DS3 (vB)

Problème 1

Soit p un réel de $]0,1[$ et $q = 1-p$. Soit X_1 et X_2 deux variables indépendantes de même loi géométrique de paramètre p (d'espérance $\frac{1}{p}$).
On pose : $Y = X_1 - X_2$, $T = \max(X_1, X_2)$ et $Z = \min(X_1, X_2)$.
On rappelle que $T + Z = X_1 + X_2$ et $T - Z = |X_1 - X_2| = |Y|$.

1. ***a)*** Rappeler sans démonstration les valeurs respectives de $\mathbb{V}(X_1)$ et de $\mathbb{P}\big([X_1 \leqslant k]\big)$, pour tout k de $X_1(\Omega)$.

b) Calculer $\mathbb{E}(X_1 + X_2)$, $\mathbb{V}(X_1 + X_2)$, $\mathbb{E}(X_1 - X_2)$, $\mathbb{V}(X_1 - X_2)$.

c) Écrire une fonction **Python** qui prend en entrée le paramètre p et qui renvoie 1 si l'événement $[X_1 = X_2]$ est réalisé et 0 sinon. On simulera les variables aléatoires X_1 et X_2 dans le programme.

d) Établir la relation : $\mathbb{P}\big([X_1 = X_2]\big) = \dfrac{p}{1+q}$.

2. ***a)*** Montrer que Z suit la loi géométrique de paramètre $1 - q^2$. En déduire $\mathbb{E}(Z)$, $\mathbb{V}(Z)$ et $\mathbb{E}(T)$.

b) Soit k un entier de $\mathbb{N}^*$. Justifier l'égalité : $[Z = k] \cup [T = k] = [X_1 = k] \cup [X_2 = k]$.
En déduire la relation suivante : $\mathbb{P}\big([T = k]\big) = 2\ \mathbb{P}\big([X_1 = k]\big) - \mathbb{P}\big([Z = k]\big)$.

c) Établir la formule : $\mathbb{V}(T) = \dfrac{q\ \left(2q^2 + q + 2\right)}{\left(1 - q^2\right)^2}$.

3. ***a)*** Préciser $(T - Z)(\Omega)$.
Exprimer pour tout j de $\mathbb{N}^*$, l'événement $[Z = j] \cap [Z = T]$ en fonction des événements $[X_1 = j]$ et $[X_2 = j]$. En déduire pour tout j de $\mathbb{N}^*$, l'expression de $\mathbb{P}\big([Z = j] \cap [Z = T]\big)$.

b) Montrer que pour tout couple (j, l) de $(\mathbb{N}^*)^2$, on a : $\mathbb{P}\big([Z = j] \cap [T - Z = l]\big) = 2\ p^2\ q^{2j+l-2}$.

c) Montrer que pour tout k de $\mathbb{Z}$, $\mathbb{P}\big([X_1 - X_2 = k]\big) = \dfrac{pq^{|k|}}{1+q}$.
(on distinguera trois cas : $k = 0$, $k > 0$ et $k < 0$)

d) En déduire la loi de la variable aléatoire $|X_1 - X_2|$.

e) Établir à l'aide des questions précédentes que les variables Z et $T - Z$ sont indépendantes.

4. ***a)*** À l'aide du résultat de la question ***3.e***, calculer $\mathrm{Cov}(Z, T)$.
Les variables Z et T sont-elles indépendantes ?

b) Calculer en fonction de q, le coefficient de corrélation linéaire ρ de Z et T.

c) Déterminer la loi de probabilité du couple (Z, T).

d) Déterminer pour tout j de $\mathbb{N}^*$, la loi de probabilité conditionnelle de T sachant l'événement $[Z = j]$.

e) Soit j un élément de $\mathbb{N}^*$. On suppose qu'il existe une variable aléatoire D_j à valeur dans $\mathbb{N}^*$, dont la loi de probabilité est la loi conditionnelle de T sachant l'événement $[Z = j]$.
Calculer $\mathbb{E}(D_j)$.

Problème 2

L'objet du problème est l'étude de la concentration en un type de bactéries d'un bassin destiné à la baignade.

Une municipalité doit effectuer un prélèvement et l'analyser afin de décider d'autoriser ou non l'utilisation du bassin.
Dans une première partie, on étudiera des propriétés reliant la loi binomiale et la loi de Poisson. Dans une deuxième partie, on regardera la modélisation de la concentration en bactéries du bassin. Enfin, la troisième partie étudiera le principe d'un test destiné à prendre une décision d'utilisation.

