PREPRINT
SEPT 1971
7

Hans Brøns, Inge Henningsen,
and Søren Tolver Jensen

A Canonical Statistical Hypothesis in the Multidimensional Normal Distribution

UNIVERSITY OF COPENHAGEN
INSTITUTE OF
MATHEMATICAL STATISTICS

Hans Brøns, Inge Henningsen, and Søren Tolver Jensen

A CANONICAL STATISTICAL HYPOTHESIS IN THE

MULTIDIMENSIONAL NORMAL DISTRIBUTION

Preprint 1971 No. 7

INSTITUTE OF MATHEMATICAL STATISTICS

UNIVERSITY OF COPENHAGEN

September 1971

Introduction.

In the statistical theory of the one-dimensional normal distribution the linear hypothesis is canonical in the sense that it is the only kind of hypothesis which gives tractable statistical models and makes it possible to answer all statistical questions concerning the models in a satisfactory way. In geometric or invariant formulation the theory has reached a high degree of perfection. For the problems allowing unknown multidimensional covariances a similar technique does not exist. The literature lists a great number of special cases, but a general theory is lacking completely.

We define a general canonical hypothesis which includes most if not all of the well-behaved normal models. The usefulness of the model lies in its invariant formulation. We therefore begin with a careful exposition of the necessary algebra and then give an invariant treatment of the distributional theory of the normal distribution, which makes it possible to derive the distributions of the maximum likelihood estimators under the canonical hypotheses.

The main algebraic tool is the tensor product which was introduced into statistics by Stein [9] . We use the whole algebraic machinery of Bourbaki [8] and some technique from category theory see e.g. MacLane and Birkhoff [7] .

The present work is an outcome of research done at the Institute of Mathematical Statistics, University of Copenhagen, during the last four years. The aim is an algebraic analysis of the normal models, but the work is not finished and several of the present formulations are tentative.

The authors want to thank Susanne Møller whose work [12] on the Wishart distribution formed the basis of the research presented here.

1. Positive tensors.

Let $\mathcal{S}$ denote the category of sets, $\mathcal{V}$ the category of vectorspaces over R, and $\mathcal{SV}$ the category of semi-vectorspaces over R_+. Here R_+ denotes the semiring of non-negative real numbers. Let

$$W : \mathcal{V} \longrightarrow \mathcal{SV}$$

$$U : \mathcal{SV} \longrightarrow \mathcal{S}$$

$$U' = UW : \mathcal{V} \longrightarrow \mathcal{S}$$

be forgetful functors and F[F'] the left adjoint functor for U[U'].

$\mathcal{SV}$ and $\mathcal{V}$ have tensorproducts which give them the structure of symmetric monoidal closed categories.

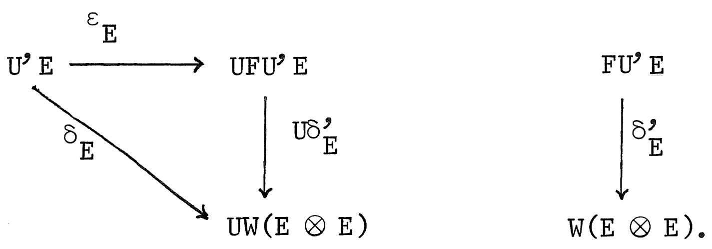
Let $E \in \mathcal{V}$. The isomorphism $\hom_{\mathcal{S}}(U'E, UW(E \otimes E)) \rightarrow \hom_{\mathcal{SV}}(FU'E, W(E \otimes E))$ takes

$$\delta_E : U'E \longrightarrow UW(E \otimes E)$$

$$x \longmapsto x \otimes x$$

into

$$\delta'_E : FU'E \longrightarrow W(E \otimes E).$$



Let (p'_E, PE) be the image of δ'_E. The elements of UPE will be called the <u>positive tensors over E</u>.

