TMA4267 Linear statistical models
Recommended exercises 3 – solutions

Problem 1 Simple calculations with the multivariate normal distribution

a) $3X_1 - 2X_2 + X_3 = (3\ {-2}\ 1)\boldsymbol{X}$, and we know from theory that $A\boldsymbol{X} + \boldsymbol{b}$ is multivariate normal if $\boldsymbol{X}$ is. In this case, $E\big((3\ {-2}\ 1)\boldsymbol{X}\big) = (3\ \ {-2}\ 1)E\boldsymbol{X} = (3\ {-2}\ 1)(2\ {-3}\ 1)^{\mathrm{T}} = 13$ and

$$\begin{aligned}\operatorname{Cov}\Big(\begin{pmatrix}3 & -2 & 1\end{pmatrix}\boldsymbol{X}\Big) &= \begin{pmatrix}3 & -2 & 1\end{pmatrix}(\operatorname{Cov}\boldsymbol{X})\begin{pmatrix}3 & -2 & 1\end{pmatrix}^{\mathrm{T}} \\ &= \begin{pmatrix}3 & -2 & 1\end{pmatrix}\begin{pmatrix}1 & 1 & 1\\ 1 & 3 & 2\\ 1 & 2 & 2\end{pmatrix}\begin{pmatrix}3\\ -2\\ 1\end{pmatrix} = 9.\end{aligned}$$

So $3X_1 - 2X_2 + X_3 \sim N(13, 9)$. (By the way, we already knew that it had a univariate normal distribution since it is a linear combination of the components of a multivariate normal vector.)

b) We may use general properties of covariances (see p. 125 in Härdle and Simar). Let $\boldsymbol{a} = (a_1\ a_3)^{\mathrm{T}}$. Then we want

$$\begin{aligned}0 = \operatorname{Cov}\left(X_2, X_2 - \boldsymbol{a}^{\mathrm{T}}\begin{pmatrix}X_1\\ X_3\end{pmatrix}\right) &= \operatorname{Cov}(X_2, X_2) + \operatorname{Cov}\left(X_2, -\boldsymbol{a}^{\mathrm{T}}\begin{pmatrix}X_1\\ X_3\end{pmatrix}\right) \\ = \operatorname{Var} X_2 - \operatorname{Cov}\left(X_2, \begin{pmatrix}X_1\\ X_3\end{pmatrix}\right)\boldsymbol{a} &= 3 - \begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}a_1\\ a_3\end{pmatrix} = 3 - a_1 - 2a_3,\end{aligned}$$

so any $\boldsymbol{a} = (3 - 2a_3\quad a_3)^{\mathrm{T}}$ will do. Since both variables are linear combinations of components of a multivariate normal vector, zero covariance implies independence.

Alternatively, we might note that

$$X_2 = \begin{pmatrix}0 & 1 & 0\end{pmatrix}\boldsymbol{X} \qquad \text{and} \qquad X_2 - \boldsymbol{a}^{\mathrm{T}}\begin{pmatrix}X_1\\ X_3\end{pmatrix} = \begin{pmatrix}-a_1 & 1 & -a_3\end{pmatrix}\boldsymbol{X}$$

and require

$$\begin{aligned}0 = \operatorname{Cov}\left(X_2, X_2 - \boldsymbol{a}^{\mathrm{T}}\begin{pmatrix}X_1\\ X_3\end{pmatrix}\right) &= \operatorname{Cov}\Big(\begin{pmatrix}0 & 1 & 0\end{pmatrix}\boldsymbol{X}, \begin{pmatrix}-a_1 & 1 & -a_3\end{pmatrix}\boldsymbol{X}\Big) \\ = \begin{pmatrix}0 & 1 & 0\end{pmatrix}(\operatorname{Cov}\boldsymbol{X})\begin{pmatrix}-a_1\\ 1\\ -a_3\end{pmatrix} &= \begin{pmatrix}0 & 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1 & 1\\ 1 & 3 & 2\\ 1 & 2 & 2\end{pmatrix}\begin{pmatrix}-a_1\\ 1\\ -a_3\end{pmatrix} = -a_1 + 3 - 2a_3,\end{aligned}$$

c) We need to partition $\boldsymbol{x}$, the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ of $\boldsymbol{X}$ into

