Compulsory Exercise 1 in TMA4267 Statistical Linear Models, Spring 2021

Sander Ruud, Alexander J Ohrt

25 februar, 2021

Problem 1 Bivariate normal distribution

a)

Let

$$A := \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Then

$$\mathrm{E}(\boldsymbol{Y}) = \mathrm{E}(A\boldsymbol{X}) = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -2/\sqrt{2} \\ 2/\sqrt{2} \end{pmatrix} = \sqrt{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

and

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

We know that the Gaussian distribution is closed under linear combinations. This means that

$$\mathbf{Y} \sim N_2\left(\sqrt{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}\right)$$

Moreover, since we know that uncorrelated variables in a random vector with a multivariate Gaussian distribution must be independent, this means that Y_1 and Y_2 are independent.

Bonus: Calculations in R.

```
mu.X <- c(0,2)
Sigma.X <- matrix(c(3,1,1,3), ncol = 2)
A <- matrix(c(1/sqrt(2), 1/sqrt(2), -1/sqrt(2), 1/sqrt(2)), ncol = 2)

mu.Y <- A %*% mu.X
mu.Y
```

```
#>           [,1]
#> [1,] -1.414214
#> [2,]  1.414214
```

```r
Sigma.Y <- A %*% Sigma.X %*% t(A)
Sigma.Y
```

```
#>      [,1] [,2]
#> [1,]    2    0
#> [2,]    0    4
```

```r
# Yes, the coordinates of Y are independent, since Y is normally distributed
# and the coordinates are uncorrelated.
```

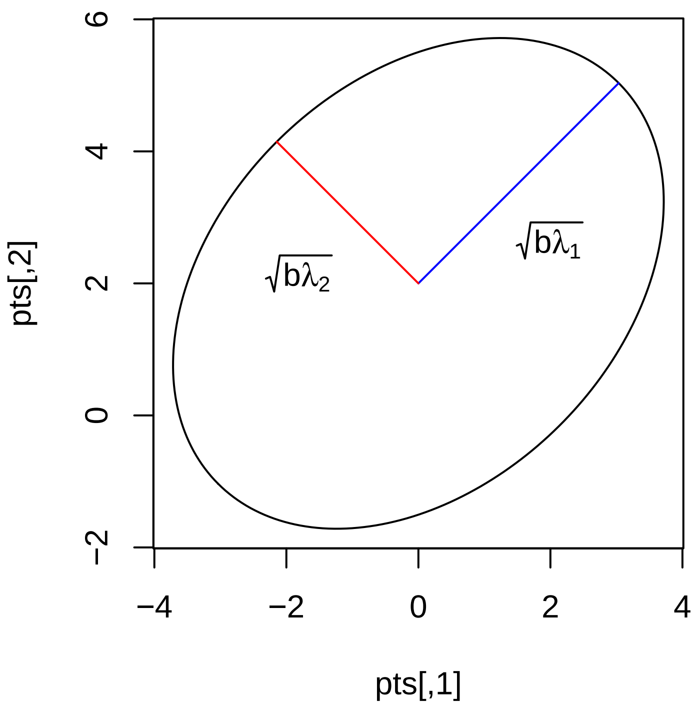
b)

By using the method of diagonalization one can show that the contours of a vector that has a multivariate Gaussian distribution are ellipsoids. Moreover, one can show that the axes of the ellipsoids have direction along the eigenvectors of the covariance matrix Σ and half-lengths $\sqrt{\lambda_i b}$ for, where λ_i are the eigenvalues of Σ. More specifically, each of the half axes are given, in descending order based on length, by $\sqrt{\lambda_i b}$ with direction following each respective eigenvector $\mathbf{p}_i$, $i = 1, \ldots, k$ for $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$, where k is the dimension of the multivariate Gaussian distribution.

Theorem B.8.1 (FKLM, 2013) states that, in general, $\boldsymbol{Y} = (\boldsymbol{X} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_2$ when $\boldsymbol{Y} \sim N_2(\mu, \Sigma)$. Hence, the probability that $\boldsymbol{X}$ falls within the given ellipse is ≈ 0.9, which is seen from the given output of `qchisq(0.9, 2)` $= 4.6051702$. This means that the quantile in the χ^2_2-distribution with the value $b \approx 4.6$ corresponds to the probability ≈ 0.9.