P is the object function of a functor $\check{V} \longrightarrow \check{S}\check{V}$, and p' is a natural transformation:

$$
\begin{array}{ccccc}
E & & PE & \xrightarrow{\quad p'_E \quad} & W(E \otimes E) \\
{\scriptstyle e}\downarrow & & {\scriptstyle P(e)}\downarrow & & \downarrow{\scriptstyle W(e \otimes e)} \\
E' & & PE' & \xrightarrow[\quad p'_{E'} \quad]{} & W(E' \otimes E').
\end{array}
$$

This follows immediately from the definition of (p'_E, PE) and the fact that F is a free object functor.

It is, moreover, immediate that δ_E has a unique factorization through $U(p'_E)$,

$$
\begin{array}{ccc}
U'E & \xrightarrow{\quad p_E \quad} & UPE \\
& {\scriptstyle \delta_E}\searrow & \downarrow{\scriptstyle U(p'_E)} \\
& & UW(E \otimes E).
\end{array}
$$

Hence, p is a natural transformation $U' \longrightarrow UP$.

Note that $P(R) = R_+$ and $P(0) = 0$.

In $\check{V}$ let $(E \oplus G, (u,v))$ be the direct sum of E and G with projections r and s,

$$
E \underset{r}{\overset{u}{\rightleftarrows}} E \oplus G \underset{s}{\overset{v}{\leftrightarrows}} G .
$$

Since P is a functor the diagram

$$
\begin{array}{ccccc}
P(E) & \xrightarrow{\quad P(u) \quad} & P(E \oplus G) & \xleftarrow{\quad P(v) \quad} & P(G) \\
& \searrow & \uparrow{\scriptstyle I_{E,G}} & \swarrow & \\
& & PE \oplus PG & &
\end{array}
$$

defines a morphism

$$L_{E,G} : P(E) \oplus P(G) \longrightarrow P(E \oplus G).$$

L is a natural transformation

$$L : \oplus \circ (P \times P) \longrightarrow P \circ \oplus.$$

L is called the <u>direct orthogonal sum transformation</u>. For $\Sigma \in UPE$ and $\Phi \in UPG$ we shall call $L_{E,G}(\Sigma \oplus \Phi)$ the direct orthogonal sum of Σ and Φ and usually write $\Sigma \, L \, \Phi$ instead of $L_{E,G}(\Sigma \oplus \Phi)$.

A positive tensor $H \in UP(E \oplus G)$ is the direct orthogonal sum of a positive tensor from UPE and a positive tensor from UPG if and only if

$$H = P(u \circ r)(H) + P(v \circ s)(H).$$

This follows immediately from the fact that P is a functor.

From associativity and commutativity of the tensor product it follows that the diagram

$$\begin{array}{ccc} U'E \times U'G & \longrightarrow & U'(E \otimes G) \\ \downarrow {\scriptstyle \delta_E \times \delta_G} & & \downarrow {\scriptstyle \delta_{E \otimes G}} \\ UW(E \otimes E) \times UW(G \otimes G) & \longrightarrow & UW((E \otimes E) \otimes (G \otimes G)) = UW(E \otimes G) \otimes (E \otimes G)) \end{array}$$

is commutative.

From this it follows that the diagram

$$\begin{array}{ccc} UPE \times UPG & \longrightarrow & UP(E \otimes G) \\ \downarrow {\scriptstyle U(p'_E) \times U(p'_G)} & & \downarrow {\scriptstyle U(p'_{E \otimes G})} \\ UW(E \otimes E) \times UW(G \otimes G) & \longrightarrow & UW((E \otimes E) \otimes (G \otimes G)) = UW((E \otimes G) \otimes (E \otimes G)) \end{array}$$

is commutative.

To show this, first use the factorization $\delta = U(\delta') \circ \varepsilon$ and then the natural transformation property of $\varepsilon: U' \longrightarrow UFU'$. The proof is completed by using the fact that F is a free object functor and the definition of p' and P.

The function $UPE \times UPG \longrightarrow UP(E \otimes G)$ is bilinear. This gives a factorization

$$\begin{array}{ccc} UPE \times UPG & \longrightarrow & U(PE \otimes PG) \\ & \searrow & \downarrow U(\tau_{E,G}) \\ & & UP(E \otimes G) \end{array} \qquad \begin{array}{c} PE \otimes PG \\ \downarrow \tau_{E,G} \\ P(E \otimes G). \end{array}$$

For $\Sigma \in UPE$ and $\Phi \in UPG$ $\tau_{E,G}(\Sigma \otimes \Phi)$ is called the <u>tensor product of the positive tensors</u> Σ and Φ. We will usually write $\Sigma \otimes \Phi$ for $\tau_{E,G}(\Sigma \otimes \Phi)$.

