Correction du Devoir Surveillé n° 1 -ESSEC/HEC
07/10/22

Partie 1 :

On note I_j la variable donnant la position de la j-ième insertion.

1. Justifier que $\forall i \in [[2, N-1]] \quad T_i = \Delta_1 + \Delta_2 + \cdots + \Delta_i$.
   Que représente l'intervalle de temps Δ_i ?
2. Loi de Δ_1.
   (a) Exprimer l'évènement $(\Delta_1 > n)$ à l'aide des évènements $[I_j < N]$.
   (b) Déterminer pour tout entier $n \in \mathbb{N}$, $\mathbf{P}(\Delta_1 > n)$.
   (c) Reconnaître la loi de Δ_1.
3. Soit $i \in [[2, N-1]]$. Loi de Δ_i.
   (a) Établir que pour tout entier $n \in \mathbb{N}$, on a $\mathbf{P}(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n$.
   (b) En déduire que Δ_i suit une loi usuelle que l'on précisera.
   (c) En déduire $E(\Delta_i) = \dfrac{N}{i}$, et $V(\Delta_i) = N\dfrac{N-i}{i^2}$.
4. Loi de T_2 . Soit $n \geq 2$.
   (a) Démontrer que $\mathbf{P}(T_2 = n) = \sum\limits_{k=1}^{n-1} \mathbf{P}(\Delta_2 = n-k)\,\mathbf{P}(\Delta_1 = k)$
   (b) Justifier que $\sum\limits_{k=1}^{n-1} \left(\dfrac{1-1/N}{1-2/N}\right)^k = N\left(1 - \dfrac{1}{N}\right)\left[\left(\dfrac{1-1/N}{1-2/N}\right)^{n-1} - 1\right]$
   (c) En déduire que l'on a : $\mathbf{P}(T_2 = n) = \frac{2}{N}\left[\left(1-\frac{1}{N}\right)^{n-1} - \left(1-\frac{2}{N}\right)^{n-1}\right]$.
5. À l'instant T_2, la carte C_N est située en position $N-2$ et deux cartes se trouvent sous elle qui ont été insérées aux instants T_1 et T_2.
   Que valent alors les probabilités, qu'à l'instant T_2 :
   (a) la carte insérée à l'instant T_1 soit en place $N-1$ et celle insérée à l'instant T_2 en place N ?
   (b) la carte insérée à l'instant T_2 soit en place $N-1$ et celle insérée à l'instant T_1 en place N ?
6. À l'instant T_3, la carte C_N est située en position $N-3$ et trois cartes, insérées aux instants T_1, T_2 et T_3, se trouvent sous elle. On note alors, pour $i \in \{1, 2, 3\}$, a_i la position de la carte ayant été insérée à l'instant T_i.
   (a) Combien y a-t-il de résultats possibles pour le triplet (a_1, a_2, a_3) ?
   (b) Quelques exemples. Donner les probabilités qu'à l'instant T_3 :
      i. on obtienne $(a_1, a_2, a_3) = (N-2, N-1, N)$?
      ii. on obtienne $(a_1, a_2, a_3) = (N-2, N, N-1)$?
7. Justifier la phrase suivante :

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

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

Partie 1 - Description et premiers résultats

1. Pour tout i, Δ_i est le nombre de coups séparant la $i-1^{ème}$ de la $i^{ème}$ remontée.
   Donc T_i nombre total de coup pour la $i^{ème}$ remontée est $T_i = T_1 + \cdots + \Delta_i = \Delta_1 + \cdots + \Delta_i$
   (on pouvait aussi le faire par récurrence, plus cohérent avec l'ordre des deux questions)
2. Loi de Δ_1.
   (a) $(\Delta_1 > n)$ signifie qu'aucune insertion ne s'est faite en $N^{ème}$ position durant les n premières insertions.
   En notant I_m la position de la $m^{ième}$ insertion on a donc $(\Delta_1 > n) = \bigcap_{m=1}^{n} (I_m < N)$

(b) Par indépendance des insertions $\mathrm{P}(\Delta_1 > n) = \prod_{m=1}^{n} \mathrm{P}(I_m < N)$ avec $\mathrm{P}(I_m < N) = \frac{N-1}{N}$ car les positions d'insertion sont équiprobables.

