TD

Thème : Lois à densité et simulation

Exercice 1. Soit f la fonction définie sur $\mathbb{R}$ par $f(x) = \dfrac{1}{2(1+|x|)^2}$.

1. Démontrer que f est la densité de probabilité d'une variable aléatoire X. On note F sa fonction de répartition.

2. On considère la variable aléatoire $\mathrm{Y} = \ln(1+|\mathrm{X}|)$ et on note G sa fonction de répartition. Exprimer G en fonction de F.

3. Reconnaitre la loi de Y.

Simulations

Comme pour les variables discrètes, la bibliothèque `numpy.random` importée via `rd` permet de simuler les lois à densité usuelles.

- La commande `rd.random()` simule une loi uniforme continue sur $[0;1[$.
- `rd.exponential(a)` renvoie la simulation d'une variable aléatoire de loi exponentielle de paramètre $1/$`a`.
- `rd.normal(mu,sigma)` renvoie la simulation d'une variable de loi normale d'espérance `mu` et d'écart-type `sigma`.
- `rd.gamma(a)` renvoie la simulation d'une variable aléatoire de loi gamma de paramètre `a`.

Exercice 2. ✧ ✎ En utilisant uniquement la commande `rd.random()`, expliquer comment simuler une variable aléatoire suivant une loi uniforme continue sur $[a, b]$.

Exercice 3. ✦✦ *D'après ESSEC 2014 E*

Dans tout le sujet, $\mathrm{I} =]a, b[$ est un intervalle ouvert non vide de $\mathbb{R}$, où a et b sont réels ou infinis. On dit qu'une densité de probabilité f vérifie l'hypothèse CSP(I) lorsque f est :

- continue sur I,
- strictement positive sur I,
- nulle en dehors de I.

On écrira alors simplement : f est CSP(I).

La plupart des langages informatiques possèdent un générateur de nombres aléatoires. En Python par exemple, on dispose de l'instruction **rd.random()** de la bibliothèque **numpy.random** (alias **rd**). Ces générateurs produisent une suite de variables aléatoires indépendantes de loi uniforme sur $]0, 1[$.
On propose dans la suite deux méthodes permettant de simuler des lois à densité quelconques en utilisant ces générateurs aléatoires.
Jusqu'à la fin du problème : on note Z une variable aléatoire continue à valeurs dans I, de fonction de répartition G et admettant une densité g qui est CSP(I).

A- Simulation par la méthode d'inversion

1. **a)** On note H la restriction de G à I. Montrer que H réalise une bijection de I sur $]0,1[$. On note H^{-1} la bijection réciproque. Dresser le tableau de variations de H^{-1}.

Soit U une variable aléatoire suivant la loi uniforme sur $]0,1\,[$. On pose $\mathrm{X}=\mathrm{H}^{-1}(\mathrm{U})$.

b) ✎ Justifier que X suit la même loi que Z.

2. *Simulation de lois exponentielles.*
On suppose dans cette question que Z suit la loi exponentielle de paramètre $\lambda > 0$.

a) Expliciter l'intervalle I et les fonctions g, G et H^{-1}.

b) Écrire une fonction Python d'argument λ qui simule la loi exponentielle de paramètre λ.

3. *Simulation de la loi de Laplace.*
On cherche dans cette question à simuler une variable aléatoire de densité g donnée par :

$$\forall x \in \mathbb{R}, \quad g(x)=\frac{1}{2}\mathrm{e}^{-|x|} \quad \text{(densité de Laplace).}$$

Soit Y une variable aléatoire suivant la loi exponentielle de paramètre 1.
Soit V une variable aléatoire indépendante de Y suivant la loi uniforme sur $\{-1,1\}$, ce qui signifie $\mathbf{P}([\mathrm{V}=-1])=\mathbf{P}([\mathrm{V}=1])=\frac{1}{2}$. On pose $\mathrm{X}=\mathrm{VY}$.

a) Vérifier que g est une densité de probabilité qui est CSP($\mathbb{R}$).

b) Établir :

- pour tout $x \geqslant 0$, $\mathbf{P}([\mathrm{X}>x])=\frac{1}{2}\mathbf{P}([\mathrm{Y}>x])$;
- pour tout $x \leqslant 0$, $\mathbf{P}([\mathrm{X}\leqslant x])=\frac{1}{2}\mathbf{P}([\mathrm{Y}\geqslant -x])$.

c) En déduire une expression de la fonction de répartition de X.

d) Conclure que X est une variable aléatoire continue admettant g comme densité.

e) Écrire une fonction Python qui simule la loi de Laplace.