Note that $1 \in P(R) = R_+$ is unit for

$$PE \otimes PR \longrightarrow P(E \otimes R) = PE.$$

By arguments analogous to those for the direct orthogonal sum it can be shown that τ is a natural transformation from $\otimes \circ (P \times P) \longrightarrow P \circ \otimes$, and the distributive law for direct orthogonal sum and tensor product of positive tensors obtains,

$$P(d) \circ \tau \circ (1 \otimes \perp) = \perp \circ (\tau \oplus \tau) \circ d'.$$

Here $d[d']$ is the distributive law for direct sum and tensor product in $\acute{V}[\acute{S}\acute{V}]$.

2. Regular positive tensors.

In this section all vector spaces will have finite dimension.

For $E \in \acute{V}$ let E^* denote the dual space

$$E^* = Hom_{\acute{V}}(E, R).$$

The canonical transformations

$$E \longrightarrow E^{**},$$

$$\mathrm{Hom}\,(E_1, G_1) \otimes \mathrm{Hom}(E_2, G_2) \longrightarrow \mathrm{Hom}(E_1 \otimes E_2,\ G_1 \otimes G_2),$$

$$E_1^* \otimes E_2^* \longrightarrow (E_1 \otimes E_2)^*,$$

$$E_1^* \oplus E_2^* \longrightarrow E_1^* \times E_2^* \ (\cong (E_1 \oplus E_2)^*),$$

defines all natural transformations between functors, and the transformations being isomorphisms we shall identify the spaces in question.

The sequence of morphisms

$$\begin{aligned}
&PE \longrightarrow \\
&W(E \otimes E) \longleftrightarrow \\
&W(E^{**} \otimes E) \longleftrightarrow \\
&W(\mathrm{Hom}(E^*, R) \otimes \mathrm{Hom}(R, E)) \longleftrightarrow \\
&W(\mathrm{Hom}(E^* \otimes R,\ R \otimes E)) \longleftrightarrow \\
&W(\mathrm{Hom}(E^*, E))
\end{aligned}$$

determines a morphism $\rho : PE \longrightarrow W(\mathrm{Hom}(E^*, E))$.

Definition: A positive tensor $\Sigma \in UPE$ is <u>regular</u>, if $\rho(\Sigma)$ is an isomorphism.

Definition: For $\Sigma \in UPE$ regular,

$$\Sigma^{-1} = P(\rho(\Sigma)^{-1})(\Sigma)$$

is called the <u>inverse positive tensor</u>. $\Sigma^{-1} \in UP(E^*)$.

The proof of the following theorem is not hard.

Theorem: If $\Sigma \in UPE$ is regular, then $\Sigma^{-1} \in UP(E^*)$ is regular, and

$$\Sigma = (\Sigma^{-1})^{-1}.$$

Theorem: If $\Sigma \in UPE$ and $\Phi \in UPG$ are regular then $\Sigma \perp \Phi$ and $\Sigma \otimes \Phi$ are regular and

$$(\Sigma \perp \Phi)^{-1} = \Sigma^{-1} \perp \Phi^{-1}$$

and
$$(\Sigma \otimes \Phi)^{-1} = \Sigma^{-1} \otimes \Phi^{-1}.$$

Proof: Since the diagrams

$$\begin{array}{ccc} PE \oplus PG & \longrightarrow & P(E \oplus G) \\ \downarrow & & \downarrow \\ \mathrm{Hom}(E^*, E) \oplus \mathrm{Hom}(G^*, G) & \longrightarrow \mathrm{Hom}(E^* \oplus G^*, E \oplus G) \cong & \mathrm{Hom}((E \oplus G)^*, (E \oplus G)) \end{array}$$

and

$$\begin{array}{ccc} PE \otimes PG & \longrightarrow & P(E \otimes G) \\ \downarrow & & \downarrow \\ \mathrm{Hom}(E^*, E) \otimes \mathrm{Hom}(G^*, G) & \longrightarrow \mathrm{Hom}(E^* \otimes G^*, E \otimes G) \cong & \mathrm{Hom}((E \otimes G)^*, (E \otimes G)) \end{array}$$

are commutative the assertions follow from $\perp$ and τ being natural transformations.

