



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2011

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

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 à cette question.

Considérons un jeu de N cartes, avec $N \geq 2$, numérotées de C_1 à C_N et disposées en un paquet sur une table. Le résultat d'un mélange est une permutation des N cartes. On note S_N l'ensemble des permutations possibles et on rappelle que $\text{Card}(S_N) = N!$.

On se place dans un espace probabilisé $(\Omega, \mathcal{A}, P)$ permettant de modéliser la suite des insertions aléatoires décrite ci-dessous. Pour toute variable aléatoire X , on notera $E(X)$ et $V(X)$ son espérance et sa variance lorsqu'elles existent.

On considère qu'un paquet est convenablement mélangé lorsque toutes ses permutations sont équiprobables, c'est-à-dire lorsque, pour toute permutation $\sigma \in S_N$, la probabilité que le tas de cartes se trouve dans la configuration σ vaut $1/N!$.

Vocabulaire et notations

Une carte située au sommet de la pile est dite en position numéro 1, celle immédiatement en dessous est en position numéro 2, etc. La carte située en position numéro N est donc celle du bas de la pile. On distinguera la position d'une carte du numéro qu'elle porte.

Partons d'un tas de cartes rangées initialement dans l'ordre suivant : pour tout $i \in \llbracket 1, N \rrbracket$, la carte C_i est en position i . La carte C_1 est donc au-dessus du paquet, tandis que C_N est en dessous.

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

Le battage par insertions consiste à effectuer une suite d'insertions aléatoires, en choisissant à chaque instant, uniformément dans $\llbracket 1, N \rrbracket$, la place de l'insertion, indépendamment des insertions précédentes. Les instants successifs d'insertion sont notés $1, 2, \dots, n, \dots$; l'instant initial est $n = 0$.

On note T_1 le premier instant où la carte du dessus est glissée en dernière position, c'est-à-dire le premier instant où C_N remonte de la position N à la position $N-1$. Plus

généralement, pour $i \in \llbracket 1, N-1 \rrbracket$, T_i est le premier instant où C_N atteint la position $N-i$. On pose $\Delta_1 = T_1$ et, pour $i \in \llbracket 2, N-1 \rrbracket$, $\Delta_i = T_i - T_{i-1}$. Enfin, on note $T = T_{N-1} + 1$.

On admet que les variables aléatoires $(\Delta_i)_{1 \leq i \leq N-1}$ sont indépendantes.

Description d'un exemple

Le tableau donne les résultats d'une expérience pour $N = 4$. La première ligne indique les instants, la deuxième les places d'insertion, et les suivantes la configuration du paquet.

Instant n	0	1	2	3	4	5	6	7
Insertion en place k	—	3	2	4	1	3	4	2
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, $T_1 = 3$, $T_2 = 5$, $T_3 = 6$ et $T = 7$.

Partie 1 — Description et premiers résultats

- 1) Justifier que, pour tout $i \in \llbracket 2, N-1 \rrbracket$, $T_i = \Delta_1 + \Delta_2 + \dots + \Delta_i$. Que représente l'intervalle de temps Δ_i ?
- 2) Loi de Δ_1 . Déterminer, pour tout entier $n \in \mathbb{N}$, $P(\Delta_1 > n)$ et reconnaître la loi de Δ_1 .
- 3) Soit $i \in \llbracket 2, N-1 \rrbracket$. Loi de Δ_i .
 - 3)(a) Établir que, pour tout $n \in \mathbb{N}$, $P(\Delta_i > n) = ((N-i)/N)^n$. En déduire que Δ_i suit

une loi usuelle que l'on précisera.

3)(b) En déduire $E(\Delta_i) = N/i$ et $V(\Delta_i) = N(N-i)/i^2$.

4) Loi de T_2 . Dans cette question, on suppose $N \geq 3$ et $n \geq 2$.

4)(a) Démontrer que

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

4)(b) Justifier que

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

4)(c) En déduire que $P(T_2 = n) = \frac{2}{N}[(1 - 1/N)^{n-1} - (1 - 2/N)^{n-1}]$.

5) On suppose $N \geq 3$. À l'instant T_2 , la carte C_N est en position $N - 2$ et deux cartes se trouvent sous elle, insérées aux instants T_1 et T_2 . Que valent les probabilités qu'à l'instant T_2 :

5)(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 ?

5)(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) On suppose $N \geq 4$. À l'instant T_3 , C_N est en position $N - 3$, et les trois cartes insérées aux instants T_1, T_2, T_3 se trouvent sous elle. Pour $i \in \{1, 2, 3\}$, on note a_i la position de la carte insérée à l'instant T_i .

