ENSIIE 2018-2019
2A UE Méthodes de Simulation

Feuille 1 de Travaux pratiques

1. Méthode d'inversion

Rappel de la méthode d'inversion. On suppose que l'on sait simuler des réalisations indépendantes de loi uniforme sur $]0, 1[$, c'est-à-dire, une suite $(U_n)_{n\geq 1}$ de variables aléatoires indépendantes de même loi $\mathcal{U}]0, 1[$. L'appel à la fonction `rand` génère une réalisation $u = U(\omega)$ de la loi uniforme sur $]0, 1[$.

Soit $U \sim \mathcal{U}(]0, 1[)$ et soit X une variable aléatoire de fonction de répartition F: $F(x) = \mathbb{P}(X \leq x)$, $\forall\, x \in \mathbb{R}$. Pour simuler des réalisations de variables aléatoires $(X_n)_{n\geq 1}$ indépendantes de même loi que X on utilise le résultat suivant: si

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

alors X et $F^{-1}(U)$ ont même loi: on note $X \overset{d}{=} F^{-1}(U)$.

Exercice 1. (*Loi de Bernoulli*) Soit U une loi uniforme sur $]0, 1[$ et soit $p \in]0, 1[$.

1. Vérifier que la variable aléatoire $X = \mathbb{1}_{\{U\leq p\}}$ suit une loi de Bernoulli de paramètre p.
2. Utiliser la question précédente pour simuler un échantillon indépendant $(X_i)_{1\leq i\leq N}$ de taille N de la loi de Bernoulli de paramètre $p = 0.3$, pour $N = 100, 1000$ puis 10000.
3. Calculer pour chaque N, la quantité
$$\frac{\#\{i \in \{1, \cdots, N\} \text{ tel que } X_i = 1\}}{N}$$
et le comparer avec p.
4. On rappelle le *Théorème Central Limite* (TCL): Pour toute suite de variables aléatoires $(X_n)_{n\geq 1}$ iid de moyenne $\mu = \mathbb{E}(X_1)$ et de variance finie σ^2, on a:
$$Z_n := \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \quad \overset{n\to\infty}{\longrightarrow} \quad Z \sim \mathcal{N}(0, 1), \tag{1}$$
où $\bar{X}_n = (X_1 + \ldots + X_n)/n$ est la moyenne empirique associée à l'échantillon.
   (a) Gérérez un échantillon $(Z_n^i)_{i=1,\ldots,N}$ de taille $N = 10$ de la loi de Z_n pour $n = 10, 30, 40$.
   (b) Gérérez un échantillon $(Z_n^i)_{i=1,\ldots,N}$ de taille $N = 10^5$ de la loi de Z_n pour $n = 10, 30, 40$. Pour tout $n \in \{10, 30, 40\}$, tracez sur le même graphique l'histogramme empirique associé à l'échantillon $(Z_n^i)_{i=1,\ldots,N}$ et la densité
$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$
d'une loi $\mathcal{N}(0, 1)$.
   (c) Commentez les graphiques obtenus dans la question précédente.

Exercice 2. (*Loi binomiale*) Soit $n = 30$ et $p = 0.1$. Soit $(U_k)_{k\geq 1}$ une suite de variables aléatoires indépendantes de loi uniforme sur $]0, 1[$.

1. Vérifiez, que la variable aléatoire X suivante suit une loi binomiale de paramètres (n, p):
$$X = \sum_{k=1}^{n} \mathbb{1}_{\{U_k\leq p\}}.$$

2. Générez un échantillon indépendant de taille $N = 10000$ de la loi binomiale de paramètres (n, p) en utilisant la question précédente.

3. Tracez l'histogramme des fréquences empiriques associées à l'échantillon simulé.

4. Tracez le graphe $k \in \{0, \cdots, n\} \mapsto p_k = \mathbb{P}(X = k)$, $k = 0, \cdots, n$ et comparez le avec l'histogramme des fréquences empiriques.

