Property (5) **Marginal Distribution**

Let $X = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix}$ where $X_{(1)} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix}$ and $X_{(2)} = \begin{bmatrix} x_{q+1} \\ x_{q+2} \\ \vdots \\ x_p \end{bmatrix}$

also μ and Σ has partitions as

$$\mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where, Σ_{11} is the upper left hand corner submatrix of Σ of order q.

similarly Σ_{12} and Σ_{22}.

If $X \sim N_p(\mu, \Sigma)$ and $\Sigma_{12} = \Sigma_{21} = 0$

then $X_{(1)}$ and $X_{(2)}$ are independently normally distributed with mean $\mu_{(1)}$ and $\mu_{(2)}$ and covariance matrix Σ_{11} and Σ_{22} respectively

Proof: Assume that $\Sigma_{12} = \Sigma_{21} = 0$

Consider the quadratic form $Q = (x-\mu)' \Sigma^{-1} (x-\mu)$

$$Q = \left[(X_{(1)} - \mu_{(1)})' \; (X_{(2)} - \mu_{(2)})' \right] \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} X_{(1)} - \mu_{(1)} \\ X_{(2)} - \mu_{(2)} \end{bmatrix}$$

$$= \left[(X_{(1)} - \mu_{(1)})' \Sigma_{11}^{-1} \quad (X_{(2)} - \mu_{(2)})' \Sigma_{22}^{-1} \right] \begin{bmatrix} X_{(1)} - \mu_{(1)} \\ X_{(2)} - \mu_{(2)} \end{bmatrix}$$

$$= (X_{(1)} - \mu_{(1)})' \Sigma_{11}^{-1} (X_{(1)} - \mu_{(1)}) + (X_{(2)} - \mu_{(2)})' \Sigma_{22}^{-1} (X_{(2)} - \mu_{(2)})$$

$$Q = Q_1 + Q_2 \quad (\text{say})$$

∴ The p.d.f of random vector X is,

$$f_x(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)}$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} Q}$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\left[(X_{(1)} - \mu_{(1)})' \Sigma_{11}^{-1} (X_{(1)} - \mu_{(1)}) + (X_{(2)} - \mu_{(2)})' \Sigma_{22}^{-1} (X_{(2)} - \mu_{(2)})\right]}$$

$$= \frac{1}{(2\pi)^{q/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}\left[(X_{(1)} - \mu_{(1)})' \Sigma_{11}^{-1} (X_{(1)} - \mu_{(1)})\right]} \cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} |\Sigma_{22}|^{1/2}} e^{-\frac{1}{2}\left[(X_{(2)} - \mu_{(2)})' \Sigma_{22}^{-1} (X_{(2)} - \mu_{(2)})\right]}$$

$$f_x(x) = \frac{1}{(2\pi)^{q/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2} Q_1} \cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} |\Sigma_{22}|^{1/2}} e^{-\frac{1}{2} Q_2}$$

$$\Rightarrow N_p(\mu, \Sigma) = N_q(\mu_{(1)}, \Sigma_{11}) \cdot N_{p-q}(\mu_{(2)}, \Sigma_{22}) \longrightarrow ①$$

∴ The marginal density function of $X_{(1)}$ is given by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_p(\mu, \Sigma) \, dx_{q+1} \cdots dx_p$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_q(\mu_{(1)}, \Sigma_{11}) \, N_{p-q}(\mu_{(2)}, \Sigma_{22}) \, dx_{q+1} \cdots dx_p$$

$$= N_q(\mu_{(1)}, \Sigma_{11}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N_{p-q}(\mu_{(2)}, \Sigma_{22}) \, dx_{q+1} \cdots dx_p$$

$$= N_q(\mu_{(1)}, \Sigma_{11})$$

Thus the marginal distribution of $X_{(1)}$ is $N_q(\mu_{(1)}, \Sigma_{11})$

and the marginal distribution of $X_{(2)}$ is $N_{p-q}(\mu_{(2)}, \Sigma_{22})$

From (1) we observed that the joint density of X is the Product of the marginal density of $X_{(1)}$ and $X_{(2)}$.

$\therefore$ The two sets of variates are independent.

Property (6)

Conditional distribution of the multivariate normal distn.

