Chapter 8
Canonical Correlations And Canonical Variables

In many practical applications, the investigator may have two large sets of variables and wishes to study the interrelationships among them. If the two sets are very large, the experimenter may wish to consider only a few linear combinations of each set. Then the experimenter would like to study those linear combinations which are most highly correlated.

Consider two sets of variables with a joint distribution and analyze the correlations between the variables of one set and those of the other set. Find a new coordinate system in the space of each set of variables in such a way that the new coordinate system displays the system of correlations unambiguously.

We find the linear combinations of variables in each set that have the maximum correlation. These combinations are the first coordinates in the new system.

Then consider the second linear combination in each set, and choose the one that maximises the correlation between the two. This procedure continues until the two new coordinate systems are fully specified. When the observations are taken on a large number of correlated variables, how can we reduce the number of variables without sacrificing too much information?

When the variables are considered as belonging to a single set, principal component analysis is used to reduce dimensionality.

When the variables naturally fall into two sets, canonical correlation analysis is used to reduce dimensionality.

The canonical correlation analysis facilitates the study of interrelationships among the sets of multiple dependent variables and multiple independent variables. Note that the multiple linear regression analysis predicts a single dependent variable from a set of multiple independent variables.

Canonical correlation simultaneously predicts multiple dependent variables from multiple independent variables. This is the difference between multiple linear regression and canonical correlation analysis.

The appropriate data for canonical correlation analysis is the two sets of variables. Assume that each set can be given some theoretical meaning, at least to the extent that one set could be defined as the independent variables and the other set as the dependent variables. Once this distinction has been made, the canonical correlation can address a wide range of objectives, e.g.,

1. Determining whether the two sets of variables are independent of one another or determining the magnitude of the relationships that may exist between the two sets.

2. Deriving a set of weights for each set of dependent and independent variables so that the linear combinations of each set are maximally correlated. Additional linear functions that maximize the remaining correlation are independent of preceding set(s) of linear combinations.

3. Explaining the nature of whatever relationships exist between the sets of dependent and independent variables, generally by measuring the relative contribution of each variable to the canonical function (relationships) that are extracted.

Definition

Let $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{pmatrix}$ where $p \times 1$ vector $\underset{\sim}{X}$ is subdivided into $p_1 \times 1$ vector $\underset{\sim}{X}_1$ and $p_2 \times 1$ vector $\underset{\sim}{X}_2$ such that $p_1 + p_2 = p$. The r^{th} pair of canonical variates is the pair of linear combinations

$$U_r = \underset{\sim}{\alpha}^{(r)'} \underset{\sim}{X}^{(1)}$$

$$\text{and} \quad V_r = \underset{\sim}{\gamma}^{(r)'} \underset{\sim}{X}^{(2)}$$

with $Var(U_r) = 1$, $Var(V_r) = 1$ and (U_r, V_r) are uncorrelated with the first $(r-1)$ pairs of canonical variates and having maximum correlation. The correlation is the r^{th}

canonical correlation.

Canonical Correlation and Variates in the Population

Suppose $\underset{\sim}{X} = (X_1, X_2, \ldots, X_p)'$ is a $p \times 1$ random vector with $E(\underset{\sim}{X}) = 0$ and positive definite covariance matrix Σ. Partition $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix}$ such that $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are p_1 and p_2 components vectors, respectively. For convenience, assume $p_1 \leq p_2$. Partition Σ similarly in p_1 rows and and p_2 columns as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Finding First Pair of Canonical Variates

Consider two arbitrary linear combinations as follows:

$$U = \underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)} \quad \text{of the components of } \underset{\sim}{X}^{(1)},$$
$$V = \underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)} \quad \text{of the components of } \underset{\sim}{X}^{(2)}.$$

Now, we find the linear functions with the highest correlation. Since the correlation of a multiple of U and a multiple of V is the same as the correlation of U and V (i.e., $Cor(kU, kV) = Cor(U, V)$, k is a scalar), so we can make an arbitrary normalization of $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$.

Therefore we need to find $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ to be such that $Var(U) = 1$ and $Var(V) = 1$, i.e.,

$$1 = E(U^2) = E(\underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}\underset{\sim}{X}^{(1)'}\underset{\sim}{\alpha}) = \underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} \tag{1}$$
$$1 = E(V^2) = E(\underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(2)'}\underset{\sim}{\gamma}) = \underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma}. \tag{2}$$

Here

$$E(U) = E(\underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}) = \underset{\sim}{\alpha}'E(\underset{\sim}{X}^{(1)}) = 0$$
$$E(V) = E(\underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}) = \underset{\sim}{\gamma}'E(\underset{\sim}{X}^{(2)}) = 0$$
$$E(UV) = E(\underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}\underset{\sim}{X}^{(2)'}\underset{\sim}{\gamma}) = \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} \tag{3}$$

The correlation coefficient is

$$Cor(U,V) = \frac{E(UV) - E(U)E(V)}{\sqrt{Var(U)}\sqrt{Var(V)}} = \frac{E(UV) - 0}{1} = E(UV).$$

Algebraically, find $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ to maximize (3), the correlation coefficient $E(UV)$ subject to (1) and (2), i.e., $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 1$ and $\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} = 1$.

Let

$$\Psi = \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} - \frac{\lambda}{2}(\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} - 1) - \frac{\mu}{2}(\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} - 1)$$

where λ and μ are the Lagrangian multipliers. Differentiating Ψ with respect to $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$, and equating to zero, we get

$$\frac{\partial \Psi}{\partial \underset{\sim}{\alpha}} = \Sigma_{12}\underset{\sim}{\gamma} - \lambda\Sigma_{11}\underset{\sim}{\alpha} = 0 \tag{4}$$

$$\frac{\partial \Psi}{\partial \underset{\sim}{\gamma}} = \Sigma_{12}'\underset{\sim}{\alpha} - \mu\Sigma_{22}\underset{\sim}{\gamma} = 0 \tag{5}$$

Premultiply (4) by $\underset{\sim}{\alpha}$ and postmultiply (5) by $\underset{\sim}{\gamma}$, we get

$$\underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} - \lambda\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 0 \Rightarrow \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} = \lambda$$

$$\underset{\sim}{\gamma}'\Sigma_{12}'\underset{\sim}{\alpha} - \mu\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} = 0 \Rightarrow \underset{\sim}{\gamma}'\Sigma_{12}'\underset{\sim}{\alpha} = \mu$$

using $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 1$ and $\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} = 1$. Thus

$$\lambda = \mu = \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma}.$$

We can write (4) and (5) as

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\gamma} \end{pmatrix} = 0 \tag{6}$$

To have a nontrivial solution (which is necessary for a solution satisfying (1) and (2)), we need to have

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} = 0, \tag{7}$$

i.e., $\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ is a nonsingular matrix.

Note that (7) is a polynomial of degree p and has p roots, say, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. It remains to show that (7) is a polynomial of degree p. First, we see from the equation

$$\underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} - \lambda\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 0$$

that $\lambda = \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} = Cor(U,V)$ where $U = \underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}$, $V = \underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}$ and $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ satisfy (6).

