Devoir Maison 3

Densité

À rendre pour le : **07 mars 2025**

Problème :

Toutes les variables aléatoires de cet exercice sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$.

Dans ce devoir, on admettra que : pour toute variable aléatoire X qui admet une fonction de répartition F strictement croissante sur son support, la variable $F^{-1}(U)$ suit la même loi que X.

Partie I Variable de référence

On considère la fonction f définie sur $\mathbb{R}$ par :

$$f(x) = \begin{cases} \dfrac{1}{x(\ln x)^2} & \text{si } x \geq e \\ 0 & \text{sinon .} \end{cases}$$

1. **a)** Montrer que f est une densité de probabilité.
b) Soit à partir de maintenant X une variable aléatoire de densité f. Déterminer un support de X ainsi que la fonction de répartition F de X.
c) Justifier rapidement que F est strictement croissante de $[e, +\infty[$ dans $[0, 1[$ et admet une réciproque $F^{-1} : [0, 1[\to [e, +\infty[$.

2. Simulation par approximation de F^{-1}.
a) Soit $y \in [0, 1[$. Admettons avoir écrit le programme ci-dessous. Expliquer ce que rend l'appel à la fonction `borne_sup`. Est-on certain que la boucle va bien s'arrêter ?

```
def F(x):
    return(1-1/np.log(x))

def borne_sup(y):
    b=10
    while F(b)<y:
        b=b*10
    return(b)
```

b) Déduire des questions précédentes une fonction Python qui prend en argument $y \in [0, 1[$ et qui rend une valeur approchée à 10^{-2} près de $F^{-1}(y)$ pour tout $y \in [0, 1[$. Afin de tester son programme, on pourra par exemple vérifier que $F^{-1}(0.95) \simeq 485\,165\,195.41$
c) Présenter pour finir une fonction Python permettant de simuler X à l'aide de l'approximation de F^{-1} faite ci-dessus. (On pourra se référer au TP sur la simulation de variables à densité.)

3. Simulation par valeur exacte
a) Déterminer l'expression exacte de la réciproque de $F : [e; +\infty[\to [0, 1[$.
b) En déduire une simulation de X par valeurs exactes de F^{-1}.

4. **a)** Montrer que pour tout $r \in \mathbb{N}^*$, il existe $\gamma > e$ tel que

$$x^r f(x) \geq \frac{1}{x} \quad \forall x \geq \gamma.$$

b) En déduire que la variable aléatoire X n'admet aucun moment.

Partie II Le deuxième plus petit

On considère, à partir de maintenant, une suite $(X_k)_{k \in \mathbb{N}^*}$ de variables aléatoires indépendantes, de même loi que X.

Pour tout entier $n \geq 3$ et pour tout $\omega \in \Omega$, on note $Y_n(\omega)$ le « deuxième plus petit » des nombres $X_1(\omega), \dots, X_n(\omega)$, c'est-à-dire le deuxième lorsqu'ils sont classés par ordre croissant (au sens large).
Par exemple, si $n = 5$ et $(X_1(\omega), \dots, X_5(\omega)) = (2.9,\ 4.78,\ 7.8,\ 5.1,\ 3.8)$, on a

$$Y_5(w) = 3.8$$

On admet que l'on définit ainsi une variable aléatoire Y_n pour tout $n \geq 3$.

5. **a)** Écrire une fonction d'entrée `n` qui simule la variable Y_n. *(On se servira bien entendu des fonctions déjà construites précédemment.)*

b) Pour tout $x \geq e$, soit U_x la variable aléatoire qui compte le nombre de variables X_i qui prennent une valeur inférieure ou égale à x. Quelle est la loi de U_x ?

c) Montrer qu'il existe $c \in \mathbb{N}$ à déterminer tel que, pour tout $x \geq e$,

$$(Y_n \leq x) \ \Leftrightarrow\ (U_x \geq c).$$

d) En déduire la fonction de répartition G_n de Y_n.

e) Montrer que Y_n est une variable à densité et montrer qu'une densité de Y_n est la fonction g_n donnée par :

$$g_n(x) = \begin{cases} \dfrac{n(n-1)}{x(\ln x)^n}\left(1 - \dfrac{1}{\ln x}\right) & \text{si } x \geq e \\ 0 & \text{sinon .} \end{cases}$$

f) En déduire que la variable aléatoire $Z_n = \ln(Y_n)$ admet une espérance et calculer sa valeur.

