# Chapter 2
# Multivariate Normal Distribution

## Multivariate Normal Distribution

A generalization of the univariate bell-shaped normal distribution is the multivariate normal distribution. Most multivariate statistical techniques assume that the data are generated from a multivariate normal distribution. We derive the multivariate analogue of the univariate normal distribution.

The probability density function (PDF) of the univariate normal distribution $N(\mu, \sigma^2)$ is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty,\ -\infty < \mu < \infty,\ 0 < \sigma^2 < \infty.$$

which can be expressed as

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right].$$

Note that $\left(\frac{1}{\sigma\sqrt{2\pi}}\right)$ is a constant and $(x-\mu)(\sigma^2)^{-1}(x-\mu)$ is expressible as a quadratic form.

Now consider $p$ random variables $X_1, X_2, \ldots, X_p$ in the form of a $p \times 1$ random vector $\underset{\sim}{X} = (X_1, X_2, \ldots, X_p)'$. To derive an analogue multivariate form of the univariate $N(\mu, \sigma^2)$, consider replacing scalar $x$ by vector $\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)'$, scalar $\mu$ by a vector $\underset{\sim}{b} = (b_1, b_2, \ldots, b_p)'$ and scalar $(\sigma^2)^{-1}$ by a positive definite matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}.$$

Assume $A$ to be a full-rank matrix (which helps later when we need to consider its inverse). Now replace

$$(x-\mu)'(\sigma^2)^{-1}(x-\mu) \quad \text{by} \quad (\underset{\sim}{x}-\underset{\sim}{b})'A(\underset{\sim}{x}-\underset{\sim}{b}).$$

Note that

$$(\underset{\sim}{x} - \underset{\sim}{b})' A (\underset{\sim}{x} - \underset{\sim}{b}) = \sum_i \sum_j a_{ij}(x_i - b_i)(x_j - b_j).$$

Consider the analogous form of the density function of a $p$-variate normal distribution as

$$f_{\underset{\sim}{X}}(x_1, x_2, \ldots, x_p) = K \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{b})' A (\underset{\sim}{x} - \underset{\sim}{b})\right],$$

where the constant $K > 0$ is chosen such that the integral over the entire $p$-dimensional Euclidean space of $x_1, x_2, \ldots, x_p$ is unity.

The next objective is to determine $K$, $\underset{\sim}{b}$ and $A$.

### (i) Determining $K$

Observe that $f_{\underset{\sim}{X}}(x_1, x_2, \ldots, x_p) \geq 0$, and since $A$ is a positive definite matrix, so

$$(\underset{\sim}{x} - \underset{\sim}{b})' A (\underset{\sim}{x} - \underset{\sim}{b}) \geq 0.$$

Thus, the density is bounded, i.e.,

$$f_{\underset{\sim}{X}}(x_1, x_2, \ldots, x_p) \leq K.$$

Determine $K$ such that

$$K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{b})' A (\underset{\sim}{x} - \underset{\sim}{b})\right] dx_p \cdots dx_1 = 1.$$

Let

$$K^* = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{b})' A (\underset{\sim}{x} - \underset{\sim}{b})\right] dx_p \cdots dx_1.$$

Note that

$$KK^* = 1.$$

Since $A$ is a positive definite matrix, there exists a non-singular matrix $C$ such that

$$C'AC = I.$$

Pre-multiplying by $(C')^{-1}$ and post-multiplying by $C'$ on both sides of $C'AC = I$, we get

$$\begin{aligned} & C'AC = I \\ \text{or } & (C')^{-1}C'ACC^{-1} = (C')^{-1}C^{-1} \\ \text{or } & A = (CC')^{-1} \\ \text{or } & A^{-1} = CC'. \end{aligned}$$

Let $\underset{\sim}{x} - \underset{\sim}{b} = C\underset{\sim}{y}$, where $\underset{\sim}{y} = (y_1, y_2, \ldots, y_p)'$. Now transform $\underset{\sim}{x}$ to $\underset{\sim}{y}$ as

$$(\underset{\sim}{x} - \underset{\sim}{b})'A(\underset{\sim}{x} - \underset{\sim}{b}) = \underset{\sim}{y}'C'AC\underset{\sim}{y} = \underset{\sim}{y}'I\underset{\sim}{y} = \underset{\sim}{y}'\underset{\sim}{y}, \quad [\text{using } C'AC = I].$$

The Jacobian of the transformation is

$$J = \text{mod } |C|,$$

where $|C| = \det(C)$, i.e., the determinant of $C$ and mod $|C|$ is the absolute value of $|C|$. Thus,

$$|J| = \frac{\partial(x_1, x_2, \ldots, x_p)}{\partial(y_1, y_2, \ldots, y_p)} = |C|$$

Therefore,

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = f_{\underset{\sim}{X}}(\underset{\sim}{x})\,|J|.$$

Then from

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = K \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{b})'A(\underset{\sim}{x} - \underset{\sim}{b})\right],$$

we get

$$\begin{aligned} f_{\underset{\sim}{Y}}(\underset{\sim}{y}) &= K \exp\left(-\frac{1}{2}\underset{\sim}{y}'C'AC\underset{\sim}{y}\right).|C| \\ &= K|CC'|^{1/2} \exp\left(-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}\right) \\ &= K|A^{-1}|^{1/2} \exp\left(-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}\right) \quad (\text{using } A' = CC'). \end{aligned}$$

Considering $f_{\underset{\sim}{Y}}(\underset{\sim}{y})$ to be the PDF,

$$\begin{aligned}
&\int_{\mathbb{R}^p} f_{\underset{\sim}{Y}}(\underset{\sim}{y})\, d\underset{\sim}{y} = 1 \\
\text{or}\qquad & K|A^{-1}|^{1/2} \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}\right) \prod_{i=1}^{p} dy_i = 1 \\
\text{or}\qquad & K|A^{-1}|^{1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{p} y_i^2\right) \prod_{i=1}^{p} dy_i = 1 \\
\text{or}\qquad & K|A^{-1}|^{1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{p} \exp\left(-\frac{1}{2}y_i^2\right) dy_i = 1 \\
\text{or}\qquad & K|A^{-1}|^{1/2} \prod_{i=1}^{p} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y_i^2\right) dy_i = 1.
\end{aligned}$$

Since $y_i'$ s are i.i.d. and $y_i \sim N(0,1)$, so

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_i^2\right) dy_i = 1 \\
\text{or}\qquad &\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y_i^2\right) dy_i = \sqrt{2\pi}.
\end{aligned}$$

Then

$$\begin{aligned}
&K|A^{-1}|^{1/2} \prod_{i=1}^{p} (\sqrt{2\pi}) = 1. \\
\text{or}\qquad & K = (2\pi)^{-p/2}|A|^{1/2} \\
\text{or}\qquad & K = \frac{|A|^{1/2}}{(2\pi)^{p/2}}.
\end{aligned}$$

Hence

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{\sqrt{|A|}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{b})'A(\underset{\sim}{x}-\underset{\sim}{b})\right].$$

Alternatively, the same result can be found as follows. Since $C'AC = I$, so

$$|C'| \cdot |A| \cdot |C| = |I|.$$

Since $|I| = 1$ and $|C'| = |C|$,

$$\text{mod } |C| = \frac{1}{|A|^{1/2}}$$

and thus,

$$K = \frac{1}{K^*} = \frac{|A|^{1/2}}{(2\pi)^{p/2}}.$$

**(ii) Finding $\underset{\sim}{b}$**

Next we find the parameter $\underset{\sim}{b}$ as follows using $\underset{\sim}{Y} = (\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \ldots, Y_p)'$ and $\underset{\sim}{y} = (y_1, y_2, \ldots, y_p)'$. Consider the marginal density function,

$$\begin{aligned}
E(Y_i) \; &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i \prod_{j=1}^{p} \left[ \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} y_j^2 \right) \right] \, dy_1 \cdots dy_p \\
&= \frac{1}{\sqrt{2\pi}} \left[ \int_{-\infty}^{\infty} y_i \exp\left( -\frac{1}{2} y_i^2 \right) dy_i \right] \prod_{\substack{j=1 \\ j \neq i}}^{p} \left[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} y_j^2 \right) dy_j \right] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i \exp\left( -\frac{1}{2} y_i^2 \right) \, dy_i \cdot 1 \\
&= 0 \qquad \text{[being an odd function of } y_i.\text{]}
\end{aligned}$$

Thus

$$E(\underset{\sim}{Y}) = \begin{pmatrix} E(\underset{\sim}{Y}) \\ E(\underset{\sim}{Y}_2) \\ \vdots \\ E(Y_p) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \underset{\sim}{0}.$$

Since

$$\underset{\sim}{X} - \underset{\sim}{b} = C\underset{\sim}{Y},$$

so

$$E(\underset{\sim}{X}) - \underset{\sim}{b} = CE(\underset{\sim}{Y}) = \underset{\sim}{0}.$$

Hence

$$E(\underset{\sim}{X}) = \underset{\sim}{b},$$

i.e, the mean vector of $\underset{\sim}{X}$ is $\underset{\sim}{b}$. So we denote it by $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$, as

$$\underset{\sim}{\mu} = E(\underset{\sim}{X}) = \underset{\sim}{b},$$

where $E(X_i) = \mu_i, \quad i = 1, 2, \ldots, p.$

**(iii) Determining** $A$

Consider the expectation of the $p \times p$ symmetric matrix $(\underset{\sim}{X} - \underset{\sim}{b})(\underset{\sim}{X} - \underset{\sim}{b})'$ as

$$\begin{aligned}
E(\underset{\sim}{X} - \underset{\sim}{b})(\underset{\sim}{X} - \underset{\sim}{b})' &= E(C\underset{\sim}{Y}\,\underset{\sim}{Y}'C') \\
&= C\,E(\underset{\sim}{Y}\,\underset{\sim}{Y}')\,C'. \\
&= C\,E\begin{pmatrix} Y_1^2 & Y_1Y_2 & \cdots & Y_1Y_p \\ Y_2Y_1 & Y_2^2 & \cdots & Y_2Y_p \\ \vdots & \vdots & \ddots & \vdots \\ Y_pY_1 & Y_pY_2 & \cdots & Y_p^2 \end{pmatrix} C'.
\end{aligned}$$

Now we find $E(\underset{\sim}{Y}\,\underset{\sim}{Y}')$ which consists of diagonal and off-diagonal elements.