6)(a) Combien y a-t-il de résultats possibles pour le triplet (a_1, a_2, a_3) ?

6)(b) Quelques exemples. Donner les probabilités qu'à l'instant T_3 :

6)(b)(i) on obtienne $(a_1, a_2, a_3) = (N - 2, N - 1, N)$;

6)(b)(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 pourra montrer par récurrence qu'à l'instant T_i , l'ordre des i cartes placées sous C_N est uniforme, même lorsque les durées d'attente et l'évolution des cartes au-dessus de C_N sont fixées. À l'insertion suivante faisant remonter C_N , les $i + 1$ positions possibles de la carte nouvellement placée sous C_N sont équiprobables. En déduire que, pour chaque entier t tel que $P(T = t) > 0$, la configuration à l'instant t , conditionnellement à $T = t$, est uniforme. Vérifier aussi qu'une insertion indépendante préserve une configuration uniforme.

Si l'on arrête le battage exactement à l'instant T , le paquet est donc convenablement mélangé. Cependant T est aléatoire : sans repérer C_N , on ne peut pas reconnaître cet instant en observant seulement le nombre d'insertions.

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

On introduit les suites $(H_n)_{n \geq 1}$ et $(u_n)_{n \geq 1}$ définies par

$$H_n = \sum_{k=1}^n 1/k$$

et $u_n = H_n - \ln n$.

8) Espérance et variance de T . Justifier $E(T) = NH_N$ et

$$V(T) = N^2 \sum_{k=1}^N 1/k^2 - NH_N.$$

9) Étude de la suite (u_n) .

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

9)(b) En déduire successivement :

9)(b)(i) la décroissance de la suite (u_n) ;

9)(b)(ii) l'encadrement $\ln(n+1) < H_n \leq \ln n + 1$ pour tout $n \geq 1$, l'inégalité de droite étant stricte pour $n \geq 2$.

9)(c) Dédurre de ce qui précède que (u_n) converge et que sa limite, notée γ , appartient à $[0, 1]$.

10)(a) Établir que, lorsque $N \rightarrow +\infty$, $E(T) \sim N \ln N$ et $E(T) = N \ln N + N\gamma + o(N)$.

10)(b) Quelle est la nature de la suite $(V(T)/N^2)_{N \geq 2}$? On prendra garde au fait que $V(T)$ dépend de N . Justifier qu'il existe une constante $\alpha > 0$ telle que $V(T) \sim \alpha N^2$ et $V(T) \leq \alpha N^2$.

11) Écart à la moyenne. On rappelle l'inégalité de Bienaymé-Tchebychev : pour une variable aléatoire X admettant un moment d'ordre deux et tout $\varepsilon > 0$, $P(|X - E(X)| \geq \varepsilon) \leq V(X)/\varepsilon^2$. Soit N fixé et une constante $c > 1$.

11)(a) Justifier que, pour toute issue ω , $|T(\omega) - N \ln N| \leq |T(\omega) - E(T)| + N$. Comparer par une inclusion les événements $[|T - N \ln N| \geq cN]$ et $[|T - E(T)| \geq N(c-1)]$.

11)(b) Démontrer que $P(|T - N \ln N| \geq cN) \leq \alpha/(c-1)^2$, où α est définie en 10)(b). Le nombre N étant fixé, que vaut $\lim_{c \rightarrow +\infty} P(|T - N \ln N| \geq cN)$?

12) Démontrer aussi que, pour tout $\varepsilon > 0$, $\lim_{N \rightarrow +\infty} P(|T - N \ln N| \geq \varepsilon N \ln N) = 0$.

Ainsi, un écart de T à $N \ln N$ de l'ordre de $N \ln N$ est asymptotiquement rare. Pour un paquet de 32 cartes, $32 \ln 32 \simeq 110$; pour un paquet de 52 cartes, $52 \ln 52 \simeq 205$.

13) Simulation informatique. Dans cette question, $N = 32$. Le paquet est représenté par une liste Python `jeu` de longueur 32. L'élément `jeu[i]`, pour i entre 0 et 31, est le

numéro de la carte située en position $i + 1$. Par exemple, `jeu[0] == 10` signifie que C_{10} se trouve en position 1. Initialement, `jeu[i] == i + 1`.

La commande `np.random.randint(0, 32)` renvoie un entier uniforme entre 0 et 31 inclus. Si k est cet entier, l'insertion est effectuée en position mathématique $k + 1$. La tranche `jeu[:] = ...` remplace tout le contenu de la liste en conservant la même liste. Une fonction sans instruction `return` peut modifier la liste reçue en argument.

On donne l'extrait suivant à compléter.