Conclusion : $\boxed{\mathrm{P}(\Delta_1 > n) = \left(\frac{N-1}{N}\right)^n}$

et comme (valeurs entières) on a alors $(\Delta_1 > n-1) = (\Delta_1 > n) \cup (\Delta_1 = n)$ donc

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

(c) On reconnaît, avec $\Delta_1(\Omega) = \mathbb{N}^*$, $\boxed{\Delta_1 \hookrightarrow \mathcal{G}\left(\frac{1}{N}\right)}$

N.B. on pouvait le voire directement avec Δ_1 rang de la première insertion en position N dans une suite d'insertions indépendantes, la probabilité à chacune étant de $1/N$

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

(a) Pour tout $n \in \mathbb{N}$

$(\Delta_i > n)$ signifie que, il y a plus de n insertion pour la $i^{ème}$ remontée de la carte C_N

Cette remontée se fait de la position $N-i-1$ (après $i-1$ remontées) à la $N-i$.

Il y a donc $N-i$ positions d'insertions permises.

$(\Delta_i > n) = \bigcap_{m=1}^{n} (I_m \leq N-i)$ d'et par indépendance des insertions

$\mathrm{P}(\Delta_i > n) = \prod_{m=1}^{n} \mathrm{P}(I_m \leq N-i)$ avec $\mathrm{P}(I_m \leq N-i) = \dfrac{N-i}{N}$

Conclusion : $\boxed{\mathrm{P}(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n}$

(b) Comme précédemment, *Conclusion :* $\boxed{\Delta_i \hookrightarrow \mathcal{G}\left(\frac{i}{N}\right)}$

(c) On a alors $E(\Delta_i) = \dfrac{1}{i/N} = \dfrac{N}{i}$ et $V(\Delta_i) = \dfrac{1-\frac{i}{N}}{\left(\frac{i}{N}\right)^2} = \dfrac{N(N-i)}{i^2}$

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

(a) $T_2 = \Delta_1 + \Delta_2$ donc

$$(T_2 = n) = \bigcup_{k=1}^{n-1} (\Delta_1 = k \cap \Delta_2 = n-k)$$

les bornes étant imposées par $\Delta_1 \geq 1$ et $\Delta_2 \geq 1$

Donc ($\cup$ incompatibles et (Δ_i) indépendants)

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

(b) On a

$$\begin{aligned}
\sum_{k=1}^{n-1} \left(\frac{1-1/N}{1-2/N}\right)^k &= \left(\frac{1-1/N}{1-2/N}\right) \frac{1-\left(\frac{1-1/N}{1-2/N}\right)^{n-1}}{1-\left(\frac{1-1/N}{1-2/N}\right)} \text{ car } \frac{1-1/N}{1-2/N} \neq 1 \\
&= (1-1/N)\, \frac{1-\left(\frac{1-1/N}{1-2/N}\right)^{n-1}}{1-2/N-1+1/N} \\
&= N\left(1-\frac{1}{N}\right)\left[1-\left(\frac{1-1/N}{1-2/N}\right)^{n-1}\right]
\end{aligned}$$

(c) On reprend

$$\begin{aligned}
\mathrm{P}(T_2=n) &= \sum_{k=1}^{n-1} \mathrm{P}(\Delta_2=n-k)\,\mathrm{P}(\Delta_1=k) \text{ et } k \text{ et } n-k\geq 1 \text{ donc}\\
&= \sum_{k=1}^{n-1}\left(1-\frac{2}{N}\right)^{n-k-1}\frac{2}{N}\left(1-\frac{1}{N}\right)^{k-1}\frac{1}{N}\\
&= \frac{2}{N^2}\left(1-\frac{2}{N}\right)^{n-1}\left(1-\frac{1}{N}\right)^{-1}\sum_{k=1}^{n-1}\left(\frac{1-1/N}{1-2/N}\right)^{k}\\
&= \frac{2}{N^2}\left(1-\frac{2}{N}\right)^{n-1}\left(1-\frac{1}{N}\right)^{-1}N\left(1-\frac{1}{N}\right)\left[1-\left(\frac{1-1/N}{1-2/N}\right)^{n-1}\right]\\
&= \frac{2}{N}\left(1-\frac{2}{N}\right)^{n-1}\left[1-\left(\frac{1-1/N}{1-2/N}\right)^{n-1}\right]\\
&= \frac{2}{N}\left[\left(1-\frac{2}{N}\right)^{n-1}-\left(1-\frac{1}{N}\right)^{n-1}\right]
\end{aligned}$$