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{\mathrm{A}} \\ \boldsymbol{x}_{\mathrm{B}} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathrm{A}} \\ \boldsymbol{\mu}_{B} \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Sigma_{\mathrm{AA}} & \Sigma_{\mathrm{AB}} \\ \Sigma_{\mathrm{BA}} & \Sigma_{\mathrm{BB}} \end{pmatrix},$$

where the A-parts relates to X_1 and the B-parts to $(X_2\ X_3)^{\mathrm{T}}$. The the conditional distribution of X_1 given $X_2 = x_2$ and $X_3 = x_3$ is

$$N\Big(\boldsymbol{\mu}_{\mathrm{A}} + \Sigma_{\mathrm{AB}}\Sigma_{\mathrm{BB}}^{-1}(\boldsymbol{x}_{\mathrm{B}} - \boldsymbol{\mu}_{\mathrm{B}}), \Sigma_{\mathrm{AA}} - \Sigma_{\mathrm{AB}}\Sigma_{\mathrm{BB}}^{-1}\Sigma_{\mathrm{BA}}\Big).$$

We find

$$\Sigma_{\mathrm{AB}}\Sigma_{\mathrm{BB}}^{-1} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} \end{pmatrix},$$

$$\boldsymbol{\mu}_{\mathrm{A}} + \Sigma_{\mathrm{AB}}\Sigma_{\mathrm{BB}}^{-1}(\boldsymbol{x}_{\mathrm{B}} - \boldsymbol{\mu}_{\mathrm{B}}) = 2 + \begin{pmatrix} 0 & \frac{1}{2} \end{pmatrix} \left(\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} -3 \\ 1 \end{pmatrix} \right) = 2 + \frac{1}{2}(x_3 - 1) = \frac{1}{2}x_3 + \frac{3}{2},$$

$$\Sigma_{\mathrm{AA}} - \Sigma_{\mathrm{AB}}\Sigma_{\mathrm{BB}}^{-1}\Sigma_{\mathrm{BA}} = 1 - \begin{pmatrix} 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2},$$

so that the conditional distribution is univariate $N(x_3/2 + 3/2, 1/2)$.

Problem 2 From correlated to independent variables

a) Using the relevant entries of Σ, we find the correlations

$$\mathrm{Corr}(X_1, X_3) = \frac{\mathrm{Cov}(X_1, X_3)}{\sqrt{(\mathrm{Var}\, X_1)(\mathrm{Var}\, X_3)}} = \frac{1}{\sqrt{1 \cdot 3}} = \frac{\sqrt{3}}{3} \approx 0.5774,$$

$$\mathrm{Corr}(X_2, X_3) = \frac{\mathrm{Cov}(X_2, X_3)}{\sqrt{(\mathrm{Var}\, X_2)(\mathrm{Var}\, X_3)}} = \frac{-1}{\sqrt{2 \cdot 3}} = -\frac{\sqrt{6}}{6} \approx -0.4082,$$

so X_3 is most correlated with X_3.

We note that $\boldsymbol{Z} = A\boldsymbol{X}$, where

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

a linear transformation of the trivariate normal vector $\boldsymbol{X}$, which means that $\boldsymbol{Z}$ is multivariate, in this case bivariate, normal. The mean vector and covariance matrix of $\boldsymbol{Z}$

are

$$E\boldsymbol{Z} = A\boldsymbol{\mu} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 & 2 \end{pmatrix},$$

$$\operatorname{Cov} \boldsymbol{Z} = A\Sigma A^{\mathrm{T}} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}.$$

b)

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{e}_1^{\mathrm{T}} \\ \boldsymbol{e}_2^{\mathrm{T}} \end{pmatrix} \boldsymbol{X},$$

so $\boldsymbol{Y}$ is a trivariate normal vector premultiplied by a 2×3 matrix, that is, bivariate normal.

