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

Exercise 3
- Proposed Solution -

Friday, November 2, 2018

Solution of Problem 1

Note that for any random variable $\boldsymbol{Y} = g(\boldsymbol{X})$ the expectation $\mathrm{E}(\boldsymbol{Y}) = \mathrm{E}(g(\boldsymbol{X}))$ is defined by

$$\mathrm{E}(\boldsymbol{Y}) = \begin{cases} \sum_i g(\boldsymbol{x}_i)\, p_{\boldsymbol{X}}(\boldsymbol{x}_i)\,, & \text{if } \boldsymbol{X} \text{ is discrete,} \\ \int_{\mathrm{supp}\{\boldsymbol{X}\}} g(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\,, & \text{if } \boldsymbol{X} \text{ is continuous} \end{cases} \tag{1}$$

Because of the linearity of both operators (sum and integral), it follows that:

a)

$$\begin{aligned} \mathrm{E}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) &= \sum_i (\boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b})\, p_{\boldsymbol{X}}(\boldsymbol{x}_i) \\ &\underset{\text{linearity}}{=} \boldsymbol{A} \sum_i \boldsymbol{x}_i\, p_{\boldsymbol{X}}(\boldsymbol{x}_i) + \boldsymbol{b} \sum_i p_{\boldsymbol{X}}(\boldsymbol{x}_i) \\ &\underset{\text{definition}}{=} \boldsymbol{A}\,\mathrm{E}(\boldsymbol{X}) + \boldsymbol{b}\,, \end{aligned}$$

b)

$$\begin{aligned} \mathrm{E}(c_X \boldsymbol{X} + c_Y \boldsymbol{Y}) &= \sum_{i,j} (c_X \boldsymbol{x}_i + c_Y \boldsymbol{y}_j)\, p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}_i, \boldsymbol{y}_j) \\ &\underset{\text{linearity}}{=} c_X \sum_{i,j} \boldsymbol{x}_i\, p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}_i, \boldsymbol{y}_j) + c_Y \sum_{i,j} \boldsymbol{y}_j\, p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}_i, \boldsymbol{y}_j) \\ &\underset{\text{independency}}{=} c_X \sum_{i,j} \boldsymbol{x}_i\, p_{\boldsymbol{X}}(\boldsymbol{x}_i)\, p_{\boldsymbol{Y}}(\boldsymbol{y}_j) + c_Y \sum_{i,j} \boldsymbol{y}_j\, p_{\boldsymbol{X}}(\boldsymbol{x}_i)\, p_{\boldsymbol{Y}}(\boldsymbol{y}_j) \\ &\underset{\text{separation}}{=} c_X \sum_i \boldsymbol{x}_i\, p_{\boldsymbol{X}}(\boldsymbol{x}_i) \sum_j p_{\boldsymbol{Y}}(\boldsymbol{y}_j) + c_Y \sum_i p_{\boldsymbol{X}}(\boldsymbol{x}_i) \sum_j \boldsymbol{y}_j\, p_{\boldsymbol{Y}}(\boldsymbol{y}_j) \\ &\underset{\text{unitary}}{=} c_X \sum_i \boldsymbol{x}_i\, p_{\boldsymbol{X}}(\boldsymbol{x}_i) + c_Y \sum_j \boldsymbol{y}_j\, p_{\boldsymbol{Y}}(\boldsymbol{y}_j) \\ &\underset{\text{definition}}{=} c_X\, \mathrm{E}(\boldsymbol{X}) + c_Y\, \mathrm{E}(\boldsymbol{Y})\,, \end{aligned}$$

c)