Les trois parties peuvent être traitées indépendamment en admettant les résultats des parties précédentes.

Toutes les variables aléatoires sont définies sur un même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. Si elles existent, on note $\mathbb{E}(T)$ et $\mathbb{V}(T)$ l'espérance et la variance d'une variable aléatoire T.

I. Lien entre loi binomiale et loi de Poisson

1. Pour tout entier n strictement positif, on se donne un réel p_n strictement positif et n variables aléatoires $(X_k)_{k\in[\![1,n]\!]}$ indépendantes et suivant une loi de Bernoulli de paramètre p_n. On suppose que np_n a une limite finie strictement positive et on pose $\lim\limits_{n\to+\infty} np_n = \lambda$.

- ***a)*** Quelle est la loi de $S_n = \sum\limits_{k=1}^{n} X_k$?
- ***b)*** Soit k un entier naturel.
  - ***(i)*** Donner l'expression de $\mathbb{P}\big([S_n = k]\big)$ pour n supérieur ou égal à k.
  - ***(ii)*** Que peut-on dire de la limite de p_n quand n tend vers l'infini ?
    Étudier la limite de $(1-p_n)^n$ quand n tend vers l'infini.
  - ***(iii)*** Montrer : $\lim\limits_{n\to+\infty} \mathbb{P}\big([S_n = k]\big) = \mathrm{e}^{-\lambda}\, \dfrac{\lambda^k}{k!}$.
- ***c)*** On pose $N_n = \max(X_1, \dots, X_n)$.
  - ***(i)*** Quelles sont les valeurs que peut prendre la variable aléatoire N_n ?
  - ***(ii)*** Calculer $\lim\limits_{n\to+\infty} \mathbb{P}\big([N_n = 0]\big)$.
  - ***(iii)*** `(CUBES UNIQUEMENT)` En déduire la limite en loi de la suite $(N_n)_{n\in\mathbb{N}^*}$.

2. Soit λ un réel strictement positif et n un entier strictement positif tels que $0 < \lambda \leqslant n$. On considère U une variable aléatoire de loi de Bernoulli de paramètre $\dfrac{\lambda}{n}$. On a donc $\mathbb{P}\big([U=1]\big) = \dfrac{\lambda}{n}$, $\mathbb{P}\big([U=0]\big) = 1 - \dfrac{\lambda}{n}$ et pour tout i entier naturel supérieur ou égal à 2, $\mathbb{P}\big([U=i]\big) = 0$.

- ***a)*** Écrire une fonction **Python** nommée `simulU(lam, n)` qui simule la variable aléatoire U (`lam` représente ici le réel λ).
- ***b)*** Montrer que pour tout réel positif u : $1 - u \leqslant \mathrm{e}^{-u}$.
- ***c)*** Montrer :

$$\sum_{k=0}^{+\infty} \left| \mathbb{P}\big([U=k]\big) - \left(\frac{\lambda}{n}\right)^k \frac{1}{k!}\, \mathrm{e}^{-\frac{\lambda}{n}} \right| = \frac{2\lambda}{n} \left(1 - \mathrm{e}^{-\frac{\lambda}{n}}\right)$$

d) En déduire :

$$\sum_{k=0}^{+\infty} \left| \mathbb{P}\big([U = k] \big) - \left(\frac{\lambda}{n}\right)^k \frac{1}{k!} \, \mathrm{e}^{-\frac{\lambda}{n}} \right| \leqslant 2 \left(\frac{\lambda}{n}\right)^2$$

3. Soit λ un réel strictement positif et n un entier strictement positif tels que $0 < \lambda \leqslant n$. Soient Z, U et V trois variables aléatoires indépendantes à valeurs dans $\mathbb{N}$. On suppose que U suit une loi de Bernoulli de paramètre $\dfrac{\lambda}{n}$ et que V suit une loi de Poisson de paramètre $\dfrac{\lambda}{n}$. On observe en particulier que pour tout entier p strictement négatif $\mathbb{P}\big([Z = p] \big) = \mathbb{P}\big([U = p] \big) = \mathbb{P}\big([V = p] \big) = 0$.

a) Soit i un entier naturel fixé. En remarquant que pour tout réels a et b, on a : $|a - b| \leqslant |a| + |b|$, montrer que la série de terme général $\big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big|$, où k décrit $\mathbb{N}$, est convergente. On note A_i sa somme, $A_i = \sum\limits_{k=0}^{+\infty} \big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big|$.