We want the ellipsoid that captures information about the covariates in a way that a certain amount of points drawn from our distribution will end up inside this ellipsoid. The ellipsoid that includes 90 % of random points generated from the bivariate normal distribution of $\boldsymbol{X}$ is drawn below.

```r
library(mixtools)
library(latex2exp)
sigma=matrix(c(3,1,1,3),ncol=2)
mu_x <- 0
mu_y <- 2

lam1 <- eigen(sigma)$values[1]
v1 <- eigen(sigma)$vectors[,1]

lam2 <- eigen(sigma)$values[2]
v2 <- eigen(sigma)$vectors[,2]

l1 = sqrt(qchisq(0.9,2)*lam1)
l2 = sqrt(qchisq(0.9,2)*lam2)

# The vectors are already normalized in R.
x1 = (v1*l1)[1]
x2 = (v1*l1)[2]

x3 = (v2*l2)[1]
x4 = (v2*l2)[2]

ellipse(c(mu_x, mu_y), sigma, alpha = 0.1, newplot=TRUE, type='l')
lines(c(mu_x, x1+mu_x), c(mu_y, x2+mu_y), col = "blue")
text(2,2.7, labels=TeX("$\\sqrt{b\\lambda_1}$"))
lines(c(mu_x, x3+mu_x), c(mu_y, x4+mu_y), col = "red")
text(-1.8,2.2, labels=TeX("$\\sqrt{b\\lambda_2}$"))
```



The major half-axis, in blue, shows $\sqrt{\lambda_1 b}\frac{\boldsymbol{p_1}}{|\boldsymbol{p_1}|}$, i.e. the length $\sqrt{\lambda_1 b}$ in the direction of the first eigenvector of Σ. Similarly, the minor half-axis, in red, shows $\sqrt{\lambda_2 b}\frac{\boldsymbol{p_2}}{|\boldsymbol{p_2}|}$.

Problem 2 Distributional results for $\overline{X}$ and S^2 for a univariate normal sample

a)

First, it will be shown that $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\mathbf{1}^T\boldsymbol{X}$

$$\overline{X} = \frac{1}{n}\mathbf{1}^T\boldsymbol{X} = \frac{1}{n}\begin{pmatrix}1 & \cdots & 1\end{pmatrix}\begin{pmatrix}X_1 \\ \vdots \\ X_n\end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Next, it will be shown that $S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\boldsymbol{X}^T C\boldsymbol{X}$, where $C = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. To this end, we will first show that $\boldsymbol{X} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\boldsymbol{X} = (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X} = \boldsymbol{X} - \overline{\boldsymbol{X}}$, where $\overline{\boldsymbol{X}}$ is a vector of the average of the components of $\boldsymbol{X}$. This can be shown by

$$(I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X} = \begin{pmatrix} 1-1/n & -1/n & \dots & -1/n \\ \vdots & \ddots & \vdots & \vdots \\ -1/n & -1/n & \dots & 1-1/n \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 - \frac{X_1}{n} - \frac{X_2}{n} - \dots - \frac{X_n}{n} \\ -\frac{X_1}{n} + X_2 - \frac{X_2}{n} - \dots - \frac{X_n}{n} \\ \vdots \\ -\frac{X_1}{n} - \frac{X_2}{n} - \dots + X_n - \frac{X_n}{n} \end{pmatrix}$$

$$= \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} -\frac{X_1}{n} - \frac{X_2}{n} - \dots - \frac{X_n}{n} \\ -\frac{X_1}{n} - \frac{X_2}{n} - \dots - \frac{X_n}{n} \\ \vdots \\ -\frac{X_1}{n} - \frac{X_2}{n} - \dots - \frac{X_n}{n} \end{pmatrix} = \boldsymbol{X} - \begin{pmatrix} \sum_{i=1}^n \frac{X_i}{n} \\ \sum_{i=1}^n \frac{X_i}{n} \\ \vdots \\ \sum_{i=1}^n \frac{X_i}{n} \end{pmatrix} = \boldsymbol{X} - \overline{\boldsymbol{X}}.$$

This means that

$$((I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X})^T((I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X}) = (\boldsymbol{X}-\overline{\boldsymbol{X}})^T(\boldsymbol{X}-\overline{\boldsymbol{X}}) = \sum_{i=1}^n (X_i - \overline{X})^2,$$

since $\boldsymbol{a}^T\boldsymbol{b} = \sum_{i=1}^n a_i b_i$ is the inner product between two vectors $\boldsymbol{a} = (a_1\ a_2\ \dots\ a_n)^T$ and $\boldsymbol{b} = (b_1\ b_2\ \dots\ b_n)^T$. Hence, this leads to the final result

$$\begin{aligned} S^2 &= \frac{1}{n-1}\boldsymbol{X}^T C \boldsymbol{X} = \frac{1}{n-1}\boldsymbol{X}^T(I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X} \\ &= \frac{1}{n-1}((I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X})^T((I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X}) \\ &= \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})(X_i-\overline{X}) = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2, \end{aligned}$$

since $C = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is idempotent and symmetric. The fact that C is idempotent and symmetric may be verified by direct calculation. The idempotent property follows by

$$\begin{aligned} (I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)^2 &= (I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)(I-\frac{1}{n}\mathbf{1}\mathbf{1}^T) = I - \frac{2}{n}\mathbf{1}\mathbf{1}^T + \frac{1}{n^2}\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T \\ &= I - \frac{2}{n}\mathbf{1}\mathbf{1}^T + \frac{1}{n}\mathbf{1}\mathbf{1}^T = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T, \end{aligned}$$

