Inge Henningsen

CANONICAL HYPOTHESES FOR LINEAR NORMAL MODELS

Preprint 1971 No. 8

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

November 1971

0. Introduction

In statistical analysis a frequently encountered situation is that of a hypothesis on the mean of n independent observations each coming from a k-dimensional normal distribution and all having the same unknown covariance.

It is well known that if the hypothesis on the mean has the form $\xi \in L^k$ for some $L \subseteq R^n$ and the dimension of L is small enough to allow for estimation of the covariance, then the maximum likelihood estimator for the mean and the covariance exists and is sufficient.

We shall show that these hypotheses are canonical in the sense that if we demand existence and sufficiency of the maximum likelihood estimator then under very mild conditions the hypothesis on the mean must have the form $\xi \in L^k$.

1. A lemma

Let $E = (R^n)^k$ be the space of all $n \times k$-matrices and P the set of all positive definite $k \times k$-matrices. $\Sigma \in P$ defines an inner product on E by

$$(x,y) \mapsto \operatorname{tr} \Sigma \quad x'y \qquad (1)$$

and consequently for any subspace $M \subseteq E$ an orthogonal projection

$$p_\Sigma : E \to M.$$

Lemma: If M is a subspace of E and p_Σ, $\Sigma \in P$, is the orthogonal projection on M with respect to the inner product defined by (1), then the following 3 conditions are equivalent:

1) p_Σ, $\Sigma \in P$, is independent of Σ.

2) M has the form $M = L^k$ for some $L \subseteq R^n$.

3) $xb \in M$ for all $x \in M$ and for all $k \times k$-matrices b.

Proof: Let

$$q_j : E \to R^n \qquad j = 1,2, \ldots , k.$$

be the coordinate transformations, let M_j be the image of M under q_j, and let

$$p_j : R^n \to M_j \qquad j = 1,2, \ldots ,k$$

be the orthogonal projection with respect to the ordinary inner product in R^n, $(w,z) \mapsto \sum_{i=1}^{n} w_i z_i$.

1) $\Rightarrow$ 2).

Assume p_Σ independent of Σ, e.g. we have

$$p : E \to M,$$

such that

$$\operatorname{tr} \Sigma(x - px)'y = 0 \quad \forall y \in M, \quad \forall \Sigma \in P \tag{2}$$

Writing (2) coordinatwise gives

$$\sum_{j=1}^{k} \sigma_{jj}(q_jx-q_jpx)'q_jy + \sum_{i>j} \sigma_{ij}((q_jx-q_jpx)'q_iy + (q_ix-q_ipx)'q_jy))=0 \;\forall y \in M, \forall \Sigma \in P.$$

$$\Leftrightarrow \begin{cases} (q_jx - q_jpx)'q_jy = 0 & \forall y \in M \quad j = 1, \ldots ,k & (3) \\ (q_jx - q_jpx)'q_iy + (q_ix - q_ipx)'q_jy = 0 & \forall y \in L \quad i=1,\ldots,k \; j=1,\ldots,k. & (4) \end{cases}$$

Now (3) is equivalent to

$$(q_jx - q_jpx)'w = 0 \qquad \forall\, w \in M_j, \quad j = 1, \ldots ,k ,$$

which implies

$$q_j p x = p_j q_j x \qquad \forall\, x \in E$$

or

$$q_j p = p_j q_j \qquad j = 1, \dots ,k.$$

Hence $$M = M_1 x \dots x M_k \ .$$

Let $w \in M_i$. Since M is a product space we can take $y \in M$, such that $q_j y = 0$, $q_i y = w$. With these values (4) gives

$$(q_j x - p_j q_j x)'w = 0 \qquad \forall\, x \in E \quad \forall\, w \in M_i \ ,$$

which is equivalent to

$$(z - p_j z)'w = 0 \qquad \forall\, z \in R^n \quad \forall w \in M_i. \qquad (5)$$

Now p_j is orthogonal projection on M_j, therefore (5) implies $M_j^\perp \subseteq M_i^\perp$, which in turn implies $M_i \supseteq M_j$. Since i and j are arbitrary we have $M_1 = \dots = M_k$.

2) ⇒ 3).

Take $x \in L^k$, $b = (\beta_{ij})_{\substack{i=1\dots k\\ j=1\dots k}}$. Clearly $xb \in E$ and we have

$$q_i(xb) = \sum_{j=1}^{k} q_j(x)\, \beta_{ij} \ ,$$

which shows that $q_i(xb)$ is a linear combination of $q_1(x), \dots ,q_k(x)$. Hence $q_i(xb) \in L$, since $q_j(x) \in L$, $j = 1, \dots ,k$.

