Applied Data Analysis

Exercise Sheet 2 - Solutions

Exercise 7

(a) According to Theorem I.2.7(3) marginals corresponding to the index set $\emptyset \neq J$ with $j = |J| \in \mathbb{N}$ of the normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$, $p \in \mathbb{N}$, $p \geq j$, are again distributed according to $N_j(\boldsymbol{\mu}_j, \Sigma_{J,J})$, i.e. the normal distribution with mean vector $\boldsymbol{\mu}_J$ and covariance matrix $\Sigma_{J,J} \in \mathbb{R}^{j\times j}_{\geq 0}$.

Therefore, by appropriate marginalization, one may assume, without loss of generality, that $K \cup L = \{1, \ldots, p\}$. Further, according to Definition 1.2.2 – after applying permuation matrices, if necessary – one may assume, without loss of generality, that

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_K \\ \boldsymbol{X}_L \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_K \\ \boldsymbol{\mu}_L \end{pmatrix}, \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{K,K} & \Sigma_{K,L} \\ \Sigma_{L,K} & \Sigma_{L,L} \end{pmatrix}.$$

Following the hint, for $\boldsymbol{x} = (\boldsymbol{x}_K, \boldsymbol{x}_L) \in \mathbb{R}^p$ the density of $\boldsymbol{X} = \boldsymbol{X}_{K\cup L}$ can be factorized into

$$f^{\boldsymbol{X}}(\boldsymbol{x}_K, \boldsymbol{x}_L) = f^{\boldsymbol{X}_K}(\boldsymbol{x}_K) \cdot f^{\boldsymbol{X}_L|\boldsymbol{X}_K=\boldsymbol{x}_K}(\boldsymbol{x}_L).$$

Application of the results of Ex. 4 yields

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{K,K}^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{pmatrix} := \begin{pmatrix} \Sigma^{K,K} & \Sigma^{K,L} \\ \Sigma^{L,K} & \Sigma^{L,L} \end{pmatrix}$$

with $E = \Sigma_{L,L} - \Sigma_{L,K}\Sigma_{K,K}^{-1}\Sigma_{K,L}$ and $F = \Sigma_{K,K}^{-1}\Sigma_{K,L}$. Notice that according to Ex. 4 the matrices $\Sigma_{K,K}^{-1}$ and E are regular if and only if Σ is regular, such that their inverses are well defined.

After some algebra the quadratic form $Q(\boldsymbol{x}_K, \boldsymbol{x}_L) = (\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ can be expressed as

$$\begin{aligned}
Q(\boldsymbol{x}_K, \boldsymbol{x}_L) =& (\boldsymbol{x}_K - \boldsymbol{\mu}_K)'\Sigma^{K,K}(\boldsymbol{x}_K - \boldsymbol{\mu}_K) \\
&\quad + 2(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{K,L}(\boldsymbol{x}_L - \boldsymbol{\mu}_L) + (\boldsymbol{x}_L - \boldsymbol{\mu}_L)'\Sigma^{L,L}(\boldsymbol{x}_L - \boldsymbol{\mu}_L) \\
=& (\boldsymbol{x}_K - \boldsymbol{\mu}_K)'(\Sigma_{K,K}^{-1} + FE^{-1}F')(\boldsymbol{x}_K - \boldsymbol{\mu}_K) \\
&\quad + 2(\boldsymbol{x} - \boldsymbol{\mu})'(-FE^{-1})(\boldsymbol{x}_L - \boldsymbol{\mu}_L) + (\boldsymbol{x}_L - \boldsymbol{\mu}_L)'E^{-1}(\boldsymbol{x}_L - \boldsymbol{\mu}_L) \\
=& (\boldsymbol{x}_K - \boldsymbol{\mu}_K)'\Sigma_{K,K}^{-1}(\boldsymbol{x}_K - \boldsymbol{\mu}_K) \\
& + (\boldsymbol{x}_K - \boldsymbol{\mu}_K)'(FE^{-1}F')(\boldsymbol{x}_K - \boldsymbol{\mu}_K) \\
& - 2(\boldsymbol{x}_K - \boldsymbol{\mu}_K)'FE^{-1}(\boldsymbol{x}_L - \boldsymbol{\mu}_L) \\
& + (\boldsymbol{x}_L - \boldsymbol{\mu}_L)'E^{-1}(\boldsymbol{x}_K - \boldsymbol{\mu}_K) \\
=& (\boldsymbol{x}_K - \boldsymbol{\mu}_K)'\Sigma_{K,K}^{-1}(\boldsymbol{x}_K - \boldsymbol{\mu}_K) \\
&\quad + [(\boldsymbol{x}_L - \boldsymbol{\mu}_L) - F'(\boldsymbol{x}_K - \boldsymbol{\mu}_K)]'E^{-1}[(\boldsymbol{x}_L - \boldsymbol{\mu}_L) - F'(\boldsymbol{x}_K - \boldsymbol{\mu}_K)]
\end{aligned}$$