Since we want the maximum correlation, we take

$$\lambda = \lambda_1.$$

Let a solution of (6) for $\lambda = \lambda_1$ be $\underset{\sim}{\alpha}^{(1)}, \underset{\sim}{\gamma}^{(1)}$ and let $U_1 = \underset{\sim}{\alpha}^{(1)'}\underset{\sim}{X}^{(1)}$, $V_1 = \underset{\sim}{\gamma}^{(1)'}\underset{\sim}{X}^{(2)}$. Then U_1 and V_1 are the normalized linear combinations of $\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{X}^{(2)}$, respectively with maximum correlation.

Before proceeding, we first show that (7) is a polynomial of degree p. Consider a Laplace expansion by minors of the first p_1 columns. One term is

$$|-\lambda\Sigma_{11}| \cdot |-\lambda\Sigma_{22}| = (-\lambda)^{p_1+p_2}|\Sigma_{11}| \cdot |\Sigma_{22}|.$$

The other terms in the expansion are of lower degree in λ because one or more rows of each minor in the first p_1 columns do not contain λ. Since Σ is positive definite, so $|\Sigma_{11}| \cdot |\Sigma_{22}| \neq 0$. This shows that (7) is a polynomial equation of degree p and has p roots, say $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.

Before deriving the second pair of canonical variates, we can derive a single matrix equation for $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ for the equations related to the first pair of canonical variates.

Consider the following (4) and (5) equations:

$$-\lambda\Sigma_{11}\underset{\sim}{\alpha} + \Sigma_{12}\underset{\sim}{\gamma} = 0$$
$$\text{and} \quad \Sigma_{21}\underset{\sim}{\alpha} - \lambda\Sigma_{22}\underset{\sim}{\gamma} = 0.$$

Premultiplying (4) by λ and (5) by Σ_{22}^{-1}, we get,

$$\lambda\Sigma_{12}\underset{\sim}{\gamma} = \lambda^2\Sigma_{11}\underset{\sim}{\alpha} \tag{8}$$

$$\Sigma_{22}^{-1}\Sigma_{21}\underset{\sim}{\alpha} = \lambda\underset{\sim}{\gamma} \tag{9}$$

Substitute (**9**) into (8) and we get

$$\begin{aligned} \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\underset{\sim}{\alpha} &= \lambda^2\Sigma_{11}\underset{\sim}{\alpha} \\ \text{or} \quad (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11})\underset{\sim}{\alpha} &= 0. \end{aligned} \tag{10}$$

The quantities $\lambda_1^2, \lambda_2^2, \ldots, \lambda_p^2$ satisfy

$$|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11}| = 0$$

and $\underset{\sim}{\alpha}^{(1)}, \underset{\sim}{\alpha}^{(2)}, \ldots, \underset{\sim}{\alpha}^{(p_1)}$ satisfy (8) for $\lambda^2 = \lambda_1^2, \lambda_2^2, \ldots, \lambda_{p_1}^2$, respectively.

Similar equation occur for $\underset{\sim}{\gamma}^{(1)}, \underset{\sim}{\gamma}^{(2)}, \ldots, \underset{\sim}{\gamma}^{(p_2)}$ when $\mu^2 = \mu_1^2, \mu_2^2, \ldots, \mu_p^2$ are substituted with

$$(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \mu^2\Sigma_{22})\underset{\sim}{\gamma} = 0$$

which is obtained by premultiplying (4) by Σ_{11}^{-1} and (5) by μ, i.e.,

$$\begin{aligned} \mu\underset{\sim}{\alpha} &= \Sigma_{11}^{-1}\Sigma_{12}\underset{\sim}{\gamma} \\ \Sigma_{21}\underset{\sim}{\alpha} &= \mu^2\Sigma_{22}\underset{\sim}{\gamma} \end{aligned}$$

which gives $\quad (\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \mu^2\Sigma_{22})\underset{\sim}{\gamma} = 0.$

Finding Second Pair of Canonical Variates

Consider finding a second linear combination of $\underset{\sim}{X}^{(1)}$, say

$$U_2 = \underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}$$

and a second linear combination of $\underset{\sim}{X}^{(2)}$, say

$$V_2 = \underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}$$

such that all linear combinations are uncorrelated with the first pair of canonical variates U_1 and V_1, and have maximum correlation. This will yield, say, $U_2 = \underset{\sim}{\alpha}^{(2)'}\underset{\sim}{X}^{(1)}$ and $V_2 = \underset{\sim}{\gamma}^{(2)'}\underset{\sim}{X}^{(2)}$ with the correlation coefficient $\lambda_2 = \lambda^{(2)}$ which is the second maximum characteristic root of (7).

The detailed steps of derivation of such a result are explained in the following when the $(r+1)^{th}$ pair of canonical variates is obtained.

Finding $(r+1)^{th}$ Pair of Canonical Variates

Continuing with the procedure, at r^{th} step, the following linear combinations are obtained:

$$
\begin{aligned}
&(U_1 = \underset{\sim}{\alpha}^{(1)'}\underset{\sim}{X}^{(1)}, V_1 = \underset{\sim}{\gamma}^{(1)'}\underset{\sim}{X}^{(2)}),\\
&(U_2 = \underset{\sim}{\alpha}^{(2)'}\underset{\sim}{X}^{(1)}, V_2 = \underset{\sim}{\gamma}^{(2)'}\underset{\sim}{X}^{(2)}), \ldots,\\
&(U_r = \underset{\sim}{\alpha}^{(r)'}\underset{\sim}{X}^{(1)}, V_r = \underset{\sim}{\gamma}^{(r)'}\underset{\sim}{X}^{(2)})
\end{aligned}
$$

with the corresponding correlation coefficients $\lambda^{(1)} = \lambda_1, \lambda^{(2)} = \lambda_2, \ldots, \lambda^{(r)} = \lambda_r$ where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the characteristic roots of (7).

We ask for a linear combination of variables in $\underset{\sim}{X}^{(1)}$, say $U = \underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}$ and a linear combination of variables in $\underset{\sim}{X}^{(2)}$, say $V = \underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}$, that among all linear combinations uncorrelated with $(U_1, V_1), (U_2, V_2), \ldots, (U_r, V_r)$ have maximum correlation.

Next, we determine the conditions under which U and V are uncorrelated with U_i and V_i.

