CONCOURS BLANC 1 MATHS II

ECG2 MATHS APPLIQUÉES

Dans tout le sujet, on suppose déjà importées sous leur alias habituels les bibliothèques Python usuelles.

```python
import numpy as np
import numpy.random as rd
import numpy.linalg as al
import matplotlib.pyplot as plt
```

Le sujet est long, les notations et le contexte difficiles.
On pourra admettre certaines questions, tout comme on pourra traiter certaines questions de milieu de sujet indépendantes de tout le reste (il y en a).
Une lecture approfondie de l'intégralité du sujet avant de se lancer dans la résolution des questions est indispensable.

Introduction : contexte et notations. 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. Plus précisément

- On rappelle qu'une **permutation** de l'ensemble $\{C_1, C_2, ..., C_N\}$ est une bijection de cet ensemble dans lui même. On note S_N l'ensemble des permutations possibles pour ce paquet de N cartes et on rappelle que $\text{Card}(S_N) = N!$.
- On se place dans un espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ avec $\Omega = S_N$ et P l'équiprobabilité sur Ω. C'est à dire que l'ensemble des issues de cet univers est l'ensemble des permutations du paquet de cartes.

On considère qu'un paquet est *convenablement mélangé* lorsque toutes les permutations sont équiprobables, c'est-à-dire lorsque pour toute permutation σ de S_N la probabilité que le tas de cartes se trouve dans la configuration σ vaut $1/N!$.
Une carte située au sommet de la pile est dite en *position* 1, celle qui se trouve immédiatement en dessous est dite en *position* 2, et ainsi de suite jusqu'à la *position* N pour désigner 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.**

On part d'un tas de cartes rangées, au moment initial $n = 0$, 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.
On introduit les variables aléatoires ci-dessous.

Date: 7 Décembre 2023 8h30-12h30.
http://louismerlin.fr.

- Pour $k \in \mathbb{N}^*$, X_k la place de la $k-$ième insertion. On peut donc supposer que les variables aléatoires (X_k) sont indépendantes ;
- T_1 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 $N-2$;
- Plus généralement, pour $i \in [\![1, N-1]\!]$, T_i le premier instant où la carte C_N atteint la position $N-i$.
- On posera également $\Delta_1 = T_1$ et, pour $i \in [\![2; N-1]\!]$, $\Delta_i = T_i - T_{i-1}$.
- 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 $(\Delta_i)_{1 \leq i \leq N-1}$ sont **indépendantes**.

Exemple 1. 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 la dernière ligne figure 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 resultats $T_1(\omega) = 3$, $T_2(\omega) = 5$ et $T_3(\omega) = 6$ et $T(\omega) = 7$.

Partie I : Simulation d'insertions avec Python

1. **Évolution des positions successives de C_N au cours du temps**.

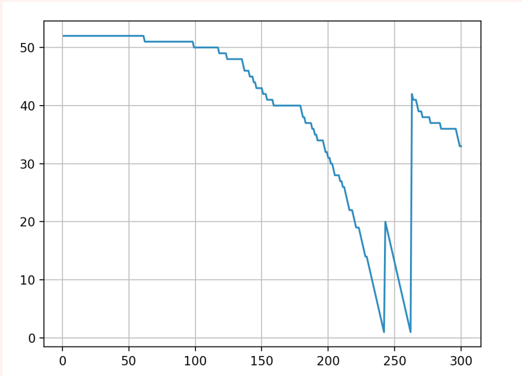
   a. Recopier et compléter la fonction ci-dessous afin qu'elle renvoie la liste des positions de C_N après n insertions.

```python
def simul_position(N, n):
    position = N
    L=[position]
    for j in range(1, n):
        if position == 1 :
            position = ................
        else :
            k=rd.randint(1, N+1)
            if k>= position :
                position = .............
        L.append(............)
    return L
```

   b. On ajoute les commandes suivantes dont l'exécution permet l'affichage ci-après. Par lecture graphique, déterminer les valeurs de T_1, T_2, T pour cette simulation.

```python
N=52
n=300
plt.grid()
plt.plot([k for k in range(1, n+1)], simul_position(N,n))
plt.show()
```