Problème :

1. **a)** La fonction f est clairement positive sur $\mathbb{R}$ et $\mathcal{C}^0$ sur $\mathbb{R}$ sauf éventuellement en e.

Étudions sa convergence sur $\mathbb{R}$:
On a
$$\int_{-\infty}^{+\infty} f(x)\ \mathrm{d}x = \int_{e}^{+\infty} \frac{1}{x(\ln x)^2}\ \mathrm{d}x$$
La fonction à intégrer étant continue sur $[e;+\infty[$, il suffit d'étudier le problème en $+\infty$. De plus, on reconnait en f une expression du type $\frac{u'}{u^2}$ (avec $u=\ln x$). Ainsi, une primitive de f est $-\frac{1}{u} = -\dfrac{1}{\ln x}$.
D'où, pour tout $t \geq e$:
$$\int_e^t \frac{1}{x(\ln x)^2}\ \mathrm{d}x = \left[-\frac{1}{\ln x}\right]_e^t = 1 - \frac{1}{\ln t} \xrightarrow[x\to+\infty]{} 1$$
Ainsi, on a bien la convergence de l'intégrale avec
$$\int_{-\infty}^{+\infty} f(x)\ \mathrm{d}x = 1$$
En conclusion, f est bien une densité de probabilités.

---

b) La fonction f nous dit en premier lieu que
$$\boxed{Supp(X) = [e,+\infty[.}$$
On sait alors que
$$F = 0 \text{ sur }]-\infty;e[.$$
La fonction f étant $\mathcal{C}^0$ sur au moins $\mathbb{R}-\{e\}$, on sait de plus que la fonction de répartition F de X est $\mathcal{C}^1$ sur au moins $\mathbb{R}-\{e\}$ et sur cet intervalle : $F' = f$. Ayant déjà calculé une primitive dans la question précédente et F étant continue en e et donc en e^+ (par propriété des v.a. à densité), on obtient l'existence d'une constante α telle que :
$$F(x) = \begin{cases} 0 & \text{si } x < e \\ \alpha - \dfrac{1}{\ln x} & \text{si } x \geq e. \end{cases}$$
Avec par exemple la propriété $\lim\limits_{+\infty} F = 1$, on trouve la propriété $\alpha = 1$ et donc, par prolongement par
$$\boxed{F(x) = \begin{cases} 0 & \text{si } x < e \\ 1 - \dfrac{1}{\ln x} & \text{si } x \geq e. \end{cases}}$$

---

c) Sur cet intervalle, $F' = f > 0$, ce qui justifie la stricte croissance. De plus, par valeur en e et propriété des fonctions de répartition,
$$F(e) = 0 \quad \lim_{+\infty} F = 1.$$
Comme de plus F est continue sur l'intervalle en question, par théorème de la bijection, F vérifie bien les propriérés demandées.

---

2. **a)** Il parcourt toutes les valeurs b de type 10^n et s'arrête dès que
$$\boxed{y \in [0, F(10^n)]}$$
On est certain que la boucle va s'arrêter car :
⋆ d'une part, $\lim_{+\infty} F = 1$. Ainsi, pour tout $y \in [0,1[$, il va exister M pour lequel $F(M) > y$.
⋆ d'autre part, ayant $\lim_{n\to+\infty} 10^n = +\infty$, il va exister $n \in \mathbb{N}$ tel que $10^n \geqslant M$
⋆ et pour finir, la fonction F est croissante, on aura bien un n tel que $F(10^n) \geqslant F(M) > y$, ce qui arrête la boucle.

