TMA4267 - Compulsory exercise 1

Anders Fagerli

February 26, 2019

1 Problem 1

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?

$$\boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \Sigma = \mathrm{Cov}(\boldsymbol{X}) = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \boldsymbol{A}\boldsymbol{X}, \quad \boldsymbol{A} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

Mean vector

$$\begin{aligned} E(\boldsymbol{Y}) = E(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{\mu} &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{bmatrix} \\ &= \underline{\underline{\begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix}}} \end{aligned}$$

Covariance matrix

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X})\boldsymbol{A}^T = \boldsymbol{A}\Sigma\boldsymbol{A}^T &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{2} & \frac{2}{\sqrt{2}} \\ -\sqrt{2} & \frac{2}{\sqrt{2}} \end{pmatrix} \\ &= \underline{\underline{\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}}} \end{aligned}$$

$\boldsymbol{Y}$ is a linear combination of a bivariate normal variable, and is thus also <u>bivariate normal</u>. This, together with the result $\mathrm{Cov}(Y_1, Y_2) = 0$, means that Y_1 and Y_2 are <u>independent variables</u>.

b) Explain the connections between the covariance matrix Σ, the chosen value of b and features of the ellipse (e.g. principal axes and their half-lengths). Mark these features on the figure (make a drawing or use the printed figure). What is the probability that $\boldsymbol{X}$ falls within the given ellipse?

A contour of f is an ellipse given by the exponent in the pdf,

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

This is an equation describing an ellipse with half-lengths,

$$l_1 = \sqrt{\lambda_1 b}, \quad l_2 = \sqrt{\lambda_2 b},$$

where λ_1 and λ_2 are the eigenvalues of the covariance matrix Σ, while the principle axis are given by the eigenvectors u_1 and u_2 of Σ. To see this, Equation 1 must be written in the form of an ellipse, described by the equation,

$$1 = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2}, \tag{2}$$

where x_1 and x_2 are the principal axis, and a and b are the half-lengths.

We start with the eigendecompostition of the symmetric and positive definite Σ, written as $\Sigma = P\Lambda P^T$, where P is a matrix consisting of orthogonal eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ of Σ, and Λ is a diagonal matrix whose entries are the eigenvalues λ_1 and λ_2 of Σ. The eigendecomposition of Σ can be calculated to,

$$\Sigma = P\Lambda P^T = \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

The exponent of f can then be written in terms of the eigendecomposition $\Sigma^{-1} = P\Lambda^{-1}P^T$.

$$\begin{aligned} b &= (\boldsymbol{x} - \boldsymbol{\mu})^T P \Lambda^{-1} P^T (\boldsymbol{x} - \boldsymbol{\mu}) \\ &= (\boldsymbol{x} - \boldsymbol{\mu})^T \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} \Lambda^{-1} \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} (\boldsymbol{x} - \boldsymbol{\mu}) \\ &= (\begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} (\boldsymbol{x} - \boldsymbol{\mu}))^T \Lambda^{-1} \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} (\boldsymbol{x} - \boldsymbol{\mu}) \end{aligned}$$

We use the substitution $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix}^T (\boldsymbol{x} - \boldsymbol{\mu})$,

$$\begin{aligned} b &= \boldsymbol{y}^T \Lambda^{-1} \boldsymbol{y} \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{pmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \frac{y_1}{\lambda_1} \\ \frac{y_2}{\lambda_2} \end{bmatrix} \\ &= \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} \end{aligned}$$

$$\Rightarrow$$

$$1 = \frac{y_1^2}{\lambda_1 b} + \frac{y_2^2}{\lambda_2 b} \tag{3}$$

We can then see from Equation 3 that the half-lengths of $(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ are $\sqrt{\lambda_1 b}$ and $\sqrt{\lambda_2 b}$, with the centre of the ellipse at the origin in the y_1, y_2-coordinate system. Substituting back to the original coordinate system, we can see that the principle axis are shifted by $(\boldsymbol{x}-\boldsymbol{\mu})$ where $\boldsymbol{\mu} = \begin{bmatrix} 0 & 2 \end{bmatrix}^T$, resulting in the center $(0, 2)$. The directions of the axis can also be seen from the substitution, given by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. The connection between the covariance matrix Σ and the features of the ellipse is thus given by the eigendecomposition of Σ, where the eigenvalues correspond to half-lengths and the eigenvectors correspond to the principle axis. We can also see that an increasing value of b results in a larger ellipse, meaning the given contour is located at a smaller value for $f(\boldsymbol{x})$.

