Cours sur les chaînes de Markov à Oran (Asselah, Chafaï)
Un exemple de convergence abrupte vers l'équilibre

Djalil Chafaï (Paris-Est Marne-la-Vallée)

Oran, mars 2011

Table des matières

1 Distance en variation

Dans toute cette section, E est un ensemble au plus dénombrable. L'ensemble des lois sur E est un espace métrique pour la distance en variation (on dit aussi *variation totale*)

$$d_V(\mu,\nu) = \sup_{A\subset E} |\mu(A)-\nu(A)|.$$

On a $0 \le d_V(\mu,\nu) \le 1$ et $d_V(\mu,\nu)=1$ si μ et ν ont des supports disjoints.

Théorème 1.1 (Expressions alternatives)**.** *Si μ et ν sont des lois sur E alors*

$$d_V(\mu,\nu) = \frac{1}{2}\sup_{f:E\to[-1,1]}\left|\int f\,d\mu - \int f\,d\nu\right| = \frac{1}{2}\sum_{x\in E}|\mu(x)-\nu(x)|$$

De plus, le supremum dans la définition de d_V est atteint pour l'ensemble

$$A_* = \{x\in E : \mu(x)\ge\nu(x)\}$$

tandis que dans l'expression variationnelle fonctionnelle de d_V il est atteint pour

$$f = \mathbf{1}_{A_*} - \mathbf{1}_{A_*^c}.$$

Démonstration. La seconde égalité provient de l'identité

$$\left|\int f\,d\mu - \int f\,d\nu\right| \le \sum_{x\in E}|f(x)||\mu(x)-\nu(x)| \le \sup_{x\in E}|f(x)|\sum_{x\in E}|\mu(x)-\nu(x)|$$

qui est saturée pour $f=\mathbf{1}_{A_*}-\mathbf{1}_{A_*^c}$. Pour la première égalité, on écrit

$$|\mu(A)-\nu(A)| = \frac{1}{2}\left|\int f_A\,d\mu - \int f_A\,d\nu\right|$$

où $f=\mathbf{1}_A-\mathbf{1}_{A^c}$, ce qui donne

$$|\mu(A)-\nu(A)| \le \frac{1}{2}\sup_{f:E\to[-1,1]}\left|\int f\,d\mu - \int f\,d\nu\right| = \frac{1}{2}\sum_{x\in E}|\mu(x)-\nu(x)|$$

qui est saturé pour $A = A_*$ car

$$2|\mu(A_*) - \nu(A_*)| = \sum_{x \in A_*} |\mu(x) - \nu(x)| + \sum_{x \in A_*^c} |\mu(x) - \nu(x)|.$$

□

Théorème 1.2 (Convergence en loi)**.** *Si (X_n) est une suite de variables aléatoires sur E et si μ_n désigne la loi de X_n, alors pour toute loi μ sur E, les propriétés suivantes sont équivalentes :*

1. $\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$ *pour toute fonction bornée* $f : E \to \mathbb{R}$
2. $\lim_{n\to\infty} \mu_n(x) = \mu(x)$ *pour tout* $x \in E$
3. $\lim_{n\to\infty} d_V(\mu_n, \mu) = 0$

Lorsqu'elles ont lieu on dit que (X_n) converge en loi vers μ quand $n \to \infty$.

Démonstration. Pour déduire *1.* de *3.* il suffit d'utiliser l'expression variationnelle fonctionnelle de d_V. Pour déduire *2.* de *1.* on peut prendre $f = \mathbf{1}_{\{x\}}$. Pour déduire *3.* de *2.* on observe que pour tout $A \subset E$,

$$\sum_{x \in E} |\mu_n(x) - \mu(x)| = \sum_{x \in A} |\mu_n(x) - \mu(x)| + \sum_{x \in A^c} |\mu_n(x) - \mu(x)|,$$

ensuite, grâce à *4.*, pour tout $\varepsilon' > 0$ il existe un entier $N = N(A, \varepsilon')$ tel que le premier terme du membre de droite est majoré par $\mathrm{card}(A)\varepsilon'$ pour tout $n \geq N$. Pour le second terme du membre de droite, on écrit

$$\sum_{x \in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x \in A^c} \mu_n(x) + \sum_{x \in A^c} \mu(x).$$

