À rendre pour le jeudi 25 janvier

DM NORMAL

EXERCICE

On munit $\mathbb{R}^3$ de sa structure euclidienne canonique et on note (e_1, e_2, e_3) sa base canonique. On note $w = 2e_1 + 3e_2 + e_3$ et Π le plan vectoriel d'équation $2x + 3y + z = 0$. On note s la symétrie orthogonale par rapport au plan Π et S la matrice de s dans la base (e_1, e_2, e_3) de $\mathbb{R}^3$.

1. Donner une base (u, v) du plan Π et justifier, sans calcul, que (u, v, w) est une base de $\mathbb{R}^3$.

2. Écrire la matrice S' de s dans la base (u, v, w).

3. En déduire la matrice S.

PROBLÈME

Une question que se pose un joueur de cartes est de savoir combien de fois il est nécessaire de battre les cartes pour que le paquet soit convenablement mélangé. Ce problème décrit un procédé très élémentaire pour mélanger les cartes et propose de répondre alors à cette question.

Considérons un jeu de N cartes numérotées de C_1 à C_N et disposées en un paquet sur une table. Un joueur bat les cartes et repose le paquet sur la table. Le résultat du mélange est une permutation de ces N cartes.
Notations et Rappel : on note $\mathcal{S}_N$ l'ensemble des permutations possibles pour ce paquet de N cartes et on rappelle que $\text{card}(\mathcal{S}_N) = N!$
On se place dans un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ avec $\Omega = \mathcal{S}_N$, $\mathcal{A} = \mathcal{P}(\mathcal{S}_N)$ l'ensemble des parties de $\mathcal{S}_N$ et $\mathbf{P}$ l'équiprobabilité sur Ω. Pour toute variable aléatoire X on notera $E(X)$ et $V(X)$ l'espérance et la variance de X lorsqu'elles existent.

On considère qu'un paquet est *convenablement mélangé* lorsque toutes les permutations sont équiprobables, c'est-à-dire lorsque pour toute permutation σ de $\mathcal{S}_N$ la probabilité que le tas de cartes se trouve dans la configuration σ vaut $1/N!$

Vocabulaire et notations :
Une carte située au sommet de la pile est dite *en position n°1*, celle qui se trouve immédiatement en dessous est dite *en position n°2*, etc. Ainsi une carte située en position n°N désigne la carte située en bas de la pile. On prendra garde à bien distinguer la position d'une carte dans le paquet du numéro qu'elle porte.

Partons d'un tas de cartes rangées initialement dans l'ordre suivant :

pour tout i élément de $[\![1, N]\!]$, la carte C_i se trouve en position i.

Ainsi, à l'instant initial, la carte C_1 se trouve sur le dessus du paquet alors que C_N se trouve donc tout en dessous du paquet.

Pour k élément de $[\![1, N]\!]$, on appelle *insertion à la k-ième place* l'opération qui consiste à prendre la carte située au-dessus du paquet et à l'insérer entre la k-ième et la $(k+1)$-ième place. Une insertion à la première place ne change pas l'ordre des cartes. Une insertion à la N-ième place consiste à faire glisser la carte située au-dessus du paquet pour la mettre sous le paquet.

Le *battage par insertions* du jeu de cartes consiste à effectuer une suite d'insertions aléatoires, en choisissant, à chaque instant, au hasard uniformément dans $\{1, \cdots, N\}$ la place à laquelle l'insertion a lieu, indépendamment des insertions précédentes.
Les instants successifs d'insertions seront notées $1, 2, \ldots, n, \ldots$, l'instant initial est $n = 0$.

Notations. Nous notons :

- T_1 le premier instant où la carte située sur le dessus du paquet est glissée en dernière position, c'est-à-dire le premier instant où la carte C_N se trouve remontée de la position N à la position $N-1$,
- T_2 le premier instant où la carte C_N se trouve remontée en position $N-2$,
- et plus généralement, pour i dans $[\![1, N-1]\!]$, T_i le premier instant où la carte C_N atteint la position $N-i$.
- On posera également $\Delta_1 = T_1$ et : $\forall\ i \in [\![2, N-1]\!]$, $\Delta_i = T_i - T_{i-1}$.
- Enfin, on notera $T = T_{N-1} + 1$.

