MULTIVARIATE ANALYSIS I

BY

CHARLES M. STEIN

(Notes prepared by M. L. Eaton)

TECHNICAL REPORT NO. 42
DECEMBER 2, 1969

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT 6681

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MULTIVARIATE ANALYSIS I

Charles M. Stein

(Notes prepared by M. L. Eaton)

Section 1.

To begin, a characterization result for some univariate distributions will be presented and then an application to the univariate normal distribution will be given. Let $p(x)$ be a density on the real line $\mathbb{R}$ with respect to Lebesgue measure and suppose $dp(x)/dx$ exists a.e. Let $\mathcal{J}$ be the class of functions on $\mathbb{R}$ such that $f \in \mathcal{J}$ implies

a. $\frac{d}{dx} f(x)$ exists a.e.,

b. $\lim_{|x| \to \infty} f(x)p(x) = 0$,

c. $\int_a^b \frac{d}{dx} [p(x)f(x)] dx < \infty$ for all finite $a$ and $b$,

and define

$$h_f(x) = f'(x) + \frac{d}{dx} \log p(x) \cdot f(x),$$

**Proposition:** If $X$ is a random variable with density $p(x)$, then $\mathbb{E}(h_f(X)) = 0$ for all $f \in \mathcal{J}$. 

1
Proof: \( \mathcal{E}(h_p(x)) = \int [f'(x) + \frac{p'(x)}{p(x)} f(x)] p(x) dx \)

\[ = \int [f'(x)p(x) + p'(x)f(x)] dx \]

\[ = \lim_{a \to -\infty} \int_a^b \frac{d}{dx} [f(x)p(x)] dx \]

\[ = \lim_{b \to +\infty} \int_a^b [f'(x)p(x)] dx \]

\[ = \lim_{a \to -\infty} \int_a^b [f'(x)p(x)] dx \]

\[ = \lim_{b \to +\infty} \int_a^b [f'(x)p(x)] dx \]

Proposition: Let \( X \) be a random variable with distribution function \( G(x) \). If \( \mathcal{E}_h_p(X) = 0 \) for \( f \in \mathcal{F} \), then \( \frac{dG}{dx} = p(x) \).

Proof: Let \( K(x) \) be any bounded function such that \( \int_{-\infty}^{\infty} K(x)p(x)dx = 0 \), and define \( f(x) \) by \( f(x) = \frac{1}{p(x)} \int_{-\infty}^{x} p(y)K(y)dy \) on the set where \( p \) does not vanish, and \( f = 0 \), otherwise. It is easy to verify that \( f \in \mathcal{F} \). Since \( f(x)p(x) = \int_{-\infty}^{x} p(y)K(y)dy \), we have \( \frac{d}{dx}(f(x)p(x)) = p(x)K(x) \), so that \( K(x) = \frac{1}{p(x)} \frac{d}{dx}(f(x)p(x)) = h_p(x) \). By assumption, \( \mathcal{E}(K(x)) = 0 \) since \( K(X) = h_p(X) \). Now, let

\[
K(x) = \begin{cases} 
1 - \int_{-\infty}^{y_0} p(y)dy, & \text{if } x \leq y_0, \\
\int_{-\infty}^{y_0} p(y)dy, & \text{if } x > y_0.
\end{cases}
\]

Then \( \int_{-\infty}^{\infty} K(x)p(x)dx = 0 \) and \( \mathcal{E}(K(x)) = 0 \) implies \( G(y_0) = \int_{-\infty}^{y_0} p(x)dx \).
Definition. A random variable $X$ has a **unit normal distribution** if the density of $X$ is $p_X(x) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}x^2\right)$. A real valued random variable $Y$ has a normal distribution if for some constants $\sigma$ and $x^0$, $Y$ has the same distribution as $\sigma X + x^0$ where $X$ is unit normal and $\sigma > 0$. If $Y$ is normal, the density of $Y$ is

$$p_{Y|\sigma,x^0}(y|x^0) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(y-x^0)^2}{2\sigma^2}}.$$ 

Let $X$ be unit normal. It is easy to show $\mathcal{E}(X) = 0$ and $\mathcal{E}(X^2) = 1$. Since $p_X(x)$ is a symmetric function, $\mathcal{E}(X^{2K-1}) = 0$ for $K = 1,2,\ldots$. Let $f(x) = x^{2K+1}$ so that

$$f'(x) + (\partial \log p_X(x)/\partial x)f(x)$$

$$= (2K+1)x^{2K} - x \cdot x^{2K+2}$$

$$= (2K+1)x^{2K} - x^{2K+2}.$$ 

For $X$ unit normal, we then have $\mathcal{E}(X^{2K+2}) = (2K+1)(x^{2K}) K = 1,2,3,\ldots$. Induction gives us $\mathcal{E}(X^{2n}) = 1 \cdot 3 \cdot 5 \ldots (2n-1)$.

**Comment:** If $X$ and $Y$ are independent random variables such that $X + Y$ is independent of $X - Y$, then both $X$ and $Y$ are normal.

**Existence and Uniqueness of the Multivariate Normal Distribution:**

Let $\mathcal{X}$ be a finite dimensional linear vector space, $\mathcal{X}'$ the dual space of $\mathcal{X}$, and $\mathcal{X}''$ the second dual of $\mathcal{X}$. Let $X$ be a random variable
Definition: $X$ has a normal distribution if for every $\xi \in \mathcal{F}'$, $\langle \xi, X \rangle$ has a normal distribution.

Proposition: If $X$ has a normal distribution in $\mathcal{F}$ and $\alpha$ is a linear transformation from $\mathcal{F}$ into a linear space $\mathcal{Y}$, then $\alpha X$ has a normal distribution in $\mathcal{Y}$.

Proof: Let $\eta \in \mathcal{Y}'$. Then $\langle \eta, \alpha X \rangle = \langle \alpha' \eta, X \rangle$ which has a normal distribution since $\alpha' \eta \in \mathcal{F}'$ for all $\eta \in \mathcal{Y}'$.

Let $X$ be a random variable in $\mathcal{F}$ and assume $\mathcal{E}|\langle \xi, X \rangle| < \infty$ for all $\xi \in \mathcal{F}'$ so that $\mathcal{E}<\xi, X>$ exists. Consider the function $\xi \rightarrow \mathcal{E}<\xi, X>$. This function is linear and hence is an element of $\mathcal{F}''$. Thus, there exists a unique $x^0 \in \mathcal{F}$ such that $\mathcal{E}\xi, X = \langle \xi, x^0 \rangle$.

We then define $\mathcal{E}(x)$ to be $x^0 \in \mathcal{F}$.

For $x \in \mathcal{F}$ define the function $\tilde{x}: \mathcal{F}' \rightarrow \mathcal{F}$ by $\tilde{x}r = rx$. Then $\tilde{x}: \mathcal{F}' \rightarrow \mathcal{F}' \rightarrow \mathcal{F}$ is $\tilde{x}' \xi = \langle \xi, x \rangle$. If $\mathcal{F}$ and $\mathcal{Y}$ are linear spaces $x \in \mathcal{F}$ and $y \in \mathcal{Y}$ then $\tilde{x} \cdot \tilde{y}' : \mathcal{Y}' \rightarrow \mathcal{F}$. Let $X$ be a random variable in $\mathcal{F}$ such that $\mathcal{E}(\langle \xi, X \rangle)^2 < \infty$ for all $\xi \in \mathcal{F}'$. The covariance operator of $X$ is a function $\Gamma_X : \mathcal{F}' \rightarrow \mathcal{F}$ defined by

$$\Gamma = \mathcal{E}(X - \mathcal{E}(X)) \cdot (X - \mathcal{E}(X))'.$$

If $\alpha : \mathcal{F} \rightarrow \mathcal{Y}$ is linear and $X$ is a random variable in $\mathcal{F}$, then

$$\Gamma_{\alpha X} = \mathcal{E}(\alpha X - \alpha \mathcal{E}(X)) \cdot (\alpha X - \alpha \mathcal{E}(X))' = \alpha \cdot \mathcal{E}(X - \mathcal{E}(X)) \cdot (\alpha X - \alpha \mathcal{E}(X))' \cdot \alpha' = \alpha \Gamma_X \alpha'.$$

Note that $\Gamma_X = \Gamma_X'$ and also $\langle \xi, \Gamma_X \xi \rangle = \mathcal{E}(\langle \xi, X - \mathcal{E}(X) \rangle)^2 \geq 0$, so that $\Gamma_X$ is positive semi-definite.
Now, let \( x^0 \in \mathcal{X} \) and \( \Gamma : \mathcal{X}' \to \mathcal{X} \) be a self-adjoint positive semi-definite linear operator on \( \mathcal{X}' \) to \( \mathcal{X} \). We want to construct a random variable \( X \) in \( \mathcal{X} \) which is normal with mean \( x^0 \) and covariance operator \( \Gamma \). Let \( x_i, i = 1, \ldots, r \) be independent unit normals where \( r = \text{rank}(\Gamma) \). Choose \( \eta_1, \ldots, \eta_r \) so that \( \{\Gamma \eta(i)\} \) is a basis for \( \text{HM}(\Gamma) \) and apply Gram-Schmidt orthogonalization to obtain \( \xi^1, \ldots, \xi^r \) such that \( <\xi^i, \xi^j> = \delta_{ij} \). Then let \( X = \sum_{i=1}^{r} X_i \Gamma \eta_i + x^0 \). It is clear that 

\[
\mathcal{E}(X) = x^0 \quad \text{and} \quad X \text{ is normal in } \mathcal{X}.
\]

Also \( \mathcal{E}(X - x_0)(X - x_0)' = \mathcal{E} 
\[
\begin{bmatrix}
\sum_{i} X_i \Gamma \eta_i \end{bmatrix}' 
\[
\begin{bmatrix}
\sum_{i} X_i \Gamma \eta_i \end{bmatrix}
\]

\[
= \sum_{i} \Gamma \eta_i ' (\Gamma \eta_i)' = \Gamma (\sum_{i} \xi_i ' \xi_i ') \Gamma = \Gamma.
\]

Thus \( \Gamma_X = \Gamma \). The uniqueness of the normal distribution \( \mathcal{X} \) is established by recalling the following two facts:

1. The distribution of \( X \) is completely determined by the family of distributions \( <\xi, X> \).

2. The mean and variance of a normal distribution on \( \mathcal{X} \) completely determine the distribution.

From this we conclude that the distribution of \( X \), when \( X \) is normal, is completely determined by \( x^0 \) and \( \Gamma \).

Let \( X_1, \ldots, X_n \) be independent and identically distributed normal random variables in \( \mathcal{X} \) such that \( \mathcal{E}(X_i) = 0 \) and \( \text{Cov}(X_i) = \Gamma \) for \( i = 1, \ldots, n \).

Definition: The random variable \( S = \sum_{i=1}^{n} X_i X_i' \) on \( \mathcal{X}' \) to \( \mathcal{X} \) is said to have a Wishart distribution with \( n \) degrees of freedom.
It is clear that $E(S) = n\Gamma$ and by the strong law of large numbers $(S_n/n) \overset{a.s.}{\rightarrow} \Gamma$. If $\alpha: \mathbb{R} \rightarrow \mathbb{R}$ is linear and $S$ is Wishart, then $\alpha S\alpha' = \sum_{i=1}^{n} \alpha_i X_i \alpha'$ is Wishart with $n$ degrees of freedom and $E(\alpha S\alpha') = n\alpha \Gamma \alpha'$. Also, if $S_1$ and $S_2$ are independent Wishart with $n_1$ and $n_2$ degrees of freedom, respectively, such that $E(S_i) = n_i \Gamma$, $i = 1,2$, then $S_1 + S_2$ will be Wishart with $n_1 + n_2$ degrees of freedom and $E(S_1 + S_2) = (n_1 + n_2)\Gamma$.

**Lemma:** Let $\mathcal{X}$ be a dimension $p$ and $X_1, \ldots, X_n$ independent normal random variables in $\mathcal{X}$, with the same covariance matrix $\Gamma$ which we assume positive definite. If $n \leq p$, then $P(X_1, \ldots, X_n$ are linearly independent) = 1.

**Proof:** $P(X_1, \ldots, X_n$ are linearly independent) = 1 - $P(X_n$ is a linear combination of $X_1, \ldots, X_{n-1}) = 1 - E[P(X_n$ is a linear combination of $X_1, \ldots, X_{n-1} | X_1, \ldots, X_{n-1}) = 1 - E(0) = 1$.

From this lemma, we have that if $n \geq p$ and $X_1, \ldots, X_n$ are independent and identically distributed $N(0, \Gamma)$ where $\Gamma$ is positive definite then $P(X_1, \ldots, X_n$ span $\mathcal{X}) = 1$ so that $P(S = \sum_{i=1}^{n} X_i X_i' \text{ is positive definite}) = 1$.

Now, let $\Gamma$ be positive definite and $\dim \mathcal{X} = p$ and consider $X_1, \ldots, X_p$ independent $N(0, 1)$ random variables. Then

$$P_{X_1, \ldots, X_p}(x_1, \ldots, x_p) = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2} \sum_{i=1}^{p} x_i^2} = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2} \text{tr} xx'}$$
where $x' = (x_1, \ldots, x_p)$. Let $X = \alpha \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} + x^0$ where $\alpha$ is a matrix whose $i^{th}$ column is $\Gamma_i^i ((i)_{i=1}^p)$ is the orthogonal basis for $\text{IM}(\Gamma)$ constructed earlier. Then we have

$$p_{X|X^0, \Gamma}(x|x^0, \Gamma) = \frac{|\alpha|^{-1}}{(2\pi)^{p/2}} e^{-\frac{1}{2} \text{tr} (\alpha^{-1})^2 (x-x^0)(x-x^0)'}.$$ 

Now

$$\alpha^2 = \Gamma \left( \sum_{i=1}^p \Gamma_i^i \Gamma_i^i \right) \Gamma = \Gamma \quad \text{so that} \quad |\alpha| = |\Gamma|^{\frac{1}{2}} \quad \text{and} \quad (\alpha^{-1})^2 = \Gamma^{-1}. \quad \text{Thus}$$

$$p_{X|X^0, \Gamma}(x|x^0, \Gamma) = \frac{|\Gamma|^{-\frac{1}{2}}}{(2\pi)^{p/2}} e^{-\frac{1}{2} \text{tr} \Gamma^{-1} (x-x^0)(x-x^0)'},$$

where $X$ has a normal distribution with mean $x^0$ and covariance operator $\Gamma$. If $X_1, \ldots, X_n$ are independent $N(0, \Gamma)$ and

$$S = \sum_{i=1}^n x_i x_i'$$

then it will be shown later that

$$p_S|\Gamma(s|\Gamma) = c|\Gamma^{-1}s|^{n/2} |s|^{-\frac{p+1}{2}} e^{-\frac{1}{2} \text{tr} \Gamma^{-1} s},$$

where $\Gamma$ is assumed to be positive definite.

Section 2.

We begin with a review of the elements of group theory.

**Definition:** A group is a set $\mathcal{G}$ together with a function, whose value at $(a, b) \in \mathcal{G} \times \mathcal{G}$ is denoted by $ab$, such that

(i) $(a, b) \in \mathcal{G} \times \mathcal{G} \Rightarrow ab \in \mathcal{G}$.

(ii) There exists $1 \in \mathcal{G}$ such that $al = la = a$ for $a \in \mathcal{G}$.
(iii) For each \( a \in \mathcal{Y} \), there exists \( a^{-1} \in \mathcal{Y} \) such that
\[ a^{-1}a = aa^{-1} = 1. \]

**Definition:** A group \( \mathcal{Y} \) is called **abelian** if \( ab = ba \) for all \( a, b \in \mathcal{Y} \).

**Definition:** A non-empty subset \( \mathcal{H} \) of a group \( \mathcal{Y} \) is a **subgroup** if the restriction of the operation in \( \mathcal{Y} \) to \( \mathcal{H} \) makes \( \mathcal{H} \) a group.

**Comment:** If \( \mathcal{H} \) is a non-empty subset of a group \( \mathcal{Y} \), then \( \mathcal{H} \) is a subgroup if and only if \( ab^{-1} \in \mathcal{H} \) for all \( a, b \in \mathcal{H} \).

**Example of a group:** Let \( \mathcal{X} \) be a set and \( \mathcal{Y}_x \) the set of all 1-1 functions of \( \mathcal{X} \) onto \( \mathcal{X} \). Define multiplication in \( \mathcal{Y}_x \) by
\[ (g_1, g_2)(x) = g_1(g_2x). \]
Then \( \mathcal{Y}_x \) is a group and is called the **permutation group** of \( \mathcal{X} \).

**Example:** If \( \mathcal{X} \) is a linear space, then under the operation of addition, \( \mathcal{X} \) is an abelian group.

**Example:** Let \( \mathcal{X} \) be a linear space of dimension \( n \). Some important groups associated with \( \mathcal{X} \) are now given.

\[ (i) \quad \text{Translation group:} \quad \text{if} \quad x_1 \in \mathcal{X}, \quad \text{define} \quad g_{x_1}(x) = x + x_1. \]

The collection of \( g_{x_1} \) form an additive abelian group.

\[ (ii) \quad \text{Let} \quad \mathcal{L}(n) \quad \text{denote the set of all non-singular linear transformations of} \quad \mathcal{X} \quad \text{onto} \quad \mathcal{X}. \quad \mathcal{L}(n) \quad \text{is obviously a group with multiplication} \]
defined as function composition.

(iii) An element of the affine group is a pair \((g,x)\) where \(g \in \mathcal{L}(n)\) and \(x \in \mathcal{X}\). Multiplication is defined by

\[(g_1,x_1)(g_2,x_2) = (g_1g_2, g_1x_2 + x_1).\]

(iv) Let \(\Gamma : \mathcal{X}^1 \rightarrow \mathcal{X}\) be symmetric and positive definite. The orthogonal group with respect to \(\Gamma\), denoted by \(\mathcal{O}(\Gamma)\), is defined to be the set of elements \(g \in \mathcal{L}(n)\) such that \(g\Gamma g' = \Gamma\).

(v) The unimodular group is the subgroup of \(\mathcal{L}(n)\) such that \(g\) is in this group if and only if \(\det(g) = 1\).

(vi) Let \(0 = \mathcal{X}_1 \subset \mathcal{X}_2 \subset \cdots \subset \mathcal{X}_k = \mathcal{X}\) be a strictly increasing sequence of linear subspaces of \(\mathcal{X}\). Let \(\mathcal{U}\) be the subgroup of \(\mathcal{L}(n)\) such that \(g \in \mathcal{U}\) if and only if \(g\mathcal{X}_i = \mathcal{X}_i\) for \(i = 1, \ldots, k\). Thus \(\mathcal{U}\) is the group of non-singular linear transformations on \(\mathcal{X}\) to \(\mathcal{X}\) which leave each \(\mathcal{X}_i\) invariant.

