
DS2 (version B) - correction (ESSEC II 2011)

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 ?

Démonstration.

D'après l'énoncé : $\forall k \in \llbracket 2, N-1 \rrbracket, T_k - T_{k-1} = \Delta_k.$

Soit $i \in \llbracket 2, N-1 \rrbracket.$ En reconnaissant une somme télescopique :

$$\sum_{k=2}^i (T_k - T_{k-1}) = \sum_{k=2}^i \Delta_k$$

||

$$T_i - T_1$$

Or $T_1 = \Delta_1,$ donc :

$$T_i = \Delta_1 + \sum_{k=2}^i \Delta_k = \sum_{k=1}^i \Delta_k$$

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

Pour tout $i \in \llbracket 1, N-1 \rrbracket,$ la v.a.r. $\Delta_i = T_i - T_{i-1}$ représente le nombre d'insertions entre l'instant où la carte C_N est en position $N - (i - 1)$ et l'instant où elle est en position $N - i.$ □

2. Loi de $\Delta_1.$

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

Démonstration.

- Soit $n \in \mathbb{N}.$ L'événement $[\Delta_1 > n]$ est réalisé si et seulement si la carte C_N monte en position $N - 1$ strictement après l'instant $n,$ c'est-à-dire si aucune des insertions de l'instant 1 à l'instant n ne s'effectue en position $N.$ On en déduit :

$$[\Delta_1 > n] = [X_1 \neq N] \cap [X_2 \neq N] \cap \dots \cap [X_n \neq N]$$

Comme les insertions sont indépendantes, les v.a.r. X_1, \dots, X_n le sont aussi. Donc :

$$\mathbb{P}([\Delta_1 > n]) = \mathbb{P}\left(\bigcap_{j=1}^n [X_j \neq N]\right) = \prod_{j=1}^n \mathbb{P}([X_j \neq N])$$

De plus, la carte du dessus du paquet est toujours insérée uniformément sur les N positions possibles. Donc : $\forall j \in \mathbb{N}^*, \mathbb{P}([X_j \neq N]) = \frac{N-1}{N}$.

Ainsi :

$$\mathbb{P}([\Delta_1 > n]) = \prod_{j=1}^n \left(\frac{N-1}{N}\right) = \left(\frac{N-1}{N}\right)^n = \left(1 - \frac{1}{N}\right)^n$$

$$\forall n \in \mathbb{N}, \mathbb{P}([\Delta_1 > n]) = \left(1 - \frac{1}{N}\right)^n$$

• Loi de Δ_1 .

× Tout d'abord : $\Delta_1(\Omega) = \mathbb{N}^*$.

En effet, l'insertion de la carte du dessus du paquet en position N peut se faire dès le 1^{er} instant, ou aux instants suivants. Elle peut également ne jamais se produire.

× De plus, d'après ce qui précède : $\forall n \in \mathbb{N}^*, \mathbb{P}([\Delta_1 > n]) = \left(1 - \frac{1}{N}\right)^n$

$$\text{On en déduit : } \Delta_1 \hookrightarrow \mathcal{G}\left(\frac{1}{N}\right).$$

Commentaire

• On rappelle pour X une v.a.r. telle que $X \hookrightarrow \mathcal{G}(p)$, on a :

$$\forall k \in \mathbb{N}^*, \mathbb{P}([X > k]) = (1-p)^k$$

• Dans cette question, on déduit directement la loi de Δ_1 à partir de :

× $\Delta_1(\Omega) = \mathbb{N}^*$

× $\forall n \in \mathbb{N}^*, \mathbb{P}([\Delta_1 > n]) = \left(1 - \frac{1}{N}\right)^n$

plutôt que de :

× $\Delta_1(\Omega) = \mathbb{N}^*$

× $\forall n \in \mathbb{N}^*, \mathbb{P}([\Delta_1 = n]) = \left(1 - \frac{1}{N}\right)^{n-1} \frac{1}{N}$

En effet, la formulation de l'énoncé suggère que l'on peut conclure rapidement.

• Il est cependant **indispensable** de savoir déterminer les réels $\mathbb{P}([\Delta_1 = n])$ à partir des réels $\mathbb{P}([\Delta_1 > n])$. On rappelle la marche à suivre.

Soit $n \geq 1$. On a l'égalité entre événements suivante :

$$\begin{aligned} [\Delta_1 > n-1] &= [\Delta_1 \geq n] && \text{(car } \Delta_1 \text{ est à} \\ & && \text{valeurs entières)} \\ &= [\Delta_1 = n] \cup [\Delta_1 > n] \end{aligned}$$

Les événements $[\Delta_1 = n]$ et $[\Delta_1 > n]$ sont incompatibles. Donc :

$$\begin{aligned} \mathbb{P}([\Delta_1 = n]) &= \mathbb{P}([\Delta_1 > n-1]) - \mathbb{P}([\Delta_1 > n]) \\ &= \left(1 - \frac{1}{N}\right)^{n-1} - \left(1 - \frac{1}{N}\right)^n = \left(1 - \frac{1}{N}\right)^{n-1} \frac{1}{N} \end{aligned}$$

□

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.

Démonstration.

• Soit $n \in \mathbb{N}$.

L'événement $[\Delta_i > n]$ est réalisé si et seulement si la carte C_N passe de la position $N-i+1$ en position $N-i$ en strictement plus de n insertions, c'est-à-dire si aucune des insertions de l'instant T_{i-1} à l'instant $T_{i-1}+n$ ne s'effectue en une position comprise entre $N-i+1$ et N . Soit $k \in \mathbb{N}^*$. On a :

$$\begin{aligned} & [T_{i-1} = k] \cap [\Delta_i > n] \\ &= [T_{i-1} = k] \cap [T_i - T_{i-1} > n] \\ &= [T_{i-1} = k] \cap [T_i > k+n] \\ &= [T_{i-1} = k] \cap [X_{k+1} \notin \llbracket N-i+1, N \rrbracket] \cap [X_{k+2} \notin \llbracket N-i+1, N \rrbracket] \cap \cdots \cap [X_{k+n} \notin \llbracket N-i+1, N \rrbracket] \\ &= [T_{i-1} = k] \cap [X_{k+1} \in \llbracket 1, N-i \rrbracket] \cap [X_{k+2} \in \llbracket 1, N-i \rrbracket] \cap \cdots \cap [X_{k+n} \in \llbracket 1, N-i \rrbracket] \end{aligned}$$

La famille $([T_{i-1} = k])_{k \in \mathbb{N}^*}$ est un système complet d'événements, car la v.a.r. T_{i-1} est à valeurs entières non nulles.

Par formule des probabilités totales :

$$\begin{aligned} & \mathbb{P}([\Delta_i > n]) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k] \cap [\Delta_i > n]) \\ &= \sum_{k=1}^{+\infty} \mathbb{P} \left([T_{i-1} = k] \cap \left(\bigcap_{j=k+1}^{k+n} [X_j \in \llbracket 1, N-i \rrbracket] \right) \right) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k]) \mathbb{P} \left(\bigcap_{j=k+1}^{k+n} [X_j \in \llbracket 1, N-i \rrbracket] \right) \quad \text{(par indépendance de } [T_{i-1} = k] \text{ et } \bigcap_{j=k+1}^{k+n} [X_j \in \llbracket 1, N-i \rrbracket]) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k]) \prod_{j=k+1}^{k+n} \mathbb{P}([X_j \in \llbracket 1, N-i \rrbracket]) \quad \text{(par indépendance mutuelle des v.a.r. } X_j) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k]) \left(\frac{N-i}{N}\right)^{k+n-(k+1)+1} \quad \text{(par équiprobabilité)} \\ &= \left(\frac{N-i}{N}\right)^n \sum_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k]) \end{aligned}$$

Comme $([T_{i-1} = k])_{k \in \mathbb{N}^*}$ est un système complet d'événements : $\sum_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k]) = 1$.

Ainsi :

$$\forall n \in \mathbb{N}, \mathbb{P}([\Delta_i > n]) = \left(1 - \frac{i}{N}\right)^n$$

Commentaire

Détaillons la justification de l'indépendance des événements $[T_{i-1} = k]$ et $\bigcap_{j=k+1}^{k+n} [X_j \in [1, N - i]]$.

- D'une part, l'événement $[T_{i-1} = k]$ est réalisé si et seulement si la carte C_N atteint la position $N - i + 1$ à l'instant k . Cet événement dépend donc uniquement des k premières insertions. Il peut s'exprimer en fonction d'événements ne faisant intervenir que les v.a.r. X_1, \dots, X_k .
- D'autre part, l'expression de l'événement $\bigcap_{j=k+1}^{k+n} [X_j \in [1, N - i]]$ ne fait intervenir que les v.a.r. X_{k+1}, \dots, X_{k+n} .

