Chapter 4 - Multivariate Normal Distribution

Univariate normal:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \qquad \begin{matrix} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{matrix}$$

Note: $\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$

Multivariate Normal $(p>1)$: $\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$ $\underset{\sim}{\mu}_X = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$ $\Sigma_X = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{1p} & \cdots & \sigma_{pp} \end{bmatrix}$

$$(\underset{\sim}{X} - \underset{\sim}{\mu}_X)' \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu}_X) \equiv 1 \times 1$$

$$f(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_X|^{1/2}} e^{-(\underset{\sim}{x} - \underset{\sim}{\mu}_x)' \Sigma_X^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_X)} \qquad \begin{matrix} -\infty < x_i < \infty \\ i = 1, \ldots, p \end{matrix}$$

$$\Rightarrow \underset{\sim}{X}_p \sim N_p(\underset{\sim}{\mu}_X, \Sigma_X)$$

Bivariate Normal $(p=2)$: $\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ $\underset{\sim}{\mu}_X = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

$$\Sigma_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \quad \text{s.t.} \quad \sigma_{12} = \rho_{12} \sqrt{\sigma_{11}} \sqrt{\sigma_{22}}$$

$$|\Sigma_X|^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

$$\Rightarrow (\underset{\sim}{x} - \underset{\sim}{\mu}_X)' \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_X) = \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix} \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}$$

Note $\sigma_{12} = \rho \sqrt{\sigma_{11}} \sqrt{\sigma_{22}}$ $\quad \sigma_{12}^2 = \rho_{12}^2 \sigma_{11} \sigma_{22}$

$$\Rightarrow \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} = \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)}$$

$$\Rightarrow (\underset{\sim}{x} - \underset{\sim}{\mu}_x)' \Sigma_x^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_x) = \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \times$$

$$\left[(x_1-\mu_1)\sigma_{22} - (x_2-\mu_2)\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \quad -(x_1-\mu_1)\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} + (x_2-\mu_2)\sigma_{11} \right] \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \Big[(x_1-\mu_1)^2\sigma_{22} - (x_1-\mu_1)(x_2-\mu_2)\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$$

$$+ (-x_1-\mu)(x_2-\mu_2)\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} + (x_2-\mu_2)^2\sigma_{11} \Big]$$

$$= \frac{(x_1-\mu_1)^2}{\sigma_{11}(1-\rho_{12}^2)} - \frac{2(x_1-\mu_1)(x_2-\mu_2)\rho_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(1-\rho_{12}^2)} + \frac{(x_2-\mu_2)^2}{\sigma_{22}(1-\rho_{12}^2)}$$

$$= \frac{1}{(1-\rho_{12}^2)} \left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) \right]$$

$$\Rightarrow (p=2) \quad f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{(\underset{\sim}{x}-\underset{\sim}{\mu}_x)'\Sigma_x^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_x)}{2} \right\}$$

$$(2\pi)^{p/2} = (2\pi)^{2/2} = 2\pi \qquad |\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$$

$$\Rightarrow f(x_1,x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \times \exp\left\{ -\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) \right] \right\}$$

General p: Constant probability density contour:

$$\{ \underset{\sim}{x} \text{ s.t. } (\underset{\sim}{x}-\underset{\sim}{\mu}_x') \Sigma_x^{-1} (\underset{\sim}{x}-\mu_x) = c^2 \}$$

$\equiv$ surface of ellipsoid centered @ $\underset{\sim}{\mu}_x$

Result 4.1 $\Sigma_x \equiv$ p.d. $\Rightarrow \Sigma_x^{-1}$ exists

$\Rightarrow \quad \Sigma_x \underset{\sim}{e} = \lambda \underset{\sim}{e} \Rightarrow \Sigma_x^{-1} \underset{\sim}{e} = \frac{1}{\lambda} \underset{\sim}{e}$

$(\lambda, \underset{\sim}{e}) \equiv$ eigenvalue-eigenvector pair for Σ_x

Corresponding to $\left(\left(\frac{1}{\lambda}\right), \underset{\sim}{e}\right)$ for Σ^{-1} ($\Sigma \equiv$ p.d.)

