DS2 bareme (version B) - sujet 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 $C_1, \ldots, C_N$ 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 $(\Omega, \mathscr{A}, \mathbb{P})$ avec $\Omega = \mathcal{S}_N$, $\mathscr{A} = \mathcal{P}(\mathcal{S}_N)$ 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 $\dfrac{1}{N!}$, 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 : $(C_1,\ C_2,\ \ldots,\ C_N)$
(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 tout en dessous du paquet.
Pour k élément de $[\![1, N]\!]$, 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 $\{1, \ldots, N\}$ la place à laquelle l'insertion a lieu, indépendamment des insertions précédentes.
Les instants successifs d'insertions seront notées 1, 2, ..., n, ..., 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 $[\![1, N-1]\!]$, T_i le premier instant où la carte C_N atteint la position $N-i$.
- × on posera également $\Delta_1 = T_1$ et $\forall i \in [\![2, N-1]\!]$, $\Delta_i = T_i - T_{i-1}$.
- × 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 $(\Delta_i)_{i\in[\![1,N-1]\!]}$ sont indépendantes.

Description d'un exemple. Dans le tableau ci-dessous, nous décrivons les résultats d'une expérience faite sur un paquet de $N = 4$ cartes. La première ligne du tableau indique les instants n, la deuxième ligne indique les positions d'insertions, et dans 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		position	1	C_1	C_2	C_3	C_2	C_2	C_1	C_4	C_2
du		position	2	C_2	C_3	C_2	C_1	C_1	C_4	C_2	C_4
paquet		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 experience, on a les resultats :
× $T_1(\omega) = 3$, $T_2(\omega) = 5$ 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 : $\forall i \in [\![2, N-1]\!]$, $T_i = \Delta_1 + \Delta_2 + \cdots + \Delta_i$.
Que représente l'intervalle de temps Δ_i ?

- **1 pt : télescopage correct (formule de Δ_1 particulière)**
- **1 pt : 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 Δ_1.
Déterminer pour tout entier $n \in \mathbb{N}$, $\mathbb{P}([\Delta_1 > n])$ et reconnaître la loi de Δ_1.

- **1 pt : 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**
- **1 pt :** $[\Delta_1 > n] \; = \; [X_1 \neq N] \cap [X_2 \neq N] \cap \cdots \cap [X_n \neq N]$
- **1 pt :** $\mathbb{P}([\Delta_1 > n]) \; = \; \mathbb{P}\left(\bigcap\limits_{j=1}^{n} [X_j \neq N]\right) \; = \; \prod\limits_{j=1}^{n} \mathbb{P}([X_j \neq N])$ **par indépendance**
- **1 pt :** $\mathbb{P}([X_j \neq N]) = \dfrac{N-1}{N}$ **par équiprobabilité**
- **1 pt :** $\mathbb{P}([\Delta_1 > n]) = \left(1 - \dfrac{1}{N}\right)^n$
- **1 pt :** $\Delta_1(\Omega) = \mathbb{N}^*$
- **1 pt :** $\Delta_1 \hookrightarrow \mathcal{G}\left(\dfrac{1}{N}\right)$ **(pas de calcul précis attendu)**

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

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

- **1 pt : formule des probabilités totales avec le système complet d'événements $([T_{i-1} = k])_{k\in\mathbb{N}^*}$ (la v.a.r. T_{i-1} est à valeurs entières non nulles)**

- **1 pt :** $[T_{i-1} = k] \cap [\Delta_i > n] = [T_{i-1} = k] \cap [X_{k+1} \in [\![1, N-i]\!]] \cap [X_{k+2} \in [\![1, N-i]\!]] \cap \cdots \cap [X_{k+n} \in [\![1, N-i]\!]]$
- **1 pt : indépendance de** $[T_{i-1} = k]$ **et** $\bigcap\limits_{j=k+1}^{n} [X_j \in [\![1, N-i]\!]]$ **par lemme des coalitions**
- **1 pt :** $\sum\limits_{k=1}^{+\infty} \mathbb{P}([T_{i-1} = k]) = 1$ **par argument de sce**
- **2 pts :** $\mathbb{P}([\Delta_i > n]) = \left(1 - \dfrac{i}{N}\right)^n$ **(seulement 1 pt si il manque un argument avant)**
- **1 pt :** $\Delta_i \hookrightarrow \mathcal{G}\left(\dfrac{i}{N}\right)$

