
DS2 (version B)

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

*Les candidat·es sont invité·es à **encadrer** dans la mesure du possible les résultats de leurs calculs.*

*Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.*

Si au cours de l'épreuve, un candidat ou une candidate repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

On suppose, et c'est valable pour toute l'épreuve, que les librairies suivantes sont importées sous leurs alias habituels :

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

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 mélange soit acceptable. 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 notées $C_1, \dots, C_N$ et regroupées en un paquet. Un joueur bat les cartes et repose le paquet sur la table. Une configuration du paquet de N cartes est un état du paquet.

Par exemple, dans le cas d'un paquet de $N = 4$ cartes, (C_1, C_2, C_3, C_4) , (C_2, C_4, C_1, C_3) , (C_3, C_1, C_2, C_4) sont trois configurations possibles. Dans le cas $N = 4$, on a $4! = 24$ configurations possibles qui correspondent à l'ensemble des permutations de $\{C_1, C_2, C_3, C_4\}$.

Notations et Rappel. On note $\mathcal{S}_N$ l'ensemble des configurations possibles pour ce paquet de N cartes. On a alors : $\text{Card}(\mathcal{S}_N) = N!$.

On se place dans un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ avec $\Omega = \mathcal{S}_N$, $\mathcal{A} = \mathcal{P}(\mathcal{S}_N)$ l'ensemble des parties de $\mathcal{S}_N$ et $\mathbb{P}$ l'équiprobabilité sur Ω . Pour toute variable aléatoire X on notera $\mathbb{E}(X)$ et $\mathbb{V}(X)$ l'espérance et la variance de X lorsqu'elles existent.

On considère qu'une méthode de mélange est *acceptable* si :

- × toutes les configurations peuvent être obtenues avec cette méthode de mélange,
- × chacune des configurations a la même probabilité d'être obtenue.

Ainsi la probabilité d'apparition de chaque configuration σ de $\mathcal{S}_N$ est $\frac{1}{N!}$, c'est-à-dire :

$$\forall \sigma \in \mathcal{S}_N, \mathbb{P}(\{\sigma\}) = \frac{1}{N!}$$

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 dans la configuration initiale suivante : $(C_1, C_2, \dots, C_N)$ (*pour tout i élément de $\llbracket 1, N \rrbracket$, 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 tout en dessous du paquet.

Pour k élément de $\llbracket 1, N \rrbracket$, on appelle *insertion* à la $k^{\text{ème}}$ place l'opération qui consiste à prendre la carte située au-dessus du paquet et à l'insérer entre la $k^{\text{ème}}$ et la $(k+1)^{\text{ème}}$ place. Une insertion à la première place ne change pas l'ordre des cartes. Une insertion à la $N^{\text{ème}}$ 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 $\{1, \dots, N\}$ la place à laquelle l'insertion a lieu, indépendamment des insertions précédentes.

Les instants successifs d'insertions seront notées $1, 2, \dots, n, \dots$, 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 $N-1$,
- × T_2 le premier instant où la carte C_N se trouve remontée en position $N-2$,
- × et plus généralement, pour tout i dans $\llbracket 1, N-1 \rrbracket$, T_i le premier instant où la carte C_N atteint la position $N-i$.
- × on posera également $\Delta_1 = T_1$ et $\forall i \in \llbracket 2, N-1 \rrbracket, \Delta_i = T_i - T_{i-1}$.
- × on notera $T = T_{N-1} + 1$.
- × on définit enfin, pour tout $j \in \mathbb{N}^*$, la variable aléatoire X_j qui correspond à la place de l'insertion effectuée à l'instant j .

On admet que les conditions de l'expérience permettent de faire l'hypothèse que les variables aléatoires $(\Delta_i)_{i \in \llbracket 1, N-1 \rrbracket}$ 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 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 résultats :

- × $T_1(\omega) = 3$, $T_2(\omega) = 5$ et $T_3(\omega) = 6$ et $T(\omega) = 7$,
- × $X_1(\omega) = 3$, $X_2(\omega) = 2$, $X_3(\omega) = 4$, $X_4(\omega) = 1$, $X_5(\omega) = 3$, $X_6(\omega) = 4$ et $X_7(\omega) = 2$.

