12 T.D. Variables aléatoires à densité.

- X admet une densité si, et seulement si, sa fonction de répartition F_X est continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sauf éventuellement en un nombre fini de points. « Dans ce contexte, donner la loi d'une variable aléatoire X, c'est justifier que X admet une densité et en donner une ».
- Sur des exemples simples, recherche de la loi de $u(X)$, X ayant une densité donnée.
- Espérance, théorème de transfert et inégalité de Markov. Variance, écart-type, moments.
- Lois usuelles : Loi uniforme, loi exponentielle et loi normale.
- Somme de variables aléatoires à densité indépendantes, la formule du produit de convolution devant être rappelée en cas de besoin.

Exercice 1 : ✶

Soit un entier naturel $n \geq 2$. Pour chaque fonction f ci-dessous, déterminer la valeur de la constante $\lambda > 0$ pour que f soit une densité de probabilité.
Déterminer la fonction de répartition associée et donner son graphe.

$$a) f(x) = \lambda x \mathbf{1}_{]0,1[}(x); \qquad b) f(x) = \lambda x(1-x)\mathbf{1}_{[0,1]}(x); \qquad c) f(x) = \lambda\,(2-|x|)\,\mathbf{1}_{[-2,2]}(x);$$

$$d) f(x) = \frac{\lambda}{x^n}\mathbf{1}_{[1,+\infty[}(x); \qquad e) f(x) = \frac{\lambda}{1+n^2x^2}; \qquad f) f(x) = \frac{\lambda}{\sqrt{1-x^2}}\mathbf{1}_{]0,1[}(x)$$

Exercice 2 : ✶✶

k étant un réel donné, on considère l'application f définie pour tout x réel par $f(x) = k\cos(x)\mathbf{1}_{[0,\pi/2]}(x)$.

① Déterminer k pour que f soit une densité de probabilité sur $\mathbb{R}$.

② Soit X, variable aléatoire de densité f. Déterminer sa fonction de répartition, l'espérance et la variance de X.

③ Soit $Y = \tan(X)$.

a. Montrer que $\sin(\arctan(y)) = \dfrac{y}{\sqrt{1+y^2}}$ pour tout $y \in \mathbb{R}_+$.
b. Déterminer la fonction de répartition de Y.
c. Calculer l'espérance de Y si elle existe.

Exercice 3 ✶✶ :

Soit f une fonction définie sur $\mathbb{R}$ par $f(x) = \dfrac{e^{-x}}{(1+e^{-x})^2}$.

① Montrer que f est une densité de variable aléatoire réelle.
Soit X, variable aléatoire de densité f. Déterminer sa fonction de répartition.

② Soit g définie par $g(x) = \dfrac{e^x-1}{e^x+1}$. Montrer que g est une bijection de $\mathbb{R}$ sur $]-1,1[$. Déterminer sa fonction réciproque g^{-1}.

③ On définit la variable aléatoire Y par $Y = \dfrac{e^X-1}{e^X+1}$. Déterminer la fonction de répartition de Y. Calculer, s'ils existent, $\mathbb{E}(X)$ et $\mathbb{V}(X)$.

Exercice 4 : ✶✶

Soit X une variable aléatoire réelle dont une densité de probabilité est f définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ xe^{-x^2/2} & \text{si } x \geq 0 \end{cases}$$

① Vérifier que f est bien une densité de probabilité.

② Quelle est la loi suivie par $Y = X^2$?

③ Calculer l'espérance et la variance de X. Généraliser ce résultat en calculant le moment d'ordre r quelconque.

Exercice 5 : ✶✶

Dans le cadre de l'implantation d'éoliennes à proximité du lac de Grand-Lieu, on cherche à simuler les vitesses quotidiennes moyennes de vents mesurées expérimentalement par MétéoFrance à une altitude de 10 mètres. On dispose pour cela de six années de données, de 2008 à 2013, soit 2190 mesures, conservées sous la forme d'un tableau de deux colonnes représentant respectivement la vitesse (en $m.s^{-1}$) et la direction du vent (0 pour le nord, 90 pour l'est,...). On trouvera ce tableau sur le site de la classe, dans le dossier « td12_Python » sous le nom « VentsMFr.csv ».

① Récupérer sous Python les vitesses de vents sous la forme d'un tableau `Tv`, exploitable par le module `numpy`.
Écrire une fonction Python permettant d'afficher la moyenne des vitesses de vent, leur écart-type, les vitesses de vents minimum et maximum rencontrées pendant ces huit années.
Tracer dans une première fenêtre une boîte à moustache permettant de synthétiser ces résultats et dans une deuxième fenêtre les fréquences cumulées des vitesses de vents avec des classes d'amplitude $0.1m.s^{-1}$.
Conserver pour la suite le tableau `Fc` des fréquences cumulées de vitesses de vents.

② Soit a et λ deux réels strictement positifs et f définie sur $\mathbb{R}$ par $f(x) = b\lambda x^{a-1}e^{-\lambda x^a}\mathbf{1}_{\mathbb{R}_+}(x)$.
Déterminer b pour que f soit une densité de probabilité d'une variable aléatoire X. On dira dans ce cas que X suit une loi de Weibull de paramètres a et λ. Reconnaissez cette loi dans le cas $a = 1$.

③ Déterminer la fonction de répartition F_X de X.