If we choose a basis \(x_1^{(1)}, \ldots, x_k^{(1)}, \ldots, x_1^{(k)}, \ldots, x_1^{(k)}\) for \(\mathcal{X}_1 - \mathcal{X}_{i-1}\) where \(k_i = \dim \mathcal{X}_i - \dim \mathcal{X}_{i-1}\), and write an element \(g \in \mathcal{U}\) in matrix form, then \(g\) has the form

\[
g = \begin{bmatrix}
g_{11} & g_{12} & \cdots & g_{1k} \\
0 & g_{22} & \cdots & g_{2k} \\
& & \ddots & \vdots \\
& & & g_{kk}
\end{bmatrix},
\]

where \(g_{ij}\) is a block of size \(k_i \times k_j\) for \(i \leq j\). When each \(k_i = 1\) we of course obtain the upper triangular group of matrices.
Definition: Let \( X \) be a set and \( Y \) be a group. \( Y \) operates on \( X \) on the left if there exists a function on \( Y \times X \) to \( X \), (whose value at \( gx \) is denoted by \( gx \) such that

(i) \( 1x = x \) for all \( x \in X \),

(ii) \( g_2(g_1x) = (g_2g_1)x \).

Comment: (i) and (ii) imply \( g \in Y \) is 1-1 on \( X \) to \( X \). To see this, suppose \( gx_1 = gx_2 = y \). Then \( g^{-1}(gx_1) = g^{-1}(gx_2) = g^{-1}y \). Using (i) and (ii), we have \( x_1 = x_2 = g^{-1}y \).

Definition: Let \( Y \) operate on \( X \). \( Y \) operates transitively on \( X \) if for every \( x_1, x_2 \in X \), there exists a \( g \in Y \) such that \( gx_1 = x_2 \).

Comment: If \( X \) is a linear space, \( Y \ell(n) \) operates transitively on \( X \) - \( \{0\} \). Also, if \( \Gamma: \mathbb{X} \to X \) and \( \Gamma \) is positive definite and symmetric, then \( \mathbf{SO}(\Gamma) \) operates transitively on the set
\[
\{x: <r^{-1}x, x> = r^2 \} \quad \text{the circle of radius} \ r \ \text{centered at} \ 0.
\]

Let \( Y \) be a group and \( H \) a subgroup of \( Y \). The space \( Y/H \) is defined to be the set of elements of the form \( g_1H = \{g_1h \mid h \in H\} \) where \( g_1 \in Y \). Note that \( g_1H = g_2H \) if and only if there exists \( h \in H \) such that \( g_1 = g_2h \) -- that is, \( g_1 \) and \( g_2 \) are equivalent modulo \( H \). Define \( Y \) operating on \( Y/H \) by \( g(g_1H) = (gg_1)H \).

Must now check that the definition of the operation of \( Y \) on \( Y/H \) does not depend on the representation of the element \( g_1H \). Thus, suppose \( g_1H = g_2H \) and consider \( g \in Y \). Must show that \( g(g_1H) = g(g_2H) \). There exists \( h \in H \) such that \( g_1 = g_2h \).
Thus \( g(g_1 H) = g(g_2 h H) = g(g_2 H) \) since \( h H = H \). Thus, for each subgroup \( H \) of \( Y \), we have defined \( Y \) operating on \( Y / H \). Note that \( Y \) operates transitively on \( Y / H \) since \( g_1 H = g_1 g_2^{-1}(g_2 H) \) for all \( g_1, g_2 \in Y \).

Consider now a group \( Y \) operating transitively on \( X \). Fix \( x_0 \in X \) and let \( H_{x_0} = \{ g \mid gx_0 = x_0 \} \). \( H_{x_0} \) is clearly a subgroup of \( Y \).

Define the mapping \( \tau_{x_0} : X \to Y / H_{x_0} \) by \( \tau_{x_0} = \{ g | gx_0 = x \} \in Y / H_{x_0} \).

Since \( Y \) acts transitively on \( X \), there exists \( g_1 \) such that \( g_1 x_0 = x \). Thus \( \tau_{x_0} x = \{ g | gx_0 = x \} = \{ g | gx_0 = g_1 x_0 \} = g_1 \{ g | gx_0 = x_0 \} = g_1 H_{x_0} \), where \( g_1 x_0 = x \). To show \( \tau_{x_0} \) is onto, consider \( g_1 H_{x_0} \in Y / H_{x_0} \).

Then \( \tau_{x_0} g_1 x_0 = g_1 H_{x_0} \) so that \( \tau_{x_0} \) is onto. Also, since each \( g \in Y \) is 1-1 on \( X \) to \( X \), \( \tau_{x_0} \) is obviously 1-1.

For \( g_1 \in Y \), \( g_1(\tau_{x_0} x) = g_1 \{ g | gx_0 = x \} = \{ g | gx_0 = g_1 x \} = \tau_{x_0}(g_1 x) \).

Hence to study \( Y \) operating transitively on \( X \), it is equivalent to study \( Y \) operating on \( Y / H_{x_0} \).

Example: Let \( X \) be a linear space and \( Y \) the affine group operating on \( X \). Choose \( x_0 = 0 \in X \) so that \( H_{x_0} = Y \ell(n) \). Obviously, \( Y / Y \ell(n) \) is identifiable with the space \( X \) since \( Y / Y \ell(n) \) is just the translation group.

Example: Let \( X \) be a linear space and \( \mathcal{Q}_X \) be the set of positive definite symmetric operators on \( X' \) to \( X \). Consider \( Y \ell(n) \) operating on \( \mathcal{Q}_X \) by \( g(\Gamma) = g\Gamma g' \) for \( \Gamma \in \mathcal{Q}_X \) and \( g \in Y \ell(n) \). Fix \( \Gamma_0 \in \mathcal{Q}_X \) so that \( H_{\Gamma_0} = Y \ell(\Gamma_0) \), since \( H_{\Gamma_0} = \{ g | g\Gamma_0 g' = \Gamma_0 \} \).

Thus \( \mathcal{Q}_X \) can be identified with \( Y \ell(n)/Y \ell(\Gamma_0) \).
If $\mathcal{Y}$ operates on $\mathcal{X}$ on the left, we write $(\mathcal{Y}, \mathcal{X})$. Consider $(\mathcal{Y}, \mathcal{X})$ and $(\mathcal{Y}, \mathcal{Y})$ and let $\mathcal{X} \times \mathcal{Y} = \{(x, y) | x \in \mathcal{X} \text{ and } y \in \mathcal{Y}\}$. Then $g_{\mathcal{X} \times \mathcal{Y}}$ operating on $\mathcal{X} \times \mathcal{Y}$ is defined by $g_{\mathcal{X} \times \mathcal{Y}}(x, y) = (g_x, g_y)$. Let $\mathcal{P} \mathcal{X}$ be the set of all subsets of $\mathcal{X}$. Define $g_{\mathcal{P} \mathcal{X}}$ operating on $\mathcal{P} \mathcal{X}$ by

$$g_{\mathcal{P} \mathcal{X}}(A) = \{g_{\mathcal{X}}x | x \in A \} \in \mathcal{P} \mathcal{X}$$

for $A \in \mathcal{P} \mathcal{X}$. Also, let $\mathcal{Y}^\mathcal{X}$ be the space of all functions on $\mathcal{X}$ to $\mathcal{Y}$ and define $g_{\mathcal{Y}^\mathcal{X}}$ operating on $\mathcal{Y}^\mathcal{X}$ by $(g_{\mathcal{Y}^\mathcal{X}}f)(x) = g_y(f(g_{\mathcal{X}}^{-1}x))$. The reason for this definition is so that $(g_1g_2)_{\mathcal{Y}} = g_1g_2_{\mathcal{Y}}$.

Now consider $(\mathcal{Y}, \mathcal{X})$ and let $\mathcal{Y}$ be an arbitrary set. Let $\mathcal{A}$ be the set of functions on $\mathcal{X}$ onto $\mathcal{Y}$ such that if $f(x_1) = f(x_2)$ then for all $g_\mathcal{X} \in \mathcal{A}$, $f(g_\mathcal{X}x_1) = f(g_\mathcal{X}x_2)$. On $\mathcal{Y}$, define $g_{\mathcal{Y}}$ by $g_{\mathcal{Y}}(f(x)) = f(g_\mathcal{X}x)$. Since $f$ is not necessarily 1-1, it is necessary to check that the definition of $g_{\mathcal{Y}}$ is independent of the representation of an element $z \in \mathcal{Y}$. Suppose $z = f(x_1) = f(x_2)$. Then $g_{\mathcal{Y}}(z) = g_{\mathcal{Y}}(f(x_1)) = f(g_\mathcal{X}x_1) = f(g_\mathcal{X}x_2) = g_{\mathcal{Y}}(f(x_2)) = g_{\mathcal{Y}}(z)$ so that $g_{\mathcal{Y}}$ is well defined. Note that $(g_1g_2)_{\mathcal{Y}} = g_1g_2_{\mathcal{Y}}$ and that $g_{\mathcal{Y}}$ depends on the choice of $f \in \mathcal{A}$.

Example: Let $\mathcal{X}$ be a linear space and consider $\mathcal{Y} \mathcal{X}(n)$ operating on $\mathcal{X}$. For the dual space $\mathcal{X}'$ of $\mathcal{X}$, define $g_{\mathcal{X}'} = g_\mathcal{X}^{-1}$ -- again, the reason for this definition is so that $(g_1g_2)_{\mathcal{X}'} = g_1g_2_{\mathcal{X}'}$. Now consider $\mathcal{X}'_x$ -- the space of all positive definite symmetric operators.
on \( \mathcal{X} \) to \( \mathcal{X} \) and let \( \mathcal{G} \Sigma(n) \) be the general linear group of \( \mathcal{X} \).

From the discussion and definitions above, the natural way to define
\( g_\mathcal{X} \) on \( \mathcal{G} \mathcal{X} \) is

\[
g_\mathcal{X} \Gamma = g\Gamma(g^{-1})^{-1} = g\Gamma g'
\]

-- the definition we have used previously.

Invariance in statistical problems is now considered. Let \((\mathcal{X}, \mathcal{A})\)
be a measurable space and \( \mathcal{R} \) be a set of points -- \( \mathcal{R} = \{\theta\} \). Consider a functions \( P \) on \( \mathcal{R} \) to the set of probability measures on
\((\mathcal{X}, \mathcal{A})\) whose value at \( \theta \) is denoted by \( P_\theta \). For this part of a statistical problem to be invariant under a group \( \mathcal{G} \), assume:

(i) \( g_\mathcal{X} : \mathcal{X} \rightarrow \mathcal{X} \) is \((\mathcal{A}, \mathcal{A})\) measurable

(ii) \( g_\mathcal{R} : \mathcal{R} \rightarrow \mathcal{R} \) is such that: if \( X \sim P_\theta \) then \( g_\mathcal{X}X \sim P_\theta g_\mathcal{R} \theta \).

An equivalent way of writing (ii) is \( P_\theta (g_\mathcal{X}X \in A) = P_\theta (g_\mathcal{R} \theta (X \in A) \setminus A \).

or \( P_\theta (g_\mathcal{X}^{-1}(A)) = P_\theta (g_\mathcal{R}^{-1}(A)) \setminus A \). Setting \( B = g_\mathcal{X}^{-1}A \), we can also write

(ii) as \( P_\theta (B) = P_\theta (g_\mathcal{R}^{-1}g_\mathcal{X}^{-1}B) \) for \( B \in \mathcal{A} \). If we define \( \tilde{P} P(B|\theta) = P(g_\mathcal{X}^{-1}B|g_\mathcal{R}^{-1}\theta) \),
then condition (ii) becomes \( \tilde{P} P = P \).

Comment: Suppose \( \theta_1 \neq \theta_2 \) implies \( P_{\theta_1} \neq P_{\theta_2} \). Then \( g_\mathcal{X} \) determines the
transformation \( g_\mathcal{R} \) and in this case the correspondence between \( g_\mathcal{X} \)
and \( g_\mathcal{R} \) is a homomorphism. In statistical problems, we usually write
\( g \) for \( g_\mathcal{X} \) and \( \tilde{g} \) for \( g_\mathcal{R} \).

Very often in statistical problems there exists a measure \( \lambda \) on \( \mathcal{A} \)
such that \( P_\theta \ll \lambda \) for all \( \theta \in \mathcal{R} \) so that there is a density function
\( p_\theta \) and

13
\[ P_\theta(A) = \int_A P_\theta(x) d\lambda(x). \]

Also, in a great many cases of interest, it is possible to choose the measure \( \lambda \) so that \( \lambda(A) = \lambda(gA) \) for all \( A \in \mathcal{B}, g \in \mathcal{G} \) -- that is, the measure \( \lambda \) is invariant under the group \( \mathcal{G} \). In this case, the condition of invariance becomes

\[ p_{\mathcal{G}}(gx) = p_\theta(x). \]

It is now convenient to give some examples of invariance measures on sets \( \mathcal{X} \) which are acted upon transitively by a group \( \mathcal{G} \).

(i) Suppose \( \mathcal{G} \) is a subgroup of the permutation group of a finite set \( \mathcal{X} \). If \( A \) is a subset of \( \mathcal{X} \), let \( \lambda(A) \) be the number of elements in the set \( A \). Then \( \lambda \) is an invariant measure on \( \mathcal{X} \) under \( \mathcal{G} \) and \( \lambda \) is unique up to a positive multiplicative constant, for \( \mathcal{G} \) transitive on \( \mathcal{X} \).

(ii) Let \( \mathcal{X} \) be Euclidean \( n \) space and let \( \mathcal{G} \) be the group of translations: \( g \in \mathcal{G} \) implies \( g_x x = x + x_1 \). It is clear that \( \mathcal{G} \) acts transitively on \( \mathcal{X} \) and that \( n \)-dimensional Lebesgue measure \( \lambda \), is invariant. Again, \( \lambda \) is unique up to multiplication by a positive constant.

(iii) Consider the group \( \mathcal{G} \) \( \ell(n) \) operating on itself. Since an element \( g \) operating on \( \mathcal{G} \ell(n) \) (as a linear space) is linear, it has a determinant with respect to this linear operation. It follows easily that the determinant of \( g \) operating on \( \mathcal{G} \ell(n) \) is \( (\det g)^n \). Now, consider the measure
$$\text{d} \lambda(y) = \frac{\prod \text{d} y_{i j}}{(\text{det } y)^n}.$$  

Then

$$\text{d} \lambda(g y) = \frac{(\text{det } g)^n \prod \text{d} y_{i j}}{(\text{det } g y)^n} = \text{d} \lambda(y)$$

so that $\lambda$ is invariant. Also this measure is the unique invariant measure up to a positive multiplicative constant.

(iv) Let $\mathcal{S}$ be the space of positive definite $p \times p$ symmetric operators and consider $\mathcal{G} \ell(n)$ operating on $\mathcal{S}$ by $g(s) = g s g^\top$. This transformation is linear and has determinant $(\text{det } g)^{p+1}$. Consider the measure $\lambda$ on $\mathcal{S}$ given by

$$\text{d} \lambda(s) = \frac{\prod \text{d} x_{i j}}{i \times j \frac{p+1}{2} (\text{det } s)^{\frac{p+1}{2}}}.$$  

Then

$$\text{d} \lambda(g s g^\top) = \frac{(\text{det } g)^{p+1} \prod \text{d} s_{i j}}{i \times j \frac{p+1}{2} (\text{det } g s g^\top)^{\frac{p+1}{2}}} = \text{d} \lambda(s)$$

so that $\lambda$ is invariant under $\mathcal{G} \ell(p)$. Also note that $\lambda$ is unique up to a positive multiplicative constant.

Before proceeding further, we state a result concerning the transformation of variables in abstract integrals. Let $(\mathcal{Z}, \mathcal{B}, \mu)$ be a measure space and let $(\mathcal{Y}, \mathcal{C})$ be a measurable space. Consider a
measurable function \( \varphi \) on \( \mathcal{X} \) to \( \mathcal{Y} \) and define the measure \( \nu \) on \( (\mathcal{Y}, \mathcal{C}) \) by

\[
\nu(A) = \mu(\varphi^{-1}(A))
\]

for all \( A \in \mathcal{C} \).

**Theorem:** If \( f \) is a real valued measurable function on \( \mathcal{X} \) such that \( f(x) = g(\varphi(x)) \) and if \( f \) is integrable, then

\[
\int_{\mathcal{X}} f(x) d\mu(x) = \int_{\mathcal{Y}} g(\varphi(x)) d\mu(x) = \int_{\mathcal{Y}} g(y) d\nu(y).
\]


**Derivation of the Wishart Distribution**

Let \( X_1, \ldots, X_n \) be independent, each \( N(0, \Gamma) \) where \( \Gamma \) is a \( p \times p \) positive definite symmetric operator and assume \( n \geq p \). We want to find the distribution of

\[
S = \sum_{i=1}^{n} X_i X_i'.
\]

The joint density of \( (X_1, \ldots, X_n) \) is

\[
p(x_1, \ldots, x_n | \Gamma) = c |\Gamma|^{-\frac{n}{2}} e^{-\frac{1}{2} \text{tr} \Gamma^{-1} \sum_{i=1}^{n} x_i x_i'}
\]

where \( c \) denotes a constant and \( |\Gamma| = \det(\Gamma) \).

Using the above theorem, we obtain
\[ P_\Gamma(S \in A) = c |r^{-1}|^{1/2} \int_{\sum_{x_i x_i' \in A}} e^{-\frac{1}{2} \text{tr} r^{-1} \sum_{i=1}^{n} x_i x_i'} \prod_{i=1}^{n} dx_i \]

\[ = c \int_{s \in A} |r^{-1}|^{1/2} e^{-\frac{1}{2} \text{tr} r^{-1} s} dm(s) \]

where \( m \) is the measure corresponding to the measure \( \nu \) of the above theorem. Define a new measure \( m^* \) by

\[ dm^*(s) = \frac{dm(s)}{|s|^2} \]

so that

\[ P_\Gamma(S \in A) = c \int_{s \in A} |r^{-1} s|^{1/2} e^{-\frac{1}{2} \text{tr} r^{-1} s} dm^*(s). \]

To find the distribution of \( S \) it suffices to find the measure \( m^* \). For this purpose, write \( \Gamma = \alpha \alpha' \) for \( \alpha \in \mathcal{D}(p) \). However from properties of the Wishart distribution we know that

\[ P_{\alpha \alpha'}(S \in A) = P_{\Gamma}(\alpha \alpha' \in A). \]

But

\[ P_{\alpha \alpha'}(S \in A) = c \int_{s \in A} |(\alpha \alpha')^{-1} s|^{1/2} e^{-\frac{1}{2} \text{tr} (\alpha \alpha')^{-1} s} dm^*(s), \]

and
\[ P_1(\tilde{\alpha} \tilde{\omega}' \in A) = c \int_{\tilde{\alpha} \tilde{\omega}' \in A} \left| s \right|^2 e^{-\frac{1}{2} \text{tr} \tilde{s}} \, \text{dm}^*(\tilde{s}) \]

\[ = c \int_{s \in A} |(\alpha \omega')_s| \left| s \right|^2 e^{-\frac{1}{2} \text{tr}(\alpha \omega')^{-1}_s} \, \text{dm}^*(\alpha^{-1}_s \omega'^{-1}_s). \]

Since this holds for all sets \( A \), we must have that for all \( g \in \mathcal{G} \ell(p) \)

\[ \text{dm}^*(s) = \text{dm}^*(g s g'). \]

By the uniqueness of the invariant measure under \( \mathcal{G} \ell(p) \), we have

\[ \prod_{i \leq j} |s|^{p_i} \left( \frac{i!}{j!} \right) \left( \frac{1}{p+1} \right)^{p+1} \]

Thus the density of the Wishart distribution is

\[ P_\Sigma | r(s | \tilde{r}) = c_2 |\tilde{r}^{-1}_s| \frac{n}{2} e^{-\frac{1}{2} \text{tr} \tilde{r}^{-1}_s} \cdot \frac{1}{|s|^{p+1}} \]

where the density is with respect to Lebesgue measure \( \prod_{i \leq j} ds_{i,j} \).

