Homework 3

Simulation and Performance Evaluation – University of Trento

DEADLINE: 17:29 on May 20, 2025

You can solve the following assignments using any programming language. Try to do the homework by yourselves, without help from AI tools. Unless otherwise stated, can use utility functions made available by the programming language of your choice, including functions to extract random numbers.

You will see a facility on Moodle to upload your homework. Please upload your code, and separately upload a **short** report where you describe your findings (no more than 3 pages). Upload **two versions** of the report: one with your names, and a second, fully anonymized one.

Exercise 1

Load the data from the CSV file `data_ex1_wt.csv`. This dataset contains measurements of some quantity over time, measured in seconds. In each line: the 1st value refers to the time of the measurement; the 2nd value is the measurement output. We are interested in modeling the statistics of the dataset by fitting a proper distribution to the data. For this, do the following:

1. Draw a scatter plot and verify that there is a trend in the data.
2. Use least squares to estimate the coefficient of a polynomial trend function, and remove the trend.
3. After having verified that you only need powers up to 5, remove the trend from the data and show the histogram of the resulting dataset.
4. Implement the Expectation-Maximization algorithm to fit a mixture of three Gaussian distributions to the data. Give the mean and variance of the distributions thus found, and plot the corresponding PDFs on top of the empirical PDF of the data (e.g., the histogram). Compare the mean and variance values you estimated against the values used to generate the dataset, as given below:

	μ	σ^2
Gaussian distribution #1	-5	3
Gaussian distribution #2	0	6
Gaussian distribution #3	4	1

   Do you observe any significant discrepancies?

5. *Facultative:* can you devise an algorithm to automatically establish that the best number of Gaussian distributions to fit the de-trended dataset is in fact 3?