Partie 1 - Description et premiers résultats

1. Justifier : $\forall i \in \llbracket 2, N-1 \rrbracket$, $T_i = \Delta_1 + \Delta_2 + \dots + \Delta_i$.

Que représente l'intervalle de temps Δ_i ?

2. Loi de Δ_1 .

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

3. Soit $i \in \llbracket 2, N-1 \rrbracket$. Loi de Δ_i .

a) Établir que pour tout entier $n \in \mathbb{N}$, on a : $\mathbb{P}([\Delta_i > n]) = \left(\frac{N-i}{N}\right)^n$. En déduire que Δ_i suit une loi usuelle que l'on précisera.

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

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

a) Démontrer : $\mathbb{P}([T_2 = n]) = \sum_{k=1}^{n-1} \mathbb{P}([\Delta_2 = n-k]) \mathbb{P}([\Delta_1 = k])$.

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

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

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 .

On admettra sans démonstration que X_{T_2} est une variable aléatoire. La v.a.r. X_{T_2} correspond à la place de l'insertion à l'instant aléatoire T_2 . En d'autres termes, X_{T_2} est la place de l'insertion à l'instant où la carte C_N passe de la position $N-1$ à la position $N-2$.

a) Déterminer $X_{T_2}(\Omega)$.

- b)* À quelles conditions nécessaires et suffisantes l'événement $[X_{T_1} = N] \cap [X_{T_2} = N]$ est-il réalisé ?
Même question pour l'événement $[X_{T_1} = N] \cap [X_{T_2} = N - 1]$.
- c)* En déduire la valeur de $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N])$ et $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N - 1])$.
- 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 admettra sans démonstration que X_{T_3} est une v.a.r. .
- a)* Déterminer $X_{T_3}(\Omega)$.
- b)* Avec le même raisonnement qu'en question **5**, déterminer :

$$\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N]) \quad \text{et} \quad \mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N - 1])$$

- 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 :

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

- 8.** Espérance et variance de T

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

- 9.** Étude de la suite (u_n)

a) Montrer, pour tout entier $k \geq 1$: $\frac{1}{k+1} \leq \int_k^{k+1} \frac{1}{t} dt \leq \frac{1}{k}$.

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 $[0, 1]$.

- 10. a)** Établir : $\mathbb{E}(T) \underset{N \rightarrow +\infty}{\sim} N \ln(N)$ et $\mathbb{E}(T) = N \ln(N) + N \gamma + o_{N \rightarrow +\infty}(N)$.

b) Quelle est la nature de la suite $\left(\frac{\mathbb{V}(T)}{N^2} \right)_{N \in \mathbb{N}^*}$?

(on prendra garde au fait que $\mathbb{V}(T)$ dépend de N).

Justifier qu'il existe une constante α , strictement positive, telle que :

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

11. Écart à la moyenne

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

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

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

a) Justifier : $\forall \omega \in \Omega, |T(\omega) - N \ln(N)| \leq |T(\omega) - \mathbb{E}(T)| + N$.

Comparer par une inclusion les événements suivants :

$$[|T - N \ln(N)| \geq cN] \quad \text{et} \quad [|T - \mathbb{E}(T)| \geq N(c-1)]$$

b) Démontrer :

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

où α a été définie à la question **10.b**).

Le nombre N étant fixé, que vaut $\lim_{c \rightarrow +\infty} \mathbb{P}(|T - N \ln(N)| \geq cN)$?

12. Démontrer aussi, pour tout $\varepsilon > 0$:

$$\lim_{N \rightarrow +\infty} \mathbb{P}(|T - N \ln(N)| \geq \varepsilon N \ln(N)) = 0$$

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

13. Simulation informatique. Dans cette question on considère un jeu de $N = 32$ cartes.

MODÉLISATION : le paquet de 32 cartes est représenté par un tableau numpy `Jeu` rempli initialement des entiers 1 à 32. La carte sur le dessus du paquet se trouve en première coordonnée du tableau, celle en dessous se trouve en dernière coordonnée.

