ENSIIE 2024-2025
2A UE Simulation methods

Practical session 1

1. Inversion method

Recall. Suppose that we may sample a sequence $(U_n)_{n\geq 1}$ of random variables (r.v.) from the uniform distribution $\mathcal{U}]0,1[$ on $]0,1[$. In practice, we simulate a random number $u = U(\omega)$ from the uniform distribution on $]0,1[$ from the function we will call by `rand`.

Let $U \sim \mathcal{U}(]0,1[)$ and let X be a random variable with cdf F: $F(x) = \mathbb{P}(X \leq x)$, $\forall\, x \in \mathbb{R}$. To simulate a sample $(X_n)_{n\geq 1}$ from X (means, the X_n's are independent and has the same distribution as X), we use the following result: if

$$F^{-1}(u) := \inf\{t \in \mathbb{R} :\ F(t) \geq u\}, \ \text{ for all } \ u \in]0,1[$$

$$\text{then} \quad X \text{ and } F^{-1}(U) \text{ has the same distribution : } X \overset{d}{=} F^{-1}(U).$$

Exercice 1. (*Bernoulli distribution*) Let U be a r.v. with uniform distribution on $]0,1[$ and let $p \in]0,1[$.

1. Show that $1-U$ and U have the same distribution: if $U \sim \mathcal{U}(]0,1[)$ then $1-U \sim \mathcal{U}(]0,1[)$.
2. Show that X has the same distribution as $\mathbb{1}_{\{U\leq p\}}$ and $\mathbb{1}_{\{1-U\leq p\}}$.
3. Use the previous question to simulate (and display) a sample $(X_k)_{1\leq k\leq N}$ of size N of the $\mathcal{B}ern(p)$, with $p = 0.6$, and $N = 10$.
4. For $N = 100$, $N = 1000$ and then $N = 10\,000$, simulate a sample $(X_k)_{1\leq k\leq N}$ of size N of the $\mathcal{B}ern(p)$, with $p = 0.8$, and, for each N, compute and compare the following quantity with p:

$$\frac{\#\{i \in \{1,\cdots,N\} \text{ s.t. } X_i = 1\}}{N}$$

5. Recall the *Central Limit Theorem* (CLT): Let $(X_n)_{n\geq 1}$ be an iid sequence of r.v. with expectation $\mu = \mathbb{E}(X_1)$ and finite variance σ^2. Then

$$Z_n := \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \quad \xrightarrow{d} \quad Z \sim \mathcal{N}(0,1), \text{ when } n \to \infty, \tag{1}$$

where $\bar{X}_n = (X_1 + \ldots + X_n)/n$ is the empirical mean associated to a sample of size n of $(X_n)_{n\geq 1}$.

(a) Generate and display a sample $(Z_n^i)_{i=1,\ldots,N}$ of size $N = 10$ of Z_n, for $n \in \{10, 30, 50\}$.

(b) For $n = 10$, plot on the same graphic the histogram associated to a sample $(Z_n^i)_{i=1,\ldots,N}$ of size $N = 10^5$ and the density function of the standard normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

(c) Take up the previous question for $n = 30$ and for $n = 50$.

(d) Comment the graphics obtained in the two previous questions.

Exercice 2. (*Discrete random variable*) Let X be a r.v. taking values in $\{a_1, a_2, a_3, a_4\}$ with $a_1 = 0.5$; $a_2 = -9$ $a_3 = -1.5$; $a_4 = 7$ and let the associated weights $(p_i)_{i=1,\cdots,4}$ be defined by

$$\begin{cases} p_1 = \mathbb{P}(X = a_1) = 1/4 \\ p_2 = \mathbb{P}(X = a_2) = 1/8 \\ p_3 = \mathbb{P}(X = a_3) = 1/8 \\ p_4 = \mathbb{P}(X = a_4) = 1/2. \end{cases}$$

1. Generate a sample of size $N = 10000$ of X and compare the empirical frequencies with their associated weights.

2. Let $\mu = \mathbb{E}(X)$, $\sigma^2 = Var(X)$ and let $(Z_n)_{n\geq 1}$ be the sequence defined for every $n \geq 1$ by

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \tag{2}$$

where $\bar{X}_n = (X_1 + \ldots + X_n)/n$ is the sample mean.