Note that the density of $\underset{\sim}{Y}$ is

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{1}{(2\pi)^{p/2}}\exp\left(-\frac{1}{2}\underset{\sim}{y}'\,\underset{\sim}{y}\right) = \prod_{j=1}^{p}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_j^2\right).$$

Consider

$$E(Y_iY_j) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} y_iy_j\prod_{h=1}^{p}\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_h^2\right)\right]dy_1\cdots dy_p.$$

When $i = j$, then

$$\begin{aligned}
E(Y_iY_j) &= E(Y_i^2) \\
&= \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y_i^2\exp\left(-\frac{1}{2}y_i^2\right)dy_i\right]\left[\prod_{\substack{h=1\\h\neq i}}^{p}\left\{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_h^2\right)dy_h\right\}\right] \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y_i^2\exp\left(-\frac{1}{2}y_i^2\right)dy_i\cdot 1 \\
&= 1 \qquad \text{[being even function of } y_i\text{]}.
\end{aligned}$$

When $i \neq j$ then,

$$\begin{aligned}
E(Y_iY_j) = &\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y_i\exp\left(-\frac{1}{2}y_i^2\right)dy_i\right]\times\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y_j\exp\left(-\frac{1}{2}y_j^2\right)dy_j\right] \\
&\times\prod_{\substack{h=1\\h\neq i,j}}^{p}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}y_h^2\right)dy_h\right] \\
= \ &0 \qquad \text{[being odd function of } y_i \text{ and } y_j,\ \text{first and second integrals are zero.]}
\end{aligned}$$

Thus

$$
\begin{aligned}
E(\underset{\sim}{X}-\underset{\sim}{b})(\underset{\sim}{X}-\underset{\sim}{b})' &= C\begin{pmatrix} E(Y_1^2) & E(Y_1Y_2) & \cdots & E(Y_1Y_p) \\ E(Y_2Y_1) & E(Y_2^2) & \cdots & E(Y_2Y_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(Y_pY_1) & E(Y_pY_2) & \cdots & E(Y_p^2) \end{pmatrix} C' \\
&= C\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} C' \\
&= CIC' \\
&= CC' \\
&= A^{-1}.
\end{aligned}
$$

Since $\underset{\sim}{b}=\underset{\sim}{\mu}$, so $E(\underset{\sim}{X}-\underset{\sim}{\mu})(\underset{\sim}{X}-\underset{\sim}{\mu})'$ is the covariance matrix of $\underset{\sim}{X}$, so we denote it as $\Sigma$.

Thus

$$A^{-1}=\Sigma.$$

and

$$E(\underset{\sim}{X}-\underset{\sim}{\mu})(\underset{\sim}{X}-\underset{\sim}{\mu})' = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

which is a positive definite symmetric matrix.

Substituting $K$, $\underset{\sim}{b}$ and $A$ in $K\exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{b})'A(\underset{\sim}{x}-\underset{\sim}{b})\right]$, we obtain the probability density function of the $p$-variate multivariate normal distribution as

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

and we write

$$\underset{\sim}{X} \sim N_p\left(\underset{\sim}{\mu},\Sigma\right).$$

Hence, we obtain the following result:

**Theorem 1:** If the density of a $p$-dimensional random vector $\underset{\sim}{X}$ is

$$\frac{|A|^{1/2}}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{b})' A (\underset{\sim}{x} - \underset{\sim}{b})\right],$$

then

$$E(\underset{\sim}{X}) = \underset{\sim}{b} \qquad \text{and} \qquad Cov(\underset{\sim}{X}) = A^{-1}.$$

Conversely, given a vector $\underset{\sim}{\mu}$ and positive definite matrix $\Sigma$, there is a multivariate normal density

$$\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu})\right]$$

such that

$$E(\underset{\sim}{X}) = \underset{\sim}{\mu} \quad \text{and} \quad Cov(\underset{\sim}{X}) = \Sigma$$

with this density.

**Note 1:** If $\Sigma$ is not of full rank, then $\Sigma$ is not uniquely invertible and the distribution is known as a *singular distribution*. In this case, the density is concentrated on a lower-dimensional space which is less than $p$.

**Note 2:** It is important to note that $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, where $X_1, X_2, \ldots, X_p$ are not necessarily independent. Suppose we use $\underset{\sim}{X} - \underset{\sim}{\mu} = C\underset{\sim}{Y}$ which implies $\underset{\sim}{Y} = C^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$, where $C$ is such that $C'\Sigma^{-1}C = I$ or $\Sigma = CC'$. Then $\underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, I)$, and $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \ldots, \underset{\sim}{Y}_p$ are now independent.

## Characteristic Function of Multivariate Normal Distribution

The characteristic function of a random variable $X$ in a univariate case is defined as

$$\phi(t) = E\left[\exp(itX)\right] = \int_{\mathbb{R}} e^{itx} f_X(x)\, dx$$

and for the multivariate case, the characteristic function of a $p$-component random vector $\underset{\sim}{X}$, is defined as

$$\phi_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\left[\exp(i\underset{\sim}{t}'\underset{\sim}{X})\right] = \int_{\mathbb{R}_p} e^{i\underset{\sim}{t}'\underset{\sim}{x}} f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d\underset{\sim}{x}.$$

It is important to note that uniqueness and inversion theorems help identify probability distributions. The uniqueness theorem states that two random variables have the same characteristic function if and only if they have the same PDF. So if we know the characteristic function, we can uniquely determine the PDF. The inversion theorem helps determine the PDF from a given characteristic function.

Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$. The characteristic function is given as

$$\phi_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\left[\exp(i\underset{\sim}{t}'\underset{\sim}{X})\right].$$

Substitute $\underset{\sim}{X} - \underset{\sim}{\mu} = C\underset{\sim}{Y}$ where the non singular matrix $C$ is determined such that $C'\Sigma^{-1}C = I$ or $\Sigma^{-1} = CC'$. Then $\underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, I)$.

The characteristic function is

$$\begin{aligned}
\phi_{\underset{\sim}{X}}(\underset{\sim}{t}) &= E\left[\exp\left\{i\underset{\sim}{t}'(C\underset{\sim}{Y} + \underset{\sim}{\mu})\right\}\right] \\
&= E\left[\exp\left(i\underset{\sim}{t}'C\underset{\sim}{\mu} + i\underset{\sim}{t}'C\underset{\sim}{Y}\right)\right] \\
&= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}\right) E\left[\exp\left(i\underset{\sim}{u}'\underset{\sim}{Y}\right)\right] \quad (\text{where } \underset{\sim}{u}' = \underset{\sim}{t}'C). \\
&= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}\right) E\left[\exp\left(i\sum_{j=1}^{p} u_j Y_j\right)\right] \\
&= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}\right) \prod_{j=1}^{p} E\left[\exp\left(iu_j Y_j\right)\right].
\end{aligned}$$

For a univariate random variable $W \sim N(\mu, \sigma^2)$, the characteristic function is given by

$$\phi_W(t) = \exp\left(it\mu - \frac{1}{2}\sigma^2 t^2\right).$$

If $\mu = 0$, $\sigma^2 = 1$, then

$$\phi_W(t) = E\left[\exp(itW)\right] = \exp\left(-\frac{t^2}{2}\right).$$

Note that $Y_j \sim N(0, 1)$, so

$$\phi_{Y_j}(t) = \exp\left(-\frac{t^2}{2}\right),$$

so

$$E\left[\exp\left(iu_j Y_j\right)\right] = \exp\left(-\frac{u_j^2}{2}\right).$$

Thus,

$$\begin{aligned}\phi_{\underset{\sim}{X}}(\underset{\sim}{u}) &= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}\right)\prod_{j=1}^{p}\exp\left(-\frac{u_j^2}{2}\right)\\ &= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}\right)\exp\left(-\frac{1}{2}\sum_{j=1}^{p}u_j^2\right).\\ &= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}\right)\exp\left(-\frac{1}{2}\underset{\sim}{u}'\underset{\sim}{u}\right)\\ &= \exp\left(i\underset{\sim}{u}'\underset{\sim}{\mu}-\frac{1}{2}\underset{\sim}{u}'\underset{\sim}{u}\right).\end{aligned}$$

Substituting

$$\underset{\sim}{u}'\underset{\sim}{u} = \underset{\sim}{t}'CC'\underset{\sim}{t} = \underset{\sim}{t}'\Sigma\underset{\sim}{t},$$

we get

$$\phi_{\underset{\sim}{X}}(\underset{\sim}{t}) = \exp\left[i\underset{\sim}{t}'\underset{\sim}{\mu}-\frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}\right].$$

## Distribution of Linear Combinations of Normally Distributed Random Variables

Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ and $\underset{\sim}{Z} = A\underset{\sim}{X}$, where $\underset{\sim}{X}$ is a $p \times 1$ vector, $A$ is an $m \times p$ matrix of coefficients of $m$ linear combinations of $\underset{\sim}{X}$, and $\underset{\sim}{Z}$ is an $m \times 1$ vector. Here $Rank(A) = m \leq p$.

Consider the characteristic function of $\underset{\sim}{Z}$ as follows:

$$\begin{aligned}\phi_{\underset{\sim}{Z}}(\underset{\sim}{t}) &= E\left[\exp\left(i\underset{\sim}{t}'\underset{\sim}{Z}\right)\right]\\ &= E\left[\exp\left(i(\underset{\sim}{t}'A)\underset{\sim}{X}\right)\right]\\ &= E\left[\exp\left(i\underset{\sim}{\theta}'\underset{\sim}{X}\right)\right] \quad (\text{where } \underset{\sim}{\theta}' = \underset{\sim}{t}'A)\\ &= \exp\left[i\underset{\sim}{\theta}'\underset{\sim}{\mu}-\frac{1}{2}\underset{\sim}{\theta}'\Sigma\underset{\sim}{\theta}\right]\\ &= \exp\left[i\underset{\sim}{t}'(A\underset{\sim}{\mu})-\frac{1}{2}\underset{\sim}{t}'(A\Sigma A')\underset{\sim}{t}\right].\end{aligned}$$

Using the uniqueness theorem of the characteristic function, the characteristic function is recognizable as

$$N_m(A\underset{\sim}{\mu}, A\Sigma A'),$$

i.e., an $m$-variate multivariate normal distribution with mean vector $A\underset{\sim}{\mu}$ and covariance matrix $A\Sigma A'$. Since the characteristic function determines the distribution uniquely, it follows that

$$\underset{\sim}{Z} \sim N_m(A\underset{\sim}{\mu}, A\Sigma A').$$

Hence, the linear transformation of a multivariate normal distribution is also a multivariate normal distribution.

