À rendre pour le jeudi 22 janvier

L'objet du problème est l'étude du nombre de records d'une permutation via la suite de variables aléatoires $(R_n)_{n\in\mathbb{N}^*}$ introduite dans la partie II. La partie III utilise les notations et des résultats de la partie II.

Partie I Une distance entre lois de variables aléatoires

Toutes les variables aléatoires de cette partie sont définies sur le même espace probabilisé $(\Omega, \mathcal{T}, \mathbf{P})$.

1. Soient X, Y et Z trois variables aléatoires à valeurs dans $\mathbb{N}$.

a. Justifier la convergence de la série de terme général $|\mathbf{P}([X=n]) - \mathbf{P}([Y=n])|$.

On note $d(X,Y) = \dfrac{1}{2}\displaystyle\sum_{n=0}^{+\infty} (|\mathbf{P}([X=n]) - \mathbf{P}([Y=n])|)$.

b. Que dire des variables X et Y quand $d(X,Y)=0$?
Donner un exemple simple de deux variables X et Y distinctes telles que $d(X,Y)=0$.

c. Prouver l'inégalité : $d(X,Z) \leqslant d(X,Y) + d(Y,Z)$.

2. ***a.*** On pose : $A = \{k \in \mathbb{N}; \mathbf{P}([X=k]) \geqslant \mathbf{P}([Y=k])\}$. Vérifier l'égalité :

$$d(X,Y) = |\mathbf{P}([X\in A]) - \mathbf{P}([Y \in A])|.$$

b. Soit B une partie de $\mathbb{N}$. Prouver l'inégalité :

$$|\mathbf{P}([X\in B]) - \mathbf{P}([Y\in B])| \leqslant d(X,Y).$$

On pourra faire intervenir les parties $B\cap A$ et $B \cap A^c$.

3. Soit $p \in [0,1]$, X une variable aléatoire suivant une loi de Bernoulli de paramètre p et Y une variable aléatoire suivant une loi de Poisson de paramètre p.
Établir l'égalié $d(X,Y) = p(1-e^{-p})$ et en déduire la majoration $d(X,Y) \leqslant p^2$.

Dans la fin de cette partie, on considère un entier n au moins égal à 3, un entier $N > n$ et on note I la matrice identité de $\mathcal{M}_N(\mathbb{R})$ et R la matrice de $\mathcal{M}_N(\mathbb{R})$ dont tous les coefficients sont nuls sauf ceux de la surdiagonale qui valent 1, c'est-à-dire :

$$R = \begin{pmatrix} 0 & 1 & & (0) \\ & \ddots & \ddots & \\ & (0) & \ddots & 1 \\ & & & 0 \end{pmatrix}.$$

On considère une n-liste $(p_1, p_2, \ldots, p_n)$ de réels de $[0,1]$, n variables aléatoires $X_1, X_2, \ldots, X_n$ (resp. $Y_1, Y_2, \ldots, Y_n$), indépendantes suivant des lois de Bernoulli (resp. Poisson) de paramètres respectifs $p_1, p_2, \ldots, p_n$.
On note, pour tout entier $i \in [\![1,n]\!]$, $P_i = (1-p_i)I + p_iR$ et $Q_i = p_i(R-I)$.

4. ***a.*** Montrer que la première ligne du produit $P_1 P_2$ est constituée des $n+1$ réels $\mathbf{P}([X_1+X_2=0]), \mathbf{P}([X_1+X_2=1]), \ldots, \mathbf{P}([X_1+X_2=n])$ suivis de termes nuls.

b. On note $U_n = \sum_{k=1}^{n} X_k$. Montrer que la première ligne du produit $P_1 P_2 \ldots P_n$ est constituée des $n+1$ réels $\mathbf{P}([U_n=0]), \mathbf{P}([U_n=1]), \ldots, \mathbf{P}([U_n=n])$ suivis de termes nuls.

5. ***a.*** Montrer que : $\exp Q_i = \sum_{j=0}^{N-1} \frac{p_i^j e^{-p_i}}{j!} R^j$.

b. On note $V_n = \sum_{k=1}^{n} Y_k$. Montrer que la première ligne du produit $\exp Q_1 \times \exp Q_2 \times \ldots \times \exp Q_n$ est constituée par les N réels $\mathbf{P}([V_n=0]), \mathbf{P}([V_n=1]), \ldots, \mathbf{P}([V_n=N-1])$.