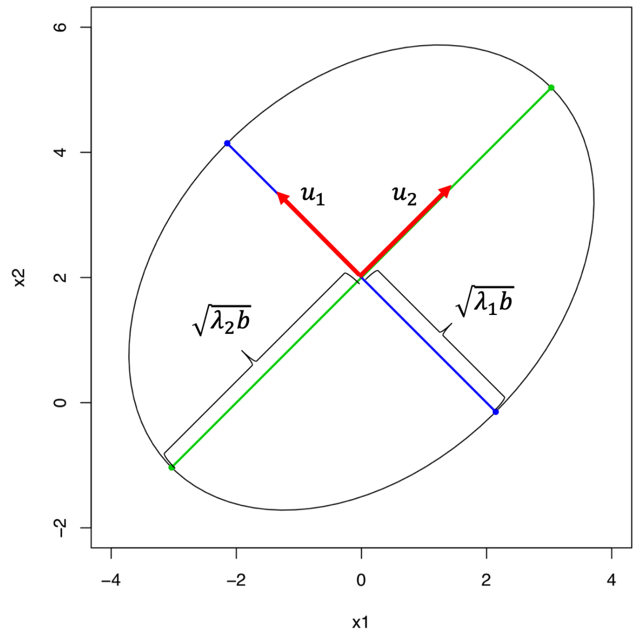

Figure 1: Ellipse with marked features

To calculate the probability of $\boldsymbol{X}$ falling within the given ellipse we must know the distribution of $(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. By **Theorem 4.7**(1, p. 149),

If $X \sim N_p(\mu, \Sigma)$, then the variable $U = (X-\mu)^T\Sigma^{-1}(X-\mu)$ has a χ^2_p distribution.

The random variable U describing the ellipse is therefore χ^2-distributed with two degrees of freedom. The probability of $\boldsymbol{X}$ falling *within* the ellipse accounts to calculating the probability,

$$P(U > b) = 1 - P(U \leqslant b) = 1 - P(U \leqslant 4.6)$$

This can be calculated by looking up values from known tables of the χ^2-distribution, or by using R. We see from the given R code that the χ^2-statistic for a probability of 90% and 2 degrees of freedom equals $4.60517 \approx b$. The probability of X falling withing the given ellipse is therefore $\underline{\underline{90\%}}$.

2 Problem 2

a) Show that $\bar{X} = \frac{1}{n}\mathbf{1}^T\boldsymbol{X}$ and that $S^2 = \frac{1}{n-1}\boldsymbol{X}^T C\boldsymbol{X}$.

$$
\begin{aligned}
\frac{1}{n}\mathbf{1}^T\boldsymbol{X} &= \frac{1}{n}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\begin{bmatrix}X_1\\ X_2\\ \vdots\\ X_n\end{bmatrix}\\
&= \frac{1}{n}(X_1 + X_2 + \cdots + X_n)\\
&= \frac{1}{n}\sum_{i=1}^{n} X_i\\
&= \underline{\underline{\bar{X}}}
\end{aligned}
$$

The centering matrix C has the effect of subtracting the mean from each component in the vector it is multiplied with, in this case the random vector $\boldsymbol{X}$.

$$
C\boldsymbol{X} = (\boldsymbol{X} - \bar{X}) = \begin{bmatrix}X_1 - \bar{X}\\ X_2 - \bar{X}\\ \vdots\\ X_n - \bar{X}\end{bmatrix}
$$

We see from the definition of S^2 that this term is squared and summed over all n observations. In matrix form, we then consider

$$
\begin{aligned}
\frac{1}{n-1}(\boldsymbol{X} - \bar{X})^T(\boldsymbol{X} - \bar{X}) &= \frac{1}{n-1}(C\boldsymbol{X})^T(C\boldsymbol{X})\\
&= \frac{1}{n-1}\boldsymbol{X}^T CC\boldsymbol{X}\\
&= \frac{1}{n-1}\boldsymbol{X}^T C^2\boldsymbol{X}, \qquad \text{C idempotent}\\
&= \frac{1}{n-1}\boldsymbol{X}^T C\boldsymbol{X}\\
&= \underline{\underline{S^2}}
\end{aligned}
$$

b) Show that $\frac{1}{n}\mathbf{1}^T C = \mathbf{0}^T$. What does this imply about $\frac{1}{n}\mathbf{1}^T\boldsymbol{X}$ and $C\boldsymbol{X}$? How can you use this to conclude that $\bar{X}$ and S^2 are independent?

