TMA4267 - Linear Statistical Models
Solutions to Recommended Exercise 2 - V2017

12 January 2017

Problem 1: Simple calculations with the multivariate normal distribution

Let $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$

a) $\boldsymbol{X}$ is multivariate normal, and thus each element of $\boldsymbol{X}$ is univariate normal. $Y = 3X_1 - 2X_2 + X_3$. Y must be univariate normally distributed since it is a linear combination of univariate normal variables.

$$\mathrm{E}(Y) = 3\mu_1 - 2\mu_2 + \mu_3 = 3 \cdot 2 - 2 \cdot (-3) + 1 = 13$$

$$\mathrm{Cov}(Y) = \begin{bmatrix} 3 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = 9$$

b) Let $\boldsymbol{a} = (a_1, a_3)^T$, and define $Z = X_2 - a_1 X_1 - a_3 X_3$. We want to find (a_1, a_3) so that X_2 and Z are independent. Since the vector $(X_2, Z)^T$ is constructed from linear combinations of multivariate normal variables, the vector is multivariate normal. For multivariate normal data independence is achieved when $\mathrm{Cov}(X_2, Z) = 0$.

$$\begin{aligned}
\mathrm{Cov}(X_2, Z) &= \mathrm{Cov}(X_2, X_2 - a_1X_1 - a_3X_3) = \mathrm{Cov}(X_2, X_2) - a_1 \mathrm{Cov}(X_2, X_1) - a_3 \mathrm{Cov}(X_2, X_3) \\
\mathrm{Cov}(X_2, Z) &= 3 - a_1 \cdot 1 - a_3 \cdot 2 \\
0 &= 3 - a_1 \cdot 1 - a_3 \cdot 2 \\
a_1 &= 3 - 2a_3
\end{aligned}$$

Choosing $a_3 = 0$ we get independence for $a_1 = 3$, and with $a_3 = 1$ we get independence with $a_1 = 1$, and so on.

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

We need to partition $\boldsymbol{X}$, the mean vector and covariance matrix of $\boldsymbol{X}$ as follows.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_A \\ \boldsymbol{X}_B \end{bmatrix} \text{ and } \boldsymbol{\mu} = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{bmatrix}$$

where the A-part contains X_1 and the B-part contains X_2 and X_3. The formula for the conditional mean $\boldsymbol{\mu}^*$ and covariance $\boldsymbol{\Sigma}^*$ of the A-part given the B-part at $(x_2, x_3)^T$ is

$$\begin{aligned}\boldsymbol{\mu}^* &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B)\\ \boldsymbol{\Sigma}^* &= \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}\end{aligned}$$

We thus need

$$\begin{aligned}\boldsymbol{\Sigma}_{BB}^{-1} &= \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 1.5 \end{bmatrix}\\ \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1} &= \begin{bmatrix} 0 & 0.5 \end{bmatrix}\\ \boldsymbol{\Sigma}^* &= \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA} = 0.5\end{aligned}$$

Then,

$$\begin{aligned}&\boldsymbol{\mu}^* = 2 + (0, 0.5)^T(\boldsymbol{x}_B - \boldsymbol{\mu}_B) = 2 + 0\cdot(x_2+3) + 0.5(x_3-1) = 2 + 0.5(x_3-1) = 0.5x_3 + 1.5\\ &\boldsymbol{\Sigma}^* = 0.5\end{aligned}$$

Problem 2: From correlated to independent variables

a) To find out which variable has the strongest correlation with X_3, we compute $\operatorname{Corr}(X_1, X_3)$ and $\operatorname{Corr}(X_2, X_3)$ from the elements in $\boldsymbol{\Sigma}$,

$$\begin{aligned}\operatorname{Corr}(X_1, X_3) &= \frac{\operatorname{Cov}(X_1, X_3)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_3)}} = \frac{1}{\sqrt{1\cdot 3}} = \frac{\sqrt{3}}{3} \approx 0.5774\\ \operatorname{Corr}(X_2, X_3) &= \frac{\operatorname{Cov}(X_2, X_3)}{\sqrt{\operatorname{Var}(X_2)\operatorname{Var}(X_3)}} = \frac{-1}{\sqrt{2\cdot 3}} = -\frac{\sqrt{6}}{6} \approx -0.4082.\end{aligned}$$