B- Préliminaires

On considère dans cette partie :

- X une variable aléatoire réelle continue à valeurs dans I, de fonction de répartition F et admettant une densité de probabilité f qui est CSP(I).
- U une variable aléatoire qui suit la loi uniforme sur $]0,1[$ et qui est indépendante de X.
- h une fonction continue sur I à valeurs dans $[0,1]$.

On se propose d'établir la formule suivante :

$$\mathbf{P}\big([\mathrm{U}\leqslant h(\mathrm{X})]\big)=\mathbf{P}\big([\mathrm{U}<h(\mathrm{X})]\big)=\int_a^b f(t)h(t)\mathrm{d}t.$$

On définit sur I la fonction Ψ par : $\forall x \in \mathrm{I}$, $\Psi(x)=\mathbf{P}\big([\mathrm{X}\leqslant x]\cap[\mathrm{U}\leqslant h(\mathrm{X})]\big)$.

1. Pour tous réels x et y dans I tels que $x<y$, on pose $\mathrm{M}(x,y)=\max\limits_{t\in[x,y]} h(t)$ et $m(x,y)=\min\limits_{t\in[x,y]} h(t)$.

a) Soit x dans I. Justifier que pour tout y dans l'intervalle $]x,b\,[$, il existe α_y dans l'intervalle $[x,y]$ tel que $\mathrm{M}(x,y)=h\left(\alpha_y\right)$.

b) En déduire : $\lim\limits_{\substack{y\to x\\ y>x}} \mathrm{M}(x,y)=h(x)$.

c) Montrer de même que pour tout y dans I, $\lim\limits_{\substack{x\to y\\ x<y}} \mathrm{M}(x,y)=h(y)$.

On montrerait de manière analogue (on ne demande pas de le vérifier) : $\lim\limits_{\substack{y\to x\\ y>x}} m(x,y)=h(x)$ et $\lim\limits_{\substack{x\to y\\ x<y}} m(x,y)=h(y)$.

2. Soient x et y des réels de I tels que $x < y$.

a) Établir l'inclusion suivante entre événements :

$$\big[x < \mathrm{X} \leqslant y\big] \cap \big[\mathrm{U} \leqslant h(\mathrm{X})\big] \subset \big[x < \mathrm{X} \leqslant y\big] \cap \big[\mathrm{U} \leqslant \mathrm{M}(x,y)\big].$$

En déduire l'inégalité :

$$\Psi(y) - \Psi(x) \leqslant \big(\mathrm{F}(y) - \mathrm{F}(x)\big)\mathrm{M}(x,y).$$

b) Établir une minoration analogue pour $\Psi(y) - \Psi(x)$, puis l'encadrement

$$\frac{\mathrm{F}(y) - \mathrm{F}(x)}{y - x} m(x,y) \leqslant \frac{\Psi(y) - \Psi(x)}{y - x} \leqslant \frac{\mathrm{F}(y) - \mathrm{F}(x)}{y - x} \mathrm{M}(x,y)$$

c) Montrer que Ψ est dérivable sur I, et exprimer sa dérivée en fonction de f et h.

3. a) En déduire que, pour tout x et y dans I :

$$\Psi(y) - \Psi(x) = \int_x^y f(t)h(t)\mathrm{d}t.$$

b) Établir : pour tout x dans I, $\Psi(x) \leqslant \mathrm{F}(x)$, puis montrer : $\lim\limits_{x \to a} \Psi(x) = 0$. En déduire :

$$\forall x \in \mathrm{I}, \qquad \Psi(x) = \int_a^x f(t)h(t)\mathrm{d}t.$$

c) Établir : pour tout x dans I, $\mathbf{P}\big([\mathrm{X} > x] \cap [\mathrm{U} \leqslant h(\mathrm{X})]\big) = \mathbf{P}\big([\mathrm{U} \leqslant h(\mathrm{X})]\big) - \Psi(x)$.
En déduire : $\lim\limits_{x \to b} \Psi(x) = \mathbf{P}\big([\mathrm{U} \leqslant h(\mathrm{X})]\big)$, puis

$$\mathbf{P}\big([\mathrm{U} \leqslant h(\mathrm{X})]\big) = \int_a^b f(t)h(t)\mathrm{d}t.$$

4. Montrer que $\mathbf{P}\big([\mathrm{U} < h(\mathrm{X})]\big) = 1 - \mathbf{P}\big([1 - \mathrm{U} \leqslant 1 - h(\mathrm{X})]\big)$, et en déduire

$$\mathbf{P}\big([\mathrm{U} < h(\mathrm{X})]\big) = \int_a^b f(t)h(t)\mathrm{d}t.$$