$$\begin{aligned} \mathrm{E}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}) &= \sum_{i,j} \boldsymbol{x}_i^{\mathrm{T}}\, \boldsymbol{y}_j\, p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}_i, \boldsymbol{y}_j) \\ &\underset{\text{independency}}{=} \sum_i \sum_j \boldsymbol{x}_i^{\mathrm{T}}\, \boldsymbol{y}_j\, p_{\boldsymbol{X}}(\boldsymbol{x}_i)\, p_{\boldsymbol{Y}}(\boldsymbol{y}_j) \\ &\underset{\text{separation}}{=} \left(\sum_i \boldsymbol{x}_i\, p_{\boldsymbol{X}}(\boldsymbol{x}_i) \right)^{\mathrm{T}} \sum_j \boldsymbol{y}_j\, p_{\boldsymbol{Y}}(\boldsymbol{y}_j) \\ &\underset{\text{definition}}{=} \mathrm{E}(\boldsymbol{X})^{\mathrm{T}}\, \mathrm{E}(\boldsymbol{Y})\,. \end{aligned}$$

Note that the covariance $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})$ between two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined by $\mathrm{E}\left([\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})][\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})]^{\mathrm{H}}\right)$ while the covariance matrix of the random variable $\boldsymbol{Z}$ is given by $\mathrm{Cov}(\boldsymbol{Z}, \boldsymbol{Z})$ or in simple notation $\mathrm{Cov}(\boldsymbol{Z})$. Hence, this leads to

d)

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{AX} + \boldsymbol{b}) &= \mathrm{Cov}(\boldsymbol{AX} + \boldsymbol{b}, \boldsymbol{AX} + \boldsymbol{b}) \\
&\underset{\text{definition}}{=} \mathrm{E}\left([\boldsymbol{AX} + \boldsymbol{b} - \mathrm{E}(\boldsymbol{AX} + \boldsymbol{b})][\boldsymbol{AX} + \boldsymbol{b} - \mathrm{E}(\boldsymbol{AX} + \boldsymbol{b})]^{\mathrm{H}}\right) \\
&\underset{\text{a)}}{=} \mathrm{E}\left([\boldsymbol{AX} + \boldsymbol{b} - \boldsymbol{A}\,\mathrm{E}(\boldsymbol{X}) - \boldsymbol{b}][\boldsymbol{AX} + \boldsymbol{b} - \boldsymbol{A}\,\mathrm{E}(\boldsymbol{X}) - \boldsymbol{b}]^{\mathrm{H}}\right) \\
&\underset{\text{apply brackets}}{=} \mathrm{E}\left(\boldsymbol{A}[\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})][\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})]^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\right) \\
&\underset{\text{a)}}{=} \boldsymbol{A}\,\mathrm{E}\left([\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})][\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})]^{\mathrm{H}}\right)\boldsymbol{A}^{\mathrm{H}} \\
&\underset{\text{definition}}{=} \boldsymbol{A}\,\mathrm{Cov}(\boldsymbol{X})\,\boldsymbol{A}^{\mathrm{H}},
\end{aligned}$$

e) Similarly to the proof in d)

