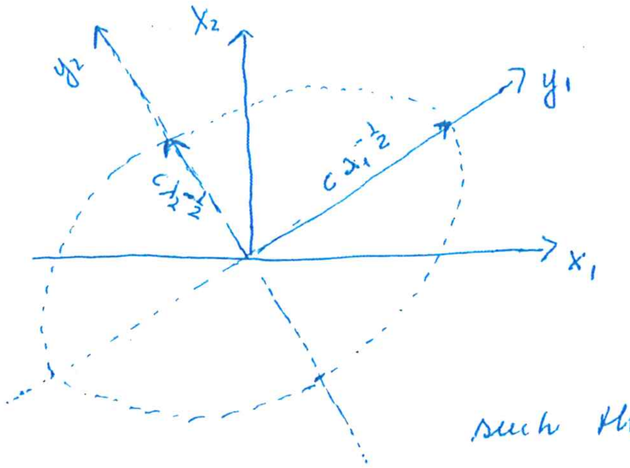
and $\underline{x}^T \underline{A}\, \underline{x} = \lambda_1 \underbrace{\underline{x}^T \underline{e}_1}_{y_1} \underline{e}_1^T \underline{x} + \lambda_2 \underline{x}^T \underline{e}_2 \underline{e}_2^T \underline{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2$

and $\underline{x}^T \underline{A}\, \underline{x} = c^2$ is the equation for an ellipse in y_1 and y_2

$$\underline{x} = c\lambda_1^{-\frac{1}{2}} \underline{e}_1 \Rightarrow \begin{cases} y_1 = c\lambda_1^{-\frac{1}{2}} \\ y_2 = 0 \end{cases}$$



$$\underline{x} = c\lambda_2^{-\frac{1}{2}} \underline{e}_2 \Rightarrow \begin{cases} y_1 = 0 \\ y_2 = c\lambda_2^{-\frac{1}{2}}. \end{cases}$$

are points on the ellipse $\lambda_1 y_1^2 + \lambda_2 y_2^2 = c^2$ such that the length of the half axes are $c\lambda_1^{-\frac{1}{2}}$ and $c\lambda_2^{-\frac{1}{2}}$. In general $(\underline{x}-\underline{\mu})^T \underline{A} (\underline{x}-\underline{\mu})$ gives the same squared distance, c^2, for all points that lie on an ellipsoid with half axes given by $c\lambda_i^{-\frac{1}{2}} \underline{e}_i$.

The multivariate normal density is constant on surfaces where $(\underline{x}-\underline{\mu})^T \underline{\Sigma}^{-1} (\underline{x}-\underline{\mu})$ is constant i.e for all $\underline{x}$ such that $(\underline{x}-\underline{\mu})^T \underline{\Sigma}^{-1} (\underline{x}-\underline{\mu}) = c^2$.

__Theorem__ (corollar 5.2 in HS). Assume $\underline{X} \sim \underline{N}(\underline{\mu}, \underline{\Sigma})$.

Then $\underline{A}\underline{X}$ and $\underline{B}\underline{X}$ are independent if and only if $\underline{A}\underline{\Sigma}\underline{B}^T = \underline{0}$

__Note__ $\underline{A}\underline{\Sigma}\underline{B}^T = \operatorname{cov}(\underline{A}\underline{X}, \underline{B}\underline{X})$

$\Rightarrow$ $\underline{A}\underline{X}$ and $\underline{B}\underline{X}$ independent $\Rightarrow \operatorname{cov}(\underline{A}\underline{X}, \underline{B}\underline{X}) = \underline{A}\underline{\Sigma}\underline{B}^T = \underline{0}$

$\Leftarrow$ $\begin{bmatrix} \underline{A} \\ \underline{B} \end{bmatrix} \underline{X} \sim N\left(\begin{bmatrix} \underline{A} \\ \underline{B} \end{bmatrix} \underline{\mu}, \begin{bmatrix} \underline{A} \\ \underline{B} \end{bmatrix} \underline{\Sigma} \begin{bmatrix} \underline{A}^T & \underline{B}^T \end{bmatrix}\right)$, $\begin{bmatrix} \underline{A} \\ \underline{B} \end{bmatrix} \underline{\Sigma} \begin{bmatrix} \underline{A}^T & \underline{B}^T \end{bmatrix} = \begin{bmatrix} \underline{A}\underline{\Sigma}\underline{A}^T & \underline{A}\underline{\Sigma}\underline{B}^T \\ \underline{B}\underline{\Sigma}\underline{A}^T & \underline{B}\underline{\Sigma}\underline{B}^T \end{bmatrix}$

with $\underline{A}\,\underline{\Sigma}\,\underline{B}^T = \underline{0}$ we get $\text{Cov}\left[\begin{bmatrix}\underline{A}\\ \underline{B}\end{bmatrix}\underline{X}\right] = \begin{bmatrix}\underline{A}\,\underline{\Sigma}\,\underline{A}^T & \underline{0}\\ \underline{0} & \underline{B}\,\underline{\Sigma}\,\underline{B}^T\end{bmatrix}$

