Combien de fois faut-il battre les cartes ?

Thomas Fernique

Sudislavl', 10-15 février 2015

1 Modélisation (40')

On numérote les cartes de 1 à n. Chaque état possible du jeu correspond donc à une permutation. On modélise alors le mélange américain comme suit (modèle de Gilbert-Reed-Shannon) :

Définition 1 (Mélange américain) *Tirer M tel que $\mathbb{P}(M = k) = C_n^k/2^n$ (nombre de "pile" sur n lancers[1]). Couper en deux paquets (les M premières et les $n-M$ dernières cartes). Entrelacer en faisant tomber les cartes, la probabilité qu'une carte tombe d'un paquet étant proportionnelle à la taille de ce paquet.*

Proposition 1 *À coupe donnée, les entrelacements possibles sont équiprobables.*

Preuve. Considérons un entrelacement σ obtenu après une coupe en M. Si p est la taille du premier paquet (et donc $k-p$ celle du deuxième paquet) quand on lâche la k-ème carte, alors c'est la bonne carte avec probabilité $p/(n-k)$ si elle doit venir du premier paquet, $(n-p-k)/(n-k)$ sinon. Au final, la probabilité d'obtenir σ est donc la fraction dont

- le dénominateur est le produit de $n, n-1, \dots, 1$ (on lâche toutes les cartes) ;
- le numérateur est le produit de $M, M-1, \dots, 1$ (on lâche toutes les cartes du premier paquet) et de $n-M, n-M-1, \dots, 1$ (on lâche toutes les cartes du second paquet).

C'est donc $M!(n-M)!/n! = 1/C_n^M$: les entrelacements sont équiprobables. □

Modélisé ainsi, le mélange américain est un processus probabiliste (X_t) sur l'espace des permutations à n éléments. Il s'agit même d'une *chaîne de Markov* : la probabilité d'aller en y à l'instant $t+1$ ne dépend que de l'état à l'instant t :

$$\mathbb{P}(X_{t+1} = y | X_t = y_t, \dots, X_0 = y_o) = \mathbb{P}(X_{t+1} = y | X_t = y_t).$$

1. Loi binômiale $B(n, 1/2)$. Rappel : $M \sim B(n,p)$ quand $\mathbb{P}(M = k) = C_n^k p^k (1-p)^k$.

Un tel processus est donc complètement décrit par la matrice P qui donne la probabilité d'aller d'un état à l'autre. La matrice du mélange américain se détermine facilement grâce à la notion de séquence montante :

Définition 2 (Séquence montante) *On appelle* séquence montante *de la permutation σ une suite maximale $i_1, \ldots, i_k$ telle que $\sigma(i_{j+1}) = \sigma(i_j)+1$, $1 \leq j < k$.*

Proposition 2 *Le mélange américain vérifie*

$$P(\sigma, \sigma') = \begin{cases} (n+1)/2^n & \text{si } \sigma' = \sigma, \\ 1/2^n & \text{si } \sigma' \circ \sigma^{-1} \text{ a exactement deux séquences montantes}, \\ 0 & \text{sinon.} \end{cases}$$

Preuve. Sans restriction de généralité, σ est la permutation identité. Alors σ' ne peut pas avoir plus de deux séquences montantes.
Si σ' a exactement deux séquences montantes, chacune provient d'un paquet et la longueur k de la coupe est donc caractérisée. Pour aller en σ' il a donc fallu couper en $M = k$ (probabilité $C_n^k/2^n$) et choisir le bon entrelacement (probabilité $1/C_n^k$), d'où une probabilité totale de $1/2^n$.
Si σ' est l'identité, c'est que les deux paquets ont juste été remis l'un sur l'autre après la coupe. On a donc pu couper pour n'importe quel k dans $\{0, \ldots, n\}$ (probabilité $C_n^k/2^n$ à chaque fois) puis il a fallu choisir le bon "entrelacement", *i.e.* celui qui n'entrelace rien du tout (probabilité $1/C_n^k$), d'où une probabilité totale de $(n+1)/2^n$. □

2 Expérience (30')

Répéter la procédure suivante en incrémentant à chaque fois t à partir de $t = 1$, jusqu'à ce que les avis soient à peu près partagés (ou que personne n'ait d'avis).

