Comparison of linear normal experiments.

by

Ole Håvard Hansen og Erik N. Torgersen

A B S T R A C T

Consider independent and normally distributed random variables $X_1,\ldots,X_n$ such that $0 < \operatorname{Var} X_i = \sigma^2$; $i = 1,\ldots,k$ and $E(X_1,\ldots,X_n)' = A'\beta$ where A' is a known $n \times k$ matrix and $\beta = (\beta_1,\ldots,\beta_k)'$ 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 $X_1,\ldots,X_n$ by $\mathcal{E}_A$. 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) $\mathcal{E}_A$ is more informative than $\mathcal{E}_B$ if and only if $AA' - BB'$ is non negative definit (and $n_A \geq n_B + \operatorname{rank}(AA' - BB')$).

Content

Page

1. Introduction.

A notion of "being more informative" for experiments was introduced by Bohnenblust, Shapley and Sherman and may be found in Blackwell [1]. We will write $\mathcal{E} \geq \mathcal{F}$ if the experiment $\mathcal{E}$ is more informative than the experiment $\mathcal{F}$.

Consider now independent and normally distributed random variables $X_1,\ldots,X_n$ such that $0 < \operatorname{Var} X_i = \sigma^2$; $i = 1,\ldots,n$ and $E(X_1,\ldots,X_n)' = A'\beta$ where A' is a known $n \times k$ matrix and $\beta = (\beta_1,\ldots,\beta_k)'$ 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 $\mathcal{E}_A$.

The purpose of this paper is to present a simple criterion for the informational inequality, $\mathcal{E}_A \geq \mathcal{E}_B$, 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 $j = 1,2,\ldots$ and $c \geq 0$ be given constants and consider the experiment $\mathcal{F}_{c,j}$ of observing independent random variables Z and W such that Z is $N(\zeta,(1+c)\sigma^2)$ distributed and $W/_{\sigma^2}$ is χ^2_j distributed. It is assumed that ζ and $\sigma^2 > 0$ are totally unknown. Then Boll proved that $\mathcal{F}_{c,j} \geq \mathcal{F}_{0,k}$ if and only if, $c > 0$ and $j \geq k+1$, or, $c = 0$ and $j \geq k$. Boll proved this as an application of a general result, proved in [2], which roughly states that "$\geq$" within "invariant pairs" of experiments may - provided certain conditions are satisfied - be based on invariant kernels. We include - since reduction by invariance is used here too - a reference to the exposition in section 2 in Torgersen [3].

2. Comparison of "reduced" linear normal models.

We shall in this section 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 non negative integer k.

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

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

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

Here $c_1,\ldots,c_l$ are known constants and we shall assume that $k \geq l$, p and q are given non negative integers. The unknown parameter is $(\beta_1,\ldots,\beta_k)$ when σ^2 is known and it is $(\beta_1,\ldots,\beta_k,\sigma^2)$ when σ^2 is unknown. If σ^2 is known then - by sufficiency - $X_{k+1},\ldots,X_{k+p}$ may be deleted from $\mathcal{E}$ and $Y_{l+1},\ldots,Y_{l+q}$ may be deleted from $\mathcal{F}$. If σ^2 is unknown then - by sufficiency - $X_{k+1},\ldots,X_{k+p}$ may be replaced by $S = X_{k+1}^2+\ldots+X_{k+p}^2$ and $Y_{l+1},\ldots,Y_{l+q}$ by $T = Y_{l+1}^2+\ldots+Y_{l+q}^2$. In order to avoid trivialities we will assume that $l \geq 1$ when σ^2 is known and assume that $k+p, l+q \geq 1$ when σ^2 is unknown.

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

<u>Proposition 2.1</u> If σ^2 is known then $\mathcal{E} \geq \mathcal{F}$ if and only if $|c_i| \leq 1$; $i = 1,\dots,l$.

If σ^2 is unknown then $\mathcal{E} \geq \mathcal{F}$ if and only if $|c_i| \leq 1$; $i = 1,\dots,l$ and $p \geq q + \#\{i : 1 \leq i \leq l, |c_i| < 1\}$.

<u>Remark</u>: The "invariance" part of the proof below is very similar to that of Boll and the proof might - as Boll did - 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 .

<u>Proof of the proposition</u>: Let $\tilde{\mathcal{E}}$ denote the experiment obtained from $\mathcal{E}$ by deleting $X_i : l < i \leq k$. Clearly $\tilde{\mathcal{E}} \geq \mathcal{F} \Rightarrow \mathcal{E} \geq \mathcal{F}$ and that the converse holds may be seen by letting the additive group R^{k-l} act as follows:

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

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

and $(Y_1,\dots,Y_{l+q}) \to (Y_1,\dots,Y_{l+q})$. Obviously $\mathcal{E}$ and $\mathcal{F}$ are both invariant under this group and any invariant kernel from $\mathcal{E}$ to $\mathcal{F}$ does not depend on $X_{l+1},\dots,X_k$. 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 $\mathcal{E}$ is X_i while the UMVU estimator for β_i in $\mathcal{F}$ is Y_i/c_i . The variances of these estimators are, respectively, σ^2 and σ^2/c_i^2 . Hence $\mathcal{E} \geq \mathcal{F} \Rightarrow |c_i| \leq 1$; $i = 1,\dots,k$. Conversely - assume -