$$
\begin{aligned}
\frac{1}{n}\mathbf{1}^T C &= \frac{1}{n}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\begin{pmatrix}1-\frac{1}{n} & \frac{-1}{n} & \dots & \frac{-1}{n}\\ \frac{-1}{n} & 1-\frac{1}{n} & \dots & \frac{-1}{n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{-1}{n} & \frac{-1}{n} & \dots & 1-\frac{1}{n}\end{pmatrix}\\
&= \frac{1}{n}\left[1-\sum_{i=1}^{n}\frac{1}{n},\quad 1-\sum_{i=1}^{n}\frac{1}{n},\dots,1-\sum_{i=1}^{n}\frac{1}{n}\right]\\
&= \frac{1}{n}\left[1-\frac{n}{n},\quad 1-\frac{n}{n},\dots,1-\frac{n}{n}\right]\\
&= \underline{\underline{\mathbf{0}^T}}
\end{aligned}
$$

We can show independence between $\frac{1}{n}\mathbf{1}^T\boldsymbol{X}$ and $C\boldsymbol{X}$ by **Corollary 5.2**(1, p. 185),

> *If $X \sim N_p(\mu, \Sigma)$ and given some matrices A and B, then $A\boldsymbol{X}$ and $B\boldsymbol{X}$ are independent if and only if $A\Sigma B^T = 0$.*

Choose $A = \frac{1}{n}\mathbf{1}^T$ and $B = C$.

$\Rightarrow$

$$
A\Sigma B^T = \frac{1}{n}\mathbf{1}^T\Sigma C = \frac{1}{n}\mathbf{1}^T\sigma^2 IC = \frac{1}{n}\mathbf{1}^T C\sigma^2 = \mathbf{0}^T,
$$

so $\frac{1}{n}\mathbf{1}^T\boldsymbol{X}$ and $C\boldsymbol{X}$ are independent. Since S^2 is a function of $C\boldsymbol{X}$ and we proved independence between $\bar{X}$ and $C\boldsymbol{X}$, $\bar{X}$ and S^2 are also independent.

c) Derive the distribution of $(n-1)S^2/\sigma^2$.

From **Theorem B.8.2**(2, p. 651),

> *Let $X \sim N_p(0, I), B$ an $(n \times p)$-matrix $(n \leq p)$ and R a symmetric idempotent $(p \times p)$-matrix with $rk(R) = r$. Then $X^T RX \sim \chi^2_p$.*

In order to use the above theorem, the matrices on either side of the idempotent matrix need to be standardized. We start by writing out the terms in $(n-1)S^2/\sigma^2$.

$$
\begin{aligned}
(n-1)S^2/\sigma^2 &= \frac{n-1}{\sigma^2}\cdot\frac{1}{n-1}\boldsymbol{X}^T C\boldsymbol{X}\\
&= \frac{1}{\sigma^2}(C\boldsymbol{X})^T C(C\boldsymbol{X})\\
&= \frac{1}{\sigma}(C\boldsymbol{X})^T C(C\boldsymbol{X})\frac{1}{\sigma}
\end{aligned}
$$

Denote $\boldsymbol{Y} = (C\boldsymbol{X})/\sigma$, so $(n-1)S^2/\sigma^2 = \boldsymbol{Y}^T C\boldsymbol{Y}$.

$$
\begin{aligned}
\boldsymbol{Y} = (C\boldsymbol{X})/\sigma &= \frac{1}{\sigma}\begin{bmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{bmatrix} \\
&= \begin{bmatrix} (X_1 - \bar{X})/\sigma \\ (X_2 - \bar{X})/\sigma \\ \vdots \\ (X_n - \bar{X})/\sigma \end{bmatrix}
\end{aligned}
$$

We now notice that every element of $\boldsymbol{Y}$ is shifted by the mean value of the sample and divided by the standard deviation. Each Y_j is therefore standardized, so $\boldsymbol{Y} \sim N_p(0, I)$. C is an idempotent and symmetric matrix, and will therefore have rank equal to its trace. $\text{Rank}(C) = n(1 - \frac{1}{n}) = n - 1$. **Theorem B.8.2** can now be used to conclude that $\underline{\underline{(n-1)S^2/\sigma^2 \text{ has a } \chi^2_{n-1}\text{-distribution}}}$.

References

[1] Wolfgang Karl Härdle, Léopold Simar. *Applied Multivariate Statistical Analysis, Fourth Edition.*

[2] Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, Brian Marx. *Regression*