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
La carte insérée à T_1 sera au dessus de celle insérée à T_2, si (et seulement si) celle insérée à T_2 l'est en position N.
Elle sera en dessous si celle insérée à T_2 l'est en position $N-1$
Les deux cas sont équiprobables. Donc à l'instant T_2
   - « la carte insérée à l'instant T_1 est en place $N-1$ et celle insérée à l'instant T_2 en place N » et
   - « la carte insérée à l'instant T_1 est en place N et celle insérée à l'instant T_2 en place $N-1$ »

   ont la même probabilité : $\dfrac{1}{2}$
6. A l'instant T_3, la carte C_N est située en position $N-3$ et trois cartes, insérées aux instants T_1, T_2 et T_3 se trouvent sous elle. On note alors, pour $i\in\{1,2,3\}$, a_i la position de la carte ayant été insérée à l'instant T_i.
   (a) Après la troisième insertion en dessous de C_N,
   la première carte insérée peut se retrouver en $\{N-2,N-1,N\}$
   la seconde ans $\{N-1,N\}$ et ma troisième est en position N.
   Il y a donc $3\cdot 2\cdot 1=6$ résultats possibles pour (a_1,a_2,a_3)
   Equiprobables ? les points d'insertion faisant remonter C_N sont équiprobables. Il y en a $1\cdot 2\cdot 3$
   (b) Quelques exemples. à l'instant T_3 :
      i. on obtient $(a_1,a_2,a_3)=(N-2,N-1,N)$ si (et seulement si) les insertions ont été faites les trois fois en position N.
      Il a donc une probabilité de $\frac{1}{6}$
      ii. on obtient $(a_1,a_2,a_3)=(N-2,N,N-1)$ si et seulement si
      la carte insérée en T_1 a été deux fois repoussée (insertion T_2 et T_3 en dessous)
      et celle insérée en T_2 n'a pas été repoussée par l'insertion en T_3 donc
      la première carte insertion est en N, la seconde en N et la dernière en $N-1$
      Il a donc une probabilité de $\frac{1}{6}$
7. « À partir de l'instant T, toutes les configurations du jeu de cartes sont équiprobables »
C'est clair ! non ?
À chaque instant T_i $(i<N)$ la carte insérée en dessous de C_N l'est équiprobablement sur chacune des positions $[[N-i+1,N]]$ et on a alors (récurrence) pour les cartes insérées en dessous, toutes les permutations équiprobables.
À l'instant $T_{N-1}+1$, on insère enfin la carte C_N qui se retrouve équiprobablement en toutes positions.
Au final, toutes les permutations seront équiprobables à l'instant T.

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

Partie 2 :

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

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

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

8. <u>Espérance et variance de T</u>

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

9. <u>Étude de la suite (u_n)</u>

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

   (b) En déduire successivement :

   i. la décroissance de la suite (u_n),

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

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

10.

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

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

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

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

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

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

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

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

   (b) Démontrer que

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

   où α a été définie à la question 10b.
   Le nombre N étant fixé, que vaut $\lim\limits_{N\to+\infty} \mathbf{P}(|T - N\ln(N)| \geq cN)$?

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

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

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

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

Modélisation : Le paquet de 32 cartes est représenté par une variable `Jeu` de contentant une matrice remplie initialement d'entiers entre 1 et 32. Donc, initialement, `Jeu[i]` contient $i+1$, c'est à dire que la carte C_i est en position $i-1$ dans la matrice (la numérotation commence à zéro). Au cours des insertions, `Jeu [i]` désigne le numéro de la carte en position numéro $i+1$. Par exemple, `Jeu [i] =10` signifie que la carte C_{11} est en position i.
On indique à la fin de cette question un extrait de programme à compléter en suivant les questions suivantes :

(a) Écrire une fonction d'en-tête `Init()` permettant de définir en sortie une matrice correspondant à la configuration initiale du paquet de cartes.

(b) Compléter la fonction `Insertion` qui simule une opération d'insertion. On rappelle que la fonction `rd.randint(0,32)` permet de tirer un nombre entier au hasard dans l'intervalle $[[0, 31]]$.

```python
import numpy as np
import numpy.random as rd

Jeu=Init()
def Insertion(Jeu):
        k=... # Position où va être insérée la carte du dessus
        Cartedessus=Jeu[0]
        if k>0:
                for i in range(k):
                        Jeu[i]=...

                Jeu[k]=...
```

(c) Que fait la fonction `T` ?

```python
def T(Jeu):
        n=0
        while Jeu[0] != 31:
                Insertion(Jeu)
                n=n+1
        return(n+1)
```

(d) Écrire un programme permettant de calculer et d'afficher la moyenne des valeurs prises par la fonction `T` sur 1000 expériences.