5. On rappelle à nouveau le *Théorème Central Limite* (TCL): Pour toute suite de variables aléatoires $(X_k)_{k\geq 1}$ iid de moyenne $\mu = \mathbb{E}(X_1)$ et de variance finie σ^2, on a:

$$Z_k := \sqrt{k}\frac{\bar{X}_k - \mu}{\sigma} \quad \overset{k\to\infty}{\longrightarrow} \quad Z \sim \mathcal{N}(0, 1), \tag{2}$$

   où $\bar{X}_k = (X_1 + \ldots + X_k)/k$ est la moyenne empirique associée à l'échantillon.

   (a) Gérérez un échantillon $(Z_k^i)_{i=1,\ldots,N}$ de taille $N = 10$ de la loi de Z_k pour $k = 10, 30, 40$.

   (b) Gérérez un échantillon $(Z_k^i)_{i=1,\ldots,N}$ de taille $N = 10^5$ de la loi de Z_n pour $k = 10, 30, 40$. Pour tout $k \in \{10, 30, 40\}$, tracez sur le même graphique l'histogramme empirique associé à l'échantillon $(Z_k^i)_{i=1,\ldots,N}$ et la densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

   d'une loi $\mathcal{N}(0, 1)$.

6. Reprendre toute la question 5. mais avec cette fois-ci $p = 0.5$. Comparez les résutats obtenus avec ceux de la question 5.

Exercice 3. (*Loi discrète quelconque*) Soit X une variable aléatoire prenant ses valeurs dans $\{a_1, a_2, a_3, a_4\}$ avec $a_1 := 0.5$; $a_2 := 1$ $a_3 := 1.5$; $a_4 := 2$ et soit $(p_i)_{i=1,\cdots,4}$ les poids associés définis par

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

1. Simuler un échantillon indépendant de taille $N = 10000$ de X et tracer l'histogramme des fréquences.

2. Comparer l'histogramme au graphe $i \in \{1, \cdots, 4\} \mapsto p_i$.

Exercice 4. Ecrivez une fonction qui considère les lois usuelles: *Bernouilli, binomiale, de Poisson, exponentielle, Weibull* et qui, pour une loi choisie:

1. Demande en entrée les paramètres de la loi.

2. Trace sur le même graphique l'histogramme des fréquences empiriques de Z_k dans (2), pour $k = 30$, et la densité d'une loi normale standard (autrement dit, qui vérifie le TCL).

2. Méthodes de rejet et de transformation

Rappel de la méthode de rejet. On veut simuler une v.a. X de densité f. On suppose qu'il existe une v.a. Y que l'on sait simuler, de densité g vérifiant pour tout $x \in \mathbb{R}$, $f(x) \leq cg(x)$, pour une constante $c \geq 1$. On considère l'algorithme suivant:

1. `Simuler` U `de loi` $\mathcal{U}([0,1])$ `et` Y `de densité` g `et poser` $h(Y) = f(Y)/(cg(Y))$;

2. `Si` $U < h(Y)$, `on pose` $X = Y$, `sinon on retourne en` 1.

Alors on montre que l'algorithme converge en temps fini et que X est de densité f.

Exercice 5. On veut générer une réalisation de la loi $\mathcal{N}(0,1)$ par la méthode de rejet.

1. Montrer que si X est de loi $\mathcal{N}(0,1)$, alors $|X|$ a pour densité
$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathbb{1}_{\{x>0\}}.$$
2. Montrer que pour tout $x > 0$,
$$f(x) \leq \sqrt{\frac{2e}{\pi}}\, e^{-x}.$$
3. En déduire un algorithme de simulation de $|X|$ par la méthode de rejet.
4. Tracer l'histogramme des fréquences pour un échantillon de taille N assez grand et le comparer avec la densité théorique de $|X|$.
5. Soit Θ une v.a. ăleurs dans $\{-1,+1\}$ avec $\mathbb{P}(\Theta = +1) = \mathbb{P}(\Theta = -1) = 1/2$. Montrer que X a la même loi de probabilité que $\Theta|X|$.
6. Simuler un échantillon indépendant de taille N de loi normale en utilisant les questions précédentes et comparer les densités théoriques et empiriques.

