Math3806 Lecture Note 3 Appendix

Heng Peng

January 29, 2021

P1. Given $X_1, X_2, \ldots, X_n$ are independent random variables, have a same distribution with mean μ and variance σ^2, then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu, \text{ as } n \to \infty$$

(Law of Large number)

$$\sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, \sigma^2), \text{ as } n \to \infty$$

(*Central Limited Theorem*)

- Law of large number and Central limited theorem can be extended to multivariate dimensional random vectors.
- logarithm of normal likelihood

$$\log \prod_{i=1}^{n} f(X_i) = \sum_{i=1}^{n} \log f(X_i) = \sum_{i=1}^{n} \{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(X_i - \mu)^2}{2\sigma^2}\}$$

P3. Rewrite the normal density function

$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}(\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)}$$

$$(2\pi)^{\frac{1}{2}} \sim (2\pi)^{\frac{p}{2}}, x \sim \boldsymbol{X}, \mu \sim \boldsymbol{\mu}, \sigma^2 \sim \Sigma \text{ or } |\Sigma|$$

P4. Quantile values of Normal population. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathrm{P}(|X - \mu| \geq \sigma) \approx 1 - .683 = .317,$$

$$\mathrm{P}(|X - \mu| \geq 2\sigma) \approx 1 - .954 = .046 < 0.05.$$

P5. Example 3.1. Bivariate normal density

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

$$\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}, \ |\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$$

$$\begin{aligned} & (\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\ = \ & [x_1-\mu_1, x_2-\mu_2]\frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \\ \times \ & \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \\ = \ & \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \end{aligned}$$

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

- If $\rho_{12} = 0$, then $f(x_1, x_2) = f(x_1)f(x_2)$, $x_{1,2}$ are not only linear independent, but also statistical independent. This result is true in general .

P5. Result3.1 is an extension of the properties of the positive definite matrix with its spectral decomposition. See lecture note 2.

P7. Constant probability density contour, i.e for all $\mathbf{x}$, their density equal to a constant, i.e. $f(\mathbf{x}) = c$.

P8. Example 4.2. First find eigenvalues and eigenvectors of Σ, by solve $|\Sigma - \lambda \mathbf{I}| = 0$

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

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

$$\lambda_1 = \sigma_{11} + \sigma_{12}, \quad \lambda_2 = \sigma_{11} - \sigma_{12}$$

- Solve the following equations when $\lambda = \lambda_1$ or λ_2

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \lambda \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

- Then get the solutions when
  $\lambda_1 = \sigma_{11} + \sigma_{12}$, $\mathbf{e}_1 = [1/\sqrt{2}, 1/\sqrt{2}]^T$,
  and
  $\lambda_2 = \sigma_{11} - \sigma_{12}$, $\mathbf{e}_2 = [1/\sqrt{2}, -1/\sqrt{2}]^T$
- To summarize, the axes of the ellipse of constant density for a bivariate normal density with $\sigma_{11} = \sigma_{22}$ are determined by

$$\pm c\sqrt{\sigma_{11} + \sigma_{12}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \text{ and } \pm c\sqrt{\sigma_{11} - \sigma_{12}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

P9. If $\mathbf{x}$ follows multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and Σ, then

$$\mathrm{P}(f(\mathbf{x}) \geq c) = 1 - \alpha$$

is equivalent to

$$\mathrm{P}\left(\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\mu)/2} \geq c\right) = 1 - \alpha$$

So

$$\mathrm{P}\left(e^{-(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\mu)/2} \geq (2\pi)^{p/2}|\Sigma|^{1/2}c\right) = 1 - \alpha$$

and

$$\mathrm{P}\left((\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\mu) \leq -2\log\{(2\pi)^{p/2}|\Sigma|^{1/2}c\}\right) = 1 - \alpha$$

$(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\mu) \sim \chi^2_p$, and $-2\log\{(2\pi)^{p/2}|\Sigma|^{1/2}c\} = \chi^2_p(\alpha)$

P11.

$$\mathrm{E}\mathbf{a}^T\boldsymbol{X} = \mathbf{a}^T\mathrm{E}\boldsymbol{X} = \mathbf{a}^T\boldsymbol{\mu}$$

$$\mathrm{Var}(\mathbf{a}^T\boldsymbol{X}) = \mathbf{a}^T\mathrm{Var}(\boldsymbol{X})\mathbf{a} = \mathbf{a}^T\Sigma\mathbf{a}$$

- Example 3.3. $\mathrm{E}\mathbf{a}^T\boldsymbol{X} = \mu_1$, $\mathrm{Var}(\mathbf{a}^T\boldsymbol{X}) = \sigma_{11}$.
-

$$\mathrm{E}\mathbf{A}\boldsymbol{X} = \mathbf{A}\mathrm{E}\boldsymbol{X} = \mathbf{A}\mu$$

$$\mathrm{Var}(\mathbf{A}\boldsymbol{X}) = \mathbf{A}\mathrm{Var}(\boldsymbol{X})\mathbf{A}^T = \mathbf{A}\Sigma\mathbf{A}^T$$

P12. Example 3.4.