1. trier le jeu de carte par ordre croissant ou décroissant (non divulgué) ;
2. faire t passes de mélange américain ;
3. laisser les élèves examiner le jeu obtenu ;
4. les sonder sur l'ordre initial du paquet (croissant ou décroissant).

L'évolution du pourcentage d'avis majoritaire est censée refléter celle de la distance à la distribution stationnaire. Faire comprendre (par exemple en suivant les deux cartes du bas du paquet) que 4 passes sont très vraisemblablement insuffisantes pour mélanger un jeu de 52 cartes.

3 Convergence (40')

Considérons une chaîne de Markov très simple (Fig. 1). Supposons qu'elle soit initialement en A. Quelle est la probabilité qu'après k étapes elle soit encore en A ? Et à la limite ?

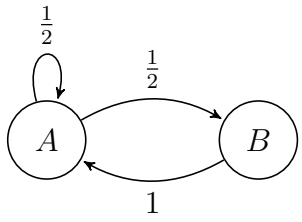

FIGURE 1 – Une chaîne de Markov.

Cet exemple montre qu'une chaîne de Markov agit sur les *distributions*. On a initialement une distribution de Dirac (tout en A) qui évolue au fur et à mesure. Dans ce cas, il y a convergence vers une distribution dite *stationnaire* au sens de la distance suivante, appelée *variation totale* :

$$||\mu - \nu|| := \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

De manière générale, on montre que si une chaîne de Markov est *irréductible* et *apériodique*[2], alors elle admet une unique distribution stationnaire π vers laquelle la convergence est exponentielle (quelle que soit la distribution initiale). Plus précisément, si la chaîne est décrite par la matrice P, on introduit

$$d(t) = \max_{x \in \Omega} ||P^t(x, \cdot) - \pi||,$$

$$\tau_{\text{mix}} = \min\{t \mid d(t) \leq 1/4\},$$

et on montre :

$$d(t) \leq 2^{-t/\tau_{\text{mix}}}.$$

La quantité τ_{mix} est appelé *temps de mélange* de la chaîne. Il correspond (en gros) au temps pour diviser par deux la distance à la distribution stationnaire (comme le temps de demi-vie d'un élément radioactif).

2. $\forall x, y \in \Omega,\ \exists t,\ P^t(x, y) > 0$ et $\text{pgcd}\{t \mid P^t(x, y) > 0\} = 1$.

Par exemple, le PageRank (original) de Google est la distribution stationnaire d'une marche aléatoire sur le graphe du Web. C'est en fait le temps moyen que passerait sur chaque page un internaute naviguant indéfiniment au hasard (théorème ergodique). La convergence est certes exponentielle en t, mais le temps de mélange dépend du nombre de pages web (plus de trois milliards). Qu'en est-il du mélange américain ?

4 Un temps stationnaire fort (40')

Il est plus commode d'étudier l'inverse du mélange américain :

Définition 3 (Mélange américain inversé) *Distribuer les cartes en deux paquets en lançant une pièce équilibré. Empiler le deuxième paquet sur le premier.*

C'est la chaîne inverse de celle du mélange américain : elle commence par "désentrelacer" en deux paquets, puis à "dé-couper" en superposant ces paquets. Irréductibilité, apériodicité, distribution stationnaire et temps de mélange sont préservés car tout est symétrique via $\sigma \to \sigma^{-1}$.

Définition 4 (Histoire) *La séquence de pile ou face qui a déterminé dans quel paquet une carte allait à chaque passe s'appelle son* histoire.

Lançons par exemple la chaîne du mélange américain inverse sur une séquence "abcdefgh" de huit cartes. Chaque ligne représente un pas de la chaîne, avec en majuscule les cartes labellées "pile" pour le pas suivant.

A	b	C	D	e	f	g	h
a	c	d	B	E	f	G	H
B	E	g	h	A	c	D	f
b	E	A	d	g	H	C	F

Après ce quatrième pas, les cartes ont toutes des histoires différentes :

AaAA bBBb CccC DdDd eEEE fffF gGgg hHhH

Le jeu obtenu, "eahcfbdg", se retrouve en remontant les histoires :

1. La dernière lettre montre que a,c,e,f,h ont été placées au dessus de b,g,d. On note acefh/bgd.
2. La troisième lettre donne ae/cfh et bd/g. Donc ae/cfh/bd/g.
3. La deuxième lettre donne e/a, h/cf et b/d. Donc e/a/h/cf/b/d/g.