Exercice 6. Soit X une variable aléatoire de loi uniforme sur la boule unité $A = \{(x,y) \in \mathbb{R}^2,\ x^2+y^2 \leq 1\}$ et soit le domaine $S = \{(x,y) \in \mathbb{R}^2,\ x \in]-1,+1[,\ y \in]-1,+1[\}$, de sorte que $A \subset S$.

1. L'algorithme suivant simule une réalisation de quelle variable aléatoire (on rappelle que `rand()` génère une réalisation d'une loi uniforme sur]0,1[):
```
do    u1 ← 2*rand() -1
      u2 ← 2*rand() -1
while (u1*u1 + u2*u2 > 1)
end
U1 ← u1 and U2 ← u2
```
2. Générer un échantillon indépendant de taille $N = 10000$ de la v.a. précédente et représenter les points simulés sur $\mathbb{R}^2$.

Exercice 7. (*Méthode de Box-Muller*). Soit R une v.a. de loi exponentielle de paramètre $1/2$: $R \sim \text{Exp}(1/2)$, et soit $\Theta \sim \mathcal{U}(]0,2\pi[)$, avec .

1. Montrer que $X_1 = \sqrt{R}\cos(\Theta)$ et $X_2 = \sqrt{R}\sin(\Theta)$ sont des v.a. gaussiennes indépendantes, centrées et réduites. On rappelle que la densité f de la $\mathcal{N}(0,1)$ est donnée pour tout $x \in \mathbb{R}$ par
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$
2. En déduire une simulation de $N = 10000$ réalisations indépendantes de la loi gaussienne
$$Z = (X_1, X_2) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$
et représenter les points générés sur $\mathbb{R}^2$.

3. Tracer la densité empirique de votre échantillon de la question précédente et le comparer avec la densité théorique.
4. En utilisant la question 1., simuler un N-échantillon de $X_1 \sim \mathcal{N}(0;1)$ de taille $N = 10000$.
5. Tracer la densité empirique de votre échantillon de la question précédente et le comparer avec la densité théorique.
6. Soit $\mu_1, \mu_2 \in \mathbb{R}$ et $\sigma_{ij} \geq 0$, pour $i, j \in \{1,2\}$. Soit

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

On montre que $Z = (Z_1, Z_2) \sim \mathcal{N}(\mu, \Sigma)$ où

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

et σ_1, σ_2, ρ sont définis par:

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

(a) On pose $\mu_1 = \mu_2 = 0$. Choisir des valeurs pour les σ_{ij} pour que ρ prenne les valeurs 0.1, 0.5, 0.9 et pour chaque valeur de ρ générer un échantillon de taille $N = 10000$ de (Z_1, Z_2).

(b) Repésenter les points des échantillons de la question précédente pour chaque valeur de ρ et commenter les graphes.

(c) Représenter les densités empiriques pour chaque valeur de ρ et commenter les graphes.

Exercice 8. (*Mélange de lois*) Soit $X_1 \sim \mathcal{N}(-3,1)$ et $X_2 \sim \mathcal{N}(3,1)$ deux v.a. indépendantes de densités respectives f_1 et f_2. Soit X la v.a. de densité

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

- Représenter graphiquement f pour $(p_1,p_2) = (1/2, 1/2)$, $(p_1,p_2) = (1/4, 3/4)$, $(p_1,p_2) = (3/4, 1/4)$.
- Générer un échantillon de taille $N = 10000$ de la loi de X pour chaque valeur de (p_1, p_2) et représenter les densités empiriques associées. Commenter les graphes obtenus.