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Hence

$$\mathrm{E}\mathbf{A}\boldsymbol{X} = \mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

$$\begin{aligned} \mathrm{Var}(\mathbf{A}\boldsymbol{X}) &= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \Sigma \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} \\ \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix} \end{aligned}$$

P13. Let $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$ where $\mathbf{A}_1 = \mathbf{I}_{q\times q}$ and $\mathbf{A}_2 = \mathbf{0}_{q\times(p-q)}$. Then applying Result 3.3. to obtain result 3.4.

- Example 3.5.

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Then

$$\mathbf{A}\boldsymbol{X} = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$$

Hence $[X_2, X_4]^T$ follows the multivariate normal distribution with mean vector

$$\boldsymbol{\mu}^* = \mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \text{ and } \Sigma^* = \mathbf{A}\Sigma\mathbf{A}^T = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

P14. If $\Sigma_{12} = \Sigma_{21} = 0$, then $|\Sigma| = |\Sigma_{11}||\Sigma_{22}|$. The density function of $[\boldsymbol{X}_1^T, \boldsymbol{X}_2^T]^T$ can be written as

$$
\begin{aligned}
& f(\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{1}{(2\pi)^{q_1/2}(2\pi)^{q_2/2}|\Sigma_{11}|^{1/2}|\Sigma_{22}|^{1/2}} \times \\
& \exp\left\{-\frac{1}{2}[(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T, (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T] \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}\right\} \\
= & \frac{1}{(2\pi)^{q_1/2}|\Sigma_{11}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)\right\} \\
\times & \frac{1}{(2\pi)^{q_2/2}|\Sigma_{22}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right\} \\
= & f(\boldsymbol{X}_1) f(\boldsymbol{X}_2)
\end{aligned}
$$

P15. Example 3.6. X_1, X_2 are not independent, but (X_1, X_2) and X_3 are independent.

- First, it is well known

$$f(\boldsymbol{X}_1 | \boldsymbol{X}_2 = \mathbf{x}_2) = \frac{f(\boldsymbol{X}_1, \boldsymbol{X}_2 = \mathbf{x}_2)}{f(\boldsymbol{X}_2 = \mathbf{x}_2)}.$$

Let

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

Then

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

- The covariance matrix $\mathbf{A}\Sigma\mathbf{A}^T$ should be

$$\begin{bmatrix} \mathbf{I}_{q\times q} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0}_{(p-q)\times q} & \mathbf{I}_{(p-q)\times(p-q)} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{q\times q} & \mathbf{0}_{q\times(p-q)} \\ (-\Sigma_{12}\Sigma_{22}^{-1})^T & \mathbf{I}_{(p-q)\times(p-q)} \end{bmatrix}$$

  After some calculation, it should equal to

$$\begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix}$$

- $\boldsymbol{X}_1 - \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{X}_2 - \boldsymbol{\mu}_2$ have zero covariance, and they are independent.
- Hence given $\boldsymbol{X}_2 = \mathbf{x}_2$, $\boldsymbol{X}_1 - \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)$ is same as unconditional distribution, and $\boldsymbol{X}_1$ is same as with distribution $\mathcal{N}_q(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.
- Check

$$\begin{aligned} & f(\boldsymbol{X}_1 | \boldsymbol{X}_2 = \mathbf{x}_2) f(\boldsymbol{X}_2 = \mathbf{x}_2) \\ = \; & \mathcal{N}_q(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \mathcal{N}_{p-q}(\boldsymbol{\mu}_2, \Sigma_{22}) \\ = \; & \mathcal{N}_p(\boldsymbol{\mu}, \Sigma) = f(\boldsymbol{X}_1, \boldsymbol{X}_2 = \mathbf{x}_2) \end{aligned}$$

P16. Example 3.7

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)} = \mathcal{N}\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right)$$

(Check the equation $f(x_1|x_2)f(x_2) = f(x_1, x_2)$)

- Given $X_1, \ldots, X_p$ independent and follow standard normal $\mathcal{N}(0, 1)$ distribution, then

$$X_1^2 + \cdots + X_p^2 \sim \chi_p^2.$$

-

$$(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})^T(\frac{1}{\lambda_1}\mathbf{e}_1\mathbf{e}_1^T + \cdots + \frac{1}{\lambda_p}\mathbf{e}_p\mathbf{e}_p^T)(\mathbf{x}-\boldsymbol{\mu})$$

$$= \frac{1}{\lambda_1}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{e}_1\mathbf{e}_1^T(\mathbf{x}-\boldsymbol{\mu}) + \cdots + \frac{1}{\lambda_p}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{e}_p\mathbf{e}_p^T(\mathbf{x}-\boldsymbol{\mu})$$