Affichage Python



2. Simulation de T et conjectures.

a. Recopier et compléter la fonction suivante qui renvoie une simulation de T.

```
def simul_T(N):
    position = N
    t=0
    while position > 1 :
        if ...................   :
            .....................
        t=t+1
    return .......
```

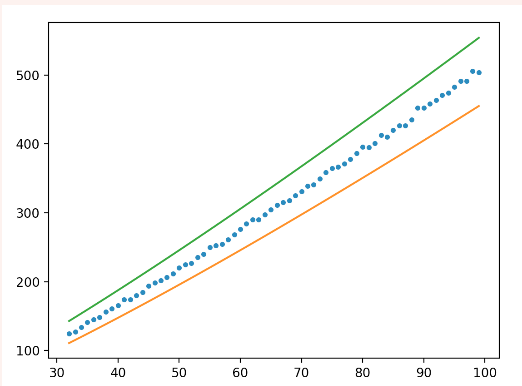
b. Que renvoie la fonction `mystere` ci-dessous ?

```
def mystere(N):
    return np.mean([simul_T(N) for k in range(1000)])
```

c. On ajoute les commandes suivantes dont l'exécution permet l'affichage ci-après. Émettre alors une conjecture.

```
abs=[k for k in range(32, 100)]
ord=[mystere(N) for N in abs]
plt.plot(abs, ord, '.')
plt.plot(abs, [k*np.log(k) for k in abs])
plt.plot(abs, [k*np.log(k)+k for k in abs])
plt.show()
```

Affichage Python



Partie II : Description et premiers résultats

3. Quelle est, pour $k \in \mathbb{N}^*$, la loi de X_k ?

4. Que représente Δ_i ? Que vaut donc $\Delta_i(\Omega)$? Justifier que, pour tout $i \in [\![2, N-1]\!]$,

$$T_i = \Delta_1 + \Delta_2 + \cdots + \Delta_i.$$

En déduire en particulier que $T_i(\Omega) = [\![i, +\infty[\![$.

5. Soient Y une variable aléatoire à valeurs dans $\mathbb{N}$ et $n \in \mathbb{N}$. Établir un lien entre

$$P(Y > n), \quad P(Y = n), \quad P(Y > n-1).$$

6. **Loi de** Δ_1. Déterminer pour tout entier $n \in \mathbb{N}$, $P(\Delta_1 > n)$

Indication : on pourra représenter l'évènement $[\Delta_1 > n]$ *à l'aide d'intersections d'évènements formés avec les variables* X_k.

Reconnaître la loi de Δ_1.

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

a. (i) Établir que pour tout $n \in \mathbb{N}$, et pour tout $m \in \mathbb{N}^*$, on a

$$P_{[T_{i-1}=m]}(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n.$$

(ii) Montrer alors que, pour tout $n \in \mathbb{N}$,

$$P(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n.$$

(iii) Conclure que Δ_i suit une loi usuelle que l'on précisera.

b. En déduire

$$E(\Delta_i) = \frac{N}{i}, \quad \text{et} \quad V(\Delta_i) = N\left(\frac{N-i}{i^2}\right).$$

8. **Loi de** T_2 . Soit $n \geq 2$.

a. Justifier que, si $k \geq j$

$$P(T_2 = j \cap \Delta_1 = k) = 0.$$

b. En déduire, à l'aide de la formule des probabilités totales, que

$$P(T_2 = n) = \sum_{k=1}^{n-1} P(\Delta_2 = n-k)\, P(\Delta_1 = k).$$

c. Justifier que

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

d. En déduire que l'on a

$$P(T_2 = n) = \frac{2}{N}\left[\left(1 - \frac{1}{N}\right)^{n-1} - \left(1 - \frac{2}{N}\right)^{n-1}\right].$$

9. 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 :

a. la carte insérée à l'instant T_1 soit en place $N-1$ et celle insérée à l'instant T_2 en place N ?

b. la carte insérée à l'instant T_2 soit en place $N-1$ et celle insérée à l'instant T_1 en place N ?

10. 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 $(a_1, a_2, a_3) = (N-2, N, N-1)$?

11. 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 d'avoir marqué la carte C_N bien sûr !