Furthermore, according to Ex. 4 the determinant of Σ is given by

$$|\Sigma| = |\Sigma_{K,K}|\,|E|.$$

Finally, the density of $\boldsymbol{X}$ can be expressed as

$$\begin{aligned}
f(\boldsymbol{x}) =& \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}Q(\boldsymbol{x})\right\} \\
=& \frac{1}{(2\pi)^{k/2}(2\pi)^{l/2}|\Sigma_{K,K}|^{1/2}|E|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_K-\boldsymbol{\mu}_K)'\Sigma_{K,K}^{-1}(\boldsymbol{x}_K-\boldsymbol{\mu}_K)\right\} \\
& \times \exp\left\{-\frac{1}{2}[(\boldsymbol{x}_L-\boldsymbol{\mu}_L)-F'(\boldsymbol{x}_K-\boldsymbol{\mu}_K)]'E^{-1}[(\boldsymbol{x}_L-\boldsymbol{\mu}_L)-F'(\boldsymbol{x}_K-\boldsymbol{\mu}_K)]\right\} \\
=& f^{\boldsymbol{X}_K}(\boldsymbol{x}_K)\cdot f^{\boldsymbol{Y}}(\boldsymbol{x}_L)
\end{aligned}$$

where $\boldsymbol{Y} \sim N_l(\boldsymbol{\mu}_L + F'(\boldsymbol{x}_K - \boldsymbol{\mu}_K, E)$, which proves the proposition.

(b) If $\boldsymbol{X}_K$ and $\boldsymbol{X}_L$ are independent, then, for any $\boldsymbol{x}_L \in \mathbb{R}^l$, the conditional distribution of $\boldsymbol{X}_K$ given $\boldsymbol{X}_L = \boldsymbol{x}_L$ is equal to the unconditional distribution by definition. Therefore, we have

$$\mathrm{E}(\boldsymbol{X}_K|\boldsymbol{X}_L = \boldsymbol{x}_L) = \mathrm{E}(\boldsymbol{X}_K) = \boldsymbol{\mu}_K.$$

If, on the other hand $\mathrm{E}(\boldsymbol{X}_K|\boldsymbol{X}_L = \boldsymbol{x}_L) = \boldsymbol{\mu}_K$ for every $\boldsymbol{x}_L \in \mathbb{R}^l$, then, according to (a), it follows that

$$\boldsymbol{\mu}_K + \Sigma_{K,L}\Sigma_{L,L}^{-1}(\boldsymbol{x}_L - \boldsymbol{\mu}_L) = \boldsymbol{\mu}_K,$$

and therefore, $\Sigma_{K,L}\Sigma_{L,L}^{-1} = 0_{k\times k}$. Then, $\Sigma_{KK|L}$ simplifies to

$$\Sigma_{KK|L} = \Sigma_{K,K}$$

and, again according to (a), the conditional distribution of

$$\boldsymbol{X}_K|\boldsymbol{X}_L = \boldsymbol{x}_L \sim N(\boldsymbol{\mu}_K, \Sigma_{K,K}),$$

which is equal to the unconditional distribution. Consequently, $\boldsymbol{X}_K$ and $\boldsymbol{X}_L$ are independent.