Proof: $\Sigma \equiv$ p.d. , $\underset{\sim}{e} \neq \underset{\sim}{0} \equiv$ eigenvector

$\Rightarrow 0 < \underset{\sim}{e}' \Sigma \underset{\sim}{e} = \underset{\sim}{e}'(\Sigma \underset{\sim}{e}) = \underset{\sim}{e}' \lambda \underset{\sim}{e} = \lambda \underset{\sim}{e}'\underset{\sim}{e} = \lambda$

Also $\underset{\sim}{e} = \Sigma^{-1}(\Sigma \underset{\sim}{e}) = \Sigma^{-1}(\lambda \underset{\sim}{e}) \Rightarrow \underset{\sim}{e} = \lambda \Sigma^{-1} \underset{\sim}{e}$

dividing by $\lambda > 0 \Rightarrow \Sigma^{-1} \underset{\sim}{e} = \frac{1}{\lambda} \underset{\sim}{e}$ $= \left(\frac{1}{\lambda}, \underset{\sim}{e}\right) \equiv$ eigval, eigvec pair for Σ^{-1}

Also: $\underset{\sim}{x}' \Sigma_x^{-1} \underset{\sim}{x} = x'\left(\sum_i \frac{1}{\lambda_i} \underset{\sim}{e}_i \underset{\sim}{e}_i'\right) x = \sum_{i=1}^{p} \frac{1}{\lambda_i} (x' e_i)^2 \geq 0$

since $\lambda_i^{-1}(x_i e_i)^2 \geq 0$ $\quad x' e_i = 0 \ \forall \ i$ only if $\underset{\sim}{x} = \underset{\sim}{0}$

$\underset{\sim}{x} \neq \underset{\sim}{0} \Rightarrow \sum_i \left(\frac{1}{\lambda_i}\right)(x' e_i)^2 > 0 \Rightarrow \Sigma_x^{-1}$ is p.d.

$\Rightarrow$ Contours of constant density for p-dim Normal are ellipsoids defined by $\underset{\sim}{x}$ s.t.

$$(\underset{\sim}{x} - \underset{\sim}{\mu}_x)' \Sigma_x^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_x) = c^2$$

Ellipsoids centered @ $\underset{\sim}{\mu}$ w/ axes $\pm c\sqrt{\lambda_i}\, e_i$ $\quad \Sigma \underset{\sim}{e}_i = \lambda_i \underset{\sim}{e}_i$, $i = 1, \ldots, p$

EXAMPLE 4.2 Contours of Bivariate Normal Density

Special Case $\sigma_{11} = \sigma_{22}$ $\quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix}$

$$0 = |\Sigma - \lambda I| = \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix} = (\sigma_{11} - \lambda)^2 - \sigma_{12}^2$$

$$= \sigma_{11}^2 - 2\sigma_{11}\lambda + \lambda^2 - \sigma_{12}^2 = (\lambda - \sigma_{11} - \sigma_{12})(\lambda - \sigma_{11} + \sigma_{12})$$

$$\Rightarrow \lambda_1 = \sigma_{11} + \sigma_{12} \qquad \lambda_2 = \sigma_{11} - \sigma_{12}$$

$$\Sigma \underset{\sim}{x}_1 = \lambda_1 \underset{\sim}{x}_1 = \begin{matrix} \sigma_{11} x_1 + \sigma_{12} x_2 = (\sigma_{11} + \sigma_{12}) x_1 \\ \sigma_{12} x_1 + \sigma_{11} x_2 = (\sigma_{11} + \sigma_{12}) x_2 \end{matrix}$$

$$\Rightarrow \sigma_{12} x_2 = \sigma_{12} x_1 \Rightarrow x_1 = x_2 \Rightarrow \underset{\sim}{e}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\Sigma \underset{\sim}{x}_2 = \lambda_2 \underset{\sim}{x}_2 \Rightarrow \begin{matrix} \sigma_{11} x_1 + \sigma_{12} x_2 = (\sigma_{11} - \sigma_{12}) x_1 \\ \sigma_{12} x_1 + \sigma_{11} x_2 = (\sigma_{11} - \sigma_{12}) x_2 \end{matrix}$$