A form which will also be useful is

\[ P_\Sigma | r(s | \tilde{r}) d\lambda(s) = c_2 |\tilde{r}^{-1}_s| \frac{n}{2} e^{-\frac{1}{2} \text{tr} \tilde{r}^{-1}_s} \, \text{dm}^*(s) \]

where \( d\lambda(s) = \prod_{i \leq j} ds_{i,j}, \) \( c_2 \) is a constant, and

\[ \text{dm}^*(s) = \frac{\prod_{i \leq j} ds_{i,j}}{|s|^{p+1}} \]
is the invariant measure on the space of $p \times p$ positive definite symmetric matrices operated by $\mathcal{G}/\ell(p)$.

**Section 3.**

We first discuss orthogonal projections. Let $\mathcal{X}$ be a real linear space of finite dimension $n$ and let $\mathcal{X}'$ be the dual space of $\mathcal{X}$. Let $\Gamma$ be a positive definite symmetric linear transformation on $\mathcal{X}'$ to $\mathcal{X}$. Now $\langle \xi_1, \xi_2 \rangle_{\mathcal{X}} = \langle \xi_1, \Gamma \xi_2 \rangle_{\mathcal{X}}$ in $\mathcal{X}$. Let $\mathcal{Y}$ be a real finite dimensional linear space such that the dimension of $\mathcal{Y}$ is less than or equal to that of $\mathcal{X}$ and let $\mathcal{Y}'$ be the dual space of $\mathcal{Y}$. Consider a 1-1 linear mapping, $\iota$, of $\mathcal{Y}$ into $\mathcal{X}$. We want to define the orthogonal projection onto the image of $\iota$ in $\mathcal{X}$ (orthogonal projection with respect to the inner product defined above):

![Diagram showing orthogonal projection]

The mapping $\iota^+_{\Gamma}$ is defined by

$$\iota^+_{\Gamma} = (\iota' \Gamma^{-1} \iota)^{-1} \iota' \Gamma^{-1}.$$

Recall the following.
Definition: If $\mathcal{X}$ and $\mathcal{Y}$ are two linear spaces and $\alpha$ is a linear transformation on $\mathcal{X}$ to $\mathcal{Y}$, then a linear transformation $\alpha^-$ on $\mathcal{Y}$ to $\mathcal{X}$ is a **pseudo inverse** of $\alpha$ if

$$\alpha \alpha^- \alpha = \alpha.$$ 

**Comment:** If $\alpha$ is 1-1, then $\alpha^- \alpha = I_\mathcal{X}$ and if $\alpha$ is onto then

$$\alpha \alpha^- = I_\mathcal{Y}.$$ 

With this definition, it is clear that $\iota^+_\Gamma$ is a pseudo inverse of $\iota$.

**Definition:** The **orthogonal projection** onto the image of $\iota$, denoted by $P_{\text{Im} \iota}$, is defined by

$$P_{\text{Im} \iota} = \iota \iota^+_\Gamma.$$ 

Now, consider $\mathcal{Y}$ as a linear subspace of $\mathcal{X}$ and let $\mathcal{Z}$ be the quotient space $\mathcal{X}/\mathcal{Y}$. Denote by $\pi$ the natural homomorphism of $\mathcal{X}$ onto $\mathcal{Z}$, and let $\iota$ be the inclusion mapping of $\mathcal{Y}$ to $\mathcal{X}$.

\[
\begin{array}{ccc}
\mathcal{Z} & \xrightarrow{\pi} & \mathcal{X} \\
\downarrow{\pi \Gamma} & & \downarrow{\iota} \\
\mathcal{Z} & \xleftarrow{\pi} & \mathcal{Y}
\end{array}
\quad
\begin{array}{ccc}
\mathcal{X} & \xleftarrow{\iota} & \mathcal{Y} \\
\downarrow{\iota \Gamma} & & \downarrow{\iota \Gamma^{-1}} \\
\mathcal{X} & \xrightarrow{\iota \Gamma} & \mathcal{Y}
\end{array}
\]
The mapping $\pi_\Gamma^+$ is defined by

$$\pi_\Gamma^+ = \Gamma \pi' (\pi \Gamma)^{-1}.$$ 

Also the mapping $P_{\text{Im}(\pi_\Gamma^+)} = \pi_\Gamma^+ \pi$ is the orthogonal projection on the image of $\pi_\Gamma^+$. Note that $\pi_\Gamma^+$ is a pseudo inverse of $\pi$. Since the image of $\pi_\Gamma^+$ is the orthogonal complement of the image of $\Gamma$, the two orthogonal projections $\pi_\Gamma^+ \pi$ and $\iota_\Gamma^+$ are complementary -- that is

$$\pi_\Gamma^+ \pi + \iota_\Gamma^+ = I_Z$$

and

$$(\pi_\Gamma^+ \pi)(\iota_\Gamma^+) = (\iota_\Gamma^+)(\pi_\Gamma^+ \pi) = 0.$$ 

**Example: Multivariate Regression**

Let $X$ be $N(x^0, \Gamma)$ and partition $X$ as $(Y, Z)$. Correspondingly, partition $x^0$ and $\Gamma$ as

$$x^0 = \begin{pmatrix} 0 \\ Y \\ 0 \\ 0 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}.$$ 

Then the conditional distribution of $Y$ given $Z$ is

$$N(y^0 + \Gamma_{12} \Gamma_{22}^{-1} (z - z^0), \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}).$$
To discuss this in terms of orthogonal projections, let \( \dim \mathcal{X} = p \) and \( \dim \mathcal{Y} = q \). Choose the matrix of \( \pi \) to be \((0, I_q)\) so that \( \pi(\mathcal{Y}) = \mathcal{Z} \). For convenience, assume \( \mathcal{E}(\mathcal{X}) = 0 \). Then

\[
\pi^+ = \pi \pi' (\pi \pi')^{-1} = \begin{pmatrix}
\Gamma_{12} & \Gamma_{12}^{-1} \\
\Gamma_{22} & 1
\end{pmatrix}
\]

so that

\[
\mathcal{E}(\mathcal{X}|\mathcal{Z}) = \pi^+ \mathcal{Z} = \pi^+ \pi \mathcal{X}.
\]

Thus, the conditional expectation of \( \mathcal{X} \) given \( \mathcal{Z} \) is just the projection of \( \mathcal{X} \) on the image of \( \pi^+ \). Also,

\[
\pi^+ \pi \pi^+ = \begin{pmatrix}
\Gamma_{12} & \Gamma_{12}^{-1} \\
1 & \Gamma_{22}
\end{pmatrix}
\Gamma_{22} (\Gamma_{22}^{-1}, 1)
\begin{pmatrix}
\Gamma_{12} & \Gamma_{12}^{-1} & \Gamma_{12} & \Gamma_{12} \\
\Gamma_{22} & \Gamma_{21} & \Gamma_{22} & \Gamma_{22}
\end{pmatrix}
\]

so that

\[
\text{Cov}(\mathcal{X}|\mathcal{Z}) = \Gamma - \pi^+ \pi \pi^+ = \begin{pmatrix}
\Gamma_{11} - \Gamma_{12} & \Gamma_{12}^{-1} & \Gamma_{21} & 0 \\
0 & \Gamma_{22} & \Gamma_{21} & 0
\end{pmatrix}.
\]

**Tensor Products**

If \( \mathcal{X} \) and \( \mathcal{Y} \) are two linear spaces, let \( \mathcal{L}(\mathcal{X}, \mathcal{Y}) \) denote the linear space of all linear transformations on \( \mathcal{X} \) to \( \mathcal{Y} \).

**Definition:** The tensor product of two linear spaces \( \mathcal{X} \) and \( \mathcal{Y} \), denoted by \( \mathcal{X} \otimes \mathcal{Y} \) is defined to be \( \mathcal{L}(\mathcal{Y}', \mathcal{X}) \) where \( \mathcal{Y}' \) is the dual space of \( \mathcal{Y} \).
If \( \dim \mathcal{X} = p \) and \( \dim \mathcal{Y} = q \), then \( \mathcal{X} \otimes \mathcal{Y} \) is of dimension \( pq \) since \( \mathcal{X} \otimes \mathcal{Y} = \mathcal{Z}(\eta', \mathcal{X}) \) which is of dimension \( pq \). We can think of the elements of \( \mathcal{X} \otimes \mathcal{Y} \) as \( p \times q \) matrices which operate on the right of column vectors in \( \mathcal{Y}' \).

For \( x \in \mathcal{X} \) and \( y \in \mathcal{Y} \), define the tensor product of \( x \) and \( y \) by \( (x \otimes y)\eta = x < \eta, y > \) for \( \eta \in \mathcal{Y}' \). Of course, in coordinate notation, \( x \otimes y \) is just \( xy' \). If \( x^1, \ldots, x^p \) is a basis for \( \mathcal{X} \) and \( y^1, \ldots, y^q \) is a basis for \( \mathcal{Y} \), it is easy to show that \( \{ x^i \otimes y^j \} \) is a basis for \( \mathcal{X} \otimes \mathcal{Y} \). Note that \( x \otimes y \) is linear in \( x \) and \( y \) and \( (x \otimes y)' = y \otimes x \).

Now, consider \( \alpha \in \mathcal{X} \otimes \mathcal{Y} \) and suppose \( \beta: \mathcal{X} \rightarrow \mathcal{X}_2 \) and \( \gamma: \mathcal{Y}_1 \rightarrow \mathcal{Y}_2 \) where \( \beta \) and \( \gamma \) are linear transformations. We want to define the tensor product of the two linear transformations \( \beta \) and \( \gamma \), say \( \beta \otimes \gamma \), so that

\[
\beta \otimes \gamma: \mathcal{X}_1 \otimes \mathcal{Y}_1 \rightarrow \mathcal{X}_2 \otimes \mathcal{Y}_2.
\]

Thus, consider the following diagram

![Diagram](image_url)

Hence, we define \( \beta \otimes \gamma \) by
$$(\beta \boxtimes \gamma)\alpha = \beta \alpha \gamma'$$

where

$$\beta \boxtimes \gamma : \mathcal{A}_1 \times \mathcal{B}_1 \rightarrow \mathcal{A}_2 \boxtimes \mathcal{B}_2$$

and $\alpha \in \mathcal{A}_1 \boxtimes \mathcal{B}_1$. From the above definition, it is clear that

$$(\beta \boxtimes \gamma)' = \beta \boxtimes \gamma'.$$

Now, consider the following:

From this diagram, we see that

$$(\beta_2 \boxtimes \gamma_2)(\beta_1 \boxtimes \gamma_1) = (\beta_2 \beta_1 \boxtimes \gamma_2 \gamma_1).$$

It then follows directly that if $\beta$ maps $\mathcal{A}_1$ into $\mathcal{A}_1$ and $\gamma$ maps $\mathcal{B}_1$ into $\mathcal{B}_1$ such that $\beta$ and $\gamma$ are invertible, then

$$(\beta \boxtimes \gamma)^{-1} = \beta^{-1} \boxtimes \gamma^{-1}.$$
Also note that if $x \in \mathbb{F}^1$ and $y \in \mathbb{Y}^1$, then

$$(\beta \times \gamma)(x \times y) = (\beta x) \times (\gamma y)$$

since $(\beta \times \gamma)(x \times y) = \beta(x \times y)\gamma = \beta x y' \gamma' = (\beta x) \times (\gamma y)$.

We now want to discuss the tensor product $\beta \times \gamma$ in terms of matrices and vectors, that is, in coordinate notation. Suppose that $\beta$ is $n \times m$ and $\gamma$ is $l \times k$ so that $\beta \times \gamma$ acts on matrices $\alpha$ of dimension $m \times k$ and $\beta \gamma' \alpha = \delta$ is $n \times l$. Consider the vector $v_1$ of dimension $mk \times 1$ defined as $v_1 = (\alpha_{11} \alpha_{21}, \ldots, \alpha_{m1}, \alpha_{12}, \ldots, \alpha_{m2}, \ldots, \alpha_{1k}, \ldots, \alpha_{mk})$ and define the vector $v_2$ as $v_2 = (\delta_{11}, \delta_{12}, \ldots, \delta_{n1}, \delta_{12}, \ldots, \delta_{n2}, \ldots, \delta_{1l}, \ldots, \delta_{nk})$ which is $n \times l$. Think of $\beta \times \gamma$ as a matrix $n \times mk$ operating on the left of the column vector $v_1$ to give the column vector $v_2$, where $\delta$ is defined by $(\beta \times \gamma)\alpha = \delta$ and we want to determine the matrix of $\beta \times \gamma$ so that $[\beta \times \gamma]v_1 = v_2$. Elementary computations show that the matrix of $\beta \times \gamma$ is given by

$$[\beta \times \gamma] = \begin{bmatrix}
\gamma_{11} \beta & \gamma_{12} \beta & \cdots & \cdots & \gamma_{1k} \beta \\
\gamma_{21} \beta & \gamma_{22} \beta & \cdots & \cdots & \gamma_{2k} \beta \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\gamma_{l1} & \gamma_{l2} \beta & \cdots & \cdots & \gamma_{lk} \beta
\end{bmatrix}$$

which is an $nl \times mk$ matrix. This matrix is commonly called the Kronecker product of the two matrices $\beta$ and $\gamma$.

Suppose $A$ is $n \times n$ and $B$ is $m \times m$ are two matrices. Let $A \times B$ denote the Kronecker product of $A$ and $B$. If $\lambda_1, \ldots, \lambda_n$ are
the eigenvalues of $A$ and $\mu_1, \ldots, \mu_m$ are the eigenvalues of $B$, then
the eigenvalues of $A \otimes B$ consist of the set \( \{\mu_i \lambda_j : i = 1, \ldots, m; j = 1, \ldots, n\} \).
From this it follows directly that \( \det(A \otimes B) = (\det A)^m(\det B)^n \) and
\( \text{tr}(A \otimes B) = (\text{tr} A)(\text{tr} B) \). Also, if $A$ and $B$ are non-singular, then
\( (A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \) as was shown previously.

Application to Multivariate Regression

Consider $Y_i$, $i = 1, \ldots, n$ independent random column vectors such
that $Y_i \sim N(\alpha z_i, \Gamma)$ where $\alpha$ is $p \times q$, $z_i$ is $q \times 1$. Let
$Y = (Y_1, Y_2, \ldots, Y_n)$ be the $p \times n$ matrix whose columns are independent.
Then we have

$$Y \sim N(\alpha z, \Gamma \otimes I)$$

where $z = (z_1, \ldots, z_n)$ is $q \times n$ and we assume rank $z = q$.

Note that $\Gamma \otimes I$ has the form

$$\Gamma \otimes I = \begin{bmatrix}
\Gamma & 0 & \ldots & 0 \\
0 & \Gamma & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & \Gamma
\end{bmatrix}$$

which says that any two columns of $Y$ are independent and the covariance
of any column of $Y$ is $\Gamma$. 

26
Proposition 1: Let \( Y \) be \( p \times n \) such that \( Y \sim N(y^0, A \otimes B) \) where \( A \) is \( p \times p \), \( B \) is \( n \times n \), and \( y^0 \) is \( p \times n \). If \( C \) is \( q \times p \) and \( D \) is \( n \times k \), then

\[
\text{CYD} \sim N(\text{Cy}^{-1}D, (CAC') \otimes (DED')).
\]

Proof: This follows easily from the definition of \( A \otimes B \).

Now, the joint density of \( Y_1, \ldots, Y_n \) is

\[
p(Y) = \prod_{i=1}^{n} p(Y_i) = \prod_{i=1}^{n} \left( \frac{1}{\sqrt{2\pi}} \right)^p |\Gamma|^{-\frac{1}{2}} e^{-\frac{1}{2} \text{tr} \Gamma^{-1}(Y_i - \alpha z_i)(Y_i - \alpha z_i)'}
\]

\[
= \left( \frac{1}{\sqrt{2\pi}} \right)^{np} \frac{n}{2} \text{tr} \Gamma^{-1} \sum_{i=1}^{n} (Y_i - \alpha z_i)(Y_i - \alpha z_i)'
\]

However, \( |\Gamma|^{-\frac{n}{2}} = |\Gamma^{-1}I|^{-\frac{n}{2}} \) and \( \sum_{i=1}^{n} (Y_i - \alpha z_i)(Y_i - \alpha z_i)' = (Y - \alpha z)(Y - \alpha z)' \)

so that

\[
p(Y) = \left( \frac{1}{\sqrt{2\pi}} \right)^{\frac{np}{2}} |\Gamma^{-1}I|^{-\frac{n}{2}} e^{-\frac{1}{2} \text{tr} (\Gamma^{-1}I)^{-1}(Y - \alpha z)(Y - \alpha z)'}
\]

\[
= \left( \frac{1}{\sqrt{2\pi}} \right)^{\frac{np}{2}} |\Gamma|^{-\frac{n}{2}} e^{-\frac{1}{2} \text{tr} \Gamma^{-1}(Y - \alpha z)(Y - \alpha z)'}
\]

To find the maximum likelihood estimate of \( \alpha \), we want to minimize

\[
\text{tr} \Gamma^{-1}(Y - \alpha z)(Y - \alpha z)'.
\]

Now
\[ \frac{\partial}{\partial \alpha} \text{tr} \Gamma^{-1}(Y-\alpha z)(Y-\alpha z)' = -\text{tr} \Gamma^{-1}z(Y-\alpha z)' \].

Setting this equal to 0, we have \( \text{tr} \Gamma^{-1}z(Y'-z'\alpha') = 0 \) so that 
\( z(Y'-z'\alpha') = 0 \) or \( \hat{\alpha} = Yz'(zz')^{-1} \).