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

- **1 pt :** $\mathbb{E}(\Delta_i) \ = \ \dfrac{1}{\frac{i}{N}} \ = \ \dfrac{N}{i}$
- **1 pt :** $\mathbb{V}(\Delta_i) \ = \ \dfrac{1 - \frac{i}{N}}{\left(\frac{i}{N}\right)^2} \ = \ \dfrac{\frac{N-i}{\cancel{N}}}{\frac{i^2}{N^{\cancel{2}}}} \ = \ N \ \dfrac{N-i}{i^2}$

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

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

- **1 pt : la famille** $([\Delta_1 = k])_{k \in \mathbb{N}^*}$ **est un système complet d'événements**
- **1 pt :** $\mathbb{P}([T_2 = n]) = \sum\limits_{k=1}^{+\infty} \mathbb{P}([\Delta_1 = k])\mathbb{P}([\Delta_2 = n-k])$ **par indépendance de** Δ_1 **et** Δ_2
- **1 pt :** $\left\{ \begin{array}{l} n-k \in \Delta_2(\Omega) = \mathbb{N}^* \\ k \geqslant 1 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} 1 \leqslant n-k \\ 1 \leqslant k \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} k \leqslant n-1 \\ 1 \leqslant k \end{array} \right. \Leftrightarrow \left\{ \ 1 \leqslant k \leqslant n-1 \right.$

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

- **1 pt :** $\dfrac{1 - \frac{1}{N}}{1 - \frac{2}{N}} \neq 1$
- **1 pt : calcul correct**

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

- **2 pts : calcul correct en utilisant les deux questions précédentes et la loi de** Δ_i

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 $X_{T_2}(\Omega)$?

- **1 pt : 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

- **1 pt :** $X_{T_2}(\Omega) = \{N-1, N\}$

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

- **1 pt : 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.**
  × **à l'instant $T_2(\omega)$, la carte du dessus du paquet est insérée en position N.**
- **1 pt : 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.**
  × **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 $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N-1])$.

- **1 pt : $[X_{T_1} = N] = \Omega$ donc $\mathbb{P}([X_{T_1} = N]) = 1$**
- **1 pt : la famille $([X_{T_2} = N-1]\,,\ [X_{T_2} = N])$ forme un système complet d'événements**
- **1 pt : la carte du dessus est insérée uniformément sur l'ensemble des positions possibles donc $\mathbb{P}([X_{T_2} = N-1]) = \mathbb{P}([X_{T_2} = N])$**
- **1 pt : $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N-1]) = \dfrac{1}{2}$ et $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N]) = \dfrac{1}{2}$**

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 $X_{T_3}(\Omega)$.

- **1 pt : 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**
- **1 pt :** $X_{T_3}(\Omega) = \{N-2, N-1, N\}$

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

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

- **1 pt : FPT**

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

- **1 pt : Or les insertions s'effectuent de manière uniforme, donc :**

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

- **1 pt : $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N]) = \dfrac{1}{6}$**
  et $\mathbb{P}([X_{T_1} = N] \cap [X_{T_2} = N] \cap [X_{T_3} = N-1]) = \dfrac{1}{6}$

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 !

- **1 pt : la v.a.r. X_{T_j} suit la loi uniforme sur $[\![N-j+1, N]\!]$.**
  Donc, à l'instant T_j, la carte C_N est en position $N-j$ et la configuration des cartes sous elle (positions $N-j+1$, ..., N) est uniforme
- **1 pt : 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**

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

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

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

8. Espérance et variance de T

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

- **1 pt :** $T = \sum_{i=1}^{N-1} \Delta_i + 1$
- **1 pt :** T **admet une espérance et une variance en tant que somme de v.a.r. qui admettent une espérance et une variance**
- **1 pt : linéarité de l'espérance citée**
- **1 pt : calcul correct de l'espérance**
- **1 pt : les v.a.r.** Δ_1, ..., Δ_{N-1} **sont indépendantes**
- **1 pt :** $\mathbb{V}(T) = \sum_{i=1}^{N-1} N\,\frac{N-i}{i^2}$
- **1 pt : calcul correct de la variance**

9. Étude de la suite (u_n)