where the third equality follows since, in $\mathbb{R}^n$, we have that $\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T = n\mathbf{1}\mathbf{1}^T$. The symmetric property follows by

$$(I-\frac{1}{n}\mathbf{1}\mathbf{1}^T)^T = I^T - \frac{1}{n}(\mathbf{1}\mathbf{1}^T)^T = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T.$$

b)

$$\frac{1}{n}\mathbf{1}^T C = \frac{1}{n}\mathbf{1}^T I - \frac{1}{n^2}\mathbf{1}^T\mathbf{1}\mathbf{1}^T = \mathbf{1}^T\frac{1}{n} - \frac{1}{n^2}\,[n\,...\,n] = \mathbf{1}^T(1/n - 1/n) = \mathbf{0}^T$$

By Corollary 5.2 (HS), this implies that $\frac{1}{n}\mathbf{1}^{\mathrm{T}}\boldsymbol{X}$ and $C\boldsymbol{X}$ are independent, since $\frac{1}{n}\mathbf{1}^T\sigma^2 I C^T = 0$, C is idempotent and $\boldsymbol{X} \sim N(\mu\mathbf{1}, \sigma^2 I)$. This is a result of the fact that two linear transforms, $A\boldsymbol{X}$ and $B\boldsymbol{X}$, of a multivariate Gaussian $\boldsymbol{X}$ are independent iff $\mathrm{Cov}(A\boldsymbol{X}, B\boldsymbol{X}) = 0$. Since $\frac{1}{n}\mathbf{1}^{\mathrm{T}}\boldsymbol{X}$ and $C\boldsymbol{X}$ are independent, $\frac{1}{n}\mathbf{1}^{\mathrm{T}}\boldsymbol{X}$ and $(C\boldsymbol{X})^T C\boldsymbol{X} = \boldsymbol{X}^T C^2 \boldsymbol{X} = \boldsymbol{X}^T C \boldsymbol{X}$ must also be independent. This implies that $\bar{X} = \frac{1}{n}\mathbf{1}^{\mathrm{T}}\boldsymbol{X}$ and $S^2 = \frac{1}{n-1}\boldsymbol{X}^T C \boldsymbol{X}$ are independent.

c)

Using the Mahalanobis transform, we know that

$$\frac{1}{\sigma^2}(\boldsymbol{X} - \mu\mathbf{1})^T C(\boldsymbol{X} - \mu\mathbf{1}) \sim \chi^2_r,$$

where $r = \text{rank}(C)$, since $\frac{1}{\sigma^2}(\boldsymbol{X} - \mu\mathbf{1}) \sim N(\mathbf{0}, I)$ and $C = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is idempotent and symmetric (Theorem B.8.2, FKLM, 2013). We want to derive the distribution of $\frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)}{(n-1)\sigma^2}\boldsymbol{X}^T C\boldsymbol{X} = \frac{1}{\sigma^2}\boldsymbol{X}^T C\boldsymbol{X}$. Expanding $\frac{1}{\sigma^2}(\boldsymbol{X} - \mu\mathbf{1})^T C(\boldsymbol{X} - \mu\mathbf{1})$ gives

$$\begin{aligned}
&\frac{1}{\sigma^2}(\boldsymbol{X} - \mu\mathbf{1})^T C(\boldsymbol{X} - \mu\mathbf{1}) \\
=&\frac{1}{\sigma^2}(\boldsymbol{X}^T C\boldsymbol{X} - \boldsymbol{X}^T C\mu\mathbf{1} - (\mu\mathbf{1})^T C\boldsymbol{X} + (\mu\mathbf{1})^T C\mu\mathbf{1}) \\
=&\frac{1}{\sigma^2}(\boldsymbol{X}^T C\boldsymbol{X} - 2\mu\boldsymbol{X}^T C\mathbf{1} + \mu^2\mathbf{1}^T C\mathbf{1}) \\
=&\frac{1}{\sigma^2}\boldsymbol{X}^T C\boldsymbol{X}
\end{aligned}$$

where the third inequality holds because $C\mathbf{1} = (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{1} = \mathbf{1} - \frac{1}{n}\mathbf{1}n = 0$. Hence, since $\frac{1}{\sigma^2}(\boldsymbol{X} - \mu\mathbf{1})^T C(\boldsymbol{X} - \mu\mathbf{1}) \sim \chi^2_r$, we now know that $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\boldsymbol{X}^T C\boldsymbol{X} \sim \chi^2_r$ also, where $r = \text{rank}(C) = \text{rank}(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T) = \text{tr}(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T) = n - \frac{n}{n} = n - 1$. Here, $\text{tr}(C) = \text{rank}(C)$ since C is idempotent. $\square$