The condition that U be uncorrelated with $U_i = \underset{\sim}{\alpha}^{(i)'}\underset{\sim}{X}^{(1)}$ is

$$0 = E(UU_i) = E(\underset{\sim}{\alpha}'\underset{\sim}{X}^{(1)}\underset{\sim}{X}^{(1)'}\underset{\sim}{\alpha}^{(i)}) = \underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} \tag{11}$$

Using (4), i.e., $\Sigma_{12}\underset{\sim}{\gamma} = \lambda\Sigma_{11}\underset{\sim}{\alpha}$, we have

$$E(UV_i) = \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma}^{(i)} = \lambda^{(i)}\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0.$$

The condition that V is uncorrelated with $V_i = \underset{\sim}{\gamma}^{(i)'}\underset{\sim}{X}^{(2)}$

$$0 = E(VV_i) = E(\underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(2)'}\underset{\sim}{\gamma}^{(i)}) = \underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma}^{(i)} \tag{12}$$

The condition that V is uncorrelated with $U_i = \underset{\sim}{\alpha}^{(i)'}\underset{\sim}{X}^{(1)}$ is

$$E(VU_i) = E(\underset{\sim}{\gamma}'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(1)'}\underset{\sim}{\alpha}^{(i)}) = \underset{\sim}{\gamma}'\Sigma_{21}\underset{\sim}{\alpha}^{(i)} = 0$$

Now we find the $(r+1)^{th}$ pair of canonical variates as follows:

Maximize $E(U_{r+1}V_{r+1})$ by choosing $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ to satisfy the following conditions, as obtained by (1), (2), (11) and (12) for $i = 1, 2, \ldots, r$,

$$\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 1, \quad \underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} = 1, \quad \underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0 \quad \text{and} \quad \underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma}^{(i)} = 0,$$

respectively.

Consider the Lagrangian function

$$\Psi_{r+1} = \underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma} - \frac{\lambda}{2}(\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} - 1) - \frac{\mu}{2}(\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} - 1) + \sum_{i=1}^{r}\delta_i\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} + \sum_{i=1}^{r}\theta_i\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma}^{(i)}$$

where $\lambda, \mu, \delta_1, \delta_2, \ldots, \delta_r, \theta_1, \theta_2, \ldots, \theta_r$ are Lagrangian multipliers. Differentiating Ψ_{r+1} partially with respect to $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ and equating them to zero, we get

$$\frac{\partial\Psi_{r+1}}{\partial\underset{\sim}{\alpha}} = \Sigma_{12}\underset{\sim}{\gamma} - \lambda\Sigma_{11}\underset{\sim}{\alpha} + \sum_{i=1}^{r}\delta_i\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0 \tag{13}$$

$$\frac{\partial\Psi_{r+1}}{\partial\underset{\sim}{\gamma}} = \Sigma_{21}\underset{\sim}{\alpha} - \mu\Sigma_{22}\underset{\sim}{\gamma} + \sum_{i=1}^{r}\theta_i\Sigma_{22}\underset{\sim}{\gamma}^{(i)} = 0. \tag{14}$$

Premultiplying (13) by $\underset{\sim}{\alpha}^{(j)'}$, we get

$$\underset{\sim}{\alpha}^{(j)'}\Sigma_{12}\underset{\sim}{\gamma} - \lambda\underset{\sim}{\alpha}^{(j)'}\Sigma_{11}\underset{\sim}{\alpha} + \sum_{i=1}^{r}\delta_i\underset{\sim}{\alpha}^{(j)'}\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0$$

in which the first term is zero due to $\underset{\sim}{\gamma}'\Sigma_{21}\underset{\sim}{\alpha}^{(j)} = 0$ and the second term is zero due to $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(j)} = 0$. In the third term, all the terms $\underset{\sim}{\alpha}^{(j)'}\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0$ where $i \neq j = 1, 2, \ldots, r$ and the j^{th} term $\underset{\sim}{\alpha}^{(j)'}\Sigma_{11}\underset{\sim}{\alpha}^{(j)} = 1$. So we finally get $\delta_j = 0$.

Similarly, premultiplying (14) by $\underset{\sim}{\gamma}^{(j)'}$, we get

$$\underset{\sim}{\gamma}^{(j)'}\Sigma_{21}\underset{\sim}{\alpha} - \mu\underset{\sim}{\gamma}^{(j)'}\Sigma_{22}\underset{\sim}{\gamma} + \sum_{i=1}^{r}\theta_i\underset{\sim}{\gamma}^{(j)'}\Sigma_{22}\underset{\sim}{\gamma}^{(i)} = 0$$

in which the first and second terms are zero due to $\underset{\sim}{\gamma}^{(j)'}\Sigma_{21}\underset{\sim}{\alpha} = 0$ and $\underset{\sim}{\gamma}^{(j)'}\Sigma_{22}\underset{\sim}{\gamma} = 0$, respectively. In the third term $\underset{\sim}{\gamma}^{(j)'}\Sigma_{22}\underset{\sim}{\gamma}^{(i)} = 0$ for $i \neq j$ and $\underset{\sim}{\gamma}^{(j)'}\Sigma_{22}\underset{\sim}{\gamma}^{(j)} = 1$. So we finally get $\theta_j = 0$.

So we observe that the obtained equations (13) and (14) are simply the equations (4) and (5) or alternatively (6).

We therefore take the largest λ_i, say $\lambda^{(r+1)}$, such that there is a solution to

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\gamma} \end{pmatrix} = 0$$

satisfying $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 1$, $\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} = 1$, $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0$ and $\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma}^{(i)} = 0$ for $i = 1, 2, \ldots, r$. Let this solution be $\underset{\sim}{\alpha}^{(r+1)}$ and $\underset{\sim}{\gamma}^{(r+1)}$ and let $U_{r+1} = \underset{\sim}{\alpha}^{(r+1)'}\underset{\sim}{X}^{(1)}$ and $V_{r+1} = \underset{\sim}{\gamma}^{(r+1)'}\underset{\sim}{X}^{(2)}$.

We continue with this procedure step by step as long as the successive solution can be found which satisfies the condition

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\gamma} \end{pmatrix} = 0$$

with $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 1$, $\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma} = 1$, $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0$, $\underset{\sim}{\gamma}'\Sigma_{22}\underset{\sim}{\gamma}^{(i)} = 0$ for $i = 1, 2, \ldots, r$.

The next question that arises is how long we can continue? The answer is that we can continue for p_1 steps ($p_1 \leq p_2$), and now we show this.

Number of Canonical Variables

Let m be the number of steps for which we can continue with the procedure. Let

$$\begin{aligned} A &= (\underset{\sim}{\alpha}^{(1)}, \underset{\sim}{\alpha}^{(2)}, \ldots, \underset{\sim}{\alpha}^{(m)}), \\ \Gamma_1 &= (\underset{\sim}{\gamma}^{(1)}, \underset{\sim}{\gamma}^{(2)}, \ldots, \underset{\sim}{\gamma}^{(m)}), \\ \text{and} \quad \Lambda &= diag(\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(m)}). \end{aligned}$$

The conditions $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha} = 1$ and $\underset{\sim}{\alpha}'\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = 0$ can be summarized as $A'\Sigma_{11}A = I$.

Since $rank(\Sigma_{11}) = p_1$ and $rank(I_m) = m$, we have $m \leq p_1$.

Now we show that $m < p_1$ leads to a contradiction by showing that in this case, there is another vector satisfying the condition.