From proposition 1, we have
\[ \hat{\alpha} \sim N(\alpha, \Gamma \otimes (zz')^{-1}). \]

Now, \( Y - \hat{\alpha} z = Y(I-z'(zz')^{-1}z) \) and

\[ Y(I-z'(zz')^{-1}z) \sim N(0, \Gamma \otimes (I-z'(zz')^{-1}z)) \]

since \( I - z'(zz')^{-1}z \) is a projection. Also, \( (I-z'(zz')^{-1}z)z'(zz')^{-1} = 0 \) which implies that \( \hat{\alpha} \) and \( Y(I-z'(zz')^{-1}z) \) are independent. Since 
\( z'(zz')^{-1}z \) is a projection of rank \( q \), \( I - z'(zz')^{-1}z \) is a projection of rank \( n - q \) which implies that
\[ (Y-\hat{\alpha} z)(Y-\hat{\alpha} z)' \sim W(n-q, \Gamma). \]

Comment: Consider \( X \sim N(x^0, \Gamma \otimes I) \) where \( X \) is \( p \times n \) and partition 
\( X \) as \( X = (Y, Z) \). Similarly, partition \( x^0 = \begin{pmatrix} y^0 \\ z^0 \end{pmatrix} \) and 
\( \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \)

as \( X \) is partitioned. Note that \( x^0 \) is a \( p \times n \) matrix. From our previous results it follows easily that the conditional distribution of 
\( Y \) given \( Z \) is
\[ N(y^0 + r_{12}^{-1} (z - z^0), \quad (r_{11}^{-1} - r_{12}^{-1} r_{22}^{-1} r_{21}) \otimes I). \]

**Some Distribution Theory**

Let \( S \) be a \( p \times p \) Wishart matrix with \( n \) degrees of freedom and covariance \( \Gamma \). Then

\[
p(S) = C|S|^{\frac{n-p-1}{2}} |\Gamma|^{-\frac{n}{2}} e^{-\frac{1}{2} \text{tr} \Gamma^{-1} S}.
\]

Let \( S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \) where \( S_{11} \) is \( q \times q \) and partition \( \Gamma \) is

\[
\Gamma = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}.
\]

Also, let \( \Lambda = r^{-1} \) and write \( \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \).

Consider the new random variables defined by

\[
W = S_{11} - S_{12} S_{22}^{-1} S_{21}
\]

\[
L = S_{12} S_{22}^{-\frac{1}{2}}
\]

\[
V = S_{22}.
\]

We want to find the joint distribution of \( W, L \) and \( V \). The Jacobian of the above transformation is \( |S_{22}|^{q/2} \). Solving the above equations for \( S_{11}, S_{12} \) and \( S_{22} \) yields

\[
S_{11} = W + LL, \quad S_{12} = LV^{\frac{1}{2}}, \quad S_{22} = V.
\]
Also, write

$$|s| = |s_{22}| |s_{11} - s_{12} s_{22}^{-1} s_{21}|, \quad |\Gamma| = |\Gamma_{22}| |\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}|$$

and

$$\text{tr} \Gamma^{-1} s = \text{tr} AS = \text{tr} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$$

$$= \text{tr} \begin{bmatrix} A_{11}s_{11} + A_{12}s_{21} & A_{11}s_{12} + A_{12}s_{22} \\ A_{21}s_{11} + A_{22}s_{21} & A_{21}s_{12} + A_{22}s_{22} \end{bmatrix}$$

$$= \text{tr}(A_{11}s_{11} + A_{12}s_{21}) + \text{tr}(A_{21}s_{12} + A_{22}s_{22})$$

$$= \text{tr}(A_{11}s_{11} + A_{12}s_{21} + s_{12}A_{21}) + \text{tr} A_{22}s_{22}$$

$$= \text{tr}(A_{11}s_{11} + A_{12}s_{21} + s_{12}A_{21} + A_{11}^{-1}A_{12}s_{22}A_{21})$$

$$+ \text{tr}(A_{22} - A_{21}A_{11}^{-1}A_{12})s_{22}.$$

Now, noting that $A_{22} - A_{21}A_{11}^{-1}A_{12} = \Gamma_{22}^{-1}$, substituting the variables $W$, $L$, and $V$ into the density of $S$, and multiplying by the Jacobian of the transformation, we have

$$p(W, L, V) = c |V|^{\frac{n-(p-q)-1}{2}} |\Gamma_{11}^{-1} \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}|^{\frac{n-(p-q)}{2}} |\Gamma_{22}|^{-n/2}$$

$$|\Gamma_{11}^{-1} \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}|^{\frac{p-q}{2}} |W|^{\frac{n-(p-q)-q-1}{2}}$$

$$\exp[-\frac{1}{2} \text{tr} \Gamma_{22}^{-1} V] \exp[-\frac{1}{2} \text{tr} A_{11} W]$$

$$\exp[-\frac{1}{2} \text{tr}(A_{11}LL' + A_{12}V_{L}^{\frac{1}{2}}L' + LV_{L}^{\frac{1}{2}}A_{21} + A_{12}VA_{21}A_{11}^{-1})].$$

30
However,

\[
\text{tr}(\Lambda_{11} LL' + \Lambda_{12} V^{\frac{1}{2}} L' + LV^{\frac{1}{2}} \Lambda_{21} + \Lambda_{12} V \Lambda_{21} \Lambda_{11}^{-1})
\]

\[= \text{tr} \Lambda_{11} [L + \Lambda_{11}^{-1} \Lambda_{12} V^{\frac{1}{2}}][L + \Lambda_{11}^{-1} \Lambda_{12} V^{\frac{1}{2}}]' \cdot \]

Noting that \( \Lambda_{11} = (\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1} \) and \( \Lambda_{11}^{-1} \Lambda_{12} = - \Gamma_{12} \Gamma_{22}^{-1} \), we have

\[ p(W,L,V) = p_1(W) p_2(L,V) p_3(V) \]

where

\[
p_1(W) = c_1 |W|^{\frac{n-(p-q)-q-1}{2}} |\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}|^{\frac{n-(p-q)}{2}} \exp[-\frac{1}{2}\text{tr}(\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})W] \]

\[
p_2(L,V) = c_2 |(\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1} \otimes \Gamma_{p-q}|^{\frac{1}{2}} \exp[-\frac{1}{2}\text{tr}(\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1}(L - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1} \Gamma_{p-q}^{-1} \Gamma_{22}^{-1} \Gamma_{21})' \}
\]

and

\[
p_3(V) = c_3 |V|^{\frac{n-(p-q)-l}{2}} |\Gamma_{22}|^{\frac{n}{2}} \exp(-\frac{1}{2}\text{tr} \Gamma_{22}^{-1} V) \cdot \]

Thus, \( W \) is \( W(n-(p-q), \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}) \), \( L/V \) is \( N(\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}^{-1}, (\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1} \otimes \Gamma_{p-q}) \) and \( V \) is \( W(n, \Gamma_{22}) \). Also, \( W \) is independent of the pair \((L, V)\). Interpreting these results in terms of the original variables, we have

\[
S_{11} - S_{22} \sim W(n-(p-q), \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}) \]

31
and is independent of \( \left( S_{12}^{-\frac{1}{2}}, S_{22} \right) \) and

\[
S_{12}^{-\frac{1}{2}}S_{22} \sim N \left( \Gamma_{12}^{-1}S_{22}, \left( \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} \right) \otimes I_{p-q} \right)
\]

and

\[
S_{22} \sim W(n, \Gamma_{22}).
\]

Also, multiplying \( S_{12}^{-\frac{1}{2}} \) by \( S_{22}^{-\frac{1}{2}} \), we have

\[
S_{12}^{-1}S_{22} \sim N \left( \Gamma_{12}^{-1}, \left( \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} \right) \otimes S_{22}^{-1} \right).
\]

We now want to obtain some further distributional results for the Wishart distribution. Recall that if \( S \) is \( W(n, \Gamma) \) and \( A \) is a \( q \times p \) matrix \( (q \leq p) \), then \( ASA' \) is \( W(n, A\Gamma A') \), where \( A \) is assumed to be of full rank. Now, consider \( A \) \( q \times p \) of rank \( q \) \( (q \leq p) \) and suppose the distribution of \( (AS^{-1}A')^{-1} \) is desired. We can write \( A = B(I_q \circ C) \)

where \( B \) is \( q \times q \), \( C \) is \( p \times p \) and both \( B \) and \( C \) are non-singular.

Then \( (AS^{-1}A')^{-1} = \left( B(I_q \circ C)S^{-1}C', \begin{pmatrix} I_q \\ 0 \end{pmatrix} \right)^{-1} = B^{-1}\left( \begin{pmatrix} I_q \\ 0 \end{pmatrix} \right)^{-1}B^{-1} \)

where \( \tilde{S} \) is \( W(n, C^{-1}\Gamma C^{-1}) = W(n, \tilde{\Gamma}) \). However, \( \left( \begin{pmatrix} I_q \\ 0 \end{pmatrix} \right)^{-1} = \)

\[
\tilde{S}_{11}^{-1} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21}
\]

which is \( W(n-(p-q), \tilde{\Gamma}_{11}^{-1} - \tilde{\Gamma}_{12}\tilde{\Gamma}_{22}^{-1}\tilde{\Gamma}_{21}) \) by our previous result. Thus, \( B^{-1}\left( \begin{pmatrix} I_q \\ 0 \end{pmatrix} \right)^{-1}B^{-1} = W(n-(p-q), B^{-1}(\tilde{\Gamma}_{11}^{-1} - \tilde{\Gamma}_{12}\tilde{\Gamma}_{22}^{-1}\tilde{\Gamma}_{21})B^{-1}) \).

Elementary computations show that \( B^{-1}(\tilde{\Gamma}_{11}^{-1} - \tilde{\Gamma}_{12}\tilde{\Gamma}_{22}^{-1}\tilde{\Gamma}_{21})B^{-1} = (A\Gamma^{-1}A')^{-1} \).

Thus \( (AS^{-1}A')^{-1} \) is \( W(n-(p-q), (A\Gamma^{-1}A')^{-1}) \) where \( A \) is \( q \times p \) \( (q \leq p) \) and is of rank \( q \).
In the case of \( q = 1 \), if \( \mathbf{a} \) is a \( 1 \times p \) vector, then

\[
\frac{a\mathbf{S}\mathbf{a}'}{a\mathbf{T}\mathbf{a}'}
\]

is \( \chi^2_n \) and

\[
\frac{a\mathbf{U}^{-1}\mathbf{a}'}{a\mathbf{S}^{-1}\mathbf{a}'}
\]

is \( \chi^2_{n-p+1} \) where \( \mathbf{S} \) is \( W(n,\mathbf{T}) \).

Section 4.

In this section, various testing problems for the mean of a multivariate normal distribution are considered. Also, the classification problem is considered since this problem motivates a testing problem which is a variant of the mean problem. The section is concluded with a discussion of multiple and partial correlation since the methods of deriving the distributions of the multiple and partial correlation coefficients are similar to those used in the mean test problems.

A few words are in order concerning the approach taken here for testing problems. Rather than compute likelihood ratio tests, we will look for a group under which the problem is invariant and then base a test on the maximal invariant computed under the group. A justification for this approach is the following. If a testing problem is invariant under a group and \( t(X) \) is a maximal invariant, then under regularity conditions, the likelihood ratio test is a function of the maximal invariant \( t(X) \). Thus, if we find a uniformly most powerful invariant test using the above approach, the likelihood ratio test can be no better since it is an invariant test.
Comment: For conditions under which the likelihood ratio test is invariant when the problem is invariant see Lehmann, Testing Statistical Hypotheses, page 252, problem 17.

Consider a space $\mathcal{X}$ and a group $\mathcal{G}$ operating on the left of $\mathcal{X}$. Let $\mathcal{G}$ be a set of points.

Definition: A function $f: \mathcal{X} \to \mathcal{Y}$ is called invariant if for all $x \in \mathcal{X}$ and $g \in \mathcal{G}$, we have

$$f(x) = f(gx).$$

Definition: A function $f: \mathcal{X} \to \mathcal{Y}$ is called maximal invariant if $f$ is invariant and if $f(x_1) = f(x_2)$ then there exists $g \in \mathcal{G}$ such that $gx_1 = x_2$.

Comment: Let $\mathcal{W}$ be a point set and suppose $h$ is a 1-1 function on $\mathcal{Y}$ to $\mathcal{W}$. If $f$ is maximal invariant on $\mathcal{X}$ to $\mathcal{Y}$, then $f \cdot h$ is maximal invariant on $\mathcal{X}$ to $\mathcal{W}$. This comment is often used to write a maximal invariant in a convenient form.

Example: Let $\mathcal{X} = (\mathcal{X}_1 \times \mathcal{X}_2)$ where $\mathcal{X}_1$ is $p$-dimensional Euclidean space and $\mathcal{X}_2$ is the set of all $p \times p$ symmetric positive definite matrices. If $x \in \mathcal{X}_1$ and $S \in \mathcal{X}_2$, write $x = \begin{pmatrix} x_1^T \\ y \end{pmatrix}$,

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where $y$ is $q \times 1$ and $S_{11}$ is $q \times q$. Let $\mathcal{G}$ be the group of all non-singular $p \times p$ matrices $A$ of the form
\[
A = \begin{pmatrix}
  A_{11} & A_{12} \\
  0     & A_{22}
\end{pmatrix}
\]

where \( A_{11} \) is \( q \times q \). The group operation is defined by

\[
A: \quad x \rightarrow Ax = \begin{pmatrix}
  A_{11} & A_{12} \\
  0     & A_{22}
\end{pmatrix}
\begin{pmatrix}
  y \\
  z
\end{pmatrix} = \begin{pmatrix}
  A_{11}y + A_{12}z \\
  A_{22}z
\end{pmatrix}
\]

A: \( S \rightarrow AS' \).

We now want to show that a maximal invariant is \( f(x, S) = (x'S^{-1}x, z'S^{-1}z) \).

First note that \( f \) is invariant and \( x'S^{-1}x = (y-s_{12}^{-1}S^{-1}S_{22})'(s_{11}-s_{12}^{-1}S^{-1}S_{22})^{-1}
(y-s_{12}^{-1}S^{-1}S_{22}z) + z'S^{-1}z \).

Now, suppose that

\[
(x_1'S^{-1}x_1', z_1'S^{-1}z_1') = (x_2'T^{-1}x_2', z_2'T^{-1}z_2').
\]

We must show there exists an \( A \in \mathcal{L} \) so that

\[
x_1 = Ax_2
\]

and

\[
S = ATA'.
\]

First choose \( B = \begin{pmatrix}
  B_{11} & B_{12} \\
  0 & B_{22}
\end{pmatrix} \) as \( B_{22} = S_{22}^{-1} \), \( B_{11} = (S_{11}-S_{12}S_{22}^{-1}S_{21})^{-1} \),

\[
B_{12} = -B_{11}S_{12}^{-1}.
\]

Then \( BSB = I \). Also, choosing \( C = \begin{pmatrix}
  C_{11} & C_{12} \\
  0 & C_{22}
\end{pmatrix} \) as

35
$C_{22} = T_{22}^{-\frac{1}{2}}, \ C_{11} = (T_{11}^{-1}T_{12}^{-1}T_{22}^{-1})^{-\frac{1}{2}}, \ C_{12} = C_{11}T_{12}^{-1}T_{22}^{-1}$ we have $CTC' = I$.

Since $z_1S_{22}^{-1}z_1 = z_2T_{22}^{-1}z_2$, there exists an orthogonal matrix

$\psi_2((p-q) \times (p-q))$ so that $B_{22}z_1 = S_{22}^{-1}z_1 = \psi_2T_{22}^{-1}z_2 = \psi_2C_{22}z_2$. Since

$x_1S_{11}^{-1}x_1 = x_2T_{12}^{-1}x_2$ and $S_{11}^{-1} = B'B, \ T_{12}^{-1} = C'C, \ z_1B'_{12}B_{22}z_1 = z_2C_{22}z_2$, we have

$$\|B_{11}y_1 + B_{12}z_1\|^2 = \|C_{11}y_2 + C_{12}z_2\|^2$$

so that there exists an orthogonal matrix $\psi_1(p \times p)$ such that

$$B_{11}y_1 + B_{12}z_1 = \psi_1(C_{11}y_2 + C_{12}z_2).$$

Now, let $\psi = \begin{pmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{pmatrix}$. Then $BSB' = I = \psi CTC' \psi'$ so that $S = B^{-1}CTC' \psi'B^{-1}$. Setting $A = B^{-1}C$, we then have $S = ATA'$.

However, direct computation shows that

$$x_1 = Ax_2,$$

so that a maximal invariant is $(xS_{12}^{-1}x', x'S_{22}^{-1}z)$.

Very often, a more convenient way to write the maximal invariant is

$$\begin{pmatrix} (y-S_{12}^{-1}S_{22}^{-1}y) \cdot (S_{11}^{-1}S_{12}^{-1}S_{22}^{-1}y)^{-1} \\
1 + z'S_{22}^{-1}z \end{pmatrix} = \begin{pmatrix} x'S_{12}^{-1}x - z'S_{22}^{-1}z \\
1 + z'S_{22}^{-1}z \end{pmatrix}.$$
Hotelling's $T^2$

Consider $X_1, \ldots, X_{n+1}$ independent $N(x^0, \Gamma)$. Reducing to the sufficient statistics, we have $X = \sum_1^{n+1} X_i$ and $S = \sum_1^{n+1} X_i X'_i - \frac{1}{(n+1)} XX'$

where $X$ is $N((n+1)x^0, \Gamma)$ and is independent of $S$ which is $W(\Gamma, n)$. Let $x^0 = (n+1)x^0$ so that we have $X \sim N(x^0, \Gamma)$, $S \sim W(\Gamma, n)$ and $S$ and $X$ are independent. Let $\mathcal{Z}$ be the linear space in which $X$ takes values and consider two subspaces of $\mathcal{Z}$, say $\mathcal{U}$ and $\mathcal{V}$, such that $\mathcal{U} \neq \mathcal{V} \subseteq \mathcal{Z}$. We are interested in the hypothesis

$$H_0: x^0 \in \mathcal{U}$$

versus

$$H_1: x^0 \in \mathcal{V}.$$ 

Note that this hypothesis testing problem remains invariant under the group of non-singular linear transformations which leave both $\mathcal{U}$ and $\mathcal{V}$ invariant. Also, the problem remains invariant under translations by any element of $\mathcal{U}$.

**Case I: $\mathcal{U} = 0$, $\mathcal{V} = \mathcal{Z}$**

For this case we want to test $H_0: x^0 = 0$ against unrestricted alternatives. The problem is invariant under the group of all non-singular linear transformations. The group operation is defined by

$$X \rightarrow AX$$

$$S \rightarrow ASA'$$
and it is elementary to show that \( T^2 = X'S^{-1}X \) is a maximal invariant (the argument is similar to that in the example above). The statistic \( T^2 \) is commonly called Hotelling's \( T^2 \)-statistic, although our definition of \( T^2 \) is off by a constant term from that originally given by Hotelling. We now want to find the distribution of \( T^2 \).

**Proposition:** If \( X \sim N(x^0, \Gamma) \) and is independent of \( S \sim W(\Gamma, n) \) where \( X \) is of dimension \( p \) \((n \geq p)\), then

\[
T^2 = X'S^{-1}X \sim \frac{x^2_p(x^0', \Gamma^{-1}x^0)}{\chi^2_{n-p+1}}.
\]

**Proof:**

\[
X'S^{-1}X = \frac{X^tS^{-1}X}{X^t\Gamma^{-1}X} \cdot X^t\Gamma^{-1}X.
\]

However, \( \frac{X^tS^{-1}X}{X^t\Gamma^{-1}X} \) has the same distribution as \( \frac{1}{\chi^2_{n-p+1}} \) independent of \( X \) (see Section 3 of these notes). Also, \( X^t\Gamma^{-1}X \) has a non-central \( x^2 \) distribution with \( p \) degrees of freedom and non-centrality parameter \( x^0'\Gamma^{-1}x^0 \).

