DVP De la manière de battre les cartes en Amérique

Antoine Dequay

21 septembre 2022

Notes

— Prof : .

— Leçons : 262, 264.

— Références :

  — Aigner, Ziegler, *Raisonnements divins.*

Proposition 1 Paradoxe des anniversaires : avec $K = 365$, la probabilité que deux personnes d'un groupe de n personnes soient nées le même jour est de $p(n, K) = 1 - \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right)$.

Preuve. Passer à l'évènement contraire, faire les hypothèses classiques et traduire le fait qu'on veut mettre les n personnes dans K boîtes avec au plus 1 personne par boîtes.

□

Définition 2 *Mesure de variation totale.* On définit une distance (mesure de variation totale) sur les distributions de probabilité définies sur $\mathfrak{S}_n$:

$$\|Q_1 - Q_2\| := \frac{1}{2} \sum_{\pi \in \mathfrak{S}_n} |Q_1(\pi) - Q_2(\pi)|.$$

On a alors

$$\|Q_1 - Q_2\| = \max_{S \subseteq \mathfrak{S}_n} |Q_1(S) - Q_2(S)|$$

avec $Q(S) := \sum_{\pi \in S} Q(\pi)$.

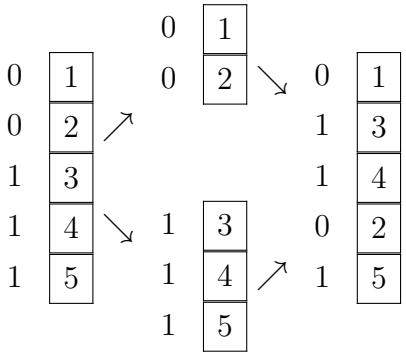
Preuve. Avec $S_1 = \{\pi \in \mathfrak{S}_n, Q_1(\pi) > Q_2(\pi)\}$, on a $\max_{S \subseteq \mathfrak{S}_n} |Q_1(S) - Q_2(S)| = |Q_1(S_1) - Q_2(S_1)|$, et :

$$\begin{aligned}
\|Q_1 - Q_2\| &= \frac{1}{2} \left(\sum_{\pi \in S_1} Q_1(\pi) - Q_2(\pi) + \sum_{\pi \in \mathfrak{S}_n \setminus S_1} Q_2(\pi) - Q_1(\pi) \right) \\
&= \frac{1}{2} \left(\sum_{\pi \in S_1} Q_1(\pi) - Q_2(\pi) + 1 - 1 - \sum_{\pi \in S_1} Q_2(\pi) - Q_1(\pi) \right) \\
&= \sum_{\pi \in S_1} Q_1(\pi) - Q_2(\pi) = |Q_1(S_1) - Q_2(S_1)|
\end{aligned}$$

□

Remarque Dans la suite, être "presque aléatoire" voudra dire "être à petite distance de la distribution uniforme" $U : \pi \mapsto \frac{1}{n!}$.

Définition 3 *Mélange à l'américaine.* On représente un jeu de cartes par l'ensemble $[\![1, n]\!]$ (avec $n = 52$ dans la pratique). Un mélange quelconque est alors une permutation $\pi \in \mathfrak{S}_n$. Plus spécifiquement, un mélange à l'américaine est une permutation telle que la séquence $(\pi(i))_{i \in [\![1,n]\!]}$ puisse se décomposer en 2 sous-séquences croissantes entrelacées (sauf dans le cas de la permutation identité, où on a une unique séquence) :



Proposition 4 On a $2^n - n$ mélanges possibles

Preuve. Pour $t \in [\![0, n]\!]$ cartes marquées à 0, on a $\binom{n}{t}$ possibilités pour choisir leurs places dans les n cases finales. Une fois les cases réservées, on ne peut choisir leur ordre. Comme pour chaque t, on peut avoir une fois la permutation identité (et que toutes les autres sont distinctes, il vient bien $\sum_{t=0}^{n} \left(\binom{n}{t} - 1 \right) + 1 = 2^n - n$ possibilités.

□

Définition 5 *Critère d'arrêt uniforme fort.* On appelle critère d'arrêt uniforme fort un critère d'arrêt ne dépendant que des opérations effectuées aléatoirement sur le jeu de cartes, tel que pour tout $k \in \mathbb{N}$:

Si le processus est interrompu après exactement k étapes, alors les permutations associées à l'ordre potentiel des cartes ont une distribution uniforme.