This shows that X_1 has the strongest correlation with X_3 in absolute value. To determine the distribution of $\mathbf{Z} = (X_2 - X_1, X_3 - X_1)^T$, observe that if we let

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

then $\mathbf{Z} = \mathbf{AX}$. In other words, $\mathbf{Z}$ is a linear transformation of the trivariate normal vector $\mathbf{X}$, which means that $\mathbf{Z}$ is bivariate normal, since $\mathbf{A}$ is a 2×3 matrix. The expectation and covariance matrix of $\mathbf{Z}$ are

$$\operatorname{E}(\mathbf{Z}) = \mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \end{bmatrix} = \begin{bmatrix} 6-2 \\ 4-2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

and

$$\begin{aligned}\operatorname{Cov}(\mathbf{Z}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T &= \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\\ &= \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 2 & -1 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}.\end{aligned}$$

b) Note that $\mathbf{Y} = (\mathbf{e}_1^T\mathbf{X}, \mathbf{e}_2^T\mathbf{X})^T = (\mathbf{e}_1^T, \mathbf{e}_2^T)^T\mathbf{X} = \mathbf{B}\mathbf{X}$ is a linear transformation of $\mathbf{X}$ where

$$\mathbf{B} = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{bmatrix}$$

is a 2×3 matrix. By the same argument as for $\mathbf{Z}$ in **a)**, it then follows that $\mathbf{Y}$ is bivariate normal. To show independence between Y_1 and Y_2, consider their covariance

$$\begin{aligned}
\mathrm{Cov}(Y_1, Y_2) &= \mathrm{Cov}(\mathbf{e}_1^T\mathbf{X}, \mathbf{e}_2^T\mathbf{X}) = \mathrm{E}\left((\mathbf{e}_1^T\mathbf{X} - \mathrm{E}(\mathbf{e}_1^T\mathbf{X}))(\mathbf{e}_2^T\mathbf{X} - \mathrm{E}(\mathbf{e}_2^T\mathbf{X}))^T\right) \\
&= \mathrm{E}(\mathbf{e}_1^T(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\mathbf{e}_2) = \mathbf{e}_1^T\,\mathrm{E}((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T)\mathbf{e}_2 \\
&= \mathbf{e}_1^T\boldsymbol{\Sigma}\mathbf{e}_2 = \mathbf{e}_1^T\lambda\mathbf{e}_2 = \lambda\mathbf{e}_1^T\mathbf{e}_2 = 0.
\end{aligned}$$

In the last line we use the orthogonality of the eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Now, since Y_1 and Y_2 are univariate normal, zero covariance implies independence.

By the proportion of variation in $\mathbf{X}$ which is explained by $\mathbf{Y}$, we mean the ratio of the total variance of $\mathbf{Y}$ to the total variance of $\mathbf{X}$. The total variance of a vector is the sum of the variances of its elements, which is equal to the trace of the covariance matrix. Write the eigenvalue decomposition of $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}^T = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \mathbf{e}_3^T \end{bmatrix}.$$

The total variance of $\mathbf{X}$ is then

$$\mathrm{tr}(\boldsymbol{\Sigma}) = \mathrm{tr}(\mathbf{V}\mathbf{D}\mathbf{V}^T) = \mathrm{tr}(\mathbf{V}^T\mathbf{V}\mathbf{D}) = \mathrm{tr}(\mathbf{I}\mathbf{D}) = \mathrm{tr}(\mathbf{D}) = \sum_{i=1}^{3} \lambda_i,$$

where we use the fact that matrices in the trace of a product can be switched. We see that the total variation is the sum of the eigenvalues of the covariance matrix. Next, we turn to the covariance matrix of $\mathbf{Y} = \mathbf{B}\mathbf{X}$, which is

$$\begin{aligned}
\mathrm{Cov}(\mathbf{Y}) &= \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T = \mathbf{B}\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{B}^T \\
&= \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \mathbf{e}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},
\end{aligned}$$

so the total variance of $\mathbf{Y}$ is $\mathrm{tr}(\mathrm{Cov}(\mathbf{Y})) = \lambda_1 + \lambda_2$. We are now ready to compute the proportion of variation in $\mathbf{X}$ explained by $\mathbf{Y}$,

$$\frac{\mathrm{tr}(\mathrm{Cov}(\mathbf{Y}))}{\mathrm{tr}(\mathrm{Cov}(\mathbf{X}))} = \frac{\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)}{\mathrm{tr}(\boldsymbol{\Sigma})} = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{3.8794 + 1.6527}{3.8794 + 1.6527 + 0.4679} = 0.922.$$

Approximately 92% of the variation in $\mathbf{X}$ is explained by $\mathbf{Y}$.