6. On note, pour toute matrice $A=(a_{i,j})_{1\leqslant i,j\leqslant N}$ de $\mathcal{M}_N(\mathbb{R})$, $\| A \| = \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} |a_{i,j}|$.

a. Prouver, pour tout couple (A,B) d'éléments de $\mathcal{M}_N(\mathbb{R})$ les inégalités :

$$\| A+B \| \leqslant \| A \| + \| B \| \text{ et } \| AB \| \leqslant \| A \| \| B \| .$$

b. Prouver, pour tout entier $i \in [\![1,n]\!]$, l'inégalité : $\| \exp Q_i \| \leqslant 1$.

c. En remarquant que :

$$\prod_{i=1}^{n} P_i - \prod_{i=1}^{n} \exp Q_i = (P_1 - \exp Q_1)\left(\prod_{i=2}^{n} P_i\right) + (\exp Q_1)\left(\prod_{i=2}^{n} P_i - \prod_{i=2}^{n} \exp Q_i\right),$$

prouver l'inégalité :

$$\left\| \prod_{i=1}^{n} P_i - \prod_{i=1}^{n} \exp Q_i \right\| \leqslant \sum_{i=1}^{n} \|P_i - \exp Q_i\|.$$

d. Prouver l'inégalité : $\|P_i - \exp Q_i\| \leqslant 2p_i^2$.

7. ***a.*** Déduire des questions précédentes l'inégalité : $d(U_n, V_n) \leqslant \sum_{i=1}^{n} p_i^2$.

b. Montrer que V_n suit une loi de Poisson de paramètre $p_1 + ... + p_n$.

c. Soit $\lambda > 0$. En appliquant les résultats précédents, retrouver la propriété d'approximation usuelle de la loi binomiale de paramètres n et $\frac{\lambda}{n}$ par une loi de Poisson, quand n tend vers $+\infty$.

Partie II Records d'une permutation

On rappelle qu'une permutation d'une ensemble non vide E est une bijection de E sur E. On note, pour tout entier naturel n non nul, $\mathcal{S}_n$ l'ensemble des permutations de l'ensemble $[\![1,n]\!]$ (des entiers k tels que $1 \leqslant k \leqslant n$) et $\sigma := (\sigma_1, \sigma_2, \ldots, \sigma_n)$ la permutation σ qui envoie l'entier 1 sur σ_1, 2 sur σ_2, ..., n sur σ_n.
On note $\mathcal{P}(\mathcal{S}_n)$ l'ensemble des parties de $\mathcal{S}_n$ et on munit l'espace probabilisable $(\mathcal{S}_n, \mathcal{P}(\mathcal{S}_n))$ de la probabilité uniforme notée $\mathbf{P}$. Ainsi, pour tout élément σ de $\mathcal{S}_n$, on a : $\mathbf{P}(\{\sigma\}) = \frac{1}{n!}$.
Soit $n \in \mathbb{N}^*$ et $k \in [\![1,n]\!]$. On dit q'une permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ présente un record au rang

k si, pour tout entier i vérifiant $1 \leqslant i \leqslant k$, on a $\sigma_i \leqslant \sigma_k$. Ainsi, en particulier, toute permutation présente un record au rang 1. On note $R_n(\sigma)$ le nombre de records que compte la permutation σ. On définit ainsi une variable aléatoire R_n sur $\mathcal{S}_n$. Bien sûr, la variable R_1 est certaine égale à 1. On note, pour tout entier n non nul : $H_n = \displaystyle\sum_{k=1}^{n} \frac{1}{k}$.

1. ***a.*** Écrire en PYTHON une fonction `record(L)` qui à une liste $[\sigma(1), ..., \sigma(n)]$ décrivant une permutation σ renvoie le nombre de records.

b. Écrire en PYTHON une fonction `permutation(n)` qui à un entier n renvoie une permutation aléatoire σ sous forme de liste $[\sigma(1), ..., \sigma(n)]$ et cela selon une prbabilité uniforme. On utilisera `import numpy.random as rd` qui permet d'utiliser la fonction `rd.randint(a,b)` qui permet de choisir un entier au hasard dans l'intervalle $[\![a, b[\![$ et de façon uniforme.

c. Mettre en évidence à l'aide de ceci que : $E(R_n) \underset{n\to+\infty}{\sim} \ln(n)$.

2. Montrer que la suite $(H_n - \ln(n))_{n\in\mathbb{N}^*}$ est monotone convergente. On note γ la limite de cette suite.

Dans toute la suite de cette partie **II** on considère un entier n au moins égal à 3.

3. Déterminer la loi de R_3, son espérance et sa variance.

4. Déterminer les probabilités $\mathbf{P}([R_n = 1])$ et $\mathbf{P}([R_n = n])$.

5. ***a.*** Soit p un entier de $[\![2, n]\!]$. Déterminer le nombre de permutations de $[\![1, n]\!]$ ayant exactement deux records lesquels sont atteints aux rangs 1 et p.

b. Prouver l'égalité : $\mathbf{P}([R_n = 2]) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n-1} \frac{1}{k}$.

c. Donner un équivalent de $\mathbf{P}([R_n = 2])$ quand n tend vers l'infini.

6. On introduit, pour tout entier i de $[\![1, n]\!]$, la variable aléatoire T_i, définie sur $(\mathcal{S}_n, \mathcal{P}(\mathcal{S}_n)), \mathbf{P}$, qui, à chaque élément σ de $\mathcal{S}_n$, associe la valeur 1 si σ présente un record au rang i et égale à 0 sinon.

a. Montrer que, pour tout entier i de $[\![1, n]\!]$, T_i suit la loi de Bernoulli de paramètre $\dfrac{1}{i}$.

b. Calculer l'espérance de la variable R_n et en donner un équivalent simple quand n tend vers l'infini.

