA' - BB')\,.$$

REMARK. Suppose σ^2 is unknown. It follows that $\mathscr{E}_A$ and $\mathscr{E}_B$ are comparable if and only if $|n_B - n_A| \geqq \operatorname{rank}(AA' - BB')$ and $AA' \geqq BB'$ or $\leqq BB'$ as $n_A \geqq n_B$ or $\leqq n_B$.

As a particular case consider the experiment $\mathscr{E}_A$ consisting in observing m independent $N(\xi, \sigma^2/2)$ variables, and the experiment $\mathscr{E}_B$ consisting in observing n independent $N(\xi, \sigma^2)$ variables. Here ξ and σ^2 are unknown parameters. Then $\mathscr{E}_A \geqq \mathscr{E}_B$ if and only if $m \geqq n + 1$, while $\mathscr{E}_B \geqq \mathscr{E}_A$ if and only if $n \geqq 2m$.

PROOF. Let I denote the $k \times k$ identity matrix. $\mathscr{E}_A$ may be considered as the experiment of observing a $N(A'\beta, \sigma^2 I)$ distributed $k \times 1$ column matrix X while $\mathscr{E}_B$ may be considered as the experiment of observing a $N(B'\beta, \sigma^2 I)$ distributed $k \times 1$ column matrix Y.

(i) *$AA' = I$ and $BB' = \Delta$, a diagonal matrix.* Then $\hat{\beta} = AX$ and $\|X - A'\hat{\beta}\|^2$ are independent variables which, together, constitute a sufficient statistic in $\mathscr{E}_A$. $\hat{\beta}$ is $N(\beta, \sigma^2 I)$ distributed and $\|X - A'\hat{\beta}\|^2/\sigma^2$ is $\chi^2_{n_A - k}$ distributed.

Let $\{i_1, \cdots, i_w\} = \{i : \Delta_i \neq 0\}$ where $w = \operatorname{rank} B$ and let β^* be any solution of the normal equations in $\mathscr{E}_B$. Then $(\Delta_{i_r})^{\frac{1}{2}}\beta^*_{i_r}$; $r = 1, \cdots, w$ and $\|Y - B'\beta^*\|^2/\sigma^2$ are independent random variables which, together, constitute a sufficient statistic in $\mathscr{E}_B$. $(\Delta_{i_r})^{\frac{1}{2}}\beta^*_{i_r}$ is $N((\Delta_{i_r})^{\frac{1}{2}}\beta_{i_r}, \sigma^2)$ distributed, $r = 1, \cdots, w$, while $\|Y - B'\beta^*\|^2/\sigma^2$ has a χ^2 distribution with $n_B - w$ degrees of freedom.

We are now within the framework of Proposition 2.1 and the proof is—in this case—completed by noting that $AA' \geqq BB'$ if and only if $|\Delta_{i_r}| \leqq 1$, $r = 1, \cdots, w$ and that $\operatorname{rank}(AA' - BB') = k - \operatorname{rank} B + \#\{r : \Delta_{i_r} \neq 1\}$.

[1] If M_1 and M_2 are symmetric matrices of the same dimension, then "$M_1 \geqq M_2$" is short for: "$M_1 - M_2$ is nonnegative definite."

(ii) *Rank* $A = k$. By a well-known result on simultaneous reduction of two quadratic forms there is a (nonsingular) $k \times k$ matrix F so that

$$F'AA'F = I \quad \text{and} \quad F'BB'F \quad \text{is a diagonal matrix.}$$

Put $\tilde{A} = F'A$ and $\tilde{B} = F'B$. It is easily seen—by reparametrization—that

$$\mathscr{E}_A \geqq \mathscr{E}_B \Leftrightarrow \mathscr{E}_{\tilde{A}} \geqq \mathscr{E}_{\tilde{B}} .$$

The theorem follows—in this case—since $\mathscr{E}_{\tilde{A}}$ and $\mathscr{E}_{\tilde{B}}$ satisfy the condition in (i) and since $\tilde{A}\tilde{A}' \geq \tilde{B}\tilde{B}' \Leftrightarrow AA' \geq BB'$ and rank $(\tilde{A}\tilde{A}' - \tilde{B}\tilde{B}')$ = rank $(AA' - BB')$.