On admet que les conditions de l'expérience permettent de faire l'hypothèse que les variables aléatoires $(\Delta_i)_{i\in[\![1,N-1]\!]}$ sont indépendantes.

Description d'un exemple. Dans le tableau ci-dessous, nous décrivons les résultats d'une expérience faite sur un paquet de $N = 4$ cartes. La première ligne du tableau indique les instants n, la deuxième ligne indique les positions d'insertions, et dans les quatre dernières lignes figurent la configuration du paquet à l'instant n.

	instant n	0	1	2	3	4	5	6	7
	insertion en place k		3	2	4	1	3	4	2
Configuration du paquet	position 1	C_1	C_2	C_3	C_2	C_2	C_1	C_4	C_2
	position 2	C_2	C_3	C_2	C_1	C_1	C_4	C_2	C_4
	position 3	C_3	C_1	C_1	C_4	C_4	C_2	C_3	C_3
	position 4	C_4	C_4	C_4	C_3	C_3	C_3	C_1	C_1

Pour cette expérience, on a les résultats $T_1(\omega) = 3$, $T_2(\omega) = 5$ et $T_3(\omega) = 6$ et $T(\omega) = 7$.

Partie 1 - Description et premiers résultats

1. Justifier que : $\forall i \in [\![2, N-1]\!]\ \ T_i = \Delta_1 + \Delta_2 + \cdots + \Delta_i$.
Que représente l'intervalle de temps Δ_i ?

2. Loi de Δ_1.
Déterminer pour tout entier $n \in \mathbb{N}$, $\mathbf{P}(\Delta_1 > n)$ et reconnaître la loi de Δ_1.

3. Soit $i \in [\![2, N-1]\!]$. Loi de Δ_i.

a. Établir que pour tout entier $n \in \mathbb{N}$, on a : $\mathbf{P}(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n$. En déduire que Δ_i suit une loi géométrique de paramètre i/N.

b. En déduire $E(\Delta_i) = \dfrac{N}{i}$ et $V(\Delta_i) = N\dfrac{N-i}{i^2}$.

4. Loi de T_2. Soit $n \geq 2$.

a. Démontrer que : $\mathbf{P}(T_2 = n) = \displaystyle\sum_{k=1}^{n-1} \mathbf{P}(\Delta_2 = n-k)\mathbf{P}(\Delta_1 = k)$.

b. Justifier que : $\displaystyle\sum_{k=1}^{n-1} \left(\frac{1-1/N}{1-2/N}\right)^k = N\left(1-\frac{1}{N}\right)\left[\left(\frac{1-1/N}{1-2/N}\right)^{n-1} - 1\right]$.

c. En déduire que l'on a : $\mathbf{P}(T_2 = n) = \frac{2}{N}\left[\left(1-\frac{1}{N}\right)^{n-1} - \left(1-\frac{2}{N}\right)^{n-1}\right]$.

5. À l'instant T_2, la carte C_N est située en position $N-2$ et deux cartes se trouvent sous elle qui ont été insérées aux instants T_1 et T_2.
Que valent alors les probabilités, qu'à l'instant T_2 :
1) la carte insérée à l'instant T_1 soit en place $N-1$ et celle insérée à l'instant T_2 en place N ?
2) la carte insérée à l'instant T_2 soit en place $N-1$ et celle insérée à l'instant T_1 en place N ?

6. À l'instant T_3, la carte C_N est située en position $N-3$ et trois cartes, insérées aux instants T_1, T_2 et T_3, se trouvent sous elle. On note alors, pour $i \in \{1,2,3\}$, a_i la position de la carte ayant été insérée à l'instant T_i.

a. Combien y a-t-il de résultats possibles pour le triplet (a_1, a_2, a_3) ?

b. Quelques exemples. Donner les probabilités qu'à l'instant T_3 :
i) on obtienne $(a_1, a_2, a_3) = (N-2, N-1, N)$?
ii) on obtienne $(a_1, a_2, a_3) = (N-2, N, N-1)$?