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

8. Espérance et variance de T
On a $T = T_{N-1} + 1 = \sum_{k=1}^{N-1} \Delta_k + 1$ donc

$$\begin{aligned}
E(T) &= 1 + \sum_{k=1}^{N-1} E(\Delta_k) \\
&= 1 + \sum_{k=1}^{N-1} \frac{N}{k} = \frac{N}{N} + \sum_{k=1}^{N-1} \frac{N}{k} \\
&= NH_N \text{ et} \\
V(T) &= \sum_{k=1}^{N-1} V(\Delta_k) \text{ par indépendance} \\
&= \sum_{k=1}^{N-1} \frac{N(N-k)}{k^2} = \sum_{k=1}^{N} \frac{N(N-k)}{k^2} - 0 \\
&= \sum_{k=1}^{N} \frac{N^2}{k^2} - \frac{Nk}{k^2} \\
&= N^2 \sum_{k=1}^{N} \frac{1}{k^2} - N \sum_{k=1}^{N} \frac{1}{k}
\end{aligned}$$

Conclusion : $\boxed{V(T) = N^2 \sum_{k=1}^{N} \frac{1}{k^2} - NH_N}$

9. Étude de la suite (u_n)

(a) Pour tout $k \geq 1$ et $t \in [k, k+1]$ on a

$$\begin{aligned} \frac{1}{k+1} &\leq \frac{1}{t} \leq \frac{1}{k} \text{ donc bornes } k \leq k+1 \\ \int_k^{k+1} \frac{1}{k+1} dt &\leq \int_k^{k+1} \frac{1}{t} dt \leq \int_k^{k+1} \frac{1}{k} dt \text{ et} \\ \frac{1}{k+1} &\leq \int_k^{k+1} \frac{1}{t} dt \leq \frac{1}{k} \end{aligned}$$

(b) On en déduit pour tout entier $n \geq 1$

$$\begin{aligned} u_{n+1} - u_n &= \frac{1}{n+1} - \ln(n+1) + \ln(n) \\ &= \frac{1}{n+1} - \int_n^{n+1} \frac{1}{t} dt \\ &\leq 0 \end{aligned}$$

donc la suite $(u_n)_{n \geq 1}$ est décroissante

Donc, comme $u_1 = H_1 - \ln(1) = 1$ alors pour tout $n \geq 1$: $u_n \leq u_1 = 1$ et $\boxed{H_n \leq \ln(n) + 1}$

Enfin en sommant les inégalités de droite du a)

$$\begin{aligned} \sum_{k=1}^{n} \int_k^{k+1} \frac{1}{t} dt &\leq \sum_{k=1}^{n} \frac{1}{k} \text{ et Chasles} \\ \int_1^{n+1} \frac{1}{t} dt &\leq H_n \end{aligned}$$

Conclusion : $\boxed{\forall n \in \mathbb{N}^*, \quad \ln(n+1) \leq H_n \leq \ln(n) + 1}$

(c) On a donc $\forall n \in \mathbb{N}^*$, $\ln(n+1) - \ln(n) \leq H_n - \ln(n) \leq 1$
et donc $0 \leq \ln\left(\frac{n+1}{n}\right) \leq u_n \leq 1$
La suite u est donc décroissante et minorée par 0 donc converge vers une limite γ et $\gamma \in [0, 1]$

10. (a) Comme $\ln(n+1) \leq H_n \leq \ln(n) + 1$ avec $\ln(n+1) = \ln(n(1 + 1/n)) = \ln(n) + \ln(1 + 1/n)$
donc

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

et par encadrement $H_n / \ln(n) \to 1$ et $H_n \sim \ln(n)$

Conclusion : $\boxed{E(T) = NH_N \sim N \ln(N) \text{ quand } N \to +\infty}$

Et comme $H_n = \ln(n) + u_n$ et avec $u_n - \gamma = \varepsilon(n) \to 0$ on a donc
$E(T) = NH_N = N(\ln(n) + \gamma + \varepsilon(N)) = N\ln(N) + N\gamma + N\varepsilon(N)$

Conclusion : $\boxed{E(T) = N\ln(N) + N\gamma + o(N)}$

(b) On a *vu* que $V(T) = N^2 \sum_{k=1}^{N} \frac{1}{k^2} - NH_N = N^2 \sum_{k=1}^{N} \frac{1}{k^2} - E(T)$
Donc

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

avec $\frac{E(T)}{N^2} = \frac{\ln(N)}{N} + \frac{\gamma}{N} + \frac{o(N)}{N^2} \to 0$ et la série $\sum_{k \geq 1} \frac{1}{k^2}$ convergente