Theorem:

Let $X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \begin{matrix} q \\ p-q \end{matrix}$ have a p variate normal

distribution with mean $\mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} \begin{matrix} q \\ p-q \end{matrix}$

and covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

Then the conditional distribution of $X^{(1)}$, given $X^{(2)}$ is q variate normal distribution

with mean $\mu_{1\cdot 2} = E\left[X^{(1)}/X^{(2)}\right] = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})$

and covariance matrix $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Proof: If $\Sigma_{12} = \Sigma_{21} = 0$ then $\Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$

The joint distribution of $X^{(1)}$ and $X^{(2)}$ is

$$f_X(x) = f(x^{(1)}, x^{(2)}) = \frac{1}{(2\pi)^{q/2} |\Sigma_{11}|^{1/2}} \exp\left[-\frac{1}{2}(x^{(1)} - \mu^{(1)})' \Sigma_{11}^{-1} (x^{(1)} - \mu^{(1)})\right] \cdot$$

$$\cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} |\Sigma_{22}|^{1/2}} \exp\left[-\frac{1}{2}(x^{(2)} - \mu^{(2)})' \Sigma_{22}^{-1} (x^{(2)} - \mu^{(2)})\right] \longrightarrow ①$$

Assume $Y^{(1)} = X^{(1)} + M X^{(2)}$ and $Y^{(2)} = X^{(2)} \longrightarrow ②$

Choosing M so that the components of $Y^{(1)}$ and $Y^{(2)}$ are uncorrelated then M must satisfy the equation

$$0 = E\left\{(Y^{(1)} - E(Y^{(1)}))\cdot(Y^{(2)} - E(Y^{(2)}))'\right\}$$

$$= E\left\{(X^{(1)} + M X^{(2)} - E(X^{(1)} + M X^{(2)}))\cdot(X^{(2)} - E(X^{(2)}))'\right\}$$

$$= E\left\{(X^{(1)} + M X^{(2)} - E(X^{(1)}) - M E(X^{(2)}))\cdot(X^{(2)} - E(X^{(2)}))'\right\}$$

$$= E\left\{\left[(X^{(1)} - E(X^{(1)})) + M(X^{(2)} - E(X^{(2)}))\right]\cdot(X^{(2)} - E(X^{(2)}))'\right\}$$

$$= E\left\{\left[(X^{(1)} - \mu^{(1)}) + M(X^{(2)} - \mu^{(2)})\right]\cdot(X^{(2)} - \mu^{(2)})'\right\}$$

$$= E\left\{(X^{(1)} - \mu^{(1)})(X^{(2)} - \mu^{(2)})' + M(X^{(2)} - \mu^{(2)})(X^{(2)} - \mu^{(2)})'\right\}$$

$$= E\left[(X^{(1)} - \mu^{(1)})(X^{(2)} - \mu^{(2)})'\right] + M \cdot E\left[(X^{(2)} - \mu^{(2)})(X^{(2)} - \mu^{(2)})'\right]$$

$$= 0 = \Sigma_{12} + M\Sigma_{22}$$

$$\Rightarrow \quad M\Sigma_{22} = -\Sigma_{12}$$

Post multiplying both sides by Σ_{22}^{-1} ~~we~~

we get : $\boxed{M = -\Sigma_{12}\Sigma_{22}^{-1}}$ $\longrightarrow$ (3)

Substitute the value of M into (2).

$$(2) \Rightarrow \quad Y^{(1)} = X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1} X^{(2)}$$

$$Y^{(2)} = X^{(2)}$$

The vector $Y = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix}$

$$= \begin{pmatrix} X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1} X^{(2)} \\ 0X^{(1)} + X^{(2)} \end{pmatrix}$$

$$Y = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

$$E(Y) = E\begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} E(X^{(1)}) \\ E(X^{(2)}) \end{pmatrix}$$

$$= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)} \\ \mu^{(2)} \end{pmatrix}$$

Note:

$$E[(X-\mu)(X-\mu)'] = \Sigma$$

~~$E[(X_i - \mu)$~~

$$E[(X^{(i)} - \mu^{(i)})(X^{(j)} - \mu^{(j)})'] = \Sigma_{ij}$$

$$\therefore \quad E(Y) = E\begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} \nu^{(1)} \\ \nu^{(2)} \end{pmatrix} = \nu \longrightarrow (4)$$

and $\quad Cov(Y) = E\left[(Y-\nu)(Y-\nu)'\right]$

$$= E\,(Y^{(i)} - \nu^{(i)})(Y^{(j)} - \nu^{(j)})'$$

$i = 1, 2 \; ; \; j = 1, 2$

$$= \begin{bmatrix} E\,(Y^{(1)} - \nu^{(1)})(Y^{(1)} - \nu^{(1)})' & E\,(Y^{(1)} - \nu^{(2)})(Y^{(2)} - \nu^{(2)})' \\ E\,(Y^{(2)} - \nu^{(2)})(Y^{(1)} - \nu^{(1)})' & E\,(Y^{(2)} - \nu^{(2)})(Y^{(2)} - \nu^{(2)})' \end{bmatrix} \longrightarrow (5)$$