Since $A'\Sigma_{11}$ is of order $m \times p_1$, so there exists a $p_1 \times (p_1 - m)$ matrix E with $rank(E) = p_1 - m$ such that $A'\Sigma_{11}$E$= 0$. Similarly there is a $p_2 \times (p_2 - m)$ matrix F with $rank(F) = p_2 - m$ such that $\Gamma_1\Sigma_{22}F = 0$. Also, we have $\Gamma_1\Sigma_{21}E = \Lambda A'\Sigma_{11}E = 0$ and $A'\Sigma_{12}F = \Lambda F'\Sigma_{22}F = 0$.

Since $rank(E) = p_1 - m$, so $E'\Sigma_{11}E$ is nonsingular if $m < p_1$ and similarly $F'\Sigma_{22}F$ is nonsingular.

Thus, there is at least one root of

$$\begin{vmatrix} -\delta E'\Sigma_{11}E & E'\Sigma_{12}F \\ F'\Sigma_{21}E & -\delta F'\Sigma_{22}F \end{vmatrix} = 0$$

because $|E'\Sigma_{11}E| \cdot |F'\Sigma_{22}F| \neq 0$.

From the preceding algebra, we see that there exist vectors $\underset{\sim}{a}$ and $\underset{\sim}{b}$ such that

$$E'\Sigma_{12}F\underset{\sim}{b} = \delta E'\Sigma_{11}E\underset{\sim}{a}$$
$$F'\Sigma_{21}E\underset{\sim}{a} = \delta F'\Sigma_{22}F\underset{\sim}{b}.$$

Let $E\underset{\sim}{a} = \underset{\sim}{g}$ and $F\underset{\sim}{b} = \underset{\sim}{h}$.

Now we show that $\delta, \underset{\sim}{g}$ and $\underset{\sim}{h}$ form a new solution $\lambda^{(m+1)}, \underset{\sim}{\alpha}^{(m+1)}$ and $\underset{\sim}{\gamma}^{(m+1)}$. Let $\Sigma_{11}^{-1}\Sigma_{12}\underset{\sim}{h} = \underset{\sim}{k}$. Since $A'\Sigma_{11}\underset{\sim}{k} = A'\Sigma_{12}F\underset{\sim}{b} = 0$, so $\underset{\sim}{k}$ is orthogonal to the rows of $A'\Sigma_{11}$ and therefore is a linear combination of columns of E, say $E\underset{\sim}{c}$. Thus $\Sigma_{12}\underset{\sim}{h} = \Sigma_{11}\underset{\sim}{k}$ can be written as

$$\Sigma_{12}F\underset{\sim}{b} = \Sigma_{11}E\underset{\sim}{c}.$$

Multiplying by E' on the left, we get

$$E'\Sigma_{12}F\underset{\sim}{b} = E'\Sigma_{11}E\underset{\sim}{c}.$$

Since $E'\Sigma_{11}E$ is nonsingular, comparing $E'\Sigma_{12}F\underset{\sim}{b} = \delta E'\Sigma_{11}E\underset{\sim}{a}$ and $E\Sigma_{12}F\underset{\sim}{b} = E'\Sigma_{11}E\underset{\sim}{c}$ shows that $\underset{\sim}{c} = \delta\underset{\sim}{a}$ and therefore $\underset{\sim}{k} = \delta\underset{\sim}{g}$. Thus $\Sigma_{12}\underset{\sim}{h} = \delta\Sigma_{11}\underset{\sim}{g}$.

In a similar fashion, we show that $\Sigma_{21}\underset{\sim}{g} = \delta\Sigma_{22}\underset{\sim}{h}$.

Therefore $\delta = \lambda^{(m+1)}, \underset{\sim}{g} = \underset{\sim}{\alpha}^{(m+1)}$ and $\underset{\sim}{h} = \underset{\sim}{\gamma}^{(m+1)}$ is another solution. But this is contrary to the assumption that $\lambda^{(m)}, \underset{\sim}{\alpha}^{(m)}$ and $\underset{\sim}{\gamma}^{(m)}$ was the last possible solution.

Thus $m = p_1$.

Compilation of Conditions and Handling Remaining Variables

The conditions on λ's, $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\gamma}$'s can be summarized as

$$A'\Sigma_{11}A = I$$
$$A'\Sigma_{12}\Gamma_1 = \Lambda = diag(\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(p_1)})$$
$$\Gamma_1'\Sigma_{22}\Gamma_1 = I$$

where $A = (\underset{\sim}{\alpha}^{(1)}, \underset{\sim}{\alpha}^{(2)}, \ldots, \underset{\sim}{\alpha}^{(p_1)})$ and $\Gamma_1 = (\underset{\sim}{\gamma}^{(1)}, \underset{\sim}{\gamma}^{(2)}, \ldots, \underset{\sim}{\gamma}^{(p_1)})$. We are now left with $(p_2 - p_1)$ variables in $\underset{\sim}{X}^{(2)}$. So for them, the Γ matrix is constructed as follows:

Let

$$\Gamma_2 = (\underset{\sim}{\gamma}^{(p_1+1)}, \underset{\sim}{\gamma}^{(p_1+2)}, \ldots, \underset{\sim}{\gamma}^{(p_2)})$$

be a $p_2 \times (p_2 - p_1)$ matrix satisfying

$$\Gamma_2'\Sigma_{22}\Gamma_1 = 0$$
$$\Gamma_2'\Sigma_{22}\Gamma_2 = I.$$

Any Γ_2 can be multiplied on the right hand by an arbitrary $(p_2-p_1)\times(p_2-p_1)$ orthogonal matrix. Such a matrix can be formed one column at a time as follows:
$\underset{\sim}{\gamma}^{(p_1+1)}$ is a vector orthogonal to $\Sigma_{22}\Gamma_1$ and normalized so $\underset{\sim}{\gamma}^{(p_1+1)'}\Sigma_{22}\underset{\sim}{\gamma}^{(p_1+1)} = 1$; $\underset{\sim}{\gamma}^{(p_1+2)}$ is a vector orthogonal to $\Sigma_{22}(\Gamma_1 \quad \underset{\sim}{\gamma}^{(p_1+1)})$ and normalized so $\underset{\sim}{\gamma}^{(p_1+2)'}\Sigma_{22}\underset{\sim}{\gamma}^{(p_1+2)} = 1$ and so forth.

Let

$$\Gamma = (\Gamma_1, \quad \Gamma_2)$$

be a square matrix and is nonsingular since $\Gamma'\Sigma_{22}\Gamma = I$.