Montrer que pour tout i on a : $A_i \leqslant 2 \left(\dfrac{\lambda}{n}\right)^2$.

b) Montrer que la série de terme général $A_i \; \mathbb{P}\big([Z = i] \big)$ est convergente.

c) Soit k un entier naturel fixé.
Montrer que la série de terme général $\big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big| \times \mathbb{P}\big([Z = i] \big)$, où i décrit $\mathbb{N}$, est convergente.

d) On pose $B_k = \sum\limits_{i=0}^{+\infty} \big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big| \times \mathbb{P}\big([Z = i] \big)$.

(i) Montrer que, pour $k \geqslant 2$, on a :

$$B_k = \left| 1 - \frac{\lambda}{n} - \mathrm{e}^{-\frac{\lambda}{n}} \right| \mathbb{P}\big([Z = k] \big) + \left| \frac{\lambda}{n} - \frac{\lambda}{n} \, \mathrm{e}^{-\frac{\lambda}{n}} \right| \mathbb{P}\big([Z = k - 1] \big) + \sum_{i=0}^{k-2} \mathbb{P}\big([V = k - i] \big) \, \mathbb{P}\big([Z = i] \big)$$

(ii) Justifier que pour $k \geqslant 2$, on a : $\sum\limits_{i=0}^{k-2} \mathbb{P}\big([V = k - i] \big) \; \mathbb{P}\big([Z = i] \big) \leqslant \mathbb{P}\big([V + Z = k] \big)$.

(iii) Montrer finalement que la série de terme général B_k, où k décrit $\mathbb{N}$, est convergente.
On admettra alors qu'on a : $\sum\limits_{k=0}^{+\infty} B_k = \sum\limits_{i=0}^{+\infty} A_i \, \mathbb{P}\big([Z = i] \big)$, c'est-à-dire :

$$\begin{aligned} & \sum_{k=0}^{+\infty} \left(\sum_{i=0}^{+\infty} \big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big| \times \mathbb{P}\big([Z = i] \big) \right) \\ = \; & \sum_{i=0}^{+\infty} \left(\sum_{k=0}^{+\infty} \big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big| \right) \times \mathbb{P}\big([Z = i] \big) \end{aligned}$$

4. On conserve dans cette question les notations et les hypothèses de la question 3 concernant les variables aléatoires Z, U et V.

a) Montrer que la série de terme général : $\big|\mathbb{P}\big([Z + U = k] \big) - \mathbb{P}\big([Z + V = k] \big)\big|$ est convergente.

b) Déduire de la question 3 :

$$\sum_{k=0}^{+\infty} \big|\mathbb{P}\big([Z + U = k] \big) - \mathbb{P}\big([Z + V = k] \big)\big| \leqslant \sum_{i=0}^{+\infty} \left(\sum_{k=0}^{+\infty} \big|\mathbb{P}\big([U = k - i] \big) - \mathbb{P}\big([V = k - i] \big)\big| \right) \mathbb{P}\big([Z = i] \big)$$

puis :

$$\sum_{k=0}^{+\infty} \big|\mathbb{P}\big([Z + U = k] \big) - \mathbb{P}\big([Z + V = k] \big)\big| \leqslant 2 \left(\frac{\lambda}{n}\right)^2$$

5. Soient $U_1, \dots, U_n, V_1, \dots, V_n$, $2n$ variables aléatoires indépendantes, telles que pour tout entier i tel que $1 \leqslant i \leqslant n$, U_i suit une loi de Bernoulli de paramètre $\dfrac{\lambda}{n}$ et V_i suit une loi de Poisson de paramètre $\dfrac{\lambda}{n}$.

a) Montrer que pour tout entier naturel k, on a :

$$\begin{aligned}
& \Big|\mathbb{P}\big([U_1 + \cdots + U_n = k]\big) - \mathbb{P}\big([V_1 + \cdots + V_n = k]\big)\Big| \\
\leqslant \quad & \Big|\mathbb{P}\big([U_1 + \cdots + U_n = k]\big) - \mathbb{P}\big([U_1 + \cdots + U_{n-1} + V_n = k]\big)\Big| \\
+ \quad & \Big|\mathbb{P}\big([U_1 + \cdots + U_{n-1} + V_n = k]\big) - \mathbb{P}\big([U_1 + \cdots + U_{n-2} + V_{n-1} + V_n = k]\big)\Big| \\
+ \quad & \cdots + \Big|\mathbb{P}\big([U_1 + V_2 + \cdots + V_n = k]\big) - \mathbb{P}\big([V_1 + \cdots + V_n = k]\big)\Big|
\end{aligned}$$