(c) Applying the permutation matrix

$$A := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

we get

$$\tilde{X} = \begin{pmatrix} X_2 \\ X_1 \end{pmatrix} = AX \sim N_2(A\boldsymbol{\mu}, A\Sigma A'),$$

where

$$\begin{aligned}
A\boldsymbol{\mu} &= \begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix}, \\
A\Sigma A' &= \begin{pmatrix} \sigma_2^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_1^2 \end{pmatrix}.
\end{aligned}$$

Furthermore,

$$|\tilde{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2) > 0.$$

Applying part (a) we get

$$P^{X_2|X_1=t} = N_2(\mu_{2\cdot 1}, \Sigma_{22\cdot 1}),$$

where

$$\mu_{2\cdot 1} = \mu_2 + \sigma_1\sigma_2\rho\frac{1}{\sigma_1^2}(t-\mu_1) = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(t-\mu_1)$$

$$\Sigma_{22\cdot 1} = \sigma_2^2 - \sigma_1\sigma_2\rho\frac{1}{\sigma_1^2}\sigma_1\sigma_2\rho = \sigma_2^2 - \sigma_2^2\rho^2 = \sigma_2^2(1-\rho^2).$$

Exercise 8

Let $r = \text{rank}(\Sigma)$.
Since Σ is an orthogonal projector, there exists an ortho-projection matrix O with $\Sigma = OO'$ (see Ex. 2). Additionally, by assumption $\boldsymbol{\mu} \in Im(\Sigma)$, and therefore we have $\boldsymbol{\mu} = O\boldsymbol{b}$ for some $b \in \mathbb{R}^p$. Let $\boldsymbol{Y} \sim N_p(b, I_p)$. Then, according to Theorem I.2.11, we have $\boldsymbol{X} \sim O\boldsymbol{Y}$.

Since $O'O = \text{diag}(\underbrace{1,\dots,1}_{r \text{ times}}, 0, \dots, 0)$ is also an orthogonal projector, we get by Theorem I.3.5, that

$$\boldsymbol{X}'\boldsymbol{X} \sim \boldsymbol{Y}'O'O\boldsymbol{Y} \sim \chi^2(r, \boldsymbol{b}'O'O\boldsymbol{b}/2).$$

Finally, observe that $\boldsymbol{b}'O'O\boldsymbol{b} = \boldsymbol{\mu}'\boldsymbol{\mu}$.

Exercise 9

(a) By symmetry of A we have $BA' = BA = 0_{p\times p}$. Therefore, application of Theorem I.2.14 yields independence of $A\boldsymbol{X}$ and $B\boldsymbol{X}$.

Then, any transformation of $A\boldsymbol{X}$ is independent of any transformation of $B\boldsymbol{X}$. Denote A^+ the Moore-Penrose inverse of A, then, by symmetry of A

$$\boldsymbol{X}'A\boldsymbol{X} = \boldsymbol{X}'AA^+A\boldsymbol{X} = \boldsymbol{X}'A'A^+A\boldsymbol{X} = (A\boldsymbol{X})'A^+A\boldsymbol{X}.$$

Therefore, $\boldsymbol{X}'A\boldsymbol{X}$ is one such transformation and the proposition follows.

(b) By analogous arguments, if we denote by B^+ the Moore-Penrose inverse of B, we have

$$\boldsymbol{X}'B\boldsymbol{X} = (B\boldsymbol{X})'B^+B\boldsymbol{X}.$$

and therefore $\boldsymbol{X}'A\boldsymbol{X}$ and $\boldsymbol{X}'B\boldsymbol{X}$ as suitable transformations of $A\boldsymbol{X}$ and $B\boldsymbol{X}$ are independent.

Exercise 9

(a) We get

$$\begin{aligned}
\text{E}(\hat{\boldsymbol{\varepsilon}}) =& \text{E}\left[(I_p - B(B'B)^{-1}B')\boldsymbol{Y}\right] \\
=& (I_p - B(B'B)^{-1}B')\text{E}\left[B\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right] \\
=& (B - BI_p)\text{E}(\boldsymbol{\beta}) + (I_p - B(B'B)^{-1}B')\text{E}(\boldsymbol{\varepsilon}) \\
=& 0 + 0 = 0
\end{aligned}$$

(b) We get

$$\begin{aligned}\mathrm{Cov}(\hat{\boldsymbol{\varepsilon}}) =& \mathrm{Cov}\left[(I_p - B(B'B)^{-1}B')(B\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right]\\ =& \mathrm{Cov}\left[(I_p - B(B'B)^{-1}B')\boldsymbol{\varepsilon}\right]\\ =& (I_p - B(B'B)^{-1}B')\sigma^2(I_p - B(B'B)^{-1}B')'\end{aligned}$$

Since $(I_p - B(B'B)^{-1}B')$ is symmetric and idempotent this simplifies to

$$\mathrm{Cov}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2(I_p - B(B'B)^{-1}B').$$