But [24]

$$E\,(Y^{(1)} - \nu^{(1)})(Y^{(1)} - \nu^{(1)})' = E\left[X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)} - (\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)})\right]\left[X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)} - (\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)})\right]'$$

$$= E\left[(X^{(1)} - \mu^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})\right]\cdot\left[(X^{(1)} - \mu^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})\right]'$$

$$= \cancel{E \quad \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}}$$

$$= E\Big\{(X^{(1)} - \mu^{(1)})(X^{(1)} - \mu^{(1)})' - \Sigma_{12}\Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})(X^{(1)} - \mu^{(1)})'$$
$$- (X^{(1)} - \mu^{(1)})\Sigma_{21}\Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})' + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})(X^{(2)} - \mu^{(2)})'\Big\}$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \Sigma_{12}\Sigma_{21}\Sigma_{22}^{-1} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{22}^{-1}\Sigma_{22}$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \Sigma_{12}\Sigma_{21}\Sigma_{22}^{-1} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\boxed{E\left(y^{(1)} - \nu^{(1)}\right)\left(y^{(1)} - \nu^{(1)}\right)' = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}} \longrightarrow (6) \quad \Sigma_{11\cdot 2} \text{ (say)} \longrightarrow (6)$$

and $Y^{(1)}$ and $Y^{(2)}$ are independent events $\Rightarrow$ $\Sigma_{12} = \Sigma_{21} = 0$

$$(5) \Rightarrow Cov(Y) = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$$

{$\because$ by using (6)

$\because$ $Y^{(2)} = X^{(2)}$

Since $Y^{(1)}$ and $Y^{(2)}$ are uncorrelated normal distribution,

$$\therefore \quad Y^{(1)} \sim N\left(\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \; \Sigma_{11\cdot 2}\right)$$

$$Y^{(2)} \sim N\left(\mu^{(2)}, \; \Sigma_{22}\right)$$

$\therefore$ joint distribution of $f(y^{(1)})$ and $f(y^{(2)})$,

$$f\left(y^{(1)}, y^{(2)}\right) = f\left(y^{(1)}\right) \cdot f\left(y^{(2)}\right)$$

$$f(y^{(1)}, y^{(2)}) = \frac{1}{(2\pi)^{q/2} |\Sigma_{11\cdot 2}|^{1/2}} \exp\left[-\frac{1}{2}\left[y^{(1)} - \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\right]' \Sigma_{11\cdot 2}^{-1}\left[y^{(1)} - \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\right]\right\}$$

$$\cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} |\Sigma_{22}|^{1/2}} \exp\left\{-\frac{1}{2}\left[(y^{(2)} - \mu^{(2)})' \Sigma_{22}^{-1} (y^{(2)} - \mu^{(2)})\right]\right\} \longrightarrow (7)$$

We have,

$$y^{(1)} = x^{(1)} + M x^{(2)}$$

$$\therefore y^{(1)} = x^{(1)} - \Sigma_{12}\Sigma_{22}^{-1} x^{(2)}$$

and $y^{(2)} = x^{(2)}$

By Jacobian transformation,

$$J = \begin{vmatrix} \frac{\partial y^{(1)}}{\partial x^{(1)}} & \frac{\partial y^{(2)}}{\partial x^{(1)}} \\ \frac{\partial y^{(1)}}{\partial x^{(2)}} & \frac{\partial y^{(2)}}{\partial x^{(2)}} \end{vmatrix} = \begin{vmatrix} I & 0 \\ -\Sigma_{12}\Sigma_{22}^{-1} & I \end{vmatrix} = 1$$

Hence the joint density of $X^{(1)}$ and $X^{(2)}$ is

$$(8) \Rightarrow f(x^{(1)}, x^{(2)}) = \frac{1}{(2\pi)^{q/2} |\Sigma_{11\cdot 2}|^{1/2}} \exp\left\{-\frac{1}{2}\left[x^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}x^{(2)} - \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\right]' \Sigma_{11\cdot 2}^{-1}\left[x^{(1)} - \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\right]\right\}$$

$$\cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} |\Sigma_{22}|^{1/2}} \exp\left\{-\frac{1}{2}\left[(x^{(2)} - \mu^{(2)})' \Sigma_{22}^{-1} (x^{(2)} - \mu^{(2)})\right]\right\}$$

$$= \frac{1}{(2\pi)^{q/2} |\Sigma_{11.2}|^{1/2}} \exp\left\{-\frac{1}{2}\left[(x^{(1)}-\mu^{(1)}) - \Sigma_{11}\Sigma_{22}^{-1}(x^{(2)}-\mu^{(2)})\right]' \Sigma_{11.2}^{-1}\left[(x^{(1)}-\mu^{(1)}) + \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\right]\right.$$

$$\cdot \frac{1}{(2\pi)^{\frac{p-q}{2}} |\Sigma_{22}|^{1/2}} \exp\left\{-\frac{1}{2}(x^{(2)}-\mu^{(2)})' \Sigma_{22}^{-1}(x^{(2)}-\mu^{(2)})\right\} \longrightarrow (8)$$