Problem 3: The bivariate normal distribution

a)

$$\mathbf{\Sigma}^{-1} = \frac{1}{\det(\mathbf{\Sigma})}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix} = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}$$

$$\begin{aligned}
(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{bmatrix} x-\mu_X & y-\mu_Y \end{bmatrix}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix} \\
&= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{bmatrix} x-\mu_X & y-\mu_Y \end{bmatrix}\begin{bmatrix} \sigma_Y^2(x-\mu_X)-\rho\sigma_X\sigma_Y(y-\mu_Y) \\ -\rho\sigma_X\sigma_Y(x-\mu_X)+\sigma_X^2(y-\mu_Y) \end{bmatrix} \\
&= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\big[\sigma_Y^2(x-\mu_X)^2-\rho\sigma_X\sigma_Y(y-\mu_Y)(x-\mu_X) \\
&\quad -\rho\sigma_X\sigma_Y(x-\mu_X)(y-\mu_Y)+\sigma_X^2(y-\mu_Y)^2\big] \\
&= \frac{1}{(1-\rho^2)}[(\frac{x-\mu_X}{\sigma_X})^2+(\frac{y-\mu_Y}{\sigma_Y})^2-2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y})] = Q(x,y)
\end{aligned}$$

b) From a) we saw that $Q(x,y) = (\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. Further, using the formla for $\det(\mathbf{\Sigma})$ we find that

$$c = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} = \frac{1}{2\pi\det(\mathbf{\Sigma})^{1/2}}$$

This gives, directly,

$$\begin{aligned}
f(x,y) &= c\exp(-\frac{1}{2}Q(x,y)) \\
&= \frac{1}{2\pi\det(\mathbf{\Sigma})^{1/2}}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \\
&= f(\boldsymbol{x})
\end{aligned}$$

c)

$$\begin{aligned}
f(\boldsymbol{x}) &= k \\
\frac{1}{2\pi\det(\mathbf{\Sigma})^{1/2}}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) &= k \\
\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) &= k\cdot 2\pi\det(\mathbf{\Sigma})^{1/2} \\
(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= -2\cdot\log(k\cdot 2\pi\det(\mathbf{\Sigma})^{1/2}) \\
(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= d^2
\end{aligned}$$

So, we may find the contours by solving the first or the last equation. We will now work with the last equation.

We recognize that this is a quadratic form. Further, $\mathbf{\Sigma}$ is a real, symmetric matrix. We know from linear algebra that the eigenvectors of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^{-1}$ are the same, while the eigenvalues are reciprocal.

We put the eigenvalues of $\boldsymbol{\Sigma}$ on the diagonal of the matrix $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2)$ and the (normalized) (2×1) eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, which we put into the (2×2) matrix $\boldsymbol{P} = [\boldsymbol{e}_1\boldsymbol{e}_2]$, such that $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$.

$$
\begin{aligned}
(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= d^2 \\
(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^T(\boldsymbol{x}-\boldsymbol{\mu}) &= d^2 \\
\frac{1}{\lambda_1}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{e}_1\boldsymbol{e}_1^T(\boldsymbol{x}-\boldsymbol{\mu}) + \frac{1}{\lambda_2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{e}_2\boldsymbol{e}_2^T(\boldsymbol{x}-\boldsymbol{\mu}) &= d^2
\end{aligned}
$$

Let $w_1 = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{e}_1$ and $w_2 = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{e}_2$.

$$
\begin{aligned}
\frac{1}{\lambda_1}w_1^2 + \frac{1}{\lambda_2}w_2^2 &= d^2 \\
\frac{1}{\lambda_1 d^2}w_1^2 + \frac{1}{\lambda_2 d^2}w_2^2 &= 1
\end{aligned}
$$

From the latter equation we see that this is an ellipse with axis in the direction of the eigenvectors of $\boldsymbol{\Sigma}$, with halflengths $\sqrt{\lambda_1}d$ and $\sqrt{\lambda_2}d$. The center of the ellipse is at $\boldsymbol{\mu}$.