7. Justifier la phrase suivante :

"À partir de l'instant T, toutes les configurations du jeu de cartes sont équiprobables."

On retiendra que si on arrête le battage des cartes par insertion exactement à l'instant T, on a un paquet convenablement mélangé. *Cependant le temps T étant aléatoire, il n'est pas possible d'arrêter de battre les cartes à cet instant précis, à moins de marquer la carte C_N bien sûr !*

Partie 2 - Estimation du nombre d'insertions pour bien mélanger les cartes

Notations : On introduit les suites $(H_n)_{n\geq 1}$ et $(u_n)_{n\in\mathbb{N}^*}$ définies par :

$$\forall n \geq 1 \quad H_n = \sum_{k=1}^{n} \frac{1}{k} \text{ et } u_n = H_n - \ln(n).$$

On admet ces résulats que l'on pourra utiliser par la suite :
i) la décroissance de la suite (u_n),
ii) l'encadrement : $\forall n \in \mathbb{N}^*,\ \ln(n+1) \leq H_n \leq \ln(n) + 1$.
iii) que la suite (u_n) est convergente et que sa limite, notée γ appartient à $[0,1]$.
(tous ces résultats se montrent par comparaison séries et intégrales).

8. Espérance et variance de T

Justifier que : $E(T) = NH_N$ et que $V(T) = N^2\left(\displaystyle\sum_{k=1}^{N}\frac{1}{k^2}\right) - NH_N$.

9. ***a.*** Établir que : $E(T) \underset{N\to+\infty}{\sim} N\ln(N)$ et $E(T) = N\ln(N) + N\gamma + o(N)$.

b. Quelle est la nature de la suite $\left(\dfrac{V(T)}{N^2}\right)_{N\in\mathbb{N}^*}$? *(on prendra garde au fait que $V(T)$ dépend de N).*
Justifier qu'il existe une constante α, strictement positive, telle que :

$$V(T) \underset{N\to\infty}{\sim} \alpha N^2 \text{ et } V(T) \leq \alpha N^2$$

10. Écart à la moyenne

On rappelle l'inégalité de Bienaymé-Chebychev valable pour une variable aléatoire X admettant une espérance et une variance :

$$\forall \varepsilon > 0, \quad \mathbf{P}(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

Soit N fixé et une constante c strictement plus grande que 1.

a. Justifier que : $\forall \omega \in \Omega, \quad |T(\omega) - N\ln(N)| \leq |T(\omega) - E(T)| + N$.
Comparer par une inclusion les événements suivants :

$$\Big(|T - N\ln(N)| \geq cN\Big) \text{ et } \Big(|T - E(T)| \geq N(c-1)\Big)$$

b. Démontrer que :

$$\mathbf{P}\Big(|T - N\ln(N)| \geq cN\Big) \leq \frac{\alpha}{(c-1)^2}$$

où α a été définie à la question **9.b.**.
Le nombre N étant fixé, que vaut $\lim_{c\to+\infty} \mathbf{P}\Big(|T - N\ln N| \geq cN\Big)$?

11. Démontrer aussi que pour tout $\varepsilon > 0$:

$$\lim_{N\to+\infty} \mathbf{P}\Big(|T - N\ln(N)| \geq \varepsilon N\ln(N)\Big) = 0$$

On peut traduire ces résultats en disant que l'événement : "T s'écarte de $N\ln(N)$ de manière significative" est un événement asymptotiquement rare.
Pour information, pour un paquet de 32 cartes, on donne $32\ln(32) \simeq 110$ *et pour un paquet de 52 cartes,* $52\ln(52) \simeq 205$.