$$\begin{aligned} \operatorname{Cov} \boldsymbol{Y} &= \operatorname{Cov}\left(\begin{pmatrix} \boldsymbol{e}_1^{\mathrm{T}} \\ \boldsymbol{e}_2^{\mathrm{T}} \end{pmatrix} \boldsymbol{X} \right) = \begin{pmatrix} \boldsymbol{e}_1^{\mathrm{T}} \\ \boldsymbol{e}_2^{\mathrm{T}} \end{pmatrix} \Sigma \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{e}_1^{\mathrm{T}}\Sigma\boldsymbol{e}_1 & \boldsymbol{e}_1^{\mathrm{T}}\Sigma\boldsymbol{e}_2 \\ \boldsymbol{e}_2^{\mathrm{T}}\Sigma\boldsymbol{e}_1 & \boldsymbol{e}_2^{\mathrm{T}}\Sigma\boldsymbol{e}_2 \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{e}_1 & \lambda_2\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{e}_2 \\ \lambda_1\boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{e}_1 & \lambda_2\boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{e}_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \end{aligned}$$

since $\Sigma\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$ and the eigenvalues constitute an orthonormal set, showing that $\operatorname{Cov}(Y_1, Y_2) = 0$. Since Y_1 and Y_2 are components of a bivariate normal vector, they are independent. (You could also argue that $(\boldsymbol{e}_1^{\mathrm{T}}\ \boldsymbol{e}_2^{\mathrm{T}})^{\mathrm{T}}\Sigma(\boldsymbol{e}_1\ \boldsymbol{e}_2)$ is an upper-left submatrix of $(\boldsymbol{e}_1^{\mathrm{T}}\ \boldsymbol{e}_2^{\mathrm{T}}\ \boldsymbol{e}_3^{\mathrm{T}})^{\mathrm{T}}\Sigma(\boldsymbol{e}_1\ \boldsymbol{e}_2\ \boldsymbol{e}_3)$, which is diagonal by orthogonal diagonalization.)

The total variance of $\boldsymbol{X}$ is defined as the sum of the variances of its components, that is, as $\operatorname{tr}\Sigma$. By theory, the trace is equal to the sum of the eigenvalues, so the total variance is $\lambda_1 + \lambda_2 + \lambda_3$. The total variance of $(\boldsymbol{e}_1^{\mathrm{T}}\ \boldsymbol{e}_2^{\mathrm{T}}\ \boldsymbol{e}_3^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{X}$ is also $\lambda_1 + \lambda_2 + \lambda_3$ by an argument as the one above, or by theory for principal components. For $\boldsymbol{Y} = (\boldsymbol{e}_1^{\mathrm{T}}\ \boldsymbol{e}_2^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{X}$, the total variance is $\lambda_1 + \lambda_2$. The proportion of the total variance explained by $\boldsymbol{Y}$ is then $(\lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2 + \lambda_3) \approx 0.922$ (see R output for eigenvalues).

(Note that the components of $\boldsymbol{Y}$ are the two first (theoretical) principal components of $\boldsymbol{X}$.)

Problem 3

a) By the formula $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)^{-1} = \frac{1}{ad-ba}\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$,

$$\Sigma^{-1} = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)} \begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}.$$

$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is the quadratic form defined by Σ^{-1} evaluated in $\boldsymbol{x}-\boldsymbol{\mu} = (x-\mu_X \;\; y-\mu_Y)^{\mathrm{T}}$,

$$
\begin{aligned}
&(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\
&\quad = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\Big(\sigma_Y^2(x-\mu_X)^2 + \sigma_X^2(y-\mu_Y)^2 - 2\rho\sigma_X^2\sigma_Y^2(x-\mu_X)(y-\mu_Y)\Big) \\
&\quad = \frac{1}{1-\rho^2}\bigg(\Big(\frac{x-\mu_X}{\sigma_X}\Big)^2 + \Big(\frac{y-\mu_Y}{\sigma_Y}\Big)^2 - 2\rho\Big(\frac{x-\mu_X}{\sigma_X}\Big)\Big(\frac{y-\mu_Y}{\sigma_Y}\Big)\bigg) = Q(x,y).
\end{aligned}
$$

b) Using the results from (a), we get

$$f(x,y) = f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{\det\Sigma}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}.$$

c) $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d^2$ if and only if $f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{\det\Sigma}}e^{-\frac{1}{2}d^2}$, and the right hand side of the last equation is a constant, so the equation defines a contour of f. Also, any contour of f can be expressed this way, since $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is always non-negative (the quadratic form is positive semidefinite).