Consider the determinant

$$\begin{aligned}
&\begin{vmatrix} A' & 0 \\ 0 & \Gamma_1' \\ 0 & \Gamma_2' \end{vmatrix} \begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} \begin{vmatrix} A & 0 & 0 \\ 0 & \Gamma_1 & \Gamma_2 \end{vmatrix} \\
&= \begin{vmatrix} -\lambda I & \Lambda & 0 \\ \Lambda & -\lambda I & 0 \\ 0 & 0 & -\lambda I \end{vmatrix} \\
&= (-\lambda)^{p_2-p_1} \begin{vmatrix} -\lambda I & \Lambda \\ \Lambda & -\lambda I \end{vmatrix} \\
&= (-\lambda)^{p_2-p_1} |-\lambda I| |-\lambda I - \Lambda(-\lambda I)^{-1}\Lambda| \\
&= (-\lambda)^{p_2-p_1} |\lambda^2 I - \Lambda^2| \\
&= (-\lambda)^{p_2-p_1} \prod |\lambda^2 - \lambda^{(i)2}|. \qquad (15)
\end{aligned}$$

Except for a constant factor, this polynomial (12) is $\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix}$. Thus the roots of $\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} = 0$ and (15) are the same.

These roots are

$$\begin{aligned}
&\lambda = \pm\lambda^{(i)}, i = 1, 2, \ldots, p_1 \\
&\lambda = 0 \quad \text{(of multiplicity } p_2 - p_1).
\end{aligned}$$

Thus

$$(\lambda_1, \lambda_2, \ldots, \lambda_{p1}) = (\lambda_1, \lambda_2, \ldots, \lambda_p, 0, 0, \ldots, 0, -\lambda_{p_1}, -\lambda_{p_2}, \ldots, -\lambda_1).$$

The set $\{\lambda^{(i)2}\}, i = 1, 2, \ldots, p_1$ is the set $\{\lambda_i\}, i = 1, 2, \ldots, p_1$, we only need to show that $\lambda^{(i)}$ is nonnegative (and therefore is one of the $\lambda_i, i = 1, 2, \ldots, p_1$).

It is clarified as follows:

Recall that $\lambda_1, \lambda_2, \ldots, \lambda_p$ are obtained through equations (4) and (5) given by

$$-\lambda\Sigma_{11}\underset{\sim}{\alpha} + \Sigma_{12}\underset{\sim}{\gamma} = 0$$

and

$$\Sigma_{21}\underset{\sim}{\alpha} - \lambda\Sigma_{22}\underset{\sim}{\gamma} = 0.$$

Now, let $\lambda = \pm\lambda^{(i)}$ and corresponding to $\lambda = \lambda^{(i)}$, we can obtain $\underset{\sim}{\alpha}^{(i)}$ and $\underset{\sim}{\gamma}^{(i)}$. It should satisfy (4) and (5), i.e., if $\lambda = \lambda^{(i)}$, them

$$\lambda^{(i)}\Sigma_{11}\underset{\sim}{\alpha}^{(i)} = \Sigma_{12}\underset{\sim}{\gamma}^{(i)}$$

and

$$\Sigma_{21}\underset{\sim}{\alpha}^{(i)} = \lambda^{(i)}\Sigma_{22}\underset{\sim}{\gamma}^{(i)}.$$

Then (4) and (5) can be rewritten as

$$\Sigma_{12}\underset{\sim}{\gamma}^{(i)} = -\lambda^{(i)}\Sigma_{11}(-\underset{\sim}{\alpha}^{(i)})$$

and

$$\Sigma_{21}(-\underset{\sim}{\alpha}^{(i)}) = -\lambda^{(i)}\Sigma_{22}\underset{\sim}{\gamma}^{(i)}.$$

Thus if $(\lambda^{(i)}, \underset{\sim}{\alpha}^{(i)}, \underset{\sim}{\gamma}^{(i)})$ is a solution, so $(-\lambda^{(i)}, -\underset{\sim}{\alpha}, \underset{\sim}{\gamma}^{(i)})$ is also a solution. If $\lambda^{(i)} < 0$ then $-\lambda^{(i)} \geq 0$ and $-\lambda^{(i)} \geq \lambda^{(i)}$. Since $\lambda^{(i)}$ is the maximum, we must have $\lambda^{(i)} \geq -\lambda^{(i)}$ and therefore $\lambda^{(i)} \geq 0$. Further, the correlation coefficient is $\underset{\sim}{\alpha}'\Sigma_{12}\underset{\sim}{\gamma}$. So if $(\lambda^{(i)}, \underset{\sim}{\alpha}^{(i)}, \underset{\sim}{\gamma}^{(i)})$ is a solution, then $\lambda^{(i)} = \underset{\sim}{\alpha}^{(i)\prime}\Sigma_{12}\underset{\sim}{\gamma}^{(i)}$. Since $\lambda^{(i)} \geq 0$, so $\underset{\sim}{\alpha}^{(i)\prime}\Sigma_{12}\underset{\sim}{\gamma}^{(i)} \geq 0$. If $(-\lambda^{(i)}, -\underset{\sim}{\alpha}^{(i)}, \underset{\sim}{\gamma}^{(i)})$ is a solution, then $-\lambda^{(i)} = (-\underset{\sim}{\alpha}^{(i)})'\Sigma_{12}\underset{\sim}{\gamma}^{(i)}$. So sign is irrelevant in the sense that it can change $\underset{\sim}{\alpha}^{(i)}$ into $-\underset{\sim}{\alpha}^{(i)}$.

We observe that

$$\Sigma_{12}\underset{\sim}{\gamma}^{(r)} = -\lambda^{(r)}\Sigma_{11}(-\underset{\sim}{\alpha}^{(r)})$$
$$\Sigma_{21}(-\underset{\sim}{\alpha}^{(r)}) = -\lambda^{(r)}\Sigma_{22}\underset{\sim}{\gamma}^{(r)},$$

thus if $\lambda^{(r)}, \underset{\sim}{\alpha}^{(r)}$ and $\underset{\sim}{\gamma}^{(r)}$ is a solution, then $-\lambda^{(r)}, -\underset{\sim}{\alpha}^{(r)}$ and $\underset{\sim}{\gamma}^{(r)}$ is also a solution. If $\lambda^{(r)}$ were negative, then $-\lambda^{(r)} > 0$ and $-\lambda^{(r)} \geq \lambda^{(r)}$. But since $\lambda^{(r)}$ was to be the maximum, we must have $\lambda^{(r)} \geq -\lambda^{(r)}$ and therefore $\lambda^{(r)} \geq 0$. Since set $\{\lambda^{(i)}\}$ is the same as $\{\lambda_i\}$, $i = 1, 2, \ldots, p_1$, we must have $\lambda^{(i)} = \lambda_i$.