We shall need the following readily seen result.

Lemma: Let $\Sigma \in UPE$ be regular and $f \in GL(E)$, then $P(f)(\Sigma)$ is regular and

$$(P(f)(\Sigma))^{-1} = P(f^{*-1})(\Sigma^{-1}).$$

3. The normal distribution.

Let $E \in \acute{V}$, let $\acute{M}_1(E)$ be the positive Radon measures with norm 1, and let $\mu \in \acute{M}_1(E)$.

If the integral $\int_E x d\mu(x)$ exists we shall say that μ has a mean. The mean is defined by

$$m(\mu) = \int_E x d\mu(x). \tag{1}$$

Note that $m(\mu) \in E$.

If μ has a mean and the integral $\int_E p_E(x-m(\mu))d\mu(x)$ exists we shall say that μ has a variance. The variance is defined by

$$\mathrm{var}(\mu) = \int_E p_E(x-m(\mu))d\mu(x). \tag{2}$$

$\mathrm{var}(\mu) \in PE$.

Note that

$$\mathrm{var}(\mu) = \int_E p_E(x)d(\delta(m(\mu))\mu)(x). \tag{3}$$

Here δ is translation to the left.

From this and the definition of P(E) it follows that the function

$$\begin{aligned} U\acute{M}_1(E) &\longrightarrow U'E \times UPE \\ \mu &\longmapsto (m(\mu), \mathrm{var}(\mu)) \end{aligned} \tag{4}$$

is a surjection.

For $f \in \mathrm{End}E$ we have

$$m(f\mu) = f(m(\mu)), \tag{5}$$

and by (3)

$$\mathrm{var}(f\mu) = P(f)(\mathrm{var}(\mu)), \tag{6}$$

since p is a natural transformation.

Let $E_1, E_2 \in \acute{V}$, $\mu_1 \in \acute{M}_1(E_1)$ and $\mu_2 \in \acute{M}_1(E_2)$, then $\mu_1 \otimes \mu_2 \in \acute{M}_1(E_1 \oplus E_2)$. If $m(\mu_1)$ and $m(\mu_2)$ exist we have

$$m(\mu_1 \otimes \mu_2) = (m(\mu_1), \; m(\mu_2)). \tag{7}$$

If $var(\mu_1)$ and $var(\mu_2)$ exist it follows from (3), (6), and remark p. 4 that

$$var(\mu_1 \otimes \mu_2) = var\,\mu_1 \perp var\,\mu_2. \tag{8}$$

For a product measure $\bigotimes_{i=1}^{n} \mu$ on $\oplus^n E$ the isomorphisms

$$\oplus^n E \cong \oplus^n (E \otimes R) \cong E \otimes \oplus^n R$$

gives

$$var(\bigotimes_{i=1}^{n} \mu) = \mathop{\perp}_{i=1}^{n} var\,\mu = \mathop{\perp}_{i=1}^{n} var\,\mu \otimes 1 = var\,\mu \otimes (\mathop{\perp}_{i=1}^{n} 1).$$

Here 1 is the unit in P(R).

Let dim E = k and let $(e_i)_{i=1..k}$ be a basis for E. Hence, $(e_i \otimes e_j)_{\substack{i=1...k \\ j=1...k}}$ is a basis for $E \otimes E$. For $\mu \in \acute{M}_1(E)$ the mean can be expressed in terms of $(e_i)_{i=1...k}$ and the variance in terms of $(e_i \otimes e_j)_{\substack{i=1...k \\ j=1...k}}$.

For a finite dimensional vectorspace with a basis the normal distribution $\varphi_{\xi,\Sigma}$ with mean ξ and variance Σ is defined in Brøns [1] chap. 6.

Since the transformation (4) is a surjection we have a distribution $\varphi_{\xi,\Sigma}$ on E with basis $(e_i)_{i=1...k}$, for every choice of $\xi \in E$, $\Sigma \in P(E)$ and $(e_i)_{i=1...k}$.