Or la suite $(X_j)_{j \in \mathbb{N}^*}$ est une suite de v.a.r. mutuellement indépendantes (car les insertions le sont).

Ainsi, d'après le lemme des coalitions, les événements $[T_{i-1} = k]$ et $\bigcap_{j=k+1}^{k+n} [X_j \in [1, N - i]]$ sont bien indépendants.

• Loi de Δ_i .

× Tout d'abord : $\Delta_i(\Omega) = \mathbb{N}^*$.

En effet, la montée de la carte C_N de la position $N - i + 1$ à la position $N - i$ peut s'effectuer dès la 1^{ère} insertion, ou aux insertions suivantes. Elle peut également ne jamais se produire.

× De plus, d'après ce qui précède : $\forall n \in \mathbb{N}^*, \mathbb{P}([\Delta_i > n]) = \left(1 - \frac{i}{N}\right)^n$

On en déduit : $\Delta_i \hookrightarrow \mathcal{G}\left(\frac{i}{N}\right)$.

□

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

Démonstration.

Comme $\Delta_i \hookrightarrow \mathcal{G}\left(\frac{N - i}{N}\right)$, on en déduit que la v.a.r. Δ_i admet une variance (et donc une espérance). De plus :

$$\mathbb{E}(\Delta_i) = \frac{1}{\frac{i}{N}} = \frac{N}{i}$$

Enfin :

$$\mathbb{V}(\Delta_i) = \frac{1 - \frac{i}{N}}{\left(\frac{i}{N}\right)^2} = \frac{\frac{N-i}{N}}{\frac{i^2}{N^2}} = N \frac{N - i}{i^2}$$

$\mathbb{E}(\Delta_i) = \frac{N}{i}$ et $\mathbb{V}(\Delta_i) = N \frac{N - i}{i^2}$

□

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

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

Démonstration.

- Tout d'abord, d'après la question 1 : $T_2 = \Delta_1 + \Delta_2$. Donc :

$$[T_2 = n] = [\Delta_1 + \Delta_2 = n]$$

- De plus, la famille $([\Delta_1 = k])_{k \in \mathbb{N}^*}$ est un système complet d'événements. Ainsi, par formule des probabilités totales :

$$\begin{aligned} \mathbb{P}([\Delta_1 + \Delta_2 = n]) &= \sum_{k=1}^{+\infty} \mathbb{P}([\Delta_1 = k] \cap [\Delta_1 + \Delta_2 = n]) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}([\Delta_1 = k] \cap [k + \Delta_2 = n]) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}([\Delta_1 = k]) \mathbb{P}([\Delta_2 = n - k]) && \text{(car les v.a.r. } \Delta_1 \text{ et } \Delta_2 \\ & && \text{sont indépendantes)} \\ &= \sum_{\substack{k=1 \\ n-k \in \Delta_2(\Omega)}}^{+\infty} \mathbb{P}([\Delta_1 = k]) \mathbb{P}([\Delta_2 = n - k]) \\ &+ \sum_{\substack{k=1 \\ n-k \notin \Delta_2(\Omega)}}^{+\infty} \mathbb{P}([\Delta_1 = k]) \mathbb{P}([\Delta_2 = n - k]) && \text{(car si } n - k \notin \Delta_2(\Omega) \\ & && \text{alors } [\Delta_2 = n - k] = \emptyset) \\ &= \sum_{k=1}^{n-1} \mathbb{P}([\Delta_1 = k]) \mathbb{P}([\Delta_2 = n - k]) \end{aligned}$$

La dernière ligne est obtenue en constatant :

$$\begin{cases} n - k \in \Delta_2(\Omega) = \mathbb{N}^* \\ k \geq 1 \end{cases} \Leftrightarrow \begin{cases} 1 \leq n - k \\ 1 \leq k \end{cases} \Leftrightarrow \begin{cases} k \leq n - 1 \\ 1 \leq k \end{cases} \Leftrightarrow \{ 1 \leq k \leq n - 1 \}$$

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

□

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

Démonstration.

La suite $\left(\left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}} \right)^k \right)_{k \in \mathbb{N}}$ est une suite géométrique de raison $\frac{1-\frac{1}{N}}{1-\frac{2}{N}} \neq 1$. Donc :

$$\begin{aligned} \sum_{k=1}^{n-1} \left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}} \right)^k &= \frac{1-\frac{1}{N}}{1-\frac{2}{N}} \frac{1-\left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}}\right)^{n-1}}{1-\frac{1-\frac{1}{N}}{1-\frac{2}{N}}} \\ &= \frac{1-\frac{1}{N}}{\cancel{1-\frac{2}{N}}} \frac{1-\left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}}\right)^{n-1}}{\frac{1-\frac{2}{N}-(1-\frac{1}{N})}{\cancel{1-\frac{2}{N}}}} \\ &= \left(1-\frac{1}{N}\right) \frac{1-\left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}}\right)^{n-1}}{-\frac{1}{N}} \\ &= N \left(1-\frac{1}{N}\right) \left(\left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}} \right)^{n-1} - 1 \right) \end{aligned}$$

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

□

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

Démonstration.

Reprenons le calcul de la question 4.a).

$$\begin{aligned} \mathbb{P}([T_2 = n]) &= \sum_{k=1}^{n-1} \mathbb{P}([\Delta_1 = k]) \mathbb{P}([\Delta_2 = n - k]) \\ &= \sum_{k=1}^{n-1} \left(1 - \frac{1}{N} \right)^{k-1} \frac{1}{N} \left(1 - \frac{2}{N} \right)^{n-k-1} \frac{2}{N} && \text{(car } \Delta_1 \hookrightarrow \mathcal{G}\left(\frac{1}{N}\right) \text{ et } \Delta_2 \hookrightarrow \mathcal{G}\left(\frac{2}{N}\right)) \\ &= \frac{2}{N^2} \frac{\left(1 - \frac{2}{N} \right)^{n-1}}{1 - \frac{1}{N}} \sum_{k=1}^{n-1} \left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}} \right)^k \\ &= \frac{2}{N^2} \frac{\left(1 - \frac{2}{N} \right)^{n-1}}{\cancel{1-\frac{1}{N}}} \times \cancel{N} \left(1 - \frac{1}{N} \right) \left(\left(\frac{1-\frac{1}{N}}{1-\frac{2}{N}} \right)^{n-1} - 1 \right) && \text{(d'après 4.b)} \\ &= \frac{2}{N} \left(\left(1 - \frac{1}{N} \right)^{n-1} - \left(1 - \frac{2}{N} \right)^{n-1} \right) \end{aligned}$$

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

□

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 $X_{T_2}(\Omega)$.

Démonstration.

Pour que la carte C_N passe de la position $N - 1$ à la position $N - 2$, il faut effectuer une insertion à la position $N - 1$ ou à la position N . Or X_{T_2} correspond à la position de l'insertion au moment précis où la carte C_N passe de $N - 1$ à $N - 2$.

$$\text{On en déduit : } X_{T_2}(\Omega) = \{N - 1, N\}.$$

□

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 $[X_{T_1} = N] \cap [X_{T_2} = N - 1]$.

Démonstration.

Soit $\omega \in \Omega$.

- ω réalise l'événement $[X_{T_1} = N] \cap [X_{T_2} = N]$ si et seulement si :
 - × à l'instant $T_1(\omega)$, la carte du dessus du paquet est insérée en position N .
On peut d'ailleurs noter que, par définition de T_1 (le premier instant où C_N passe de la position N à la position $N - 1$), l'insertion à l'instant $T_1(\omega)$ se fait toujours en position N (seule position de l'insertion qui permet de faire monter C_N de la position N à $N - 1$).
 - × à l'instant $T_2(\omega)$, la carte du dessus du paquet est insérée en position N .
- De même, ω réalise l'événement $[X_{T_1} = N] \cap [X_{T_2} = N - 1]$ si et seulement si :
 - × à l'instant $T_1(\omega)$, la carte du dessus du paquet est insérée en position N (c'est de toute façon la seule insertion possible à l'instant $T_1(\omega)$ comme expliqué précédemment),
 - × à l'instant $T_2(\omega)$, la carte du dessus du paquet est insérée en position $N - 1$.

□

c) En déduire la valeur de $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N])$ et $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N - 1])$.

Démonstration.