**Alternative Method to Derive the Distribution of Linear Combinations of Normally Distributed Random Variables**

Consider $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ and $\underset{\sim}{Z} = C\underset{\sim}{X}$, where $\underset{\sim}{Z}$ is an $p \times 1$ vector and $C$ is an $m \times p$ nonsingular full rank matrix of the coefficients of $m$ linear combinations of $\underset{\sim}{X}$, and $\underset{\sim}{X}$ is a $p \times 1$ vector.

Here

$$\underset{\sim}{X} = C^{-1}\underset{\sim}{Z},$$

and we want to obtain the density of $\underset{\sim}{Z}$. The Jacobian of transformation is

$$\begin{aligned}
|J| &= \text{ mod}\left|C^{-1}\right| \\
&= \frac{1}{\text{mod}\,|C|} \\
&= \sqrt{\frac{1}{|C|^2}} \\
&= \sqrt{\frac{|\Sigma|}{|C|\,|\Sigma|\,|C'|}} \\
&= \frac{|\Sigma|^{1/2}}{|C\Sigma C'|^{1/2}}.
\end{aligned}$$

Consider the quadratic form from the exponent of $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ as

$$\begin{aligned}
(\underset{\sim}{x} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}) &= (C^{-1}\underset{\sim}{z} - \underset{\sim}{\mu})'\Sigma^{-1}(C^{-1}\underset{\sim}{z} - \underset{\sim}{\mu}) \\
&= (C^{-1}\underset{\sim}{z} - C^{-1}C\underset{\sim}{\mu})'\Sigma^{-1}(C^{-1}\underset{\sim}{z} - C^{-1}C\underset{\sim}{\mu}) \\
&= \left[C^{-1}(\underset{\sim}{z} - C\underset{\sim}{\mu})\right]'\Sigma^{-1}\left[C^{-1}(\underset{\sim}{z} - C\underset{\sim}{\mu})\right] \\
&= (\underset{\sim}{z} - C\underset{\sim}{\mu})'(C\Sigma C')^{-1}(\underset{\sim}{z} - C\underset{\sim}{\mu}),
\end{aligned}$$

since $CC' = I$ and $(C^{-1})' = (C')^{-1}$. Thus the density of $\underset{\sim}{Z}$ is

$$\begin{aligned}
f_{\underset{\sim}{Z}}(\underset{\sim}{z}) &= f_{\underset{\sim}{X}}(\underset{\sim}{x})\,|J| \\
&= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{z} - C\underset{\sim}{\mu})'(C\Sigma C')^{-1}(\underset{\sim}{z} - C\underset{\sim}{\mu})\right] \cdot \frac{|\Sigma|^{1/2}}{|C\Sigma C'|^{1/2}} \\
&= \frac{1}{(2\pi)^{p/2}|C\Sigma C'|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{z} - C\underset{\sim}{\mu})'(C\Sigma C')^{-1}(\underset{\sim}{z} - C\underset{\sim}{\mu})\right].
\end{aligned}$$

Thus

$$\underset{\sim}{Z} \sim N_m(C\underset{\sim}{\mu},\, C\Sigma C').$$

## Marginal Distributions

Let $\underset{\sim}{X} = (X_1, X_2, \ldots, X_p)'$ be a $p \times 1$ vector of random variables $X_1, X_2, \ldots, X_p$. If it is partitioned into two sub-vectors containing $q$ random variables in the first sub-vector and $(p-q)$ random variables in the second sub-vector, i.e.,

$$\underset{\sim}{X} = (X_1, X_2, \ldots, X_q, X_{q+1}, \ldots, X_p)' = (\underset{\sim}{X}_1, \underset{\sim}{X}_2)',$$

where $\underset{\sim}{X}_1 = (X_1, X_2, \ldots, X_q)'$ and $\underset{\sim}{X}_2 = (X_{q+1}, X_{q+2}, \ldots, X_p)'$.

Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ and $\underset{\sim}{\mu}$ as well as $\Sigma$ be partitioned in a similar way as $\underset{\sim}{\mu} = (\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2)'$, and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ such that

- $E(\underset{\sim}{X}_1) = \underset{\sim}{\mu}_1$ is a $q \times 1$ vector,
- $E(\underset{\sim}{X}_2) = \underset{\sim}{\mu}_2$ is a $(p-q) \times 1$ vector,

- $E\big[(\underset{\sim}{X}_1-\underset{\sim}{\mu}_1)(\underset{\sim}{X}_1-\underset{\sim}{\mu}_1)'\big]=\Sigma_{11}$ is a $q\times q$ matrix,
- $E\big[(\underset{\sim}{X}_2-\underset{\sim}{\mu}_2)(\underset{\sim}{X}_2-\underset{\sim}{\mu}_2)'\big]=\Sigma_{22}$ is a $(p-q)\times(p-q)$ matrix and
- $E\big[(\underset{\sim}{X}_2-\underset{\sim}{\mu}_2)(\underset{\sim}{X}_1-\underset{\sim}{\mu}_1)'\big]=\Sigma_{21}$ is a $(p-q)\times q$ matrix.

Also, $\underset{\sim}{X}$ can be partitioned into two subvectors as $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ in two ways - both sub-vectors are independently distributed and both sub-vectors are non-independently distributed. We consider both cases.

## Case I: When $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are Independently Distributed

If $\Sigma_{12}=0$, a null matrix, then $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are independently distributed. So consider

$$\Sigma=\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{pmatrix}.$$

Then

$$\Sigma^{-1}=\begin{pmatrix}\Sigma_{11}^{-1} & 0\\ 0 & \Sigma_{22}^{-1}\end{pmatrix},\qquad \text{and}\qquad |\Sigma|=|\Sigma_{11}|\,|\Sigma_{22}|.$$

Consider

$$\begin{aligned}
Q&=(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\\
&=\Big((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)' \quad (\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Big)\begin{pmatrix}\Sigma_{11}^{-1} & 0\\ 0 & \Sigma_{22}^{-1}\end{pmatrix}\begin{pmatrix}\underset{\sim}{x}_1-\underset{\sim}{\mu}_1\\ \underset{\sim}{x}_2-\underset{\sim}{\mu}_2\end{pmatrix}\\
&=(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)'\Sigma_{11}^{-1}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)+(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\\
&=Q_1+Q_2 \quad \text{(say.)}
\end{aligned}$$

where $Q_1=(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)'\Sigma_{11}^{-1}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)$ and $Q_2=(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)$.

**Joint PDF**

Consider the PDF of $\underset{\sim}{X}$ and substitute the partitioned values:

$$\begin{aligned}
f_{\underset{\sim}{X}}(\underset{\sim}{x}) &= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right) \\
&= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}Q\right) \\
&= \frac{1}{(2\pi)^{q/2}(2\pi)^{(p-q)/2}|\Sigma_{11}|^{1/2}|\Sigma_{22}|^{1/2}} \exp\left(-\frac{1}{2}(Q_1+Q_2)\right) \\
&= \frac{1}{(2\pi)^{q/2}|\Sigma_{11}|^{1/2}} \exp\left(-\frac{Q_1}{2}\right) \cdot \frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}} \exp\left(-\frac{Q_2}{2}\right) \\
&= N_q(\underset{\sim}{\mu}_1, \Sigma_{11}) \cdot N_{p-q}(\underset{\sim}{\mu}_2, \Sigma_{22}).
\end{aligned}$$

**Marginal distribution of $\underset{\sim}{X}_1$**

The marginal distribution of $\underset{\sim}{X}_1$ is now found as follows:

$$\begin{aligned}
f_{\underset{\sim}{X}_1}(\underset{\sim}{X}_1) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_p(\underset{\sim}{x}|\underset{\sim}{\mu}, \Sigma)\, dx_{q+1} \cdots dx_p \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_q(\underset{\sim}{x}_1|\underset{\sim}{\mu}_1, \Sigma_{11})\, N_{p-q}(\underset{\sim}{x}_2|\underset{\sim}{\mu}_2, \Sigma_{22})\, dx_{q+1} \cdots dx_p \\
&= N_q(\underset{\sim}{x}_1|\underset{\sim}{\mu}_1, \Sigma_{11}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_{p-q}(\underset{\sim}{x}_2|\underset{\sim}{\mu}_2, \Sigma_{22})\, dx_{q+1} \cdots dx_p \\
&= N_q(\underset{\sim}{x}_1|\underset{\sim}{\mu}_1, \Sigma_{11}) \cdot 1 \\
&= N_q(\underset{\sim}{x}_1|\underset{\sim}{\mu}_1, \Sigma_{11})
\end{aligned}$$

Therefore $\underset{\sim}{x}_1 \sim N_q(\underset{\sim}{\mu}_1, \Sigma_{11})$.

**Marginal distribution of $\underset{\sim}{X}_2$**

Similarly, the marginal distribution of $\underset{\sim}{X}_2$ is now found as follows:

$$\begin{aligned} f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_q(\underset{\sim}{x}_1|\underset{\sim}{\mu}_1, \Sigma_{11})\, N_{p-q}(\underset{\sim}{x}_2|\underset{\sim}{\mu}_2, \Sigma_{22})\, dx_1 \cdots dx_p \\ &= N_{p-q}(\underset{\sim}{x}_2|\underset{\sim}{\mu}_2, \Sigma_{22}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_q(\underset{\sim}{x}_1|\underset{\sim}{\mu}_1, \Sigma_{11})\, dx_1 \cdots dx_q \\ &= N_{p-q}(\underset{\sim}{x}_2|\underset{\sim}{\mu}_2, \Sigma_{22}) \cdot 1 \\ &= N_{p-q}(\underset{\sim}{x}_2|\underset{\sim}{\mu}_2, \Sigma_{22}) \end{aligned}$$

Therefore $\underset{\sim}{x}_2 \sim N_{p-q}(\underset{\sim}{\mu}_2, \Sigma_{22})$.

