Chair for Theoretical Information Technology | RWTH AACHEN UNIVERSITY

Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 4
- Proposed Solution -

Friday, November 9, 2018

Solution of Problem 1

The multivariate normal (or Gaussian) distribution of a random vector $\mathbf{Y} \in \mathbb{R}^p$ has the following pdf:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\},$$

where $\mathbf{y} = (y_1, \ldots, y_p)^T \in \mathbb{R}^p$, and the parameters: $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, where $\boldsymbol{\Sigma} \succ 0$.

a) In our case we have that $p = 2$, yielding

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}.$$

We start by calculating the determinant of $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times 2}$ as $|\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2 - \sigma_1^2\sigma_2^2\rho^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$. This leads to $|\boldsymbol{\Sigma}|^{1/2} = \sigma_1\sigma_2\sqrt{1-\rho^2}$ and

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.$$

Finally, we calculate

$$\begin{aligned}
&-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\
&\quad = -\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix}(y_1-\mu_1) & (y_2-\mu_2)\end{bmatrix} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix}(y_1-\mu_1) \\ (y_2-\mu_2)\end{bmatrix} \\
&\quad = -\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix}(y_1-\mu_1) & (y_2-\mu_2)\end{bmatrix} \begin{bmatrix} \sigma_2^2(y_1-\mu_1) - \rho\sigma_1\sigma_2(y_2-\mu_2) \\ \sigma_1^2(y_2-\mu_2) - \rho\sigma_1\sigma_2(y_1-\mu_1) \end{bmatrix} \\
&\quad = -\frac{\sigma_2^2(y_1-\mu_1)^2 + \sigma_1^2(y_2-\mu_2)^2 - 2\rho\sigma_1\sigma_2(y_1-\mu_1)(y_2-\mu_2)}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \\
&\quad = -\frac{1}{2(1-\rho^2)}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right],
\end{aligned}$$

this gives us the final expression for $f_{\mathbf{Y}}(\mathbf{y})$ as

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}.$$

b) From the definition of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ we directly get $Y_1 \sim N(\mu_1, \sigma_1)$ and $Y_2 \sim N(\mu_2, \sigma_2)$.

c) As stated in theorem 3.5 of the lecture's script, the conditional density $f_{Y_1|Y_2}(y_1|y_2)$ is given by the normal distribution $f_{Y_1|Y_2}(y_1|y_2) \sim N_1(\mu_{1|2}, \Sigma_{1|2})$, where $\mu_{1|2}$ is

$$\begin{aligned}\mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)\\ &= \mu_1 + (\rho\sigma_1\sigma_2)(1/\sigma_2^2)(y_2 - \mu_2)\\ &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y_2 - \mu_2)\end{aligned}$$

and $\Sigma_{1|2}$ is

$$\begin{aligned}\Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\\ &= \sigma_1^2 - (\rho\sigma_1\sigma_2)(1/\sigma_2^2)(\rho\sigma_1\sigma_2)\\ &= \sigma_1^2 - \rho^2\sigma_1^2 = \sigma_1^2(1-\rho^2)\,.\end{aligned}$$

Solution of Problem 2

Note that an estimator $\hat{X}$ of a parameter X is unbiased if its expected value equals X. Therefore it is enough to show:

$$\mathbb{E}(\overline{\mathbf{X}}) = \boldsymbol{\mu} = \mathbb{E}(\mathbf{X}), \quad \mathbb{E}(\mathbf{S}_n) = \boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}).$$

First see that:

$$\mathbb{E}(\overline{\mathbf{X}}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\mathbf{X}_i\right) = \frac{1}{n}n\mathbb{E}\left(\mathbf{X}\right) = \mathbb{E}\left(\mathbf{X}\right).$$

For the sample covariance matrix, we have:

$$\begin{aligned}\mathbb{E}(\mathbf{S}_n) &= \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right)\\ &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left((\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right)\end{aligned}$$

Next see that:

$$\begin{aligned}&\mathbb{E}\left((\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T\right) = \mathbb{E}\left((\mathbf{X}_i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j)(\mathbf{X}_i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j)^T\right)\\ &= \mathbb{E}\left((\mathbf{X}_i - \boldsymbol{\mu} - \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu}))(\mathbf{X}_i - \boldsymbol{\mu} - \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu}))^T\right)\\ &= \mathbb{E}\left(\left(\frac{n-1}{n}(\mathbf{X}_i - \boldsymbol{\mu}) - \frac{1}{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)\left(\frac{n-1}{n}(\mathbf{X}_i - \boldsymbol{\mu}) - \frac{1}{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_j - \boldsymbol{\mu})\right)^T\right)\end{aligned}$$