- D'après la question 5.a), la famille $([X_{T_2} = N - 1], [X_{T_2} = N])$ forme un système complet d'événements.
 Ainsi, par formule des probabilités totales :

$$\mathbb{P}([X_{T_1} = N]) = \mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N - 1]) + \mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N])$$

- De plus, comme expliqué en question précédente : pour tout $\omega \in \Omega$, la seule insertion possible à l'instant $T_1(\omega)$ est en position N . Autrement dit : $\forall \omega \in \Omega, \omega \in [X_{T_1} = N]$. Donc :

$$[X_{T_1} = N] = \Omega$$

Ainsi :

$$\begin{array}{rcccc} \mathbb{P}([X_{T_1} = N]) & = & \mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N - 1]) & + & \mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N]) \\ \parallel & & \parallel & & \parallel \\ 1 & = & \mathbb{P}(\Omega) & & \mathbb{P}([X_{T_2} = N]) \end{array}$$

- Enfin, comme la carte du dessus est insérée uniformément sur l'ensemble des positions possibles, on a :

$$\mathbb{P}([X_{T_2} = N - 1]) = \mathbb{P}([X_{T_2} = N])$$

- On obtient :

$$1 = \mathbb{P}([X_{T_2} = N - 1]) + \mathbb{P}([X_{T_2} = N]) = 2\mathbb{P}([X_{T_2} = N])$$

Finalement : $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N - 1]) = \frac{1}{2}$
 et $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N]) = \frac{1}{2}$.

□

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 $X_{T_3}(\Omega)$.

Démonstration.

Pour que la carte C_N passe de la position $N - 2$ à la position $N - 3$, il faut effectuer une insertion à la position $N - 2$, $N - 1$ ou N . Or X_{T_3} correspond à la position de l'insertion au moment précis où la carte C_N passe de $N - 2$ à $N - 3$.

On en déduit : $X_{T_3}(\Omega) = \{N - 2, N - 1, N\}$.

□

b) Avec le même raisonnement qu'en question 5, déterminer :

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

Démonstration.

- Rappelons tout d'abord : $[X_{T_1} = N] = \Omega$. Donc :

$$\begin{aligned} [X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N] &= [X_{T_2} = N] \cap [X_{T_3} = N] \\ [X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N - 1] &= [X_{T_2} = N] \cap [X_{T_3} = N - 1] \\ [X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N - 2] &= [X_{T_2} = N] \cap [X_{T_3} = N - 2] \end{aligned}$$

- La famille $([X_{T_3} = N - 2], [X_{T_3} = N - 1], [X_{T_3} = N])$ forme un système complet d'événements. Ainsi, par formule des probabilités totales :

$$\begin{aligned} \frac{1}{2} = \mathbb{P}([X_{T_2} = N]) &= \mathbb{P}([X_{T_3} = N - 2] \cap [X_{T_2} = N]) \\ &+ \mathbb{P}([X_{T_3} = N - 1] \cap [X_{T_2} = N]) \\ &+ \mathbb{P}([X_{T_3} = N] \cap [X_{T_2} = N]) \end{aligned}$$

Or les insertions s'effectuent de manière uniforme, donc :

$$\mathbb{P}([X_{T_3} = N - 2] \cap [X_{T_2} = N]) = \mathbb{P}([X_{T_3} = N - 1] \cap [X_{T_2} = N]) = \mathbb{P}([X_{T_3} = N] \cap [X_{T_2} = N])$$

On obtient alors :

$$\frac{1}{2} = 3\mathbb{P}([X_{T_3} = N] \cap [X_{T_2} = N])$$

Finalement : $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N]) = \frac{1}{6}$
 et $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N - 1]) = \frac{1}{6}$.

□

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 !

Démonstration.

- D'après la question **5.c)**, la v.a.r. X_{T_2} suit la loi uniforme sur $\{N - 1, N\}$.
Donc, à l'instant T_2 , la carte C_N est en position $N - 2$ et la configuration des cartes sous elle (positions $N - 1$ et N) est uniforme.
- D'après la question **6.b)** et la formule des probabilités totales appliquée sur le système complet d'événements associé à X_{T_2} (*i.e.* la famille $([X_{T_2} = N - 1], [X_{T_2} = N])$), la v.a.r. X_{T_3} suit la loi uniforme sur $\{N - 2, N - 1, N\}$.
Donc, à l'instant T_3 , la carte C_N est en position $N - 3$ et la configuration des cartes sous elle (positions $N - 2, N - 1$ et N) est uniforme.
- On démontrerait de même par récurrence que la v.a.r. X_{T_j} suit la loi uniforme sur $\llbracket N - j + 1, N \rrbracket$.
Donc, à l'instant T_j , la carte C_N est en position $N - j$ et la configuration des cartes sous elle (positions $N - j + 1, \dots, N$) est uniforme.
- Ainsi, la v.a.r. $X_{T_{N-1}}$ suit la loi uniforme sur $\llbracket 2, N \rrbracket$.
Donc, à l'instant T_{N-1} , la carte C_N est en position 1 et la configuration des cartes sous elle (positions 2, \dots , N) est uniforme.
- À l'instant T , on prend la carte du dessus, c'est-à-dire C_N , et on la place au hasard dans le paquet. Il est donc légitime de penser qu'alors toutes les configurations du jeu sont équiprobables.

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

Commentaire

Détaillons l'obtention du résultat $X_{T_3} \hookrightarrow \mathcal{U}(\llbracket N - 2, N \rrbracket)$.

- Tout d'abord, d'après la question **6.a)** : $X_{T_3}(\Omega) = \llbracket N - 2, N \rrbracket$.
- Soit $k \in \llbracket N - 2, N \rrbracket$.

La famille $([X_{T_2} = N - 1], [X_{T_2} = N])$ forme un système complet d'événements.

Ainsi, par formule des probabilités totales :

$$\begin{aligned} \mathbb{P}([X_{T_3} = k]) &= \mathbb{P}([X_{T_2} = N - 1] \cap [X_{T_3} = k]) + \mathbb{P}([X_{T_2} = N] \cap [X_{T_3} = k]) \\ &= \frac{1}{6} + \frac{1}{6} \end{aligned}$$

(d'après la question 6.b)

Finalement, on a bien : $X_{T_3} \hookrightarrow \mathcal{U}(\llbracket N - 2, N \rrbracket)$. □

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

Démonstration.

- Tout d'abord :

$$\begin{aligned} T &= T_{N-1} + 1 && \text{(par définition)} \\ &= \sum_{i=1}^{N-1} \Delta_i + 1 && \text{(d'après la question 1)} \end{aligned}$$

Or, d'après les questions 2 et 3.a), pour tout $i \in \llbracket 1, N-1 \rrbracket$, $\Delta_i \hookrightarrow \mathcal{G}\left(\frac{i}{N}\right)$. Ces v.a.r. admettent donc une variance (donc une espérance).

On en déduit que la v.a.r. T admet une variance (donc une espérance) en tant que somme de v.a.r. qui en admettent une.

Commentaire

On notera que l'égalité « $T = T_{N-1} + 1$ » est une égalité entre variables aléatoires. La notation « 1 » désigne donc la **variable aléatoire** constante égale à 1.

- On obtient :

$$\begin{aligned} \mathbb{E}(T) &= \mathbb{E}\left(\sum_{i=1}^{N-1} \Delta_i + 1\right) \\ &= \sum_{i=1}^{N-1} \mathbb{E}(\Delta_i) + 1 && \text{(par linéarité de l'espérance)} \\ &= \sum_{i=1}^{N-1} \frac{N}{i} + 1 && \text{(d'après les questions 2 et 3.b)} \\ &= N \sum_{i=1}^{N-1} \frac{1}{i} + \frac{N}{N} \\ &= N \left(\sum_{i=1}^{N-1} \frac{1}{i} + \frac{1}{N} \right) \\ &= N \sum_{i=1}^N \frac{1}{i} = N H_N \end{aligned}$$

$$\mathbb{E}(T) = N H_N$$

• Enfin :

$$\begin{aligned}
 \mathbb{V}(T) &= \mathbb{V}\left(\sum_{i=1}^{N-1} \Delta_i + 1\right) \\
 &= \mathbb{V}\left(\sum_{i=1}^{N-1} \Delta_i\right) && \text{(par propriété de la variance)} \\
 &= \sum_{i=1}^{N-1} \mathbb{V}(\Delta_i) && \text{(car les v.a.r. } \Delta_1, \dots, \Delta_{N-1} \text{ sont indépendantes)} \\
 &= \mathbb{V}(\Delta_1) + \sum_{i=2}^{N-1} \mathbb{V}(\Delta_i)
 \end{aligned}$$

Or :

$$\mathbb{V}(\Delta_1) = \frac{1 - \frac{1}{N}}{\left(\frac{1}{N}\right)^2} = N^2 \frac{N-1}{N} = N \frac{N-1}{1}$$

Ainsi, d'après la question **3.b)** :

$$\begin{aligned}
 \mathbb{V}(T) &= N \frac{N-1}{1^2} + \sum_{i=2}^{N-1} N \frac{N-i}{i^2} \\
 &= \sum_{i=1}^{N-1} N \frac{N-i}{i^2} = N^2 \sum_{i=1}^{N-1} \frac{1}{i^2} - N \sum_{i=1}^{N-1} \frac{1}{i}
 \end{aligned}$$

Or :

$$\begin{aligned}
 N^2 \sum_{i=1}^N \frac{1}{i^2} - N H_N &= N^2 \sum_{i=1}^N \frac{1}{i^2} - N \sum_{i=1}^N \frac{1}{i} \\
 &= N^2 \sum_{i=1}^{N-1} \frac{1}{i^2} + \cancel{N^2 \frac{1}{N^2}} - N \sum_{i=1}^{N-1} \frac{1}{i} - \cancel{N \frac{1}{N}} \\
 &= N^2 \sum_{i=1}^{N-1} \frac{1}{i^2} - N \sum_{i=1}^{N-1} \frac{1}{i} \\
 &= \mathbb{V}(T)
 \end{aligned}$$

$$\mathbb{V}(T) = N^2 \sum_{i=1}^N \frac{1}{i^2} - N H_N$$

□

9. Étude de la suite (u_n)

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

Démonstration.