$\therefore$ The Conditional density of $x^{(1)}$ given $x^{(2)}$,

$f(x^{(1)}/x^{(2)})$ is the quotient of equation (8)

$\therefore$ The quotient is

$$f(x^{(1)}/x^{(2)}) = \frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}} \exp\left\{-\frac{1}{2}\left[(x^{(1)}-\mu^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)}-\mu^{(2)})\right]' * \Sigma_{11.2}^{-1}\right[$$

$\therefore$ The density of $x^{(1)}/x^{(2)}$ is the q variate normal density with mean $\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)}-\mu^{(2)})$ and covariance matrix $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Hence the Proof.

Maximum ~~Likelifood~~ Likelihood Estimation of Parameters μ and Σ.

Theorem: If $x_1, x_2 \dots x_N$ Constitute a sample from $N(\mu, \Sigma)$ with $p < N$, the MLE of μ and Σ are

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha \quad \text{and}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \quad \text{respectively.}$$

Proof. Let $x_1, x_2 \dots x_N$ be random samples of size N from $N_p(\mu, \Sigma)$

Since $x_1\ x_2 \dots x_N$ are mutually independent and each has the distribution $N_p(\mu, \Sigma)$

The joint density of $x_1, x_2 \dots x_N$ is

$$\prod_{\alpha=1}^{N} \left\{ \frac{1}{(2\pi)^{p/2} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu) \right] \right\}$$

$\therefore$ The likelihood function

$$L = \frac{1}{(2\pi)^{\frac{NP}{2}} |\Sigma|^{N/2}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{N} (x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu) \right] \quad \to (1)$$

Since, in the likelihood function $x_1\ x_2 \dots x_N$ are fixed at the sample values ~~of~~ ~~and~~ and L is a function of μ and Σ. To emphasize that μ and Σ are variables (not Parameters) we shall denote them by μ^* and Σ^*.

Then the log of the Likelihood function is

$$\log L = -\frac{NP}{2} \log(2\pi) - \frac{N}{2} \log |\Sigma^*| - \frac{1}{2} \sum_{\alpha=1}^{N} (x_\alpha - \mu^*)' (\Sigma^*)^{-1} (x_\alpha - \mu^*) \quad \to (2)$$

Consider the exponent term of ②, we get

$$(x_\alpha-\mu^*)' \Sigma^{*^{-1}} (x_\alpha-\mu^*)$$
$$= tr\left[(x_\alpha-\mu^*)' \Sigma^{*^{-1}} (x_\alpha-\mu^*)\right]$$
$$= tr\left[\Sigma^{*^{-1}} (x_\alpha-\mu^*)(x_\alpha-\mu^*)'\right]$$

$\{\because x'Ax = tr(x'Ax) = tr(Axx')$

$tr(A) = \sum_{i=1}^{k} \lambda_i$

λ_i - eigen values$\}$

$$\sum_{\alpha=1}^{N} (x_\alpha-\mu^*)' \Sigma^{*^{-1}} (x_\alpha-\mu^*) = tr\, \Sigma^{*^{-1}} \left\{\sum_{\alpha=1}^{N} (x_\alpha-\mu^*)(x_\alpha-\mu^*)'\right\} \rightarrow ③$$

$$= tr(\Sigma^*)^{-1} \left\{\sum_{\alpha=1}^{N} (x_\alpha-\bar{x}+\bar{x}-\mu^*)(x_\alpha-\bar{x}+\bar{x}-\mu^*)'\right\}$$

Note: $\boxed{\Sigma(x_\alpha-\bar{x})=0}$

$$= tr(\Sigma^*)^{-1} \left\{\sum_{\alpha=1}^{N} (x_\alpha-\bar{x})(x_\alpha-\bar{x})' + \sum_{\alpha=1}^{N} (\bar{x}-\mu^*)\cancel{(x_\alpha-\bar{x})'}_{0} + \sum_{\alpha=1}^{N} \cancel{(x_\alpha-\bar{x})}_{0}(\bar{x}-\mu^*)' + \sum_{\alpha=1}^{N} (\bar{x}-\mu^*)(\bar{x}-\mu^*)'\right\}$$

$$= tr(\Sigma^*)^{-1} \left\{A + N(\bar{x}-\mu^*)(\bar{x}-\mu^*)'\right\} \quad (A)$$

$$= tr(\Sigma^*)^{-1} A + N(\bar{x}-\mu^*)(\Sigma^*)^{-1}(\bar{x}-\mu^*)' \rightarrow ④$$

$$② \Rightarrow \log L = -\frac{NP}{2}\log(2\pi) - \frac{1}{2}N\log|\Sigma^*| - \frac{1}{2}tr(\Sigma^*)^{-1}A - \frac{1}{2}N(\bar{x}-\mu^*)(\Sigma^*)^{-1}(\bar{x}-\mu^*)' \rightarrow ⑤$$