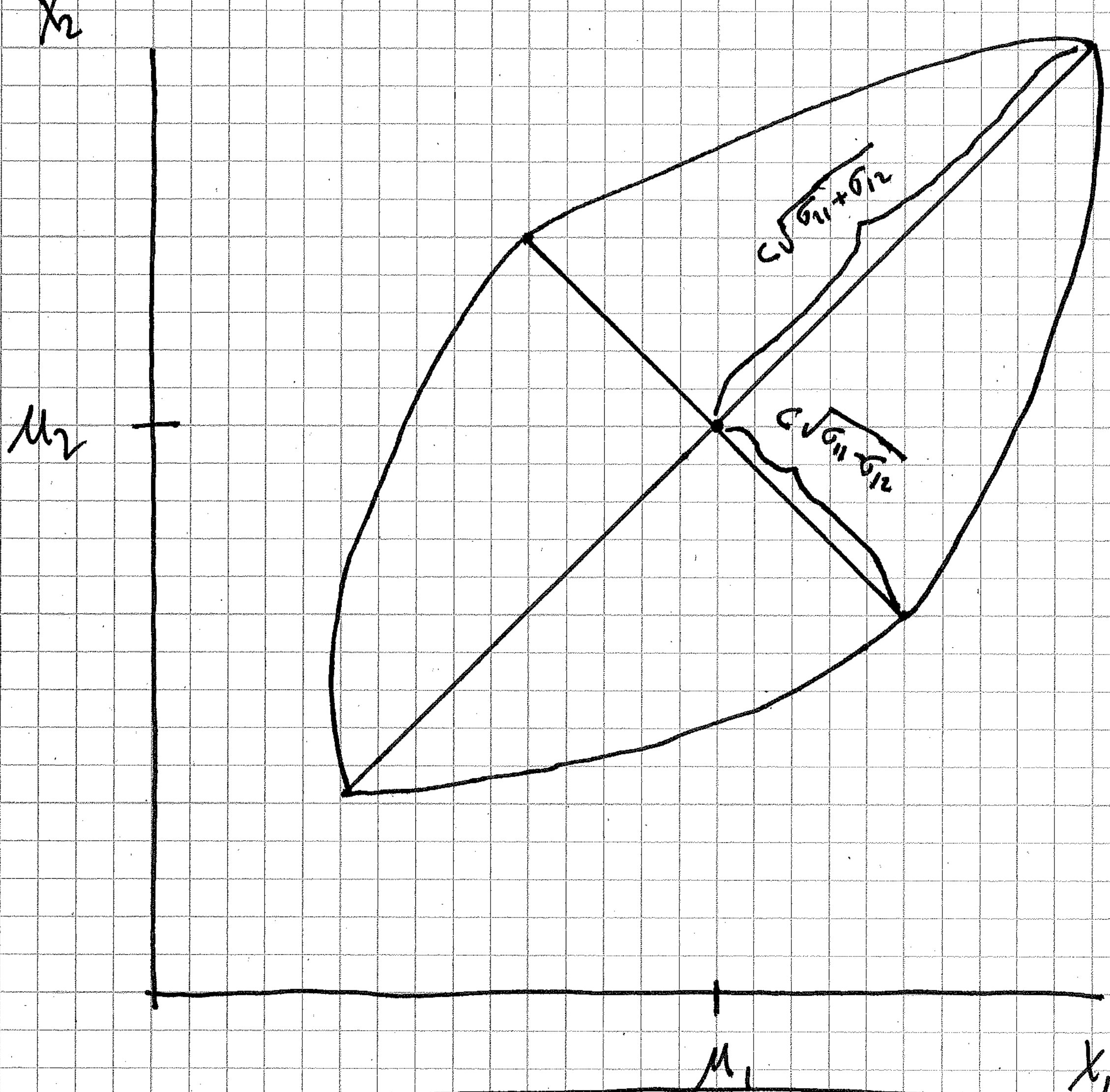
$$\Rightarrow \sigma_{12} x_2 = -\sigma_{12} x_1 \Rightarrow x_1 = -x_2 \quad \underset{\sim}{e}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