$$\underline{\Sigma}^{-1} = \begin{bmatrix}(\underline{A}\,\underline{\Sigma}\,\underline{A}^T)^{-1} & \underline{0}\\ \underline{0} & (\underline{B}\,\underline{\Sigma}\,\underline{B}^T)^{-1}\end{bmatrix}$$ and

$$\left[(\underline{A}(\underline{X}-\underline{\mu}))^T\ (\underline{B}(\underline{X}-\underline{\mu}))^T\right]\ \underline{\Sigma}^{-1}\begin{bmatrix}\underline{A}(\underline{X}-\underline{\mu})\\ \underline{B}(\underline{X}-\underline{\mu})\end{bmatrix}$$

$$= \left[\underline{A}(\underline{X}-\underline{\mu})\right]^T\left[\underline{A}\,\underline{\Sigma}\,\underline{A}^T\right]^{-1}\left[\underline{A}(\underline{X}-\underline{\mu})\right] + \left[(\underline{B}(\underline{X}-\underline{\mu}))\right]^T\left[\underline{B}\,\underline{\Sigma}\,\underline{B}^T\right]^{-1}\left[\underline{B}(\underline{X}-\underline{\mu})\right]$$

$$|\Sigma| = |\underline{A}\,\underline{\Sigma}\,\underline{A}^T|\,|\underline{B}\,\underline{\Sigma}\,\underline{B}^T|$$

and the densities for $\underline{A}\underline{X}$ and $\underline{B}\underline{X}$ factorize which shows independence.

Corollar

Let $\underline{X} = \begin{bmatrix}X_{1\ q\times 1}\\ X_{2\ (p-q)\times 1}\end{bmatrix}$ Then $\underline{X}_1$ and $\underline{X}_2$ are independent $\iff \text{Cov}(\underline{X}_1, \underline{X}_2) = 0$

Proof. Let $A_{q\times p} = [\underline{I}_{q\times q} \vdots \underline{0}_{q\times(p-q)}]$ and

$\underline{B}_{(p-q)\times p} = [\underline{0}_{(p-q)\times q} \vdots \underline{I}_{(p-q)\times(p-q)}]$

Then $\underline{X}_1 = \underline{A}\,\underline{X}$ and $\underline{X}_2 = \underline{B}\,\underline{X}$ and $\underline{X}_1$ and $\underline{X}_2$ are independent $\iff \underline{A}\,\underline{\Sigma}\,\underline{B}^T = \underline{0} \iff \text{Cov}(\underline{X}_1, \underline{X}_2) = \underline{0}$

Theorem 5.1 and 5.3 (HS)

Let $\underline{X} = \begin{bmatrix}X_{1\ q\times 1}\\ X_{2\ (p-q)\times 1}\end{bmatrix} \sim N_p(\underline{\mu}, \underline{\Sigma})$ where $\underline{\mu} = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$

and $\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$ with $|\underline{\Sigma}_{22}| > 0$

Then

a) $\underline{X}_1 - \underline{\mu}_1 - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{X}_2 - \underline{\mu}_2)$ and $(\underline{X}_2 - \underline{\mu}_2)$ are independent.

b) $\underline{X}_1 | \underline{X}_2 = \underline{x}_2 \sim N_q\left(\underline{\mu}_1 + \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{x}_2 - \underline{\mu}_2),\ \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21}\right)$

Proof. Define $\underline{A}_{p\times p} = \begin{bmatrix} \underline{I}_{q\times q} & -\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}{}_{(q\times(p-q))} \\ \underline{0} & \underline{I}_{(p-q)\times(p-q)} \end{bmatrix}$

Then $\underline{A}(\underline{X} - \underline{\mu}) = A\begin{bmatrix} \underline{X}_1 - \underline{\mu} \\ \underline{X}_2 - \underline{\mu} \end{bmatrix} = \begin{bmatrix} \underline{X}_1 - \underline{\mu}_1 - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{X}_2 - \underline{\mu}_2) \\ \underline{X}_2 - \underline{\mu}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}, \underline{\Sigma}^*\right)$

where $\underline{\Sigma}^* = \underline{A}\,\underline{\Sigma}\,\underline{A}^T = \begin{bmatrix} \underline{I} & -\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1} \\ 0 & \underline{I} \end{bmatrix}\begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}\begin{bmatrix} \underline{I} & \underline{0} \\ (-\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1})^T & \underline{I} \end{bmatrix}$

$$= \begin{bmatrix} \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21} & \underline{0} \\ \underline{0} & \underline{\Sigma}_{22} \end{bmatrix}$$

This shows that $\underline{Y} = (\underline{X}_1 - \underline{\mu}_1) - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{X}_2 - \underline{\mu}_2)$ and $\underline{X}_2 - \underline{\mu}_2$ are independent and that $\underline{Y} \sim N_q(\underline{0}, \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21})$