**Comment:** Recall that \( x^2_p(x^0', \Gamma^{-1}x^0) \) has the same distribution as \( x^2_{p+2K} \) where \( K \) is a Poisson random variable with parameter

\[
\lambda = \frac{x^0'\Gamma^{-1}x^0}{2}.
\]

Now it is well known that \( \frac{x^2_p(x^0', \Gamma^{-1}x^0)}{\chi^2_{n-p+1}} \) has a monotone likelihood.
ratio in the non-centrality parameter $x^0 \Gamma^{-1} x^0$. Since our hypothesis is $x^0 \Gamma^{-1} x^0 = 0$ against the alternative $x^0 \Gamma^{-1} x^0 > 0$, we conclude that the test which rejects for large values of $T^2$ is the uniformly most powerful invariant test under the full linear group. Note that this test is equivalent to the likelihood ratio test.

Case II: $0 \notin \mathcal{U} \neq \emptyset, \mathcal{V} = \emptyset$

Let $\mathcal{U}$ be of dimension $q$ and let $\mathcal{Z}$ be the quotient space $\mathbb{R} / \mathcal{U}$, of dimension $p - q$. Consider the diagram

$$
\begin{array}{ccc}
\mathcal{U} & \xrightarrow{\iota} & \mathbb{R} \\
\downarrow & & \downarrow \pi \\
\mathcal{Z} & & \\
\end{array}
$$

where $\iota$ is the inclusion mapping and $\pi$ is the natural homomorphism of $\mathbb{R}$ onto $\mathcal{Z}$ such that $\text{Im}(\iota) = \text{Ker}(\pi)$. For computational convenience, we pick a coordinate system so that the matrix of $\iota$ is $\begin{pmatrix} I_q \\ 0 \end{pmatrix}$ where $I_q$ is the $q \times q$ identity and the matrix of $\pi$ is $\begin{pmatrix} 0, I_{p-q} \end{pmatrix}$.

Now, write $X = \begin{pmatrix} Y \\ z \end{pmatrix}$ and $x^0 = \begin{pmatrix} y^0 \\ z^0 \end{pmatrix}$, $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ where $Y \in \mathcal{U}$, $y^0 \in \mathcal{U}$ and $S_{11}$ is $q \times q$. Then our hypothesis is $H_0: z^0 = 0$ against an unrestricted alternative. This testing problem is invariant under the following group of transformations:

$$(A, b)X = AX + b$$

$$(A, b)S = ASA'$$
where \( A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \), \( b = \begin{pmatrix} b_1 \\ 0 \end{pmatrix} \) is in \( \text{Im}(\iota) \) and the group operation is defined by \( (A_2, b_2)(A_1, b_1) = (A_2 A_1, A_2 b_1 + b_2) \).

It is not hard to show that the maximal invariant under this group of transformations is \( Z'S_{22}^{-1}Z \). Note that this is just Hotelling's \( T^2 \) based on \( \pi X \) and \( \pi S \pi' \) so that we have \( Z'S_{22}^{-1}Z = (\pi X)'(\pi S \pi')^{-1}(\pi X) \).

From the treatment of Case I, it is clear that the distribution of \( Z'S_{22}^{-1}Z \) is that of

\[
\chi^2_{p-q} \frac{Z'S_{22}^{-1}Z}{Z'S_{22}^{-1}22}
\]

As in Case I, we conclude that the test which rejects for large values of \( Z'S_{22}^{-1}Z \) is the uniformly most powerful invariant test under the above group. Again, this test is the same as the likelihood ratio test.

**Case III:** \( \mathcal{U} = 0, \ 0 \neq \mathcal{V} \subsetneq \mathcal{Z} \):

Let \( \mathcal{V} \) be of dimension \( q \) and \( \mathcal{Z} \) of dimension \( p \) and consider the diagram:

\[
\mathcal{V} \xrightarrow{\iota} \mathcal{X} \xrightarrow{\pi} \mathcal{Y}
\]

where \( \iota \) is the inclusion mapping and \( \pi \) is the natural homomorphism of \( \mathcal{X} \) onto \( \mathcal{Z} \). Again, choose a coordinate system so that the matrix of \( \iota \) is \( \begin{pmatrix} I_q \\ 0 \end{pmatrix} \) and the matrix of \( \pi \) is \( \begin{pmatrix} 0 \\ I_{p-q} \end{pmatrix} \).

40
Write \( X = (Y, Z) \), \( x^0 = (y^0, z^0) \), and \( S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \) where

\( Y \in \mathcal{V} \), \( y^0 \) is in \( \mathcal{V} \), and \( S_{11} \) is \( q \times q \) so that the hypothesis is \( H_0: y^0 = 0, z^0 = 0 \) and the alternative is \( H_1: z^0 = 0 \).

This testing problem is invariant under the following transformations:

\[ X \rightarrow AX \]
\[ S \rightarrow ASA' \]

where \( A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \) and \( A_{11} \) is \( q \times q \), \( A \) is non-singular.

In a previous example, the maximal invariant was shown to be

\[
\left( \frac{X'S^{-1}X - Z'S^{-1}Z}{1 + Z'S^{-1}Z}, Z'S^{-1}Z \right).
\]

We now want to find the joint distribution of the above vector. For simplicity, let

\[
W_1 = \frac{X'S^{-1}X - Z'S^{-1}Z}{1 + Z'S^{-1}Z}
\]

and

\[ W_2 = Z'S^{-1}Z. \]

We can write

\[
W_1 = \frac{(Y - S_{12}S^{-1}Z)(S_{11} - S_{12}S^{-1}S_{21})^{-1}(Y - S_{12}S^{-1}Z)}{1 + Z'S^{-1}Z}.
\]
It has been shown that

\[ S_{11} - S_{12} S_{22}^{-1} S_{22} \sim W(n-(p-q), \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}) \]

and is independent of \((S_{12}, S_{22}). Also Y|S_{22}, Z \sim N(y^0 + \Gamma_{12} \Gamma_{22}^{-1} (z-z^0), \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})\) and is independent of

\[ S_{12} S_{22}^{-1} Z|S_{22}, Z \sim N(\Gamma_{12} \Gamma_{22}^{-1} Z, Z S_{22}^{-1} Z \cdot (\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})) \]

so that

\[
\frac{Y - S_{12} S_{22}^{-1} Z}{\sqrt{1 + Z S_{22}^{-1} Z}} \mid S_{22}, Z \sim N\left(\frac{y^0 - \Gamma_{12} \Gamma_{22}^{-1} Z^0}{\sqrt{1 + Z S_{22}^{-1} Z}}, \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}\right).
\]

From a previous proposition, it then follows directly that

\[
\chi^2_{q} \left(\frac{(y^0 - \Gamma_{12} \Gamma_{22}^{-1} Z^0)(\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1}(y^0 - \Gamma_{12} \Gamma_{22}^{-1} Z^0)}{1 + Z S_{22}^{-1} Z}\right) \frac{\chi^2_{n-p+1}}{n-p+1}
\]

Since the conditional distribution of \(W_1|S_{22}, Z\) depends on \((S_{22}, Z)\) only through \(Z S_{22}^{-1} = W_2\), we have that

\[
\chi^2_{q} \left[\frac{(y^0 - \Gamma_{12} \Gamma_{22}^{-1} Z^0)(\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1}(y^0 - \Gamma_{12} \Gamma_{22}^{-1} Z^0)}{1 + W_2}\right] \frac{\chi^2_{n-p+1}}{n-p+1}
\]

Further, it is clear that
\[ W_2 \sim \frac{\chi^2_{p-q}(z^0, \Gamma^{-1}z^0)}{\chi^2_{n-(p-q)+1}}. \]

Returning to the testing problem, note that \( z^0 = 0 \) under both \( H_0 \) and \( H_1 \) so that

\[
W_1 | W_2 \sim \frac{\chi^2_{q} \left( \frac{y^0 (\Gamma_{11} - \Gamma_{12} \Gamma^{-1}_{22} \Gamma_{21}) y^0}{1 + W_2} \right)}{\chi^2_{n-p+1}}
\]

and

\[ W_2 \sim \frac{\chi^2_{p-q}}{\chi^2_{n-(p-q)+1}}. \]

Before proposing a possible test statistic, we digress for a moment to discuss the principle of conditionality.

The Principle of Conditionality

Let \( E_a, a \in \mathfrak{A} \) be a set of experiments and let \( Q \) be a probability measure on \( \mathfrak{A} \). Suppose for \( a \in \mathfrak{A} \), we observe \( X_a \) with a distribution \( P_\theta \) such that when \( \theta = \theta_0 \), \( P_{\theta_0} \) does not depend on \( a \in \mathfrak{A} \). Suppose also that an appropriate level \( \alpha \) test for \( H_0: \theta = \theta_0 \) based on \( E_a \) is to reject if \( X_a \in S_a \).

Let \( A \) be a random variable distributed according to \( Q \) in \( \mathfrak{A} \). Suppose that we choose \( A \) according to \( Q \) and then perform \( E_A \) to obtain \( X_A \). The principle of conditionality asserts that an appropriate level \( \alpha \) test is to reject \( H_0 \) if \( X_A \in S_A \).
For our problem, we let \( W_2 \) correspond to \( A \) and \( W_1 \mid W_2 \) correspond to \( X_a \). Under \( H_0 \), \( W_1 \) and \( W_2 \) are independent and hence the appropriate test is to reject for large values of \( W_1 \). Thus, using the principle of conditionality, we reject for large values of

\[
\frac{(Y - S_{12}S_{22}^{-1}Z)'(S_{11} - S_{12}S_{22}^{-1}S_{21})^{-1}(Y - S_{12}S_{22}^{-1}Z)}{1 + Z'S_{22}^{-1}Z}.
\]

We also note that the likelihood ratio test is equivalent to the above test.

Now, the question is: Is the above test the only possibility? Consider \( H_0^1: y^0 = 0, z^0 = 0 \) against the alternative

\( H_1^1: y^0'(\tau_{11} - \tau_{12}T_{22}T_{21})^{-1}y^0 = c_0, z^0 = 0 \) where \( c_0 \) is specified, \( c_0 > 0 \). Since this problem is invariant under the group considered above, we can use the Neyman-Pearson Lemma to compute the uniformly most powerful invariant of \( H_0 \) versus \( H_1^1 \) since we have a simple alternative for the invariant problem. Let \( p_\lambda(W_1 \mid W_2) \) be the conditional density of \( W_1 \mid W_2 \) when \( y^0'(\tau_{11} - \tau_{12}T_{22}T_{21})^{-1}y^0 = \lambda \) and let \( q(W_2) \) be the density of \( W_2 \). Then the joint density of \( (W_1,W_2) \) is

\[
f_\lambda(W_1,W_2) = p_\lambda(W_1 \mid W_2)q(W_2).
\]

Now, the uniformly most powerful invariant test is to reject if

\[
\frac{f_{c_0}(W_1,W_2)}{f_0(W_1,W_2)} \geq b_\alpha
\]

where \( b_\alpha \) is computed to make the test level \( \alpha \). Thus, we want to reject if
\[ \frac{p_c(\mathbf{w}_1 | \mathbf{w}_2)}{p_0(\mathbf{w}_1 | \mathbf{w}_2)} \geq b^* \]

However, this test is different than the test which rejects for large values of \( \mathbf{w}_1 \).

Case IV: \( 0 \notin \mathcal{U} \subsetneq \mathcal{V} \subsetneq \mathcal{X} \):

Let \( \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{pmatrix} \), \( \mathbf{x}^0 = \begin{pmatrix} \mathbf{x}_1^0 \\ \mathbf{x}_2^0 \\ \mathbf{x}_3^0 \end{pmatrix} \) and \( \mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \)

where \( \mathbf{X} \), \( \mathbf{x}^0 \) and \( \mathbf{S} \) are partitioned according to the dimension of \( \mathcal{U} \) and \( \mathcal{V} \). Of course, we assume that a convenient coordinate system has been chosen. Our problem is to test \( H_0: \mathbf{x}_2^0 = 0 \), \( \mathbf{x}_3^0 = 0 \) against \( H_1: \mathbf{x}_3^0 = 0 \). The problem is invariant under the transformations

\[
(A, b) \mathbf{X} = \mathbf{A} \mathbf{X} + b
\]

\[
(A, b) \mathbf{S} = \mathbf{A} \mathbf{S} \mathbf{A}^T
\]

where \( \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{pmatrix} \) and \( b = \begin{pmatrix} b_1 \\ 0 \\ 0 \end{pmatrix} \). A maximal invariant for this problem is the pair \( (\mathbf{w}_1, \mathbf{w}_2) \) where

\[
\mathbf{w}_1 = \frac{(\mathbf{x}_2^0 - s_{22}^{-1} s_{23}^{-1} s_{33}^{-1} \mathbf{x}_3^0) \cdot (s_{22}^{-1} s_{23}^{-1} s_{33}^{-1})^{-1} (\mathbf{x}_2^0 - s_{22}^{-1} s_{23}^{-1} s_{33}^{-1} \mathbf{x}_3^0)}{1 + \mathbf{x}_3^0 s_{22}^{-1} s_{23}^{-1} s_{33}^{-1} \mathbf{x}_3^0}
\]

and

\[
\mathbf{w}_2 = \mathbf{x}_3^0 s_{33}^{-1} \mathbf{x}_3^0.
\]

45
It is now clear that the distribution theory given in Case III applies directly to the present case. Also, the remarks given in Case III carry over without change to the present case.

The Classification Problem:

Consider two completely specified densities \( p_1 \) and \( p_2 \) and let \( Z \) be an observations which we know comes from either \( p_1 \) or \( p_2 \). The problem is to decide from which population \( Z \) comes.

Let action \( a_1 \) correspond to deciding that \( Z \) comes from \( p_1 \) and let \( a_2 \) correspond to deciding that \( Z \) comes from \( p_2 \). The states of nature are \( p_1 \) and \( p_2 \) which we denote by 1 and 2. Consider the loss table

<table>
<thead>
<tr>
<th></th>
<th>( a_1 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( c_2 )</td>
</tr>
<tr>
<td>2</td>
<td>( c_1 )</td>
<td>0</td>
</tr>
</tbody>
</table>

where \( c_1 \) and \( c_2 \) are both positive. Let \(( \pi_1, 1-\pi_1)\) be the prior probability distribution on the states of nature. Then the posterior probability \(( \xi_1, 1-\xi_1)\) given \( Z \) on the states of nature is given by

\[
\xi_1 = \frac{\pi_1 p_1(Z)}{\pi_1 p_1(Z) + (1-\pi_1)p_2(Z)}.
\]

Thus, the posterior expected loss for taking action 1 is
\[
\frac{(1-\pi_1)p_2(z)c_1}{\pi_1 p_1(z) + (1-\pi_1)p_2(z)}
\]

and for taking action 2 is

\[
\frac{\pi_1 p_1(z)c_2}{\pi_1 p_1(z) + (1-\pi_1)p_2(z)}.
\]

We then take action 2 if

\[
\frac{\pi_1 p_1 c_2}{\pi_1 p_1 + (1-\pi_1)p_2} < \frac{(1-\pi_1)p_2 c_1}{\pi_1 p_1 + (1-\pi_1)p_2}
\]

or take action 2 if

\[
\frac{p_2(z)}{p_1(z)} > \frac{\pi_1 c_2}{(1-\pi_1)c_1} \equiv k
\]

where we must randomize if there is equality.

Now, suppose \( p_1 \sim N(x^0, \Gamma) \) and \( p_2 \sim N(y^0, \Gamma) \) where \( x^0, y^0 \), and \( \Gamma \) are known. Thus, we take action 2 if

\[
Z'(\Gamma^{-1}(y^0 - x^0) \geq \frac{1}{2} \text{tr}(y^0 y^0' - x^0 x^0') \Gamma^{-1} + \log k.
\]

The linear functional \( \Gamma^{-1}(y^0 - x^0) \) is called Fisher's Discriminant Function. In the case where \( \Gamma^{-1}, x^0 \) and \( y^0 \) are unknown, we can consider estimation and testing problems for \( \Gamma^{-1}(y^0 - x^0) \).

Consider \( X_1, \ldots, X_m \) independent \( N(x^0, \Gamma) \) and \( Y_1, \ldots, Y_n \) independent \( N(y^0, \Gamma) \). Invariance and sufficiency considerations lead us to
consider

\[ z = \frac{1}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \frac{\bar{y} - \bar{x}}{n} \]

and

\[ s = \sum_{1}^{m} (x_i - \bar{x})(x_i - \bar{x})' + \sum_{1}^{n} (y_i - \bar{y})(y_i - \bar{y})' \]

for inferences about \( \Gamma^{-1}(y_0 - x) \). Thus, in canonical form, we have (relabeling variables) \( X \sim N(x^0, \Gamma) \) and \( S \sim W(\Gamma, n) \) and want to consider problems concerning \( \Gamma^{-1}x^0 \).

Tests of Hypotheses for \( \Gamma^{-1}x^0 \)

Let \( X \sim N(x^0, \Gamma) \) and \( S \sim W(\Gamma, n) \) and consider the linear space \( \mathcal{Z} \) in which \( X \) takes values. Let \( \mathcal{Z}' \) be the dual space of \( \mathcal{Z} \) and consider two subspaces \( \mathcal{U} \subset \mathcal{V} \subset \mathcal{Z}' \). We are interested in the hypothesis \( H_0: \Gamma^{-1}x^0 \in \mathcal{U} \) versus the alternative \( H_1: \Gamma^{-1}x^0 \in \mathcal{V} \). As in the mean testing problems we again consider cases. Note that the case of \( \mathcal{U} = 0 \) does not seem to be of much interest in the context of linear discriminant functions but is included for completeness.

Case I: \( \mathcal{U} = 0, \mathcal{V} = \mathcal{Z}' \)

In this case, \( H_0: \Gamma^{-1}x^0 = 0 \) versus \( H_1: \Gamma^{-1}x \in \mathcal{Z}' \) is equivalent to \( H_0: x^0 = 0 \) versus \( H_1: x^0 \in \mathcal{Z} \) which has already been discussed.
Case II: $0 \neq \mathcal{U}, \mathcal{V} = \mathbb{R}$

This case will follow from the results presented for Case IV.

