TMA4267 Linear Statistical Models
Compulsory Exercise Part 1: Multivariate random variables and the multivariate normal distribution

Due date: 3 February 2017 at 6pm
Upload pdf-file with handwritten solutions in Blackboard

Text in red added on January 27.

Problem 1

Assume that $\boldsymbol{X}$ is a bivariate normal random variable with $\mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ and $\mathrm{Cov}(\boldsymbol{X}) = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Let $\boldsymbol{C} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ and $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \boldsymbol{C}\boldsymbol{X}$.

a) Find the mean vector and covariance matrix of $\boldsymbol{Y}$.
What is the distribution of $\boldsymbol{Y}$.
Are Y_1 and Y_2 independent random variables?

b) Let $f(\boldsymbol{x})$ be the joint pdf for $\boldsymbol{X}$. Contours of $f(\boldsymbol{x})$ are the solutions to $f(\boldsymbol{x}) = a$ for some constant $a > 0$, or equivalently to

$$(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = b \qquad (1)$$

for a corresponding constant $b > 0$. In Figure 1 you see the graphical solution (black ellipse) to Equation (1) for $b = 4.6$.

Explain what is the connections between properties of the covariance matrix $\boldsymbol{\Sigma}$, the chosen value of b and features of the figure (e.g. principal axes and half-lengths). Please make a drawing and/or mark where these features are on the printed figure.
What is the probability that a random observation $\boldsymbol{X}$ from $f(\boldsymbol{x})$ will fall within the given ellipse?
The following information might be useful:

```
> sigma=matrix(c(3,1,1,3),ncol=2)
> sigma
     [,1] [,2]
[1,]    3    1
[2,]    1    3
> eigen(sigma)
$values
[1] 4 2
$vectors
           [,1]       [,2]
```

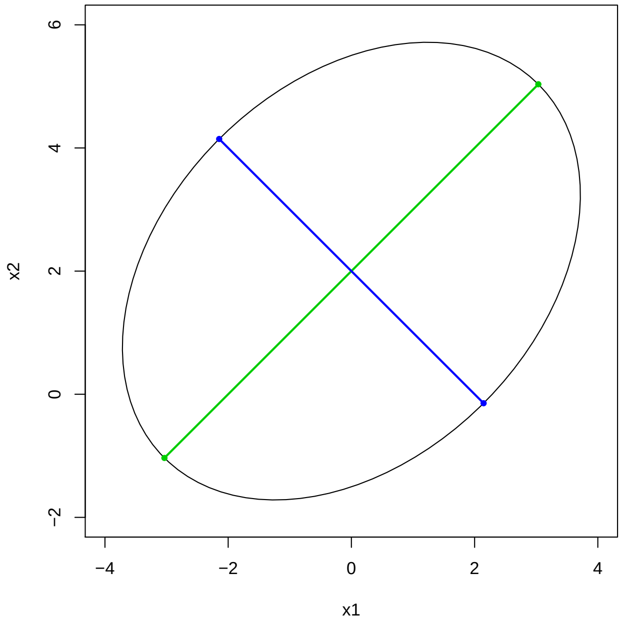

Figure 1: Graphical solution to Equation (1).

```
[1,] 0.7071068 -0.7071068
[2,] 0.7071068  0.7071068
> qchisq(0.9,2)
[1] 4.60517
```

Problem 2

Let $X_1, X_2, \dots, X_n$ be a (univariate) random sample from some population. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Futher, let $\mathbf{1}$ be an n-dimensional vector of 1s. Then $\mathbf{1}\mathbf{1}^T$ is an $n \times n$ matrix of 1s. The matrix $\boldsymbol{R} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is called the *centering matrix*.

$$\boldsymbol{R} = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix}$$

a) Show that $\bar{X} = \boldsymbol{A}\boldsymbol{X}$ where $\boldsymbol{A} = \frac{1}{n}\mathbf{1}^T$, and that $S^2 = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{R}\boldsymbol{X}$.
Hint: start by looking at the ith element of $\boldsymbol{R}\boldsymbol{X}$, and observe that $\boldsymbol{R}$ is symmetric and idempotent.

Now, in addition assume that the random sample is taken from the univariate normal distribution with mean μ and variance σ^2. In the notation of TMA4267 we would let $(X_1, X_2, \cdots, X_n)^T = \boldsymbol{X}$, and write $\boldsymbol{X} \sim N_n(\mathbf{1}\mu, \sigma^2\boldsymbol{I})$.

In our first statistics course we met the T-statistic

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

and were told that the T-statistic followed a t-distribution with parameter $n-1$ (named "degrees of freedom"), however we did not show that $\bar{X}$ in the numerator and S^2 in the denominator were independent (as is required for the t-distribution). With our new skills on the multivariate normal distribution we will do that now.

b) Show that $\boldsymbol{A}\boldsymbol{R} = \mathbf{0}$. What does this imply about $\boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{R}\boldsymbol{X}$? How can you use this to conclude that $\bar{X}$ og S^2 are independent?

Finally, we have worked with quadratic forms, see Fährmeir, Kneib, Lang and Marx (2013): Theorem B.8 on page 651. In the next question Theorem B.8.2 is useful: Let $\boldsymbol{R}$ be a symmetric and idempotent matrix with rank r, and $\boldsymbol{Y} \sim N_p(\mathbf{0}, \boldsymbol{I})$, then $\boldsymbol{Y}^T\boldsymbol{R}\boldsymbol{Y} \sim \chi^2_r$.

c) Use the fact that $S^2 = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{R}\boldsymbol{X}$, where $\boldsymbol{R}$ is symmetric and idempotent, to derive the distribution of $\frac{(n-1)S^2}{\sigma^2}$. Hint: a similar problem is given in the TMA4267V2014 exam, Problem 1b.