## Case II: When $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are Not Independent

In this case,

$$\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2)', \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\Sigma_{11}$ is a non-null matrix.

It is usually easier to handle the case when the random variables are independent than when they are not. So we aim to construct two random vectors $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ as a function of $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ such that $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ are independent, whereas $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are not independent.

Consider a nonsingular linear transformation to sub-vectors:

$$\begin{aligned} \underset{\sim}{Y}_1 &= \underset{\sim}{X}_1 + B\underset{\sim}{X}_2 \\ \underset{\sim}{Y}_2 &= \underset{\sim}{X}_2 \end{aligned}$$

and choose $B$ such that the components of $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ are uncorrelated, where $B$ is a

$q \times (p-q)$ fixed but unknown matrix.

$$\begin{aligned}
0 &= E[\{\underset{\sim}{Y}_1 - E(\underset{\sim}{Y}_1)\}\{\underset{\sim}{Y}_2 - E(\underset{\sim}{Y}_2)\}'] \\
&= E[\{\underset{\sim}{X}_1 + B\underset{\sim}{X}_2 - E(\underset{\sim}{X}_1) - BE(\underset{\sim}{X}_2)\}\{\underset{\sim}{X}_2 - E(\underset{\sim}{X}_2)\}'] \\
&= E[\{(\underset{\sim}{X}_1 - E(\underset{\sim}{X}_1)) + B(\underset{\sim}{X}_2 - E(\underset{\sim}{X}_2))\}\{\underset{\sim}{X}_2 - E(\underset{\sim}{X}_2)\}'] \\
&= \Sigma_{12} + B\Sigma_{22}
\end{aligned}$$

which implies

$$B = -\Sigma_{12}\Sigma_{22}^{-1}.$$

Hence,

$$\underset{\sim}{Y}_1 = \underset{\sim}{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{X}_2,$$

and we can write

$$\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \underset{\sim}{X}.$$

This is a nonsingular transformation of $\underset{\sim}{X}$, and therefore has a multivariate normal distribution with mean vector

$$\begin{aligned}
E\begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \end{pmatrix} &= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} E(\underset{\sim}{X}_1) \\ E(\underset{\sim}{X}_2) \end{pmatrix} \\
&= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix} \\
&= \begin{pmatrix} \underset{\sim}{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}_2 \\ \underset{\sim}{\mu}_2 \end{pmatrix} \\
&= \begin{pmatrix} \underset{\sim}{\nu}_1 \\ \underset{\sim}{\nu}_2 \end{pmatrix} = \underset{\sim}{\nu}, \quad \text{say.}
\end{aligned}$$

First find the covariance matrix of $\underset{\sim}{Y}_1$ as follows:

$$
\begin{aligned}
Cov(\underset{\sim}{Y}_1) &= E[(\underset{\sim}{Y}_1 - \underset{\sim}{\nu}_1)(\underset{\sim}{Y}_1 - \underset{\sim}{\nu}_1)'] \\
&= E\left[\left\{(\underset{\sim}{X}_1 - \underset{\sim}{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)\right\}\left\{(\underset{\sim}{X}_1 - \underset{\sim}{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)\right\}'\right] \\
&= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} \\
&= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
&= \Sigma_{11.2}, \text{ say.}
\end{aligned}
$$

The covariance matrix of $\underset{\sim}{Y}$ is

$$
\begin{aligned}
Cov(\underset{\sim}{Y}) &= E[(\underset{\sim}{Y} - \underset{\sim}{\nu})(\underset{\sim}{Y} - \underset{\sim}{\nu})'] \\
&= \begin{pmatrix} E[(\underset{\sim}{Y}_1 - \underset{\sim}{\nu}_1)(\underset{\sim}{Y}_1 - \underset{\sim}{\nu}_1)'] & E[(\underset{\sim}{Y}_1 - \underset{\sim}{\nu}_1)(\underset{\sim}{Y}_2 - \underset{\sim}{\nu}_2)'] \\ E[(\underset{\sim}{Y}_2 - \underset{\sim}{\nu}_2)(\underset{\sim}{Y}_1 - \underset{\sim}{\nu}_1)'] & E[(\underset{\sim}{Y}_2 - \underset{\sim}{\nu}_2)(\underset{\sim}{Y}_2 - \underset{\sim}{\nu}_2)'] \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}.
\end{aligned}
$$

Thus $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ are independently distributed. The marginal distribution of $\underset{\sim}{Y}_2 = \underset{\sim}{X}_2$ is $N_{p-q}(\underset{\sim}{\mu}_2, \Sigma_{22})$.

It may be noted that $E(\underset{\sim}{Y}_1) = \underset{\sim}{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}_2$. Since $\underset{\sim}{Y}_1$ is a nonsingular linear transformation of a multivariate normally distributed random vector, we have

$$\underset{\sim}{Y}_1 \sim N_q\left(\underset{\sim}{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}_2,\ \Sigma_{11.2}\right).$$

It can be observed that by considering $\underset{\sim}{Y}_2 = \underset{\sim}{X}_2$, we have derived the marginal distribution of $\underset{\sim}{X}_2$. Similarly, in order to derive the marginal distribution of $\underset{\sim}{X}_1$, consider $\underset{\sim}{Y}_1^*$ and $\underset{\sim}{Y}_2^*$ by interchanging the subscripts 1 and 2 and proceeding in the same way by considering

$$\underset{\sim}{Y}_1^* = \underset{\sim}{X}_1, \qquad \underset{\sim}{Y}_2^* = \underset{\sim}{X}_2 + B^*\underset{\sim}{X}_1.$$

Then we find $B^*$ such that $\underset{\sim}{Y}_1^*$ and $\underset{\sim}{Y}_2^*$ are uncorrelated. Construct $\underset{\sim}{Y}_1^*$ and $\underset{\sim}{Y}_2^*$ and proceed further on the same lines as in the case of $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$.

It is also important to note that the two subvectors $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ contain those variables out of $X_1, X_2, \ldots, X_p$ which the experimenter or the statistician decides.

**Theorem 2:** If $X_1, X_2, \ldots, X_p$ have a joint normal distribution, a necessary and sufficient condition for one subset of the random variables and the subset consisting of the remaining variables to be independent is that each covariance of a variable from one set and a variable from the other set is zero.

**Proof:** Let $X_i$ and $X_j$ be two random variables from the first and second sets, respectively. Then their covariance is

$$\begin{aligned}\sigma_{ij} &= E\big[(X_i - \underset{\sim}{\mu}_i)(X_j - \underset{\sim}{\mu}_j)\big] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f_{X_1, X_2, \ldots, X_p}(x_1, \ldots, x_p)\, dx_1 \cdots dx_p.\end{aligned}$$

Since the joint density factors as

$$f_{X_1, X_2, \ldots, X_p}(x_1, \ldots, x_p) = f_{X_1, X_2, \ldots, X_q}(x_1, \ldots, x_q) \cdot f_{X_{q+1}, X_{q+2}, \ldots, X_p}(x_{q+1}, \ldots, x_p),$$

we have

$$\begin{aligned}\sigma_{ij} = &\left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_i) f_{X_1, X_2, \ldots, X_q}(x_1, \ldots, x_q)\, dx_1 \cdots dx_q\right] \\ &\times \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_j - \mu_j) f_{X_{q+1}, X_{q+2}, \ldots, X_p}(x_{q+1}, \ldots, x_p)\, dx_{q+1} \cdots dx_p\right] \\ = &\, 0\end{aligned}$$

as each of the integrals is zero.

On the other hand,

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j.$$

Since $\Sigma$ is nonsingular, $\sigma_i \neq 0$ and $\sigma_j \neq 0$. Therefore, $\sigma_{ij} = 0$ only when $\rho_{ij} = 0$. Thus, if one set of variates is uncorrelated with the remaining variates, the two sets are independent.

It is important to understand that the implication of independence from a lack of correlation depends on the assumption of normal distribution, but the converse is always true.

## Conditional Distributions

The conditional distributions are of a simple nature because the means depend only linearly on the variates held fixed, and variance and covariance do not depend at all on the values of the fixed variates.

To formulate the conditional distribution, we assume $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ and partition $\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2)'$ where $\underset{\sim}{X}_1$ is a $(q \times 1)$ vector of random variables $X_1, X_2, \ldots, X_q$ and $\underset{\sim}{X}_2$ is a $(p-q) \times 1$ vector of random variables $X_{q+1}, X_{q+2}, \ldots, X_p$.

Consider a nonsingular transformation

$$\underset{\sim}{Y}_1 = \underset{\sim}{X}_1 + B\underset{\sim}{X}_2$$
$$\underset{\sim}{Y}_2 = \underset{\sim}{X}_2$$

and choose $B$ such that $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ are uncorrelated.

This exercise has already been done when the marginal distributions were found when $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are partitioned such that they were not independent.

We found $B = -\Sigma_{12}\Sigma_{22}^{-1}$, and $\underset{\sim}{Y}_1 = \underset{\sim}{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{X}_2$, which follows $\underset{\sim}{Y}_1 \sim N_q\left(\underset{\sim}{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}_2, \Sigma_{11.2}\right)$ where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, and $\underset{\sim}{Y}_2 \sim N_{p-q}(\underset{\sim}{\mu}_2, \Sigma_{22})$.

We also proved that $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2 = \underset{\sim}{X}_2$ are independently distributed. So the joint density of $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ is

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}_1, \underset{\sim}{y}_2) = f_{\underset{\sim}{Y}_1}(\underset{\sim}{y}_1) \cdot f_{\underset{\sim}{Y}_2}(\underset{\sim}{y}_2).$$

Also, $f_{\underset{\sim}{X}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2) \sim N_p(\underset{\sim}{\mu}, \Sigma)$.

Suppose we want to find the conditional distribution of $\underset{\sim}{X}_1 | \underset{\sim}{X}_2 = \underset{\sim}{x}_2$.