Let

$$
\begin{aligned}
U &= (U_1, U_2, \ldots, U_{p_1})' = A'\underset{\sim}{X}^{(1)}, \\
V^{(1)} &= (V_1, V_2, \ldots, V_{p_1})' = \Gamma_1'\underset{\sim}{X}^{(2)}, \\
V^{(2)} &= (V_{p_1+1}, V_{p_2+2}, \ldots, V_{p_2})' = \Gamma_2'\underset{\sim}{X}^{(2)}.
\end{aligned}
$$

The components of U are the set of canonical variates and components of $(V^{(1)'}V^{(2)'})'$ are the other set of canonical variates. Then we have

$$
\begin{aligned}
E\begin{pmatrix} U \\ V^{(1)} \\ V^{(2)} \end{pmatrix}(U'V^{(1)'}V^{(2)'}) &= E\begin{pmatrix} A' & 0 \\ 0 & \Gamma_1' \\ 0 & \Gamma_2' \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} A & 0 & 0 \\ 0 & \Gamma_1 & \Gamma_2 \end{pmatrix} \\
&= E\begin{pmatrix} A'\underset{\sim}{X}^{(1)}\underset{\sim}{X}^{(1)'}A & A'\underset{\sim}{X}^{(1)}\underset{\sim}{X}^{(2)'}\Gamma_1 & A'\underset{\sim}{X}^{(1)}\underset{\sim}{X}^{(2)'}\Gamma_2 \\ \Gamma_1'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(1)'}A & \Gamma_1'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(2)'}\Gamma_1 & \Gamma_1'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(2)'}\Gamma_2 \\ \Gamma_2'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(1)'}A & \Gamma_2'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(2)'}\Gamma_1 & \Gamma_2'\underset{\sim}{X}^{(2)}\underset{\sim}{X}^{(2)'}\Gamma_2 \end{pmatrix} \\
&= \begin{pmatrix} A'\Sigma_{11}A & A'\Sigma_{12}\Gamma_1 & A'\Sigma_{12}\Gamma_2 \\ \Gamma_1'\Sigma_{21}A & \Gamma_1'\Sigma_{22}\Gamma_1 & \Gamma_1'\Sigma_{22}\Gamma_2 \\ \Gamma_2'\Sigma_{21}A & \Gamma_2'\Sigma_{22}\Gamma_1 & \Gamma_2'\Sigma_{22}\Gamma_2 \end{pmatrix} \\
&= \begin{pmatrix} I_{p_1} & \Lambda & 0 \\ \Lambda & I_{p_1} & 0 \\ 0 & 0 & I_{p_2-p_1} \end{pmatrix}
\end{aligned}
$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_{p_1})$.

Example

Suppose $\underset{\sim}{X}$ is a 12-component vector, i.e., $p = 12$. It is partitioned into 5 and 7 component vectors $\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{X}^{(2)}$ respectively. Thus $p_1 = 5, p_2 = 7$ and the number of canonical correlations is $\min(p_1, p_2) = \min(5, 7) = 5$.

Standardize the variables and find the characteristic roots. Let their values are $\lambda_1 = 0.55, \lambda_2 = 0.23, \lambda_3 = 0.12, \lambda_4 = 0.08$ and $\lambda_5 = 0.05$ with $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \lambda_5$.

Corresponding to each λ_i, find the respective characteristic vectors $\hat{\underset{\sim}{\alpha}}$ and $\hat{\underset{\sim}{\gamma}}$. Then $U_i = \hat{\underset{\sim}{\alpha}}^{(i)'} \underset{\sim}{X}^{(1)}$ and $V_i = \hat{\underset{\sim}{\gamma}}^{(i)'} \underset{\sim}{X}^{(2)}$ $(i = 1, 2, 3, 4, 5)$ give pairs of canonical variables (U_i, V_i) and canonical correlation λ_i. Note that if $\lambda_i < 0$ then it is difficult to decide. That is why we consider the positive square root of λ_i^2. Choose the number of canonical variables by adding up correlations to have the maximum contribution.

Estimation of Canonical Correlations and Canonical Variables

Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N$ be a sample from $N_p(\underset{\sim}{\mu}, \Sigma)$. Partition $\underset{\sim}{X} = (\underset{\sim}{X}_\alpha^{(1)'} \quad \underset{\sim}{X}_\alpha^{(2)})'$ where $\underset{\sim}{X}_\alpha^{(1)}$ and $\underset{\sim}{X}_\alpha^{(2)}$ have p_1 and p_2 components, respectively. The maximum likelihood estimator of Σ is

$$
\begin{aligned}
\hat{\Sigma} &= \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} \\
&= \frac{1}{N} \begin{pmatrix} \sum_{\alpha=1}^N (\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' & \sum_{\alpha=1}^N (\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})' \\ \sum_{\alpha=1}^N (\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' & \sum_{\alpha=1}^N (\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})(\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})' \end{pmatrix}
\end{aligned}
$$

The maximum likelihood estimates of the matrix of canonical correlations Λ and the canonical variates are $\hat{A}$ and $\hat{\Gamma}$, respectively. The matrices $\Lambda, \hat{A}$ and $\hat{\Gamma}$ are uniquely defined if we assume that the canonical correlations are different and that the first nonzero element of each column of $\hat{A}$ is positive.

The indeterminacy in $\hat{\Gamma}_2$ allows multiplication on right by a $(p_2 - p_1) \times (p_2 - p_1)$ orthogonal matrix. This indeterminacy can be removed by various types of requirements, e.g., a submatrix formed by the lower $(p_2 - p_1)$ rows is upper or lower triangular with positive diagonal elements.

The maximum likelihood estimates of $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the characteristic roots of

$$
\begin{vmatrix} -\hat{\lambda}\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -\hat{\lambda}\hat{\Sigma}_{22} \end{vmatrix} = 0
$$

and j^{th} column of $\hat{A}$ and $\hat{\Gamma}_1$ satisfy

$$
\begin{pmatrix} -\hat{\lambda}_j\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -\hat{\lambda}_j\hat{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \hat{\underset{\sim}{\alpha}}^{(j)} \\ \hat{\underset{\sim}{\gamma}}^{(j)} \end{pmatrix} = 0
$$

$$\hat{\underset{\sim}{\alpha}}^{(j)'}\hat{\Sigma}_{11}\hat{\underset{\sim}{\alpha}}^{(j)} = 1$$

$$\hat{\underset{\sim}{\gamma}}^{(j)'}\hat{\Sigma}_{22}\hat{\underset{\sim}{\gamma}}^{(j)} = 1, \quad j = 1, 2, \ldots, p_1$$

and $\hat{\Gamma}_2$ satisfies

$$\hat{\Gamma}_2'\hat{\Sigma}_{22}\hat{\Gamma}_1 = 0$$

$$\hat{\Gamma}_2'\hat{\Sigma}_{22}\hat{\Gamma}_2 = I.$$

When the restrictions on Γ_2 are made, then $\hat{A}, \hat{\Gamma}$ and $\hat{\Lambda}$ are uniquely defined.