Puisqu'on a

$$\sum_{x \in A^c} \mu_n(x) = \sum_{x \in A} \mu(x) - \sum_{x \in A} \mu_n(x) + \sum_{x \in A^c} \mu(x)$$

on obtient

$$\sum_{x \in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x \in A} |\mu_n(x) - \mu(x)| + 2 \sum_{x \in A^c} \mu(x).$$

Puisque $\mu \in \mathcal{P}$, pour tout $\varepsilon'' > 0$, on peut choisir A fini tel que $\mu(A^c) \leq \varepsilon''$. □

Remarque 1.3 (Dispersion à l'∞)**.** *Si (μ_n) sont des lois et $\mu(x) := \lim_{n\to\infty} \mu_n(x)$ alors μ n'est pas forcément une loi, sauf si E est fini. En effet, lorsque E est infini, il peut se produire un phénomène de dispersion de la masse à l'infini. Contre exemple : $E = \mathbb{N}$ et μ_n affecte la masse $1/n$ aux singletons $\{1\}, \ldots, \{n\}$, ce qui donne μ identiquement nulle.*

Théorème 1.4 (Autre expression et cas extrémal)**.** *Si μ et ν sont des lois sur E alors*

$$d_V(\mu, \nu) = 1 - \sum_{x \in E} (\mu(x) \wedge \nu(x)).$$

En particulier, $d_V(\mu, \nu) = 1$ si et seulement si μ et ν ont des support disjonits.

Démonstration. Il suffit d'écrire

$$\sum_{x \in E} (\mu(x) \wedge \nu(x)) = \frac{1}{2} \sum_{x \in E} (\mu(x) + \nu(x) - |\mu(x) - \nu(x)|) = 1 - d_V(\mu, \nu).$$

□

Théorème 1.5 (Couplage)**.** *Si μ et ν sont des lois sur E alors*

$$d_V(\mu,\nu) = \inf_{(X,Y)} \mathbb{P}(X \neq Y)$$

où l'infimim porte sur les couples de v.a. sur $E \times E$ de lois marginales μ et ν. De plus, il existe un couple de ce type pour lequel l'égalité est atteinte (i.e. l'infimum est un minimum).

Démonstration. Soit (X,Y) est un couple de variables aléatoires sur $E \times E$ de lois marginales μ et ν. Comme $\mathbb{P}(X = x, Y = x) \leq \mu(x) \wedge \nu(x)$ pour tout $x \in E$ on a

$$1 - d_V(\mu,\nu) = \sum_{x \in E} (\mu(x) \wedge \nu(x)) \geq \sum_{x \in E} \mathbb{P}(X = x, Y = x) = \mathbb{P}(X = Y).$$

Il suffit donc de construire un couple (X,Y) pour lequel l'égalité est atteinte. Posons

$$p = 1 - d_V(\mu,\nu) \in [0,1].$$

Cas où $p = 0$. On a alors $d_V(\mu,\nu) = 1$ et μ et ν ont des supports disjoints. Cela donne $\mathbb{P}(X = Y) = \sum_{x \in E} \mu(x)\nu(x) = 0$. On prend (X,Y) avec $X \sim \mu$ et $Y \sim \nu$ indépendantes.
Cas où $p = 1$. On a alors $d_V(\mu,\nu) = 0$ et donc $\mu = \nu$. On prend (X,X) où $X \sim \mu$.
Cas où $0 < p < 1$. Soit (U,V,W) un triplet de variables aléatoire de lois respectives

$$p^{-1}(\mu \wedge \nu), \quad (1-p)^{-1}(\mu - (\mu \wedge \nu)), \quad (1-p)^{-1}(\nu - (\mu \wedge \nu)).$$

Notons que $p = \sum_{x \in E}(\mu(x) \wedge \nu(x))$. Soit B une variable aléatoire de Bernouill, indépendante de (U,V,W), telle que $\mathbb{P}(B = 1) = 1 - \mathbb{P}(B = 0) = p$. Définissons $(X,Y) = (U,U)$ si $B = 1$ et $(X,Y) = (V,W)$ si $B = 0$. On a alors $X \sim \mu$ et $Y \sim \nu$, et puisque les lois de V et W ont des supports disjoints, on a $\mathbb{P}(V = W) = 0$, et donc $\mathbb{P}(X = Y) = \mathbb{P}(B = 1) = p$. $\square$

