TP 4 – Simulation de variables aléatoires à densité

1 Simulation des lois usuelles

1.1 Fonctions Python

On commencera par importer les librairies :

```
import numpy as np
import numpy.random as rd
```

Les lois à densité au programme peuvent être simulées à l'aide des commandes suivantes.

Loi	*Syntaxe*
$\mathcal{U}([0,1[)$	`rd.random()` ou `rd.rand()`
$\mathcal{N}(0,1)$	`rd.normal()`
$\mathcal{E}(\lambda)$	`rd.exponential(1/lamb)`
$\gamma(\nu)$	`rd.gamma(nu)`

1.2 Loi uniforme

Exercice 1. Soient a et b deux réels avec $a < b$. On rappelle que si $U \hookrightarrow \mathscr{U}([0,1])$, alors on a $(b-a)U + a \hookrightarrow \mathscr{U}([a,b])$.

Écrire une fonction `uniforme(a,b)` simulant la loi $\mathscr{U}([a,b])$ à l'aide de la fonction `rd.rand()` uniquement.

1.3 Loi normale

Rappel. Soient $m \in \mathbb{R}$ et $\sigma \in]0,+\infty[$. Si X suit la loi $\mathscr{N}(0,1)$, alors $\sigma X + m \hookrightarrow \mathscr{N}\left(m, \sigma^2\right)$.

Exercice 2. Écrire une fonction `normale(m,s)` simulant la loi $\mathscr{N}\left(m, s^2\right)$ à partir de la fonction `rd.normal()`.

2 Méthode d'inversion

2.1 Principe

Théorème 1. *Soient X une variable aléatoire à densité et F_X sa fonction de répartition.*

On suppose que F_X est strictement croissante sur $]a,b[$, où $-\infty \leqslant a < b \leqslant +\infty$, et $F_X(]a,b[) =]0,1[$. Alors, si $U \hookrightarrow \mathscr{U}(]0,1[)$,

$$F_X{}^{-1}(U) \text{ suit la même loi que } X,$$

où $F_X^{-1} :]0,1[\to]a,b[$ est la bijection réciproque de F_X.

Exercice 3.

1. Rappeler l'expression de la fonction de répartition de $U \hookrightarrow \mathscr{U}([0,1])$.
2. Démontrer le théorème précédent.

Ceci fournit une manière de simuler une loi à densité, dans le cas où sa fonction de répartition vérifie les hypothèses du théorème.

Simulation à l'aide de la méthode d'inversion.
Soient X une variable aléatoire à densité et F_X sa fonction de répartition. Si on dispose d'une expression explicite de $F_X{}^{-1}$, on peut simuler X en procédant comme suit :

1. on fait une simulation de la loi $\mathcal{U}([0,1])$: `u=rd.rand()`
2. on retourne $F_X{}^{-1}(u)$.

2.2 Simulation de la loi exponentielle

Exercice 4.

1. Rappeler l'expression de la fonction de répartition d'une loi exponentielle, et montrer qu'elle réalise une bijection de $\mathbb{R}_+^\star$ sur $]0,1[$. Déterminer sa bijection réciproque.
2. Écrire une fonction `exponentielle(lamb)` simulant une loi $\mathscr{E}(\lambda)$ à partir de la fonction `rd.rand()`.
3. Écrire la fonction `Exponentielle(lamb,N)` donnant un échantillon de taille N de la loi $\mathscr{E}(\lambda)$.
4. Créer un vecteur de taille 10 000 contenant 10 000 simulations d'une variable aléatoire suivant la loi $\mathscr{E}(1/2)$.
5. En utilisant les commandes `np.mean` et `np.var`, vérifier que la moyenne et l'écart-type sont bien conformes à ce qu'on attend.

2.3 Simulation de la loi de Cauchy

Exercice 5. Une variable aléatoire X suit la loi de Cauchy si elle admet pour fonction de répartition

$$F : x \mapsto \frac{1}{\pi}\left(\arctan x + \frac{\pi}{2}\right),$$

et donc pour densité $f : t \mapsto \dfrac{1}{\pi}\dfrac{1}{1+t^2}$.

1. Montrer que F réalise une bijection de $\mathbb{R}$ sur $]0,1[$, et déterminer F^{-1}.

2. *a.* Écrire une fonction `cauchy()` qui simule la loi de Cauchy, à partir de la fonction `rd.rand()`.

   b. Écrire une fonction `Cauchy(N)` qui donne un échantillon de taille `N` de la loi de Cauchy.

3. Créer un tableau contenant 10 000 simulations de la loi de Cauchy, et déterminer sa moyenne. Recommencer avec plusieurs échantillons. Qu'en déduire sur l'espérance de la loi de Cauchy ?

3 Représentation graphique

À l'aide d'un grand nombre de simulations d'une loi à densité (échantillon), on peut utiliser la représentation d'un histogramme pour approcher la courbe d'une densité de cette loi. Il s'agit de :

- séparer les données de l'échantillon en classes de valeurs, délimitées par $c_0 < \ldots < c_n$,
- tracer pour chaque i le rectangle de largeur $[c_i, c_{i+1}]$, et d'aire la fréquence associée à la classe de valeurs $[c_i, c_{i+1}]$, c'est-à-dire la proportion dans l'échantillon de données dans cette classe.

Pour tracer l'histogramme des fréquences à partir d'un échantillon `A`, la commande Python est

```
import matplotlib.pyplot as plt
hist(A,n,density=True)
```

où `n` désigne le nombre de classes. Il peut alternativement être remplacé par le tableau des c_i.