Lemme 6 On note Q^{*k} la distribution obtenue après k mélanges décrits par une distribution de probabilité $Q : \mathfrak{S}_n \to \mathbb{R}$ et on se donne un critère d'arrêt uniforme fort de temps d'arrêt T. Alors on a, pour $k \in \mathbb{N}$:

$$\left\| Q^{*k} - U \right\| \leq \mathbb{P}[T > k].$$

Preuve. Pour $X \sim Q$, on note $Q(S) = \mathbb{P}[X \in S]$. Pour $Q = U$, on a $U(S) = \mathbb{P}[X \in S] = \dfrac{|S|}{n!}$, et plus généralement :

$$\begin{aligned}
Q^{*k}(S) = \mathbb{P}[X_k \in S] &= \sum_{j \leq k} \mathbb{P}[X_k \in S \wedge T = j] + \mathbb{P}[X_k \in S \wedge T > j] \text{ par les probas totales,} \\
&= \sum_{j \leq k} U(S)\mathbb{P}[T = j] + \mathbb{P}[X_k \in S | T > k]\mathbb{P}[T > k] \\
&= U(S)(1 - \mathbb{P}[T > k]) + \mathbb{P}[X_k \in S | T > k] \cdot \mathbb{P}[T > k] \\
&= U(S) + \underbrace{(\mathbb{P}[X_k \in S | T > k] - U(S))}_{|\cdot| \leq 1} \cdot \mathbb{P}[T > k]
\end{aligned}$$

D'où :

$$\left| Q^{*k}(S) - U(S) \right| \leq \mathbb{P}[T > k].$$

□

Définition 7 Rif. Avec ce qui précède, on définit naturellement la distribution de probabilité suivante :

$$\text{Rif} : \left(\begin{array}{ccc} \mathfrak{S}_n & \longrightarrow & \mathbb{R} \\ \pi & \longmapsto & \begin{cases} \frac{n+1}{2^n} & \text{si } \pi = Id, \\ \frac{1}{2^n} & \text{si } \pi \text{ se décompose en 2 séquences croissantes,} \\ 0 & \text{sinon.} \end{cases} \end{array} \right).$$

Remarque On remarque que cette distribution de probabilité caractérise ces deux situations :

— On coupe le paquet en 2 partie avec une probabilité $\frac{1}{2^n}\binom{n}{t}$. On prend le premier paquet en main droite et le second en main gauche et on appelle respectivement d et g le nombre de cartes dans chaque main. Alors on définit la probabilité de poser la carte "suivante" (respectant l'ordre de départ) depuis la main droite comme étant $\frac{d}{d+g}$ et celle venant de la main gauche comme $\frac{g}{d+g}$. On recommence jusqu'à ce qu'il n'y ai plus de cartes (Attention, cette phrase est fausse mais utile : cela correspond à une version "uniforme" du mélange à l'américaine).

— Le cas du mélange à l'américaine inverse (dessiner en rouge les flèches dans l'autre sens) : On attribue un 0 ou un 1 à chaque cartes avec une probabilité $\frac{1}{2}$, puis on ramène tous les 0 au dessus des 1, en gardant l'ordre relatif. Cela correspond aux permutations croissantes, à l'exception d'une décroissance (qui n'existe pas dans le cas de l'identité).

En effet, On doit choisir $t \in [\![0, n]\!]$ carte assignées à 0 parmis n, puis les "remonter" (rien n'est changé si on choisit les t premières cartes).

Plus précisément, la distribution de probabilité est ici $\overline{\text{Rif}}(\pi) := \text{Rif}\left(\pi^{-1}\right)$.

On cherche à savoir à parti de combien de mélanges on peut considérer notre jeu de carte comme mélangé "presque aléatoirement".

Théorème 8 On a :

$$\left\|\text{Rif}^{*k} - U\right\| \leq 1 - \prod_{i=1}^{n-1}\left(1 - \frac{1}{2^k}\right).$$

Preuve. Comme $\pi \mapsto \pi^{-1}$ est une bijection de $\mathfrak{S}_n$ et que $U(\pi) = U\left(\pi^{-1}\right)$, on a le lemme de Reed :

$$\left\|\text{Rif}^{*k} - U\right\| = \left\|\overline{\text{Rif}}^{*k} - U\right\|.$$

A partir de maintenant, on raisonne donc sur les mélanges inverses.

A chaque mélange, on note sur chaque carte de droite à gauche ne numéros b_i affecté à celle-ci (avec i le numéro du mélange). On définit le critère d'arrêt suivant : Arrêter dès que les cartes ont des chaînes $b_k \dots b_1$ distinctes.

Les cartes sont alors triées en fonction des nombres binaires qui leurs sont associés.

Faire exemple en direct.

Le temps requis par le critère d'arrêt est alors distribué selon le paradoxe des anniversaires avec $K = 2^k$. En effet, les boîtes sont étiquetées par les nombres binaires à k chiffres, et on a donc $\mathbb{P}[T > k] = 1 - \prod i = 1^{n-1} \left(1 - \frac{1}{2^k}\right)$. D'où le résultat !

□

Remarque — Avec $k = 2\ln_2(cn)$ et $c \geq 1$, on a $\mathbb{P}[T > k] \simeq 1 - e^{-\frac{1}{2c^2}} \simeq \frac{1}{2c^2}$. Demander pourquoi. On aura donc besoin de plus de $2\ln_2(n)$ mélanges pour de grandes valeurs de n. (pour $n = 52$, cela fait environ 11) On passe sous les 0.28 à partir de $k = 12$, ce qui est "acceptable",

— Plus précisément, pour $k = 10$, le majorant vaut 0.043, ce qui est "acceptable",

— Penser à se souvenir de l'allure de la courbe,

— On a un résultat plus fort qui dit que c'est "acceptable" au bout de $k = 7$.