b) La méthode de la Dichomotie pour une fonction φ croissante qui s'annule dans un intervalle $[a,b]$, permet de déterminer une valeur approchée de x_0 telle que $\varphi(x_0)=0$.Si on cherche plutôt, à y fixé, une valeur de $F^{-1}(y)$, on veut en réalité x_0 tel que $F(x_0)=y$. Ainsi, en posant $\varphi(x)=F(x)-y$ et en cherchant une racine de φ par Dichotomie, on obtient alors une valeur approchée de $F^{-1}(y)$.
On en déduit le programme suivant :

```
def inverse(y):
    '''Pour y dans [0,1[, recherche F^{-1}(y) dans
       l'intervalle [e,borne_sup].'''
    a=np.exp(1)
    b=borne_sup(y)
    while b-a>0.01:
        c=(a+b)/2
        if (F(a)-y)*(F(c)-y)<0: # signes contraires
            , le zéro est dans [a,c[
            b=c
        else: # le zéro est dans [c,b]
            a=c
    return((a+b)/2)
```

c) On peut proposer le programme suivant :

```
import random as rd
def simul_X():
    U=rd.random()
    return(inverse(U))
```

3. a) Soit $y\in[0,1[$. On cherche l'expression de $F^{-1}(y)$ dont on connait l'existance. On pose alors $x=F^{-1}(y)$ et on a les égalités suivantes :

$$\begin{aligned} x=F^{-1}(y) &\Leftrightarrow F(x)=y \\ &\Leftrightarrow 1-\frac{1}{\ln x}=y \\ &\Leftrightarrow \underbrace{1-y}_{\neq 0}=\frac{1}{\ln x} \\ &\Leftrightarrow \ln x=\frac{1}{1-y} \\ &\Leftrightarrow x=e^{\frac{1}{1-y}} \end{aligned}$$

b) Voici le programme très simple suivant !

```
def simul_exacte_X():
    U=rd.random()
    resu=np.exp(1/(1-U))
    return(resu)
```

4. a) Par croissances comparées, on a, pour tout $r\in\mathbb{N}^*$,

$$x\times x^r f(x)=\frac{x^r}{(\ln x)^2}\xrightarrow[x\to+\infty]{}+\infty.$$

Il existe donc $\gamma>e$ tel que

$$x\times x^r f(x)\geq 1\quad \forall x\geq\gamma.$$

La divison par x donne le résultat souhaité :

$$\boxed{x^r f(x)\geq\frac{1}{x}\quad\forall x\geq\gamma}$$

b) D'après le théorème de transfert, X admet un moment d'ordre r si et seulement si l'intégrale

$$\int_{-\infty}^{+\infty} |x^r| f(x)\,\mathrm{d}x = \int_{e}^{+\infty} x^r f(x)\,\mathrm{d}x$$

converge absolument.
Or, on a d'après la question précédente, pour tout $x \geq \gamma > 0$:

$$x^r f(x) \geq \frac{1}{x}$$

Comme l'intégrale $\displaystyle\int_{\gamma}^{+\infty} \frac{1}{x}\,\mathrm{d}x$ diverge d'après le cours et qu'on a

$$\forall x \geq \gamma : x^r f(x) \geq \frac{1}{x} \geq 0,$$

par théorème de comparaison des intégrales de fonctions positives, on obtient que $\displaystyle\int_{\gamma}^{+\infty} x^r f(x)\,\mathrm{d}x$ diverge et donc $\displaystyle\int_{-\infty}^{+\infty} |x^r| f(x)\,\mathrm{d}x$ diverge également.
Donc, par théorème de transfert, X^r n'a pas d'espérance et ceci pour aucun $r \in \mathbb{N}^*$, donc

$$\boxed{\text{donc } X \text{ n'admet aucun moment.}}$$

---

5. a)

```
def simul_Y(n):
    L=[]
    for _ in range(n):
        L+=[simul_X()]
    # recherche de Y_n
    L.remove(min(L))
    return(min(L))
```