Determining the Number of Useful Canonical Variables

Now we would like to develop a hypothesis test to determine the number of useful canonical variables. Similar results were developed in the case of principal components based on the test of $H_0 : \Sigma_{ij} = 0$. For a p component random vector $\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_k)'$ was partitioned into k subvectors of $p_1, p_2, \ldots, p_k$ components. Further, the Σ and A were suitably partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \ldots & \Sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \ldots & \Sigma_{kk} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \ldots A_{1k} \\ A_{21} & A_{22} \ldots A_{2k} \\ \vdots & \vdots \ddots \vdots \\ A_{k1} & A_{k2} \ldots A_{kk} \end{pmatrix}.$$

The likelihood ratio test statistic was derived as

$$\Lambda = \frac{|A|^{\frac{N}{2}}}{\prod_{i=1}^{k} |A_{ii}|^{\frac{N}{2}}}.$$

Now, in the case of canonical variables, there are two sets of variables $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ having p_1 and p_2 components, respectively, with $p_1 \leq p_2$. Then the likelihood ratio test statistic becomes

$$\begin{aligned} \Lambda &= \left(\frac{|A|}{|A_{11}||A_{22}|} \right)^{\frac{N}{2}} \\ &= |I - A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}|^{\frac{N}{2}} \\ &= \prod_{i=1}^{p_1} (1 - r_i^2)^{\frac{N}{2}}. \end{aligned}$$

[Note and recall that the second equation is the same as that obtained earlier for $\lambda_1^2, \lambda_2^2, \ldots, \lambda_{p_1}^2$ for finding out the canonical correlation]

Thus H_0 for testing of hypothesis for the canonical correlations $\rho_1, \rho_2, \ldots, \rho_{p_1}$ given by

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_{p_1}$$

is equivalent to testing

$$H_0 : \Sigma_{12} = 0.$$

If $H_0 : \rho_1 = \rho_2 = \ldots = \rho_{p_1}$ is accepted, then it clearly indicates that there are no useful canonical variables.

If $H_0 : \rho_1 = \rho_2 = \ldots = \rho_{p_1} = 0$ is rejected, then it is possible that $\rho_1 > \rho_2 = \rho_3 = \ldots = \rho_{p_1} = 0$ which indicates that only first canonical variable is useful. This process of testing is continued further.

In practice, the following hypothesis is tested to test the sequence of the null hypothesis:

$$H_k : \rho_{k+1} = \rho_{k+2} = \ldots = \rho_{p_1} = 0; \quad k = 0, 1, \ldots, p-1$$

. The null hypothesis $H_0 : \rho_1 = \rho_2 = \ldots = \rho_{p_1} = 0$ can be tested using

$$\Lambda_1 = \prod_{i=1}^{p_1} (1 - r_i^2).$$

The statistic

$$\Lambda = \frac{|A|}{|A_{11}| \cdot |A_{22}|} = \frac{|R|}{|R_{11}| \cdot |R_{22}|}$$

follows Wilks Lambda distribution, denoted as $\Lambda(p_1, p_2, n-1-p_2)$ and the decision rule is to reject H_0 at ε level of significance if $\Lambda \leq \Lambda_\varepsilon$.

Alternatively, we can use F approximation as follows

$$F = \frac{1 - \Lambda_1^{1/t}}{\Lambda_1^{1/t}} \cdot \frac{df_2}{df_1}$$

which follows F-distribution $F(df_1, df_2)$ under H_0 where

$$\begin{aligned}
df_1 &= p_1p_2 \\
df_2 &= \omega t - \frac{p_1p_2}{2} + 1 \\
\omega &= n - \frac{p_1 + p_2 + 3}{2} \\
t &= \sqrt{\frac{p_1^2 p_2^2 - 4}{p_1^2 + p_2^2 - 5}}\,.
\end{aligned}$$

The decision rule is to reject H_0 if $F > F_\varepsilon$ at ε level of significance.

This test of asymptotic level ε has the decision rule to reject H_0 if

$$-\left[n - \frac{p_1 + p_2 + 1}{2}\right] \log \Lambda_1 > C_{p_1p_2}(\varepsilon)$$

where $C_{p_1p_2}(\varepsilon)$ is the upper $\varepsilon\%$ point on χ^2 with p_1p_2 degrees of freedom at ε level of significance.

In terms of canonical correlations,

$$\Lambda \equiv \Lambda_1 = \prod_{i=1}^{p}(1 - r_i^2).$$

If one or more r_i^2 is large, then Λ_1 is small.

Similarly, we can define $\Lambda_2, \Lambda_3, \ldots$ to test other canonical correlations as

$$\Lambda_k = \prod_{i=k+1}^{p_1}(1 - r_i^2).$$

If the parameters exceed the range of critical values for Wilks Λ, we can use χ^2 approximation given as

$$\chi^2 = -\left[n - \frac{p_1 + p_2 + 3}{2}\right] \ln \Lambda,$$

which follows $\chi^2_{p_1p_2}$ under H_0.

Fujikoshi (1974) showed that the likelihood ratio test of H_k rejects H_k for small values of the statistic

$$\Lambda_k = \prod_{i=k+1}^{p_1}(1 - r_i^2).$$

The asymptotic distribution of Λ_k as $n \to \infty$ is

$$-n \ln \Lambda_k \sim \chi^2_{(p_1-k)(p_2-k)}$$

when H_k is true. An improvement over the statistic is

$$-\left[n - \frac{p_1+p_2+1}{2}\right] \ln \Lambda_k.$$

Further improvement, when H_k is true, the asymptotic distribution is

$$L_k = -\left[n - k - \frac{p_1+p_2+1}{2} + \sum_{i=1}^{k} \Lambda_i \cdot \frac{1}{r_i^2}\right] \ln \Lambda_k$$

and

$$L_k \sim \chi^2_{(p_1-k)(p_2-k)}.$$

Also

$$E(L_k) = (p_1-k)(p_2-k) + O(n^{-2})$$

Some other multivariate tests are Pillai's test, Lawley-Hotelling statistic and Roy's largest root statistic.

1. Pillai's Test

The Pillai's test for the significance of canonical correlation is conducted by

$$V^{(s)} = \sum_{i=1}^{s} r_i^2.$$

The upper percentage points of $V^{(s)}$ are available in Tables with $s = \min(p_1, p_2)$, $m = \frac{1}{2}(|p_2 - p_1| - 1)$ and $N = \frac{1}{2}(n - p_2 - p_1 - 2)$. This can also be approximated by F test.

2. Lawley-Hotelling Statistic

The Lawley-Hotelling statistic for testing the canonical correlations is

$$U^{(s)} = \sum_{i=1}^{s} \frac{r_i^2}{1-r_i^2}.$$

The upper percentage points for $\frac{\nu_E}{\nu_H} U^{(s)}$ are available in tables with $\nu_H = p_2$, $\nu_E = n - p_2 - 1$.