4. La première lettre donne enfin c/f.

C'est en fait général :

Proposition 3 *Quand les histoires sont différentes, elles caractérisent le jeu.*

Au moment T où les histoires sont différentes, on a alors la même probabilité d'avoir n'importe quel jeu, *i.e.*, le mélange est parfait ! Considérons en effet deux cartes, disons i et j. Leurs histoires ayant été choisies uniformément au hasard parmi les 2^T histoires possibles, i a autant de chance d'être au dessus de j qu'en dessous. Cela valant pour toute paire de cartes, on a bien autant de chances d'avoir une permutation qu'une autre. [3]

La variable aléatoire T est ce qu'on appelle un *temps stationnaire fort* pour le mélange américain inverse. Plus généralement, étant donnée une chaîne de Markov (X_t) de distribution stationnaire π, une variable aléatoire T telle que

– l'évènement $\{T = t\}$ ne dépende que de $X_0,\ldots,X_t$,

– $\mathbb{P}(X_t = x |\ t \geq T) = \pi(x)$,

est appelée un *temps stationnaire fort* pour cette chaîne. Il donne un algorithme d'*échantillonage parfait*. Son étude permet aussi de borner le temps de mélange grâce au résultat suivant (relativement simple mais ici admis) :

Proposition 4 *Si T est un temps stationnaire fort, alors*

$$d(t) \leq \mathbb{P}(t < T).$$

5 Première analyse (15')

Proposition 5 *Le mélange américain vérifie pour n assez grand*

$$\tau_{mix} \leq 2\log_2(4n/3).$$

Preuve. Après t passe, on a 2^t histoires possibles. La probabilité que la $k^{\text{ème}}$ carte ait une histoire différente de celles des $k-1$ premières cartes est $1 - k/2^t$. La probabilité que toutes les histoires soient différentes après t passes est donc

$$\mathbb{P}(t \geq T) = \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t}\right).$$

3. Et ce, bien que deux ensembles d'histoires puissent caractériser un seul et même jeu (il suffit par exemple de rajouter des passes où on écrit la même chose sur toutes les cartes).

Soit alors c tel que $2^t = n^2/c^2$. Pour tout $k \leq n$, on a $c^2k/n^2 = o(1)$. Donc en utilisant $\log(1+x) = x + O(x^2)$ avec $x = c^2k/n^2 = o(1)$, on calcule

$$\begin{aligned}
\log \mathbb{P}(t \geq T) &= \sum_{k=0}^{n-1} \log\left(1 - \frac{k}{2^t}\right) \\
&= -\sum_{k=0}^{n-1} \left(\frac{c^2 k}{n^2} + O\left(\frac{c^4 k^2}{n^4}\right)\right) \\
&= -\frac{c^2 n(n-1)}{2n^2} + O\left(\frac{c^4 (n-1)n(2n-1)}{6n^4}\right). \\
&= -\frac{c^2}{2} + O\left(\frac{1}{n}\right).
\end{aligned}$$

Or, d'après la proposition 4 :

$$d(t) \leq \mathbb{P}(t < T) = 1 - \mathbb{P}(t \geq T).$$

Donc pour c tel que $1 - \exp(-c^2/2) < 1/4$, τ_{mix} est borné par $t = 2\log_2(n/c)$ pour n assez grand. Avec $c = 3/4$ on obtient la borne annoncée. □

Pour $n = 52$, on obtient un peu plus de 12 (en supposant n assez grand).