④ Considérons V variable aléatoire égale aux vitesses de vents moyens quotidiens sur le site qui nous intéresse et supposons que les vitesses de vent observées sont des réalisations indépendantes de cette variable aléatoire. On souhaite tester l'hypothèse : *les vitesses de vents autour du lac de Grand-Lieu suivent une loi de Weibull.*Si, c'est le cas, alors :

$$F_c(v_i) \cong \mathbb{P}(V \leq v_i) = F_V(v_i),\ \forall v_i \in \{vmin, ..., vmax\} \cong X(\Omega)$$

a. Montrer que l'hypothèse précédente conduit à la relation : $ln(v_i) \cong \frac{1}{a}ln(-ln(1 - F_c(v_i)) - \frac{ln(\lambda)}{a}$.

b. Proposez un changement de variable permettant à la fois de tester notre hypothèse et d'obtenir une valeur de a et λ par la méthode des moindres carrés.

c. Vérifier graphiquement vos résultats en traçant dans une même fenêtre les fréquences cumulées empiriques et la fonction de répartition F_V pour $V \hookrightarrow W(a, \lambda)$.

d. Proposez une modélisation informatique de la loi de Weibull obtenue et validez votre simulation en comparant sur un grand nombre de simulation les moyennes et les écart-types.

Exercice 6 : ✶

En supposant que X, Y et Z suivent la loi uniforme sur $]0, 1[$ et sont indépendantes, déterminer la loi de $X+Y+Z$.

Exercice 7 : ✶✶✶

On rappelle que si S et T sont deux variables aléatoires réelles à densité et indépendantes, alors $S + T$ est une variable à densité dont une densité est donnée, sous réserve de convergence de l'intégrale, par la formule :

$$f_{S+T}(x) = \int_{-\infty}^{+\infty} f_S(t)f_T(x - t)dt$$

Préliminaire : Soit $\lambda \in \mathbb{R}_+^*$ et $(p_k)_{k\in\mathbb{N}}$ une suite définie par $p_k = e^{-\lambda}\frac{\lambda^k}{k!}$.

① Montrer que les p_k définissent les coefficients d'une loi de probabilité.
Par la suite, si N est une variable aléatoire discrète telle que $N(\Omega) = \mathbb{N}$ et $\mathbb{P}(N = k) = p_k$, $\forall k \in \mathbb{N}$, alors on dira que N suit une loi de Poisson de paramètre λ et on écrira $X \hookrightarrow \mathcal{P}(\lambda)$.

② $\mathbb{E}(N)$ existe si la série $\sum k\mathbb{P}(N = k)$ converge absolument et dans ce cas elle vaut : $\mathbb{E}(N) = \displaystyle\sum_{k=0}^{+\infty} k\mathbb{P}(N = k)$.
Montrer que $\mathbb{E}(N)$ existe et donner sa valeur.

Première partie :

① Soit U une variable aléatoire suivant une loi uniforme sur l'intervalle $[0, 1[$ et soit λ un réel strictement positif. On note $X = -\dfrac{1}{\lambda}\ln(1 - U)$. Vérifier que X suit une loi exponentielle dont on précisera le paramètre.

② On note $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi que X. On note $S_0 = 0$ et pour $n \geq 1$, $S_n = X_1 + X_2 + \cdots + X_n$. Montrer que pour $n \geq 1$, S_n admet pour densité la fonction f_n donnée par :

$$\forall t \in \mathbb{R},\ f_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ \dfrac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} & \text{si } t \geq 0 \end{cases}$$

Deuxième partie : On suppose qu'à un arrêt de bus, les différences entre les temps de passage successifs d'un autobus sont indépendantes, et de même loi exponentielle de paramètre λ.
On définit un instant 0, puis on note $S_1, S_2, \cdots$ les temps de passages successifs des autobus. On note alors, pour $t > 0$, N_t la variable égale au nombre d'autobus qui sont passés entre l'instant 0 et l'instant t à l'arrêt de bus. Autrement dit on a : $\forall n \geq 0$, $(N_t = n) = (S_n \leq t < S_{n+1})$.

① Pour $n \geq 0$, exprimer l'événement $(N_t \geq n)$ à l'aide de la variable S_n. Justifier alors que pour tout $n \in \mathbb{N}$, on a :

$$\mathbb{P}(N_t = n) = \mathbb{P}(S_n \leq t) - \mathbb{P}(S_{n+1} \leq t)$$

② En déduire que N_t suit une loi de Poisson de paramètre λt.

③ On suppose désormais que les temps de passages entre deux autobus ont pour moyenne 10 minutes et qu'un étudiant en BCPST arrive à l'arrêt à l'instant $T = 100$ pour prendre le bus. Il se pose la question de savoir combien de temps, en moyenne, il devra attendre le prochain bus mais aussi combien de temps en moyenne s'écoule entre le prochain bus et celui qui l'a précédé. Pour y répondre, il écrit rapidement le programme Python suivant :

```
from math import log
from random import random

def autobus():
    a = 0 ; b = 0 ; N = 10000
    for k in range(N):
       s = 0
       while s < 100 :
          r = s
          s = s - 10*log(random())
       u = s-100 ; v = s-r
       a = a+u ; b = b+v
    print(a/N,b/N)
```

④ Expliquer ce que représentent les variables r, s, u et v dans le programme.

⑤ Le programme affiche finalement les valeurs suivantes : `10.062252 20.315494`
En quoi les valeurs affichées sont-elles paradoxales vis à vis de la situation ?