

# Advanced Statistics - Lab 07

- 1) Using build functions randomly generate  $N = 10^4$  Gaussian vectors with length  $d = 10^2$  from the distribution  $\mathcal{N}(0, \mathbf{I})$ , where  $\mathbf{I}$  is the  $d \times d$  identity matrix. Then, do PCA of these  $N$  vectors in three dimensions, without using build in functions for PCA and only use build in functions to obtain the required eigenvalues and eigenvectors. Plot the results of the PCA of these  $N$  vectors in a three dimensional plot.
- 2) Using build functions randomly generate  $N = 10^4$  Gaussian vectors with length  $d = 10^2$  from  $\mathcal{N}(0, \mathbf{\Sigma})$ , where  $\mathbf{\Sigma}$  is the  $d \times d$  covariance matrix obtained randomly (such that it satisfies the conditions to be a covariance matrix). Then, do PCA of these  $N$  vectors in three dimensions, without using build in functions for PCA and only use build in functions to obtain the required eigenvalues and eigenvectors. Plot the results of the PCA of these  $N$  vectors in a three dimensional plot.
- 3) Discuss the difference of the two plots/results and explain the causes for the difference.

**Upload the results on Moodle in a single PDF file or as the script itself that contains explanations, the code, and figures.**

---

**Important note:** By failing to do the following, you will loose points:

- You must provide clear explanation of what your program is doing.
- Use a log-scale when plotting probabilities or tails.
- You must provide comments in your code in order for anyone to understand the code.
- You must not use in-build functions for obtaining the PDF, mean, variance, and probability.
- You must use different colors, lines, and markers in the plots, along with legends for each curve and suitable line-widths of the curves so that the figure is understandable.
- You must clearly define what are the  $x$  and  $y$  axis in your figures.
- Finally, you must use caption that fully explains the figure.