Case III: $\mathcal{U} = 0, \mathcal{V} \neq \mathcal{U} = \mathbb{R}$

We have $X \sim N(x^0, \Gamma)$, $S \sim W(\Gamma, n)$ and $H_0: \Gamma^{-1}x^0 = 0$ versus $H_1: \Gamma^{-1}x^0 \in \mathcal{V}$. Let $\xi^0 = \Gamma^{-1}x^0$, $q = \dim \mathcal{V}$,

$$\xi^0 = \begin{pmatrix} \xi^0_1 \\ 0 \\ \xi^0_2 \end{pmatrix}$$

and consider the following diagram:

$$\begin{array}{c}
\mathcal{V} \\
\downarrow \iota \\
\mathbb{R}\
\end{array}$$

where $\iota$ is the inclusion mapping. Choose a coordinate system so that the matrix of $\iota$ is $\begin{pmatrix} I_q \\ 0 \end{pmatrix}$. In terms of this coordinate system, we write

$$X = \begin{pmatrix} Y \\ Z \end{pmatrix}, S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, x^0 = \begin{pmatrix} y^0 \\ z^0 \end{pmatrix}, H_0: \xi^0_1 = 0, H_1: \xi^0_2 = 0.$$

Consider the group of transformations

$$X \rightarrow AX$$
$$S \rightarrow ASA'$$

where $A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$. Under this group, we have

$$x^0 \rightarrow Ax^0$$
$$\Gamma \rightarrow A\Gamma A'$$

so that
\[ \xi^0 \rightarrow (A')^{-1}\xi^0. \]

Noting that \((A')^{-1} = B\) has the form
\[
B = \begin{pmatrix}
B_{11} & B_{12} \\
0 & B_{22}
\end{pmatrix}
\]
we see that the above hypothesis testing problem remains invariant under this group of transformations.

Comment: Note that \(A\) is the transpose of the type of transformation leaving the corresponding mean problem invariant. The reason for this is that \(x^0 \in \mathcal{X}\) while \(\xi^0 \in \mathcal{X}'.\)

A maximal invariant is easily found to be the pair \((W_1, W_2)\) where
\[
W_1 = Y'S_{11}^{-1}Y
\]
and
\[
W_2 = \frac{(Z - S_21S_{11}^{-1}Y)'(S_{22} - S_21S_{11}^{-1}S_{12})^{-1}(Z - S_21S_{11}^{-1}Y)}{1 + Y'S_{11}^{-1}Y}.
\]

It is clear that (see Case III of the mean testing problem)
\[
W_1 \sim \frac{x^2(y^0, R_{11}^{-1}y^0)}{\chi^2_{n-q+p}}
\]
and
\[
W_2 | W_1 \sim \frac{x^2}{\chi^2_{n-p+1}} \left[ \frac{(z^0 - R_{21}^{-1}r_{11}^{-1}y^0)'(r_{22} - R_{21}r_{12}^{-1}r_{11}^{-1})^{-1}(z^0 - R_{21}r_{12}^{-1}r_{11}^{-1}y^0)}{1 + W_1} \right].
\]

Noting that \(R_{11}y^0 = x^0\), we have under \(H_0\) and \(H_1\) that \(y^0 = R_{11}^0_1\) and \(r_{21}^0_1 = x^0\). Thus, under both \(H_0\) and \(H_1\),
\[ W_1 \sim \frac{\chi^2(g_1, T, \tau_{11}, \tau_{11}^{0})}{\chi^2(n-q+1)} \]

\[ W_2 \sim \frac{\chi^2(p-q)}{\chi^2(n-p+1)} \]

and \( W_1 \) is independent of \( W_2 \). Since \( W_1 \) and \( W_2 \) are independent under both \( H_0 \) and \( H_1 \) and the distribution of \( W_2 \) does not depend on any parameters, we conclude that the uniformly most powerful invariant test is the test which rejects for large values of \( W_1 \). Of course, we have used the fact that \( W_1 \) has a monotone likelihood ratio in \( \tau_{11}^{0}, T, \tau_{11} \).

Comment: The statistic \( W_1 \) can also be written as \( (t'X)'(t'St)^{-1}(t'X) \) which is Hotelling's \( T^2 \) for \( (t'X) \) and \( t'St \).

Another approach to this problem is the following: Let \( X = (Y, Z) \), \( \xi^0 = T^{-1}X^0 \), \( \Lambda = T^{-1} \) and write the density of \( X \) and \( S \) as

\[ p(x, s) = c |s|^{-\frac{n-p-1}{2}} |\Lambda|^{-\frac{n+1}{2}} e^{-\frac{1}{2} tr \Lambda(s+xx') + \xi^0_x s - \frac{1}{2} \xi^0, \Lambda^{-1} \xi^0} \]

Writing \( \xi^0 = \begin{pmatrix} \xi^0_0 \\ \xi^0_1 \\ \xi^0_2 \end{pmatrix} \) note that under both \( H_0 \) and \( H_1 \), \( \xi^0_2 = 0 \) so that the pair \( ((s+xx'), y) \) is sufficient for \( (\Lambda, \xi^0) \) where \( x = (Y, Z) \).

Setting \( T = S + XX' \), consider the group of transformations

\[ T \rightarrow ATA' \]

\[ Y \rightarrow A_1 Y \]

51
\[ A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} \] and

\[ A \rightarrow A^{-1}AA^{-1} \]
\[ \xi \rightarrow A^{-1}\xi A. \]

As before, the problem is invariant under this group of transformations.

However, the maximal invariant is now

\[ W_1^* = Y'T_{11}^{-1}Y \]

where

\[ T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}. \]

Noting that \( T_{11} = S_{11} + YY' \), we have

\[ W_1^* = Y'(S_{11} + YY')^{-1}Y. \]

However,

\[ (S_{11} + YY')^{-1} = S_{11}^{-1}(I + YY'S_{11}^{-1})^{-1} = S_{11}^{-1}\left[I - \frac{YY'S_{11}^{-1}}{I + Y'S_{11}^{-1}Y}\right]. \]

Thus,

\[ W_1^* = \frac{Y'S_{11}^{-1}Y}{1 + Y'S_{11}^{-1}Y}. \]
which is a 1-1 function of

$$W_1 = Y^T S_{11}^{-1} Y.$$ 

Thus, we take $W_1$ to be a maximal invariant. Now,

$$W_1 \sim \frac{x^2(\xi_0^T, \Gamma_{11}, \xi_0)}{x^2(n-q+1)}$$

so that a uniformly most powerful invariant test of $H_0$ versus $H_1$ is to reject for large values of $W_1$.

Case IV: $0 \subset \mathcal{U} \subset \mathcal{V} \subset \mathcal{Z}'$

Again we have $X \sim N(x^0, \Gamma)$ and $S \sim W(\Gamma, n)$ with $X$ taking values in $\mathcal{Z}$, and we want to test $H_0: \Gamma^{-1}x^0 \in \mathcal{U}$ versus $H_1: \Gamma^{-1}x^0 \in \mathcal{V}$.

Consider the diagram

$$\mathcal{U} \xrightarrow{\iota_1} \mathcal{V} \xrightarrow{\iota_2} \mathcal{Z}'$$

where $\iota_1$ is the inclusion mapping of $\mathcal{U}$ into $\mathcal{V}$ and $\iota_2$ is the inclusion mapping of $\mathcal{V}$ into $\mathcal{Z}'$. Let $\mathcal{U}$ be of dimension $q$, $\mathcal{V}$ be of dimension $p$, and $\mathcal{Z}'$ be of dimension $r$, where $0 < q < p < r$. We choose a coordinate system so that the matrix of $\iota_1$ is $\begin{pmatrix} I_q \\ 0 \end{pmatrix}$ ($p \times q$)

and the matrix of $\iota_2$ is $\begin{pmatrix} I_p \\ 0 \end{pmatrix}$ ($r \times p$).

Let $T = S + XX'$ and write $\Lambda = \Gamma^{-1}$, $\xi_0 = \Lambda x^0$. 

53
\[ X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad T = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix}, \quad \xi^0 = \begin{pmatrix} \xi_1^0 \\ \xi_2^0 \\ \xi_3^0 \end{pmatrix} \]

so that the null hypothesis is \( H_0: \xi_2^0 = 0, \xi_3^0 = 0 \) and the alternative is \( H_1: \xi_3^0 = 0 \). Now, write the joint density of \( X \) and \( S \) as

\[ p(x,s) = c|s|^{n-2} |\Lambda|^{n+1} e^{-\frac{1}{2} \text{tr} \Lambda (s + xx')} + \xi_0^0 x - \frac{1}{2} \xi_0^0 A^{-1} \xi_0. \]

Then the pair \( (T, \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}) \) are sufficient for \( (\Lambda, \xi^0) \) since \( \xi_3^0 = 0 \) under both \( H_0 \) and \( H_1 \).

Consider the group of transformations

\[
\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \rightarrow \begin{pmatrix} A_{11} X_1 \\ A_{21} X_1 + A_{22} X_2 \end{pmatrix}, \quad T \rightarrow \Lambda T A'\Lambda^{-1}
\]

where

\[ A = \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \]

and

\[ A \rightarrow A^{-1} \Lambda A^{-1}, \quad \xi^0 \rightarrow A^{-1} \xi^0. \]

The problem remains invariant under this group. A maximal invariant is easily shown to be
\[(x_1^1 T_{11}^{-1} x_1, (x_2 - T_{21} T_{11}^{-1} x_1)' (T_{22} - T_{21} T_{11}^{-1} T_{12})^{-1} (x_2 - T_{21} T_{11}^{-1} x_1))\]

or equivalently, a maximal invariant is
\[
\left(\begin{array}{c}
  x_1 \\
  x_2
\end{array}\right) \quad \left(\begin{array}{cc}
  T_{11}^{-1} T_{12} & -1 \\
  T_{21} T_{22} & 0
\end{array}\right) \quad \left(\begin{array}{c}
  x_1 \\
  x_2
\end{array}\right)
\]

Noting that \( T = S + XX^t \), we have that a maximal invariant is (See Case III above)

\[
\left(\begin{array}{c}
  x_1 \\
  x_2
\end{array}\right) \quad \left(\begin{array}{cc}
  S_{11} & S_{12} \\
  S_{21} S_{22} & 0
\end{array}\right) \quad \left(\begin{array}{c}
  x_1 \\
  x_2
\end{array}\right)
\]

so that a maximal invariant is

\[
\left[\begin{array}{c}
  x_1^s T_{11}^{-1} x_1 \\
  1 + x_1^s T_{11}^{-1} x_1
\end{array}\right] \quad \left(\begin{array}{c}
  (x_2 - S_{21} S_{11}^{-1} x_1)' (S_{22} - S_{21} S_{11}^{-1} S_{12})^{-1} (x_2 - S_{21} S_{11}^{-1} x_1) \\
  1 + x_1^s T_{11}^{-1} x_1
\end{array}\right)
\]

Comment: In coordinate-free notation, let \( T^2 = (t'_2 x)' (t'_2 S_{22}^{-1} t'_2)^{-1} (t'_2 x) \) and \( T^2 = (t'_1 x)' (t'_1 S_{11}^{-1} t'_1)^{-1} (t'_1 x) \). Then we can write the above maximal invariant as

\[
\left(\begin{array}{c}
  T^2 \\
  T^2 - T^2
\end{array}\right) \quad \left(\begin{array}{c}
  T^2 \\
  1 + T^2
\end{array}\right)
\]

55
Now, let \( W_1 = T_1^2 \) and \( W_2 = (T_2^2 - T_1^2)/(1 + T_1^2) \). The distribution theory for \( (W_1, W_2) \) is similar to that presented in Case III for the mean problem and we easily obtain

\[
W_1 \sim \frac{\chi^2(n, \mathbf{x}_1 \mathbf{x}_1^T)}{\chi^2(n-q+1)}
\]

and

\[
W_2 | W_1 \sim \frac{\chi^2(p-q, \mathbf{x}_2^T \mathbf{r}_2^T \mathbf{x}_2)}{\chi^2(n-p+1)}
\]

Noting that \( x_1^0 = \Gamma_{11} \xi_1^0 + \Gamma_{12} \xi_2^0 \) and \( x_2^0 = \Gamma_{21} \xi_1^0 + \Gamma_{22} \xi_2^0 \), we have

\[
W_1 \sim \frac{\chi^2(n, \mathbf{r}_1^T \mathbf{A}_1^{-1} \mathbf{r}_1)}{\chi^2(n-q+1)}
\]

and

\[
W_2 | W_1 \sim \frac{\chi^2(p-q, \mathbf{r}_2^T \mathbf{A}_2^{-1} \mathbf{r}_2)}{\chi^2(n-p+1)}
\]

Now, let \( \delta_1 = \xi_1^0, (\mathbf{A}_1^{-1})^{-1} \xi_1^0 \) and \( \delta_2 = \xi_2^0, (\mathbf{A}_2^{-1})^{-1} \xi_2^0 \). Under \( H_0 \), \( \delta_2 = 0 \) and \( W_1 \) is sufficient for \( \delta_1 \). Also, under \( H_0 \), the distribution of \( W_1 \) is boundedly complete. Thus, all similar tests of \( H_0 \) versus \( H_1 \) will have Neyman Structure. However, we know that the most powerful test based on \( W_2 | W_1 \) is to reject for large values of \( W_2 \) and the rejection point does not depend on \( W_1 \) since \( W_1 \) and \( W_2 \) are independent under \( H_0 \). Thus, we conclude that the uniformly most powerful invariant similar test is to reject for large values of \( W_2 \).
Summarizing, we have shown that for testing \( H_0: \varphi_2^0 = 0, \varphi_3^0 = 0 \) versus \( H_1: \varphi_2^0 = 0 \), the test which rejects for large values of

\[
\frac{T_2^2 - T_1^2}{1 + T_1^2}
\]

where \( T_2^2 = (\iota_2' X)' (\iota_2' \delta_2) ^{-1} (\iota_2' X) \) and \( T_1^2 = (\iota_1' \iota_2' X)' (\iota_1' \delta_1 \iota_2) ^{-1} (\iota_1' \iota_2' X) \) is uniformly most powerful invariant similar. The reader should contrast this situation with that occurring in Case III of the mean testing problem.

Correlation

In this section, we are concerned with the distribution of correlation coefficients in normal populations. We first treat the bivariate normal correlation coefficient.

Let \( S \) be \( W(\Gamma, n) \) where \( S \) is \( 2 \times 2 \). Recall that the population correlation coefficient is

\[
\rho = \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}} \sqrt{\Gamma_{22}}}
\]

and the sample correlation coefficient is

\[
r = \frac{S_{12}}{\sqrt{S_{11}} S_{22}}.
\]

To study the distribution of \( r \), consider the following. Let
\[
A_{11} = S_{11} - S_{12} S_{22}^{-1} S_{21}
\]

\[
A_{12} = S_{12} S_{22}^{-\frac{1}{2}}
\]

\[
A_{22} = S_{22}
\]

so that \( A_{11} \) is independent of \((A_{12}, A_{22})\), \( A_{11} \) is \( W(\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}, n-1) \).

\[
A_{12} | A_{22} \sim N(\Gamma_{12} \Gamma_{22}^{-\frac{1}{2}} A_{22}^{-\frac{1}{2}} \Gamma_{21} \Gamma_{11}^{-1} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}, \Gamma_{11} \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}) \]

and \( A_{22} \sim W(\Gamma_{22}, n) \). Thus

\[
r = \frac{S_{12}}{\sqrt{S_{11} S_{22}}} = \frac{A_{12}}{A_{11} + A_{12}^2}.
\]

Setting \( u = A_{12} / A_{11} \), we have

\[
r = \frac{u}{\sqrt{1 + u^2}}
\]

so that

\[
u = \frac{r}{\sqrt{1 - r^2}}
\]

and so it suffices to consider the distribution of \( u \). From the distributions of \( A_{12} \) and \( A_{11} \), we easily find that

\[
u \sim \frac{Y}{\sqrt{\chi^2_{n-1}}}
\]

58
where \( Y \) is \( N \left( \frac{\rho}{1-\rho^2}, \chi^2_{n-1} \right) \) and is independent of \( \chi^2_{n-1} \). When \( \rho = 0 \), we obtain the well known result that \( \frac{r}{\sqrt{1-r^2}} \) has a \( t_{n-1} \)-distribution. When \( \rho \neq 0 \), the density of \( u \) (and correspondingly for \( r \)) is expressed as an infinite series. For explicit expressions, the reader is referred to T. W. Anderson, *Introduction to Multivariate Statistical Analysis*.

We now turn to discussion of multiple correlation. Let \( X = \begin{pmatrix} Y \\ Z \end{pmatrix} \) be a random vector of dimension \( p \) and suppose \( Y \) is of dimension 1.

Let \( X \) have covariance matrix \( \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \). The multiple correlation between \( Y \) and \( Z \) is defined to be the maximum correlation between \( Y \) and all linear functions of \( Z \). Thus

\[
\rho = \max_a \text{Corr}_\Gamma(Y, a'Z) = \max_a \frac{a'\Gamma_{21}}{\sqrt{a'\Gamma_{11}a}}
\]

where \( a \) runs over all \( p-1 \) dimensional vectors. Setting \( b = \frac{\Gamma_{12}^{\frac{1}{2}}}{\sqrt{\Gamma_{22}}}a \),

\[
\rho = \max_b \frac{b'\Gamma_{22}^{-\frac{1}{2}}\Gamma_{21}}{\sqrt{\Gamma_{11}b'b}} = \max_b \frac{b'\Gamma_{22}^{-\frac{1}{2}}\Gamma_{21}}{\sqrt{b'b}} \leq \frac{\sqrt{b'b}}{\sqrt{\Gamma_{11}}} \frac{\sqrt{\Gamma_{12}^{-1}\Gamma_{22}^{-1}}} {\sqrt{\Gamma_{11}}} = \frac{\Gamma_{12}^{-\frac{1}{2}}}{\sqrt{\Gamma_{11}}}
\]

since \( b'b = 1 \), where we have used the Cauchy-Schwarz Inequality. Thus, we have equality if and only if

\[
b = \frac{\Gamma_{22}^{-\frac{1}{2}}\Gamma_{21}}{\sqrt{\Gamma_{12}^{-1}\Gamma_{22}^{-1}}}.
\]
Therefore

\[ \rho = \text{Corr}_\Gamma (Y, \Gamma^{-\frac{1}{2}} Z) = \sqrt{\frac{\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}}{\Gamma_{11}}} \]

so that

\[ \rho^2 = \frac{\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}}{\Gamma_{11}}. \]

Now, consider \( S \sim W(\Gamma, n) \) where \( S \) is \( p \times p \) and write

\[ S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \]

and

\[ \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}, \]

where \( S_{11} \) is \( 1 \times 1 \) and \( \Gamma_{11} \) is \( 1 \times 1 \). The sample multiple correlation coefficient is defined by

\[ R = \sqrt{\frac{S_{12} S_{22}^{-1} S_{21}}{S_{11}}} \]

Thus

\[ \frac{R^2}{1-R^2} = \frac{S_{12} S_{22}^{-1} S_{21}}{S_{11} - S_{12} S_{22}^{-1} S_{21}}. \]

However,

\[
\frac{S_{11} - S_{12}S_{22}^{-1}S_{21}}{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}
\]

is

\[
\chi^2_{n-p+1}
\]

and is independent of \((S_{12}, S_{22})\). Thus

\[
\frac{R^2}{1-R^2} \sim \frac{1}{\chi^2_{n-p+1}} \cdot \frac{S_{12}S_{22}^{-1}S_{21}}{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}.
\]

However, \(\frac{S_{12}S_{22}^{-1/2}}{\sqrt{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}} \) is \(\mathcal{N} \left( \frac{\Gamma_{12}^{-1}\Gamma_{22}^{-1/2}S_{22}}{\sqrt{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}} , I \right)\) so that

\[
\frac{R^2}{1-R^2} \sim \frac{1}{\chi^2_{n-p+1}} \chi^2_{p-1} \left( \frac{\Gamma_{12}^{-1}\Gamma_{22}^{-1/2}S_{22}}{\sqrt{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}} \right).
\]

Also, \(\frac{\Gamma_{12}^{-1}\Gamma_{22}^{-1/2}}{\sqrt{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}} \sim \chi^2_n\) and noting that \(\frac{\Gamma_{12}^{-1}\Gamma_{22}^{-1/2}}{\sqrt{\Gamma_{11} - \Gamma_{12}^{-1}\Gamma_{22}^{-1}\Gamma_{21}}} = \frac{\rho^2}{1-\rho^2}\),

\[
\frac{R^2}{1-R^2} \sim \frac{1}{\chi^2_{n-p+1}} \chi^2_{p-1} \left( \frac{\rho^2}{1-\rho^2} \chi^2_n \right).
\]

Since \(\chi^2_m(\lambda) \sim \chi^2_{m+2K}\), where \(K\) is Poisson with parameter \(\frac{\lambda}{2}\), it follows that \(\chi^2_{p-1} \left( \frac{\rho^2}{1-\rho^2} \chi^2_n \right) \sim \chi^2_{p-1+2K}\) where \(K|\chi^2_n\) is Poisson with
parameter $\frac{1}{2} \frac{\sigma^2}{1-\rho^2} \chi^2_n$. Setting $\frac{\lambda}{2} = \frac{1}{2} \frac{\rho^2}{1-\rho^2}$, we have

$$P(K = k) = \int_0^\infty e^{-\frac{\lambda}{2} x} \left( \frac{\lambda}{2} \right)^{\frac{n-k}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+k}{2}\right)} e^{-\frac{x}{2}} dx = \frac{\Gamma\left(\frac{n+k}{2}\right)}{k! \Gamma\left(\frac{n}{2}\right)} \left(\frac{\lambda}{2}\right)^k \frac{n}{(1-\lambda)^{n+k}}$$

Thus

$$\frac{R^2}{1-R^2} \sim \frac{\chi^2_{p+2k}}{\chi^2_{n-p-1}}$$

where

$$P(K = k) = \frac{\Gamma\left(\frac{n+k}{2}\right)}{k! \Gamma\left(\frac{n}{2}\right)} \left(\frac{\rho^2}{1-\rho^2}\right)^k \frac{n}{(1-\rho^2)^n},$$

for $k = 0, 1, 2, \ldots$. Note that $K$ has a negative binomial distribution.