Proposition 6 *Soit $\delta > 0$. Le mélange américain vérifie pour n assez grand*

$$\tau_{mix} \geq (1 - \delta)\log_2(n).$$

Preuve. Soit Ω_x^t l'ensemble des permutations atteignables à partir de x par au plus t passes. Comme à chaque passe on peut atteindre au plus 2^n nouvelles permutations (selon les n lancers de pièce), on a $|\Omega_x^t| \leq 2^{nt}$. D'où :

$$d(t) = \max_{x \in \Omega} ||P^t(x, \cdot) - \pi|| \geq P^t(x, \Omega_x^t) - \pi(\Omega_x^t) = 1 - \frac{|\Omega_x^t|}{|\Omega|} \geq 1 - \frac{2^{nt}}{n!}.$$

Donc $d(t) > \frac{1}{4}$, *i.e.*, $t < \tau_{\text{mix}}$, tant que $1 - \frac{2^{nt}}{n!} > \frac{1}{4}$. On en déduit la borne annoncée via la formule de Stirling, qui donne $\ln(n!) \sim n\ln(n)$. □

Pour $n = 52$ on a $1 - \frac{2^{nt}}{n!} > \frac{1}{4}$ pour $t \leq 4$ passes.

6 Avec plus de mains (30')

Le mélange américain inversé admet une généralisation simple qui sera utile :

Définition 5 (a-mélange américain inversé) *Distribuer les cartes en a paquets en lançant un dé équilibré à a faces. Empiler du premier au dernier paquet.*

Le a-mélange américain (non-inversé) correspond alors au mélange américain que ferait un être doté de a mains ! Plus précisément, cela revient à couper en a paquets selon la loi multinomiale équilibrée [4] puis à entrelacer en relâchant les cartes en dessous de chaque paquet avec une probabilité proportionnelle à la taille de ce paquet (ou, de manière équivalente, à choisir uniformément au hasard parmi tous les entrelacements possibles).

Proposition 7 *Faire un a- suivi d'un b-mélange américain inversé équivaut, en terme de probabilité, à faire un seul et unique ab-mélange américain inversé.*

Preuve. Faire un a-mélange américain inversé, c'est écrire sur chaque $i^{\text{ème}}$ carte un chiffre $x_i \in \{0, \dots, a-1\}$ puis à trier le paquet selon les x_i. Faire ensuite un b-mélange américain inversé, c'est écrire sur chaque $j^{\text{ème}}$ carte un chiffre $y_j \in \{0, \dots b-1\}$ à gauche du chiffre déjà inscrit - qui est $x_{\sigma_a(i)}$ où σ_a est la permutation réalisée par le a-mélange - puis trier selon les y_j. Cela revient donc à écrire sur chaque $i^{\text{ème}}$ carte du paquet initial la concaténation $y_i \cdot x_{\sigma_a(i)}$ puis à trier par ordre lexicographique. Or les concaténations possibles sont en bijection avec $\{0, \dots, ab-1\}$ via

$$y_i \cdot x_{\sigma_a(i)} \mapsto z_i := ay_i + x_{\sigma_a(i)}.$$

Cela revient donc en fait à écrire sur chaque $i^{\text{ème}}$ carte du paquet initial un chiffre $z_i \in \{0, \dots, ab-1\}$ puis à trier le paquet selon les z_i, c'est-à-dire à faire précisément un ab-mélange américain inversé. □

Proposition 8 *La probabilité d'obtenir par un a-mélange américain inversé une permutation donnée ayant r séquences montantes est C^n_{n+a-r}/a^n.*

Preuve. Une fois les coupes définies, il y a une unique façon d'entrelacer les cartes pour obtenir la permutation donnée. Or il y a $r-1$ coupes imposées par les séquences montantes (chacune entre la dernière carte d'une séquence montante et la première de la suivante). Il y a donc $a-r$ coupes libres, à placer dans

4. Rappel : $N \sim M(n, (p_1, \dots, p_a))$ quand $\mathbb{P}(N_1 = n_1, \dots, N_a = n_a) = \frac{n!}{n_1! \cdots n_a!} p_1^{n_1} \cdots p_a^{n_a}$.

Id	x	σ_3	x_{σ_3}	y	$\sigma_4 \circ \sigma_3$
1	1	1	1	4	3
2	1	2	1	3	5
3	2	5	1	2	8
4	3	3	2	1	2
5	1	8	2	2	6
6	3	4	3	4	7
7	3	6	3	3	1
8	2	7	3	3	4

Id	z	σ_{12}
1	10	3
2	7	5
3	2	8
4	12	2
5	4	6
6	9	7
7	9	1
8	5	4

FIGURE 2 – Illustration de la proposition 7 avec $a = 3$ et $b = 4$.