$\sigma_{12} > 0 \Rightarrow \rho_{12} > 0 \Rightarrow$ $\sigma_{11} + \sigma_{12}$ is largest eigenvalue

$$\pm c\sqrt{\lambda_1}\, \underset{\sim}{e}_1 = \pm c\sqrt{\sigma_{11} + \sigma_{12}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$ Major axis

$$\pm c\sqrt{\lambda_2}\, \underset{\sim}{e}_2 = \pm c\sqrt{\sigma_{11} - \sigma_{12}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$ minor axis

4.5



Setting $c^2 = \chi^2_p(\alpha)$ s.t. $P\left(\chi^2_p \geq \chi^2_p(\alpha)\right) = \alpha$

$\Rightarrow$ Solid ellipsoid of $(\underset{\sim}{x} - \underset{\sim}{\mu}_x)' \Sigma_x^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_x) < \chi^2_p(\alpha)$

has probability $1-\alpha$.

Additional Properties of MVN Distribution

$$\underset{p\times 1}{\underset{\sim}{X}} \sim MVN(\underset{\sim}{\mu}_x, \Sigma_x)$$

1) Linear Combinations of $\underset{\sim}{X}$ are normally distributed

$$\underset{1\times p}{\underset{\sim}{a}'} \cancel{\underset{\sim}{X}} \sim \text{Normal}(\underset{\sim}{a}'\underset{\sim}{\mu}_x,\ \underset{\sim}{a}'\Sigma_x\underset{\sim}{a})$$

$$\underset{q\times p}{A}\underset{\sim}{X} \sim N_q(A\underset{\sim}{\mu}_x,\ A\Sigma_x A')$$

2) All subsets of components of $\underset{\sim}{X}$ have multivariate normal distribution

3) zero Covariance $\Rightarrow$ Components are independent.

4) Conditional Distributions of Components are MVN.

$$X = \begin{bmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{bmatrix} \sim N_p[\underset{\sim}{\mu}_x, \Sigma_x] \quad \mu_x = \begin{bmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{bmatrix}$$

$$\Sigma_x = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} \quad w/ \quad |\Sigma_{22}| > 0$$

$\Rightarrow X_1 | X_2 = \underset{\sim}{x}_2 \sim$ Normal w/ Mean $= \underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)$

Covariance: $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Proof of Conditional Distribution (Indirect)

Let
$$\underset{p\times p}{A} = \begin{bmatrix} I_{q\times q} & -\Sigma_{12}\Sigma_{22}^{-1} \\ & q\times(p-q) \\ 0 & I \\ (p-q)\times q & (p-q)\times(p-q) \end{bmatrix}$$

$$A(\underset{\sim}{X}-\underset{\sim}{\mu}) = A\begin{bmatrix} \underset{\sim}{X}_1-\underset{\sim}{\mu}_1 \\ \underset{\sim}{X}_2-\underset{\sim}{\mu}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_1-\underset{\sim}{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}_2-\underset{\sim}{\mu}_2) \\ \underset{\sim}{X}_2-\underset{\sim}{\mu}_2 \end{bmatrix}$$

$$A\Sigma A' = \begin{bmatrix} \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \overbrace{\Sigma_{12}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}}^{0} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$\times \begin{bmatrix} I & 0' \\ -\Sigma_{22}^{-1}\Sigma_{12}' & I \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0' \\ \underbrace{\Sigma_{21}-\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21}}_{0} & \Sigma_{22} \end{bmatrix}$$

(*)

$$\Rightarrow X_1-\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}_2-\underset{\sim}{\mu}_2) \perp \underset{\sim}{X}_2-\underset{\sim}{\mu}_2$$

$$\Rightarrow X_1-\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2) \sim N_q(\underset{\sim}{0}, \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Conditonal on $X_2 = x_2$,