From the above, it follows that if $\rho = 0$, then $R^2/(1-R^2)$ has an $F$-distribution (up to a constant multiple).

Now, let $X = \begin{pmatrix} Y \\ Z \end{pmatrix}$ where $Y$ is of dimension $q > 1$ and $Z$ is of dimension $p - q$. If $X$ is $N(0, \Gamma)$ and $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$, then

$$Y|Z \sim N(\Gamma_{11} \Gamma_{22}^{-1} Z, \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}).$$
Let $Y = (Y_1, Y_2, \ldots, Y_q)$. The partial multiple correlation of $Y_1$ with $Y_2 \ldots Y_q$ holding $Z$ fixed is defined to be the multiple correlation of $Y_1$ with $Y_2, \ldots, Y_q$ in the conditional distribution of $Y$ given $Z$. For $q = 2$, the partial correlation of $Y_1$ with $Y_2$ holding $Z$ fixed is just the correlation of $Y_1$ and $Y_2$ in the conditional distribution of $(Y_1, Y_2)$ given $Z$. Of course, all of the distribution theory developed previously holds for the partial multiple correlation coefficient and the partial correlation coefficient, where in this case we work with the conditional distribution and the conditional covariance matrix.

Section 5.

The main concern of this section is the distribution of the roots of certain determinantal equations. However, related topics such as non-central distributions, the orthogonal group, and canonical correlations are discussed. The section is concluded with an outline of a large sample result concerning the characteristic roots of a Wishart matrix. Certain results on the Jacobians of matrix transformations discussed in notes by I. Olkin are assumed known in this section.

Let $A$ and $B$ be $p \times p$ symmetric matrices which are $W(\Sigma, p, n_1)$ and $W(\Sigma, p, n_2)$ respectively where $n_i \geq p$ for $i = 1, 2$. We are interested in the distribution of the roots of the equation

$$|A - \lambda B| = 0.$$ 

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ denote the ordered roots. Rather than find
the distribution of \( \{ \lambda_i \} \) directly, it is convenient to first find the
law of \( \{ f_i \} \) where \( f_i = \lambda_i / (1 + \lambda_i) \). Thus the \( \{ f_i \} \) are the roots of

\[
|A - f(A+B)| = 0.
\]

Let \( F \) be a \( p \times p \) diagonal matrix with diagonal elements \( f_i \). Then
write

\[
A + B = WW' \\
A = WFW'
\]

where \( w_{ii} \geq 0, \ i = 1, \ldots, p \). This transformation is 1-1 with prob-
ability 1 since the \( f_i \) are distinct, and \( w_{ii} \geq 0, \ i = 1, \ldots, p \).
To find the Jacobian of this transformation, let \( C = A + B \) so that

\[
J((A,B) \rightarrow (W,F)) = J((A,B) \rightarrow (A,C)) \ J((A,C) \rightarrow (W,F)).
\]

Clearly \( J((A,B) \rightarrow (A,C)) = 1 \) and it is well known that

\[
J((A,C) \rightarrow (W,F)) = 2^p |W|^{p+2} \prod_{i<j} (f_i - f_j).
\]

Now, without loss of generality take \( \Sigma = I \) so that the joint density of
\( A \) and \( B \) is

\[
p(A,B) = C_p(\nu_1)C_p(\nu_2)|A|^{-\frac{1}{2}}|B|^{-\frac{1}{2}}(\nu_1 - p - 1)^{\frac{1}{2}}(\nu_2 - p - 1)^{\frac{1}{2}} e^{-\frac{1}{2}tr(A+B)}
\]

where
\[ C_p(n) = \left[ \frac{np}{2^p \pi^{p/2}} \prod_{i=1}^p \Gamma \left( \frac{n-i+1}{2} \right) \right]^{-1} \]

is the constant in the Wishart distribution. Hence the joint density of \( W \) and \( F \) is

\[
p(F, W) = C_p(n_1)C_p(n_2) \left( \prod_{i=1}^p f_i \right)^{\frac{n_1-p-1}{2}} \left( \prod_{i=1}^p (1 - f_i) \right)^{\frac{n_2-p-1}{2}} \times \prod_{i<j} (f_i - f_j) |WW'|^{-\frac{1}{2}} 2^{p-\frac{1}{2}} e^{-\frac{1}{2}trWW'}.
\]

We now want to integrate out \( W \) to find the density of \( F \). Thus we want to find the constant

\[
C_1 = \left( \sqrt{2\pi} \right)^p \int \cdots \cdots \frac{n_1+n_2-p}{2} \frac{n_1+n_2-p}{2} e^{-\frac{1}{2}trWW'} dW
\]

\[ = \left( \sqrt{2\pi} \right)^p \mathcal{E} \left( \frac{n_1+n_2-p}{2} \frac{n_1+n_2-p}{2} |WW'|^{-\frac{1}{2}} \right) \]

where \( W \) is a \( p \times p \) matrix of independent \( N(0, I) \) variables. Thus \( WW' \) is \( W(I, p, p) \) so that \( S = WW' \) has a density given by

\[ p(S) = C_p(p) |S|^{-\frac{1}{2}} e^{-\frac{1}{2}trS} dS.\]

Hence
\[ c_1 = (\sqrt{2\pi})^p \mathcal{C} \left( \frac{n_1 + n_2 - p}{2} \right) = (\sqrt{2\pi})^p c_p(p) \int |s|^{\frac{n_1 + n_2 - p - 1}{2}} e^{-\frac{1}{2} \text{tr} s s} ds \]

\[ = (\sqrt{2\pi})^p c_p(p)/c_p(n_1 + n_2). \]

Thus the density of \( F \) is

\[ p(F) = (\sqrt{2\pi})^p \frac{c_p(n_1) c_p(n_2) c_p(p)}{c_p(n_1 + n_2)} \left( \prod_{i=1}^p \frac{n_i - p - 1}{2} \right) \left( \prod_{i=1}^p (1 - f_i) \right) \left( \prod_{i<j} (f_i - f_j) \right) \]

where \( 1 > f_1 > f_2 > \ldots > f_p > 0 \). Simplification of the constant yields

\[ p(F) = c_2 \left( \prod_{i=1}^p f_i \right)^{\frac{n_1 - p - 1}{2}} \left( \prod_{i=1}^p (1 - f_i) \right)^{\frac{n_2 - p - 1}{2}} \prod_{i<j} (f_i - f_j) \]

where

\[ c_2 = \frac{\pi^p}{\prod_{i=1}^p \Gamma \left( \frac{n_i - 1}{2} \right) \Gamma \left( \frac{1}{2} \right)} \prod_{i=1}^p \Gamma \left( \frac{n_i + n_{i+1} \cdot 1}{2} \right). \]

Now, we let \( f_i = \frac{l_i}{1 + l_i} \) so that

\[ df_i = \frac{1}{(1 + l_i)^2} dl_i \]

and the density of \( L = (l_1, l_2, \ldots, l_p) \) is
\[ p(L) = c_2 \left( \prod_{i=1}^{p} \ell_i \right)^{n_1-p-1} \left( \prod_{i=1}^{p} \left( 1 + \ell_i \right) \right)^{\frac{n_1-n_2}{2}} \prod_{i<j} (\ell_i-\ell_j) \]

where the \( \{\ell_i\} \) are the roots of

\[ |A - \ell B| = 0 \]

and

\[ \ell_1 > \ell_2 > \ldots > \ell_p > 0. \]

We now consider a \( p \times p \) matrix \( A \) which is \( W(I,p,n) \) and the
distribution of the roots of the equation \( |A - \lambda I| = 0 \) is desired.

Let \( \lambda_1 > \lambda_2 > \ldots > \lambda_p > 0 \) denote the roots, and consider \( B \) distributed as \( W(I,p,n_2) \). Since \( \frac{B}{n_2} \rightarrow I \) a.s. as \( n_2 \rightarrow \infty \), the roots of the

equation \( |A - \lambda \frac{B}{n_2}| = 0 \) converge a.s. to the roots of the equation

\[ |A - \lambda I| = 0. \]

Thus, to find the density of the roots of \( |A - \lambda I| = 0 \),
it is sufficient to find the limit as \( n_2 \rightarrow \infty \) of the density of the
roots of the equation \( |A - \lambda \frac{B}{n_2}| = 0 \). As derived above, the density of

the roots of \( |A - \lambda \frac{B}{n_2}| = 0 \) is

\[
p_{n_2}(\lambda) = c_2 \left( \prod_{i=1}^{p} \lambda_i \right)^{\frac{n-p-1}{2}} \frac{1}{\left( \frac{n-p-1}{2} \right)} \prod_{i=1}^{p} \left( 1 + \frac{\lambda_i}{n_2} \right)^{\frac{n+2}{2}} \prod_{i<j} \left( \frac{\lambda_i-\lambda_j}{n_2} \right)
\]

\[
= c_2 \prod_{i=1}^{p} \lambda_i^{\frac{n-p-1}{2}} \prod_{i=1}^{p} \left( 1 + \frac{\lambda_i}{n_2} \right)^{\frac{n+2}{2}} \prod_{i<j} \left( \frac{\lambda_i-\lambda_j}{n_2} \right).
\]
Clearly,

\[
\lim_{n_2 \to \infty} \prod_{i=1}^{p} \left(1 + \frac{\lambda_i}{n_2} \right)^{\frac{n+n_2}{2}} = e^{\frac{p}{2} \sum_{i=1}^{p} \lambda_i}.
\]

However, it is easy to show that

\[
\lim_{n_2 \to \infty} \frac{1}{n_2} \frac{n+n_2-1}{\Gamma\left(\frac{n+n_2}{2}\right)} = \frac{1}{\sqrt{2}}
\]

so that

\[
C_3 = \lim \frac{c_2(n_2)}{\frac{n_2}{\pi} \frac{p}{\sqrt{2} \prod_{i=1}^{p} \Gamma\left(\frac{n_2+n_2-i+1}{2}\right)}}.
\]

Hence the density of the roots to the equation \(|A - \lambda I| = 0\) is given by

\[
p(\lambda) = C_3 \left( \prod_{i=1}^{p} \lambda_i \right)^{\frac{n-p-1}{2}} e^{-\frac{p}{2} \sum_{i=1}^{p} \lambda_i} \prod_{i<j} (\lambda_i - \lambda_j)
\]

with \(C_3\) given above.

Remark: Another approach to finding the distribution of the roots of \(|A - \lambda I|\) where \(A\) is \(W(I,p,n)\) is the following. Write

\[
A = \Gamma \Delta \Gamma'
\]
where $\gamma_{ii} \geq 0$ for $i = 1, \ldots, p$ so that the decomposition of $\mathbf{A}$ is unique with probability one. It is shown in notes by I. Olkin that

$$J(A \to (\Gamma, D_\lambda)) = h(\Gamma) \prod_{i < j} (\lambda_i - \lambda_j)$$

where $h$ is a function of $\Gamma$. Thus the joint density of $\Gamma$ and $D_\lambda$ is

$$p(\Gamma, D_\lambda) = C_p(n) \prod_{i=1}^{p} \lambda_i^{n-p-1} \prod_{i < j} \lambda_i \lambda_j^{p-1} e^{-\frac{1}{2} \sum_{i=1}^{p} \lambda_i} \prod_{i < j} (\lambda_i - \lambda_j) h(\Gamma)$$

so that $\Gamma$ and $D_\lambda$ are independent. Integrating out $\Gamma$ gives

$$p(\lambda) = C_4 \prod_{i=1}^{p} \lambda_i^{n-p-1} \prod_{i < j} \lambda_i \lambda_j^{p-1} e^{-\frac{1}{2} \sum_{i=1}^{p} \lambda_i} \prod_{i < j} (\lambda_i - \lambda_j)$$

However, this method does not yield the value of the constant $C_4$.

We now digress for a moment to consider the problem of finding the density of a maximal invariant under very special conditions. Let $(\mathcal{X}, \mathfrak{B}, \lambda)$ be a measure space and suppose $p$ is a density w.r.t. $\lambda$. Consider a group $\mathcal{G}$ operating on the left of $\mathcal{X}$ and suppose the measure $\lambda$ is invariant; that is,

$$\lambda(gB) = \lambda(B)$$

for $B \in \mathfrak{B}$, $g \in \mathcal{G}$. Let $(\mathcal{Y}, \mathfrak{A})$ be a measurable space and suppose
\( \tau \) is a measurable mapping from \( \mathcal{X} \) into \( \mathcal{Y} \) such that \( \tau \) is a maximal invariant; that is,

1. \( \tau(gx) = \tau(x) \) for \( x \in \mathcal{X} \), \( g \in \mathcal{G} \),
2. \( \tau(x_1) = \tau(x_2) \) implies there exists a \( g \in \mathcal{G} \) such that \( g \cdot x_1 = x_2 \).

Let \( \lambda^* \) be the measure induced in \( (\mathcal{Y}, \mathcal{B}_2) \) by \( \lambda \) and \( \tau \); that is,

\[
\lambda^*(D) = \lambda(\tau^{-1}(D))
\]

for \( D \in \mathcal{B}_2 \). Also, let \( P_2 \) be the probability measure on \( (\mathcal{Y}, \mathcal{B}_2) \) defined by

\[
P_2(D) = \int_{\tau^{-1}(D)} p(x) \, d\lambda(x)
\]

so that \( P_2 \) has a density, say \( p^* \), with respect to \( \lambda^* \). Hence

\[
\int_D p^*(\tau) \, d\lambda^*(\tau) = \int_{\tau^{-1}(D)} p(x) \, d\lambda(x).
\]

We now assume there exists an invariant probability measure \( \mu \) on the group \( \mathcal{G} \).

**Lemma:** The density \( p^* \) of the maximal invariant \( \tau \) (with respect to \( \lambda^* \)) is given by

\[
p^*(\tau(x)) = \int_{\mathcal{G}} p(g^{-1}x) \, d\mu(g).
\]
Proof: Let \( f \) be any integrable function on \((y, B_n)\). Then

\[
\int_y f(\tau) \ p^*(\tau) \ d\lambda^*(\tau) = \int_y f(\tau(x)) \ p(x) \ d\lambda(x).
\]

However,

\[
\int_y f(\tau(x)) \int_y p(g^{-1}x) \ d\mu(g) \ d\lambda(x) \\
= \int_y \int_y f(\tau(x)) \ p(g^{-1}x) \ d\lambda(x) \ d\mu(g) = (y = g^{-1}x) \\
\int_y \int_y f(\tau(y)) \ p(y) \ d\lambda(y) \ d\mu(g) \\
= \int_y f(\tau(y)) \ p(y) \ d\lambda(y).
\]

Then, since the function \( \int_y p(g^{-1}x) \ d\mu(g) \) is invariant, we have

\[
p^*(\tau(x)) = \int_y p(g^{-1}x) \ d\mu(g). \|
\]

For future application, we now prove the existence of an invariant probability measure on the group of \( p \times p \) orthogonal matrices, \( \mathcal{O}(p) \). Let \( X = (X_1, \ldots, X_p) \) be a \( p \times p \) random matrix with the \( i \)-th column \( X_i \) being \( N(0, I_p) \). Apply the Gram-Schmidt orthogonalization process to the columns of \( X \) to obtain the random orthogonal matrix \( G = (Y_1, \ldots, Y_p) \) where

\[
Y_i = \frac{X_i - \sum_{j=1}^{i-1} (X_i, X_j) Y_j}{\|X_i - \sum_{j=1}^{i-1} (X_i, X_j) Y_j\|}.
\]

\[ Y_i, \ldots, Y_p \]
Let $T$ denote the above transformation, that is, $G = T(X)$. Let $g \in O(p)$ so that we must show $gG$ and $G$ have the same distribution. But, $gG = (gY_1, \ldots, gY_p)$ so that if we let $X_1^* = gX_1$, then $gG = T(X_1^*, \ldots, X_p^*)$. However, $X_1$ and $X_1^*$ have the same distribution so that $G$ and $gG$ have the same distribution for each $g \in O(p)$. Thus the probability measure of $G$ defined by

$$P_G(A) = P_X(T^{-1}(A))$$

is invariant on $O(p)$.

We now give two examples.