3) ⇒ 1)

Take $\Sigma \in P$. We will show that p_Σ is orthogonal projection under the inner product (1) for any $\Sigma_0 \in P$. Take $y \in M$, then

$$\operatorname{tr} \Sigma_0^{-1} (x - p_\Sigma x)'y = \operatorname{tr} \Sigma^{-1} \Sigma\Sigma_0^{-1} (x - p_\Sigma x)'y$$

$$= \operatorname{tr} \Sigma_0^{-1} (x - p_\Sigma x)'(y\Sigma^{-1}\Sigma) = 0 ,$$

since by assumption $y \Sigma^{-1}\Sigma \in M$ for $y \in M$.

Definition: A subspace $M \subseteq E$ will be called <u>canonical</u> if $M = L^k$ for some $L \subseteq R^n$.

Take S a subset of E, $0 \in S$. The <u>canonical span</u> of S will be the smallest canonical subspace containing S and the <u>canonical dimension</u> of S will be the dimension of the canonical span of S.

By a continuity argument we get the following.

Corollary: To prove a subspace $M \subseteq E$ canonical it is enough to establish the existence of a $p : E \cap N^c \to M$, such that

$$p_\Sigma(x) = p(x) \qquad \forall x \in E \cap N^c , \quad \forall \Sigma \in Q,$$

where $N \subseteq E$ has Lebesgue measure zero , and Q is a subset of P dense in the norm defined by $\operatorname{dist}(\Sigma_1,\Sigma_2) = \max_{\substack{1\le i\le k \\ 1\le j\le k}} |\sigma_{ij}^{(1)} - \sigma_{ij}^{(2)}|$.

p extended by linearity to E is the canonical projection, i.e. orthogonal projection on M with respect to all $\Sigma \in P$.

2. Statement of the problem.

Let S be a subset of E. For $(\xi,\Sigma) \in S \times P$ the density

$$\varphi_{\xi,\Sigma}(x) = (2\pi\det\Sigma)^{-\frac{n}{2}} e^{-\frac{1}{2}\operatorname{tr} \Sigma^{-1}(x-\xi)'(x-\xi)} \qquad (6)$$

with respect to Lebesgue measure on E, defines a set of normal distributions parameterized by $S \times P$.

It is no loss of generality to assume $0 \in S$, since any problem $S \times P$ can be reduced to a problem of this type by a translation.

If the set S is a canonical subspace of E we have for $q : E \to S$ the canonical projection

$$\operatorname{tr} \Sigma^{-1}(x-\xi)'(x-\xi) = \operatorname{tr} \Sigma^{-1}(x-qx)'(x-qx)+\operatorname{tr}\Sigma^{-1}(qx-\xi)'(qx-\xi) \quad \forall\Sigma\in P,\ \forall\xi\in S, \quad (7)$$

hence,

$$\varphi_{\xi,\Sigma}(x)=(2\pi\det\Sigma)^{-\frac{n}{2}} e^{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}(x-qx)'(x-qx)-\frac{1}{2}\operatorname{tr}\Sigma^{-1}(qx-\xi)'(qx-\xi)} \quad (8).$$

If $\dim S \leq k(n-k)$ it is well known (see Rao p.449) that the maximum likelihood estimator for (ξ,Σ) exists with probability 1 and is given by

$$(\hat{\xi},\hat{\Sigma}) = (qx, \frac{1}{n}(x - qx)'(x - qx)) \quad , \qquad x \in E \cap N^c .$$

It is evident from (8) that the maximum likelihood estimator is sufficient. It is, moreover, a surjection.

3. The canonical hypothesis .

We shall say that two problems parametrized by $S_1 \times P$ and $S_2 \times P$ and both allowing for maximum likelihood estimation are equivalent if the two maximum likelihood estimators are identical except on a set of Lebesgue measure zero.

Definition: An estimation problem $(\xi,\Sigma) \in S \times P$ will be called canonical if $S \subseteq E$ is a canonical subspace.

Now let $S \subseteq E$ be any subset and consider the problem of estimating the mean and the covariance in the class of normal distributions having density defined by (6).

Theorem: Let the estimationproblem be given by $(\xi,\Sigma) \in S \times P$. If the canonical dimension of S is less than or equal to $k(n - k)$, then the problem is equivalent to a canonical one if the maximum likelihood estimator exists and is sufficient.

Proof: Define

$$l(\xi,\Sigma)(x) = -\log \varphi_{\xi,\Sigma}(x) \quad .$$

(In what follows N with various subscripts shall denote a subset of E having Lebesgue meausure zero.)

