TP – Probability Distributions and Simulation in Python

Master 1 – Computer Science

Youssef SALMAN

November 26, 2025

Instructions

In this practical session, you will use Python (NumPy, SciPy, Matplotlib) to simulate probability distributions, compute empirical statistics, and compare them with the theoretical values. Each exercise includes simulation tasks and interpretation questions.

Exercise 1 — Simulation of a Discrete Distribution and a Normal Distribution

Consider the following two random variables:

- A discrete random variable X taking values $0, 1, 2$ with probabilities
$$P(X = 0) = 0.2, \quad P(X = 1) = 0.5, \quad P(X = 2) = 0.3.$$
- A continuous random variable Z following a standard normal distribution:
$$Z \sim \mathcal{N}(0, 1).$$

(a) Simulate $n = 10\,000$ realizations of X and Z in Python.

(b) Estimate the empirical probabilities $\widehat{P}(X = k)$ for $k = 0, 1, 2$ and compare them with the theoretical values.

(c) Estimate empirically the probability
$$\widehat{P}(-1 \leq Z \leq 1)$$
and compare it with the theoretical probability
$$P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1),$$
where Φ is the cumulative distribution function of the standard normal distribution.

Exercise 2 — Empirical Mean and Density of a Normal Distribution

Let X be a random variable following a normal distribution:

$$X \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu = 2,\ \sigma = 3.$$

(a) Simulate $n = 1\,000$ realizations of X.

(b) Compute the empirical mean $\overline{X}$ and empirical variance, then compare them to the theoretical values $\mu = 2$ and $\sigma^2 = 9$.

(c) Plot the normalized histogram of the simulated data and superimpose the theoretical density of the normal distribution $\mathcal{N}(2, 9)$.

(d) Repeat the simulation for $n = 10\,000$ and comment on how sample size affects:

- the accuracy of the empirical mean and variance,
- the quality of the histogram approximation.

Exercise 3 — Binomial Distribution: Simulation vs. Theory

Let Y be a binomial random variable:

$$Y \sim \mathcal{B}(n, p), \qquad n = 20,\ p = 0.4.$$

(a) Recall the theoretical formulas for the expectation and variance of a binomial distribution:

$$E[Y] = np, \qquad Var(Y) = np(1-p).$$

(b) Simulate $N = 5\,000$ realizations of Y.

(c) Compute the empirical mean and empirical variance, and compare them with the theoretical values.

(d) Estimate empirically the probabilities $\widehat{P}(Y = k)$ for $k = 0, 1, \ldots, 20$.

(e) Plot on the same graph:

- a bar chart of the empirical probabilities,
- the theoretical PMF of the binomial distribution.

Comment on the accuracy of the empirical approximation.