Soit $k \geq 1$.

• Soit $t \in [k, k+1]$.

Comme $k \leq t \leq k+1$

alors $\frac{1}{k} \geq \frac{1}{t} \geq \frac{1}{k+1}$ *(par décroissance de la fonction $t \mapsto \frac{1}{t}$ sur $]0, +\infty[$)*

- La fonction $t \mapsto \frac{1}{t}$ est continue sur le segment $[k, k + 1]$.

On en déduit que l'intégrale $\int_k^{k+1} \frac{1}{t} dt$ est bien définie.

De plus, par croissance de l'intégration, les bornes étant rangées dans l'ordre croissant ($k \leq k + 1$) :

$$\int_k^{k+1} \frac{1}{k+1} dt \leq \int_k^{k+1} \frac{1}{t} dt \leq \int_k^{k+1} \frac{1}{k} dt$$

$$\parallel \qquad \qquad \qquad \parallel$$

$$(k+1-k) \frac{1}{(k+1)} \qquad \qquad \qquad (k+1-k) \frac{1}{k}$$

$$\forall k \geq 1, \frac{1}{k+1} \leq \int_k^{k+1} \frac{1}{t} dt \leq \frac{1}{k}$$

□

b) En déduire successivement :

- (i) la décroissance de la suite (u_n) ,
- (ii) l'encadrement : $\forall n \in \mathbb{N}^*, \ln(n+1) \leq H_n \leq \ln(n) + 1$.

Démonstration.

(i) Soit $n \in \mathbb{N}^*$.

$$\begin{aligned} u_{n+1} - u_n &= H_{n+1} - \ln(n+1) - (H_n - \ln(n)) \\ &= \frac{1}{n+1} - \ln(n+1) + \ln(n) \end{aligned}$$

Appliquons la partie gauche de l'encadrement de la question précédente pour $k = n$:

$$\frac{1}{n+1} \leq \int_n^{n+1} \frac{1}{t} dt$$

$$\parallel$$

$$[\ln(|t|)]_n^{n+1} = \ln(n+1) - \ln(n)$$

On en déduit :

$$u_{n+1} - u_n = \frac{1}{n+1} - \ln(n+1) + \ln(n) \leq 0$$

La suite $(u_n)_{n \in \mathbb{N}^*}$ est donc décroissante.

(ii) Soit $n \in \mathbb{N}^*$.

Reprenons l'encadrement de la question précédente.

× On obtient, par sommation des inégalités de gauche pour k variant de 1 à $n - 1$:

$$\sum_{k=1}^{n-1} \frac{1}{k+1} \leq \int_1^n \frac{1}{t} dt$$

$$\parallel \qquad \qquad \qquad \parallel$$

$$H_n - 1 = \sum_{k=2}^n \frac{1}{k} \qquad [\ln(|t|)]_1^n = \ln(n) - \ln(1)$$

En ajoutant 1 de part et d'autre de l'inégalité : $H_n \leq \ln(n) + 1$.

× Par sommation des inégalités de droite pour k variant de 1 à n :

$$\int_1^{n+1} \frac{1}{t} dt \leq \sum_{k=1}^n \frac{1}{k}$$

$$\parallel \qquad \parallel$$

$$\ln(n+1) - \ln(1) \qquad H_n$$

Ainsi : $\ln(n+1) \leq H_n$. □

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

Démonstration.

• Soit $n \in \mathbb{N}^*$. d'après la question précédente :

$$\ln(n+1) \leq H_n$$

Donc :

$$\ln(n+1) - \ln(n) \leq H_n - \ln(n)$$

$$\forall \qquad \parallel$$

$$0 \qquad u_n$$

(l'inégalité $\ln(n+1) - \ln(n) \geq 0$ est obtenue par croissance de la fonction \ln sur $]0, +\infty[$)

On en déduit que la suite (u_n) est minorée par 0.

- La suite (u_n) est donc :
 - × décroissante (d'après la question précédente),
 - × minorée par 0.

On en déduit que la suite (u_n) converge vers un réel γ vérifiant $\gamma \geq 0$.

- Montrons enfin : $\gamma \leq 1$.
Soit $n \in \mathbb{N}^*$. Toujours d'après la question précédente :

$$H_n \leq \ln(n) + 1$$

On en déduit : $u_n = H_n - \ln(n) \leq 1$.

Par passage à la limite : $\gamma \leq 1$. □

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

Démonstration.

- Soit $n \geq 2$. D'après la question **9.b)** :

$$\ln(n+1) \leq H_n \leq \ln(n) + 1$$

Donc, comme $\ln(n) > 0$:

$$\frac{\ln(n+1)}{\ln(n)} \leq \frac{H_n}{\ln(n)} \leq 1 + \frac{1}{\ln(n)}$$

Or :

$$\times \lim_{n \rightarrow +\infty} \frac{\ln(n+1)}{\ln(n)} = 1. \text{ En effet :}$$

$$\frac{\ln(n+1)}{\ln(n)} = \frac{\ln\left(n\left(1 + \frac{1}{n}\right)\right)}{\ln(n)} = \frac{\ln(n) + \ln\left(1 + \frac{1}{n}\right)}{\ln(n)} = 1 + \frac{\ln\left(1 + \frac{1}{n}\right)}{\ln(n)}$$

$$\times \lim_{n \rightarrow +\infty} 1 + \frac{1}{\ln(n)} = 1$$

Par théorème d'encadrement : $\lim_{n \rightarrow +\infty} \frac{H_n}{\ln(n)} = 1.$

$$\text{On en déduit : } H_n \underset{n \rightarrow +\infty}{\sim} \ln(n). \text{ Donc } \mathbb{E}(T) = N H_N \underset{N \rightarrow +\infty}{\sim} N \ln(N).$$

- On a les équivalences suivantes :

$$\mathbb{E}(T) = N \ln(N) + N \gamma + \underset{N \rightarrow +\infty}{o}(N) \Leftrightarrow N H_N = N \ln(N) + N \gamma + \underset{N \rightarrow +\infty}{o}(N)$$

$$\Leftrightarrow N (H_N - \ln(N) - \gamma) = \underset{N \rightarrow +\infty}{o}(N)$$

$$\Leftrightarrow \lim_{N \rightarrow +\infty} \frac{\cancel{N} (H_N - \ln(N) - \gamma)}{\cancel{N}} = 0$$

$$\Leftrightarrow \lim_{N \rightarrow +\infty} H_N - \ln(N) = \gamma$$

$$\Leftrightarrow \lim_{N \rightarrow +\infty} u_N = \gamma$$

Cette dernière égalité est vérifiée d'après la question **9.c**.)

$$\text{Ainsi : } \mathbb{E}(T) = N \ln(N) + N \gamma + \underset{N \rightarrow +\infty}{o}(N).$$

□

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 :

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

Démonstration.

- Tout d'abord :

$$\frac{\mathbb{V}(T)}{N^2} = \sum_{k=1}^N \frac{1}{k^2} - \frac{H_N}{N}$$

Or :

× La suite $\left(\sum_{k=1}^n \frac{1}{k^2}\right)_{n \in \mathbb{N}^*}$ est la suite des sommes partielles de la série de Riemann d'exposant 2 (avec $2 > 1$). Donc elle converge.