If

$$m : E \cap N_0^{\,c} \rightarrow S \times P$$

is the maximum likelihood estimator for (ξ,Σ), then we have functions h and k such that

$$l(\xi,\Sigma)(x) = h(x) + k(m(x),\xi,\Sigma), \; x \notin N_{\xi,\Sigma} \; , \qquad (9)$$

since m by assumption is sufficient.

Let $\xi,\xi' \in S$, $\Sigma \in P$ and $x,x' \in E \cap (N_{\xi,\Sigma} \cup N_{\xi',\Sigma})^c$.

It follows from (9) that

$$m(x)=m(x') \Rightarrow l(\xi,\Sigma)(x)-l(\xi,\Sigma)(x')-l(\xi',\Sigma)(x) + l(\xi',\Sigma)(x') = 0$$

or

$$m(x)=m(x') \Rightarrow \mathrm{tr}\Sigma^{-1}(x-\xi)'(x-\xi)-\mathrm{tr}\Sigma^{-1}(x'-\xi)'(x'-\xi)-\mathrm{tr}\Sigma^{-1}(x-\xi')'(x-\xi') + \mathrm{tr}\Sigma^{-1}(x'-\xi')'(x'-\xi') = 0$$

or

$$m(x) = m(x') \Rightarrow \mathrm{tr}\,\Sigma^{-1}(x-x')'(\xi-\xi') = 0 \quad . \qquad (10)$$

Let M be the span of S, L the canonical span of S and $q : E \to L$ the canonical projection.

We want to prove M = L and S = L a.s.

Take $\xi_1, \dots, \xi_m$ to span M and let Q be any countable dense set in P. Define

$$N = \bigcup_{\substack{\xi_i, i=1-m \\ \Sigma \in Q}} (N_{\xi_i,\Sigma} \cup N_{0,\Sigma}) \quad .$$

N is a null set.

Take

$$m' : E \to L \times \overline{P}$$
$$x \mapsto (qx, \frac{1}{n}(x - qx)'(x - qx)) = (qx, sx).$$

m' is a surjection, hence for $x \in E \cap N^c$ the set

$$S_x = \{x' \in E \cap N^c \mid m(x) = m'(x')\}$$

is non-empty. Moreover, $x_1, x_2 \in S_x \Rightarrow qx_1 = qx_2$, since $m'(x_1) = m'(x_2)$.

We can therefor define the function

$$p : E \cap N^c \to M$$
$$x \to qx' \quad \text{for } x' \in S_x \ .$$

From

$$l(\xi,\Sigma)(x) = \frac{n}{2}\log 2\pi \det\Sigma + \frac{1}{2}\mathrm{tr}\Sigma^{-1}(x-qx)'(x-qx) + \frac{1}{2}\mathrm{tr}\Sigma^{-1}(qx-\xi)'(qx-\xi) \qquad (11)$$

it is seen that m' is maximum likelihood estimator for the problem $(\xi,\Sigma) \in L \times P$.

Since $S \times P \subseteq L \times P$, we have

$$m'(x) \in S \times P \Rightarrow m'(x) = m(x)$$

or

$$q(x) \in S \Rightarrow m'(x) = m(x).$$

Hence, $x' \in S_x \Rightarrow m(x) = m'(x') = m(x')$.

Therefore, if $x \in E \cap N^c$, $x' \in S_x$ (10) yields

$$tr\Sigma^{-1}(x-x')'y = 0 \qquad \forall y \in M, \qquad \forall \Sigma \in Q$$

or

$$tr\ ^{-1}(x-qx')'y = 0 \qquad \forall y \in M, \qquad \forall \Sigma \in Q,$$

since $M \subseteq L$. Therefore

$$tr\Sigma^{-1}(x-px)'y = 0 \qquad \forall x \in E \cap N^c, \forall y \in M, \forall \Sigma \in Q,$$

and from the corollary p4 it follows that M is canonical, e.g. M = L, and p extended by linearity is the canonical projection. Hence p = q.

But the image of $E \cap N^c$ under p is contained in S, since $x \in E \cap N^c$ and $x' \in S_x \Rightarrow (qx', sx') = m(x) \in S \times P$, e.g. $px \in S, \forall x \in E \cap N^c$. Noting that p is a projection and N is a null set we can infer S = L a.e. and m = m' except on a null set. $L \times P$, however, is canonical which completes the proof.

This result for the multivariate normal distribution is a special case of a general principle suggested to the author by Hans Brøns.

References:

Rao, C.R.: Linear Statististical Inference and its Applications. Wiley, New York (1965).