Conclusion : $\boxed{\text{la suite } \left(\frac{V(T)}{N^2}\right)_{N \in \mathbb{N}^*} \text{ est convergente}}$

En notant $\alpha = \sum_{k=1}^{+\infty} \frac{1}{k^2}$ on a donc $\frac{V(T)}{N^2} \to \alpha$

Conclusion : $\boxed{V(T) \sim \alpha N^2 \text{ avec } \alpha = \sum_{k=1}^{+\infty} \frac{1}{k^2}}$

Enfin

$$V(T) - \alpha N^2 = -N^2 \sum_{k=N+1}^{+\infty} \frac{1}{k^2} - E(T) \leq 0$$

Conclusion : $\boxed{V(T) \leq \alpha N^2}$

11. Écart à la moyenne

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

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

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

(a) On a $T(\omega) - N\ln(N) = T(\omega) - E(T) + E(T) - N\ln(N)$ et par inégalité triangulaire

$$|T(\omega) - N\ln(N)| \leq |T(\omega) - E(T)| + |E(T) - N\ln(N)|$$

et $E(T) - N\ln(N) = NH_N - N\ln(N) = N(H_n - \ln(N))$ et on a vu que
$0 \leq u_n = H_n - \ln(N) \leq 1$ donc
$|E(T) - N\ln(N)| = E(T) - N\ln(N) \leq N$ et donc

Conclusion : $\boxed{|T(\omega) - N\ln(N)| \leq |T(\omega) - E(T)| + N}$

Donc, **si** $|T - N\ln(N)| \geq cN$ **alors** $|T - E(T)| + N \geq |T - N\ln(N)| \geq cN$
et $|T - N\ln(N)| \geq cN - N$

Conclusion : $\boxed{(|T - N\ln(N)| \geq cN) \subset (|T - E(T)| \geq N(c-1))}$

(b) D'après l'inégalité de Bienaymé&Tchebichev,

On a $\mathrm{P}\left(|T - E(T)| \geq (c-1)N\right) \leq \dfrac{V(T)}{(c-1)^2 N^2}$ et $V(T) \leq \alpha N^2$ (10.b)) donc

$\mathrm{P}\left(|T - E(T)| \geq (c-1)N\right) \leq \dfrac{\alpha}{(c-1)^2}$

et comme $(|T - N\ln(N)| \geq cN) \subset (|T - E(T)| \geq N(c-1))$ alors
$\mathrm{P}\left(|T - N\ln(N)| \geq cN\right) \leq \mathrm{P}\left(|T - E(T)| \geq N(c-1)\right) \leq \dfrac{\alpha}{(c-1)^2}$

Conclusion : $\boxed{\mathrm{P}\left(|T - N\ln(N)| \geq cN\right) \leq \dfrac{\alpha}{(c-1)^2}}$

Le nombre N étant fixé, $\dfrac{\alpha}{(c-1)^2} \to 0$ quand $c \to +\infty$ donc, par encadrement (une probabilité est positive)

Conclusion : $\boxed{\mathrm{P}\left(|T - N\ln(N)| \geq cN\right) \to 0 \text{ quand } c \to +\infty}$

12. Soit $\varepsilon > 0$.

La question précédente pousserait le professeur de mathématique à jouer avec les ε.

Plus simplement :

avec $c = \varepsilon\ln(N)$, on a, pour $N > \exp^{1/\varepsilon}$, $c > 1$ donc$\mathrm{P}\left(|T - N\ln(N)| \geq \varepsilon\ln(N)N\right) \leq \dfrac{\alpha}{(\varepsilon\ln(N) - 1)^2}$

Conclusion : $\boxed{\text{par encadrement, } \lim_{N\to+\infty} \mathrm{P}\left(|T - N\ln(N)| \geq \varepsilon\ln(N)N\right) = 0}$

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

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

Modélisation : On définit en `PASCAL` le `TYPE Paquet=ARRAY[1. .32] OF INTEGER;` Le paquet de 32 cartes est représenté par une variable `Jeu` de `TYPE Paquet` rempli initialement d'entiers entre 1 et 32 ; donc, initialement, `Jeu[i]` contient i, c'est à dire que la carte C_i est en position i. Au cours des insertions, `Jeu[i]` désigne le numéro de la carte en position numéro i. Par exemple, `Jeu [i] =10` signifie que la carte C_{10} est en position i.