d) Let $\sigma_X = \sigma_Y = \sigma$, and we have

$$
\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}
$$

We find the eigenvalues of $\boldsymbol{\Sigma}$ by solving $\det(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}) = 0$ to be

$$
\begin{aligned}
\lambda_1 &= \sigma^2(1+\rho) \\
\lambda_2 &= \sigma^2(1-\rho)
\end{aligned}
$$

and the corresponding normalized eigenvalues by solving $\boldsymbol{\Sigma}\boldsymbol{e} = \lambda\boldsymbol{e}$, to be

$$
\begin{aligned}
\boldsymbol{e_1} &= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
\boldsymbol{e_2} &= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}
\end{aligned}
$$

Observe that these two directions do not depend on ρ, and will always give axes which form $45°$ angles with the horizontal axis.

If $\rho > 0$ the major axis (the axis with the longest halflength) is in the direction $\boldsymbol{e}_1$, and if $\rho < 0$ the major axis is in the direction $\boldsymbol{e}_2$. If $\rho = 0$ the ellipse becomes a circle (both axes equal in length).

e) If $\sigma_X = \sigma_Y$ the axes of the ellipse are always as explained in d), and the effect of an increasing $|\rho|$ is that the ellipse becomes narrower.

If $\sigma_X \neq \sigma_Y$ then the direction of the ellipse axes will depend on all of $(\sigma_X, \sigma_Y, \rho)$. But keeping σ_X and σ_Y fixed will also result in narrower ellipses for increasing $|\rho|$.

Run the `R`-commands below (also available in an `.R` file on the course website).

```
par(mfrow=c(2,2),pty="m") # fit 4 plots in a 2 by 2 manner
# first
plot(ellipse(0.5, scale=c(1,1),centre=c(0,0)),type="l")
abline(0,1); abline(0,-1) #adding the ellipse axes
title("SigmaX=SigmaY=1, rho=0.5")
# second
plot(ellipse(-0.3, scale=c(1,1),centre=c(0,0)),type="l")
abline(0,1); abline(0,-1)
title("SigmaX=SigmaY=1, rho=-0.5")

# third
sigma <- matrix(c(1,0.5*1*2,0.5*1*2,2^2),ncol=2)
res <- eigen(sigma)
plot(ellipse(sigma,centre=c(0,0)),type="l")
title("SigmaX=1,SigmaY=2, rho=0.5")

#forth
sigma <- matrix(c(1,-0.9*1*2,-0.9*1*2,2^2),ncol=2)
plot(ellipse(sigma,centre=c(0,0)),type="l")
title("SigmaX=1,SigmaY=2, rho=-0.9")

# optional -- add axes
# want the axes to be perpendicular - then need to make plotting region square
# AND also use equal range for x and y - which must be set separately
par(mfrow=c(1,1),pty="s") # one graph and square region
muvec <- c(0,0)
eig <- eigen(sigma) # eigenvalues and vectors
const <- sqrt(qchisq(0.95,2)) # choose a constant so that 95% probability of being inside
# (more the distribution of this quadratic from later)
eobj <- ellipse(sigma,centre=muvec) # generate points on the ellipse
apply(eobj,2,range) # check which of x or y have the largest range,
#choose the one with the largest for the plot below,
#here this was y, and thus range(eboj[,2])
plot(eobj,xlim=range(eobj[,2]),ylim=range(eobj[,2]),type="l")
#plot(eobj,type="l") would give different scales for x and y and
#not make this pretty! try to see
lambda1 <- eig$values[1] # first eigenvalue
e1 <- eig$vectors[,1] # first eigenvector
pkt1R <- muvec+const*sqrt(lambda1)*e1 # point on ellipse major axis
points(pkt1R[1],pkt1R[2],col=3,pch=20) # add the point to plot, green
pkt1L <- muvec-const*sqrt(lambda1)*e1 # point on ellipse
points(pkt1L[1],pkt1L[2],col=3,pch=20) # add point to plot, green
lines(c(pkt1R[1],pkt1L[1]),c(pkt1R[2],pkt1L[2]),lwd=2,col=3) # add line between points in green
# do the same with the minor axes, now in blue
lambda2 <- eig$values[2]
e2 <- eig$vectors[,2]
pkt2R <- muvec+const*sqrt(lambda2)*e2
points(pkt2R[1],pkt2R[2],col=4,pch=20)
pkt2L <- muvec-const*sqrt(lambda2)*e2
```

```
points(pkt2L[1],pkt2L[2],col=4,pch=20)
lines(c(pkt2R[1],pkt2L[1]),c(pkt2R[2],pkt2L[2]),lwd=2,col=4)
```

Problem 4: Normal marginals, but not multivariate normal?