× De plus, d'après la question précédente : $\frac{H_N}{N} \underset{N \rightarrow +\infty}{\sim} \frac{\ln(N)}{N} \xrightarrow{N \rightarrow +\infty} 0$

$$\text{On en déduit que la suite } \left(\frac{\mathbb{V}(T)}{N^2}\right)_{N \in \mathbb{N}^*} \text{ est convergente.}$$

- D'après le point précédent :

$$\lim_{N \rightarrow +\infty} \frac{\mathbb{V}(T)}{N^2} = \sum_{k=1}^{+\infty} \frac{1}{k^2}$$

On note alors : $\alpha = \sum_{k=1}^{+\infty} \frac{1}{k^2}$. Ainsi : $\lim_{N \rightarrow +\infty} \frac{\mathbb{V}(T)}{\alpha N^2} = 1$

$$\text{On en déduit : } \mathbb{V}(T) \underset{N \rightarrow +\infty}{\sim} \alpha N^2.$$

- Ensuite, on sait : $\frac{H_N}{N} \geq 0$. Donc :

$$\sum_{k=1}^N \frac{1}{k^2} - \frac{H_N}{N} \leq \sum_{k=1}^N \frac{1}{k^2} \leq \sum_{k=1}^{+\infty} \frac{1}{k^2} = \alpha$$

||

$$\frac{\mathbb{V}(T)}{N^2}$$

$$\text{Ainsi, comme } N^2 \geq 0 : \mathbb{V}(T) \leq \alpha N^2.$$

□

11. Écart à la moyenne

On admet 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 :

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

Démonstration.

- Soit $\omega \in \Omega$.

$$\begin{aligned} |T(\omega) - N \ln(N)| &= |T(\omega) - \mathbb{E}(T) + \mathbb{E}(T) - N \ln(N)| \\ &\leq |T(\omega) - \mathbb{E}(T)| + |\mathbb{E}(T) - N \ln(N)| \quad (\text{par inégalité triangulaire}) \end{aligned}$$

- De plus :

$$|\mathbb{E}(T) - N \ln(N)| = |N H_N - N \ln(N)| = N |H_N - \ln(N)|$$

Or, d'après la question **9.b)(ii)** : $0 \leq H_N - \ln(N) \leq 1$. Ainsi :

$$|T(\omega) - N \ln(N)| \leq |T(\omega) - \mathbb{E}(T)| + N \times 1$$

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

- Montrons $[|T - N \ln(N)| \geq cN] \subset [|T - \mathbb{E}(T)| \geq N(c-1)]$.
Soit $\omega \in \Omega$.
Supposons $\omega \in [|T - N \ln(N)| \geq cN]$, autrement dit : $|T(\omega) - N \ln(N)| \geq cN$.
D'après le point précédent :

$$\begin{aligned} cN &\leq |T(\omega) - N \ln(N)| \\ &\leq |T(\omega) - \mathbb{E}(T)| + N \end{aligned}$$

On obtient : $|T(\omega) - \mathbb{E}(T)| \geq cN - N = N(c-1)$.

Autrement dit : $\omega \in [|T - \mathbb{E}(T)| \geq N(c-1)]$.

$D'o\grave{u} : [|T - N \ln(N)| \geq cN] \subset [|T - \mathbb{E}(T)| \geq N(c-1)].$

□

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 $\lim_{c \rightarrow +\infty} \mathbb{P}(|T - N \ln(N)| \geq cN)$?

Démonstration.

- Tout d'abord :

comme $[|T - N \ln(N)| \geq cN] \subset [|T - \mathbb{E}(T)| \geq N(c-1)]$

alors $\mathbb{P}(|T - N \ln(N)| \geq cN) \leq \mathbb{P}(|T - \mathbb{E}(T)| \geq N(c-1))$

Ainsi :

$$\begin{aligned} \mathbb{P}(|T - N \ln(N)| \geq cN) &\leq \mathbb{P}(|T - \mathbb{E}(T)| \geq N(c-1)) \\ &\leq \frac{\mathbb{V}(T)}{(N(c-1))^2} && \text{(d'après l'inégalité de Bienaymé-Chebychev)} \\ &\leq \frac{\alpha \cancel{N^2}}{\cancel{N^2} (c-1)^2} = \frac{\alpha}{(c-1)^2} && \text{(d'après 10.b)} \end{aligned}$$

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

- On a donc :

$$0 \leq \mathbb{P}(|T - N \ln(N)| \geq cN) \leq \frac{\alpha}{(c-1)^2}$$

Or :

× $\lim_{c \rightarrow +\infty} 0 = 0$,

× $\lim_{c \rightarrow +\infty} \frac{\alpha}{(c-1)^2} = 0$

$Par\ théorème\ d'encadrement : \lim_{c \rightarrow +\infty} \mathbb{P}(|T - N \ln(N)| \geq cN) = 0.$

□

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

Démonstration.

• D'après la question 11.b) :

$$\forall c > 1, \quad 0 \leq \mathbb{P}(|T - N \ln(N)| \geq cN) \leq \frac{\alpha}{(c-1)^2}$$

• On souhaite donc appliquer cet encadrement avec $c = \varepsilon \ln(N)$.
On commence alors par vérifier $\varepsilon \ln(N) > 1$.

$$\begin{aligned} \varepsilon \ln(N) > 1 &\Leftrightarrow \ln(N) > \frac{1}{\varepsilon} \\ &\Leftrightarrow N > e^{\frac{1}{\varepsilon}} \quad (\text{par stricte croissance de la fonction exp sur } \mathbb{R}) \end{aligned}$$

• On en déduit :

$$\forall N > e^{\frac{1}{\varepsilon}}, \quad 0 \leq \mathbb{P}(|T - N \ln(N)| \geq \varepsilon N \ln(N)) \leq \frac{\alpha}{(\varepsilon \ln(N) - 1)^2}$$

Or :

$$\times \lim_{N \rightarrow +\infty} 0 = 0,$$

$$\times \lim_{N \rightarrow +\infty} \frac{\alpha}{(\varepsilon \ln(N) - 1)^2} = 0.$$

Par théorème d'encadrement : $\lim_{N \rightarrow +\infty} \mathbb{P}(|T - N \ln(N)| \geq \varepsilon N \ln(N)) = 0.$

□

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.

Démonstration.

Initialement, les cartes sont rangées dans l'ordre : du numéro 1 à 32.

Le tableau correspondant à la configuration initiale du paquet de cartes est donc :
`Jeu = np.arange(1, 33).`

□

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
    
```

Démonstration.

```

1 def Insertion(Jeu):
2     k = rd.randint(1,33) # 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] = Jeu[i+1]
8             J[k-1] = cartedessus
9     return J
    
```

- La position où l'on insère la carte du dessus du paquet est uniforme sur les 32 positions possibles. On stocke cette valeur dans une variable `k` avec la commande suivante :

```

2     k = rd.randint(1,33)
    
```

- La configuration initiale du paquet est stockée dans le paramètre `Jeu`. On initialise donc la variable `J` de la façon ci-dessous :

```

3     J = Jeu
    
```

- On stocke ensuite dans la variable `cartedessus` le numéro de la carte du dessus du paquet (oh surprise...), puisqu'il s'agit de la carte qui va être insérée dans le paquet. Le numéro de la première carte du paquet est stocké dans la première coordonnée du vecteur `Jeu`.

```

4     cartedessus = Jeu[0]
    
```

- Deux cas se présentent alors.
 - × Si `k = 1`, c'est-à-dire si l'insertion s'effectue en position 1, alors la configuration du paquet reste identique. On ne modifie donc pas le vecteur `J`.
 - × Si `k > 1`, c'est-à-dire si l'insertion s'effectue en position `k`, on code cette condition par la structure conditionnelle suivante :

```

5     if k > 1 :
    
```

On souhaite alors :

- 1) remonter les cartes des positions 2 à $k - 1$ d'un rang, c'est-à-dire : à la position i , on affecte la carte qui était en position $i + 1$.

```

6         for i in range(k-1):
7             J[i] = Jeu[i+1]

```

- 2) positionner la carte du dessus en position k .

```

9         J[k] = cartedessus

```

Commentaire

Afin de permettre une bonne compréhension des mécanismes en jeu, on a détaillé la réponse à cette question. Cependant, compléter correctement le programme **Python** démontre la bonne compréhension de la simulation demandée et permet certainement d'obtenir la majorité des points alloués à cette question. On procèdera de même dans les autres questions **Python**. □

- 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

```

Démonstration.

• **Début du programme**

On commence par initialiser un vecteur `Jeu` et la variable `n` :

- × le vecteur `Jeu` du vecteur correspondant à la configuration initiale du paquet de cartes : elles sont rangées de C_1 à C_{32} ,
- × la variable `n` correspond au numéro de l'insertion. On démarre à l'instant 0 : aucune insertion n'a encore été effectuée.

```

2     Jeu = np.arange(1,33)
3     n = 0

```

• **Structure itérative**

Les lignes 4 à 6 suivantes consistent à effectuer une succession d'insertions jusqu'à ce que la carte C_N atteigne le dessus du paquet, c'est-à-dire jusqu'à ce que la première carte du paquet soit C_{32} . Autrement dit, on fait cette succession d'insertions tant que la première carte du paquet n'est pas C_{32} . Pour cela on met en place une structure itérative (boucle **while**).

```

4     while Jeu[0] != 32:

```

On effectue ensuite les insertions. Pour cela, on utilise la fonction `Insertion` définie en question précédente.

```

5         Jeu = Insertion(Jeu)

```

On met ensuite à jour la variable n en conséquence : on ajoute 1 pour signaler qu'une nouvelle insertion a eu lieu.

```
6         n = n + 1
```

• **Fin du programme**

À l'issue de cette boucle **while**, la variable n contient le nombre d'insertions nécessaires pour que la carte C_{32} soit en position 1, c'est-à-dire la variable n contient une réalisation de la variable aléatoire T_{N-1} .