$|c_i| \leq 1$; $i = 1,\dots,k$. Let $Z_1,\dots,Z_k$ be independently and normally distributed random variables such that: $(Z_1,\dots,Z_k)$ is independent of $(X_1,\dots,X_k)$, $EZ_i = 0$; $i = 1,\dots,k$ and $\operatorname{var} Z_i = (1-c_i^2)\sigma^2$; $i = 1,\dots,k$. Then $Z_i + c_iX_i$ has the same distribution as $Y_1,\dots,Y_k$.

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

Let G be the group of transformations of Θ of the form:

$$(\beta_1,\dots,\beta_k,\sigma^2) \rightarrow (g_1+g\beta_1,\dots,g_k+g\beta_k,g^2\sigma^2)$$

where $g_1,\dots,g_k$ and g are constants. If $k = (g_1,\dots,g_k,g) \in G$ then we let it move $(X_1,\dots,X_k,S)$ to $(g_1+gX_1,\dots,g_k+gX_k,g^2S)$ and $(Y_1,\dots,Y_k,T)$ to $(g_1+gY_1,\dots,g_k+gY_k,g^2T)$. It may then be checked that $\mathcal{E}$ and $\mathcal{F}$ are both invariant under G . Moreover - since G has an invariant mean - we may restrict attention to invariant kernels (see section 2 in [3]). It follows that we may assume that $(X_1,\dots,X_k,S)$, $(Y_1,\dots,Y_k,T)$ has a joint distribution where the conditional distribution, M , of $(Y_1,\dots,Y_k,T)$ given $(X_1,\dots,X_k,S)$ satisfies

$$(*) \quad M(B_1 \times \dots \times B_k \times B | X_1,\dots,X_k,S) =$$
$$M((g_1c_1+gB_1) \times \dots \times (g_kc_k+gB_k) \times g^2B | g_1+gX_1,\dots,g_k+gX_k,g^2S)$$

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

$$P(Y_1 \in B_1,\dots,Y_k \in B_k | X_1,\dots,X_k)$$
$$= P(Y_1 \in g(B_1-c_1X_1) + c_1X_1,\dots,Y_k \in g(B_k-c_kX_k) + c_kX_k | X_1,\dots,X_k)$$
$$= P(\tfrac{1}{g}(Y_1-c_1X_1) + c_1X_1 \in B_1,\dots,\tfrac{1}{g}(Y_k-c_kX_k) + c_kX_k \in B_k | X_1,\dots,X_k)$$

It follows - by letting $g \to \infty$ - that the conditional distribution of $(Y_1,\dots,Y_k)$ given $X_1,\dots,X_k$ is the one point distribution in $(c_1X_1,\dots,c_kX_k)$ i.e. we may as well assume $Y_i = c_iX_i$; $i = 1,\dots,k$. Hence $\sigma^2 = \text{var } Y_i = c_i^2 \text{ var } X_i = c_i^2\sigma^2$, so that $c_i^2 = 1$; $i = 1,\dots,k$. This proves the desired unequality when $p = 0$.

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

$$E\, e^{\sum_{j=1}^{k} it_jc_jX_j} \left(\prod_{j=1}^{k} E\, e^{it_j\sqrt{S}\,U_j}\right) E\, e^{itSU}$$

$$= \left(\prod_{j=1}^{k} E\, e^{it_jc_jX_j}\, E\, e^{it_j\sqrt{S}\,U_j}\right) . E\, e^{itSU}$$

$$= \left(\prod_{j=1}^{k} E\, e^{it_jY_j}\right) . E\, e^{itT}$$

$$= E\left[\prod_{j=1}^{k} e^{it_jY_j} . e^{itT}\right]$$

$$= E\left(\prod_{j=1}^{k} e^{it_jc_jX_j}\right) E\left(\prod_{j=1}^{k} e^{it_j\sqrt{S}\,U_j}\right) E\, e^{itSU}$$

It follows that $\sqrt{S}\,U_1,\dots \sqrt{S}\,U\ ,SU$ are independent. Hence: $\sigma^{-2}\left[\sum_{i:|c_i|<1} \dfrac{SU_i^2}{1-c_i^2} + SU\right]$ has a χ^2 distribution with $q + \#\{i : |c_i| < 1\}$ degrees of freedom. This yield an unbiased randomized estimator of σ^2 based on S with variance $[q + \#\{i : |c_i| < 1\}]^{-1}2\sigma^4$. Hence - since the UMVU estimator based on S has variance $p^{-1}2\sigma^4$ - $p \geq q + \#\{i : |c_i| < 1\}$.

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

$Y_i = c_i X_i$ if $|c_i| = 1$,

$Y_{i_r} = c_{i_r} X_{i_r} + \sqrt{1-c_{i_r}^2}\, X_{k+r}$, $r = 1,\dots,m$

and

$Y_{k+j} = X_{k+m+j}$; $j = 1,\dots,q$. Hence $\mathcal{E} \geq \mathcal{F}$.

