TMA4267 Linear statistical models
Recommended exercises 3

Problem 1 Simple calculations with the multivariate normal distribution

Let $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}\right)$.

a) Find the distribution of $3X_1 - 2X_2 + X_3$.

b) Find a 2×1 vector $\boldsymbol{a}$ such that X_2 and $X_2 - \boldsymbol{a}^{\mathrm{T}} \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$ are independent.

c) Find the conditional distribution of X_1 given $X_2 = x_2$ and $X_3 = x_3$.

Problem 2 From correlated to independent variables

(Exam TMA4267, May 2013, Problem 1, slightly modified)

Assume that the random vector $\boldsymbol{X} = (X_1\ X_2\ X_3)^{\mathrm{T}}$ has a trivariate normal distribution with mean vector $\boldsymbol{\mu} = (2\ 6\ 4)^{\mathrm{T}}$ and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{pmatrix}.$$

a) Find out which of the random variables X_1 and X_2 is most correlated (in absolute value) with X_3. What is the distribution of the random vector $\boldsymbol{Z} = (X_2 - X_1 \quad X_3 - X_1)^{\mathrm{T}}$?

A company is measuring three quality characteristics in order to control the quality of a product. Their respective random variables can be arranged in a random vector $\boldsymbol{X} = (X_1\ X_2\ X_3)^{\mathrm{T}}$. Based on previous experience, it is reasonable to assume that $\boldsymbol{X}$ is trivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix Σ, as given above.

The company would like to have a simplified quality control procedure where they only consider a bivariate random vector instead of a trivariate one. The eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, of Σ, as well as their respective eigenvectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$, are given below as R output.

```
$values
[1] 3.8793852 1.6527036 0.4679111

$vectors
           [,1]       [,2]       [,3]
[1,] -0.2931284 -0.4490988  0.8440296
[2,]  0.4490988 -0.8440296 -0.2931284
[3,] -0.8440296 -0.2931284 -0.4490988
```

b) Define the bivariate vector $\boldsymbol{Y} = (\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{X} \;\; \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{X})^{\mathrm{T}}$. Why does $\boldsymbol{Y}$ have a bivariate normal distribution?

Show that $Y_1 = \boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{X}$ and $Y_2 = \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{X}$ are independent. How much of the total variance in $\boldsymbol{X}$ is explained by $\boldsymbol{Y}$? Hint: The total variance is the trace of the covariance matrix, that is, the sum of the variances. Also, there is a relationship between the trace and eigenvalues of a matrix – which relationship?

Problem 3 The bivariate normal distribution

Let X and Y be random variables with joint pdf f parameterized by $(\mu_X, \mu_Y, \rho, \sigma_X^2, \sigma_Y^2)$,

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x,y)}, \quad \text{where}$$
$$Q(x,y) = \frac{1}{1-\rho^2}\Big(\Big(\frac{x-\mu_X}{\sigma_X}\Big)^2 + \Big(\frac{y-\mu_Y}{\sigma_Y}\Big)^2 - 2\rho\Big(\frac{x-\mu_X}{\sigma_X}\Big)\Big(\frac{y-\mu_Y}{\sigma_Y}\Big)\Big).$$

This is the bivariate normal distribution.

a) Show that $Q(x,y)$ can be written $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, where $\boldsymbol{x} = (x \; y)^{\mathrm{T}}$, $\boldsymbol{\mu} = (\mu_X \; \mu_Y)^{\mathrm{T}}$ and

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \operatorname{Var} X & \operatorname{Cov}(X,Y) \\ \operatorname{Cov}(Y,X) & \operatorname{Var} Y \end{pmatrix}.$$

Remark: Σ is called the variance–covariance matrix of $\boldsymbol{X}$.

b) Rewrite f using $\boldsymbol{\mu}$ and Σ. Note that $\det\Sigma = \sigma_X^2\sigma_Y^2(1-\rho^2)$.

Now we turn to look at contours of f, that is, curves in the xy-plane along which f has a constant value.

c) Why can the contours be seen as solutions to the equation $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d^2$ for a given constant d?

Use the method of diagonalization (spectral decomposition) to explain that the contours are ellipses with centre in $\boldsymbol{\mu}$, axes in the direction of the eigenvectors of Σ, with half-lengths $\sqrt{\lambda_1}d$ and $\sqrt{\lambda_2}d$ of the axes, where λ_1 and λ_2 are the eigenvalues of Σ.

Remark: Remember that there is a simple connection between the eigenvectors and eigenvalues of Σ and Σ^{-1}.

d) For the special case that $\sigma_X = \sigma_Y$, find the eigenvalues and eigenvectors of Σ. Make a drawing of some contours by hand.

e) Now we turn to R to draw ellipses. This can be done using

```
install.packages("ellipse") # if not already installed
library(ellipse)
```

First look at $\mu_X = \mu_Y = 1$, and $\sigma_X = \sigma_Y = 1$ and $\rho = 0.5$.

```
mu1 <- mu2 <- 1
sigma1 <- 1
sigma2 <- 1
rho <- 0.5
plot(ellipse(rho, scale=c(sigma1, sigma2), centre=c(mu1, mu2)), type = "l")
```

Try varying the parameters in Σ and observe.

Problem 4 Normal marginals, but not multivariate normal

Let X and Y be independent standard normally distributed variables, and define

$$Z = \begin{cases} X & \text{if } XY \geq 0, \\ -X & \text{if } XY < 0. \end{cases}$$

a) Show that Z has the standard normal distribution.

b) Show that $(Y\ Z)^{\mathrm{T}}$ does not have the bivariate normal distribution. Hint: Consider the signs of Y and Z.