Comme $\mathbb{P}(X \neq Y) = \mathbb{E}(d(X,Y))$ pour la distance atomique $d(x,y) = \mathbf{1}_{x \neq y}$, on a

$$d_V(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \int_{E \times E} d(x,y)\, d\pi(x,y)$$

où $\Pi(\mu,\nu)$ est l'ensemble des lois sur $E \times E$ de marginales μ et ν. Cela permet de voir d_V comme la distance de couplage de Wasserstein W_1 pour la distance atomique sur E.

2 Collectionneur de coupons

Le collectionneur de coupons constitue un modèle stochastique fondamental important, à ranger dans la même boite à outils que le jeu de pile ou face, auquel il est intimement relié. Un grand nombre de situations concrètes sont modélisables par le collectionneur de coupons ou une de ses variantes. Nous nous limitons ici à la variante la plus simple.

Il faut jouer un nombre de fois (aléatoire) géométrique à pile ou face pour voir apparaître les deux côtés de la pièce. Si on remplace la pièce de monnaie par un dé à $r \geq 2$ faces, combien de fois faut-il lancer le dé pour voir apparaître les r faces différentes ? On modélise cela, pour un entier fixé $r \geq 2$, en considérant la variable aléatoire

$$T = \min\{n \geq 1 : \{X_1, \ldots, X_n\} = \{1, \ldots, r\}\} = \min\{n \geq 1 : \mathrm{Card}\{X_1, \ldots, X_n\} = r\}$$

où $(X_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d. de loi uniforme sur $\{1, \ldots, r\}$. La variable aléatoire T est le temps de complétion de la collection. Le nom *collectionneur de coupons* provient des coupons à collectionner présents dans certains paquets de céréales.

Théorème 2.1 (Expression combinatoire de la loi)**.** *On a $T \geq r$ et pour tout $n \geq r$,*

$$\mathbb{P}(T = n) = \frac{r!}{r^n} \left\{ \begin{matrix} n-1 \\ r-1 \end{matrix} \right\}$$

où la notation en accolades désigne le nombre de Stirling de seconde espèce, c'est-à-dire le nombre de manières de partitionner $n-1$ objets en $r-1$ paquets non vides.

Démonstration. On a $X_T \notin \{X_1, \ldots, X_{T-1}\}$ car le coupon qui termine la collection n'a forcément jamais été vu auparavant. Si on fixe $n \geq r$, l'événement $\{T = n\}$ correspond à choisir le type du dernier coupon puis à répartir les $n-1$ coupons restants sur les $r-1$ types restants. Le résultat désiré en découle car la loi des type est uniforme. □

Le théorème 2.1 n'est malgré tout pas très parlant. Le résultat intuitif suivant va beaucoup nous aider à étudier T, et montre en particulier que $\mathbb{P}(T < \infty) = 1$.

Lemme 2.2 (Décomposition)**.** *On a $T = G_1 + \cdots + G_r$ où $G_1, \ldots, G_r$ sont des v.a. indépendantes, géométriques sur $\{1, 2, \ldots\}$, avec, pour tout $1 \leq i \leq r$, G_i de paramètre*

$$\pi_i := \frac{r - i + 1}{r}.$$

Démonstration. On pose $G_1 \equiv 1$ et pour tout $1 < i \leq r$,

$$G_i = \min\{n \geq 1 : X_{G_{i-1}+n} \notin \{X_1, \ldots, X_{G_{i-1}}\}\}.$$

On a $\mathrm{Card}(\{X_1, \ldots, X_{G_i}\}) = i$ pour tout $1 \leq i \leq n$. Les variables aléatoires $G_1, G_1 + G_2, \ldots, G_1 + \cdots + G_r$ sont les temps d'apparition des r premiers gains dans un jeu de pile ou face spécial dans lequel la probabilité de gagner change après chaque gain : cette probabilité vaut successivement $1, (r-1)/r, (r-2)/r, \ldots, 1/r$. Cela témoigne du fait qu'il est de plus en plus difficile d'obtenir un coupon d'un nouveau type au fil de la collection. □