Since Σ^* is positive definite, $\Sigma^{*^{-1}}$ is positive definite and $N(\bar{x}-\mu^*)'(\Sigma^*)^{-1}(\bar{x}-\mu^*) \geq 0$ and ~~[illegible]~~ is 0 iff $\mu^* = \bar{x}$

MLE of $\mu = \bar{x} = \frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha = \begin{bmatrix} \frac{1}{N}(\Sigma x_{1\alpha}) \\ \frac{1}{N}(\Sigma x_{2\alpha}) \\ \vdots \\ \frac{1}{N}(\Sigma x_{p\alpha}) \end{bmatrix}$

$$= \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

i.e., $\boxed{\hat{\mu} = \bar{x}}$

To maximize the 2^{nd} and 3^{rd} terms of (5) we use the following lemma;

Lemma: If D is Positive definite of order 'p', the maximum of $f(G) = -N \log|G| - tr(G^{-1}D)$ occurs at $G = \left(\frac{1}{N}\right)D$ with respect to positive definite matrices G exists.

using this lemma, the 2^{nd} and 3^{rd} terms of (5) Considered as a function,

$$f(\Sigma^*) = -\frac{N}{2}\log|\Sigma^*| - \frac{1}{2} tr(\Sigma^*)^{-1} A$$

The maximum value of $f(\Sigma^*)$ occurs at $\Sigma^* = \left(\frac{1}{N}\right)A$

But, $\Sigma^* = $ MLE of $\Sigma = \hat{\Sigma} = \left(\frac{1}{N}\right)A$

$$\boxed{\text{MLE of } \Sigma = \hat{\Sigma} = \left(\frac{1}{N}\right)\sum_{\alpha=1}^{N}(x_\alpha - \bar{x})(x_\alpha - \bar{x})'}$$

MLE of $\mu = \hat{\mu} = \frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha$

Lemma: Let $x_1, x_2, \ldots, x_N$ be N (p-components) vectors and let $\bar{x}$ be $\frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha = \begin{bmatrix} \frac{1}{N}\sum_{\alpha=1}^{N} x_{1\alpha} \\ \vdots \\ \frac{1}{N}\sum_{\alpha=1}^{N} x_{p\alpha} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$

then for any vector b

$$\sum_{\alpha=1}^{N} (x_\alpha - b)(x_\alpha - b)' = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' + N(\bar{x}-b)(\bar{x}-b)'$$

Proof:

$$\sum_{\alpha=1}^{N} (x_\alpha - b)(x_\alpha - b)' = \sum_{\alpha=1}^{N} [(x_\alpha - \bar{x}) + (\bar{x}-b)][(x_\alpha - \bar{x}) + (\bar{x}-b)]'$$

$$= \sum \left\{ (x_\alpha-\bar{x})(x_\alpha-\bar{x})' + (x_\alpha-\bar{x})(\bar{x}-b)' + (\bar{x}-b)(x_\alpha-\bar{x})' + (\bar{x}-b)(\bar{x}-b)' \right\}$$

$$= \sum (x_\alpha-\bar{x})(x_\alpha-\bar{x})' + (\bar{x}-b)' \sum (x_\alpha-\bar{x}) + (\bar{x}-b) \sum (x_\alpha-\bar{x})' + N(\bar{x}-b)(\bar{x}-b)'$$

$$= \sum (x_\alpha-\bar{x})(x_\alpha-\bar{x})' + (\bar{x}-b)'(0) + (\bar{x}-b)(0) + N(\bar{x}-b)(\bar{x}-b)'$$

$$\sum (x_\alpha-b)(x_\alpha-b)' = \sum (x_\alpha-\bar{x})(x_\alpha-\bar{x})' + N(\bar{x}-b)(\bar{x}-b)'$$

Hence the proof.

Remark: Put $b=0$ in the above lemma we get

$$\sum x_\alpha \cdot x_\alpha' = \sum (x_\alpha-\bar{x})(x_\alpha-\bar{x})' + N\bar{x}\bar{x}'$$

$$\Rightarrow \boxed{\sum (x_\alpha-\bar{x})(x_\alpha-\bar{x})' = \sum x_\alpha x_\alpha' - N\bar{x}\bar{x}'}$$

The Distribution of the sample mean vector :

Inference Concerning the mean when the Covariance matrix Σ Known

Theorem: Suppose $X_1, X_2, \dots X_N$ are independent, where X_α is distributed according to $N(\mu_\alpha, \Sigma)$. Let $C = (C_{\alpha\beta})$ be an $N \times N$ orthogonal matrix then $Y_\alpha = \sum_{\beta=1}^{N} C_{\alpha\beta} X_\beta$ is distributed according to $N(\nu_\alpha, \Sigma)$ where $\nu_\alpha = \sum_{\beta=1}^{N} C_{\alpha\beta} \mu_\beta$

$\alpha = 1, 2 \dots N$ and $Y_1, Y_2, \dots Y_N$ are independent.