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2) \equiv \text{Constant}$$

by independence (*)

Conditional distn of $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underline{x}_2 - \underline{\mu}_2)$

is same as unconditional distn

of $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$

$$X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \sim N(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$\Rightarrow$ so is $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underline{\mu}_2)$ when $\underline{X}_2 = \underline{x}_2$

$$X_1 | X_2 = x_2 \sim N_q(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Special Case: $p=2$ $X_1 | X_2 = x_2$

$$f(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

$$\Rightarrow X_1 | X_2 = x_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right)$$

(For proof see Probability Distributions powerpoint slides)

Results for Multivariate Normal Distributions

1) All conditional distributions are MVN.
2) Conditional Means of form

$$\cancel{\mu_1 + \sum_{i=q+1}^{p} \Sigma_{12}\Sigma_{22}^{-1}} \quad \mu_1 + \sum_{i=q+1}^{p} \beta_{1i}(x_i - \mu_i)$$
$$\vdots$$
$$\mu_q + \sum_{i=q+1}^{p} \beta_{qi}(x_i - \mu_i)$$

where $\Sigma_{12}\Sigma_{22}^{-1} = \begin{bmatrix} \beta_{1,q+1} & \cdots & \beta_{1p} \\ \vdots & \ddots & \vdots \\ \beta_{q,q+1} & \cdots & \beta_{qp} \end{bmatrix}$

3) Conditional Covariance does not depend on value(s) of Conditioning Variables : $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$$X \sim N_p(\mu_x, \Sigma_x) \qquad |\Sigma_x| > 0$$
$$\Rightarrow (\underset{\sim}{X} - \underset{\sim}{\mu}_x)'\Sigma_x^{-1}(\underset{\sim}{X} - \mu_x) \sim \chi_p^2$$

Sampling from MVN & ML ESTIMATION

$$\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n \quad \text{iid} \quad N_p(\mu_x, \Sigma_x)$$

Joint Density of $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ $(@ \underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n)$:

$$\prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}|\Sigma_x|^{1/2}} e^{-(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)'\Sigma_x^{-1}(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)} =$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^{n} (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)' \Sigma_x^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)}$$

Once $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ observed, function of $\underset{\sim}{\mu}_x$, Σ_x => "likelihood"

Result 4.9: $A \equiv k \times k$ symmetric matrix, $X \equiv k \times 1$ vector

a) $\underset{\sim}{x}' A \underset{\sim}{x} = tr(\underset{\sim}{x}' A \underset{\sim}{x}) = tr(A \underset{\sim}{x} \underset{\sim}{x}')$

b) $tr(A) = \sum_{i=1}^{k} \lambda_i$ $\quad \lambda_i \equiv$ eigenvalues of A

$$(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)' \Sigma_x^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x) = tr(\cdot)$$

$$= tr\left(\Sigma_x^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)' \right)$$

$$\sum_j (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)' \Sigma_x^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x) = \sum_j tr(\cdot)$$

$$= \sum_j tr\left(\Sigma_x^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)' \right)$$

$$= tr\left(\Sigma_x^{-1} \sum_j (\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x)' \right) \quad (*)$$

$$(\underset{\sim}{x}_j - \underset{\sim}{\mu}_x) = \underset{\sim}{x}_j - \bar{\underline{x}} + \bar{\underline{x}} - \underset{\sim}{\mu}_x$$

$$\Rightarrow \cancel{(*) = tr\left(\Sigma_x^{-1} \sum_j (\underset{\sim}{x}_j - \bar{\underline{x}} + \bar{\underline{x}} - \underline{\mu}_x)(\underline{x}_j - \bar{\underline{x}} + \bar{\underline{x}} - \underline{\mu}_x) \right)}$$

$$\sum_j (\underline{x}_j - \underline{\mu}_x)(\underline{x}_j - \underline{\mu}_x)' = \sum_j (\underline{x}_j - \bar{\underline{x}} + \bar{\underline{x}} - \underline{\mu}_x)(\underline{x}_j - \bar{\underline{x}} + \bar{\underline{x}} - \underline{\mu}_x)'$$