Théorème 2.3 (Principle d'inclusion-exclusion)**.** *Pour tous $A_1, \ldots, A_r \in \mathcal{F}$*

$$\mathbb{P}(\cup_{1 \leq i \leq r} A_i) = \sum_{k=1}^{r} (-1)^{k+1} S_k \quad \text{où} \quad S_k = \sum_{1 \leq i_1 < \cdots < i_k \leq r} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_r}).$$

Pour $r = 2$, on retrouve $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$, et pour $r = 3$,

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) \\ - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Le principe d'inclusion-exclusion est rarement utilisé pour $r > 2$. Il l'est cependant dans l'étude de la fluctuation asymptotique du collectionneur de coupons (théorème 2.6).

Démonstration. On procède par récurrence sur r, en observant que

$$\begin{aligned} \mathbb{P}(\cup_{1 \leq i \leq r+1} A_i) &= \mathbb{P}(\cup_{1 \leq i \leq r} A_i) + \mathbb{P}(A_{r+1}) - \mathbb{P}((\cup_{1 \leq i \leq r} A_i) \cap A_{r+1}) \\ &= \mathbb{P}(\cup_{1 \leq i \leq r} A_i) + \mathbb{P}(A_{r+1}) - \mathbb{P}(\cup_{1 \leq i \leq r} (A_i \cap A_{r+1})) \end{aligned}$$

ce qui permet d'utiliser l'hypothèse de récurrence (pour le premier et dernier terme). □

Théorème 2.4 (Autre formule pour queue de distribution)**.** *Pour tout $n \geq 1$,*

$$\mathbb{P}(T > n) = \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k} \Big(1 - \frac{k}{r}\Big)^n.$$

Démonstration. On a

$$\{T > n\} = E_{n,1} \cup \cdots \cup E_{n,r} \quad \text{où} \quad E_{n,i} := \{X_1 \neq i, \ldots, X_n \neq i\}.$$

Si $i_1, \ldots, i_k$ sont des éléments distincts de $\{1, \ldots, r\}$ alors, avec $R := \{1, \ldots, n\} \backslash \{i_1, \ldots, i_k\}$,

$$\mathbb{P}(E_{n,i_1} \cap \cdots \cap E_{n,i_k}) = \mathbb{P}(X_1 \in R) \cdots \mathbb{P}(X_n \in R) = \Big(\frac{r-k}{r}\Big)^n = \Big(1 - \frac{k}{r}\Big)^n.$$

Le résultat désiré découle alors du principe d'inclusion-exclusion. □

Théorème 2.5 (Déviation)**.** *Pour tout réel $t > 0$,*

$$\mathbb{P}(T > 1 + \lfloor tr \log(r) \rfloor) \leq r^{-t+1}.$$

Démonstration. Pour tout entier $n \geq 1$, on peut écrire

$$\mathbb{P}(T > n) = \mathbb{P}(\cup_{i=1}^{r} E_{n,i}) \leq \sum_{i=1}^{r} \mathbb{P}(E_{n,i}) \quad \text{où} \quad E_{n,i} = \{X_1 \neq i, \ldots, X_n \neq i\}.$$

Comme $\mathbb{P}(E_{n,i}) = (1 - 1/r)^n \leq e^{-n/r}$, le choix $n = 1 + \lfloor tr \log(r) \rfloor$ mène au résultat. □

À présent, pour $\alpha = 0.05$ et r fixé, on peut choisir t assez grand pour que $r^{-t+1} \leq \alpha$, ce qui fournit l'intervalle de confiance $[r, \lfloor tr \log(r) \rfloor + 1]$ de niveau α.