Then we know

$$f_{\underset{\sim}{X}_1|\underset{\sim}{X}_2}(\underset{\sim}{x}_1|\underset{\sim}{x}_2) = \frac{f_{\underset{\sim}{X}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)}{f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2)}.$$

So we find $f_{\underset{\sim}{X}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ from $f_{\underset{\sim}{Y}}(\underset{\sim}{y}_1, \underset{\sim}{y}_2)$ as

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2) = f_{\underset{\sim}{Y}}(\underset{\sim}{y}_1, \underset{\sim}{y}_2)|J|$$

where the Jacobian of transformation is $|J| = 1$. Then

$$\begin{aligned} f_{\underset{\sim}{X}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2) &= f_{\underset{\sim}{Y}}(\underset{\sim}{y}_1, \underset{\sim}{y}_2) \cdot 1 \\ &= f_{\underset{\sim}{Y}_1}(\underset{\sim}{y}_1)\, f_{\underset{\sim}{Y}_2}(\underset{\sim}{y}_2) \end{aligned}$$

$$f_{\underset{\sim}{Y}_2}(\underset{\sim}{y}_2) = f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2).$$

Then,

$$\begin{aligned} f(\underset{\sim}{X}_1|\underset{\sim}{X}_2 = \underset{\sim}{x}_2) &= \frac{f_{\underset{\sim}{X}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)}{f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2)} \\ &= \frac{f_{\underset{\sim}{Y}_1}(\underset{\sim}{y}_1)\, f_{\underset{\sim}{Y}_2}(\underset{\sim}{y}_2)}{f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2)} \\ &= f_{\underset{\sim}{Y}_1}(\underset{\sim}{y}_1) \end{aligned}$$

where

$$\begin{aligned} f_{\underset{\sim}{Y}_1}(\underset{\sim}{y}_1) = & \frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}} \\ & \times \exp\left[-\frac{1}{2}\left\{(\underset{\sim}{x}_1 - \underset{\sim}{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)\right\}' \Sigma_{11.2}^{-1} \left\{(\underset{\sim}{x}_1 - \underset{\sim}{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)\right\}\right]. \end{aligned}$$

Thus,

$$f(\underset{\sim}{X}_1|\underset{\sim}{X}_2 = \underset{\sim}{x}_2) \sim N_q\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2),\ \Sigma_{11.2}\right),$$

where $\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)$ is the conditional mean vector and $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is the conditional covariance matrix of conditional random vector $\underset{\sim}{X}_1|\underset{\sim}{X}_2 = \underset{\sim}{x}_2$.

Similarly, $f(\underset{\sim}{X}_2|\underset{\sim}{X}_1 = \underset{\sim}{x}_1)$ will also follow a $(p-q)$ variate multivariate normal distribution. The conditional mean vector and conditional covariance matrix of $\underset{\sim}{X}_2|\underset{\sim}{X}_1 = \underset{\sim}{x}_1$ can be easily found as $\underset{\sim}{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\underset{\sim}{x}_1 - \underset{\sim}{\mu}_1)$ and $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, respectively.

One may note that $E(\underset{\sim}{X}_1|\underset{\sim}{X}_2 = \underset{\sim}{x}_2)$ is a linear function of $\underset{\sim}{x}_2$ and $Cov(\underset{\sim}{X}_1|\underset{\sim}{X}_2 = \underset{\sim}{x}_2)$ does not depend on $\underset{\sim}{x}_2$.

## Regression Function

The expectation of conditional distribution is the regression function, i.e., $E(\underset{\sim}{X}_1|\underset{\sim}{X}_2 = \underset{\sim}{x}_2) = \underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)$ is the regression function. Here $B = \Sigma_{12}\Sigma_{22}^{-1}$ is the matrix of regression coefficients of $\underset{\sim}{X}_1$ on $\underset{\sim}{x}_2$. An element in $i^{th}$ row and $(k-q)^{th}$ column of $B$ is denoted by $\beta_{ik.q+1\ldots,k-1,k+1,\ldots,p}$.

The partial covariance between $X_i$ and $X_j$ holding $X_{q+1},\ldots,X_p$ fixed, is the $(i,j)^{th}$ element of $\Sigma_{11.2}$ and is denoted as $\sigma_{ij.q+1,\ldots,p}$. When $i=j$, then $i^{th}$ diagonal element of $\Sigma_{11.2}$ is the partial variance and denoted as $\sigma_{ii.q+1,\ldots,p}$. The partial correlation coefficient between $X_i$ and $X_j$ holding $X_{q+1},\ldots,X_p$ fixed is given as

$$\rho_{ij.q+1,\ldots,p} = \frac{\sigma_{ij.q+1,\ldots,p}}{\sqrt{\sigma_{ii.q+1,\ldots,p}.\sigma_{jj.q+1,\ldots,p}}}.$$

The vectors of residuals of $\underset{\sim}{X}_1$ from its regression on $\underset{\sim}{X}_2$ is given by

$$\underset{\sim}{X}_{1.2} = (\underset{\sim}{X}_1 - \underset{\sim}{\mu}_1) - \underset{\sim}{B}(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2).$$

The maximum correlation between $X_i$ and the linear combination of the random variable in $\underset{\sim}{X}_2$, i.e., $\alpha'\underset{\sim}{X}_2$ is called as the multiple correlation coefficient between $X_i$ and $\underset{\sim}{X}_2$. It is denoted as

$$\begin{aligned}
\bar{R}_{i.q+1,\ldots,p} &= \frac{E(\underset{\sim}{\beta}_i'(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)(X_i - \mu_i))}{\sqrt{\sigma_{ii}}\sqrt{E(\underset{\sim}{\beta}_{(i)}'(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)'\underset{\sim}{\beta}_{(i)})}} \\
&= \frac{\underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}\underset{\sim}{\sigma}_{(i)}}{\sqrt{\sigma_{ii}}\sqrt{\underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}\underset{\sim}{\sigma}_{(i)}}} \\
&= \sqrt{\frac{\underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}\underset{\sim}{\sigma}_{(i)}}{\sigma_{ii}}}
\end{aligned}$$

where $\underset{\sim}{\sigma}_{(i)}'$ is the $i^{th}$ row of $\Sigma_{12}$ and $\underset{\sim}{\beta}_{(i)}' = \underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}$ is the $i^{th}$ row of $B$ matrix. Then we find

$$\begin{aligned}
1 - R^2_{i.q+1,\ldots,p} &= \frac{\sigma_{ii} - \underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}\underset{\sim}{\sigma}_{(i)}}{\sigma_{ii}} \\
&= \frac{|\Sigma_i|}{\sigma_{ii}|\Sigma_{22}|}
\end{aligned}$$

where

$$\Sigma_i = \begin{pmatrix} \sigma_{ii} & \underset{\sim}{\sigma}'_{(i)} \\ \underset{\sim}{\sigma}_{(i)} & \Sigma_{22} \end{pmatrix}.$$

## Estimation of Mean Vector and Covariance Matrix

The multivariate normal distribution is specified by the mean vector $\underset{\sim}{\mu}$ and covariance matrix $\Sigma$. We consider the estimation of $\underset{\sim}{\mu}$ and $\Sigma$ on the basis of a random sample. It may be recalled that in the univariate normal distribution $N(\mu, \sigma^2)$, the maximum likelihood estimators of $\mu$ and $\sigma^2$ are the sample arithmetic mean and sample variance (the sum of squares of deviation of observations from the sample mean divided by the number of observations). Now our aim is to find the maximum likelihood estimators of $\underset{\sim}{\mu}$ and $\Sigma$ from $N_p(\underset{\sim}{\mu}, \Sigma)$.

First, we consider some basic results which will help derive the maximum likelihood estimators of $\underset{\sim}{\mu}$ and $\Sigma$.

**Lemma 3:** Let $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ be $N$ vectors, each of $p$ components, and let $\bar{\underset{\sim}{x}} = (\bar{\underset{\sim}{x}}_1, \bar{\underset{\sim}{x}}_2, \ldots, \bar{\underset{\sim}{x}}_p)'$. Then for any vector $\underset{\sim}{b}$,

$$\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{b})(\underset{\sim}{x}_\alpha - \underset{\sim}{b})' = \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' + N(\bar{\underset{\sim}{x}} - \underset{\sim}{b})(\bar{\underset{\sim}{x}} - \underset{\sim}{b})'.$$

**Proof:** Consider

$$\begin{aligned}
& \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{b})(\underset{\sim}{x}_\alpha - \underset{\sim}{b})' \\
&= \sum_{\alpha=1}^{N} \left[(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) + (\bar{\underset{\sim}{x}} - \underset{\sim}{b})\right]\left[(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) + (\bar{\underset{\sim}{x}} - \underset{\sim}{b})\right]' \\
&= \sum_{\alpha=1}^{N} \left[(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' + (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}} - \underset{\sim}{b})' + (\bar{\underset{\sim}{x}} - \underset{\sim}{b})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' + (\bar{\underset{\sim}{x}} - \underset{\sim}{b})(\bar{\underset{\sim}{x}} - \underset{\sim}{b})'\right] \\
&= \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' + \left[\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})\right](\bar{\underset{\sim}{x}} - \underset{\sim}{b})' + (\bar{\underset{\sim}{x}} - \underset{\sim}{b})\left[\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})\right]' + N(\bar{\underset{\sim}{x}} - \underset{\sim}{b})(\bar{\underset{\sim}{x}} - \underset{\sim}{b})' \\
&= \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' + 0 + 0 + N(\bar{\underset{\sim}{x}} - \underset{\sim}{b})(\bar{\underset{\sim}{x}} - \underset{\sim}{b})' \qquad \left(\text{since} \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) = 0\right) \\
&= A + N(\bar{\underset{\sim}{x}} - \underset{\sim}{b})(\bar{\underset{\sim}{x}} - \underset{\sim}{b})'
\end{aligned}$$

where $A = \sum_{\alpha=1}^{N}(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})'$.

**Result 1:** An example for understanding the operations over a lower triangular matrix is as follows:

Let $T = \begin{bmatrix} a & 0 & 0 \\ b & d & 0 \\ c & e & f \end{bmatrix}$ be a lower triangular matrix.

Then, the square of the determinant is

$$\begin{aligned} |T|^2 &= |T| \cdot |T| \\ &= \begin{vmatrix} a^2 & 0 & 0 \\ ab+bd & d^2 & 0 \\ ac+be+cf & de+ef & f^2 \end{vmatrix} \\ &= a^2d^2f^2. \end{aligned}$$

Note that $|T|^2$ is the product of the squares of the diagonal elements of $T$. Then

$$\log |T|^2 = \log a^2 + \log d^2 + \log f^2$$

and

$$\begin{aligned} TT' &= \begin{bmatrix} a & 0 & 0 \\ b & d & 0 \\ c & e & f \end{bmatrix} \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix} \\ &= \begin{bmatrix} a^2 & ab & ac \\ ab & b^2+d^2 & bc+de \\ ac & bc+de & c^2+e^2+f^2 \end{bmatrix}. \end{aligned}$$

Then the trace is

$$tr(TT') = a^2 + b^2 + c^2 + d^2 + e^2 + f^2$$

which is the sum of squares of all the elements of $T$.