3. For $n = 10$, plot on the same graphic the histogram associated to a sample $(Z_n^i)_{i=1,\ldots,N}$ of size $N = 10^5$ and the density function of the standard normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

4. Take up the previous question for $n = 30$ and for $n = 50$.

5. Comment the graphics obtained in the two previous questions.

Exercice 3. (*Choose randomly groups*) Let $n \geq 2$, $p < n$ and $E = \{1, \ldots, n\}$. We want to compose randomly p groups in the following way. The set of group numbers is $G = \{g_1, \ldots, g_p\}$ where g_i stands for the group i, for any $i \in G$. We suppose initially that for any $i \in G$, $g_i = \emptyset$. When p is not a divisor of n: $n = q \times p + r$, with $r \neq 0$, one of the groups will be of size $q + r$ and the others will be of size q. We repeat the following steps until $card(E) > 1$:

1. We choose randomly a group g following the uniform distribution $\mathcal{U}(G)$ on G: $g \sim U(G)$.

2. We then sample a random number $x \sim \mathcal{U}(E)$ and set $E = E\backslash\{x\}$ and $g = g \cup \{x\}$.

3. If the previously randomly chosen group g is such that $card(g) = q$, then we set $G = G\backslash\{g\}$ and suppose that the group g is full.

Write down a function that implements the previous steps for $(p, n) = (3, 9)$ and then for $(p, n) = (10, 101)$.

Exercice 4. (*Geometric distribution*) Let X be a r.v. with geometric distribution with success parameter p. Write down a function that simulate a sample of a given size n of X and implement it for $n = 10$ and $p = 0.4$.

2. Rejection and transformation method

Exercice 5. We want to sample a random number from the $\mathcal{N}(0, 1)$ distribution using the rejection method.

1. Show that if X has the $\mathcal{N}(0, 1)$ distribution then, $|X|$ has the density

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathbb{1}_{\{x>0\}}.$$

2. Show that for any $x > 0$,

$$f(x) \leq \sqrt{\frac{2e}{\pi}}\, e^{-x}.$$

3. Write down a function that sample a random number from $|X|$ using the rejection method. Generate a sample of size $N = 10^5$ and compare the associated histogram with the density of $|X|$.

4. Let Θ be a r.v. valued in $\{-1, +1\}$ with $\mathbb{P}(\Theta = +1) = \mathbb{P}(\Theta = -1) = 1/2$. Show that $X \overset{d}{=} \Theta\,|X|$.

5. Generate a sample of size $N = 10^5$ of the standard gaussian distribution and compare the associated histogram with the density of X.

Exercice 6. Let X be a r.v. with uniform distribution on the unit sphere $A = \{(x, y, z) \in \mathbb{R}^3,\ x^2 + y^2 + z^2 \leq 1\}$ and let $S = \{(x, y, z) \in \mathbb{R}^3,\ x \in]-1, +1[,\ y \in]-1, +1[,\ z \in]-1, +1[\}$, so that $A \subset S$.

Generate and plot a sample of size $N = 10000$ of X.

Exercice 7. (*Box-Muller method*). Let R be a r.v. with exponential distribution with parameter $1/2$: $R \sim Exp(1/2)$ and let $\Theta \sim \mathcal{U}(]0, 2\pi[)$.