3. Comparison of linear normal models.

For each given $n_A \times k$ matrix A' let $\mathcal{E}_A$ be the experiment obtained by observing n_A independent and normally distributed random variables such that: $\operatorname{Var} X_i = \sigma^2$; $i = 1,\dots,k$ and $E(X_1,\dots,X_k)' = A'(\beta_1,\dots,\beta_k)'$. 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:

$$\mathcal{E}_A \geq \mathcal{E}_B \iff AA' \geq BB'$$

If σ^2 is unknown then:

$$\mathcal{E}_A \geq \mathcal{E}_B \iff AA' \geq BB' \text{ and } n_A \geq n_B + \operatorname{rank}(AA'-BB')$$

Proof: Let I denote the $k \times k$ identity matrix. $\mathcal{E}_A$ may be considered as the experiment of observing a $N(A'\beta, \sigma^2 I)$ distributed $k \times 1$ column matrix X while $\mathcal{E}_B$ may be considered as the experiment of observing a $N(B'\beta, \sigma^2 I)$ distributied $k \times 1$ 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, constitutes a sufficient statistic in $\mathcal{E}_A$. $\hat{\beta}$ is $N(\beta, \sigma^2 I)$ distributed and $\|X - A'\hat{\beta}\|^2/\sigma^2$ is $\chi^2_{n_A - k}$ distributed.

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

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

(ii) $\operatorname{rank} A = k$.

By a well known result on simultanous reduction of two quandratic forms there is a (non singular) $k \times k$ matrix F so that

$$F'AA'F = I$$

and

$$F'BB'F = I$$

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

$$\mathcal{E}_A \geq \mathcal{E}_B \iff \mathcal{E}_{\tilde{A}} \geq \mathcal{E}_{\tilde{B}}$$

The theorem follows - in this case - since $\mathcal{E}_{\tilde{A}}$ and $\mathcal{E}_{\tilde{B}}$ satisfies the condition in (i) and since $\tilde{A}\tilde{A}' \geq \tilde{B}\tilde{B}' \iff AA' \geq BB'$ and $\operatorname{rank}(\tilde{A}\tilde{A}' - \tilde{B}\tilde{B}') = \operatorname{rank}(AA' - BB')$.

(iii) The general case.

If $\mathcal{E}_A \geq \mathcal{E}_B$ then - by the estimability criterion for linear functions of β - $\operatorname{row}[B'] \subseteq \operatorname{row}[A']$. Suppose now that $AA' \geq BB'$ and that $x \perp \operatorname{row}[A']$ then $A'x = 0$ so that $0 = x'AA'x \geq x'BB'x$. Hence $B'x = 0$ so that $x \perp \operatorname{row}[B']$. It follows again that $\operatorname{row}[B'] \subseteq \operatorname{row}[A']$. Hence we may, without loss of generality, assume that $\operatorname{row}[B'] \subseteq \operatorname{row}[A']$. Write $A' = (a_1', \dots, a_{n_A}')'$ and $B' = (b_1', \dots, b_{n_B}')'$ where $a_1', \dots, a_{n_A}'$ and $b_1', \dots, b_{n_B}'$ are, respectively, the row vectors of A' and B' . Let $v_1', \dots, v_r'$ be a basis in $\operatorname{row}[A']$ and write $p_i = v_i'\beta$; $i = 1, \dots, r$. 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 $r = \operatorname{rank} S'$ columns. It is not difficult to check that $\mathcal{E}_A \geq \mathcal{E}_B \iff \mathcal{E}_S \geq \mathcal{E}_T$, $AA' \geq BB' \iff SS' > TT'$ and that $\operatorname{rank}(AA' - BB') = \operatorname{rank}(SS' - TT')$. 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 $\beta_1,\ldots,\beta_k$ 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". [4]

In the case of unknown σ^2 the ordering $\geq$ of the Fisher information matrices of $\mathcal{E}_A$ and $\mathcal{E}_B$ is the ordering: $AA' \geq BB'$ and $n_A \geq n_B$. 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 $\mathcal{E}_A$ is more informative than n replicates of $\mathcal{E}_B$ if and only if * $n(n_A - n_B) \geq \text{rank}(AA' - BB')$. This may be seen by noting that the experiment obtained by combining (independently)

---

* If A is a matrix, then n_A denotes the number of columns in A .

experiments $\mathcal{E}_{A_1}, \mathcal{E}_{A_2}, \ldots, \mathcal{E}_{A_s}$ is equivalent with the experiment $\mathcal{E}_A$ where $AA' = \Sigma A_i A_i'$ and $n_A = \Sigma n_{A_i}$.

5. References.

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

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

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

[4] Torgersen, E.N. (1972). Unpublished manuscripts on local comparison of experiments. Abstract of "Local comparison of Experiments where the Parameter set is one dimensional" in I.M.S. Bulletin Vol.1, 127-128.