Let P be an orthogonal matrix such that $P^{\mathrm{T}}\Sigma P = \Lambda$ is diagonal, with the eigenvalues λ_1 and λ_2 of Σ on the diagonal. Then $P^{\mathrm{T}}\Sigma^{-1}P = \Lambda^{-1}$. Make a change of variable, $P\boldsymbol{y} = \boldsymbol{x}-\boldsymbol{\mu}$ (translation followed by an orthogonal change of variable), so that $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{y}^{\mathrm{T}}P^{\mathrm{T}}\Sigma^{-1}P\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{y}$. Assume $\boldsymbol{y} = (y_1\; y_2)^{\mathrm{T}}$. The contour is the graph of $\boldsymbol{y}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{y} = d^2$, or $y_1^2/\lambda_1 + y_2^2/\lambda_2 = d^2$, that is,

$$\frac{y_1^2}{\lambda_1 d^2} + \frac{y_2^2}{\lambda_2 d^2} = 1.$$

In the y_1–y_2-coordinate system, this is an ellipse with centre in the origin and axes along the coordinate axes, with half-lengths $\sqrt{\lambda_1}d$ and $\sqrt{\lambda_2}d$. In the original coordinate system, the centre is in $P\mathbf{0}+\boldsymbol{\mu} = \boldsymbol{\mu}$. The axes has directions $P(1\;0)^{\mathrm{T}}$ and $P(0\;1)^{\mathrm{T}}$, that is, the eigenvectors given by the columns of P (corresponding to λ_1 and λ_2, respectively).

d) In the case that $\sigma_X = \sigma_Y = \sigma$, $\Sigma = \sigma^2\left(\begin{smallmatrix}1 & -\rho\\ -\rho & 1\end{smallmatrix}\right)$, we find the eigenvalues, e.g. by solving the characteristic equation, $\det(\lambda I - \Sigma) = 0$, to be $\sigma^2(1\pm\rho)$. Eigenvectors can be found by solving $(\lambda I - \Sigma)\mathbf{x} = \mathbf{0}$ for $\mathbf{x}$, where λ is an eigenvalue. We find linearly independent, normalized eigenvectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \pm 1\end{pmatrix}$$

(they also work for $\rho = 0$, when σ^2 is the only eigenvalue). Note that the directions of the two axes of a contour ellipse do not depend on ρ and are at 45° with the coordinate axes. If $\rho > 0$, the major axis is in the direction of $(1\;1)^{\mathrm{T}}$, and if $\rho < 0$, the major axis is in the direction of $(1\;-1)^{\mathrm{T}}$. The ratio of the half-length of the major axis to that of the minor axis will grow when ρ approaches 1 or -1, thus the eccentricity of the ellipse will increase. For $\rho = 0$, both axes have the same length, and the contour is a circle.

Problem 4 Normal marginals, but not multivariate normal

a) By the law of total probability,

$$\begin{aligned} P(Z \leq z) &= P(Z \leq z \mid XY \geq 0)P(XY \geq 0) + P(Z \leq z \mid XY < 0)P(XY < 0) \\ &= \frac{1}{2}P(X \leq z) + \frac{1}{2}P(-X \leq z) = \frac{1}{2}P(X \leq z) + \frac{1}{2}P(X \leq z) = P(X \leq z), \end{aligned}$$

so Z has the same cdf as X, and is thus $N(0, 1)$.

b) $(Y \ Z)^{\mathrm{T}}$ cannot have the bivariate normal distribution, because Y and Z always have the same sign, which follows from the definition by inspection of the possible signs of X and Y.