C- Application : Simulation par la méthode du rejet

Dans la méthode dite du rejet, pour simuler la loi de Z de densité g, on commence par déterminer une loi de probabilité que l'on sait simuler, de densité f qui est CSP(I), et qui vérifie : il existe une constante $c > 0$ telle que :

$$\forall x \in \mathrm{I}, \qquad g(x) \leqslant cf(x).$$

1. Montrer qu'il existe une fonction h continue sur I et à valeurs dans $[0, 1]$ telle que, pour tout $x \in \mathrm{I}, g(x) = cf(x)h(x)$ On considère alors :

- une suite de variables aléatoires $(\mathrm{U}_k)_{k \in \mathbb{N}^*}$ qui suivent la loi uniforme sur $]0, 1[$.
- une suite de variables aléatoires $(\mathrm{X}_k)_{k \in \mathbb{N}^*}$ à valeurs dans $]a, b[$, ayant toutes la même loi de densité de probabilité f et de fonction de répartition F.

On suppose de plus que pour tout entier $n \geqslant 1$, les variables $\mathrm{X}_1, \ldots, \mathrm{X}_n, \mathrm{U}_1, \ldots, \mathrm{U}_n$ sont mutuellement indépendantes.
On définit N la variable aléatoire prenant comme valeur le premier indice k vérifiant $\mathrm{U}_k \leqslant h\left(\mathrm{X}_k\right)$.

2. a) En utilisant la partie précédente, prouver l'égalité, pour tout $k \in \mathbb{N}^*$, $\mathbf{P}\left(\left[\mathrm{U}_k \leqslant h\left(\mathrm{X}_k\right)\right]\right) = \frac{1}{c}$.

b) En déduire que N suit une loi géométrique dont on précisera le paramètre, l'espérance et la variance.

On définit la variable aléatoire X comme étant la valeur de X_N, c'est-à-dire la valeur de X_k pour le premier indice k vérifiant $\mathrm{U}_k \leqslant h\left(\mathrm{X}_k\right)$.

3. Soit $x \in \mathrm{I}$.

a) Soit $n \in \mathbb{N}^*$. Exprimer l'événement $[\mathrm{X} \leqslant x] \cap [\mathrm{N} = n]$ à partir des événements $[\mathrm{X}_n \leqslant x] \cap [\mathrm{U}_n \leqslant h(\mathrm{X}_n)]$ et $[\mathrm{U}_k > h(\mathrm{X}_k)]$ pour $k \in [\![1, n-1]\!]$.

b) En utilisant la question B.3.(b), montrer que pour tout $n \in \mathbb{N}^*$:

$$\mathbf{P}\left([\mathrm{X}_n \leqslant x] \cap [\mathrm{U}_n \leqslant h(\mathrm{X}_n)]\right) = \frac{1}{c}\mathrm{G}(x).$$

c) En déduire $\mathbf{P}([\mathrm{X} \leqslant x] \cap [\mathrm{N} = n])$ en fonction de c et de $\mathrm{G}(x)$.

d) Montrer finalement : $\mathbf{P}([\mathrm{X} \leqslant x]) = \mathrm{G}(x)$.

4. Conclure.

5. *Simulation de la loi normale.*

Dans cette question, Z suit la loi normale centrée et réduite, donc $\mathrm{I} = \mathbb{R}$.
Soit f la densité de Laplace (question 3), définie par $\forall x \in \mathbb{R},\ f(x) = \frac{1}{2}\mathrm{e}^{-|x|}$.

a) Donner une densité g de Z qui est CSP($\mathbb{R}$).

b) Étudier les variations sur $[0, +\infty[$ de la fonction $a : x \mapsto \mathrm{e}^{x - \frac{x^2}{2}}$.

c) Expliciter une constante $c > 0$ telle que, pour tout $x \geqslant 0$: $g(x) \leqslant \frac{c}{2}\mathrm{e}^{-x}$.

d) En déduire que pour tout x réel, $g(x) \leqslant cf(x)$.

e) Expliquer alors comment mettre en place la méthode du rejet pour simuler la loi normale centrée et réduite. On explicitera en particulier la fonction h introduite à la question 4.

f) Compléter la fonction Python suivante pour qu'elle simule la loi normale centrée réduite :

Editeur

```
def simulnormaleCR():
    ...
    while ... :
       x=laplace()
       u=rd.random()
           ...
    return normale=
```

g) Comment simuler une loi normale $\mathcal{N}(\mu, \sigma^2)$?