On effectue ensuite la commande :

```
8         return n + 1
```

On renvoie donc une réalisation de la v.a.r. $T_{N-1} + 1$, c'est-à-dire de la v.a.r. T .

La fonction **Simu** renvoie une réalisation de la v.a.r. T . □

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.

Démonstration.

```
1 N = 100 # Nombre d'expériences
2 S = 0
3 for k in range(N):
4     S = S + Simu()
5 print(S/N)
```

Détaillons l'obtention de ce programme.

Pour effectuer la moyenne des valeurs prises par la v.a.r. T sur 100 expériences, il faut commencer par obtenir 100 réalisations de cette v.a.r. .

Or le programme **Simu** de la question précédente fournit justement une réalisation de la v.a.r. T à chaque exécution. On procède donc comme suit :

1) On définit une variable qui contient le nombre d'expériences à simuler.

```
1 N = 100 # Nombre d'expériences
```

2) On initialise ensuite une variable S à 0. Cette variable contiendra en fin de boucle **for** la somme de 100 réalisations de T .

```
2 S = 0
```

3) On met ensuite en place une structure itérative (boucle **for**) pour calculer cette somme. Pour cela, on utilise donc la fonction **Simu** de la question précédente.

```
3 for k in range(N):
4     S = S + Simu()
```

4) On calcule ensuite la moyenne de ces 100 réalisations, puis on affiche cette valeur.

```
5 print(S/N)
```

□

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 :

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

Commentaire

On peut noter que ce maximum existe bien puisqu'on considère le maximum sur $A \subset \mathcal{S}_N$, c'est-à-dire sur l'ensemble des parties de \mathcal{S}_N , qui est un ensemble de cardinal fini.

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

Démonstration.

- D'après la question 7, à partir de l'instant T , toutes les configurations sont équiprobables. Ainsi, si l'événement $[T \leq n]$ est réalisé, alors, à l'instant n , on est certain que toutes les configurations sont équiprobables. On en déduit :

$$\mathbb{P}_{[T \leq n]}(E_n) = \frac{\text{Card}(A)}{N!} = \pi(A)$$

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

- Par formule des probabilités composées :

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

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

□

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

Démonstration.

Comme $E_n \cap [T > n] \subset [T > n]$

alors $\mathbb{P}(E_n \cap [T > n]) \leq \mathbb{P}([T > n])$

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

□

c) Montrer :

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

Démonstration.

- Par définition de μ_n , $\mu_n(A)$ est la probabilité qu'à l'instant n , le tas de cartes soit dans une configuration qui appartient à la partie A . Autrement dit :

$$\mu_n(A) = \mathbb{P}(E_n)$$

- La famille $([T \leq n], [T > n])$ forme un système complet d'événements. Ainsi, par formule des probabilités totales :

$$\begin{aligned} \mathbb{P}(E_n) &= \mathbb{P}(E_n \cap [T \leq n]) + \mathbb{P}(E_n \cap [T > n]) \\ &= \pi(A) \mathbb{P}([T \leq n]) + \mathbb{P}(E_n \cap [T > n]) && \text{(d'après 14.a)} \\ &\leq \pi(A) \mathbb{P}([T \leq n]) + \mathbb{P}([T > n]) && \text{(d'après 14.b)} \\ &\leq \pi(A) + \mathbb{P}([T > n]) && \text{(car } \mathbb{P}([T \leq n]) \leq 1) \end{aligned}$$

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

□

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

Démonstration.

μ_n et π sont des probabilités. Donc :

$$\mu_n(\bar{A}) - \pi(\bar{A}) = (1 - \mu_n(A)) - (1 - \pi(A)) = \pi(A) - \mu_n(A)$$

$$\boxed{\mu_n(\bar{A}) - \pi(\bar{A}) = \pi(A) - \mu_n(A)}$$

□

b) Dédurre des questions précédentes la majoration :

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

Démonstration.

- Tout d'abord, d'après la question 14.c) :

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

- On applique ensuite le même résultat avec \bar{A} :

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

$$\text{donc } \pi(A) - \mu_n(A) \leq \mathbb{P}([T > n]) \quad \text{(d'après 15.a)}$$

$$\text{d'où } \mu_n(A) - \pi(A) \geq -\mathbb{P}([T > n])$$

- Finalement :

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

$$\boxed{\text{Ainsi : } |\mu_n(A) - \pi(A)| \leq \mathbb{P}([T > n]).}$$

□

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

Démonstration.

- Soit $n \in \mathbb{N}^*$.

Comme l'ensemble des événements A vérifiant $A \subset \mathcal{S}_N$ est de cardinal fini (car $\mathcal{P}(\mathcal{S}_N)$ est de cardinal fini), il existe $A_0 \subset \mathcal{S}_N$ tel que :

$$d(\mu_n, \pi) = \max_{A \subset \mathcal{S}_N} (|\mu_n(A) - \pi(A)|) = |\mu_n(A_0) - \pi(A_0)|$$

- Or, d'après la question précédente :

$$\forall A \subset \mathcal{S}_N, 0 \leq |\mu_n(A) - \pi(A)| \leq \mathbb{P}([T > n])$$

En appliquant cet encadrement à l'événement A_0 (puisqu'il est vérifié pour tout événement $A \subset \mathcal{S}_N$), on obtient :

$$0 \leq |\mu_n(A_0) - \pi(A_0)| \leq \mathbb{P}([T > n])$$

On en déduit : $\forall n \in \mathbb{N}^*$, $0 \leq d(\mu_n, \pi) \leq \mathbb{P}([T > n])$.

- Soit $n \in \mathbb{N}^*$.

Tout d'abord : $[T > n] = \overline{[T \leq n]}$. Donc :

$$\mathbb{P}([T > n]) = \mathbb{P}(\overline{[T \leq n]}) = 1 - \mathbb{P}([T \leq n]) = 1 - F_T(n)$$

en notant F_T la fonction de répartition de T .

- Or, comme F_T est une fonction de répartition : $\lim_{n \rightarrow +\infty} F_T(n) = 1$.

Ainsi : $\lim_{n \rightarrow +\infty} \mathbb{P}([T > n]) = 1 - 1 = 0$.

- D'après le point précédent :

$$\forall n \in \mathbb{N}^*, 0 \leq d(\mu_n, \pi) \leq \mathbb{P}([T > n])$$

Or :

$$\times \lim_{n \rightarrow +\infty} 0 = 0,$$

$$\times \lim_{n \rightarrow +\infty} \mathbb{P}([T > n]) = 0.$$

Par théorème d'encadrement : $\lim_{n \rightarrow +\infty} d(\mu_n, \pi) = 0$.

□

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 .

Démonstration.

Déterminons d'abord $S_1(\Omega)$.

La v.a.r. S_1 correspond au nombre aléatoire de jours que doit attendre le collectionneur pour obtenir son premier timbre. Il obtient ce dernier forcément dès le premier jour.

Ainsi : $S_1(\Omega) = \{1\}$.

On en déduit que la v.a.r. S_1 est la variable aléatoire constante égale à 1.

□

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

Démonstration.

Soit $k \in \llbracket 2, N \rrbracket$.

- Tout d'abord : $S_k(\Omega) = \mathbb{N}^*$.

En effet, l'obtention du $k^{\text{ème}}$ timbre distinct quand on a déjà $k-1$ timbres distincts peut s'effectuer dès le 1^{er} jour, ou aux jours suivants. Elle peut également ne jamais se produire.

- Soit $i \in \mathbb{N}^*$.

L'événement $[S_k = i]$ est réalisé si et seulement si le nombre de jours entre la possession de $k-1$ timbres distincts et k distincts est égal à i , c'est-à-dire si, à partir du moment où l'on a reçu le $(k-1)^{\text{ème}}$ timbre, on reçoit pendant $i-1$ jours des timbres que l'on avait déjà puis un nouveau timbre le $i^{\text{ème}}$ jour.