Proof:

The set of vectors $Y_1, Y_2 \dots Y_N$ have a joint normal distribution, because the entire set of Components is a set of linear Combinations of the Components of $X_1, X_2 \dots X_N$ which have a joint normal distribution.

The expected value of Y_α is

$$E(Y_\alpha) = E\left[\sum_{\beta=1}^{N} C_{\alpha\beta} X_\beta\right]$$

$$= \sum_{\beta=1}^{N} C_{\alpha\beta} E(X_\beta)$$

$$= \sum_{\beta=1}^{N} C_{\alpha\beta} \mu_\beta$$

$$\boxed{E(Y_\alpha) = \nu_\alpha}$$

The Covariance matrix Y_α and Y_γ is

$$E(Y_\alpha, Y_\gamma') = E\left[(Y_\alpha - E(Y_\alpha))(Y_\gamma - E(Y_\gamma))'\right]$$

$$= E\left[(Y_\alpha - \nu_\alpha)(Y_\gamma - \nu_\alpha)'\right]$$

$$= E\left\{\left[\sum_{\beta=1}^{N} C_{\alpha\beta} X_\beta - \sum_{\beta=1}^{N} C_{\alpha\beta}\mu_\beta\right]\left[\sum_{\varepsilon=1}^{N} C_{\nu\varepsilon} X_\varepsilon - \sum_{\varepsilon=1}^{N} C_{\nu\varepsilon}\mu_\varepsilon\right]'\right\}$$

$$= E\left\{\left[\sum_{\beta=1}^{N} C_{\alpha\beta}(X_\beta - \mu_\beta)\right]\left[\sum_{\varepsilon=1}^{N} C_{\nu\varepsilon}(X_\varepsilon - \mu_\varepsilon)'\right]\right\}$$

$$= \sum_{\beta,\varepsilon=1}^{N} C_{\alpha\beta} C_{\nu\varepsilon}\, E\left[(X_\beta - \mu_\beta)(X_\varepsilon - \mu_\varepsilon)'\right]$$

$$= \sum_{\beta,\varepsilon=1}^{N} C_{\alpha\beta} C_{\nu\varepsilon}\, \delta_{\beta\varepsilon}\, \Sigma$$

$$= \sum_{\beta=1}^{N} C_{\alpha\beta} C_{\nu\beta}\, \Sigma$$

Where $\delta_{\alpha\nu}$ is the Kronecker delta

$$\delta_{\alpha\nu} = \begin{cases} 1 & ; \alpha = \nu \\ 0 & ; \alpha \neq \nu \end{cases}$$

This shows that Y_α is independent of Y_ν ; $\alpha \neq \nu$ and Y_α has the covariance matrix Σ.

General lemma

If $C = (C_{\alpha\beta})$ is orthogonal matrix then

$$\sum_{\alpha=1}^{N} X_\alpha X_\alpha' = \sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' \quad \text{where} \quad Y_\alpha = \sum_{\beta=1}^{N} C_{\alpha\beta} X_\beta \,.$$

$\alpha = 1, 2 \ldots N$

Proof

$$\sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' = \sum_{\alpha=1}^{N}\left\{\left(\sum_{\beta=1}^{N} C_{\alpha\beta} X_\beta\right)\left(\sum_{\nu=1}^{N} C_{\alpha\nu} X_\nu\right)'\right\}$$

$$= \sum_{\beta,\nu=1}^{N}\left\{\sum_{\alpha=1}^{N} C_{\alpha\beta}\, C_{\alpha\nu}'\,(X_\beta X_\nu')\right\}$$

$$= \sum_{\beta,\gamma=1}^{N} \delta_{\beta\gamma} (x_\beta x_\gamma')$$

$$= \sum_{\beta=1}^{N} (x_\beta x_\beta') \qquad \{\because \delta_{\beta\gamma} \text{ is Kronecker delta}$$

$$\boxed{\sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' = \sum_{\alpha=1}^{N} (x_\alpha x_\alpha')}$$

Hence the Proof.

Theorem: If m Component vector Y is distributed according to $N(\underset{\sim}{0}, T)$ where T is non Singular then $Y' T^{-1} Y$ is distributed according to χ^2 distribution with m degrees of freedom.