On indique à la fin de cette question un extrait de programme à compléter en suivant les questions suivantes :

(a) Écrire la procédure `Init` permettant de définir une variable `Jeu` correspondant à la configuration initiale du paquet de cartes.

```python
import numpy as np
import numpy.random as rd

def Init():
        jeu=[1]
        for i in range(1,32):
                jeu.append(i+1)
        return(jeu)
```

(b) Compléter la procédure `Insertion` qui simule une opération d'insertion.

```python
import numpy as np
import numpy.random as rd

Jeu=Init()
def Insertion(Jeu):
        k=rd.randint(0,32)+1 # Position où va être insérée la carte du dessus
        Cartedessus=Jeu[0]
        if k>0:
                for i in range(k):
                        Jeu[i]=Jeu[i+1]

                Jeu[k]=Cartedessus
        return(Jeu)
```

(c)

```python
def T(Jeu):
        n=0
        while Jeu[0] != 31:
                Insertion(Jeu)
                n=n+1
        return(n+1)
```

Cette procédure effectue les insertions jusqu'à ce que la carte 32 (celle du dessous) se retrouve au dessus. `n` compte le nombre d'insertions nécessaires pour y parvenir.
Donc `T` simule T_{N-1} et non pas T ...piège !
C'est donc un de moins que le nombre d'insertions pour que le jeu soit battu.

(d)

```python
def mT():
        M=0
        for i in range(0,1000):
                Jeu=Init()
                M=M+T(Jeu)
        return(M/1000)
```

Partie 3 :

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 :

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

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

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

1. Soient A une partie de $\mathcal{S}_N$, $n \in \mathbb{N}^*$ et E_n l'événement : "à l'instant n le paquet de cartes se trouve dans une configuration qui appartient à la partie A."
   (a) Expliquer, en utilisant la question 7, l'égalité suivante : $\mathbf{P}_{(T\leq n)}(E_n) = \pi(A)$.
   En déduire $\mathbf{P}(E_n \cap (T \leq n)) = \pi(A)\,\mathbf{P}(T \leq n)$.
   (b) Établir que $\mathbf{P}(E_n \cap (T > n)) \leq \mathbf{P}(T > n)$.
   (c) Montrer que
   $$\mu_n(A) \leq \pi(A) + \mathbf{P}(T > n)$$
2. Soit A une partie de $\mathcal{S}_N$ et $n \in \mathbb{N}^*$. On note $\overline{A}$ l'événement contraire de A.

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

(b) Déduire des questions précédentes la majoration :

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

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

Partie 3 - Distance variationnelle à la loi uniforme

Notations :

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

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

- On note également μ_n la probabilité sur S_N définie comme suit :
  pour chaque configuration σ de 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 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 :

$$d(\mu_n, \pi) = \max\{|\mu_n(A) - \pi(A)|\ ,\ A \subset S_N\}$$

14. Soient A une partie de 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) A l'instant T, le paquet a été mélangé et toutes les configurations sont équiprobables.
Et pour les mêmes raisons, à tout instant ultérieur, les configurations seront toutes équiprobables.
Donc sachant l'instant $n \geq T$, la probabilité est l'équiprobable : π.
Conclusion : $\boxed{\mathrm{P}_{T\leq n}(E_n) = \pi(A)}$
Conclusion : $\boxed{\mathrm{P}(E_n \cap (T \leq n)) = \mathrm{P}(T \leq n)\,\mathrm{P}_{T\leq n}(E_n) = \pi(A)\,\mathrm{P}(T \leq n)}$

(b) Si $(E_n \cap T > n)$ alors $(T > n)$ donc $(E_n \cap T > n) \subset (T > n)$
Conclusion : $\boxed{\mathrm{P}(E_n \cap T > n) \leq \mathrm{P}(T > n)}$

(c) Pour tout $\sigma, \mu_n(\sigma) = \mathrm{P}(\text{´ } \sigma \text{ à l'instant } n \text{ ˇ})$ donc $\mu_n(A) = \mathrm{P}(E_n)$.
et comme $(T > n, T \leq n)$ est nu système complet d'événements,

$$\begin{aligned}\mathrm{P}(E_n) &= \mathrm{P}(E_n \cap (T \leq n)) + \mathrm{P}(E_n \cap T > n)\\ &= \pi(A)\,\mathrm{P}(T \leq n) + \mathrm{P}(E_n \cap T > n)\end{aligned}$$

avec $\mathrm{P}(T \leq n)$ donc $\pi(A)\,\mathrm{P}(T \leq n) \leq \pi(A)$ et $\mathrm{P}(E_n \cap T > n) \leq \mathrm{P}(T > n)$
Conclusion : $\boxed{\mu_n(A) \leq \pi(A) + \mathrm{P}(T > n)}$