Partie III : 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 :

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

12. Espérance et variance de T

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

13. Étude de la suite (u_n)

a. Montrer que pour tout entier $k \geq 1$, on a

$$\frac{1}{k+1} \leq \int_k^{k+1} \frac{\mathrm{d}t}{t} \leq \frac{1}{k}.$$

b. En déduire successivement :

(i) la décroissance de la suite (u_n),

(ii) l'encadrement : $\forall n \in \mathbb{N}^*$, $\ln(n+1) \leq H_n \leq \ln(n) + 1$.

c. Déduire de ce qui précède que la suite (u_n) est convergente et que sa limite, notée γ appartient à $[0,1]$.

14. **a.** Établir que

$$E(T) \underset{+\infty}{\sim} N\ln(N) \qquad \text{et} \qquad E(T) = N\ln(N) + N\gamma + o(N), \quad N \to +\infty.$$

b. Quelle est la nature de la suite $(V(T)/N^2)$? (*On n'oubliera pas que $V(T)$ dépend de N.*)
Justifier qu'il existe une constante α, strictement positive, telle que

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

15. Écart à la moyenne.
On (rappelle ou) **admet** l'inégalité de *Bienaymé-Tchebychev*

Théorème : Inégalité de Bienaymé-Tchebychev

Pour toute variable aléatoire Y admettant une espérance et une variance, on a :

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

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

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

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

b. Démontrer que

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

où α a été définie à la Question 14b.
Le nombre N étant fixé, que vaut $\lim\limits_{c\to+\infty} P\left(|T - N\ln\left(N\right)| \geq cN\right)$?

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

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

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,* $52\ln(52) \simeq 205$.

Partie IV : Distance variationnelle à la loi uniforme

On précise et introduit dans cette partie quelques notations.

- Une partie $A \subset \mathcal{S}_N$ est un ensemble de configurations du paquet de cartes (c'est à dire un ensemble de permutations des cartes) ;
- 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 :

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

- On note également μ_n la probabilité sur $\mathcal{S}_N$ définie comme suit :
pour chaque configuration σ de $\mathcal{S}_N$, $\mu_n\left(\{\sigma\}\right)$ 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 :

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

17. 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 11, l'égalité suivante :

$$P_{(T\leq n)}\left(E_n\right) = \pi\left(A\right).$$

En déduire

$$P\left(E_n \cap \left(T \leq n\right)\right) = \pi\left(A\right) P\left(T \leq n\right).$$

b. Justifier que

$$P\left(E_n \cap \left(T > n\right)\right) \leq P\left(T > n\right).$$

c. Montrer que
$$\mu_n(A) \leq \pi(A) + P(T > n).$$

18. 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 $\mu_n(A) - \pi(A)$.

b. Déduire des questions précédentes la majoration :
$$|\mu_n(A) - \pi(A)| \leq P(T > n).$$

19. Montrer que, pour tout $n \in \mathbb{N}^*$,
$$0 \leq d(\mu_n, \pi) \leq P(T > n).$$
Déterminer la limite
$$\lim_{n \to +\infty} d(\mu_n, \pi).$$

Partie V : Une majoration de $P(T > n)$

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. Au 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 ;
- $S = S_1 + S_2 + \cdots + S_N$ 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.

20. Déterminer la loi de S_1.

21. Déterminer pour tout entier $k \in [\![1; N]\!]$ la loi de la variable S_k.

22. 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 $P(T > n)$.

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

a. Exprimer l'événement $(S > m)$ à l'aide des événements $B_1^m, B_2^m, ..., B_N^m$.

b. Que vaut $P(B_j^m)$ pour tout entier $j \in [\![1; N]\!]$?

c. (i) Montrer que si C et D sont deux évènements, alors $P(C \cup D) \leq P(C) + P(D)$.

(ii) En déduire que
$$P(S > m) \leq \sum_{j=1}^{N} P(B_j^m)$$
puis que
$$P(S > m) \leq N\left(1 - \frac{1}{N}\right)^m.$$

24. **a.** Montrer que $\ln(1 + x) \leq x$ pour tout $x \in]-1, +\infty[$.

b. Déduire des résultats précédents que, pour tout $m \in \mathbb{N}$,
$$P(T > m) \leq Ne^{-\frac{m}{N}}.$$

25. 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 ?

On donne $\ln(160) \simeq 5.08$.