Proof: Let C be a non Singular matrix Such that $CTC' = I$

and we define $Z = CY \longrightarrow (1)$

$$\Rightarrow Z = CY \sim N(\underset{\sim}{0}, CTC') \qquad \{\because Y \sim N(0,T)$$

$$\therefore Z \sim N(\underset{\sim}{0}, I)$$

$(\underset{\sim}{0}$: zerovector$)$

$$Y' T^{-1} Y = (C^{-1} Z)' T^{-1} (C^{-1} Z) \qquad \{\because \text{from } (1)$$

$$= Z' (C^{-1})' T^{-1} (C^{-1} Z)$$

$$= Z' (C')^{-1} T^{-1} C^{-1} Z$$

$$= Z' (CTC')^{-1} Z$$

$$= Z' I Z$$

$Y' T^{-1} Y = Z'Z$ which is Sum of Squares of the Components of Z

The Components of Z are independently distributed according to $N(0, I)$

$\therefore$ ~~ZZ~~ $z'z = Y'T^{-1}Y$ is a χ^2 variate with 'm' d.f.

Hence the proof.

<u>Note:</u> The square of standard normal variate is a χ^2 with 1 d.f.

<u>Theorem:</u> The mean of a sample of size N from $N(\mu, \Sigma)$ is distributed according to $N(\mu, \frac{\Sigma}{N})$ and independent of $\hat{\Sigma}$ the MLE ~~independent~~ of Σ. $N\Sigma$ is distributed as

$\sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'$ where Z_α is distributed according to $N(\underset{\sim}{0}, \Sigma)$ $\alpha = 1, 2 \ldots N-1$ and $Z_1\ Z_2 \ldots Z_{N-1}$ are independent.

<u>Proof:</u> Let $X_1\ X_2 \ldots X_N$ be independent each distributed according to $N(\mu, \Sigma)$ then there exists an $N \times N$ orthogonal matrix $B = (b_{\alpha\beta})$ with the last row such that $\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots \frac{1}{\sqrt{N}}\right)$

let $A = N\hat{\Sigma}$ $\left\{\because \hat{\Sigma} = \frac{A}{N}\right.$

and let $Z_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta} X_\beta$ ($\because$ By general Theorem

$$Z_N = \sum_{\beta=1}^{N} b_{N\beta} X_\beta$$

$$= b_{N1}X_1 + b_{N2}X_2 + \ldots + b_{NN}X_N$$

$$= \frac{1}{\sqrt{N}}X_1 + \frac{1}{\sqrt{N}}X_2 + \ldots + \frac{1}{\sqrt{N}}X_N$$

$$= \frac{1}{\sqrt{N}}(X_1 + X_2 + \ldots + X_N)$$

$$= \frac{\sqrt{N}}{N}(x_1+x_2+\dots+x_N)$$

$$\boxed{Z_N = \sqrt{N}\,\bar{x}}$$

W.K.T $\quad A = \sum_{\alpha=1}^{N} x_\alpha x_\alpha' - N\bar{x}\bar{x}'$

$$= \sum_{\alpha=1}^{N} z_\alpha z_\alpha' - z_N z_N'$$

$$\boxed{A = \sum_{\alpha=1}^{N-1} z_\alpha z_\alpha'}$$

Since z_n is independent of $z_1, z_2 \dots z_{N-1}$,

$\therefore$ the mean vector $\bar{x}$ is independent of A.

Since $\quad Z_N = \sum_{\beta=1}^{N} b_{N\beta}\, x_\beta$

$$E(Z_N) = \sum_{\beta=1}^{N} b_{N\beta}\, E(x_\beta)$$

$$= \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}}\mu$$

$$= N\frac{1}{\sqrt{N}}\mu$$

$$\boxed{E(Z_N) = \sqrt{N}\,\mu}$$

$$Z_N \sim N(\sqrt{N}\mu, \Sigma) \qquad \{\because \operatorname{Cov}(Z_N) = \Sigma$$

By theorem

Since $\quad Z_N = \sqrt{N}\,\bar{x}$

$$\boxed{\bar{x} = \frac{Z_N}{\sqrt{N}}}$$

$$E(\bar{x}) = \frac{1}{\sqrt{N}}E(Z_N)$$

$$= \frac{1}{\sqrt{N}} \sqrt{N} \mu$$

$$\boxed{E(\bar{x}) = \mu}$$

$$Cov(\bar{x}) = \left(\frac{1}{\sqrt{N}}\right)^2 Cov(Z_N)$$

$$\boxed{Cov(\bar{x}) = \frac{1}{N} \Sigma}$$

$$\therefore \bar{x} \sim N\left(\mu, \frac{1}{N}\Sigma\right)$$

$$\cancel{E X_\beta} \quad E(Z_\alpha) = \sum_{\beta=1}^{N} b_{\alpha\beta} \cancel{E(X_\beta)} E(X_\beta) = \sum_{\beta=1}^{N} b_{\alpha\beta} \mu =$$