12. Dans cette question, on considère un jeu de $N = 32$ cartes. Le paquet de 32 carte est représenté en PYTHON par une liste `Jeu` rempli initialement d'entiers entre 0 et 31. Attention donc dans cette question, les cartes sont numérotées $C_0, C_1, ..., C_{31}$. Ainsi, initialement, `Jeu(i)` renvoit i, c'est-à-dire que la carte C_i est en position i dans la liste (avec $i \in [\![0, 31]\!]$). Au cours des insertions, `Jeu(i)` désigne le numéro de la carte en position i dans la liste. Par exemple `Jeu(i)=10` signifie que la carte C_{10} est en position i dans la liste.

a. Créer en PYTHON la liste `Init` qui prend la configuration initiale, soit pour tout i de $[\![0, 31]\!]$, on a : `Init(i)=i`.

b. Créer une fonction `Insertion` en PYTHON qui à une liste de taille n prend son premier élément et l'insère aléatoirement dans $[\![0, n]\!]$.
On importera la bibliothèque `numpy.random` à l'aide de `import numpy.random as rd`.
L'expression `rd.randint(a, b)` permet de choisir un entier au hasard dans l'intervalle $[\![a, b-1]\!]$.

c. Que permet de faire le programme suivant :

```
Jeu=Init
mystere=1
while Jeu[0]!=31:
    Jeu=Insertion(Jeu)
    mystere+=1
print(mystere)
```

d. Écrire un programme en PYTHON qui permet d'afficher la moyenne des valeurs prises par la variable aléatoire T sur 100 expériences.

Partie 3 - Distance variationnelle à la loi uniforme

Notations :

- On note π l'équiprobabilité sur $\mathcal{S}_N$, c'est-à-dire l'application de $\mathcal{P}\left(\mathcal{S}_N\right)$ dans $[0,1]$ telle que :

$$\forall A \subset \mathcal{S}_N \quad \pi\left(A\right) = \frac{card\left(A\right)}{N!} \text{ ; en particulier, } \forall \sigma \in \mathcal{S}_N \quad \pi\left(\{\sigma\}\right) = \frac{1}{N!}$$

- On note également μ_n la probabilité sur $\mathcal{S}_N$ définie comme suit :
pour chaque configuration σ de $\mathcal{S}_N$, $\mu\left(\{\sigma\}\right)$ désigne la probabilité qu'à l'instant n le tas de cartes se trouve dans la configuration σ.
On a alors pour pour toute partie A de $\mathcal{S}_N$, $\mu_n\left(A\right) = \sum\limits_{\sigma\in A} \mu_n\left(\sigma\right)$.
On peut mesurer la qualité du mélange à un instant donné n en estimant l'écart entre μ_n et π.

Une *distance* d entre ces probabilités est définie de la manière suivante :

$$d\left(\mu_n,\pi\right) = \max\left\{\left|\mu_n\left(A\right) - \pi\left(A\right)\right|,\ A \subset \mathcal{S}_N\right\}$$

13. Soient A une partie de $\mathcal{S}_N$, $n \in \mathbb{N}^*$ et E_n l'événement : "à l'instant n le paquet de cartes se trouve dans une configuration qui appartient à la partie A."

a. Expliquer, en utilisant la question 7°, l'égalité suivante : $\mathbf{P}_{(T\leq n)}\left(E_n\right) = \pi\left(A\right)$.
En déduire : $\mathbf{P}\left(E_n \cap (T \leq n)\right) = \pi\left(A\right)\mathbf{P}\left(T \leq n\right)$.

b. Établir que : $\mathbf{P}\left(E_n \cap (T > n)\right) \leq \mathbf{P}\left(T > n\right)$.

c. Montrer que :

$$\mu_n\left(A\right) \leq \pi\left(A\right) + \mathbf{P}\left(T > n\right)$$

14. Soit A une partie de $\mathcal{S}_N$ et $n \in \mathbb{N}^*$. On note $\overline{A}$ l'événement contraire de A.

a. Exprimer $\mu_n\left(\overline{A}\right) - \pi\left(\overline{A}\right)$ en fonction de $\mu_n\left(A\right) - \pi\left(A\right)$.

b. Déduire des questions précédentes la majoration :

$$\left|\mu_n\left(A\right) - \pi\left(A\right)\right| \leq \mathbf{P}\left(T > n\right)$$

15. Montrer que : $\forall\ n \in \mathbb{N}^*,\ 0 \leq d\left(\mu_n,\pi\right) \leq \mathbf{P}\left(T > n\right)$. Déterminer la limite $\lim\limits_{n\to+\infty} d\left(\mu_n,\pi\right)$.