Some results on Distributions of Canonical Correlations

Theorem 1: Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ and $\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2)'$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ where $\underset{\sim}{X}_1$ and $\underset{\sim}{X}^{(2)}$ are p_1 and p_2 components column vectors. Let $r_1^2, r_2^2, \ldots, r_{p_1}^2$ are the sample canonical correlation coefficients which are the characteristic roots of $A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}$ and $A \sim W_{p_1+p_2}(\Sigma, n)$, $n \geq p_1 + p_2$. Then the joint density function of $r_1^2, r_2^2, \ldots, r_{p_1}^2$ is

$$\prod_{i=1}^{p}(1-\rho_i^2)^{n/2} {}_2F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{p_2}{2}; P^2, R^2\right) \cdot \frac{\pi^{p_1^2/2}}{\Gamma(\frac{p_1}{2})} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-p_2}{2})\Gamma(\frac{p_2}{2})}$$
$$\times \frac{\pi^{p_1^2/2}}{\Gamma(\frac{p_1}{2})} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-p_2}{2})\Gamma(\frac{p_2}{2})} \cdot \prod_{i=1}^{p_1}\left[(r_i^2)^{\frac{p_2-p_1-1}{2}}(1-r_i^2)^{\frac{n-p_1-p_2-1}{2}}\right]$$
$$\times \prod_{i<j}^{p_1}(r_i^2 - r_j^2) \quad ; 0 < r_{p_1}^2 < r_{p_1-1}^2 < \ldots < r_1^2 < 1$$

where $\rho_1^2, \rho_2^2, \ldots, \rho_{p_1}^2$ are the characteristic roots of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $P^2 = diag(\rho_1^2, \rho_2^2, \ldots, \rho_{p_1}^2)$ and $R^2 = diag(r_1^2, r_2^2, \ldots, r_{p_1}^2)$. The Hypergeometric function is

$$\begin{aligned} &{}_pF_q(a_1, a_2, \ldots, a_{p_1}; b_1, b_2, \ldots, b_{p_2}; z) \\ &= \sum_{k=0}^{\infty} \frac{(a_1)_k \ldots (a_{p1})_k}{(b_1)_k \ldots (b_{p_2})_k} \cdot \frac{z^k}{k!}; \end{aligned}$$

$(a)_k = a(a+1)\ldots(a+k-1)$.

Corollary 1.1: When $P = 0$, the joint density of $r_1^2, r_2^2, \ldots, r_{p_1}^2$ is

$$\frac{\pi^{p_1^2/2}\Gamma(\frac{n}{2})}{\Gamma(\frac{p_1}{2})\Gamma(\frac{n-p_2}{2})\Gamma(\frac{p_2}{2})} \prod_{i=1}^{p_1}\left[(r_i^2)^{\frac{p_2-p_1-1}{2}}(1-r_i^2)^{\frac{n-p_2-p_1-1}{2}}\right] \prod_{i<j}^{p_1}(r_i^2 - r_j^2)$$
$$; 0 < r_{p1}^2 < \ldots < r_1^2 < 1.$$

In general, the marginal distribution of a single canonical correlation, or of any subset of the sample canonical correlations, can be obtained from this result. However, the integrals involved are not particularly tractable.

Theorem 2: An asymptotic representation for large n of the joint density function of $r_1^2, r_2^2, \ldots, r_{p_1}^2$ when the population canonical correlation coefficient satisfy

$$0 = \rho_{p_2} = \rho_{p_2-1} = \ldots = \rho_{k+1} < \rho_k < \ldots < \rho_2 < \rho_1 < 1$$

is

$$K_2 \prod_{i=1}^{k} \left[(1 - r_i \rho_i)^{-n + \frac{p_1+p_2+1}{2}} (r_i^2)^{\frac{p_2-p_1}{4} - \frac{1}{2}} (1 - r_i^2)^{\frac{n-p_2-p_1-1}{2}} \right]$$

$$\times \prod_{i<j}^{k} \left(\frac{r_i^2 - r_j^2}{\rho_i^2 - \rho_j^2} \right)^{\frac{1}{2}} \prod_{i=1}^{k} \prod_{j=k+1}^{p_1} (r_i^2 - r_j^2)^{\frac{1}{2}}$$

$$\times \prod_{i=k+1}^{p_1} \left[(r_i^2)^{\frac{p_2-p_1-1}{2}} (1 - r_i^2)^{\frac{n-p_2-p_1-1}{2}} \right] \cdot \prod_{\substack{k+1 \\ i<j}}^{p_1} (r_i^2 - r_j^2)$$

where,

$$K_2 = K_1 \frac{\Gamma(\frac{p_1-k}{2}) \pi^{\frac{k}{2}}}{\Gamma(\frac{p_1-k}{2}) \Gamma(\frac{n-p_2-k}{2}) \Gamma(\frac{p_2-k}{2})} \prod_{i=1}^{k} \left[(1 - \rho_i^2)^{\frac{n}{2}} \rho_i^{k - \frac{p_1+p_2}{2}} \right]$$

$$K_1 = \frac{\left(\frac{n}{2}\right)^{-k\left(\frac{p_1+p_2-k-1}{2}\right)}}{2^k \pi^{\frac{k(k+1)}{2}}} \Gamma\left(\frac{p_2}{2}\right) \Gamma\left(\frac{p_1}{2}\right).$$

Corollary 2.1: Under the conditions of Theorem 1, the asymptotic conditional density function for large n of $r_{k+1}^2, r_{k+2}^2, \ldots, r_{p_1}^2$ [i.e., the square of smallest $(p_1 - k)$ sample correlation coefficients] given the k largest coefficients $r_1^2, r_2^2, \ldots, r_k^2$ is

$$K \prod_{i=1}^{k} \prod_{j=k+1}^{p_1} (r_i^2 - r_j^2)^{\frac{1}{2}} \prod_{i=k+1}^{p} \left[(r_i^2)^{\frac{p_2-p_1-1}{2}} (1 - r_i^2)^{\frac{n-p_2-p_1-1}{2}} \right] \prod_{\substack{k+1 \\ i<j}}^{p} (r_i^2 - r_j^2)$$

where K is a constant. **CHECK**

Note that this asymptotic conditional density function does not depend on $\rho_1^2, \rho_2^2, \ldots, \rho_k^2$, the nonzero population correlation coefficient, so that $r_1^2, r_2^2, \ldots, r_k^2$ are asymptotically sufficient for $\rho_1^2, \rho_2^2, \ldots, \rho_k^2$.

Corollary 2.2: Assuming the conditions of Theorem 1 hold, and put

$$x_i = \frac{\sqrt{n}(r_i^2 - \rho_i^2)}{2\rho_i(1 - \rho_i^2)} \ , \ i = 1, 2, \ldots, k$$

$$x_j = n r_j^2 \ , \ j = k+1, k+2, \ldots, p_1.$$

Then the limiting joint density function of $x_1, x_2, \ldots, x_{p_1}$ as $n \to \infty$ is

$$\prod_{i=1}^{k} \phi(x_i) \cdot \frac{\pi^{\frac{(p_1-k)^2}{2}} \exp\left[-\frac{1}{2} \sum_{j=k+1}^{p_1} x_j \right]}{2^{\frac{(p_1-k)(p_2-k)}{2}} \Gamma\left(\frac{p_2-p_1}{2}\right) \Gamma\left(\frac{p_1-k}{2}\right)} \times \prod_{j=k+1}^{p_1} x_j^{\frac{p_2-p_1-1}{2}} \prod_{\substack{k=1 \\ i<j}}^{p} (x_i - x_j),$$

where $\phi(\cdot)$ denotes the standard normal density function.