For $f \in EndE$ Brøns [1] gives

$$f\varphi_{\xi,\Sigma} = \varphi_{f(\xi),P(f)(\Sigma)}, \tag{9}$$

and using (5) and (6) we see that $\varphi_{\xi,\Sigma}$ is independent of the basis $(e_i)_{i=1...k}$.

Moreover, for $E = E_1 \oplus E_2$ and $\xi_1 \in E_1, \xi_2 \in E_2$, $\Sigma_1 \in P(E_1)$, $\Sigma_2 \in P(E_2)$ we get

$$\varphi_{\xi_1 \oplus \xi_2, \Sigma_1 \perp \Sigma_2} = \varphi_{\xi_1, \Sigma_1} \otimes \varphi_{\xi_2, \Sigma_2} \quad . \tag{10}$$

When Σ is regular $\varphi_{\xi,\Sigma}$ has a density

$$\frac{1}{a(\Sigma)} e^{-\frac{1}{2}\Sigma^{-1}((x-\xi)\otimes(x-\xi))} \tag{11}$$

with respect to Lebesgue measure λ on E. ($\Sigma^{-1}(x \otimes x)$ is defined by the contraction

$$E^* \otimes E^* \otimes E \otimes E \longrightarrow R \otimes R = R.)$$

The constant $a(\Sigma)$ depends on λ. From lemma p. 7 it follows that

$$a(P(f))(\Sigma)) = a(\Sigma)\ \mathrm{mod}_E f = a(\Sigma)|\det f|, \tag{12}$$

$f \in GLE$.

For $E = E_1 \oplus E_2$ and $\Sigma = \Sigma_1 \perp \Sigma_2$ the relation

$$a(\Sigma_1 \perp \Sigma_2) = a(\Sigma_1) \cdot a(\Sigma_2) \tag{13}$$

obtains.

4. Canonical Hypotheses.

The normal distributions on the finite dimensional vector space E is parametrized by $U'(E) \times UP(E)$.

A <u>canonical hypothesis</u> is a decomposition

$$E = \bigoplus_{i \in I} (E_i \otimes (L_i \oplus L'_i)) \tag{1}$$

of E together with a parametrization

$$\bigoplus_{i \in I} (E_i \otimes L'_i) \longrightarrow \bigoplus_{i \in I} (E_i \otimes (L_i \oplus L'_i)) \tag{2}$$

determined by the injections

$$E_i \otimes L'_i \longrightarrow E_i \otimes (L_i \oplus L'_i)$$

and

$$\bigoplus_{i \in I} P(E_i) \longrightarrow P(E). \tag{3}$$

The injection (3) is defined by

$$(\Sigma_i)_{i \in I} \longmapsto \perp_{i \in I} (\Sigma_i \otimes (\Phi_i \perp \Phi'_i)),$$

where $\Phi_i \in P(L_i)$ and $\Phi'_i \in P(L'_i)$ are known regular positive tensors.

Sometimes we restrict our attention to regular distributions. The canonical hypothesis is the same except that (3) is restricted to regular positive tensors. $P_r(E)$ shall denote the set of regular positive tensors on E.

Under a canonical hypothesis the variance is a direct orthogonal sum of positive tensors, each of which is a tensorproduct of a completely unknown positiv tensor and a known positive tensor. The mean-value structure is similar to the variance structure. The hypothesis on the mean-value is given as a subspace which is a direct sum of canonical subspaces i.e. subspaces of the form $E_i \otimes L'_i$ where the variance on E_i is completely unknown.

At first sight the hypothesis seems rather restrictive but because of its invariant formulation it actually comprises all well-behaved hypotheses known to the authors. It includes all hypotheses treated in Anderson [2], and all variance component models are canonical corresponding to 1-dimensional E_i's. Also tractable symmetry hypotheses are special cases. They are usually defined by symmetry conditions for the variance matrix in a fixed basis. The conditions are equivalent to requiring that the hypothesis is invariant under a finite group of isomorphisms of E, and it follows in a rather straight forward way from the representation theory for finite groups that these hypotheses are canonical. In literature the statistical problems connected with these hypotheses are treated individually see e.g. Consul [3], Olkin and Press [4], Votaw [5], and Wilks [6].