**Lemma 4:** If $D$ is a positive definite matrix of order $p \times p$, the maximum of

$$f(G) = -N \log |G| - \text{tr } G^{-1}D$$

with respect to a positive definite matrix $G$ exists and occurs at

$$G = \frac{1}{N}D$$

and has the value

$$f\left(\frac{1}{N}D\right) = pN\log N - N\log|D| - pN.$$

**Proof:** Since $D$ is a positive definite matrix, so there exists $E$ such that $D = EE'$ and $E'G^{-1}E = H$. Then

$$\begin{aligned}
& (E'G^{-1}E)^{-1} = H^{-1} \\
\text{or} \quad & E^{-1}GE'^{-1} = H^{-1} \\
\text{or} \quad & EE^{-1}GE'^{-1}E' = EH^{-1}E' \\
\text{or} \quad & G = EH^{-1}E' \ ,
\end{aligned}$$

$$\begin{aligned}
|G| &= |E||H^{-1}||E'| \\
&= |H^{-1}||EE'| \\
&= |H^{-1}||D| \\
&= \frac{|D|}{|H|}
\end{aligned}$$

and

$$\begin{aligned}
tr(G^{-1}D) &= tr(G^{-1}EE') \\
&= tr(E'G^{-1}E) \\
&= \text{tr } H.
\end{aligned}$$

The function to be maximized with respect to a positive definite matrix $H$ is

$$\begin{aligned}
f(G) &= -N\log|G| - \text{tr } G^{-1}D \\
&= -N\log\frac{|D|}{|H|} - \text{tr } H \\
&= -N\log|D| + N\log|H| - \text{tr } H.
\end{aligned}$$

Let $H = TT'$ where $T$ is a lower triangular matrix. This is due to the Cholesky decomposition theorem, which states that "If $A$ is a positive definite matrix, there exists a unique lower triangular matrix $T$ $(t_{ij} = 0, i < j)$ with positive diagonal elements such that $A = TT'$."

So we can write

$$\begin{aligned} f &= -N \log |D| + N \log |T|^2 - \text{tr } TT' \\ &= -N \log |D| + \sum_{i=1}^{p} N \log t_{ij}^2 - \left( \sum_{i=1}^{p} t_{ii}^2 + \sum_{i} \sum_{j(i<j)} t_{ij}^2 \right) \\ &= -N \log |D| + \sum_{i=1}^{p} (N \log t_{ii}^2 - t_{ii}^2) - \sum_{i} \sum_{j(i<j)} t_{ij}^2. \end{aligned}$$

A clarification to derive these steps is illustrated in the example in Result 1.

Using the principle of maxima/minima, we find that

$$\frac{\partial f}{\partial t_{ii}^2} = \frac{N}{t_{ii}^2} - 1 = 0$$

which gives $N = t_{ii}^2$ ; $i = 1, 2, \ldots p$ and

$$\frac{\partial^2 f}{\partial t_{ii}^4} = -\frac{N}{t_{ii}^4} = -\frac{1}{t_{ii}^2}$$

at $N = t_{ii}^2$ implying that the function is maximized at $N = t_{ii}^2$ . Combine all the values into a matrix. Thus, the maximum occurs at $H = NI$.

Then

$$G = EH^{-1}E' = \frac{1}{N}EE' = \frac{1}{N}D.$$

Further

$$f(G) = -N \log |G| - \text{tr } G^{-1}D,$$

so at $G = \frac{1}{N}D$,

$$\begin{aligned} f(G) = f\left(\frac{1}{N}D\right) &= -N \log \frac{|D|}{|N|} - \text{tr } \frac{D^{-1}}{N^{-1}} \cdot D \\ &= -N \log \frac{|D|}{N^p} - N \text{ tr } I_p \\ &= -N \log |D| + pN \log N - Np. \end{aligned}$$

Hence proved.

## Maximum Likelihood Estimation of $\underset{\sim}{\mu}$ and $\Sigma$

Let $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ be a sample of observations on $\underset{\sim}{X} \sim \mathcal{N}_p(\underset{\sim}{\mu}, \Sigma)$ where $N > p$. It is important to note that the sample size is $N$ (and not $n$, the notation $n$ will be used later).

Let the sample mean vector be

$$\bar{\underset{\sim}{x}} = \frac{1}{N}\sum_{\alpha=1}^{N} \underset{\sim}{x}_\alpha = \left(\frac{1}{N}\sum_{\alpha=1}^{N} x_{1\alpha}, \frac{1}{N}\sum_{\alpha=1}^{N} x_{2\alpha}, \ldots, \frac{1}{N}\sum_{\alpha=1}^{N} x_{p\alpha}\right)' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)'$$

where $\underset{\sim}{x}_\alpha = (x_{1\alpha}, x_{2\alpha}, \ldots, x_{p\alpha})'$ and $\bar{\underset{\sim}{x}}_i = \frac{1}{N}\sum_{\alpha=1}^{N} x_{i\alpha}$.

Let

$$A = \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' = \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) \qquad i, j = 1, 2, \ldots, p$$

The likelihood function of $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ is

$$L = \frac{1}{(2\pi)^{\frac{Np}{2}} |\Sigma|^{\frac{N}{2}}} \exp\left(-\frac{1}{2}\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})\right)$$

$$\log L = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu}).$$

We simplify the term inside the exponent of the likelihood function by using the result for two matrices $C$ and $D$,

$$tr(CD) = tr(DC).$$

Note that, $\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})$ is a scalar quantity. Consider

$$\begin{aligned}
\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu}) &= tr\sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})\Sigma^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})' \\
&= tr\sum_{\alpha=1}^{N} \Sigma^{-1}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})' \qquad (\text{using} \quad trCD = trDC) \\
&= tr\Sigma^{-1}\left[A + N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'\right] \quad \left(\text{using Lemma 3 with } \underset{\sim}{b} = \underset{\sim}{\mu}\right) \\
&= tr\Sigma^{-1}A + tr\Sigma^{-1}N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' \\
&= tr\Sigma^{-1}A + N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}).
\end{aligned}$$

The log likelihood can be written as

$$\begin{aligned}\log L &= -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}tr\Sigma^{-1}A - \frac{N}{2}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) \\ &= -\frac{Np}{2}\log(2\pi) - \text{term (I)} - \text{term (II)} - \text{term (III)} \qquad \text{(say)}\end{aligned}$$

where term (I) $= \frac{N}{2}\log|\Sigma|$, term (II) $= \frac{1}{2}tr\Sigma^{-1}A$ and term (III) $= \frac{N}{2}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})$.

We consider now the optimization of terms (I) , (II) and (III) separately.

Consider the minimization of term (III) $= (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})$.

Since $\Sigma$ is a positive definite matrix, so $\Sigma^{-1}$ is also a positive definite matrix. Hence $N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) > 0$ and is a positive definite quadratic term and $N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) = 0$ if and only if $\underset{\sim}{\mu} = \underset{\sim}{\bar{x}}$.

Next we consider the maximization of terms (I) and (II). We use the Lemma 4 and by substituting $G = \Sigma$ and $D = A$, we get

$$f(G) = f(\Sigma) = -N\log|\Sigma| - tr\Sigma^{-1}A$$

whose maximum occurs at $\left(G = \frac{1}{N}D, \text{i.e.,}\right)$

$$\Sigma = \frac{1}{N}A = \frac{1}{N}\sum_{\alpha=1}^{N}(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})'.$$

The log likelihood function

$$\begin{aligned}\log L &= -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}tr\Sigma^{-1}A - \frac{N}{2}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}) \\ &= \text{constant} - \frac{1}{2}f(\Sigma) - \frac{N}{2}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})\end{aligned}$$

which is maximized at

$$\underset{\sim}{\hat{\mu}} = \underset{\sim}{\bar{x}}$$

and

$$\hat{\Sigma} = \frac{1}{N}A = \frac{1}{N}\sum_{\alpha=1}^{N}(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})'.$$

Thus $\underset{\sim}{\hat{\mu}}$ and $\hat{\Sigma}$ are the maximum likelihood estimators of $\underset{\sim}{\mu}$ and $\Sigma$, respectively.

**Result:** $E(S) = \Sigma$ where $S = \frac{A}{N-1}$.

**Proof:** Consider the expectations of diagonal and off-diagonal elements of $S$, or equivalently the matrix $A$.

Consider the $i$-th diagonal element of $A$, i.e., $a_{ii}$ as follows:

$$\begin{aligned} a_{ii} &= \sum_{j=1}^{N}(x_{ji} - \bar{x}_i)^2 \\ &= \sum_{j=1}^{N}[(x_{ji} - \mu_i) - (\bar{x}_i - \mu_i)]^2 \\ &= \sum_{j=1}^{N}(x_{ji} - \mu_i)^2 - N(\bar{x}_i - \mu_i)^2. \end{aligned}$$

Taking expectations on both sides, we get

$$\begin{aligned} E(a_{ii}) &= N\sigma_{ii} - NVar(\bar{x}_i) \\ &= N\sigma_{ii} - N\left(\frac{\sigma_{ii}}{N}\right) \\ &= (N-1)\sigma_{ii}. \end{aligned}$$

Now consider

$$\begin{aligned} &\sum_{\alpha=1}^{N}[(x_{\alpha i} - \bar{x}_i) + (x_{\alpha j} - \bar{x}_j)]^2 \\ &= \sum_{\alpha=1}^{N}(x_{\alpha i} - \bar{x}_i)^2 + \sum_{\alpha=1}^{N}(x_{\alpha j} - \bar{x}_j)^2 + 2\sum_{\alpha=1}^{N}(x_{\alpha i} - \bar{x}_i)(x_{\alpha j} - \bar{x}_j) \\ &= a_{ii} + a_{jj} + 2a_{ij} \\ &= (N-1)(s_{ii} + s_{jj} + 2s_{ij}). \end{aligned}$$

Next consider

$$\begin{aligned} Var(x_{\alpha i} + x_{\alpha j}) &= Var(X_{\alpha i}) + Var(X_{\alpha j}) + 2Cov(x_{\alpha i}, x_{\alpha j}) \\ &= \sigma_{ii} + \sigma_{jj} + 2\sigma_{ij} \end{aligned}$$

and

$$E(s_{ii} + s_{jj} + 2s_{ij}) = \sigma_{ii} + \sigma_{jj} + 2\sigma_{ij}.$$

Since we already proved $E(s_{ii}) = \sigma_{ii}$ and $E(s_{jj}) = \sigma_{jj}$, it follows that

$$E(s_{ij}) = \sigma_{ij}.$$

Finally, consider $E(S)$ as

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{pmatrix}$$

$$\begin{aligned} E(S) &= \begin{pmatrix} E(s_{11}) & E(s_{12}) & \dots & E(s_{1p}) \\ E(s_{21}) & E(s_{22}) & \dots & E(s_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(s_{p1}) & E(s_{p2}) & \dots & E(s_{pp}) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \\ &= \Sigma. \end{aligned}$$

Thus, note that $\hat{\Sigma} = S = \frac{1}{N-1}A$ is an unbiased estimator of $\Sigma$.