Soit $j \in \mathbb{N}^*$. On introduit les événements C_j^{k-1} suivants :

$$C_j^{k-1} = \text{« recevoir le } j^{\text{ème}} \text{ jour l'un des } k-1 \text{ timbres déjà possédés »}$$

On en déduit :

$$\begin{aligned} & [S_1 + \dots + S_{k-1} = j] \cap [S_k = i] \\ &= [S_1 + \dots + S_{k-1} = j] \cap C_{j+1}^{k-1} \cap C_{j+2}^{k-1} \cap \dots \cap C_{i+j-1}^{k-1} \cap \overline{C_{i+j}^{k-1}} \end{aligned}$$

On note que la v.a.r. $S_1 + \dots + S_{k-1}$ correspond au nombre de jours pour que le collectionneurs obtienne $k-1$ timbres distincts.

La famille $([S_1 + \dots + S_{k-1} = j])_{j \in \mathbb{N}^*}$ est un système complet d'événements, car la variable aléatoire $S_1 + \dots + S_{k-1}$ est à valeurs entières non nulles.

Ainsi, par formule des probabilités totales :

$$\begin{aligned} & \mathbb{P}([S_k = i]) \\ &= \sum_{j=1}^{+\infty} \mathbb{P}([S_1 + \dots + S_{k-1} = j] \cap [S_k = i]) \\ &= \sum_{j=1}^{+\infty} \mathbb{P} \left([S_1 + \dots + S_{k-1} = j] \cap \left(\bigcap_{\ell=j+1}^{i+j-1} C_{\ell}^{k-1} \right) \cap \overline{C_{i+j}^{k-1}} \right) \\ &= \sum_{j=1}^{+\infty} \mathbb{P}([S_1 + \dots + S_{k-1} = j]) \mathbb{P} \left(\left(\bigcap_{\ell=j+1}^{i+j-1} C_{\ell}^{k-1} \right) \cap \overline{C_{i+j}^{k-1}} \right) \end{aligned}$$

(par indépendance de $[S_1 + \dots + S_{k-1} = j]$ et $\bigcap_{\ell=j+1}^{i+j-1} \overline{C_{\ell}^{k-1}} \cap C_{i+j}^{k-1}$)

- × Or chaque jour le timbre servant à affranchir la lettre est choisi uniformément parmi les N timbres possibles. Donc : $\forall \ell \in \mathbb{N}^*, \mathbb{P}(C_\ell^{k-1}) = \frac{k-1}{N}$.

Ainsi par indépendance des arrivages chaque jour :

$$\mathbb{P}\left(\left(\bigcap_{\ell=j+1}^{i+j-1} C_\ell^{k-1}\right) \cap \overline{C_{i+j}^{k-1}}\right) = \left(\prod_{\ell=j+1}^{i+j-1} \mathbb{P}(C_\ell^{k-1})\right) \times \mathbb{P}(\overline{C_{i+j}^{k-1}}) = \left(\frac{k-1}{N}\right)^{i-1} \times \frac{N-(k-1)}{N}$$

- × On en déduit :

$$\begin{aligned} & \mathbb{P}([S_k = i]) \\ &= \sum_{j=1}^{+\infty} \mathbb{P}([S_1 + \dots + S_{k-1} = j]) \mathbb{P}\left(\left(\bigcap_{\ell=j+1}^{i+j-1} C_\ell^{k-1}\right) \cap \overline{C_{i+j}^{k-1}}\right) \\ &= \sum_{j=1}^{+\infty} \mathbb{P}([S_1 + \dots + S_{k-1} = j]) \left(\frac{k-1}{N}\right)^{i-1} \frac{N-(k-1)}{N} \\ &= \left(\frac{k-1}{N}\right)^{i-1} \frac{N-(k-1)}{N} \sum_{j=1}^{+\infty} \mathbb{P}([S_1 + \dots + S_{k-1} = j]) \end{aligned}$$

- × Comme $([S_1 + \dots + S_{k-1} = j])_{j \in \mathbb{N}^*}$ est un système complet d'événements :

$$\sum_{j=1}^{+\infty} \mathbb{P}([S_1 + \dots + S_{k-1} = j]) = 1$$

Donc :

$$\mathbb{P}([S_k = i]) = \left(\frac{k-1}{N}\right)^{i-1} \frac{N-(k-1)}{N}$$

On en déduit : $\forall k \in \llbracket 2, N \rrbracket, S_k \hookrightarrow \mathcal{G}\left(\frac{N-k+1}{N}\right)$.

Commentaire

- Afin de permettre une bonne compréhension des mécanismes en jeu, on a rédigé rigoureusement la réponse à cette question. Cependant, expliquer en quelques phrases pourquoi la v.a.r. S_k suit la loi géométrique de paramètre $\frac{N-k+1}{N}$ permettrait très certainement d'obtenir tous les points alloués à cette question.

- Détaillons la justification de l'indépendance des événements $[S_1 + \dots + S_{k-1} = j]$ et $\bigcap_{\ell=j+1}^{i+j-1} C_\ell^{k-1} \cap \overline{C_{i+j}^{k-1}}$.

- × D'une part, l'événement $[S_1 + \dots + S_{k-1} = j]$ est réalisé si et seulement si on obtient $k-1$ timbres distincts à l'instant j . Cet événement dépend donc uniquement des j premiers arrivages..

- × D'autre part, l'expression de l'événement $\bigcap_{\ell=j+1}^{i+j-1} C_\ell^{k-1} \cap \overline{C_{i+j}^{k-1}}$ ne fait intervenir que les arrivages $j+1$ à $i+j$.

Or les arrivages sont mutuellement indépendants.

Ainsi, les événements $[S_1 + \dots + S_{k-1} = j]$ et $\bigcap_{\ell=j+1}^{i+j-1} C_\ell^{k-1} \cap \overline{C_{i+j}^{k-1}}$ sont bien indépendants.

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

Démonstration.

- D'une part, d'après la question **17** :

$$\begin{aligned} S &= S_1 + S_2 + \cdots + S_{N-1} + S_N \\ &= 1 + S_2 + \cdots + S_{N-1} + S_N \end{aligned}$$

Donc $S - 1$ est une somme de v.a.r. indépendantes suivant les lois :

$$\mathcal{G}\left(\frac{N-2+1}{N}\right), \mathcal{G}\left(\frac{N-3+1}{N}\right), \dots, \mathcal{G}\left(\frac{N-(N-1)+1}{N}\right), \mathcal{G}\left(\frac{N-N+1}{N}\right)$$

c'est-à-dire :

$$\mathcal{G}\left(\frac{N-1}{N}\right), \mathcal{G}\left(\frac{N-2}{N}\right), \dots, \mathcal{G}\left(\frac{2}{N}\right), \mathcal{G}\left(\frac{1}{N}\right)$$

- D'autre part :

$$\begin{aligned} T &= T_{N-1} + 1 \\ &= \Delta_1 + \Delta_2 + \cdots + \Delta_{N-2} + \Delta_{N-1} + 1 \end{aligned}$$

Donc, d'après les questions **2** et **3**, la variable aléatoire $T-1$ est une somme de v.a.r. indépendantes suivants les lois :

$$\mathcal{G}\left(\frac{1}{N}\right), \mathcal{G}\left(\frac{2}{N}\right), \dots, \mathcal{G}\left(\frac{N-2}{N}\right), \mathcal{G}\left(\frac{N-1}{N}\right)$$

Donc les v.a.r. $S - 1$ et $T - 1$ ont même loi.

On en déduit que les v.a.r. S et T ont même loi.

□

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

- a)** Exprimer l'événement $[S > m]$ à l'aide des événements $B_1^m, B_2^m, \dots, B_N^m$.

Démonstration.

L'événement $[S > m]$ est réalisé si et seulement si, à l'instant m , au moins l'un des N timbres n'a pas encore été reçu par le collectionneur, c'est-à-dire si, à l'instant m , le collectionneur ne possède pas :

- × soit le timbre numéro 1,
- × soit le timbre numéro 2,
- × ...
- × soit le timbre numéro N .

Ainsi : $[S > m] = B_1^m \cup B_2^m \cup \cdots \cup B_N^m$.

□

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

Démonstration.

Soit $j \in \llbracket 1, N \rrbracket$.

On introduit les événements suivants :

$$\forall i \in \llbracket 1, m \rrbracket, D_j^i = \text{« ne pas recevoir le timbre numéro } j \text{ le jour } i \text{ »}$$

L'événement B_j^m est réalisé si et seulement si le timbre numéro j n'a jamais été reçu lors des m premiers jours. On en déduit :

$$B_j^m = D_j^1 \cap D_j^2 \cap \dots \cap D_j^m$$

On obtient :

$$\begin{aligned} \mathbb{P}(B_j^m) &= \mathbb{P}(D_j^1 \cap D_j^2 \cap \dots \cap D_j^m) \\ &= \mathbb{P}(D_j^1) \times \mathbb{P}(D_j^2) \times \dots \times \mathbb{P}(D_j^m) && \text{(par indépendance} \\ & && \text{des arrivages)} \\ &= \prod_{i=1}^m \mathbb{P}(D_j^i) \end{aligned}$$

Or chaque jour le timbre servant à affranchir une lettre est choisi uniformément parmi les N positions possibles. Ainsi :

$$\forall i \in \llbracket 1, m \rrbracket, \mathbb{P}(D_j^i) = \frac{N-1}{N}$$

On en déduit :

$$\mathbb{P}(B_j^m) = \prod_{i=1}^m \frac{N-1}{N} = \left(\frac{N-1}{N} \right)^m$$

$\forall j \in \llbracket 1, N \rrbracket, \mathbb{P}(B_j^m) = \left(\frac{N-1}{N} \right)^m = \left(1 - \frac{1}{N} \right)^m$

□

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

Démonstration.