- Let $y_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i^T(\mathbf{x}-\boldsymbol{\mu})$, then

$$(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = y_1^2 + \cdot + y_p^2.$$

  Furthermore for $i = 1, \ldots, p$, $\mathrm{E}y_i = 0$ and

$$\mathrm{Var}(y_i) = \mathrm{Var}(\frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i^T(\mathbf{x}-\boldsymbol{\mu})) = \frac{1}{\lambda_i}\mathbf{e}_i^T\Sigma\mathbf{e}_i = 1$$

- So $y_i \sim \mathcal{N}(0,1), i = 1, \ldots, p$,

$$y_1^2 + \cdots + y_p^2 \sim \chi_p^2.$$

  and

$$\mathrm{P}((\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \le \chi_p^2(\alpha)) = 1-\alpha.$$

P17. Let $\boldsymbol{X} = (\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_n^T)^T$ and $\mathbf{A} = [c_1\mathbf{I}_{p\times p}, \cdots, c_n\mathbf{I}_{p\times p}]_{p\times(np)}$, then

$$\mathbf{V}_1 = \mathbf{A}\boldsymbol{X}.$$

Hence $\mathbf{V}_1$ follows multivariate normal distribution with mean vector

$$\mathbf{A}\mathrm{E}\boldsymbol{X} = [c_1\mathbf{I}_{p\times p}, \cdots, c_n\mathbf{I}_{p\times p}]\begin{bmatrix}\mathrm{E}\boldsymbol{X}_1\\ \vdots \\ \mathrm{E}\boldsymbol{X}_n\end{bmatrix} = \sum_{i=1}^n c_i\mathrm{E}\boldsymbol{X}_i = \sum_{i=1}^n c_i\boldsymbol{\mu}_i$$

- Because of independence of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$,

$$\mathrm{Var}(\boldsymbol{X}) = \mathrm{diag}(\Sigma_{p\times p}, \ldots, \Sigma_{p\times p})_{np\times np}$$

$$\begin{aligned}\mathrm{Var}(\mathbf{A}\boldsymbol{X}) &= \mathbf{A}\mathrm{diag}(\Sigma_{p\times p}, \ldots, \Sigma_{p\times p})_{np\times np}\mathbf{A}^T = \sum_{i=1}^n c_i^2\mathbf{I}_{p\times p}\Sigma\mathbf{I}_{p\times p}\\ &= \sum_{i=1}^n c_i^2\Sigma = \left(\sum_{i=1}^n c_i^2\right)\Sigma\end{aligned}$$

- Define $\mathbf{B} = [b_1\mathbf{I}_{p\times p}, \cdots, b_n\mathbf{I}_{p\times p}]_{p\times(np)}$, and then $(\mathbf{V}_1^T, \mathbf{V}_2^T)^T = (\mathbf{A}^T, \mathbf{B}^T)^T\boldsymbol{X}$, So the joint distribution of $\mathbf{V}_1$ and $\mathbf{V}_2$ follow the multivariate normal distribution. Their covariance matrix should be

$$\mathrm{Cov}(\mathbf{A}\boldsymbol{X}, \mathbf{B}\boldsymbol{X}) = \mathbf{A}\mathrm{Var}(\boldsymbol{X})\mathbf{B}^T = \left(\sum_{i=1}^{n} c_i b_i\right)\Sigma = \mathbf{b}^T\mathbf{c}\Sigma.$$

P18. Example 3.8.(a)

$$\mathrm{E}\mathbf{a}'\boldsymbol{X} = 3a_1 - a_2 + a_3$$

and

$$\mathrm{Var}(\mathbf{a}^T\boldsymbol{X}_1) = \mathbf{a}^T\Sigma\mathbf{a} = 3a_1^2 + a_2^2 + 2a_3^2 - 2a_1a_2 + 2a_1a_3$$

- (b). $\mathbf{A} = (\frac{1}{2}\mathbf{I}, \frac{1}{2}\mathbf{I}, \frac{1}{2}\mathbf{I}, \frac{1}{2}\mathbf{I})$ and $\mathbf{B} = (\mathbf{I}, \mathbf{I}, \mathbf{I}, -3\mathbf{I})$. Hence

$$\mathrm{E}\mathbf{A}\boldsymbol{X} = 2\boldsymbol{\mu} = (6, -2, 2)^T$$

$$\mathrm{E}\mathbf{B}\boldsymbol{X} = (3-3)\boldsymbol{\mu} = (3, -1, 1)^T = (0, 0, 0)^T$$

$$\mathrm{Var}(\mathbf{A}\boldsymbol{X}) = (\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4})\Sigma = \Sigma,$$

and

$$\mathrm{Var}(\mathbf{B}\boldsymbol{X}) = (1 + 1 + 1 + 3^2)\Sigma = \begin{bmatrix} 36 & -12 & 12 \\ -12 & 12 & 0 \\ 12 & 0 & 24 \end{bmatrix}$$

Because $\mathbf{a}^T\mathbf{b} = 0$, the covariance between these two linear combination is zero. $\mathbf{A}\boldsymbol{X}$ and $\mathbf{B}\boldsymbol{X}$ are independent.