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

(a) μ_n est une probabilité donc $\mu_n\left(\overline{A}\right) = 1 - \mu_n(A)$ et de même $\pi\left(\overline{A}\right) = 1 - \pi(A)$
Conclusion : $\boxed{\mu_n\left(\overline{A}\right) - \pi\left(\overline{A}\right) = \pi(A) - \mu_n(A)}$

(b) De $\mu_n(A) \leq \pi(A) + \mathrm{P}(T > n)$ on tire $\pi(A) - \mu_n(A) \geq -\mathrm{P}(T > n)$
l'inégalité précédente étant vraie pour toute partie A de S_N, elle l'est aussi pour $\overline{A}$
Donc $\pi\left(\overline{A}\right) - \mu_n\left(\overline{A}\right) \geq -\mathrm{P}(T > n)$ soit $\mu_n\left(\overline{A}\right) - \pi\left(\overline{A}\right) \leq \mathrm{P}(T > n)$
et comme d'autre part $\pi(A) - \mu_n(A) = \mu_n\left(\overline{A}\right) - \pi\left(\overline{A}\right)$ on a donc
$-\mathrm{P}(T > n) \leq \pi(A) - \mu_n(A) \leq \mathrm{P}(T > n)$ soit
Conclusion : $\boxed{|\pi(A) - \mu_n(A)| \leq \mathrm{P}(T > n)}$

16. Les quantités de $\{|\mu_n(A) - \pi(A)|\ ,\ A \subset S_N\}$ sont toutes comprises entre 0 et $\mathrm{P}(T > n)$.
Donc le maximum $d(\mu_n, \pi)$ également
Conclusion : $\boxed{0 \leq d(\mu_n, \pi) \leq \mathrm{P}(T > n)}$
T étant une variable aléatoire, sa fonction de répartition tend vers 1 en $+\infty$.
Et donc $\mathrm{P}(T > n) = 1 - \mathrm{P}(T \leq n) \to 1 - 1 = 0$ quand $n \to +\infty$
Conclusion : $\boxed{\text{par encadrement } \lim_{n\to+\infty} d(\mu_n, \pi) = 0}$

Partie 3 :

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

Dans cette partie, nous nous intéressons provisoirement à un collectionneur de timbres. Celui-ci reçoit chaque jour une lettre afIranchie 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.

1. Déterminer la loi de S_1.
2. Déterminer pour tout entier $k \in [[2, N]]$ la loi de la variable S_k.
3. En déduire que la variable S suit la même loi de probabilité que la variable T étudiée dans les parties précédentes.
   Ce résultat sera utilisé pour estimer la quantité $P(T > n)$.
4. Soit $m \in \mathbb{N}^*$.
   (a) Exprimer l'événement $(S > m)$ à l'aide des événements $B_1^m, B_2^m, \ldots, B_N^m$.
   (b) Que vaut $\mathbf{P}(B_j^m)$ pour tout entier $j \in [[1, N]]$?
   (c) On rappelle que pour tout entier $n \geq 2$ et pour toute famille d'événements $A_1, \ldots, A_n$, on a l'inégalité :
   $$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbf{P}(A_i)$$
   En déduire $\mathbf{P}(S > m) < N\left(1 - \frac{1}{N}\right)^m$
5.
   (a) Montrer que $\ln(1+x) \leq x$ pour tout $x \in \left]-1, +\infty\right[$
   (b) Déduire des résultats précédents la majoration
   $$\forall m \in \mathbb{N},\ \mathbf{P}(T > m) \leq N e^{-\frac{m}{N}}$$
6. On reprend les notations introduites dans la partie précédente.
   (a) Soit $c > 0$ fixé. Montrer que pour n entier supérieur ou égal à $N \ln N + cN$ on a : $d(\mu_n, \pi) \leq e^{-c}$.
   (b) *Application numérique.* On estime qu'une distance en variation à la loi uniforme de 0,2 est acceptable.
   Avec un jeu de 32 cartes, combien de battages par insertions doit-on faire pour considérer le paquet mélangé de façon acceptable ?

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

Dans cette partie, 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 1a 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 de $[[1, N]]$, 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. S_1 est le temps d'attente du premier timbre donc $S_1 = 1$, variable certaine.

