TMA4267 - Linear Statistical Models
Solutions to Exercise 1 - V2014

Problem 1: Covariance and correlation

a)

$$\begin{aligned}
\mathrm{E}(aX+b) &= a\mu_X + b \\
\mathrm{E}(cY+d) &= c\mu_Y + d \\
\mathrm{Cov}(aX+b, cY+d) &= \mathrm{E}([aX+b-\mathrm{E}(aX+b)][cY+d-\mathrm{E}(cY+d)]) \\
&= \mathrm{E}((aX+b-a\mu_x+b)(cY-d-c\mu_Y+d)) \\
&= \mathrm{E}(ac(X-\mu_X)(Y-\mu_Y)) \\
&= ac\,\mathrm{E}((X-\mu_X)(Y-\mu_Y)) = ac\,\mathrm{Cov}(X,Y)
\end{aligned}$$

b)

$$\begin{aligned}
\mathrm{E}(aX+b) &= a\mu_X + b \\
\mathrm{E}(cY+d) &= c\mu_Y + d \\
\mathrm{Var}(aX+b) &= a^2\sigma_X^2 \\
\mathrm{Var}(cY+d) &= c^2\sigma_Y^2 \\
\mathrm{Corr}(aX+b, cY+d) &= \frac{\mathrm{Cov}(aX+b, cY+d)}{a\sigma_X \cdot c\sigma_Y} = \frac{ac\,\mathrm{Cov}(X,Y)}{a\sigma_X \cdot c\sigma_Y} \\
&= \mathrm{Corr}(X,Y)
\end{aligned}$$

c)

$$\begin{aligned}
\mathrm{Cov}(aX+bY, cZ+dV) &= \mathrm{E}([aX+bY-a\mu_X-b\mu_Y][cZ+dV-c\mu_Z-d\mu_V]) \\
&= \mathrm{E}([a(X-\mu_X)+b(Y-\mu_Y)][(c(Z-\mu_Z)+d(V-\mu_V)]) \\
&= \mathrm{E}(ac(X-\mu_X)(Z-\mu_Z)+ad(X-\mu_X)(V-\mu_V) \\
&+ bc(Y-\mu_Y)(Z-\mu_Z)+bd(Y-\mu_Y)(V-\mu_V)) \\
&= ac\,\mathrm{Cov}(X,Z)+ad\,\mathrm{Cov}(X,V)+bc\,\mathrm{Cov}(Y,Z)+bd\,\mathrm{Cov}(Y,V)
\end{aligned}$$

d) Prove that

$$\mathrm{Cov}(X,Y)^2 \le \mathrm{Var}(X)\,\mathrm{Var}(Y).$$

$$
\begin{aligned}
Z &= X - Y \cdot \operatorname{Cov}(X,Y)/\operatorname{Var}(Y) \\
\operatorname{Var}(Z) &= \operatorname{Var}(X - Y \cdot \operatorname{Cov}(X,Y)/\operatorname{Var}(Y)) \\
&= \operatorname{Var}(X) + (\frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)})^2 \operatorname{Var}(Y) - 2(\frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}) \operatorname{Cov}(X,Y) \\
&= \operatorname{Var}(X) - \frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}(Y)}
\end{aligned}
$$

We know that 0 is a lower limit for the variance.

$$
\begin{aligned}
0 \le \operatorname{Var}(Z) &= \operatorname{Var}(X) - \frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}(Y)} \\
\operatorname{Cov}(X,Y)^2 &\le \operatorname{Var}(X) \cdot \operatorname{Var}(Y)
\end{aligned}
$$

Problem 2: The bivariate normal distribution

a)

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

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

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

This gives, directly,

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

c)

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

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, $\boldsymbol{\Sigma}$ is a real, symmetric matrix. We know from linear algebra that the eigenvectors of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ 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 $\boldsymbol{\Sigma e} = \lambda \boldsymbol{e}$, to be

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

Observe that these two directions does'nt depend on ρ, and will always give axis with 45 degree angle with the horisonal 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 (equal length of both axis).

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

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

```
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)
```