

Concours Blanc n°1
Maths 2 - ESSEC2, 2011
30/11/2023
Durée : 4h

Mélange de jeux de cartes

Une question que se pose un joueur de cartes est de savoir combien de fois il est nécessaire de battre les cartes pour que le paquet soit convenablement mélangé. Ce problème décrit un procédé très élémentaire pour mélanger les cartes et propose de répondre alors à cette question.

Considérons un jeu de N cartes numérotées de C_1 à C_N et disposées en un paquet sur une table. Un joueur bat les cartes et repose le paquet sur la table. Le résultat du mélange est une permutation de ces N cartes.

Notations et rappel :

On note S_N l'ensemble des ordres possibles pour ce paquet de N cartes et on rappelle que $\text{Card}(S_N) = N!$. On se place dans un espace probabilisé (Ω, \mathcal{A}, P) avec $\Omega = S_N$, $\mathcal{A} = \mathcal{P}(S_N)$ l'ensemble des parties de S_N et P l'équiprobabilité sur Ω . Pour toute variable aléatoire X on notera $E(X)$ et $V(X)$ l'espérance et la variance de X lorsqu'elles existent.

On considère qu'un paquet est *convenablement mélangé* lorsque toutes les permutations sont équiprobables, c'est-à-dire lorsque pour tout ordre de S_N , la probabilité que le tas de cartes se trouve dans cet ordre vaut $\frac{1}{N!}$.

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 initialement dans l'ordre suivant : 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 donc tout en dessous du paquet.

Pour k élément de $\llbracket 1, N \rrbracket$, on appelle *insertion à la k-ième place* l'opération qui consiste à prendre la carte située au-dessus du paquet et à l'insérer entre la k -ième et la $(k+1)$ -ième place. Une insertion à la première place ne change pas l'ordre des cartes. Une insertion à la N -ième place consiste à faire glisser la carte située au-dessus du paquet pour la mettre sous le paquet.

Le *battage par insertions* du jeu de cartes consiste à effectuer une suite d'insertions aléatoires, en choisissant, à chaque instant, au hasard uniformément dans $\{1, \dots, N\}$ la place à laquelle l'insertion a lieu, indépendamment des insertions précédentes. Les instants successifs d'insertions seront notés $1, 2, \dots, n, \dots$; 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 la position $N-1$,
- T_2 le premier instant où la carte C_N se trouve remontée en position $N-2$,
- et plus généralement, pour 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 $\forall i \in \llbracket 2, N-1 \rrbracket$, $\Delta_i = T_i - T_{i-1}$.
- Enfin, on notera $T = T_{N-1} + 1$.

On admet que les conditions de l'expérience permettent de faire l'hypothèse que les variables aléatoires $(\Delta_i)_{i \in \llbracket 1, N-1 \rrbracket}$ 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 $k = \dots$							
		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 $T_1(\omega) = 3$, $T_2(\omega) = 5$, $T_3(\omega) = 6$ et $T(\omega) = 7$.

Partie 1 - Description et premiers résultats

- Justifier que $\forall i \in \llbracket 2, N-1 \rrbracket$, $T_i = \Delta_1 + \Delta_2 + \dots + \Delta_i$.
Que représente l'intervalle de temps Δ_i ?

- Lois des variables Δ_i .**

(a) Justifier que $\Delta_1 \hookrightarrow \mathcal{G}\left(\frac{1}{N}\right)$; puis que, pour tout $i \in \llbracket 1, N-1 \rrbracket$, $\Delta_i \hookrightarrow \mathcal{G}\left(\frac{i}{N}\right)$.

(b) En déduire, pour tout $i \in \llbracket 1, N-1 \rrbracket$, $E(\Delta_i) = \frac{N}{i}$ et $V(\Delta_i) = N \frac{N-i}{i^2}$.

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

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

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

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

- 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 :

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

- 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 :

- on obtienne $(a_1, a_2, a_3) = (N-2, N-1, N)$?
- on obtienne $(a_1, a_2, a_3) = (N-2, N, N-1)$?

- Justifier la phrase suivante :

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

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

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

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

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

7. Espérance et variance de T

- (a) Justifier que $E(T) = N H_N$.
 (b) On admet que si X_1, \dots, X_n sont des variables aléatoires indépendantes admettant des variances, alors $X_1 + \dots + X_n$ admet une variance, et on a :

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k)$$

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

8. Étude de la suite (u_n) .

- (a) Montrer : $\forall x > -1, \ln(1+x) \leq x$.
 (b) En déduire : $\forall k \in \mathbb{N}^*, \frac{1}{k+1} \leq \ln(k+1) - \ln(k) \leq \frac{1}{k}$.
 On pourra, pour l'une des deux inégalités, poser $x = -\frac{1}{k+1}$.
 (c) En déduire successivement :
 i. la décroissance de la suite (u_n) ,
 ii. l'encadrement : $\forall n \in \mathbb{N}^*, \ln(n+1) \leq H_n \leq \ln(n) + 1$.
 (d) Déduire de ce qui précède que la suite (u_n) est convergente et que sa limite, notée γ appartient à $[0, 1]$.