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

- **1 pt :** $\frac{1}{k} \geqslant \frac{1}{t} \geqslant \frac{1}{k+1}$
- **1 pt : 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**
- **1 pt : par croissance de l'intégrale, les bornes étant dans l'ordre croissant** $(k \leqslant k+1)$

b) En déduire successivement :

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

- **1 pt :** $u_{n+1} - u_n = \frac{1}{n+1} - \ln(n+1) + \ln(n)$
- **1 pt :** $\ln(n+1) - \ln(n) = \int_n^{n+1} \frac{1}{t}\,dt$

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

- **1 pt : sommation des inégalités de gauche pour** k **variant de** 1 **à** $n-1$
- **1 pt : sommation des inégalités de droite pour** k **variant de** 1 **à** n

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

- **1 pt :** $u_n \geqslant \ln(n+1) - \ln(n) \geqslant 0$
- **1 pt : utilisation justifiée du théorème de convergence monotone**

- **1 pt : pour tout $n \in \mathbb{N}^*$, $u_n \leqslant 1$**

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

- **1 pt :** $\dfrac{\ln(n+1)}{\ln(n)} \leqslant \dfrac{H_n}{\ln(n)} \leqslant 1 + \dfrac{1}{\ln(n)}$ **(0 pt si $\ln(n) > 0$ n'apparait pas)**
- **1 pt :** $\lim\limits_{n\to+\infty} \dfrac{\ln(n+1)}{\ln(n)} = 1$
- **1 pt : Par théorème d'encadrement :** $\lim\limits_{n\to+\infty} \dfrac{H_n}{\ln(n)} = 1$
- **2 pts :** $\mathbb{E}(T) = N\,\ln(N) + N\,\gamma + \underset{N\to+\infty}{o}(N) \iff \lim\limits_{N\to+\infty} u_N = \gamma$

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

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

- **1 pt : La suite $\left(\sum\limits_{k=1}^{n} \dfrac{1}{k^2}\right)_{n\in\mathbb{N}^*}$ est la suite des sommes partielle de la série de Riemann d'exposant 2 (avec $2 > 1$). Donc elle converge.**
- **1 pt :** $\dfrac{H_N}{N} \underset{N\to+\infty}{\sim} \dfrac{\ln(N)}{N} \underset{N\to+\infty}{\longrightarrow} 0$
- **1 pt : on a :** $\lim\limits_{N\to+\infty} \dfrac{\mathbb{V}(T)}{N^2} = \sum\limits_{k=1}^{+\infty} \dfrac{1}{k^2}$ **et on note alors** $\alpha = \sum\limits_{k=1}^{+\infty} \dfrac{1}{k^2}$
- **1 pt :** $\sum\limits_{k=1}^{N} \dfrac{1}{k^2} - \dfrac{H_N}{N} \leqslant \sum\limits_{k=1}^{N} \dfrac{1}{k^2} \leqslant \sum\limits_{k=1}^{+\infty} \dfrac{1}{k^2} = \alpha$

11. Écart à la moyenne
On admet l'inégalité de *Bienaymé-Chebychev* valable pour une variable aléatoire X admettant une espérance et une variance :

$$\forall \varepsilon > 0,\ \mathbb{P}\left([|X - \mathbb{E}(X)| \geqslant \varepsilon]\right) \leqslant \frac{\mathbb{V}(X)}{\varepsilon^2}$$

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

a) Justifier : $\forall \omega \in \Omega,\ |T(\omega) - N\,\ln(N)| \leqslant |T(\omega) - \mathbb{E}(T)| + N$.
Comparer par une inclusion les événements suivants :

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

- **1 pt :** $|T(\omega) - N\,\ln(N)| \leqslant |T(\omega) - \mathbb{E}(T)| + |\mathbb{E}(T) - N\,\ln(N)|$ **par inégalité triangulaire**
- **1 pt :** $|\mathbb{E}(T) - N\,\ln(N)| = |N\,H_N - N\,\ln(N)| = N\,|u_N| \leqslant N$
- **2 pts :** $[|T - N\,\ln(N)| \geqslant c\,N] \subset [|T - \mathbb{E}(T)| \geqslant N\,(c-1)]$

b) Démontrer :

$$\mathbb{P}\left([|T - N\,\ln(N)| \geqslant c\,N]\right) \leqslant \frac{\alpha}{(c-1)^2}$$

où α a été définie à la question 10b.
Le nombre N étant fixé, que vaut $\lim\limits_{c\to+\infty} \mathbb{P}\left([|T - N\,\ln(N)| \geqslant c\,N]\right)$?