$n + (a - r)$ cases (n cartes et $a - r$ coupes), soit $C^{a-r}_{n+a-r} = C^n_{n+a-r}$ possibilités. Comme il y a au total a^n façons d'étiqueter les n cartes avec un dé à a faces, on a le résultat annoncé. □

La probabilité d'obtenir par un 12-mélange américain inversé la permutation σ_{12} représentée Fig. 2, qui a 5 séquences montantes, est donc

$$C^n_{n+a-r}/a^n = C^8_{8+12-5}/12^8 \simeq 0,0000145.$$

7 Deuxième analyse (15')

D'après la proposition 7, faire t passes du mélange américain inversé revient donc à faire un seul et unique 2^t-mélange américain inversé. D'après la proposition 8, la probabilité d'obtenir alors une permutation avec r séquences montantes est $C^n_{2^t+n-r}/2^{nt}$. En notant $A_{n,r}$ le nombre de permutations à n éléments avec r séquences montantes, on a donc

$$d(t) = \frac{1}{2} \sum_{r=1}^{n} A_{n,r} \left| \frac{C^n_{2^t+n-r}}{2^{nt}} - \frac{1}{n!} \right|.$$

Proposition 9 *Les $A_{n,r}$ satisfont la relation de récurrence*

$$A_{n,1} = 1 \qquad et \qquad A_{n,r} = r^n - \sum_{j=1}^{r-1} C^n_{n+r-j} A_{n,j}.$$

Preuve. Le terme r^n est le nombre de séquences à n éléments et au plus r montées, qui sont obtenues par $r - 1$ coupes (chaque élément est codé par sa

coupe). Par exemple, 35826714 est obtenu par 1/2/34/567/8 et se code donc par 34524413. Il faut alors retrancher le nombre de séquences avec $j < r$ montées, ce qui se fait comme dans la preuve de la proposition. 8. □

Cela permet le calcul effectif des $A_{n,r}$ et donc de la distance à la distribution uniforme après t passes du mélange américain (voir aussi Fig. 7) :

$t \leq 4$	5	6	7	8	9	10	11	12
1.000	.9237	.6135	.3341	.1672	.0854	.0429	.0215	.0108

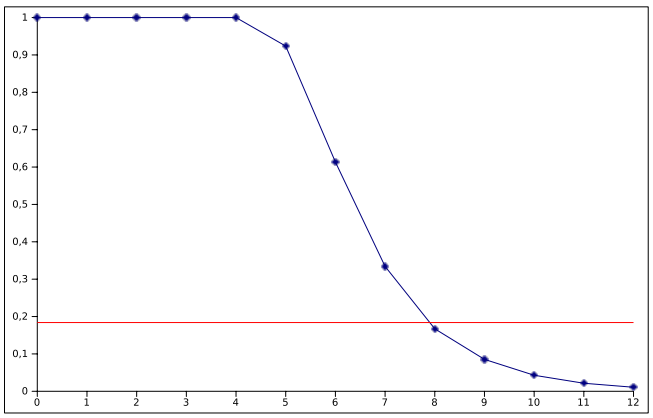

FIGURE 3 – Mélange américain de 52 de cartes : graphe de $t \to d(t)$.

Pour 52 cartes, il faut donc 8 passes (4 pour un singe – il suffit de remplacer 2 par 4 dans la formule de $d(t)$). En travaillant plus, on peut en déduire une estimation fine du temps de mélange quand le nombre n de cartes devient grand :

Théorème 1 (Bayer-Diaconis, 1992) *On a* $\tau_{mix} = \Theta(\frac{3}{2}\log_2(n))$ *avec* cutoff *:*

$$\lim_{|\Omega|\to\infty} d\left(c\frac{3}{2}\log_2(n)\right) = \begin{cases} 1 & si\ c < 1, \\ 0 & si\ c > 1. \end{cases}$$

Références

[1] Bayer, D. and Diaconis, P., *Trailing the dovetail shuffle to its lair*, Annales of applied probability **2** (1992), pp. 294–313.

[2] Levin, D. A., Peres, Y. and Wilmer, E. L., *Markov chains and mixing times*, American Mathematical Society, 371p., 2009.

[3] Mann, B., *How many times should you shuffle a deck of cards ?*, in Topics in Contemporary Probability and its Applications, pp. 261-289. Ed. by J. Laurie Snell, CRC press, 1995.