Multivariate distributions

Question 1: Data and distributional assumptions

(a) Let us assume that we are given data *with all metric columns* of the following form:

$$\begin{array}{c|cccc} & M_1 & M_2 & \cdots & M_m \\ \hline 1: & x_{11} & x_{12} & \cdots & x_{1n} \\ 2: & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n: & x_{n1} & x_{n2} & \cdots & x_{nn} \end{array} .$$

How would, i.e. as what mathematical objects and using which probabilistic assumptions, would we model the elements of this data to then be able to make inferences about the behaviour/characteristics of new row-wise observations, like $[x_{(n+1)1}, x_{(n+1)2}, \ldots, x_{(n+1)m}]$?

(b) Show that the arithmetic mean is an unbiased estimate of the expected value, *given that we are viewing the points we are averaging over as realizations of random variables whose distributions all have the same expected value.*
Do we additionally need to assume that the random variables of which we have realizations are i.i.d.? Explain your answer.

(c) Given the setting of (a), consider the case $m = 1$, i.e. that we only have the data of column M_1, but are otherwise making the same modelling choices. Show that

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2$$

is an unbiased estimate for the variance of the distribution we are assuming.
Why are we looking only at the case of $m = 1$ here instead of considering the x_is to be vectors in $\mathbb{R}^m$ in the above equation?

Question 2: Eigenvalue decomposition

Consider the random vector $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)^T$ with covariance

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

a) Determine the eigenvalues λ_1 and λ_2 and the (normalized) eigenvectors of the matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$.

b) Use the result of (a) to determine a random vector $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)^T$, whose components $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are linear combinations of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and vor which additionally holds that

$$\text{Cov}(\boldsymbol{y}) = \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Question 3: Multivariate normal distribution

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)^T$ be a p-dimensional multivariate-normal distributed random vector. The corresponding density is given by

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu}$ denotes the expected value $\mathbb{E}[\boldsymbol{x}]$ and $\boldsymbol{\Sigma}$ the covariance $\text{Cov}(\boldsymbol{x})$.

a) Write out the form of this density for the the case $p = 2$, using the parameters $\sigma_i^2 = var(\boldsymbol{x}_i)$, $i = 1, 2$, and $\rho = \frac{cov(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\sigma_1 \sigma_2}$. Conclude from this that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent if they are uncorrelated.

b) Plot the density for $\boldsymbol{\mu} = \mathbf{0}$, $\sigma_1 = 1$, $\sigma_2 = 3$ and different values of ρ using `R`. (Tip: The function `persp` in combination with the function `manipulate` from the package of the same name is well suited for this).

Question 4: Dermination of marginal distributions

Consider the random variable $\boldsymbol{z} = (\boldsymbol{y}\ \boldsymbol{x})^T \sim \mathcal{N}_{q+p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{y}} \\ \boldsymbol{\mu}_{\boldsymbol{x}} \end{pmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{y}} & \boldsymbol{\Sigma}_{\boldsymbol{yx}} \\ \boldsymbol{\Sigma}_{\boldsymbol{xy}} & \boldsymbol{\Sigma}_{\boldsymbol{x}} \end{pmatrix}.$$

Derive the marginal distributions for the component vectors $\boldsymbol{y}$ and $\boldsymbol{x}$.