In b) the data looks normal marginally, in c) the data does not look multivariate normal since we don't see elliptic shapes.

```
print("a) Load data")
ccdata <- dget("https://www.math.ntnu.no/emner/TMA4267/2017v/ccdata.dd")
dim(ccdata)

print("b) Marginal plots")

par(mfcol=c(2,3))
apply(ccdata,2,hist)
apply(ccdata,2,boxplot)
apply(ccdata,2,qqnorm)

# qqplot with lines more in detail...

par(mfrow=c(1,2))
qqnorm(ccdata[,1],main=""); qqline(ccdata[,1],col=2); title(main="Normal QQ
plot var1", cex.main=.8)
qqnorm(ccdata[,2],main=""); qqline(ccdata[,2],col=2); title(main="Normal QQ
plot var2", cex.main=.8)

print("c) Simultane plots")

library(ellipse)
par(mfrow=c(1,1))
plot(ccdata,main="Scatter plot",cex.main=.8,xlab="x",ylab="y")
lines(ellipse(matrix(c(1,0.8,0.8,1),2,2)),lwd=2,col=2)
sink()
```

Problem 5: The chi-square, t and F-distribution

a) Let $V \sim \chi^2_p$ and $W \sim \chi^2_q$, where V and W are independent. The joint distribution is then the product of the two marginal distributions.

$$f_{V,W}(u,v) = \frac{1}{\Gamma(p/2)2^{p/2}} v^{(p/2)-1} e^{-v/2} \cdot \frac{1}{\Gamma(q/2)2^{q/2}} w^{(q/2)-1} e^{-w/2}$$

Let then $F = \frac{V/p}{W/q}$, and $G = W$, and use the multivariate transformation formula to find the joint pdf of F and G. We start with the inverse functions and the Jacobian.

$$\begin{aligned}
V &= \frac{p}{q}FG \\
W &= G \\
J &= \det\begin{bmatrix} \frac{p}{q}g & \frac{p}{g}f \\ 0 & 1 \end{bmatrix} = \frac{p}{q}g \\
f_{F,G}(f,g) &= f_{V,W}(\frac{p}{q}fg, g)\cdot\frac{p}{q}g \\
&= \frac{1}{\Gamma(p/2)\Gamma(q/2)2^{(p+q)/2}}(\frac{p}{q}fg)^{(p/2)-1}g^{q/2-1}e^{-(\frac{p}{q}f+1)g/2}\frac{p}{q}g \\
&= \frac{1}{\Gamma(p/2)\Gamma(q/2)2^{(p+q)/2}}(\frac{p}{q})^{p/2}f^{p/2-1}g^{(p+q)/2-1}e^{-(\frac{p}{q}f+1)g/2}
\end{aligned}$$

Then, find the marginal distribution of F from this joint distibution.

$$\begin{aligned}
f_F(f) &= \frac{1}{\Gamma(p/2)\Gamma(q/2)2^{(p+q)/2}}(\frac{p}{q})^{p/2}f^{p/2-1}\int_0^\infty g^{(p+q)/2-1}e^{-(\frac{p}{q}f+1)g/2}dg \\
u &= (\frac{p}{q}f+1)g \text{ and } du = (\frac{p}{q}f+1)dg \\
f_F(f) &= \frac{1}{\Gamma(p/2)\Gamma(q/2)2^{(p+q)/2}}(\frac{p}{q})^{p/2}f^{p/2-1}\int_0^\infty \frac{u^{(p+q)/2-1}}{(\frac{p}{q}f+1)^{(p+q)/2-1}}e^{-u/2}\frac{du}{\frac{p}{q}f+1} \\
&= \frac{(\frac{p}{q})^{p/2}f^{p/2-1}}{\Gamma(p/2)\Gamma(q/2)2^{(p+q)/2}(\frac{p}{q}f+1)^{(p+q)/2}}\int_0^\infty u^{(p+q)/2-1}e^{-u/2}du \\
&= \frac{2^{(p+q)/2}\Gamma(\frac{p+q}{2})(\frac{p}{q})^{p/2}f^{p/2-1}}{2^{(p+q)/2}\Gamma(p/2)\Gamma(q/2)}(\frac{p}{q}f+1)^{(p+q)/2}\int_0^\infty \frac{1}{2^{(p+q)/2}\Gamma(\frac{p+q}{2})}u^{(p+q)/2-1}e^{-u/2}du \\
&= \frac{\Gamma(\frac{p+q}{2})(\frac{p}{q})^{p/2}}{\Gamma(p/2)\Gamma(q/2)}\frac{f^{p/2-1}}{(\frac{p}{q}f+1)^{(p+q)/2}}
\end{aligned}$$