$$= \sum_j (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' + \overset{0}{\cancel{\sum_j (\underline{x}_j - \bar{\underline{x}})(\bar{\underline{x}} - \underline{\mu}_x)'}}$$

$$+ \underset{0}{\cancel{\sum_j (\bar{\underline{x}} - \underline{\mu}_x)(\underline{x}_j - \bar{\underline{x}})'}} + \sum_j (\bar{\underline{x}} - \underline{\mu}_x)(\bar{\underline{x}} - \underline{\mu}_x)'$$

$$= \sum_j (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' + \sum_j (\bar{\underline{x}} - \underline{\mu}_x)(\bar{\underline{x}} - \underline{\mu}_x)'$$

$\Rightarrow$ Joint Density of $X_1, \ldots, X_n \equiv$

$$(2\pi)^{-np/2} |\Sigma_x|^{-n/2} \exp\left\{ -tr\left[\Sigma_x^{-1} \left(\sum_j (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' + n(\bar{\underline{x}} - \underline{\mu}_x)(\bar{\underline{x}} - \underline{\mu}_x)' \right) \right] / 2 \right\}$$

$\equiv$ Likelihood function for μ_x, Σ_x for observed $\underline{x}_1, \ldots, \underline{x}_n$

N.B.: $$tr\left[\Sigma_x^{-1} \left(\sum_j (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' + n(\bar{\underline{x}} - \underline{\mu}_x)(\bar{\underline{x}} - \underline{\mu}_x)' \right) \right]$$

$$= tr\left[\Sigma_x^{-1} \left(\sum_j (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' \right) \right] + n\, tr\left[\Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}_x)(\bar{\underline{x}} - \underline{\mu}_x)' \right]$$

$$= tr\left[\Sigma_x^{-1} \left(\sum_j (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' \right) \right] + n (\bar{\underline{x}} - \underline{\mu}_x)' \Sigma_x^{-1} (\bar{\underline{x}} - \underline{\mu}_x)$$

Maximum Likelihood Estimators for $\underline{\mu}_x, \Sigma_x$

Result 4.10 $B \equiv p \times p$ p.d. matrix, scalar $b > 0$

$$\frac{1}{|\Sigma|^b} e^{-\mathrm{tr}(\Sigma^{-1}B)/2} \leq \frac{1}{|B|^b} (2b)^{pb} e^{-pb}$$

$\forall$ p.d. $\underset{p \times p}{\Sigma_x}$ w/ equality iff $\Sigma_x = \frac{1}{2b} B$

Let $\bar{X}, S \equiv$ Sample mean and variance-covariance matrix for sampled data.

Result 4.11 $\underline{X}_1, \ldots, \underline{X}_n$ iid $N_p(\underline{\mu}_x, \Sigma_x)$ random samp.

$$\hat{\mu} = \bar{X} \qquad \hat{\Sigma} = \frac{1}{n} \sum_j (\underline{x}_j - \underline{\bar{x}})(\underline{x}_j - \underline{\bar{x}})'$$

Proof: Exponential portion of Likelihood function:

$$-\frac{1}{2}\left[\mathrm{tr}\left(\Sigma_x^{-1}\left(\sum_j (x_j - \bar{x})(x_j - \bar{x})'\right)\right) + n(\underline{\bar{X}} - \underline{\mu}_x)' \Sigma_x^{-1} (\underline{\bar{X}} - \underline{\mu}_x)\right]$$

Maximize exponent $\Rightarrow$ minimize ↑

wrt $\mu_x \Rightarrow$ let $\hat{\mu}_x = \bar{X}$ (Last term $= 0$)

Σ_x^{-1} p.d. $\Rightarrow n(\underline{\bar{X}} - \underline{\mu}_x)' \Sigma_x^{-1} (\underline{\bar{X}} - \underline{\mu}_x) \geq 0$