In particular for a given value of $\underline{X}_2$, $\underline{x}_2$,

$\underline{Y} | \underline{X}_2 = \underline{x}_2$ is also $N_q(\underline{0}, \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21})$

and $\underline{X}_1 | \underline{X}_2 = \underline{x}_2 = \underline{Y} | \underline{X}_2 = \underline{x}_2 + \underline{\mu}_1 + \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{x}_2 - \underline{\mu}_2)$

is $N_q(\underline{\mu}_1 + \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{x}_2 - \underline{\mu}_2),\ \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21})$

Theorem 4.8. Let $\underline{X}$ be a p-variate normal distributed random vector with $E[\underline{X}] = \underline{\mu}$ and $Cov(\underline{X}) = \Sigma$. The moment generating function is given by $M_{\underline{X}}(\underline{t}) = e^{\underline{t}^T\underline{\mu} + \frac{1}{2}\underline{t}^T\Sigma\underline{t}}$

Proof. From Cramer-Wold $M_{\underline{X}}(\underline{t}) = M_{\underline{t}^T\underline{X}}(1)$

$M_{\underline{t}^T\underline{X}}(s) = e^{s\mu + \frac{1}{2}s^2\sigma^2}$ where $\mu = E[\underline{t}^T\underline{X}] = \underline{t}^T\underline{\mu}$ and

$\sigma^2 = Var[\underline{t}^T\underline{X}] = \underline{t}^T\Sigma\underline{t}$. We get $M_{\underline{X}}(\underline{t}) = e^{\underline{t}^T\underline{\mu} + \frac{1}{2}\underline{t}^T\Sigma\underline{t}}$,

Let $\underline{A}\underline{X}$ and $\underline{B}\underline{X}$ be independent, $\underline{X}$ multivariate ~~norm~~ normal distributed and let $\underline{Y} = \begin{bmatrix} A \\ B \end{bmatrix}\underline{X}$

$$M_{\underline{Y}}(\underline{t}) = e^{\underline{t}^T\begin{bmatrix} A \\ B \end{bmatrix}\underline{\mu} + \frac{1}{2}\underline{t}^T\begin{bmatrix} A\Sigma A^T & 0 \\ 0 & B\Sigma B^T \end{bmatrix}\underline{t}}$$

$$= e^{[\underline{t}_1^T\ \underline{t}_2^T]\begin{bmatrix} A \\ B \end{bmatrix}\underline{\mu} + \frac{1}{2}[\underline{t}_1, \underline{t}_2]\begin{bmatrix} A\Sigma A^T & 0 \\ 0 & B\Sigma B^T \end{bmatrix}\begin{bmatrix} \underline{t}_1 \\ \underline{t}_2 \end{bmatrix}}$$

$$= e^{\underline{t}_1^T A\underline{\mu} + \frac{1}{2}\underline{t}_1^T A\Sigma A^T\underline{t}_1} \cdot e^{\underline{t}_2^T B\underline{\mu} + \frac{1}{2}\underline{t}_2^T B\Sigma B^T\underline{t}_2}$$

as the moment generating function of independent $\underline{A}\underline{X}$ and $\underline{B}\underline{X}$

Estimation of $\underline{\mu}$ and $\underline{\Sigma}$ in a multivariate normal distribution (HS 3.3, 4.5)

Univariate $X_1, \dots, X_n$ random sample

Estimators $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$, $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

Multivariate

Let $\underline{X} = [\underline{X}_1, \dots, \underline{X}_n]^T$ be a random sample from a p-variate $N(\underline{\mu}, \underline{\Sigma})$

$$\underline{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p}.$$

How shall we estimate $\underline{\mu}$ and $\underline{\Sigma}$

$$\hat{\underline{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \underline{X}_i = \bar{\underline{X}}_{p \times 1}, \quad \underline{S} = \frac{1}{(n)} \sum_{i=1}^{n} (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})^T$$

$(n) \rightarrow$ or $(n-1)$

We have

$$\bar{\underline{X}} = \frac{1}{n} \underline{X}^T \underline{1} \quad \text{where } \underline{1}^T = [1, 1, \dots 1]_{1 \times n}$$

and

$$\frac{1}{n} \underline{1}\,\underline{1}^T \underline{X} = \frac{1}{n} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \underline{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & & \vdots \\ \vdots & \vdots & & \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \end{bmatrix}_{n \times p}$$

such that

$$\underline{X} - \frac{1}{n} \underline{1}\,\underline{1}^T \underline{X} = \begin{bmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1p} - \bar{x}_p \\ \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}_{n \times p}$$

$$= (\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T) \underline{X}$$

and we can write

~~$S = \underline{X} - \frac{1}{n} \underline{1}\,\underline{1}^T \underline{X}$~~

$$S = \frac{\underline{X}'(\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T)(\underline{I} - \frac{1}{n} \underline{1}\,\underline{1}^T) \underline{X}}{n}$$