b) Let $U \sim N(0,1)$ and $V \sim \chi^2_p$, and U and V are independent. Find the pdf of the random variable $T = \frac{U}{\sqrt{V/p}}$.

First, the joint pdf of U and V, by multiplying the marginal pdfs.

$$f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}\cdot\frac{1}{\Gamma(p/2)2^{p/2}}v^{(p/2)-1}e^{-v/2}$$

Now , the inverse of the transformation $t = \frac{u}{\sqrt{v/p}}$ and $w = v$ is $u = t\sqrt{w/p}$ and $v = w$, with Jacobian $\sqrt{w/p}$. This gives joint distribution $f_{T,W}(t,w)$:

$$f_{T,W}(t,w) = f_{U,V}(t(\sqrt{w/p}), w)\cdot\sqrt{w/p}$$

The marginal pdf of T is

$$
\begin{aligned}
f_T(t) &= \int_0^\infty f_{U,V}(t\sqrt{w/p}, w) \cdot \sqrt{w/p}\, dw \\
&= \frac{1}{(2\pi)^{1/2}\Gamma(\frac{p}{2})2^{p/2}} \int_0^\infty e^{-(1/2)t^2 w/p} w^{p/2-1} e^{-w/2} \left(\frac{w}{p}\right)^{1/2} dw \\
&= \frac{1}{(2\pi)^{1/2}\Gamma(\frac{p}{2})2^{p/2}p^{1/2}} \int_0^\infty e^{-(1/2)(1+t^2/p)w} w^{(p+1)/2-1} dw
\end{aligned}
$$

The trick now it to recognice that the integrand is the pdf of the gamma distribution $\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, with parameters $\alpha = (p+1)/2$ and $\beta = 2/(1+t^2/p)$, so the integral is 1.

$$
\begin{aligned}
f_T(t) &= \frac{\Gamma(\frac{p+1}{2})(\frac{2}{1+t^2/p})^{(p+1)/2}}{(2\pi)^{1/2}\Gamma(\frac{p}{2})2^{p/2}p^{1/2}} \int_0^\infty \frac{1}{\Gamma(\frac{p+1}{2})(\frac{2}{1+t^2/p})^{(p+1)/2}} e^{-1/(1+t^2/p)}\, w^{(p+1)/2-1} dw \\
&= \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+t^2/p)^{(p+1)/2}}
\end{aligned}
$$

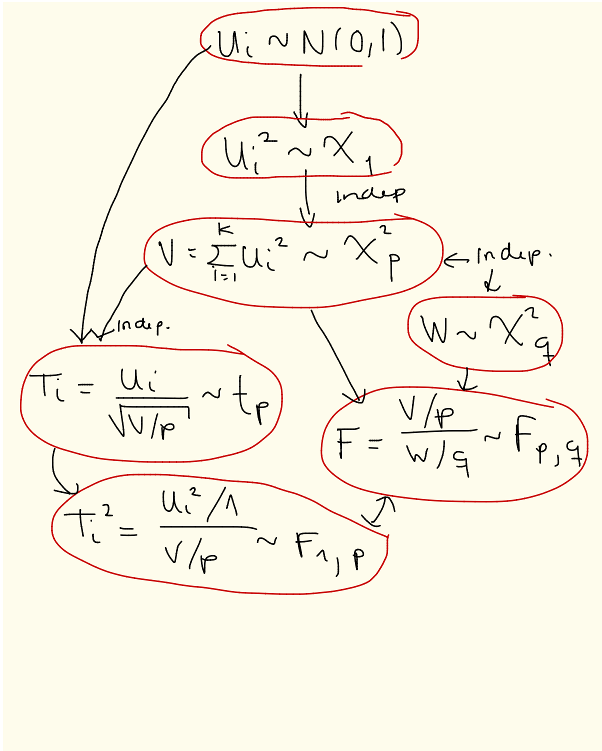
which is the T pdf.