7. ***a.*** Soit (i, j) un couple d'entiers vérifiant $2 \leqslant i < j \leqslant n$.
En calculant la probabilité $\mathbf{P}([T_i = 1] \cap [T_j = 1])$, justifier l'indépendance des variables T_i et T_j.

b. Calculer la variance de la variable R_n et en donner un équivalent simple quand n tend vers l'infini.

8. On **admet** l'indépendance mutuelle des variables $T_1, T_2, \ldots, T_n$.

a. Établir, pour tout entier k de $[\![2, n]\!]$, l'égalité :

$$\mathbf{P}([R_n = k]) = \sum_{2\leqslant i_2<i_3<\cdots<i_k\leqslant n} \frac{1}{i_2}\frac{1}{i_3}\cdots\frac{1}{i_k} \prod_{\substack{2 \leqslant j \leqslant n \\ j \notin \{i_2, \ldots, i_k\}}} \left(1 - \frac{1}{j}\right).$$

b. En déduire l'égalité : $\mathbf{P}([R_n = 3]) = \dfrac{1}{n}\displaystyle\sum_{1\leqslant i<j\leqslant n-1} \frac{1}{ij}$.

c. Prouver l'équivalence : $\mathbf{P}([R_n = 3]) \sim \dfrac{1}{2}\dfrac{\ln^2 n}{n}$.

Partie III Deux résultats asymptotiques

1. Un premier résultat

a. Soit $\varepsilon > 0$. Prouver, pour tout entier n assez grand, l'inclusion entre évènements :

$$\left[\left|\frac{R_n}{\ln n} - 1\right| \geqslant \varepsilon\right] \subset \left[\left|\frac{R_n}{\ln n} - \frac{H_n}{\ln n}\right| \geqslant \frac{\varepsilon}{2}\right].$$

b. Soit $\varepsilon > 0$.

i Établir l'égalité : $\lim\limits_{n\to+\infty} \mathbf{P}\left(\left[\left|\frac{R_n}{\ln n} - \frac{H_n}{\ln n}\right| \geqslant \varepsilon\right]\right) = 0.$

ii En déduire l'égalité : $\lim\limits_{n\to+\infty} \mathbf{P}\left(\left[\left|\frac{R_n}{\ln n} - 1\right| \geqslant \varepsilon\right]\right) = 0.$

2. Un second résultat

a. Déterminer la fonction génératrice d'une variable aléatoire suivant une loi $\mathcal{P}(\lambda)$

b. Soient $Y, X_1, X_2, ...$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$ telles que pour tout tout réel t de $[0,1]$, la suite $(G_{X_n}(t))_{n\in\mathbb{N}^*}$ converge vers $G_Y(t)$.

i. Pour $n \in \mathbb{N}$ et $s \in]0,1] \setminus \{0\}$, on pose $H_{n,0}(s) = G_{X_n}(s)$, puis pour $l \in \mathbb{N}^*$, on pose :

$$H_{n,l}(s) = \frac{H_{n,l-1}(s) - P(X_n = l-1)}{s}.$$

De même pour $s \in]0,1]$, on pose $H_0(s) = G_Y(s)$, puis pour $l \in \mathbb{N}^*$, on pose :

$$H_l(s) = \frac{H_{l-1}(s) - P(Y = l-1)}{s}.$$

Montrer que les applictions $H_{n,l}$ et H_l sont des fonctions prolongeables en série entière sur $[-1,1]$ et exprimer ces séries entières à l'aide des $P(X_n = k)$ et $P(Y = k)$.
On notera encore $H_{n,l}$ et H_l ces fonctions prolongées.

ii. En déduire que pour tout entier naturel k,

$$\lim_{n\to+\infty} P(X_n = k) = P(Y = k).$$

On pourra montrer par récurrence que sur l que $(H_{n,l})_{n\in\mathbb{N}}$ converge simplement vers H_l sur $[0,1]$. On dit que l'on a convergence en loi de $(X_n)_{n\in\mathbb{N}}$ vers Y.

c. Soit m un entier naturel au moins égal à 2 et $n = 2m$. On conserve les notations de la partie II et on pose $W_n = \sum_{k=m+1}^{2m} T_k$ (qui compte le nombre aléatoire de records que présente une permutation de $[\![1, 2m]\!]$ entre les rangs $m+1$ et $2m$).
Prouver, pour tout réel t de $[0,1]$, l'égalité : $G_{W_n}(t) = \prod_{i=m+1}^{2m}\left(1 + \frac{t-1}{i}\right)$.

d. En déduire que, lorsque l'entier m tend vers $+\infty$, la suite de terme général W_n converge en loi vers une variable aléatoire que l'on identifiera.