We shall not deal with problems of testing statistical hypotheses, but only mention that there seems to be four types of tests. Test for mean-value zero, test for identity of variances (Bartlett's test), test for orthogonal decomposition of the variance (test of independence), and test for the decomposition of the variance in a tensor product of an unknown and a known positive tensor. The first three types are well-known while the last test is not treated in the literature.

5. Estimation.

From the relation

$$\varphi_{\underset{i\in I}{\perp}} \Sigma_i \otimes (\Phi_i \perp \Phi'_i) = \bigotimes_{i\in I} (\varphi_{\Sigma_i \otimes \Phi_i} \otimes \varphi_{\Sigma_i \otimes \Phi'_i})$$

it follows that the canonical estimation-problem decomposes into a product of estimationproblems.

Therefore, we can restrict ourselves to the simple canonical hypothesis

$$(E \otimes L) \oplus (E \otimes L') \text{ with } \Phi \in P(L),\ \Phi' \in P(L') \text{ regular,}$$

without loss of generality.

Since we are only interested in the maximum likelihood estimator, we shall further restrict the hypothesis to distributions with regular variance.

We have the distribution

$$\varphi_{0,\Sigma \otimes \Phi} \otimes \varphi_{\xi,\Sigma \otimes \Phi'} \quad , \quad \xi \in (E \otimes L') \text{ and } \Sigma \in P_r(E)$$

having density

$$\frac{1}{a(\Sigma \otimes \Phi)} e^{-\frac{1}{2}(\Sigma^{-1} \otimes \Phi^{-1})(x_1 \otimes x_1)} \cdot \frac{1}{a(\Sigma \otimes \Phi')} e^{-\frac{1}{2}(\Sigma^{-1} \otimes \Phi^{-1})((x_2-\xi)\otimes(x_2-\xi))}$$

with respect to Lebesgue measure $\lambda_1 \otimes \lambda_2$, $\lambda_1[\lambda_2]$ being Lebesgue measure on $E \otimes L$ $[E \otimes L']$.

Since

$$(\Sigma^{-1} \otimes \Phi^{-1})(x_1 \otimes x_1) = \Sigma^{-1} \circ (1_{E \otimes E} \otimes \Phi^{-1})(x_1 \otimes x_1),$$

we can define the function

$$s : E \otimes L \longrightarrow E \otimes E$$

$$x_1 \longmapsto 1_{E \otimes E} \otimes \Phi^{-1}(x_1 \otimes x_1).$$

$s(x_1)$ is the <u>unnormed empirical variance</u>.

Remark: $(s(x_1), x_2)$ is a sufficient statistic for (Σ, ξ).

We have $\rho(\Phi^{-1}) \in \mathrm{Hom}(L, L^*)$ and

$$x_1 \in E \otimes L \cong E \otimes L^{**} \cong \mathrm{Hom}(R, E) \otimes \mathrm{Hom}(L^*, R)$$

$$\cong \mathrm{Hom}(L^*, E).$$

Contraction corresponds to composition of morphisms, hence we have

$$s(x_1) = P(x_1)(\Phi^{-1}).$$

This shows that the empirical variance is an element in P(E).

Suppose dim E $\leqq$ dim L. The elements of $E \otimes L = \mathrm{Hom}(L^*, E)$ of full rank constitutes an open, overall dense subset $(E \otimes L)_r$. Hence, $s(x_1)$ is regular with probability 1. If we restrict s to $(E \otimes L)_r$,

$$s : (E \otimes L)_r \longrightarrow P_r(E)$$

is a surjection.