- **1 pt :** $\mathbb{P}\big([|T - N\ \ln(N)| \geqslant c\,N]\big) \leqslant \mathbb{P}\big([|T - \mathbb{E}(T)| \geqslant N\,(c-1)]\big)$
- **1 pt : utilisation de l'inégalité de Bienaymé-Chebychev pour avoir l'inégalité voulue**
- **1 pt : par théorème d'encadrement** $\displaystyle\lim_{c\to+\infty} \mathbb{P}\left([|T - N\ \ln(N)| \geqslant c\,N]\right) = 0$

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

$$\lim_{N\to+\infty} \mathbb{P}\left([|T - N\ \ln(N)| \geqslant \varepsilon\,N\ \ln(N)]\right) = 0$$

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

- **1 pt : choix** $c = \varepsilon\ \ln(N)$
- **1 pt :** $\forall N > \mathrm{e}^{\frac{1}{\varepsilon}}$, $\quad 0 \leqslant \mathbb{P}\big([|T - N\ \ln(N)| \geqslant \varepsilon\,N\ \ln(N)]\big) \leqslant \dfrac{\alpha}{(\varepsilon\ \ln(N) - 1)^2}$ **(0 pt si aucune quantification correcte sur le** N**)**
- **1 pt : par théorème d'encadrement** $\displaystyle\lim_{N\to+\infty} \mathbb{P}\left([|T - N\ \ln(N)| \geqslant \varepsilon\,N\ \ln(N)]\right) = 0$

13. Simulation informatique. Dans cette question on considère un jeu de $N = 32$ cartes.
MODÉLISATION : le paquet de 32 cartes est représenté par un vecteur `Jeu` rempli initialement des entiers 1 à 32. La carte sur le dessus du paquet se trouve en première coordonnée du vecteur, celle en dessous se trouve en dernière coordonnée.
De manière générale, `Jeu(i)` désigne la carte (représentée par son numéro) qui se trouve en $i^{\text{ème}}$ position dans le paquet. Par exemple, si `Jeu(i)` contient 10 c'est que la carte C_{10} est en position i.
Initialement, `Jeu` est rempli de telle sorte que `Jeu(i) = i` pour tout `i` dans $[\![1, N]\!]$.

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

- **1 pt :** `Jeu = np.arange(1,33)`

b) Compléter la fonction `Insertion` qui simule une opération d'insertion. Elle prend en paramètre un vecteur qui donne la configuration du jeu avant insertion et renvoie un vecteur 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 $[\![1, 32]\!]$.

```
1  def Insertion(Jeu):
2      k = ____________________ # position où 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] = _______________
9      return J
```

- **1 pt :** 2 `k = rd.randint(1,33)`
- **1 pt :** 7 `J[i] = Jeu[i+1]`
- **1 pt :** 8 `J[k] = cartedessus`

c) Que fait la fonction `Simu` suivante ?

```
def Simu():
    Jeu = np.arange(1,33)
    n = 0
    while Jeu[0] != 32:
        Jeu = Insertion(Jeu)
        n = n + 1
    return n + 1
```

- **1 pt : La fonction `Simu` renvoie une réalisation de la v.a.r. T**
- **2 pts : explications**

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.

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

- **1 pt : initialisation `S=0`**
- **2 pts : boucle for**
- **1 pt : `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 $\mathcal{S}_N$: $\mu_n(A) = \sum\limits_{\sigma \in A} \mu_n(\{\sigma\})$.
  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 :

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

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

- **1 pt : D'après la question *7.*, à partir de l'instant T, toutes les configurations sont équiprobables**
- **1 pt : rédaction correcte (si l'événement $[T \leqslant n]$ est réalisé, alors, ...)**
- **1 pt : formule des probabilités composées**

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

- 1 pt : $E_n \cap [T > n] \subset [T > n]$

- 1 pt : croissance de l'application probabilité

c) Montrer :

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

- 1 pt : $\mu_n(A) \ = \ \mathbb{P}(E_n)$

- 1 pt : La famille $([T \leqslant n], [T > n])$ **forme un système complet d'événements**