Partie 4- Une majoration de $\mathbf{P}(T > n)$

Dans cette partie, nous nous intéressons provisoirement à un collectionneur de timbres. Celui-ci reçoit chaque jour une lettre affranchie avec un timbre choisi au hasard uniformément parmi les N timbres en vigueur. On étudie ici le nombre de jours que doit attendre le collectionneur pour posséder la collection complète des N timbres. Le jour 0 il n'a aucun timbre.

On note alors :

- pour tout entier $k \in [\![1, N]\!]$, S_k le nombre aléatoire de jours que doit attendre le collectionneur pour que le nombre de timbres différents qu'il possède passe de $k-1$ à k,
- $S = S_1 + S_2 + \cdots + S_N$, soit la variable aléatoire correspondant au nombre de jours à attendre pour posséder la collection complète des N timbres,
- en supposant les N timbres en vigueur numérotés de 1 à N, pour tout $j \in [\![1, N]\!]$, B_j^m l'événement "le jour m, le collectionneur n'a toujours pas reçu de lettre affranchie avec le timbre numéro j."

On admet que les variables aléatoires $(S_k)_{k\in[\![1,N]\!]}$ sont indépendantes.

16. Déterminer la loi de S_1.

17. Déterminer pour tout entier $k \in [\![2, N]\!]$ la loi de la variable S_k.

18. En déduire que la variable S suit la même loi de probabilité que la variable T étudiée dans les parties précédentes.

Ce résultat sera utilisé pour estimer la quantité $\mathbf{P}(T > n)$.

19. Soit $m \in \mathbb{N}^*$.

- ***a.*** Exprimer l'événement $(S > m)$ à l'aide des événements $B_1^m,\ B_2^m,\ \dots,\ B_N^m$.
- ***b.*** Que vaut $\mathbf{P}(B_j^m)$ pour tout entier $j \in [\![1, N]\!]$?
- ***c.*** Montrer que : $\mathbf{P}(S > m) \leq N\left(1 - \dfrac{1}{N}\right)^m$.

20.

- ***a.*** Montrer que $\ln(1 + x) \leq x$ pour tout $x \in]-1, +\infty[$.
- ***b.*** Déduire des résultats précédents la majoration

$$\forall m \in \mathbb{N},\ \mathbf{P}(T > m) \leq Ne^{-\frac{m}{N}}$$

21. On reprend les notations introduites dans la partie précédente.

- ***a.*** Soit $c > 0$ fixé. Montrer que pour n entier supérieur ou égal à $N \ln N + cN$ on a : $d(\mu_n, \pi) \leq e^{-c}$.
- ***b.*** *Application numérique.* On estime qu'une distance en variation à la loi uniforme de $0,2$ est acceptable.
  Avec un jeu de 32 cartes, combien de *battages par insertions* doit-on faire pour considérer le paquet mélangé de façon acceptable ?

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_1P_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 $P_1 P_2 \dots P_n$ est constituée des $n+1$ réels $\mathbf{P}([U_n = 0]), \mathbf{P}([U_n = 1]), \dots, \mathbf{P}([U_n = n])$ 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
$\exp Q_1 \times \exp Q_2 \times \dots \times \exp Q_n$ est constituée par les N réels
$\mathbf{P}([V_n = 0]), \mathbf{P}([V_n = 1]), \dots, \mathbf{P}([V_n = N-1])$.

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 :

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

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 appliquent 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 $\sigma := (\sigma_1, \sigma_2, \dots, \sigma_n)$ 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 $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ 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 : $H_n = \sum_{k=1}^{n} \frac{1}{k}$.

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 : $\mathbf{P}([R_n = 2]) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n-1}\dfrac{1}{k}$.

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 $T_1, T_2, \dots, T_n$.

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

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

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

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.