GL(E) acts on $(E \otimes L)_r$ and $P_r(E)$ by $(f, x_1) \longmapsto (f \otimes 1)(x_1)$ and $(f, y) \longrightarrow P(f)(y)$, respectively.

s commutes with these actions, since

$$s((f \otimes 1)(x_1)) = (1_{E \otimes E} \otimes \Phi^{-1})(f \otimes 1 \otimes f \otimes 1)(x_1 \otimes x_1)$$

$$= (f \otimes f)\ (1_{E \otimes E} \otimes \Phi^{-1})(x_1 \otimes x_1)$$

$$= (f \otimes f)(s(x_1)) = P(f)(s(x_1)).$$

Let μ be image by s of λ_1,

$$\mu = s\lambda_1 \ .$$

Lemma: The measure ν on $P_r(E)$ defined by

$$d\nu(y) = \frac{1}{a(y \otimes \Phi)}\, d\mu(y)$$

is the measure on $P_r(E)$ invariant under the action of GL(E). See e.g. Bourbaki [11].

Proof: For $f \in GL(E)$ we have $\delta(f)\lambda_1 = |\det\ f \otimes 1\,|\lambda_1$. Since s commutes with the action of GL(E) it follows that $\delta(f)\mu = |\det\ f \otimes 1|\mu$ and we have

$$\delta(f)\ \frac{1}{a(y \otimes \Phi)}\ d\mu(y) = \frac{1}{a(P(f)(y) \otimes \Phi)}\ d(\delta(f)\mu)(y)$$

$$= \frac{1}{a(y \otimes \Phi)|\det\ f \otimes 1|}\ |\det\ f \otimes 1|\, d\mu(y) = \frac{1}{a(y \otimes \Phi)}\ d\mu(y).$$

The following argument gives the distribution of $s(x_1)$.

$$d(s\varphi_{\Sigma \otimes \Phi})(y) = \frac{1}{a(\Sigma \otimes \Phi)} e^{-\frac{1}{2}\Sigma^{-1}(y)} d(s\lambda_1)(y)$$

$$= \frac{1}{a(\Sigma \otimes \Phi)} e^{-\frac{1}{2}\Sigma^{-1}(y)} d\mu(y) = \frac{a(y \otimes \Phi)}{a(\Sigma \otimes \Phi)} e^{-\frac{1}{2}\Sigma^{-1}(y)} d\nu(y).$$

The distribution

$$dw(y) = \frac{a(y \otimes \Phi)}{a(\Sigma \otimes \Phi)} e^{-\frac{1}{2}\Sigma^{-1}(y)} d\nu(y)$$

is the <u>Wishart-distribution</u> with parameters (Σ, Φ).

Theorem: The maximum likelihood estimator for (Σ, ξ) is

$$\left(\frac{1}{(\Phi \perp \Phi')^{-1}(\Phi \perp \Phi')} s(x_1), x_2\right).$$

Proof: The density of the sufficient statistic $(s(x_1), x_2)$ is

$$\frac{a(s(x_1) \otimes \Phi)}{a(\Sigma \otimes \Phi)} e^{-\frac{1}{2}\Sigma^{-1}(s(x_1))} \cdot \frac{1}{a(\Sigma \otimes \Phi')} e^{-\frac{1}{2}(\Sigma \otimes \Phi')^{-1}((x_2-\xi)\otimes(x_2-\xi))}.$$

It is evident that $\hat{\xi} = x_2$ and it only remains to maximize

$$\frac{a(s(x_1) \otimes \Phi)}{a(\Sigma \otimes \Phi)} \frac{1}{a(\Sigma \otimes \Phi')} e^{-\frac{1}{2}\Sigma^{-1}(s(x_1))}$$

$$= \frac{a(s(x_1) \otimes (\Phi \perp \Phi'))}{a(\Sigma \otimes (\Phi \perp \Phi'))} e^{-\frac{1}{2}\Sigma^{-1}(s(x_1))} \cdot \frac{1}{a(s(x_1) \otimes \Phi')}.$$

$\frac{1}{a(s(x_1) \otimes \Phi')}$ does not depend on Σ, so the problem is reduced to the problem of maximizing the density of the Wishart distribution with parameters $(\Sigma, \Phi \perp \Phi')$.

We consider therefore the canonical hypothesis

$$E \otimes (L \otimes L')$$

for $(\Phi \perp \Phi') \in P(L \oplus L')$.