It is easy to see that:

$$\mathbb{E}\left((\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T\right) = \delta_{ij}\boldsymbol{\Sigma}.$$

Using this fact, it is easy to see that:

$$
\begin{aligned}
&\mathbb{E}\left(\left(\frac{n-1}{n}(\mathbf{X}_i-\boldsymbol{\mu})-\frac{1}{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_j-\boldsymbol{\mu})\right)\left(\frac{n-1}{n}(\mathbf{X}_i-\boldsymbol{\mu})-\frac{1}{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_j-\boldsymbol{\mu})\right)^T\right)\\
&=\frac{(n-1)^2}{n^2}\mathbb{E}\left((\mathbf{X}_i-\boldsymbol{\mu})(\mathbf{X}_i-\boldsymbol{\mu})^T\right)+\frac{1}{n^2}\sum_{j=1,j\neq i}^{n}\mathbb{E}\left((\mathbf{X}_j-\boldsymbol{\mu})(\mathbf{X}_j-\boldsymbol{\mu})^T\right)\\
&=\frac{(n-1)^2}{n^2}\boldsymbol{\Sigma}+\frac{n-1}{n^2}\boldsymbol{\Sigma}=\frac{n-1}{n}\boldsymbol{\Sigma}.
\end{aligned}
$$

Therefore $\mathbb{E}\left((\mathbf{X}_i-\overline{\mathbf{X}})(\mathbf{X}_i-\overline{\mathbf{X}})^T\right)=\frac{n-1}{n}\boldsymbol{\Sigma}$. We can finally find the expected value of sample covariance as follows:

$$\mathbb{E}(\mathbf{S}_n)=\frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left((\mathbf{X}_i-\overline{\mathbf{X}})(\mathbf{X}_i-\overline{\mathbf{X}})^T\right)=\frac{1}{n-1}\sum_{i=1}^{n}\frac{n-1}{n}\boldsymbol{\Sigma}=\boldsymbol{\Sigma}.$$

Solution of Problem 3

Consider four samples in $\mathbb{R}^3$ given as follows:

$$\mathbf{x}_1=\begin{bmatrix}1\\2\\-3\end{bmatrix},\mathbf{x}_2=\begin{bmatrix}3\\-1\\-2\end{bmatrix}\mathbf{x}_3=\begin{bmatrix}-4\\2\\2\end{bmatrix}\mathbf{x}_4=\begin{bmatrix}-3\\-1\\4\end{bmatrix}.$$

a) The sample mean can be easily found as:

$$\overline{\mathbf{x}}=\begin{bmatrix}-0.75\\0.5\\0.25\end{bmatrix}$$

To find the sample covariance, we have:

$$\mathbf{S}_n=\frac{1}{3}\sum_{i=1}^{4}(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})^T=\frac{1}{3}\begin{bmatrix}32.75&-4.5&-28.25\\-4.5&9&-4.5\\-28.25&-4.5&32.75\end{bmatrix}.$$

b) Step 1: find the sample covariance matrix $\mathbf{S}_n$ (previous part)

Step 2: find the eigenvalues and eigenvectors of the matrix. Sort them out and pick 2 orthonormal eigenvectors corresponding to 2 highest eigenvalues

$$\lambda_1=20.333333,\lambda_2=4.5,\lambda_3=0.$$

$$\mathbf{v}_1=\begin{bmatrix}-\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{bmatrix},\mathbf{v}_2=\begin{bmatrix}\frac{1}{\sqrt{6}}\\-\frac{2}{\sqrt{6}}\\\frac{1}{\sqrt{6}}\end{bmatrix},\mathbf{v}_3=\begin{bmatrix}\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}.$$

Step 3: Construct $\mathbf{Q}=\mathbf{v}_1\mathbf{v}_1^T+\mathbf{v}_2\mathbf{v}_2^T$.
Following this procedure, we have:

$$\mathbf{Q}=\frac{1}{3}\begin{bmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{bmatrix}.$$

c) Note that all the points are already on the same plane $x + y + z = 0$, so intuitively, the projection should be the projection on the same plane. This projection leaves those points untouched (Check!). Each $\mathbf{y} \in \text{Im}(\mathbf{Q})$ is also on this plane. To see that assume that $\mathbf{y} = \mathbf{Q}\mathbf{x}$. Then $y_1 + y_2 + y_3 = 0$. Another way, is to observe that the kernel of $\mathbf{Q}$ is spanned by the vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, the last eigenvector. Note, how the eigenvalue is zero for this eigenvector. Therefore its image is the orthogonal complement of this vector which is the plane $x + y + z = 0$.