b) En déduire :

$$\sum_{k=0}^{+\infty} \Big|\mathbb{P}\big([U_1 + \cdots + U_n = k]\big) - \mathbb{P}\big([V_1 + \cdots + V_n = k]\big)\Big| \leqslant 2\frac{\lambda^2}{n}$$

c) Soit X une variable aléatoire de loi binomiale de paramètres n et $\dfrac{\lambda}{n}$. Conclure de ce qui précède :

$$\sum_{k=0}^{+\infty} \left|\mathbb{P}\big([X = k]\big) - \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\right| \leqslant 2\frac{\lambda^2}{n}$$

II. Modélisation de la concentration en bactéries.

Le bassin qu'on étudie est supposé de volume V, en m^3, et on effectue un prélèvement de volume ΔV, en m^3. Dans cette situation, la probabilité pour qu'une bactérie spécifique du bassin se trouve dans le prélèvement est égale à $\dfrac{\Delta V}{V}$.

Supposons que le bassin contienne n bactéries numérotées de 1 à n, $n \in \mathbb{N}^*$. On considère alors n variables aléatoires $X_1, \dots, X_n$ à valeurs 0 ou 1 telles que $X_i = 1$ si la bactérie i se trouve dans le prélèvement et 0 sinon. Les variables en question sont supposées indépendantes.

On pose $c = \dfrac{n}{V}$ qui représente la concentration en bactéries du bassin par m^3.

6. ***a)*** Quelle est la loi de X_i pour $i \in [\![1, n]\!]$?

b) Soit N le nombre de bactéries présentes dans le prélèvement.
Montrer que N suit une loi binomiale de paramètres n et $\dfrac{\Delta V}{V}$.

7. Soit U une variable aléatoire qui suit la loi de Poisson de paramètre $a > 0$.
On appelle F sa fonction de répartition.

a) Montrer que, pour tout $x \in \mathbb{R}$, $F(x) = \mathrm{e}^{-a} \displaystyle\sum_{k=0}^{\lfloor x \rfloor} \frac{a^k}{k!}$.

b) Recopier et compléter la fonction **Python** suivante pour qu'elle prenne en argument d'entrées les réels x et a et qu'elle renvoie la valeur de $F(x)$. On remarquera qu'on s'attache dans cette fonction à minimiser le nombre d'opérations.

```
def repartition_Poisson(x, a):
    S = 0
    terme = 1
    k = 0
    while __________
        S = S + terme
        k = _________
        terme = ___________________
    return __________________
```

8. Soit U une variable aléatoire de loi de Poisson de paramètre $c\Delta V$. Soit $K \in \mathbb{N}^*$ fixé.

a) Montrer : $\left|\mathbb{P}\big([N \leqslant K]\big) - \mathbb{P}\big([U \leqslant K]\big)\right| \leqslant \dfrac{2c\,(\Delta V)^2}{V}$.

b) On suppose que $V = 1000$, V volume en m^3, et que le prélèvement est de volume égal à 1 litre soit $\Delta V = 10^{-3}$.
Trouver un majorant de l'erreur commise en approximant $\mathbb{P}\big([N \leqslant K]\big)$ par $\mathbb{P}\big([U \leqslant K]\big)$.

c) On suppose que la concentration dans le bassin reste inférieure à 10^6 bactéries par mètre cube, et que le prélèvement réalisé est encore de 1 litre.
Quelle est la valeur minimale du volume V garantissant que l'erreur commise dans l'approximation précédente soit inférieure à 10^{-6} ?

On admettra dans la suite que le résultat de la question ***5.c)*** peut être amélioré de la façon suivante :

$$\sum_{k=0}^{+\infty} \left|\mathbb{P}\big([X = k]\big) - \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\right| \leqslant 2\lambda\,\frac{\min(2,\lambda)}{n}$$

d) Montrer que si U suit une loi de Poisson de paramètre $c\Delta V$:

$$\left|\mathbb{P}\big([N \leqslant K]\big) - \mathbb{P}\big([U \leqslant K]\big)\right| \leqslant 4\frac{\Delta V}{V}$$

e) On suppose toujours que le prélèvement réalisé est de un litre. À l'aide de l'inégalité précédente, trouver la valeur minimale du volume V garantissant que l'erreur commise dans l'approximation précédente soit inférieure a 10^{-6}.

III. Construction d'une procédure de test.

Non proposé ici.