---

b) Les variables X_i sont toutes indépendantes. Ainsi, U_x compte le nombre de succès "être inférieur ou égal à x" dans une série de n répétitions indépendantes d'une même expérience : "résultat de la variable X." Ainsi,

$$\boxed{U_x \text{ suit la loi } \mathcal{B}(n,p) \text{ avec } p = P(X \leq x) = F(x).}$$

---

c) L'événement $(Y_n \leq x)$ signifie "la deuxième valeur est inférieure ou égale à x".
Toutes les autres valeurs importent peu. Étant donné que toutes les valeurs sont ordonnées, ceci équivaut à "Il y a **au moins** deux valeurs inférieures ou égale à x" (la plus basse et Y_n). D'où

$$\boxed{(Y_n \leq x) \ \Leftrightarrow \ (U_x \geq 2)}$$

---

d) Au vu du support de X, il est clair que Y_n est à valeurs dans $[e, +\infty[$, donc

$$G_n(x) = 0 \text{ si } x < e.$$

Pour tout $x \geq e$, on a

$$\begin{aligned} G_n(x) &= P(Y_n \leq x) \\ &= P(U_x \geq 2) \\ &= 1 - P(U_x \leq 1) \\ &= 1 - P(U_x = 0) - P(U_n = 1) \\ &= 1 - \left(\frac{1}{\ln x}\right)^n - n\left(1 - \frac{1}{\ln x}\right)\left(\frac{1}{\ln x}\right)^{n-1} \\ &= 1 - \frac{n}{(\ln x)^{n-1}} + \frac{n-1}{(\ln x)^n}. \end{aligned}$$

D'où pour conclure :

$$\boxed{G_n(x) = \begin{cases} 0 & \text{si } x < e \\ 1 - \dfrac{n}{(\ln x)^{n-1}} + \dfrac{n-1}{(\ln x)^n} & \text{si } x \geq e \end{cases}}$$

e) La fonction G_n est $\mathcal{C}^0$ sur $]-\infty; e[$ et $[e; +\infty[$. On constate de plus

$$G(e) = 1 - n + (n-1) = 0 = \lim_{e^-} G_n$$

La fonction G_n est donc continue sur $\mathbb{R}$. De plus, G_n est $\mathcal{C}^1$ sur au moins $\mathbb{R} - \{e\}$, donc Y_n est une variable à densité. On note g_n sa densité.
On sait alors que sur au moins $\mathbb{R} - \{e\}$, on a $G'_n = g_n$. De plus, on peut choisir une valeur arbitraire en $x = e$ (par exemple, en prolongeant g_n par continuité en e^+, ce qui est possible ici.) Ainsi,

$$\boxed{g_n(x) = \begin{cases} 0 & \forall x < e \\ \dfrac{n(n-1)}{x(\ln x)^n}\left(1 - \dfrac{1}{\ln x}\right) & \forall x \geq e, \end{cases}}$$

---

f) D'après la densité, on a $Supp(X) \subset]e, +\infty[$ et la fonction ln est continue sur cet intervalle. D'après le théorème de transfert, la variable aléatoire Z_n admet une espérance si, et seulement si, l'intégrale $\displaystyle\int_e^{+\infty} |\ln x| g_n(x)\ \mathrm{d}x = \int_e^{+\infty} (\ln x) g_n(x)\ \mathrm{d}x$ converge absolument, ce qui est le cas car :

$$\int_e^{+\infty} |\ln x| g_n(x)\ \mathrm{d}x = \int_e^{+\infty} \frac{n}{n-2}\, g_{n-1}(x)\ \mathrm{d}x = \frac{n}{n-2}\int_e^{+\infty} g_{n-1}.$$

Or, g_{n-1} est une densité. Donc l'intégrale est convergente. Ainsi, $\mathbb{E}[Z_n]$ existe et on peut calculer son espérance. Le calcul est d'ailleurs précédemment effectué car :

$$E(Z_n) = \int_e^{+\infty} \ln x\, g_n(x)\ \mathrm{d}x = \frac{n}{n-2}\int_e^{+\infty} g_{n-1} = \frac{n}{n-2} \times 1$$

Donc :

$$\boxed{E(Z_n) = \frac{n}{n-2}.}$$