; Only 0 when $\underline{\bar{X}} - \underline{\mu}_x = 0$

(using Result 4.10)

maximizing $L(\mu, \Sigma)$ wrt Σ_x

Let $b = \frac{n}{2}$, $B = \sum_j (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'$

$$\Rightarrow \frac{1}{|\Sigma_x|^{n/2}} e^{-tr\left(\Sigma_x^{-1} \sum_j (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'\right)/2}$$

$$\leq \frac{1}{\left|\sum_j (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'\right|^{n/2}} (n)^{\frac{pn}{2}} e^{-pn/2}$$

w/ equality iff $\hat{\Sigma}_x = \frac{1}{n} \sum_j (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'$

Notes:

ML estimator of $h(\underset{\sim}{\theta})$ is ~~[illegible]~~ $\widehat{h(\hat{\underset{\sim}{\theta}})}^{MLE}$

ML estimator of $\underset{\sim}{\mu}_x' \Sigma_x^{-1} \underset{\sim}{\mu}_x = \hat{\underset{\sim}{\mu}}_x' \hat{\Sigma}_x^{-1} \hat{\underset{\sim}{\mu}}_x$

MLE for $\sqrt{\sigma_{ii}} = \sqrt{\hat{\sigma}_{ii}} = \sqrt{\frac{1}{n} \sum_j (X_{ji} - \bar{X}_i)^2}$

Sufficient Stats:

Joint density depends on $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ only

through $\bar{\underset{\sim}{X}}$ and $\sum_j (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})' = (n-1)S \Rightarrow \bar{X}, S$ sufficient statistics

Sampling Distributions of $\bar{X}$, S

Under Normality Assumption $\underset{\sim}{X}_j \sim N_p(\underset{\sim}{\mu}_x, \Sigma_x)$:

$\underset{\sim}{\bar{X}} \sim N_p(\underset{\sim}{\mu}_x, \frac{1}{n}\Sigma_x)$

$(n-1)S \sim$ Wishart $(n-1)$ df

$\underset{\sim}{\bar{X}}$, S are independent

Large-Sample Behavior of $\bar{X}$, S

$\bar{X} \xrightarrow{Prob} \underset{\sim}{\mu}$ $\qquad$ $S, \hat{\Sigma} \xrightarrow{Prob} \Sigma$

($Y_n \xrightarrow{Prob} \mu$ if $\forall\ \varepsilon > 0$, $\lim_{n\to\infty} P[-\varepsilon \le Y_n - \mu \le \varepsilon] = 1$)

Central Limit Theorem:

$\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ indep w/ $E\{\underset{\sim}{X}\} = \mu_x$ $V\{\underset{\sim}{X}\} = \Sigma$ (finite)

$\Rightarrow \sqrt{n}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}_x) \rightsquigarrow N(\underset{\sim}{0}, \Sigma_x)$

$n(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}_x)' S^{-1}(\bar{X} - \underset{\sim}{\mu}_x) \dot{\sim} \chi^2_p$ $\qquad$ for large $n-p$

Assessing Normality

Univariate – Normal prob plots, histograms, Shapiro-Wilk test
– fractions of observations w/in $\pm 1s$, $\pm 2s$ of mean

Bivariate – percentage of points w/in some distance of the mean

$$\{ x : (\underline{x}-\overline{\underline{x}})' S^{-1} (\underline{x}-\overline{\underline{x}}) \le \chi^2_2(p^*) \}$$

should be p^* of data

(use "elliptic function" in R)

In General look for outliers by extreme values of $(\underline{x}_j-\overline{\underline{x}})' S^{-1} (\underline{x}_j-\overline{\underline{x}})$

(extreme relative to ~~standard~~ χ^2_p distribution)

Transformations to Approximate Normality

Counts (y) – $\sqrt{y}$

Proportions $(\hat{p})$ – $\text{logit}(\hat{p}) = \frac{1}{2} \ln\left(\frac{\hat{p}}{1-\hat{p}}\right)$ or $\arcsin\sqrt{\hat{p}}$

Correlations (r) – Fishers z-transformation $\frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$

Box-Cox power transformation