Par le lemme 2.2 et la linéarité de l'espérance, on a

$$\mathbb{E}(T) = \sum_{i=1}^{r} \mathbb{E}(G_i) = \sum_{i=1}^{r} \frac{1}{\pi_i} = \sum_{i=1}^{r} \frac{r}{r-i+1} = r \sum_{i=1}^{r} \frac{1}{i} = r(\log(r) + \gamma + o_{r\to\infty}(1))$$

où $\gamma = \lim_{n\to\infty}(\sum_{i=1}^{n} 1/i - \log(n)) \approx 0.577$ est la constante d'Euler. Comme les $G_1, \ldots, G_r$ sont indépendantes avec $\mathrm{Var}(G_i) = (1 - \pi_i)/{\pi_i}^2 = r(i-1)/(r-i+1)^2$, on a également

$$\mathrm{Var}(T) = \sum_{i=1}^{r} \mathrm{Var}(G_i) = r \sum_{i=1}^{r-1} \frac{r-i}{i^2} = \frac{\pi^2}{6} r^2 - r\log(r) - r + o_{r\to\infty} r^2.$$

L'inégalité de Markov donne alors pour tout $t > 0$,

$$\mathbb{P}\Big(\frac{|T - r\log(r)|}{r} > t\Big) \leq \frac{\mathbb{E}(|T - r\log(r)|^2)}{r^2 t^2} \leq \frac{2\mathrm{Var}(T) + 2|\mathbb{E}(T) - r\log(r)|^2}{r^2 t^2} = \frac{O_{r\to\infty}(1)}{t^2}.$$

Par conséquent, on a pour tout $t > 0$,

$$\mathbb{P}(T \in [r\log(r) - rt, r\log(r) + rt]) \geq 1 - \frac{O_{r\to\infty}(1)}{t^2}.$$

À présent, pour $\alpha = 0.05$ et r fixé, il faut choisir t assez grand pour que le second membre soit égal à $1 - \alpha$. L'intervalle de confiance est de largeur $2rt$, et se dégrade quand t croit.

On rappelle que la loi de Gumbel a pour fonction de répartition $t \in \mathbb{R} \mapsto e^{-e^{-t}}$.

Théorème 2.6 (Fluctuations asymptotiques)**.** *On a*

$$\frac{T - r\log(r)}{r} \xrightarrow[r\to\infty]{\text{loi}} \text{Gumbel.}$$

Démonstration. Il suffit d'établir que pour tout $t \in \mathbb{R}$ on a

$$\lim_{r\to\infty} \mathbb{P}(T > r\log(r) + tr) = S(t) = 1 - e^{-e^{-t}}.$$

Fixons donc $t \in \mathbb{R}$ et supposons que r est assez grand pour que $r\log(r) + tr > r$. Introduisons l'entier $n_{t,r} = \lceil r\log(r) + tr \rceil$ (entier supérieur ou égal). Le théorème 2.4 donne

$$\mathbb{P}(T > r\log(r) + tr) = \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}}.$$

Comme $\binom{r}{k} \le r^k/k!$ et $1 - u \le e^{-u}$ pour tout $u \ge 0$, on a

$$\binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}} \xrightarrow[r\to\infty]{\le} \frac{e^{-tk}}{k!}.$$

Enfin, par convergence dominée, on obtient

$$\lim_{r\to\infty} \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}} = \sum_{k=1}^{m} (-1)^{k-1} \frac{e^{-tk}}{k!} = S(t).$$

□

Le théorème 2.6 fournit un intervalle de confiance pour T : pour tout réel $t \ge 0$,

$$\lim_{r\to\infty} \mathbb{P}(T \in [r\log(r) - rt, r\log(r) + rt]) = e^{-e^{-t}} - e^{-e^{t}}.$$

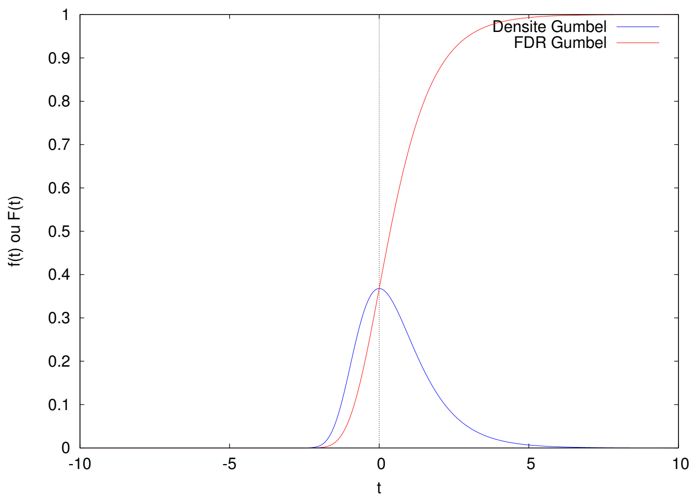

FIGURE 1 – Densité et fonction de répartition de la loi de Gumbel.