Alternatively, one can also derive $E(s_{ij}) = \sigma_{ij}$ by considering a random variable $U = X + Y$.

Based on samples $u_1, u_2, \dots, u_N$, we can express the sample variance $s_u^2$ and then

$$E(s_u^2) = \frac{1}{N-1} E\left[\sum_{i=1}^{N} (u_i - \bar{u})^2\right] = \sigma_u^2.$$

Substituting $u_i = x_i + y_i$, we can write $s_u^2$ as

$$\frac{1}{N-1} E\left[\sum_{i=1}^{N} \{(x_i - \bar{x}) + (y_i - \bar{y})\}^2\right] = Var(X+Y)$$

$$\frac{1}{N-1} E\left[\sum_{i=1}^{N} (x_i - \bar{x})^2 + \sum_{i=1}^{N} (y_i - \bar{y})^2 + 2\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})\right] = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}$$

This simplifies to

$$E(s_x^2) + E(s_y^2) + 2E(s_{xy}) = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}.$$

Since $E(s_x^2) = \sigma_x^2$ and $E(s_y^2) = \sigma_y^2$, this implies

$$E(s_{xy}) = \sigma_{xy}.$$

Then the result $E(S) = \Sigma$ follows.

## Distribution of Sample Mean Vector

It may be recalled that in the univariate normal distribution $N(\mu, \sigma^2)$, the maximum likelihood estimators are $\bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$. The distribution of $\bar{x}$ is $N(\mu, \sigma^2/n)$. Further, $\bar{X}$ and $\hat{\sigma}^2$ are independently distributed.

Now we aim to derive similar results for the case of a multivariate normal distribution. We will use the concept of an orthogonal matrix.

An orthogonal matrix $C$ satisfies $CC' = I$. To understand the interrelationship among the elements of such matrices, consider a simple example.

Let $C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$ be an orthogonal matrix such that $CC' = I$ and we can write

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Multiplying the matrices, we get

$$\begin{pmatrix} c_{11}^2 + c_{12}^2 & c_{11}c_{21} + c_{12}c_{22} \\ c_{21}c_{11} + c_{22}c_{12} & c_{21}^2 + c_{22}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} \sum_{k=1}^2 c_{1k}^2 & \sum_{k=1}^2 c_{1k}c_{2k} \\ \sum_{k=1}^2 c_{2k}c_{1k} & \sum_{k=1}^2 c_{2k}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The element-wise comparison of the left and right-hand sides shows that

- $\sum_{j=1}^2 c_{\alpha j}^2 = 1$ for $\alpha = 1, 2$
- $\sum_{k=1}^2 c_{\alpha k}c_{\beta k} = 0$ for $\alpha \neq \beta$.

Such relationships can be generalized to any orthogonal matrix $C$ of order $N \times N$ as

$$\sum_{\beta=1}^{N} c_{\alpha\beta} c_{\gamma\beta} = \begin{cases} 1 & \text{if } \alpha = \gamma \\ 0 & \text{if } \alpha \neq \gamma. \end{cases}$$

Next, we consider the following theorem involving the linear combinations of multivariate normal distributed random vectors constructed by using the elements of an orthogonal matrix.

**Theorem 5:** Suppose $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N$ are independently distributed where $\underset{\sim}{X}_\alpha \sim N_p(\underset{\sim}{\mu}_\alpha, \Sigma)$ for $\alpha = 1, 2, \ldots, N$. Let $C = ((c_{\alpha\beta}))$ be an orthogonal matrix. Then $\underset{\sim}{Y}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} \underset{\sim}{X}_\beta$ is distributed as $N_p(\underset{\sim}{\gamma}_\alpha, \Sigma)$ where $\underset{\sim}{\gamma}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} \underset{\sim}{\mu}_\beta$, and $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \ldots, \underset{\sim}{Y}_N$ are independently distributed.

**Proof:** We first show the multivariate normality of $\underset{\sim}{Y}_\alpha$'s, then find the mean vector and covariance matrix.

Note that $\{\underset{\sim}{Y}_\alpha\}$ is a set of linear combinations of $\{\underset{\sim}{X}_\beta\}$ which are following a multivariate normal distribution. Since $\underset{\sim}{X}_\alpha$ have a joint normal distribution, $\{\underset{\sim}{Y}_\alpha\}$ also have a joint normal distribution.

Now we find the mean vector:

$$E(\underset{\sim}{Y}_\alpha) = E\left[\sum_{\beta=1}^{N} c_{\alpha\beta} \underset{\sim}{X}_\beta\right] = \sum_{\beta=1}^{N} c_{\alpha\beta} \underset{\sim}{\mu}_\beta = \underset{\sim}{\gamma}_\alpha.$$

Next, we find the covariance matrix of $\underset{\sim}{Y}$s as follows:

$$\begin{aligned} Cov(\underset{\sim}{Y}_\alpha, \underset{\sim}{Y}'_\gamma) &= E[(\underset{\sim}{Y}_\alpha - \underset{\sim}{\gamma}_\alpha)(\underset{\sim}{Y}_\gamma - \underset{\sim}{\gamma}_\gamma)'] \\ &= E\left[\left\{\sum_{\beta=1}^{N} c_{\alpha\beta}(\underset{\sim}{X}_\beta - \underset{\sim}{\mu}_\beta)\right\}\left\{\sum_{\epsilon=1}^{N} c_{\gamma\epsilon}(\underset{\sim}{X}_\epsilon - \underset{\sim}{\mu}_\epsilon)\right\}'\right] \\ &= \sum_{\beta=1}^{N}\sum_{\epsilon=1}^{N} c_{\alpha\beta} c_{\gamma\epsilon} E\left[(\underset{\sim}{X}_\beta - \underset{\sim}{\mu}_\beta)(\underset{\sim}{X}_\epsilon - \underset{\sim}{\mu}_\epsilon)'\right]. \end{aligned}$$

Note that $E[(\underset{\sim}{X}_\beta - \underset{\sim}{\mu}_\beta)(\underset{\sim}{X}_\epsilon - \underset{\sim}{\mu}_\epsilon)'] = \delta_{\beta\epsilon}\Sigma$, where $\delta_{\beta\epsilon}$ is the Kronecker delta defined as

$$\delta_{\beta\epsilon} = \begin{cases} 1 & \text{if } \beta = \epsilon \\ 0 & \text{if } \beta \neq \epsilon. \end{cases}$$

Thus

$$\begin{aligned}
Cov(\underset{\sim}{Y}_\alpha, \underset{\sim}{Y}'_\gamma) &= \sum_{\beta=1}^{N}\sum_{\epsilon=1}^{N} c_{\alpha\beta}c_{\gamma\epsilon}\delta_{\beta\epsilon}\Sigma \\
&= \sum_{\beta=1}^{N} c_{\alpha\beta}\left(\sum_{\epsilon=1}^{N} c_{\gamma\epsilon}\delta_{\beta\epsilon}\right)\Sigma \\
&= \sum_{\beta=1}^{N} c_{\alpha\beta}c_{\gamma\beta}\Sigma \\
&= \delta_{\alpha\gamma}\Sigma
\end{aligned}$$

where

$$\delta_{\alpha\gamma} = \begin{cases} 1 & \text{if } \alpha = \gamma \\ 0 & \text{if } \alpha \neq \gamma \end{cases}$$

is the Kronecker delta and

$$\sum_{\beta=1}^{N} c_{\alpha\beta}c_{\gamma\beta} = \begin{cases} 1 & \text{if } \alpha = \gamma \\ 0 & \text{if } \alpha \neq \gamma, \end{cases}$$

using the property of an orthogonal matrix.

This shows that $\underset{\sim}{Y}_\alpha$ is independently distributed of $\underset{\sim}{Y}_\gamma$ (for $\alpha \neq \gamma$) and has a covariance matrix $\Sigma$.

We consider the next general lemma, which will also be used later.

**Lemma 6:** If $C = ((c_{\alpha\beta}))$ is an orthogonal matrix of order $N \times N$, then

$$\sum_{\alpha=1}^{N} \underset{\sim}{X}_\alpha \underset{\sim}{X}'_\alpha = \sum_{\alpha=1}^{N} \underset{\sim}{Y}_\alpha \underset{\sim}{Y}'_\alpha$$

where $\underset{\sim}{Y}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}\underset{\sim}{X}_\beta$.