18. Pour tout $k \in [[2, N]]$, S_k est le temps d'attente d'un timbre différent des $k-1$ qu'il possède déjà, avec une probabilité $\frac{N-(k-1)}{N}$ chaque jour, les arrivages étant indépendants.
Donc $S_k \hookrightarrow \mathcal{G}\left(1-\frac{k-1}{N}\right)$ loi de Δ_{N-k+1}
19. Donc $S = S_1+S_2+\cdots+S_N$ est une somme de variables indépendantes de même lois (en ordre inversés) que $T = \Delta_1+\cdots+\Delta_{N-1}+1$.
Conclusion : $\boxed{S \text{ suit la même loi que } T}$
Ce résultat sera utilisé pour estimer la quantité $\mathrm{P}(T > n)$.
20. Soit $m \in \mathbb{N}^*$
    (a) $(S > m)$ signifie que le $m^{ième}$ jour, le collectionneur n'a toujours pas les N timbres. Donc que l'un au moins des timbre n'a pas été reçu au jour m : $(S > m) = \bigcup_{j=1}^{N} \overline{B_j^m}$
    (b) Pour tout entier $j \in [[1, N]]$, $\overline{B_j^m}$ signifie que le timbre j n'a pas été reçu pendant les m jours (intersection d'événements indépendants)
    Donc $\mathrm{P}\left(\overline{B_j^m}\right) = \left(1-\frac{1}{N}\right)^m$ et $\mathrm{P}\left(B_j^m\right) = 1-\left(\frac{N-1}{N}\right)^m$
    (c) On rappelle que pour tout entier $n \geq 2$ et pour toute famille d'événements $A_1, \ldots, A_n$, on a l'inégalité : $\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathrm{P}(A_i)$.
    $(S > m) = \bigcup_{j=1}^{N} \overline{B_j^m}$ donc $\mathrm{P}(S > m) \leq \sum_{i=1}^{N} \mathrm{P}\left(\overline{B_j^m}\right) = N\left(1-\frac{1}{N}\right)^m$
21. (a) La fonction $x \to \ln(1+x)$ étant concave, elle sa courbe est en dessous de sa tangente en 0, d'équation $y = x$
    Conclusion : $\boxed{\ln(1+x) \leq x \text{ pour tout } x \in]-1, +\infty[}$
    Or
$$\left(1-\frac{1}{N}\right)^m = \exp\left(m\ln\left(1-\frac{1}{N}\right)\right) \text{ et}$$
$$\ln\left(1-\frac{1}{N}\right) \leq -\frac{1}{N} \text{ car } \frac{-1}{N} \in]-1, +\infty[\text{ donc}$$
$$\left(1-\frac{1}{N}\right)^m \leq \exp\left(-\frac{m}{N}\right) \text{ et}$$
$$\mathrm{P}(S > s) = N\left(1-\frac{1}{N}\right)^m \leq N\exp\left(-\frac{m}{N}\right)$$
    (b) Et comme S et T ont la même loi , $\mathrm{P}(T > s) = \mathrm{P}(S > s)$
    Conclusion : $\boxed{\mathrm{P}(T > m) \leq Ne^{-\frac{m}{N}}}$
22. On reprend les notations introduites dans la partie précédente.
    (a) Soit $c > 0$ fixé.
    On a vu (15.b) que $0 \leq d(\mu_n, \pi) \leq \mathrm{P}(T > n)$ pour tout entier $n \in \mathbb{N}^*$
    et à la question précédente $\mathrm{P}(T > n) \leq Ne^{-\frac{n}{N}}$ pour tout $n \in \mathbb{N}^*$
    donc $d(\mu_n, \pi) \leq Ne^{-\frac{n}{N}}$.
    Et pour tout entier $n \geq N\ln(N) + cN$ on a : $-\frac{n}{N} \leq -\ln(N) + c$
    d'où $e^{-\frac{n}{N}} \leq e^{-\ln(N)+c} = \frac{1}{N}e^{-c}$
    Conclusion : $\boxed{\begin{array}{l} d(\mu_n, \pi) \leq e^{-c} \\ \text{pour tout entier } n \geq N\ln(N) + cN \end{array}}$
    (b) Application numérique. On estime qu'une distance en variation à la loi uniforme de $0,2$ est acceptable.
    Avec un jeu de 32 cartes, combien de battages par insertions doit-on faire pour considérer le paquet mélangé de façon acceptable ?
    On veut $d(\mu_n, \pi) \leq 0,2$. Il suffit pour cela que $e^{-c} \leq 0,2$
    soit $c \geq -\ln(0,2) = -\ln\left(\frac{1}{5}\right) = \ln(5)$
    Il suffit donc de prendre $n \geq 32\ln(32) + 32c \geq 32(\ln(32) + \ln(5)) = 32\ln(160) \simeq 162$ pour que le paquet soit mélangé de façon acceptable.