Simulation and Performance Evaluation
Homework 1

Blascovich Alessio and Di Noia Matteo

1st April 2025

Exercise 1

```
1  def main():
2      # Definition of normal distribution parameters
3      choice_mean = [-2, 4, 10, 15]
4      choice_var = [2, 1, 3, 2]
5      choice_prob = [0.15, 0.25, 0.35, 0.25]
6      vals = []
7
8      rng = np.random.default_rng()
9      RANGE = 1000000
10     for _ in range(RANGE):
11         mu, var = choices(list(zip(choice_mean, choice_var)), choice_prob)[0]
12         vals.append(rng.normal(mu, np.sqrt(var)))
13     print(np.mean(vals))
14     print(np.var(vals))
15     # [...]
16
```

This is the simulation we run to firstly extract randomly the normal distribution parameters, line 11, from `choice_var` and `choice_mean`, with the probability vector `choice_prob`. Then we generate a sample from the created normal distribution, this sample is appended to a vector of values collecting all the sampling, The parameters pick and the sampling process was repeated 1000000 times. On the resulting vector of values we computed arithmetic mean and arithmetic variance, line 13 and line 14.

```
def compute_mean(values, probs):
    return sum(v * p for v, p in zip(values, probs))

def compute_var(values,  probs):
    squared = map(lambda x: x**2, values)
    return compute_mean(squared, probs) - compute_mean(values, probs)**2

def main():
    # [...]
    print("EXPECTED ARE:")
    expected_mean = compute_mean(choice_mean, choice_prob)
    expected_var = (compute_var(choice_mean, choice_prob)
                 + compute_mean(choice_var, choice_prob))
    print(expected_mean)
    print(expected_var)
```

To validate the result against the expected we computed the theoreticial mean and variance of the conditional distribution. For the expected value we simply used the formula of the conditional mean, meanwhile for the variance we applied the conditional variance formula.

We call X the random variable obtained by conditioning the normal distribution parameters with probability vector given.

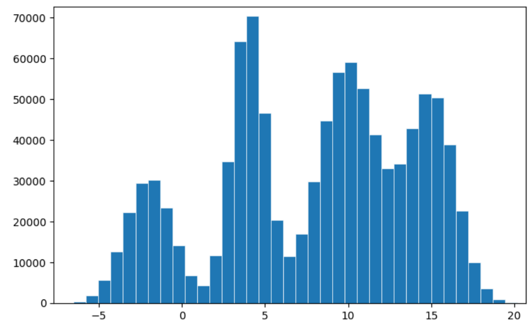
$$\mathbb{E}[X] = \sum_{i=1}^{4} \mathbb{E}[X|Y_i]P(Y_i)$$

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y])$$

The Python script described above produced:

```
EXPECTED ARE:
7.95
34.747499999999995
COMPUTING...
7.951787799337237
34.71597133688908
```

Which leads us to conclude resulting dataset obeys the conditional mean and conditional variance formulae.



The plot of the random values obtained during the sampling phase presents four peaks in correspondence of the four normal means and the height of these peaks is proportional to the probability of choosing that distribution parameters.

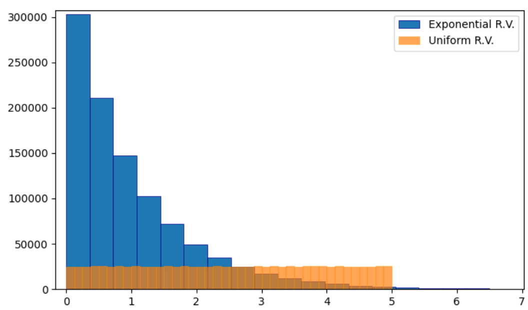
Exercise 2

```
RANGE = 1000000
vals_exp: list = []
vals_uni: list = []
rng = np.random.default_rng()

# Generate 1M random value of exponential and uniform
for _ in range(RANGE):
vals_exp.append(rng.exponential(1))
vals_uni.append(rng.uniform(0, 5))

# Count how many exponential are grater than the corresponding uniform
greater = sum(1 for i in range(RANGE) if vals_exp[i] > vals_uni[i])
print(greater / RANGE)
```

We used the above code to generate 1000000 draws from each distribution. We considered 1000000 an acceptable sample size since it was the same size suggested by the professor to obtain statisticial meaningful data in exercise 1. Then, we counted how many exponential draws were larger that the corresponding uniform draw. The result was then divided by the length of the dataset to obtain the probability for which the exponential random variable is greater that the uniform random variable. In our tests the probability was 0.198286.



Facultative

Let $X \sim \text{Exp}(\lambda = 1)$ and let $Y \sim U(0, 5)$, this exercise ask to quantify $P(X > Y)$. We have:

$$f_X(x) = e^{-x} \quad \text{and} \quad F_X(x) = 1 - e^{-x} \quad \text{with} \quad x \in [0, \infty) \tag{1}$$

$$f_Y(y) = \frac{1}{5} \quad \text{and} \quad F_Y(y) = \frac{y}{5} \quad \text{with} \quad y \in [0, 5] \tag{2}$$

We can now define:

$$P(X > Y) = \int_0^5 P(X > y) F_Y(y) dy + \int_5^\infty f_X(x) dx$$

Our goal can be decompose in, the probability that X is greater than a given value y plus the probability that the exponential random variable has a value greater than 5. We can simply the above formula using the definition of probability and the definitions in 1 and 2.

$$P(X > Y) = \int_0^5 (1 - F_X(y)) \frac{y}{5} dy + (1 - F_X(5)) = \int_0^5 e^{-y} \frac{y}{5} dy + e^{-5} \tag{3}$$

Solving the first integral by part we obtain:

$$\frac{1}{5} \int_0^5 y e^{-y} dy = \frac{1}{5} \left[-y e^{-y} + \int e^{-y} dy \right]_0^5 = \frac{1}{5} \left[-y e^{-y} - e^{-y} \right]_0^5 = \frac{1 - 6e^{-5}}{5}$$

Applying the above result to 3, the aimed probability results in:

$$P(X > Y) = \frac{1 - 6e^{-5}}{5} + e^{-5} \approx 0.19865$$

This calculated value is in line with the experimental value.