The sufficient statistic $y = s(x)$ will have a Wishart distribution

$$dw(y) = \frac{a(y \otimes (\Phi \perp \Phi'))}{a(\Sigma \otimes (\Phi \perp \Phi'))}\, e^{-\frac{1}{2}\Sigma^{-1}(y)}\, d\nu(y).$$

This is an exponential family and the maximum likelihood estimator is defined by

$$m(w) = y. \tag{1}$$

Now

$$m(w) = \int y\,dw(y) = \int s(x)\,d\varphi(x)$$

$$= \int (1 \otimes (\Phi \perp \Phi')^{-1})(x \otimes x)\, d\varphi(x).$$

For $g, h \in \mathrm{Hom}\,(E \otimes (L \oplus L'),\ E)$

$$\int (g \otimes h)(x \otimes x)\, d\varphi(x) = (g \otimes h) \int (x \otimes x)\, d\varphi(x)$$

$$= (g \otimes h)\, \mathrm{var}(\varphi) = (g \otimes h)(\Sigma \otimes (\Phi \perp \Phi')).$$

This relation can be extended by linearity and

$$\mathrm{Hom}(E \otimes (L \oplus L'),E) \otimes \mathrm{Hom}(E \otimes (L \oplus L'),E) \longrightarrow \mathrm{Hom}(E \otimes (L \oplus L') \otimes E \otimes (L \oplus L'), E \otimes E)$$

being an isomorphism we get

$$m(w) = (1 \otimes (\Phi \perp \Phi')^{-1})(\Sigma \otimes (\Phi \perp \Phi')) = \Sigma \otimes (\Phi \perp \Phi')^{-1}(\Phi \perp \Phi')$$

$$= \Sigma \cdot (\Phi \perp \Phi')^{-1}(\Phi \perp \Phi').$$

From this it follows that the equation (1) has the solution

$$\Sigma = \frac{y}{(\Phi \perp \Phi')^{-1}(\Phi \perp \Phi')}$$

Hence,

$$\hat{\Sigma} = \frac{s(x_1)}{(\Phi \perp \Phi')^{-1}(\Phi \perp \Phi')} \quad .$$

We have proved that for the canonical hypothesis the maximum likelihood estimator exists and is sufficient. In Henningsen [10] it is shown that if we have a problem

$$\begin{aligned} S &\longrightarrow U'(E \otimes F) \\ P(E) &\longrightarrow P(E \otimes F) \\ \Sigma &\longmapsto \Sigma \otimes \Phi \quad , \end{aligned}$$

where Φ is known and S is an arbitrary subset, then existence and sufficiency of the maximum likelihood estimator under very mild conditions imply that the hypothesis is canonical.

References.

[1] Brøns, H. Forelæsninger i matematisk statistik og sandsynligheds-regning 1-8. Institut for matematisk statistik, Københavns universitet. København 1966.

[2] Anderson, T.W. An Introduction to Multivariate Statistical Analysis. Wiley, New York. 1958.

[3] Consul, P.C. On the exact distributions of Votaw's criteria for testing compound symmetry of a covariance matrix. Ann. Math. Statist. 40, 1969. 836-43.

[4] Olkin, J. and Press, S.J. Testing and estimation for a circular stationary model. Ann. Math. Statist. 40, 1969. 1358-73.

[5] Votaw, D.F. Testing compound symmetry in a normal multivariate distribution. Ann. Math. Statist. 19, 1948. 447-73.

[6] Wilks, S.S. Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. Ann. Math. Statist. 17, 1946, 257-81.

[7] MacLane, S. and Birkhoff, G. Algebra. The MacMillan Company, New York. 1967.

[8] Bourbaki, N. Éléments de Mathématique, Algèbre, Chapitre 1 à 3. Hermann, Paris. 1970.

[9] Stein, C.M. Lectures on Multivariate Analysis. 1966. Unpublished.

[10] Henningsen, I. A canonical hypothesis for linear normal models. Preprint No. 8, Institute of Mathematical Statistics, University of Copenhagen, København 1971.

[11] Bourbaki, N. Éléments de Mathématique, Intégration, Chapitre 7, Hermann, Paris. 1963.

[12] Møller, S. Anvendelse af Haarmålsteorien på ortogonale grupper og flerdimensionale normale fordelinger. Institut for matematisk Statistik, Københavns Universitet, 1970. Unpublished.

/at