Tout comme pour la loi gaussienne, la densité de la loi de Gumbel est unimodale et très concentrée. Sa fonction de répartition fait apparaître une montée abrupte de 0 à 1, comme le montre la figure 2. Cela donne un phénomène de seuil pour T. La quantité $\mathbb{P}(T > n)$ passe abruptement de ≈ 1 à ≈ 0 autour de $n = r\log(r)$ si $r \gg 1$:

Théorème 2.7 (Convergence abrupte autour de $n = r\log(r)$)**.** *Pour tout réel $c > 0$,*

$$\mathbb{P}(T > r\log(r) + cr) \leq e^{-c}.$$

Si $(c_r)_{r\geq 1}$ dans $\mathbb{R}_+$ vérifie $\lim_{r\to\infty} c_r = \infty$ et $r\log(r) - rc_r > 0$ pour tout $r \geq 1$, alors

$$\lim_{r\to\infty} \mathbb{P}(T > r\log(r) - rc_r) = 1.$$

Démonstration. Avec $n = r(\log(r) + c)$ on obtient

$$\mathbb{P}(T > n) \leq \sum_{i=1}^{r} \mathbb{P}(E_{n,i}) = r(1 - 1/r)^n \leq re^{-n/r} = e^{-c}.$$

D'autre part, $(T - r\log(r))/r$ converge en loi (Gumbel) quand $r \to \infty$, donc

$$\mathbb{P}(T > r\log(r) - rc_r) = \mathbb{P}\Big(\frac{T - r\log(r)}{r} > -c_r\Big) \underset{r\to+\infty}{\longrightarrow} F(+\infty) = 1.$$

$\square$

Remarque 2.8 (Temps de recouvrement)**.** *Le collectionneur de coupons est un cas particulier du problème du recouvrement : si $(X_n)_{n\geq n_0}$ est une suite de v.a. pas nécessairement indépendantes ou de même loi, mais prenant leurs valeurs dans un même ensemble fini E, alors on définit le* temps de recouvrement *de E par*

$$T = \inf\{n \geq n_0 : \{X_{n_0}, \ldots, X_n\} = E\}.$$

3 Une marche aléatoire sur le groupe symétrique Σ_r

On considère $r \geq 2$ cartes à jouer empilées en un paquet vertical, et numérotées de 1 à r. On dit que la carte du dessus est en position 1, etc, et que celle du dessous est en position r. Une configuration du paquet correspond donc à un élément $\sigma \in \Sigma_r$, de sorte que $\sigma(k)$ désigne la position de la carte numéro k. On étudie une manière particulière de mélanger le parquet de cartes [1]. Plus précisément, on considère la k-insertion qui consiste à prendre la carte du sommet du paquet et à l'insérer entre la k^{e} et $k+1^{\text{e}}$ positions. La 1-insertion n'a aucun effet. On convient que la r-insertion place la carte du sommet sous le paquet. Une k-insertion fait passer de la configuration σ à la configuration $(k, k-1, \ldots, 1)\sigma$ où

$$(k, k-1, \ldots, 1)$$

désigne la permutation correspondant au cycle $k \to k-1 \to \cdots \to 1 \to k$. Si on choisit d'effectuer des k-insertions successives en utilisant une suite i.i.d. uniforme sur $\{1, \ldots, r\}$ pour choisir k alors, en notant X_n la configuration du paquet à l'instant n, on obtient une suite aléatoire $(X_n)_{n\geq 0}$ de Σ_r vérifiant pour tout $n \geq 0$,

$$X_{n+1} = \varepsilon_{n+1}X_n = \varepsilon_{n+1}\cdots\varepsilon_1 X_0$$

où $(\varepsilon_n)_{n\geq 1}$ sont i.i.d. de loi uniforme sur $C := \{(k, k-1, \ldots, 1) : 1 \leq k \leq r\} \subset \Sigma_r$. La suite $(X_n)_{n\geq 0}$ est une marche aléatoire à gauche sur le groupe (non abélien) Σ_r, d'incréments C. C'est aussi une chaîne de Markov d'espace d'état fini Σ_r (de cardinal $r!$) et de noyau

$$\mathbf{P}(\sigma, \sigma') = \frac{\mathbf{1}_C(\sigma'\sigma^{-1})}{|C|} = \frac{\mathbf{1}_C(\sigma'\sigma^{-1})}{r}.$$