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

- (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).
 Déterminer une constante α strictement positive (à écrire sous forme d'une somme infinie), telle que $V(T) \underset{N \rightarrow +\infty}{\sim} \alpha N^2$. Montrer qu'alors $V(T) \leq \alpha N^2$.

10. Écart à la moyenne

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

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

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

- (a) Justifier que $\forall \omega \in \Omega, |T(\omega) - N \ln(N)| \leq |T(\omega) - E(T)| + N$ (on remarquera que $N \ln(N) = E(T) - N u_N$).
 Montrer l'inclusion entre événements :

$$(|T - N \ln(N)| \geq cN) \subset (|T - E(T)| \geq N(c-1))$$

- (b) Démontrer que :

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

où α a été définie à la question (9b).

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

11. Démontrer aussi que pour tout $\epsilon > 0$:

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

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 $32 \ln(32) \approx 110$, et pour un paquet de 52 cartes, $52 \ln(52) \approx 205$.

12. Simulation informatique

On veut implémenter cette méthode de mélange en Python. On suppose les imports usuels effectués :

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

On considère un paquet de 32 cartes ; on représente le paquet par une liste de 32 éléments, qui sont les numéros des cartes. Le premier élément de la liste est la carte du dessus ; le dernier est l'élément du bas de la pile. Initialement, le paquet sera donc représenté par la liste $[1, 2, 3, \dots, 31, 32]$.

Une insertion dans le paquet modifie cette liste ; si on reprend les insertions exposées dans le tableau de début de problème, pour un jeu de 4 cartes, la liste représentant le paquet vaudra successivement :

$[1, 2, 3, 4]$, $[2, 3, 1, 4]$, $[3, 2, 1, 4]$, $[2, 1, 4, 3]$, $[2, 1, 4, 3]$, $[1, 4, 2, 3]$, $[4, 2, 3, 1]$, $[2, 4, 3, 1]$

(a) Modélisation d'une insertion.

Compléter le programme suivant pour qu'il effectue une insertion de la carte du dessus à la position k .

(Attention : si le paquet est représenté par une liste L , alors la carte « à la position k » est $L[k-1]$).

```
def insertion(paquet, k):
    # k dans [1,n] : position où la nouvelle carte est insérée
    paquet_final=paquet
    carte_dessus=paquet_final[0]
    if k>1 :
        for i .... :
            paquet_final[i] = paquet_final[i+1]
        ....
    return paquet_final
```

(b) On rappelle que la comparaison « différent de » se note $!=$.

Compléter le programme suivant pour qu'il effectue un mélange suivant le protocole du problème, et renvoie le paquet mélangé, et le temps T nécessaire à ce mélange.

```
def temps_melange(N): # N : taille du paquet
    paquet= ... # etat initial [1,2,...,N]
    t=0
    while ..... :
        ....
    ...
    ...
    return paquet, t
```

(NB : le nombre de lignes n'est pas imposé).

On souhaite tester le caractère uniforme de ce mélange (c'est-à-dire que tous les ordres possibles du paquet apparaissent de manière équiprobable).

On se place dans le cas d'un paquet de 4 cartes.

(c) Le script suivant :

```
n=0
for k in range(100000):
    if temps_melange(4)[0] == [1,2,3,4]:
        n=n+1
print(n/100000)
print(1/24)
```

produit l'affichage :

```
0.04153
0.041666666666666664
```

Qu'en pensez-vous ?

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 :

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

- 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 , $\mu_n(A) = \sum_{\sigma \in A} \mu_n(\sigma)$.

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 :

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

13. 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 .» On a donc $\mu_n(A) = P(E_n)$.

- (a) Expliquer, en utilisant la question (6), l'égalité suivante : $P_{T \leq n}(E_n) = \pi(A)$.
En déduire $P(E_n \cap (T \leq n)) = \pi(A)P(T \leq n)$.

- (b) Établir que $P(E_n \cap (T > n)) \leq P(T > n)$.

- (c) Montrer que :

$$\mu_n(A) \leq \pi(A) + P(T > n)$$

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

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

- (b) Déduire des questions précédentes la majoration :

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

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

Partie 4- Une majoration de $P(T > n)$

Dans cette partie, nous nous intéressons provisoirement à un collectionneur de timbres. Celui-ci reçoit chaque jour une lettre affranchie avec un timbre choisi au hasard uniformément parmi les N timbres en vigueur. On étudie ici le nombre de jours que doit attendre le collectionneur pour posséder la collection complète des N timbres. 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.

16. Déterminer la loi de S_1 .

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

18. 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 $P(T > n)$.

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

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

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

(c) On rappelle que pour tout entier $n \geq 2$ et pour toute famille d'événements A_1, \dots, A_n , on a l'inégalité :

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i). \text{ En déduire } P(S > m) \leq N \left(1 - \frac{1}{N}\right)^m.$$

20. Déduire des résultats précédents et de la question (8a) la majoration :

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

21. 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 $\frac{1}{5}$ est acceptable.

Avec un jeu de 32 cartes, combien de *battages par insertions* doit-on faire pour considérer le paquet mélangé de façon acceptable ? (on donne $\ln 5 \simeq 1,6$ et $\ln(32) \simeq 3,4$).