$$\begin{aligned}
\mathrm{Cov}(c_X\boldsymbol{X} + c_Y\boldsymbol{Y}) &= \mathrm{E}\left([c_X\boldsymbol{X} + c_Y\boldsymbol{Y} - \mathrm{E}(c_X\boldsymbol{X} + c_Y\boldsymbol{Y})][c_X\boldsymbol{X} + c_Y\boldsymbol{Y} - \mathrm{E}(c_X\boldsymbol{X} + c_Y\boldsymbol{Y})]^{\mathrm{H}}\right) \\
&= \mathrm{E}\left([c_X\boldsymbol{X} + c_Y\boldsymbol{Y} - c_X\,\mathrm{E}(\boldsymbol{X}) - c_Y\,\mathrm{E}(\boldsymbol{Y})][c_X\boldsymbol{X} + c_Y\boldsymbol{Y} - c_X\,\mathrm{E}(\boldsymbol{X}) - c_Y\,\mathrm{E}(\boldsymbol{Y})]^{\mathrm{H}}\right) \\
&= \mathrm{E}\left([c_X(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})) + c_Y(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))][c_X(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})) + c_Y(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))]^{\mathrm{H}}\right) \\
&= \mathrm{E}\left(|c_X|^2[\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})][\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})]^{\mathrm{H}} + |c_Y|^2[\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})][\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})]^{\mathrm{H}}\right. \\
&\quad \left. + c_X c_Y^{\mathrm{H}}[\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})][\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})]^{\mathrm{H}} + c_Y c_X^{\mathrm{H}}[\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})][\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})]^{\mathrm{H}}\right) \\
&= |c_X|^2\,\mathrm{Cov}(\boldsymbol{X}) + |c_Y|^2\,\mathrm{Cov}(\boldsymbol{Y}) \\
&\quad + \mathrm{E}\left(c_X c_Y^{\mathrm{H}}[\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})][\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})]^{\mathrm{H}} + c_Y c_X^{\mathrm{H}}[\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})][\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})]^{\mathrm{H}}\right) \\
&= |c_X|^2\,\mathrm{Cov}(\boldsymbol{X}) + |c_Y|^2\,\mathrm{Cov}(\boldsymbol{Y}) \\
&\quad + c_X c_Y^{\mathrm{H}}\,\mathrm{E}\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)\mathrm{E}\left(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\right)^{\mathrm{H}} + c_Y c_X^{\mathrm{H}}\,\mathrm{E}\left(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\right)\mathrm{E}\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)^{\mathrm{H}} \\
&= |c_X|^2\,\mathrm{Cov}(\boldsymbol{X}) + |c_Y|^2\,\mathrm{Cov}(\boldsymbol{Y}) \\
&\quad + c_X c_Y^{\mathrm{H}}\left[\mathrm{E}(\boldsymbol{X}) - \mathrm{E}(\boldsymbol{X})\right]\left[\mathrm{E}(\boldsymbol{Y}) - \mathrm{E}(\boldsymbol{Y})\right]^{\mathrm{H}} + c_Y c_X^{\mathrm{H}}\left[\mathrm{E}(\boldsymbol{Y}) - \mathrm{E}(\boldsymbol{Y})\right]\left[\mathrm{E}(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right]^{\mathrm{H}} \\
&= |c_X|^2\,\mathrm{Cov}(\boldsymbol{X}) + |c_Y|^2\,\mathrm{Cov}(\boldsymbol{Y}).
\end{aligned}$$

Solution of Problem 2

We have that $X \sim f_X(x)$ where

$$f_X(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}, \quad \text{for } \lambda > 0.$$

a) From the definition of the log-likehood function we obtain

$$\ell(\mathbf{x}, \lambda) = \sum_{i=1}^{n} \log f(x_i;\lambda) = \sum_{i=1}^{n} \log \lambda e^{-\lambda x_i} = \sum_{i=1}^{n} \log\lambda - \lambda x_i = n\log\lambda - \lambda\sum_{i=1}^{n} x_i\,,$$

with support $x_i \in (0, \infty)$ for all $i = 1, \ldots, n$.

b) In MLE, the estimate $\hat{\lambda}$ is obtained by solving

$$\hat{\lambda} = \arg\max_{\lambda} \ell(\mathbf{x}, \lambda) = \arg\max_{\lambda} n\log\lambda - \lambda \sum_{i=1}^{n} x_i \,.$$

Then, to find the λ that maximizes $\ell(\mathbf{x}, \lambda)$ we take the partial derivative $\frac{\partial}{\partial\lambda}\ell(\mathbf{x}, \lambda)$ and set it to zero. This leads to

$$\frac{\partial}{\partial\lambda}\ell(\mathbf{x}, \lambda) = \frac{1}{\lambda}n - \sum_{i=1}^{n} x_i \stackrel{!}{=} 0 \quad \Rightarrow \quad \lambda = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} \,.$$

Therefore, $\hat{\lambda} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$ is the MLE of λ.