1. Cette manière de mélanger assez peu usuelle a le mérite de permettre une analyse mathématique simple et rapide qui mène au phénomène de convergence abrupte à l'équilibre. Ce phénomène est commun à toutes les manières de mélanger les cartes qui ont été étudiées. On peut penser par exemple au *riffle shuffle* : on coupe le paquet en deux et on fusionne les deux sous-paquets en intercalant leurs cartes.

Théorème 3.1 (Convergence en loi)**.** *La chaîne* $(X_n)_{n\geq 0}$ *est récurrente irréductible apériodique. Son unique loi invariante est la loi uniforme sur* Σ_r *:*

$$\mu = \sum_{\sigma\in\Sigma_r} \frac{1}{|\Sigma_r|}\delta_\sigma = \frac{1}{r!}\sum_{\sigma\in\Sigma_r}\delta_\sigma.$$

La chaîne $(X_n)_{n\geq 0}$ *converge en loi vers* μ *quelque soit la loi initiale* $\mathcal{L}(X_0)$.

Démonstration. Les transpositions spéciales $\mathcal{T} := \{(r, r-1), \ldots, (2,1), (1,r)\}$ engendrent Σ_r. Comme les k-insertions $(r, r-1, \ldots, 1)$ et $(2,1)$ correspondant à $k = r$ et à $k = 2$ engendrent $\mathcal{T}$, on en déduit qu'elles engendrent Σ_r. Donc C engendre Σ_r et la chaîne est irréductible. Comme l'identité (1) est également une k-insertion $(k = 1)$, la diagonale de la matrice de transition de la chaîne est > 0 et donc la chaîne est apériodique. Comme Σ_r est un groupe, il y a exactement r états qui conduisent à chaque état, donc les colonnes de la matrice de transition ont exactement r entrées non nulles, toutes égales à $1/r$. Ainsi, la transposée de la matrice de transition est également une matrice de transition [2] et donc la loi uniforme est invariante. Or toute chaîne finie récurrente apériodique possède une unique loi invariante vers laquelle elle converge en loi quelque soit sa loi initiale. □

Supposons que $X_0 = (1)$ **(toutes les cartes sont dans l'ordre au départ)**. Au temps 0 la carte r est en position r (tout en bas du paquet), et subit une remontée au fil du temps, jusqu'au sommet. Cela conduit à définir des v.a.r. $T_1, \ldots, T_{r-1}$ vérifiant, pour tout $1 \leq k \leq r-1$, conditionnellement à $T_1, \ldots, T_{k-1}$ (avec $T_1 + \cdots + T_{k-1} = 0$ si $k = 1$)

$$\begin{aligned} T_k &= \min\{n \geq 1 : X_{T_1+\cdots+T_{k-1}+n}(r) = r-k\} \\ &= \min\{n \geq 1 : X_n(r) = r-k\} - (T_1 + \cdots + T_{k-1}). \end{aligned}$$

Au temps $T_1 + \cdots + T_{r-1}$ la carte r est en position 1 (sommet du paquet). Cette remontée est de plus en plus rapide. La variable aléatoire $T := 1 + T_1 + \cdots + T_{r-1}$ suit la loi du collectionneur de coupons de r coupons de probabilité d'apparition uniforme. La loi $\mathcal{L}(T_k \,|\, T_{k-1}, \ldots, T_1)$ suit la loi géométrique sur $\mathbb{N}^*$ de paramètre $(r-k)/r$, pour tout $1 \leq k \leq r-1$. Notons que $T \geq r$.

Théorème 3.2 (Bon mélange après remontée)**.** *Pour tout* $n \geq 0$*, les variables aléatoires* X_{T+n} *et* T *sont indépendantes. De plus,* X_{T+n} *suit la loi uniforme* μ *sur* Σ_r.