L'histogramme obtenu est alors approximativement superposable à la courbe d'une densité f de la loi considérée. En effet, pour tout i,

$$\mathbb{P}(c_i \leqslant X \leqslant c_{i+1}) \;=\; \int_{c_i}^{c_{i+1}} f(x)\,\mathrm{d}x.$$

Si l'échantillon est assez grand, on sait que la proportion de données dans $[c_i, c_{i+1}]$ est une approximation de $\mathbb{P}(c_i \leqslant X \leqslant c_{i+1})$. Ainsi, l'aire du rectangle de base $[c_i, c_{i+1}]$ dans l'histogramme approche l'aire sous la courbe de f entre c_i et c_{i+1}.

Exercice 6.

1. Simuler avec la fonction `Exponentielle` $N = 10\,000$ valeurs de la loi $\mathscr{E}(0.5)$.

2. Tracer la courbe représentative de la densité f de la loi $\mathscr{E}(0.5)$.

3. Tracer l'histogramme des fréquences de l'échantillon obtenu (on prendra pour cela 100 classes de même taille). Comparer l'histogramme des fréquences de l'échantillon à la courbe représentative de f.

   Que remarque-t-on ?

3.1 Comparaison fonction de répartition empirique – théorique

À partir d'un échantillon, on peut aussi approcher graphiquement la fonction de répartition de la loi considérée. Pour ce faire, on peut avoir recours à histogramme *cumulatif* des fréquences :

```
hist(A,n,density=True,cumulative=True)
```

Exercice 7.

1. Simuler 10 000 réalisations de la loi $\mathscr{E}(1)$.

2. Tracer sur une même graphique l'histogramme des fréquences cumulées et la fonction de répartition de la loi $\mathscr{E}(1)$.

4 Exercices

Exercice 8. Soient $a \in \mathbb{R}_+^\star$ et $(X_1, \dots, X_n)$ une famille de variables aléatoires indépendantes, identiquement distribuées suivant la loi $\mathscr{U}([0,a])$. On pose :

$$U = \min(X_1, \dots, X_n) \quad \text{et} \quad V = \max(X_1, \dots, X_n).$$

1. Déterminer les lois de U et V.

2. On considère la fonction suivante :

```
def simulation(a,n):
        nb_sim=10000
        A=a*rd.rand(nb_sim,n)
        return np.max(A,axis=1)
```

   Que renvoie la fonction `simulation` ?

3. Prenons $n = 10$ et $a = 1$. Illustrer graphiquement la qualité de cette simulation.

Exercice 9. Soit $(X_k)_{k \geqslant 1}$ une famille de variables aléatoires indépendantes et de même loi. On note

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k, \quad \text{et} \quad \overline{X}_n^\star = \frac{\overline{X}_n - \mathbb{E}\left(\overline{X}_n\right)}{\sqrt{\mathbb{V}\left(\overline{X}_n\right)}}.$$

1. On suppose dans cette question que les variables X_k suivent toutes la loi $\mathscr{U}([\![1,6]\!])$.
   a. Écrire une fonction `ncd(n,N)` donnant `N` réalisations de la variable $\overline{X}_n^\star$ dans ce cas.
   b. Comparer l'histogramme des fréquences et la courbe représentative de la densité de la loi $\mathscr{N}(0,1)$, selon les valeurs de n. Pour quelles valeurs de n cette simulation semble pertinente?

2. Mêmes questions en supposant que les variables X_k suivent toutes la loi $\mathscr{E}(1)$.

Exercice 10.

1. Soit X une variable aléatoire suivant la loi binomiale de paramètres n et p, et $X^\star$ la variable aléatoire centrée réduite associée à X. Créer une fonction `simuBicr` d'arguments n, p et x, simulant $\mathbb{P}(X^\star \leqslant x)$.

2. Créer une fonction `compBiNor` d'arguments n et p, donnant en sortie une matrice avec 2 lignes et 30 colonnes : la première ligne donne les valeurs $P(X^* \leqslant x)$ et la seconde les valeurs $P(Y \leqslant x)$ pour $x \in [-3, 3[$ avec un pas de 0,2.
   On importera la librairie `scipy.special` :

```
import scipy.special as sp
```

   et on utilisera la fonction la commande `sp.ndtr(x)`, qui donne une approximation de $\mathbb{P}(Y \leqslant x)$.

3. Tester la fonction `compBiNor` pour :

$$(n,p) \in \{(10,0{,}1), (10,0{,}4), (30,0{,}1), (30,0{,}4), (50,0{,}1), (50,0{,}4), (100,0{,}4), (100,0{,}1)\}.$$

   Que conclure ?

Exercice 11. Soient $k \in \mathbb{N}^\star$ et $\lambda > 0$.

1. Déterminer $r \in \mathbb{R}$ pour que la fonction

$$f_\lambda \ : \ x \ \mapsto \begin{cases} 0 & \text{si } x \leqslant \lambda, \\ \frac{r}{x^{k+1}} & \text{sinon} \end{cases}$$

soit une densité de probabilité.

Si une v.a.r. X admet pour densité f_λ, on dit que X suit la loi de Pareto de paramètres λ et k.

2. Déterminer la fonction de répartition d'une variable suivant la loi de Pareto de paramètres λ et k.

3. En utilisant la méthode d'inversion, simuler une variable aléatoire suivant une loi de Pareto de paramètres λ et k.

4. Soient $X_1, \ldots, X_k$ des variables aléatoires indépendantes suivant toutes la loi uniforme sur $[0,1]$. On pose alors $Y = \dfrac{\lambda}{\max(X_1, \ldots, X_k)}$.

   a. Montrer que Y suit une loi de Pareto de paramètres λ et k.

   b. En déduire une autre méthode pour simuler la loi de Pareto.