**Example 1:** Let $X = (X_1, \ldots, X_n)$ be a $p \times n$ matrix $(n \geq p)$ where each column $X_i$ is $N(\mu_i, I_p)$ and independent of the remaining columns of $X$. Let $\mu$ be the $p \times n$ matrix $(\mu_1, \ldots, \mu_n)$. The density of $X$ with respect to $p^n$ dimensional Lebesgue measure is

$$p(x, \mu) = \frac{1}{p^n} e^{-\frac{1}{2} \text{tr} XX'} e^{-\frac{1}{2} \text{tr} \mu \mu'} e^{\text{tr} X \mu}.$$ 

Now, it follows directly from Vinograd's Theorem that $S = XX'$ is a maximal invariant under the transformation

$$X \rightarrow X \Gamma$$

where $\Gamma \in O(n)$, the group of $n \times n$ orthogonal matrices. Let $q(S, \mu)$

72
denote the density of $S$ at the parameter value $\mu$ with respect to the induced measure, $d\lambda^*(S)$. To find $d\lambda^*(S)$, we first note that

$$q(S,0) \ d\lambda^*(S) = C_p(n)|S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\text{tr} S} dS$$

since $S$ has a central Wishart distribution when $\mu = 0$. However, we also know that

$$q(S,0) \ d\lambda^*(S) = \int_{\mathcal{O}(n)} p(x\Gamma,0) \ dv(\Gamma) \ d\lambda^*(S)$$

$$= \frac{1}{pn\sqrt{2\pi}} e^{-\frac{1}{2}\text{tr} S} \left[ \int_{\mathcal{O}(n)} dv(\Gamma) \right] d\lambda^*(S)$$

$$= \frac{1}{pn\sqrt{2\pi}} e^{-\frac{1}{2}\text{tr} S} d\lambda^*(S)$$

where $v$ is the invariant probability measure on $\mathcal{O}(n)$. Hence we conclude that

$$d\lambda^*(S) = \frac{pn}{\sqrt{2\pi}} C_p(n)|S|^{\frac{n-p-1}{2}} dS.$$ 

Thus we have that the density of $S$ at the parameter value $\mu$ with respect to the measure $d\lambda^*(S)$ is given by

$$q(S,\mu) = e^{-\frac{1}{2}\text{tr} S} e^{-\frac{1}{2}\text{tr} \mu \mu'} \int_{\mathcal{O}(n)} e^{\text{tr} X \Gamma \mu'} dv(\Gamma)$$

$$= e^{-\frac{1}{2}\text{tr} S} e^{-\frac{1}{2} \text{tr} \mu \mu'} h(x,\mu).$$
Note that \( h(X,\mu) = h(X^\Gamma,\mu^\Psi) \) for all \( \Gamma,\Psi \in \Theta(n) \). Thus we can write
\[
h(X,\mu) = f(S,\mu') .
\]

The above density \( q(S,\mu) \) is called the density function of the non-central Wishart distribution with non-centrality parameter \( \mu' \). The reader is referred to the papers of A. T. James for more explicit (but very complicated) expressions for the function \( f(S,\mu') \).

**Example 2:** Let \( S \) be \( W(\Sigma,p,n) \) and consider the roots of the equation \(|S-\lambda I| = 0\), say \( \lambda_1 > \lambda_2 > \ldots > \lambda_p > 0 \) are the roots, and let \( \lambda = (\lambda_1,\ldots,\lambda_p) \). We have found the density of \( \lambda \) when \( \Sigma = I \),
\[
p(\lambda,I) d\eta^*(\lambda) = C_3 \left( \prod_{i=1}^{p} \frac{n-p-1}{2} \lambda_i \right)^{-\frac{p}{2}} e^{-\frac{1}{2} \sum_{i=1}^{p} \lambda_i} \prod_{i<j} (\lambda_i - \lambda_j) d\lambda
\]
where \( d\lambda \) is Lebesgue measure. However, we know that \( \lambda \) is a maximal invariant under the transformation
\[
S \rightarrow \Gamma S \Gamma'
\]
where \( \Gamma \in \mathcal{O}(p) \). Thus
\[
p(\lambda,I) d\eta^*(\lambda) = C_p(n) \left( \prod_{i=1}^{p} \lambda_i \right)^{\frac{n-p-1}{2}} e^{-\frac{1}{2} \text{tr} \Gamma S \Gamma} \int_{\mathcal{O}(p)} e^{-\frac{1}{2} \text{tr} \Gamma S \Gamma} d\eta(\Gamma) d\eta^*(\lambda)
\]
\[
= C_p(n) \left( \prod_{i=1}^{p} \lambda_i \right)^{\frac{n-p-1}{2}} e^{-\frac{1}{2} \sum_{i=1}^{p} \lambda_i} d\eta^*(\lambda)
\]

74
so that

\[
d\eta^*(\lambda) = \frac{C_3}{C_p(n)} \prod_{i < j} (\lambda_i - \lambda_j) d\lambda.
\]

Thus we know that the non-central density of \( \lambda \) (i.e., \( \Sigma \neq I \)) with respect to \( d\eta^* \) is given by

\[
p(\lambda, \Sigma) = |\Sigma|^{-\frac{n}{2}} \left( \prod_{i=1}^{p} \lambda_i \right) \frac{n-p-1}{2} e^{\frac{1}{2} \text{tr} \Sigma^{-1} D_{\lambda} \Sigma} \mathcal{N}(\mu, \Sigma) d\nu(\Gamma)
\]

where \( D_{\lambda} \) is a \( p \times p \) diagonal matrix with the \( \lambda_i \) on the diagonal.

Again, it is clear that the density \( p(\lambda, \Sigma) \) depends on \( \Sigma \) only through the characteristic roots of \( \Sigma \). Again we refer the reader to the work of A. T. James for explicit expressions for the integral in the above representation of \( p(\lambda, \Sigma) \).

Canonical Correlations:

Consider a random vector \( X = (Y, Z) \) where \( Y \) is \( p \times 1 \) and \( Z \) is \( q \times 1 \) with \( p \leq q \). Assume \( X \) has a covariance matrix \( \Sigma \) which we partition (as \( X \) is partitioned)

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

where \( \Sigma_{11} \) is \( p \times p \), \( \Sigma_{12} \) is \( p \times q \) and \( \Sigma_{22} \) is \( q \times q \). Recall that if \( p = 1 \), then the multiple correlation coefficient is the largest correlation attainable between \( X \) and linear combinations of \( Z \). For \( p > 1 \), a natural generalization of the multiple correlation coefficient is the largest canonical correlation, say \( p_1 \), defined to be the largest correlation attainable between linear combinations of \( Y \) and linear combinations of \( Z \).
Thus

$$\rho_1^2 = \sup_{a \in \mathbb{R}^p} \sup_{b \in \mathbb{R}^q} \text{Correl}(a'y, b'z)$$

$$= \sup_{a \in \mathbb{R}^p} \sup_{b \in \mathbb{R}^q} \frac{(a'\Sigma_{12}b)^2}{a'a b'b} = \sup_{a \in \mathbb{R}^p} \sup_{b \in \mathbb{R}^q} \frac{(a'\Sigma_{12}^{\frac{1}{2}} \Sigma_{22}^{\frac{1}{2}} b)^2}{a'a b'b}$$

Let \( \Lambda = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} \) so that \( \Lambda \) is \( p \times q \) and \( \Lambda \) can be written

\[ \Lambda = \Gamma \Lambda' \Psi' \]

\( \Gamma \) is \( p \times p \) and orthogonal, \( \Lambda' \) is \( p \times p \) and diagonal and the elements of \( \Lambda' \) are the eigenvalues of \( \Lambda \Lambda' \), and \( \Psi \) is \( q \times p \) and is column orthogonal. Let \( \theta_1 \geq \theta_2 \geq \cdots \geq \theta_p \geq 0 \) be the ordered roots of \( \Lambda \Lambda' \), and let \( \gamma_1, \ldots, \gamma_p \) denote the \( p \) columns of \( \Lambda' \), and \( \Psi_1, \ldots, \Psi_p \) denote the \( p \) columns of \( \Psi \). Now, it is clear that

$$\rho_1^2 = \sup_{a \in \mathbb{R}^p} \sup_{b \in \mathbb{R}^q} \frac{(a'b)^2}{a'a b'b} \leq \sup_{\|a\|=1} (a'\Lambda a)^2 = \theta_1^2$$

However, if we choose \( a = \gamma_1 \) and \( b = \Psi_1 \) the value of \( \theta_1^2 \) is attained.

Hence, we have

$$\rho_1^2 = \sup_{a \in \mathbb{R}^p} \sup_{b \in \mathbb{R}^q} \frac{(a'\Sigma_{12}b)^2}{a'a b'b} = \theta_1^2$$
where the $\theta_i^2$, $i = 1, \ldots, p$ are the roots of the matrix $\Lambda \Lambda'$ and
\[ \Lambda = \Sigma^{-\frac{1}{2}}_{11} \Sigma_{12} \Sigma^{-\frac{1}{2}}_{22} . \]
If we write $\Lambda = \Gamma D \Psi$, then the maximizing $a$ and $b$ are given by $a_1 = \Sigma_{11}^{-\frac{1}{2}} \gamma_1$ and $b_1 = \Sigma_{22}^{-\frac{1}{2}} \psi_1$.

Now let us maximize the correlation between $a'Y$ and $b'Z$ subject to the side condition that $a'Y$ and $a_1'Y$ be uncorrelated and $b'Z$ and $b_1'Z$ be uncorrelated. This new maxim correlation, say $\rho_2$, is defined to be the 2nd largest canonical correlation. Thus
\[
\rho_2^2 = \sup_{a, b} \frac{(a' \Sigma_{12} b)^2}{a' \Sigma_{11} a \cdot b' \Sigma_{22} b} = \sup_{a, b} \frac{(a' \Lambda b)^2}{a' \Lambda a \cdot b' \Lambda b} = \sup_{a, b} \frac{(a' \Lambda b)^2}{a' a \cdot b' b},
\]
\[
\text{subject to } a_1 = 0 \quad \text{and } b_1 = 0,
\]
\[
\text{subject to } \gamma_1 = 0 \quad \text{and } \psi_1 = 0
\]
\[< \sup_{\|a\| = 1 \quad a_1' \gamma_1 = 0} (a' \Lambda a)^2 = \theta_2^2.
\]

However this upper bound is attained for $a_2 = \Sigma_{11}^{-\frac{1}{2}} \gamma_2$ and $b_2 = \Sigma_{22}^{-\frac{1}{2}} \psi_2$.

Thus $\rho_2^2 = \theta_2^2$. Continuing in the obvious manner, we obtain vectors
\[a_i = \Sigma_{11}^{-\frac{1}{2}} \gamma_i, \quad i = 1, \ldots, p\]
\[b_i = \Sigma_{22}^{-\frac{1}{2}} \psi_i, \quad i = 1, \ldots, p\]
such that if we set $W_i = a_i'Y$ and $V_i = b_i'Z$, $W = (W_1, \ldots, W_p)'$, $V = (V_1, \ldots, V_p)'$, then the covariance matrix of the vector $T = (W, V)'$ has the form
\[
\Sigma_T = \begin{pmatrix} I_p & D \theta \\ D \theta^T & I_p \end{pmatrix}.
\]

Also, the correlation between $W_i$ and $V_i$ has been maximized subject to the conditions that $W_i$ (resp. $V_i$) be a linear function of $X$ (resp. $Y$)
and that $W_i$ (resp. $V_i$) be uncorrelated with $(W_1', ..., W_{i-1})$
(resp. $(V_1', ..., V_{i-1})$).

These $p$ correlations, $\rho_1 \geq \rho_2 \geq ... \geq \rho_p \geq 0$, are called the canonical correlations between $Y$ and $Z$. Also, we have seen that these canonical correlations are the square roots of the roots of the equation

$$|\Lambda' - \theta I| = 0 = |\Sigma^{-\frac{1}{2}} \Sigma_{12} \Sigma^{-1} \Sigma_{21} \Sigma^{-\frac{1}{2}} - \theta I|$$

or equivalently,

$$|\Sigma_{12} \Sigma^{-1} \Sigma_{21} - \theta \Sigma_{11}| = 0.$$ 

Now, let $S$ be $W(\Sigma, r, n)$ and write $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ where $S_{11}$ is $p \times p$, $S_{22}$ is $q \times q$ and $p \leq q$, $p + q = r$. We then define the sample canonical correlations squared to be the roots of the equation

$$|S_{12} S^{-1} S_{21} - \lambda S_{11}| = 0$$

which we can also write as

$$|B - \lambda (A + B)| = 0$$

where $B = S_{12} S^{-1} S_{21}$ and $A = S_{11} - S_{12} S^{-1} S_{21}$. Now, it has been shown previously that $A$ and $B$ are independent,

$$A \text{ is } W(\Sigma_{11}, \Sigma_{12} \Sigma^{-1} \Sigma_{22}, p, n-q)$$

and

$$S_{12} S^{-\frac{1}{2}} | S_{22} \text{ is } N(\Sigma_{11}, \Sigma_{12} \Sigma^{-1} \Sigma_{21} \otimes I_q)$$

where $B = S_{12} S^{-1} S_{21}$. Hence, if $\Sigma_{12} = 0$, then

$$A \text{ is } W(\Sigma_{11}, p, n-q)$$

78
and \( B \) is \( W(S_{11}, p, q) \). Thus, in the case \( \Sigma_{12} = 0 \), the square of the sample canonical correlation is the roots to the equation

\[
|B - \lambda(A + B)| = 0
\]

where \( A \) and \( B \) are independent Wisharts with the same covariance matrix. The distribution of these roots, say \( \lambda_1 > \lambda_2 > \ldots > \lambda_p \), has already been found and is given by

\[
p(\lambda) = C_2 \left( \prod_{i=1}^{p} \lambda_i \right)^{\frac{n-p-1}{2}} \left( \prod_{i=1}^{p} (1-\lambda_i) \right)^{\frac{n-p-1}{2}} \prod_{i<j} (\lambda_i - \lambda_j)
\]

where

\[
C_2 = \frac{\pi^{p/2}}{\prod_{i=1}^{p} \Gamma \left( \frac{p-i+1}{2} \right) \Gamma \left( \frac{1}{2} \right)} \frac{\prod_{i=1}^{p} \Gamma \left( \frac{n-p+q-i+1}{2} \right)}{\Gamma \left( \frac{n-p+1}{2} \right)}
\]

Note that the roots \( \lambda_1, \ldots, \lambda_p \) are a maximal invariant under the transformations

\[
\Sigma \rightarrow A' \Sigma A
\]

where \( A \) has the form

\[
A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}
\]

\( A_1 \) is \( p \times p \) nonsingular and \( A_2 \) is \( q \times q \) non-singular.

We now discuss a large sample result concerning the roots of a determinantal equation. Let \( X \) be a \( p \times n \) matrix of independent random variables \( X_{ij} \) such that \( \mathcal{E}(X_{ij}) = 0 \), \( \mathcal{E}(X_{ij}^2) = 1 \), and \( \mathcal{E}(|X_{ij}|^k) \leq C_k < +\infty \) for each \( k \). Let \( L_1 \leq L_2 \leq \ldots \leq L_p \) denote the roots of \( S = XX' \) and let \( g(t) \) be the empirical cumulative distribution function.
based on \( L_1, \ldots, L_p \). That is,

\[
G(t) = \frac{1}{p} \sum_{i=1}^{p} K(t, L_i)
\]

where

\[
K(t, u) = \begin{cases} 
0 & \text{if } t < u \\
1 & \text{if } t \geq u 
\end{cases}
\]

We want to find a large sample approximation for \( \mathcal{E}(G(t)) \) as \( n \to \infty \), \( p \to \infty \), and \( \frac{n}{p} \to \beta > 1 \). Let \( \gamma(t) = \mathcal{E}(G(t)) \) and consider

\[
\frac{r_k}{p} = \int t^k d\gamma(t) = \int t^k d\mathcal{E}(G(t)) = \mathcal{E} \int t^k dG(t) = \frac{1}{p} \mathcal{E} \left( \sum_{i=1}^{p} l_i^k \right) = \frac{1}{p} \mathcal{E}(\text{tr } S^k).
\]

Then, \( r_0 = p, r_1 = pn, r_2 = n^2 p + p^2 n + np(C_2 - 2) = n^2 p + p^2 n + o(n^2) \).

It can be shown that for \( k \geq 2 \),

\[
r_k = (p+n)r_{k-1} + \sum_{\alpha=1}^{k-2} r_{\alpha} r_{k-1-\alpha} + o(n^k).
\]

Now consider

\[
\rho_k = \lim_{n \to \infty} \frac{r_k}{pn^k} \text{ as } \frac{n}{p} \to \beta > 1
\]

Then it follows easily that \( \rho_0 = \rho_1 = 1 \) and

\[
\rho_k = (1 + \frac{1}{\beta})\rho_{k-1} + \frac{1}{\beta} \sum_{\alpha=1}^{k-2} \rho_{\alpha} \rho_{k-1-\alpha} \text{ for } k \geq 2.
\]

Let \( F \) be a distribution function which has the moment sequence \( \{\rho_k\} \) (assuming such an \( F \) exists) so that

\[
\rho_k = \int_0^\infty x^k dF(x) = \mathcal{E}(Z^k).
\]
To find \( F \), we introduce the transform \( \psi(\lambda) \) given by

\[
\psi(\lambda) = \mathcal{E} \left[ \frac{1}{1-\lambda Z} - 1 \right] = \sum_{k=1}^{\infty} \lambda^k \rho_k.
\]

Thus,

\[
\psi(\lambda) = \lambda + \sum_{k=2}^{\infty} \lambda^k \left[ (1 + \frac{1}{\beta}) \rho_{k-1} + \frac{1}{\beta} \sum_{\alpha=2}^{k-2} \rho_{\alpha} \rho_{k-1-\alpha} \right]
\]

\[
= \lambda + (1 + \frac{1}{\beta}) \lambda \psi(\lambda) + \frac{\lambda}{\beta} \psi^2(\lambda).
\]

Solving for \( \psi(\lambda) \), we have

\[
\psi(\lambda) = \frac{\beta - (\beta+1) \pm \sqrt{(\beta+1 - \frac{\beta}{\lambda})^2 - 4 \beta}}{2}
\]

\[
= -\frac{\beta+1}{2} + \frac{1}{2} \frac{\beta}{\lambda} \left[ 1 \pm \sqrt{(1-\lambda a)(1-\lambda b)} \right]
\]

where \( a = (1 - \sqrt{\frac{1}{\beta}})^2 \) and \( b = (1 + \sqrt{\frac{1}{\beta}})^2 \). Since \( \lim_{\lambda \to 0^+} \psi(\lambda) = 0 \), we must choose the minus sign so that

\[
\psi(\lambda) = -\frac{\beta+1}{2} + \frac{1}{2} \frac{\beta}{\lambda} \left[ 1 - \sqrt{(1-\lambda a)(1-\lambda b)} \right].
\]

This transform can be inverted by methods developed in Shohat and Tamarkin (The Problem of Moments) to show that \( \psi \) corresponds to a distribution function which has a density given by

\[
p(x) = \begin{cases} 
0 & \text{if } x < a \\
\frac{\beta}{2x} \sqrt{(x-a)(b-x)} & \text{if } a \leq x \leq b \\
0 & \text{if } x > b
\end{cases}
\]

where \( a = (1 - \sqrt{\frac{1}{\beta}})^2 \), \( b = (1 + \sqrt{\frac{1}{\beta}})^2 \), and \( \beta > 1 \).