De manière générale, et puisque **Python** commence la numérotation à 0, pour tout $i \in \llbracket 0, 31 \rrbracket$, `Jeu[i]` désigne la carte (représentée par son numéro) qui se trouve en $(i+1)^{\text{e}}$ position dans le paquet. Par exemple, si `Jeu[i]` contient 10 c'est que la carte C_{10} est en position $i+1$. Initialement, `Jeu` est rempli de telle sorte que `Jeu[i] = i+1` pour tout i dans $\llbracket 0, N-1 \rrbracket$.

a) Écrire la commande **Python** permettant de définir le tableau correspondant à la configuration initiale du paquet de cartes.

b) Compléter la fonction `Insertion` qui simule une opération d'insertion. Elle prend en paramètre un tableau qui donne la configuration du jeu avant insertion et renvoie un tableau qui contient la configuration du jeu après insertion.

On rappelle que la fonction `rd.randint(1,33)` permet de tirer un nombre entier au hasard dans l'intervalle $\llbracket 1, 32 \rrbracket$.

```

1 def Insertion(Jeu):
2     k = _____ # position où l'on va insérer la carte du dessus
3     J = Jeu
4     cartedessus = Jeu[0]
5     if k > 1:
6         for i in range(k-1):
7             J[i] = _____
8             J[k-1] = _____
9     return J

```

c) Que fait la fonction `Simu` suivante ?

```

1 def Simu():
2     Jeu = np.arange(1,33)
3     n = 0
4     while Jeu[0] != 32:
5         Jeu = Insertion(Jeu)
6         n = n + 1
7     return n + 1

```

d) Écrire en **Python** le programme principal permettant de calculer et 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 $\mathcal{P}(\mathcal{S}_N)$ dans $[0, 1]$ telle que :

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

- On note également μ_n l'application probabilité définie sur $\mathcal{S}_N$ comme suit :
pour chaque configuration σ de $\mathcal{S}_N$, $\mu_n(\{\sigma\})$ désigne la probabilité qu'à l'instant n le tas de cartes se trouve dans la configuration σ .

On a alors pour toute partie A de $\mathcal{S}_N$: $\mu_n(A) = \sum_{\sigma \in A} \mu_n(\{\sigma\})$.

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 :

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

14. 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 : $\mathbb{P}_{[T \leq n]}(E_n) = \pi(A)$.

En déduire : $\mathbb{P}(E_n \cap [T \leq n]) = \pi(A) \mathbb{P}([T \leq n])$.

b) Établir : $\mathbb{P}(E_n \cap [T > n]) \leq \mathbb{P}([T > n])$.

c) Montrer :

$$\mu_n(A) \leq \pi(A) + \mathbb{P}([T > n])$$

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

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

b) Déduire des questions précédentes la majoration :

$$|\mu_n(A) - \pi(A)| \leq \mathbb{P}([T > n])$$

16. Montrer que $\forall n \in \mathbb{N}^*, 0 \leq d(\mu_n, \pi) \leq \mathbb{P}([T > n])$. Déterminer la limite $\lim_{n \rightarrow +\infty} d(\mu_n, \pi)$.

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

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 $k \in \llbracket 1, n \rrbracket$ 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 ,
- × $S = S_1 + S_2 + \dots + S_N$, 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 $j \in \llbracket 1, N \rrbracket$, 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 $(S_k)_{k \in \llbracket 1, N \rrbracket}$ sont indépendantes.

17. Déterminer la loi de S_1 .

18. Déterminer, pour tout entier $k \in \llbracket 2, N \rrbracket$, la loi de la variable S_k .

19. 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 $\mathbb{P}([T > n])$.

20. 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 $\mathbb{P}(B_j^m)$ pour tout entier $j \in \llbracket 1, N \rrbracket$?

c) Démontrer que, pour tout entier $n \geq 2$ et pour toute famille d'événements $A_1, \dots, A_n$, on a l'inégalité : $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i)$.

d) En déduire : $\mathbb{P}([S > m]) \leq N \left(1 - \frac{1}{N}\right)^m$.

21. a) Montrer : $\forall x \in]-1, +\infty[, \ln(1+x) \leq x$.

b) Déduire des résultats précédents la majoration :

$$\forall m \in \mathbb{N}, \mathbb{P}([T > m]) \leq N e^{-\frac{m}{N}}$$

22. 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 que le paquet est mélangé de façon acceptable ?