1 Exercise 1

1.1 Problem statement

In the first exercise, the parameters of four Gaussian distributions (the means and the variances) and the probabilities to draw from each distribution were given.

The task was to generate samples conditionally to the probabilities and then compare the observed mean and variance of the resulting dataset with the theoretical values provided by the conditional mean and variance formulas.

We approached the problem by interpreting it as sampling from a random variable X, which, depending on the assigned probabilities, behaves according to one of the four Gaussian distributions.

First of all we take into account a random variable Y that basically takes four values with the following probability:

$$Y = \begin{cases} 1 & \text{with probability } 0.15 \\ 2 & \text{with probability } 0.25 \\ 3 & \text{with probability } 0.35 \\ 4 & \text{with probability } 0.25 \end{cases}$$

Then, we modeled our Gaussians in this way:

1. $X|Y = 1 \sim \mathcal{N}(-2, 2)$
2. $X|Y = 2 \sim \mathcal{N}(4, 1)$
3. $X|Y = 3 \sim \mathcal{N}(10, 3)$
4. $X|Y = 4 \sim \mathcal{N}(15, 2)$

1.2 Results and theoretical interpretation

To sample from our probability distribution X, we first generated a value from Y according to its given probabilities. This value was then used to determine which Gaussian distribution to sample from.

After having generated $N = 1\,000\,000$ samples we have computed the mean and the variance of the dataset (using the functions provided in the *numpy* module).

To verify our results, we simply applied the theoretical formulas:

$$E[X] = E[X|Y = 1] \cdot P\{Y = 1\} + \cdots + E[X|Y = 4] \cdot P\{Y = 4\}$$

and $Var(X) = E[Var(X|Y)] + Var(E[X|Y])$.

From the theoretical calculations, we obtained values of 7.95 and 34.7475 for the mean and variance, respectively. In practice, the results varied slightly across different simulations but remained consistent within a precision of 10^{-3}.

2 Exercise 2

2.1 Problem Statement

In this problem, it is required to compute the probability, via simulation, that an exponential random variable with mean $\mu = 1$ (that we called X) is larger than a uniformly distributed random variable in the interval $[0, 5]$ (that we called Y).

2.2 Simulation Methodology

To estimate $P(X > Y)$, we do the following:

1. Generate $N = 5\,000\,000$ samples from the exponential distribution with mean 1.
2. Generate $N = 5\,000\,000$ samples from the uniform distribution on $[0, 5]$.
3. For each pair (x_i, y_i), check whether $x_i > y_i$.
4. Compute the probability as the fraction of pairs for which $x_i > y_i$.

2.3 Theoretical Analysis

To compute the probability that the exponential random variable X is larger than the uniform random variable Y, we can express it theoretically as follows:

$$P(X > Y) = \int_0^5 P(X > y) f_Y(y)\, dy = \int_0^5 e^{-y} \cdot \frac{1}{5}\, dy = \frac{1}{5} \int_0^5 e^{-y}\, dy,$$

$$\int_0^5 e^{-y}\, dy = \left[-e^{-y}\right]_0^5 = -e^{-5} + 1 = 1 - e^{-5},$$

$$\Rightarrow P(X > Y) = \frac{1}{5}(1 - e^{-5}) \approx \frac{0.9933}{5} \approx 0.1987.$$

2.4 Results and Discussion

In the simulation process, we generated $N = 5\,000\,000$ samples from both the exponential and the uniform distributions, and then estimated $P(X > Y)$ as the fraction of times when the exponential sample exceeds the uniform sample. As the number of samples increases, the Law of Large Numbers ensures that the empirical probability converges to the theoretical value.

The theoretical analysis shows that the probability is driven by the exponential decay factor e^{-y} averaged over the interval $[0, 5]$. This is intuitive because for small y, e^{-y} is near 1, but it decreases rapidly as y increases.
The close match between the simulated and theoretical results provides strong evidence that our simulation method accurately captures the underlying probability distribution and validates the theoretical approach.

2.5 Conclusion

In summary, this exercise illustrates the effectiveness of Monte Carlo simulation in estimating probabilities and verifying theoretical results. By comparing the empirical probability computed from a large sample set with the theoretically derived value, we confirmed (with a high accuracy) that:

$$P(X > Y) = \frac{1}{5}\left(1 - e^{-5}\right) \approx 0.1987.$$