Montrons par récurrence : $\forall k \geq 2, \mathcal{P}(k)$

où $\mathcal{P}(k)$: « pour tout A_1, \dots, A_k des événements, $\mathbb{P}\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k \mathbb{P}(A_i)$ ».

Initialisation :

Soient A_1, A_2 deux événements.

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) && \text{(par formule du crible)} \\ &\leq \mathbb{P}(A_1) + \mathbb{P}(A_2) && \text{(car } \mathbb{P}(A_1 \cap A_2) \geq 0) \end{aligned}$$

D'où $\mathcal{P}(2)$.

Hérédité : soit $k \geq 2$. Supposons $\mathcal{P}(k)$. Montrons $\mathcal{P}(k+1)$.

Soient A_1, \dots, A_{k+1} des événements. Notons $C_k = \bigcup_{i=1}^k A_i$.

$$\begin{aligned}
 \mathbb{P}\left(\bigcup_{i=1}^{k+1} A_i\right) &= \mathbb{P}(C_k \cup A_{k+1}) \\
 &= \mathbb{P}(C_k) + \mathbb{P}(A_{k+1}) - \mathbb{P}(C_k \cap A_{k+1}) && \text{(par formule du crible)} \\
 &\leq \mathbb{P}(C_k) + \mathbb{P}(A_{k+1}) && \text{(car } \mathbb{P}(C_k \cap A_{k+1}) \geq 0) \\
 &= \mathbb{P}\left(\bigcup_{i=1}^k A_i\right) + \mathbb{P}(A_{k+1}) && \text{(par définition de } C_k) \\
 &\leq \sum_{i=1}^k \mathbb{P}(A_i) + \mathbb{P}(A_{k+1}) && \text{(par hypothèse de récurrence)} \\
 &= \sum_{i=1}^{k+1} \mathbb{P}(A_i)
 \end{aligned}$$

D'où $\mathcal{P}(k+1)$.

On a donc montré par récurrence : pour tout $k \geq 2$, si A_1, \dots, A_k sont des événements, alors $\mathbb{P}\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k \mathbb{P}(A_i)$.

□

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

Démonstration. On utilise les questions précédentes :

$$\begin{aligned}
 \mathbb{P}([S > m]) &= \mathbb{P}\left(\bigcup_{j=1}^N B_j^m\right) && \text{(d'après 20.a)} \\
 &\leq \sum_{j=1}^N \mathbb{P}(B_j^m) && \text{(d'après 20.c)} \\
 &= \sum_{j=1}^N \left(1 - \frac{1}{N}\right)^m && \text{(d'après 20.b)} \\
 &= N \left(1 - \frac{1}{N}\right)^m
 \end{aligned}$$

□

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

Démonstration.

La fonction $x \mapsto \ln(1+x)$ est concave sur $]-1, +\infty[$. Sa courbe représentative est donc située sous ses tangentes, notamment celle au point d'abscisse 0, droite d'équation :

$$y = x$$

On en déduit : $\forall x \in]-1, +\infty[, \ln(1+x) \leq x$.

□

b) Dédire des résultats précédents la majoration :

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

Démonstration.

Soit $m \in \mathbb{N}$.

- Tout d'abord, d'après la question **19**, les v.a.r. T et S ont même loi. Donc :

$$\mathbb{P}([T > m]) = \mathbb{P}([S > m])$$

Commentaire

On fera bien attention à ne pas écrire $T = S$! En effet, si les v.a.r. T et S ont même loi, elles ne sont pas égales !

× T et S ont même loi si et seulement si : $\forall x \in \mathbb{R}, F_S(x) = F_T(x)$ (où F_S et F_T sont les fonctions de répartition de S et T)

× les v.a.r. T et S sont égales si et seulement si : $\forall \omega \in \Omega, T(\omega) = S(\omega)$.

- D'après la question **20.d**), on a donc :

$$\begin{aligned} \mathbb{P}([T > m]) &\leq N \left(1 - \frac{1}{N}\right)^m \\ &\quad \parallel \\ &N e^{m \ln\left(1 - \frac{1}{N}\right)} \end{aligned}$$

- Or, comme $N \geq 2$: $-\frac{1}{N} > -1$. On peut donc appliquer le résultat de la question précédente :

$$\begin{aligned} \ln\left(1 - \frac{1}{N}\right) &\leq -\frac{1}{N} \\ \text{alors } m \ln\left(1 - \frac{1}{N}\right) &\leq -m \frac{1}{N} \quad (\text{car } m \geq 0) \\ \text{d'où } e^{m \ln\left(1 - \frac{1}{N}\right)} &\leq e^{-m \frac{1}{N}} \quad (\text{par croissance de la fonction exp sur } \mathbb{R}) \end{aligned}$$

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

□

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 : $d(\mu_n, \pi) \leq e^{-c}$.

Démonstration.

Soit $n \geq N \ln(N) + cN$.

- D'après la question **16** :

$$\begin{aligned} d(\mu_n, \pi) &\leq \mathbb{P}([T > n]) \\ &\leq N e^{-\frac{n}{N}} \quad (\text{d'après } \mathbf{21.b}) \end{aligned}$$

• Or :

$$\begin{aligned}
 N e^{-\frac{n}{N}} \leq e^{-c} &\Leftrightarrow \ln\left(N e^{-\frac{n}{N}}\right) \leq \ln(e^{-c}) && \text{(par stricte croissance de la fonction } \ln \text{ sur }]0, +\infty[) \\
 &\Leftrightarrow \ln(N) - \frac{n}{N} \leq -c \\
 &\Leftrightarrow N \ln(N) - n \leq -cN && \text{(car } N > 0) \\
 &\Leftrightarrow N \ln(N) + cN \leq n
 \end{aligned}$$

La dernière inégalité est vérifiée.

Donc, par équivalence : $\forall n \geq N \ln(N) + cN, d(\mu_n, \pi) \leq e^{-c}$.

□

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 ?

Démonstration.

• Une distance à la loi uniforme (c'est-à-dire la distance à π) acceptable est de 0, 2. On souhaite donc trouver $n_0 \in \mathbb{N}^*$ tel que :

$$n_0 \geq N \ln(N) + cN \quad \text{avec} \quad e^{-c} \leq 0, 2$$

En effet, on aura alors par transitivité :

$$d(\mu_{n_0}, \pi) \leq e^{-c} \leq 0, 2 = \frac{1}{5}$$

• Or :

$$\begin{aligned}
 e^{-c} \leq \frac{1}{5} &\Leftrightarrow -c \leq \ln\left(\frac{1}{5}\right) && \text{(par stricte croissance de } \ln \text{ sur }]0, +\infty[) \\
 &\Leftrightarrow c \geq \ln(5)
 \end{aligned}$$

On choisit donc $c = \ln(5)$ (on a bien $c > 0$).

• Il faut donc choisir ensuite un entier n_0 tel que : $n_0 \geq N \ln(N) + cN$.

• Comme $N = 32$, on doit donc trouver un majorant de $32 \ln(32) + 32 \ln(5)$.

× On sait déjà d'après la question 12 : $32 \ln(32) \approx 110$.

× De plus :

$$\begin{aligned}
 e \leq 5 \leq e^2 &\Leftrightarrow \ln(e) \leq \ln(5) \leq \ln(e^2) && \text{(par stricte croissance de la fonction } \ln \text{ sur }]0, +\infty[) \\
 &\Leftrightarrow 1 \leq \ln(5) \leq 2 \\
 &\Leftrightarrow 32 \leq 32 \ln(5) \leq 64 \\
 &\Leftrightarrow 143 \leq 110 + 32 \ln(5) \leq 174
 \end{aligned}$$

On choisit donc de faire 174 battages par insertions pour considérer le paquet mélangé de façon acceptable.

Commentaire

On notera que l'approximation $32 \ln(32)$ nous empêche de répondre de manière convenable à la question. En effet, c'est une majoration de $32 \ln(32)$ qui est nécessaire pour trouver un n_0 vérifiant $d(\mu_{n_0}, \pi) \leq \frac{1}{5}$.

□