c) Let $T \sim t_q$ (t-distribution with q degrees of freedom). Then show that $T^2 \sim F_{1,q}$.

$$
\begin{aligned}
T^2 &= \left(\frac{U}{\sqrt{V/p}}\right)^2 \\
&= \frac{U^2/1}{V/p}
\end{aligned}
$$

We see that the numerator is χ^2_1 and the denominator is χ^2_p, and from 1d we see that T^2 is Fisher with 1 and p degrees of freedom.

d)



Problem 6: N, Chi-square, t and F by simulation - in R

```
B <- 10000
n <- 10

# a
x <- rnorm(B,0,1)
y <- x^2
range(y)
hist(y,nclass=100,prob=TRUE)
dchisq1 <- function(x) return(dchisq(x,df=1))
curve(dchisq1,min(y),max(y),add=TRUE,col=2)
# curve only takes a function with ONE argument, needed to make a df=1 version of dchisq
abline(v=qchisq(0.1,1),col=3)
abline(v=qchisq(0.9,1),col=3)

# b
x <- rnorm(B)
y <- rchisq(B,df=n-1)
t <- x/sqrt(y/(n-1))
hist(t,nclass=50,prob=TRUE)
dt9 <- function(x) return(dt(x,df=9))
curve(dt9,min(t),max(t),add=TRUE,col=2)
# alternative to curve - plot two vectors
xvec <- seq(min(t),max(t),length=100)
yvec <- dt(xvec,df=n-1)
lines(xvec,yvec,col=4)
abline(v=qt(0.15,n-1),col=5)
abline(v=qt(0.85,n-1),col=5)
```

```
# c
f <- t^2
hist(f,nclass=50,prob=TRUE)
xvec <- seq(min(f),max(f),length=100)
yvec <- df(xvec,1,n-1)
lines(xvec,yvec,col=2)
abline(v=qf(0.05,1,n-1),col=5)
abline(v=qf(0.95,1,n-1),col=5)

# d more F
n1 <- 5
n2 <- 40
u <- rchisq(B,df=n1)
v <- rchisq(B,df=n2)
f <- u*n2/(v*n1)
hist(f,nclass=50,prob=TRUE)
xvec <- seq(min(f),max(f),length=100)
yvec <- df(xvec,n1,n2)
lines(xvec,yvec,col=2)
abline(v=qf(0.05,n1,n2),col=3)
abline(v=qf(0.95,n1,n2),col=3)
```

Problem 7: Linear combinations and quadratic forms

a) Find the expected value of $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$.

We may use the trace-formula:

$$\mathrm{E}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu} = \mathrm{tr}\,\boldsymbol{A} + \boldsymbol{\mu}^T\boldsymbol{0} = 3\cdot\frac{2}{3} = 2$$

b) $\boldsymbol{A}$ is clearly symmetric, which we can see by $\boldsymbol{A}^T = \boldsymbol{A}$. A projection matrix is an idempotent matrix, that is, $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$. We have already seen this in **a)**.

The rank of a symmetric idempotent matrix equals its trace, which we found in **a)** to be $\mathrm{tr}(\boldsymbol{A}) = 2$.

Derive the distribution of $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$.
From the lecture it is known that if $\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ and $\boldsymbol{A}$ is a symmetric, idempotent matrix with rank r then $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} \sim \sigma^2\chi^2_r$.

We have that $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{I})$, and $\boldsymbol{A}$ is symmetric and idempotent with rank 2. We need to rewrite our expression so that we have a normally distributed random vector with mean zero and identity covariance matrix.

We subtract the mean and write

$$(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{A}(\boldsymbol{X}-\boldsymbol{\mu}) = \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} - \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$$

since $\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{0}$. Define $\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\mu}$, where $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I})$. We may thus use the above theorem, to find that $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{Z} \sim \chi^2_2$, that is χ^2 with 2 degrees of freedom.

Tabeller og formler i statistikk, page 5, we see that 6 is the critical value in the χ^2-distribution with 2 degrees of freedom and probability 0.05 (value in table is 5.991). The probability that the quadratic form is smaller than 6 is thus 95%.