1. Using the fact that the vector $(X_1, X_2) = (\sqrt{R}\cos(\Theta), \sqrt{R}\sin(\Theta))$ is gaussian $\mathcal{N}(0; I_2)$, where I_2 is the identity matrix on $\mathbb{R}^2$, generate and display a sample of size $N = 10$ of the bi-dimensional gaussian distribution.
2. Generate a sample $(X_1^k, X_2^k)_{\{1 \leq k \leq N\}}$ of (X_1, X_2) of size $N = 10^5$ plot the sampled point cloud $\{(X_1^k, X_2^k),\ 1 \leq k \leq N\}$.
3. Plot in two different graphics, the density of (X_1, X_2) and the histogram associated to a sample of size $N = 10^5$ of (X_1, X_2).
4. Using the item 1., generate a sample of $X_1 \sim \mathcal{N}(0; 1)$ of size $N = 10^5$. Plot the associated histogram and compare it (in the same graph) with the density of the $\mathcal{N}(0; 1)$.
5. Let $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_{ij} \geq 0$, for $i, j \in \{1, 2\}$. Let

$$\begin{cases} Z_1 = \mu_1 + \sigma_{11}X_1 + \sigma_{12}X_2 \\ Z_2 = \mu_2 + \sigma_{21}X_1 + \sigma_{22}X_2. \end{cases}$$

We show that $Z = (Z_1, Z_2) \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and σ_1, σ_2, ρ are defined by:

$$\sigma_1^2 = \sigma_{11}^2 + \sigma_{12}^2, \qquad \sigma_2^2 = \sigma_{21}^2 + \sigma_{22}^2, \qquad \rho = \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\sigma_1\sigma_2}.$$

(a) We set $\mu_1 = \mu_2 = 0$. Choose the values of σ_{ij} such that $\rho = 0.1$, $\rho = 0.5$ and then $\rho = 0.9$. Generate and plot a sample of size $N = 10^4$ of (Z_1, Z_2), for each $\rho \in \{0.1, 0.5, 0.9\}$. Comment the plots.

(b) For each $\rho \in \{0.1, 0.5, 0.9\}$, represent the density of (Z_1, Z_2) which reads:

$$f_{(Z_1,Z_2)}(z_1, z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big(-\frac{z}{2(1-\rho^2)}\Big), \quad \text{avec } z = \frac{z_1^2}{\sigma_1^2} - \frac{2\rho z_1 z_2}{\sigma_1\sigma_2} + \frac{z_2^2}{\sigma_2^2}.$$

Comment the plots.

Exercice 8. (*Mixed density*) Let $(p_1, p_2, p_3) = (1/6, 1/3, 1/2)$ and X be a r.v. with density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + p_3 f_3(x)$$

where $f_1(x) = \mathbb{1}_{]0,1]}(x),\ f_2(x) = \frac{1}{2}(2x - 1)\mathbb{1}_{]1,2]}(x),\ f_3(x) = \frac{2}{3}(-3x + 9)\mathbb{1}_{]2,3]}(x)$.

Generate a sample of size $N = 10^5$ of X and compare (in the same graph) the associated histogram with the density of X.

Exercice 9. Let $X_1 \sim \mathcal{N}(-3, 1)$ and $X_2 \sim \mathcal{N}(3, 1)$ two independent r.v. with resp. densities f_1 and f_2. Let X be the r.v. l with density

$$f(x) = p_1 f_1(x) + p_2 f_2(x), \qquad p_1, p_2 \in [0, 1],\ p_1 + p_2 = 1.$$

Generate a sample of size $N = 10^5$ of X and compare (in the same graph) the associated histogram with the density of X in the following cases: $(p_1, p_2) = (1/2, 1/2)$, $(p_1, p_2) = (1/8, 7/8)$, $(p_1, p_2) = (7/8, 1/8)$.

Exercice 10. Consider the file `ToBeGropuedRandomly.xlx` and implement a code that compose randomly 5 groups of 3 individuals where, at each step $k \leq 5$ and for each group,

- 1 individual of the group is chosen randomly from the uniform distribution among the $5 - k + 1$ first individuals of the file
- the 2 other individual of the group are chosen randomly from the uniform distribution among the remaining individuals

Display the groups with their set of individuals.