(iii) *The general case.* It follows from the estimability criterion for linear functions of β that row $[B'] \subseteq [A']$ provided $\mathscr{E}_A \geqq \mathscr{E}_B$. Suppose now that $AA' \geqq BB'$ and that x is orthogonal on row $[A']$. Then $A'x = 0$ so that $0 = x'AA'x \geqq x'BB'x$. Hence $B'x = 0$ so that x is orthogonal on row $[B']$. It follows again that row $[B'] \subseteq$ row $[A']$. Hence we may, without loss of generality, assume that row $[B'] \subseteq$ row $[A']$. Write $A' = (a_1', \cdots, a_{n_A}')'$ and $B' = (b_1', \cdots, b_{n_B}')'$ where $a_1', \cdots, a_{n_A}'$ and $b_1', \cdots, b_{n_B}'$ are, respectively, the row vectors of A' and B'. Let $v_1', \cdots, v_r'$ be a basis in row $[A']$ and write $p_i = v_i'\beta$; $i = 1, \cdots, r$. Define matrices $S = \{s_{ij}\}$ and $T = \{t_{ij}\}$ by: $a_i' = \sum_{j=1}^{r} s_{ij} v_j'$ and $b_i' = \sum t_{ij} v_j'$. Then $A'\beta = S'p$, $B'\beta = T'p$ and S' has r = rank S' columns. It is not difficult to check that $\mathscr{E}_A \geqq \mathscr{E}_B \Leftrightarrow \mathscr{E}_S \geqq \mathscr{E}_T$, $AA' \geqq BB' \Leftrightarrow SS' \geqq TT'$ and that rank $(AA' - BB')$ = rank $(SS' - TT')$. The theorem follows now from (ii). □

4. Comparison by Fisher information matrices. Replicates. If X is $N(A'\beta, \sigma^2 I)$ then the Fisher information matrix is $\sigma^{-2}AA'$ if σ^2 is known and it is

$$\sigma^{-2}\begin{pmatrix} AA' & 0 \\ 0 & 2n_A\sigma^{-2} \end{pmatrix}$$

if σ^2 is unknown. It follows that the comparison criterion in the case of known σ^2 is just the usual ordering of the Fisher information matrices. This criterion could also have been obtained by noting that the Bayes risk for quadratic error for the problem of estimating a given linear combination $t'\beta$ of β when $\beta_1, \cdots, \beta_k$ are independently and normally (0, 1) distributed is $t'(I + AA')^{-1}t$.

It may be shown quite generally that the ordering "being more informative" is stronger than the ordering of Fisher information matrices. In fact there is an intermediate ordering of "being locally more informative" ([7]).

In the case of unknown σ^2 the ordering $\geqq$ of the Fisher information matrices of $\mathscr{E}_A$ and $\mathscr{E}_B$ is the ordering: $AA' \geqq BB'$ and $n_A \geqq n_B$. It follows that this ordering is strictly weaker than the ordering "being more informative."

Ordering of Fisher information matrices of a fixed number, say n, of replicates does not depend on n. In contradistinction to this we have, in the case of unknown σ^2, that n replicates of $\mathscr{E}_A$ is more informative than n replicates of $\mathscr{E}_B$ if and only if[2] $n(n_A - n_B) \geqq$ rank $(AA' - BB')$ and $AA' \geqq BB'$. This may be seen

[2] If A is a matrix, then n_A denotes the number of columns in A.

by noting that the experiment obtained by combining (independently) experiments $\mathscr{E}_{A_1}, \mathscr{E}_{A_2}, \cdots, \mathscr{E}_{A_s}$ is equivalent with the experiment $\mathscr{E}_A$ where $AA' = \sum A_i A_i'$ and $n_A = \sum n_{A_i}$.

REFERENCES

[1] BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[2] BOLL, C. (1955). Comparison of experiments in the infinite case. Ph. D. dissertation, Stanford Univ.

[3] HEYER, HERBERT (1969). Erschöpftheit und Invarianz beim Vergleich von Experimenten. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **12** 21-55.

[4] LECAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419-1455.

[5] TORGERSEN, E. N. (1970). Comparison of experiments when the parameter space is finite. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 219-249.

[6] TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383-1399.

[7] TORGERSEN, E. N. (1972). Unpublished manuscripts on local comparison of experiments. Local comparison of experiments when the parameter set is one dimensional (abstract). *I.M.S. Bulletin* **1** 127-128.

MØRE AND ROMSDALE REGIONAL COLLEGE
MOLDE, NORWAY

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OSLO
BLINDERN, OSLO, NORWAY