Python

```
import numpy as np

def init():
    ...

def insertion(jeu):
    k = ...
    carte_dessus = jeu[0]
    for i in range(k):
        jeu[i] = ...
    jeu[k] = ...

def temps(jeu):
    jeu[:] = init()
    n = 0
    while jeu[0] != 32:
        insertion(jeu)
        n = n + 1
    return n

# Programme principal à compléter.
```

13)(a) Écrire la fonction `init` permettant de définir une liste correspondant à la configuration initiale du paquet.

- 13)(b) Compléter la fonction insertion qui simule une opération d'insertion.
- 13)(c) Que fait la fonction temps ? Préciser le lien entre sa valeur renvoyée et la variable mathématique T .
- 13)(d) Écrire le programme principal permettant de calculer et d'afficher la moyenne des valeurs renvoyées par temps sur 100 expériences. Initialiser les variables nécessaires.

Partie 3 — Distance variationnelle à la loi uniforme

On note π l'équiprobabilité sur S_N : pour toute partie $A \subset S_N$, $\pi(A) = \text{Card}(A)/N!$. En particulier, pour toute permutation σ , $\pi(\{\sigma\}) = 1/N!$.

On note μ_n la loi de la configuration après n insertions. Ainsi, $\mu_n(\{\sigma\})$ est la probabilité que le paquet soit dans la configuration σ à l'instant n , et

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

On mesure la qualité du mélange à un instant donné par la distance

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

- 14)** Soient $A \subset S_N$, $n \in \mathbb{N}^*$ et E_n l'événement : «à l'instant n , le paquet se trouve dans une configuration appartenant à A ».

14)(a) Lorsque $P(T \leq n) > 0$, expliquer, en utilisant 7), l'égalité $P_{[T \leq n]}(E_n) = \pi(A)$. En déduire, pour tout $n \geq 1$, y compris lorsque $P(T \leq n) = 0$, $P(E_n \cap [T \leq n]) = \pi(A)P(T \leq n)$.

14)(b) Établir $P(E_n \cap [T > n]) \leq P(T > n)$.

14)(c) Montrer que $\mu_n(A) \leq \pi(A) + P(T > n)$.

- 15)** Soient $A \subset S_N$, $n \in \mathbb{N}^*$ et $\bar{A} = S_N \setminus A$.

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

15)(b) Dédurre des questions précédentes $|\mu_n(A) - \pi(A)| \leq P(T > n)$.

16) Montrer que, pour tout $n \geq 1$, $0 \leq d(\mu_n, \pi) \leq P(T > n)$. Déterminer $\lim_{n \rightarrow +\infty} d(\mu_n, \pi)$.

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

Un collectionneur reçoit chaque jour une lettre affranchie d'un timbre choisi au hasard uniformément parmi les N timbres en vigueur, indépendamment des jours précédents. On étudie le nombre de jours nécessaires pour posséder la collection complète. Le jour 0, il ne possède aucun timbre.

Pour $k \in \llbracket 1, N \rrbracket$, on note S_k le nombre de jours à attendre pour que le nombre de timbres différents possédés passe de $k - 1$ à k . On pose $S = S_1 + \dots + S_N$, nombre de jours nécessaires pour compléter la collection.

Les timbres étant numérotés de 1 à N , on note B_j^m l'événement : « au jour m , le collectionneur n'a encore jamais reçu le timbre numéro j ».

On admet que les variables aléatoires $(S_k)_{1 \leq k \leq N}$ sont indépendantes.

17) Déterminer la loi de S_1 .

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

19) En déduire que S a la même loi de probabilité que la variable T étudiée dans les parties précédentes. Ce résultat sera utilisé pour estimer $P(T > n)$.

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

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

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

20)(c) On rappelle que, pour tout entier $r \geq 2$ et toute famille d'événements $A_1, \dots, A_r$,

$$P\left(\bigcup_{i=1}^r A_i\right) \leq \sum_{i=1}^r P(A_i).$$

En déduire $P(S > m) \leq N(1 - 1/N)^m$.

21)(a) Montrer que $\ln(1 + x) \leq x$ pour tout $x \in]-1, +\infty[$.

21)(b) Dédurre des résultats précédents la majoration $P(T > m) \leq Ne^{-m/N}$, pour tout $m \in \mathbb{N}$.

22) On reprend les notations de la partie précédente.

22)(a) Soit $c > 0$. Montrer que, pour tout entier $n \geq N \ln N + cN$, $d(\mu_n, \pi) \leq e^{-c}$.

22)(b) Application numérique. On estime qu'une distance variationnelle à la loi uniforme de 0,2 est acceptable. Avec un jeu de 32 cartes, combien de battages par insertion doit-on faire pour considérer le paquet mélangé de façon acceptable ?