- 1 pt : $\mathbb{P}([T \leqslant n]) \leqslant 1$

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

- 1 pt : $\mu_n(\overline{A}) - \pi(\overline{A}) \ = \ \pi(A) - \mu_n(A)$

b) Déduire des questions précédentes la majoration :

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

- 1 pt : $\mu_n(A) - \pi(A) \ \leqslant \ \mathbb{P}([T > n])$

- 1 pt : $\mu_n(A) - \pi(A) \geqslant -\mathbb{P}([T > n])$ **en appliquant le résultat précédent à** $\overline{A}$

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

- 1 pt : passage au max dans l'inégalité précédente

- 1 pt : $\lim\limits_{n \to +\infty} \mathbb{P}([T > n]) = 1 - 1 = 0$ **par propriété des fonctions de répartition**

- 1 pt : théorème d'encadrement

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

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

17. Déterminer la loi de S_1.

- 1 pt : $S_1(\Omega) = \{1\}$

- 1 pt : la v.a.r. S_1 **est la variable aléatoire constante égale à** 1

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

- **1 pt :** $S_k(\Omega) = \mathbb{N}^*$
- **1 pt :** $\forall k \in [\![2, N]\!]$**,** $S_k \hookrightarrow \mathcal{G}\left(\dfrac{N-k+1}{N}\right)$
- **3 pts : niveau de détails de l'argumentation**

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

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

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

- **1 pt :** $T - 1$ **est une somme de v.a.r. indépendantes suivants les lois :**

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

20. 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.

- **1 pt : 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**
- **1 pt :** $[S > m] = B_1^m \ \cup\ B_2^m\ \cup\ \cdots\ \cup\ B_N^m$

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

- **1 pt :** $B_j^m \ = \ D_j^1 \cap D_j^2 \cap \cdots \cap D_j^m$
- **1 pt : indépendance citée**
- **1 pt :** $\mathbb{P}\big(B_j^m\big) \ = \ \left(\dfrac{N-1}{N}\right)^m \ = \ \left(1 - \dfrac{1}{N}\right)^m$

c) On rappelle que pour tout entier $n \geqslant 2$ et pour toute famille d'événements A_1, …, A_n, on a l'inégalité : $\mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right) \ \leqslant\ \sum\limits_{i=1}^{n} \mathbb{P}(A_i)$. En déduire : $\mathbb{P}([S > m]) \leqslant N\left(1 - \dfrac{1}{N}\right)^m$.

- **1 pt : calcul correct**

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

- **1 pt : La fonction** $x \mapsto \ln(1+x)$ **est concave sur** $]-1, +\infty[$
- **1 pt : Sa courbe représentative est donc située sous ses tangentes, notamment celle au point d'abscisse** 0
- **1 pt : la tangente en** 0 **a pour équation** $y = x$

b) Déduire des résultats précédents la majoration :

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

- **1 pt : les v.a.r.** T **et** S **ont même loi donc :** $\mathbb{P}([T > m]) \ = \ \mathbb{P}([S > m])$
- **1 pt :** $\mathbb{P}([T > m]) \leqslant N\, \mathrm{e}^{m \ln\left(1 - \frac{1}{N}\right)}$
- **1 pt :** $N \geqslant 2$ **:** $-\dfrac{1}{N} > -1$**. On peut donc appliquer le résultat de la question précédente**

- 1 pt : fin calcul bien justifiée

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

a) Soit $c > 0$ fixé. Montrer que pour n entier supérieur ou égal à $N\ \ln(N) + c\,N$ on a : $d(\mu_n, \pi) \leqslant \mathrm{e}^{-c}$.

- 1 pt : $d(\mu_n, \pi) \leqslant N\,\mathbf{e}^{-\frac{n}{N}}$

- 1 pt : $N\,\mathbf{e}^{-\frac{n}{N}} \leqslant \mathbf{e}^{-c} \iff N\ \ln(N) + c\,N \leqslant n$

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 ?

- 1 pt : $\mathbf{e}^{-c} \leqslant \dfrac{1}{5} \iff c \geqslant \ln(5)$

- 1 pt : $143 \leqslant 110 + 32\ \ln(5) \leqslant 174$

- 1 pt : explications raisonnables