P20. Let $\{p_\theta : \theta \in \Theta\}$ be a collection of subprobability densities such that $p_\theta \neq p_{\theta_0}$ if $\theta \neq \theta_0$, then $M(\theta) = \mathrm{E}_{\theta_0} \log p_\theta / p_{\theta_0}$ attains its maximum uniquely at θ_0.

- By $\log x \leq 2(\sqrt{x} - 1)$ for every $x > 0$,

$$\begin{aligned} \mathrm{E}_{\theta_0} \log \frac{p_\theta}{p_{\theta_0}} &\leq 2\mathrm{E}_{\theta_0} \left(\sqrt{\frac{p_\theta}{p_{\theta_0}}} - 1 \right) = 2 \int \sqrt{p_\theta p_{\theta_0}} - 2 \\ &\leq -\int (\sqrt{p_\theta} - \sqrt{p_{\theta_0}})^2 d\mu. \end{aligned}$$

- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$
- $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(P\Lambda P') = \mathrm{tr}(\Lambda P' P) = \mathrm{tr}(\Lambda) = \sum_{i=1}^{n} \lambda_i$

P21. $\mathrm{E}\mathbf{S} = \Sigma$ and $\mathrm{E}\hat{\Sigma} = \frac{n-1}{n}\Sigma$.

P23. $f(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{\mu}, \Sigma) = f_1(\bar{\boldsymbol{X}}, \mathbf{S}, \boldsymbol{\mu}, \Sigma) \cdot f_2(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$

P23. $(n-1)s^2 = \sigma^2\chi^2_{n-1}$

P24. $\bar{\boldsymbol{X}} = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n$, $\mathbf{b}^* = \frac{1}{n}\mathbf{b} = \frac{1}{n}(1, \ldots, 1)^T$,

$\|\mathbf{b}^*\|^2 = \frac{1}{n^2}\sum_{i=1}^{n} b_i^2 = \frac{1}{n}$.

- $\bar{\boldsymbol{X}} = \mathbf{b}^T\boldsymbol{X}$, $(n-1)\mathbf{S} = \boldsymbol{X}^T(\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{b}\mathbf{b}^T)(\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{b}\mathbf{b}^T)^T\boldsymbol{X}$

$$\mathbf{b}^T(\mathbf{I}_{n\times n} - \frac{1}{n}\mathbf{b}\mathbf{b}^T) = 0$$

P37. Example 3.11. $\bar{\boldsymbol{X}} = .770$

$$\sum_{j=1}^{10}(X_{(j)} - \bar{\boldsymbol{X}})q_{(j)} = 8.584, \ \sum_{j=1}^{10}(X_{(j)} - \bar{\boldsymbol{X}})^2 = 8.472, \ \sum_{j=1}^{10} q_{(j)}^2 = 8.795$$

since $\bar{q} = 0$,

$$r_Q = \frac{8.584}{\sqrt{8.4728.795}} = .994 > .9351$$

Do not reject the hypothesis of normality.

P39. Example 3.12.

$$\bar{X} = \begin{bmatrix} 155.60 \\ 14.70 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{bmatrix}$$

So

$$\begin{aligned} \mathbf{S}^{-1} &= \frac{1}{103,623,12} \begin{bmatrix} 26.19 & -303.62 \\ -303.62 & 7476.45 \end{bmatrix} \\ &= \begin{bmatrix} .000253 & -.002930 \\ -.002930 & .072148 \end{bmatrix} \end{aligned}$$

Because $\chi_2^2(.5) = 1.39$, Any observation $\boldsymbol{X}^T = [X_1, X_2]$ satisfy

$$\begin{bmatrix} X_1 - 155.60 \\ X_2 - 14.70 \end{bmatrix}^T \begin{bmatrix} .000253 & -.002930 \\ -.002930 & .072148 \end{bmatrix} \begin{bmatrix} X_1 - 155.60 \\ X_2 - 14.70 \end{bmatrix} \leq 1.39$$

is on or insider the estimated 50% contour. Otherwise the observation is outside this contour. The generalized distances of the ten observations are $1.61, 0.30, 0.62, 1.79, 1.30, 4.38, 1.64, 3.53, 1.71, 1.16$.