$$\cancel{E(X_\beta)} = \qquad \left\{ \because b_{N\beta} = \frac{1}{\sqrt{N}} \right.$$

$$\Rightarrow E(Z_\alpha) = 0$$

This shows that $Z_1, Z_2 \ldots Z_{N-1}$ are independent

Theorem: An estimator t of a parameter θ, $\theta = (\theta_1 \ \theta_2 \ldots \theta_p)'$ is unbiased iff $E(t) = \theta$

Proof: Since $\bar{x} = \frac{1}{N} \Sigma x_\alpha$

$$E(\bar{x}) = \frac{1}{N} E\left[\Sigma x_\alpha\right]$$

$$= \frac{1}{N} \sum_{\alpha=1}^{N} E(x_\alpha)$$

$$= \frac{1}{N}\left[E(x_1) + E(x_2) + \cdots + E(x_N)\right]$$

$$= \frac{1}{N}\left[\mu + \mu + \cdots + \mu\right]$$

(N times)

$$= \frac{1}{N}(N\mu)$$

$$\boxed{E(\bar{x}) = \mu}$$

$\therefore$ $\bar{x}$ is an unbiased estimator of μ

Similarly, W.K.T $\frac{1}{N} A = \hat{\Sigma}$

$$E\left(\frac{1}{N} A\right) = \frac{1}{N} E(A)$$

$$= \frac{1}{N} E\left\{\sum_{\alpha=1}^{N-1} z_\alpha z_\alpha'\right\}$$

where $z_\alpha \sim N(0, \Sigma)$

$$= \frac{1}{N} \sum_{\alpha=1}^{N-1} E(z_\alpha z_\alpha')$$

$$= \frac{1}{N} \sum_{\alpha=1}^{N-1} (\Sigma)$$

$$= \frac{N-1}{N} \Sigma$$

$\therefore$ $\frac{1}{N} A$ is an unbiased estimator of Σ.

Let $E\left[\frac{N}{N-1}\frac{1}{N}A\right] = \hat{\Sigma}$

$\Rightarrow E\left[\frac{1}{N-1}A\right] = \hat{\Sigma}$

$\therefore \frac{1}{N-1}A$ is an unbiased estimator of Σ

Note: $\frac{1}{N-1}A = S$, where S is called the sample covariance matrix.

Theorem: Test and confidence region for mean vector when the covariance matrix is known.

Statement: If $\bar{x}$ is a mean of the samples $x_1, x_2 \ldots x_n$ drawn from $N_p(\mu, \Sigma)$ and Σ is known then $N(\bar{x}-\mu_0)'\Sigma^{-1}(\bar{x}-\mu_0) > \chi_p^2(\alpha)$ gives the critical region of size α for testing $H_0: \mu = \mu_0$ and $N(\bar{x}-\mu_0)'\Sigma^{-1}(\bar{x}-\mu_0) \leq \chi_p^2(\alpha)$ gives the confidence region for μ of confidence $(1-\alpha)$ where $\chi_p^2(\alpha)$ is chosen to satisfy $P_r\left[\chi_p^2 > \chi_p^2(\alpha)\right] = \alpha$

Proof:

$X \sim N(\mu, \Sigma)$

WKT $\bar{X} \sim N\left(\mu, \frac{1}{N}\Sigma\right)$

$\Rightarrow \sqrt{N}\,\bar{x} \sim N(\sqrt{N}\mu, \Sigma)$ $\left\{\because Z_N = \sqrt{N}\,\bar{x} \sim N(\sqrt{N}\mu, \Sigma\right.$

$(\sqrt{N}\bar{x} - \sqrt{N}\mu) \sim N(0, \Sigma)$

$\sqrt{N}(\bar{x}-\mu) \sim N(0, \Sigma)$

$\sqrt{N}(\bar{x}-\mu)'\Sigma^{-1}\sqrt{N}(\bar{x}-\mu) \sim \chi_p^2(\alpha)$

Let $\chi_p^2(\alpha)$ is a real number ~~sta~~ such that

$$Pr\left[\chi^2_p > \chi^2_p(\alpha)\right] = \alpha \quad ; 0<\alpha<1$$

$$Pr\left[\sqrt{N}(\bar{x}-\mu)' \Sigma^{-1} \sqrt{N}(\bar{x}-\mu) > \chi^2_p(\alpha)\right] = \alpha$$

Note: If $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ then $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{N}}\right)^2 \sim \chi^2_{(\alpha)}{}_{1\cdot d\cdot f)}$