**Proof:** Consider

$$\begin{aligned}\sum_{\alpha=1}^{N} \underset{\sim}{Y}_\alpha \underset{\sim}{Y}'_\alpha &= \sum_{\alpha=1}^{N}\left(\sum_{\beta=1}^{N} c_{\alpha\beta}\underset{\sim}{X}_\beta\right)\left(\sum_{\gamma=1}^{N} c_{\alpha\gamma}\underset{\sim}{X}_\gamma\right)' \\ &= \sum_{\beta=1}^{N}\sum_{\gamma=1}^{N}\left(\sum_{\alpha=1}^{N} c_{\alpha\beta}c_{\alpha\gamma}\right)\underset{\sim}{X}_\beta \underset{\sim}{X}'_\gamma \\ &= \sum_{\beta=1}^{N}\sum_{\gamma=1}^{N}\delta_{\beta\gamma}\underset{\sim}{X}_\beta \underset{\sim}{X}'_\gamma\end{aligned}$$

where, due to the orthogonal matrix $C$, we have

$$\sum_{\beta=1}^{N} c_{\alpha\beta}c_{\gamma\beta} = \begin{cases} 1 & \text{if } \alpha=\gamma \\ 0 & \text{if } \alpha\neq\gamma \end{cases}$$

and

$$\delta_{\beta\gamma} = \begin{cases} 1 & \text{if } \beta=\gamma \\ 0 & \text{if } \beta\neq\gamma \end{cases}$$

is the Kronecker delta.

So we get

$$\sum_{\alpha=1}^{N} \underset{\sim}{Y}_\alpha \underset{\sim}{Y}'_\alpha = \sum_{\beta=1}^{N}\left(\sum_{\gamma=1}^{N}\delta_{\beta\gamma}\underset{\sim}{X}_\beta \underset{\sim}{X}'_\gamma\right) = \sum_{\beta=1}^{N}\underset{\sim}{X}_\beta \underset{\sim}{X}'_\beta\ .$$

Before considering the proof of independence of $\hat{\underset{\sim}{\mu}}$ and $\hat{\Sigma}$, we consider the following result, which will help us in the proof.

**Result 2:** Let $\underset{\sim}{X}_\alpha \sim N_p(\underset{\sim}{\mu}, \Sigma)$, $\alpha = 1, 2, \ldots, N$ be independently distributed. There exists an $N \times N$ orthogonal matrix $B = ((b_{\alpha\beta}))$ whose last row has all the elements $(1/\sqrt{N}, 1/\sqrt{N}, \ldots, 1/\sqrt{N})$.

So

$$B = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1N} \\ \vdots & \vdots & \ldots & \vdots \\ b_{N-1,1} & b_{N-1,2} & \ldots & b_{N-1,N} \\ \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} & \ldots & \frac{1}{\sqrt{N}} \end{pmatrix}.$$

Let $\underset{\sim}{Z}_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta}\underset{\sim}{X}_\beta$.

Then for the $N^{th}$ vector:

$$\underset{\sim}{Z}_N = \sum_{\beta=1}^{N} b_{N\beta}\underset{\sim}{X}_\beta = \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}}\underset{\sim}{X}_\beta = \sqrt{N}\bar{\underset{\sim}{X}}.$$

Now we consider $A$ as

$$A = (N-1)S = \sum_{\alpha=1}^{N}(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}})' = \sum_{\alpha=1}^{N}\underset{\sim}{X}_\alpha\underset{\sim}{X}_\alpha' - N\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}'.$$

Using the result $\sum_{\alpha=1}^{N}\underset{\sim}{X}_\alpha\underset{\sim}{X}_\alpha' = \sum_{\alpha=1}^{N}\underset{\sim}{Z}_\alpha\underset{\sim}{Z}_\alpha'$ and defining $\underset{\sim}{Z}_N = \sqrt{N}\bar{\underset{\sim}{X}}$, we can express $A$ as

$$A = \sum_{\alpha=1}^{N}\underset{\sim}{Z}_\alpha\underset{\sim}{Z}_\alpha' - \underset{\sim}{Z}_N\underset{\sim}{Z}_N' = \sum_{\alpha=1}^{N-1}\underset{\sim}{Z}_\alpha\underset{\sim}{Z}_\alpha'$$

which no more depends upon $\underset{\sim}{Z}_N$ and note that $\underset{\sim}{Z}_N$ is a function of $\bar{\underset{\sim}{X}}$.

Note that $A$ is a function of $\underset{\sim}{Z}_1, \underset{\sim}{Z}_2, \ldots, \underset{\sim}{Z}_{N-1}$ and does not depend on $\underset{\sim}{Z}_N$. On the other hand, the sample mean vector $\bar{\underset{\sim}{X}}$ depends only upon $\underset{\sim}{Z}_N$ (since $\bar{\underset{\sim}{X}} = \frac{1}{\sqrt{N}}\underset{\sim}{Z}_N$). So we can say that the mean vector $\bar{\underset{\sim}{X}}$ is independent of $A$.

We also note that

$$\begin{aligned} E(\underset{\sim}{Z}_\alpha) &= \sum_{\beta=1}^{N} b_{\alpha\beta}E(\underset{\sim}{X}_\beta) \\ &= \sum_{\beta=1}^{N} b_{\alpha\beta}\underset{\sim}{\mu} \\ &= \sum_{\beta=1}^{N} b_{\alpha\beta}.\frac{\sqrt{N}}{\sqrt{N}}.\underset{\sim}{\mu} \\ &= \sum_{\beta=1}^{N} b_{\alpha\beta}b_{N\beta}.\sqrt{N}.\underset{\sim}{\mu} \\ &= 0 \end{aligned}$$

as $B = (b_{ij})$ is an orthogonal matrix and the elements in $N^{th}$ row are all equal (typically $1/\sqrt{N}$), the orthogonality condition implies that for $\alpha \neq N$, the row sum $\sum_\beta b_{\alpha\beta} = 0$. Thus

$$E(\underset{\sim}{Z}_\alpha) = \underset{\sim}{0} \quad \text{for } \alpha = 1, 2, \ldots, N-1.$$

Also, consider the expectation of the $N^{th}$ transformed vector $\underset{\sim}{Z}_N$ as follows:

$$\begin{aligned}
E(\underset{\sim}{Z}_N) &= \sum_{\beta=1}^{N} b_{N\beta} E(\underset{\sim}{X}_\beta) \\
&= \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}} \underset{\sim}{\mu} \\
&= \sqrt{N} \underset{\sim}{\mu}.
\end{aligned}$$

Thus $\underset{\sim}{Z}_N \sim N_p(\sqrt{N}\underset{\sim}{\mu}, \Sigma)$. Since $\underset{\sim}{\bar{X}} = \frac{\underset{\sim}{Z}_N}{\sqrt{N}}$, the distribution of the sample mean is

$$\underset{\sim}{\bar{X}} \sim N_p\left(\underset{\sim}{\mu}, \frac{\Sigma}{N}\right).$$

This can also be demonstrated using the likelihood function approach. The likelihood function of the sample $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N$ is $L(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_N)$ given as

$$L = \frac{1}{(2\pi)^{\frac{Np}{2}} |\Sigma|^{\frac{N}{2}}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu})\right].$$

Consider the exponent of the likelihood function and express it as a function of $\underset{\sim}{Z}_\alpha$. We simplify it by considering the exponent term as follows:

$$\begin{aligned}
\sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu}) &= tr\left[\Sigma^{-1} \sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu})(\underset{\sim}{X}_\alpha - \underset{\sim}{\mu})'\right] \\
&= tr\left[\Sigma^{-1} \sum_{\alpha=1}^{N} \{(\underset{\sim}{X}_\alpha - \underset{\sim}{\bar{X}}) + (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\}\{(\underset{\sim}{X}_\alpha - \underset{\sim}{\bar{X}}) + (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\}'\right] \\
&= tr\left[\Sigma^{-1} \left\{\sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \underset{\sim}{\bar{X}})(\underset{\sim}{X}_\alpha - \underset{\sim}{\bar{X}})' + N(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\right\}\right] \\
&= tr\left[\Sigma^{-1}\{A + N(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\}\right].
\end{aligned}$$

Substituting $A = \sum_{\alpha=1}^{N-1} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}'_\alpha$, we get

$$\begin{aligned}
\sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{X}_\alpha - \underset{\sim}{\mu}) &= tr\left[\Sigma^{-1} \sum_{\alpha=1}^{N-1} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}'_\alpha\right] + tr\left[\Sigma^{-1} N(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\right] \\
&= \sum_{\alpha=1}^{N-1} \underset{\sim}{Z}'_\alpha \Sigma^{-1} \underset{\sim}{Z}_\alpha + (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})' \left(\frac{\Sigma}{N}\right)^{-1} (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}).
\end{aligned}$$

Thus, the likelihood function is written as

$$L = \frac{1}{(2\pi)^{\frac{Np}{2}}|\Sigma|^{\frac{N}{2}}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N-1} \underset{\sim}{Z}'_\alpha \Sigma^{-1} \underset{\sim}{Z}_\alpha - \frac{1}{2}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\left(\frac{\Sigma}{N}\right)^{-1}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\right]$$

which is factorized into two parts as

$$\begin{aligned} L \;&= \frac{1}{(2\pi)^{\frac{Np}{2}}|\Sigma|^{\frac{N}{2}}}\left[\exp\left(-\frac{1}{2}\sum_{\alpha=1}^{N-1} \underset{\sim}{Z}'_\alpha \Sigma^{-1} \underset{\sim}{Z}_\alpha\right)\right]\left[\exp\left(-\frac{1}{2}N(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N-1} \underset{\sim}{Z}'_\alpha \Sigma^{-1} \underset{\sim}{Z}_\alpha\right] \times \exp\left[-\frac{1}{2}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\left(\frac{\Sigma}{N}\right)^{-1}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\right]. \end{aligned}$$

Thus $\underset{\sim}{\bar{X}}$ and $(\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_{N-1})$ are independently distributed.

Further, the distribution of $\underset{\sim}{\bar{X}}$ is

$$\text{constant} \times \exp\left[-\frac{1}{2}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\left(\frac{\Sigma}{N}\right)^{-1}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\right]$$

which can be written as $N_p(\underset{\sim}{\mu}, \frac{\Sigma}{N})$.

As $N\hat{\Sigma}$ is distributed as $A = \sum_{\alpha=1}^{N-1} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}'_\alpha$, where $\underset{\sim}{Z}_\alpha \sim N_p(\underset{\sim}{0}, \Sigma)$ for $\alpha = 1, 2, \ldots, N-1$ and $\underset{\sim}{Z}_1, \underset{\sim}{Z}_2, \ldots, \underset{\sim}{Z}_{N-1}$ are independent.

Moreover $\underset{\sim}{\bar{X}}$ and $A$ (or equivalently the sample covariance matrix $S$) are independently distributed.

The distribution of $A$ is the Wishart Distribution, which is discussed in the next chapter.