On dit alors que T est un temps fort de stationnarité (d'après Aldous et Diaconis).

Démonstration. La loi uniforme sur Σ_r peut s'obtenir en tirant uniformément sans remise les images de $1, \ldots, r$. D'autre part, la loi uniforme sur Σ_r est invariante par translation.

Au temps T_1, la carte r se trouve pour la première fois en position $r-1$, car la carte numéro 1 a été glissée sous le paquet. Au temps $T_1 + T_2$, la carte r se trouve pour la première fois en position $r-2$ car la carte numéro 2 a été glissée sous ou sur la carte 1. Les numéros des deux cartes sous le paquet sont $1, 2$ ou $2, 1$ avec probabilité $1/2$. Par récurrence, au temps $T_1 + \cdots + T_{r-1} = T - 1$, la carte r se trouve au sommet du paquet pour la première fois et les $r-1$ cartes qui sont sous elles ont des numéros répartis uniformément sans remise dans $\{1, \ldots, r-1\}$. Au temps T, la carte r est placée aléatoirement et uniformément dans le paquet à une position entre 1 et r et donc X_T est uniforme. La probabilité $\mathbb{P}(X_T = \sigma, T = k)$ ne dépend pas de σ et en particulier T et X_T sont indépendantes. Comme μ est invariante, on obtient $X_{T+n} \sim \mu$ pour tout $n \geq 0$. □

2. Elle est bistochastique ou doublement stochastique. L'ensemble des matrices bistochastiques $n \times n$ est un polytope (intersection de demi-espaces) convexe et compact à $(n-1)^2$ degrés de liberté. Ses points extrémaux sont les matrices de permutations (Birkhoff et von Neumann). Encore le groupe symétrique !

Pour bien mélanger le paquet de cartes, on pourrait s'arrêter au temps T. Malheureusement, on ne connaît pas T en pratique ! Alternativement, on pourrait chercher à déterminer une valeur de n déterministe, aussi petite que possible, telle que $d_V(\mathcal{L}(X_n),\mu)$ est proche de zéro, où d_V est la distance en variation totale. Il s'avère que pour r assez grand, la quantité $d_V(\mathcal{L}(X_n),\mu)$ passe de 1 à 0 de manière abrupte autour de $n = r\log(r)$. Ce phénomène de convergence abrupte est quantifié par le théorème suivant.

Théorème 3.3 (Convergence abrupte autour de $n = r\log(r)$)**.** *Pour tout réel $c > 0$,*

$$d_V(\mathcal{L}(X_{r\log(r)+cr}),\mu) \leq e^{-c}.$$

Si $(c_r)_{r\geq 1}$ dans $\mathbb{R}_+$ vérifie $\lim_{r\to\infty} c_r = \infty$ et $r\log(r) - rc_r > 0$ pour tout $r \geq 1$, alors

$$\lim_{r\to\infty} d_V(\mathcal{L}(X_{r\log(r)-rc_r}),\mu) = 1.$$

Démonstration. Grâce au théorème 3.2, on a, pour tout $A \subset \Sigma_r$,

$$\begin{aligned}
\mathbb{P}(X_n \in A) &= \sum_{k=0}^{n} \mathbb{P}(X_n \in A, T = k) + \mathbb{P}(X_n = \sigma, T > n) \\
&= \sum_{k=0}^{n} \mathbb{P}(X_{T+n-k} \in A, T = k) + \mathbb{P}(T > n) \\
&= \sum_{k=0}^{n} \mu(A)\mathbb{P}(T = k) + \mathbb{P}(T > n) \\
&= \mu(A)\mathbb{P}(T \leq n) + \mathbb{P}(T > n) \\
&\leq \mu(A) + \mathbb{P}(T > n).
\end{aligned}$$

d'où $\mathbb{P}(X_n \in A) - \mu(A) \leq \mathbb{P}(T > n)$. Appliqué à A et A^c, cela donne, pour tout $n \geq 0$,

$$d_V(\mathcal{L}(X_n),\mu) \leq \mathbb{P}(T > n).$$

Le résultat voulu découle à présent du théorème 2.7 sur le collectionneur de coupons. □

Pour aller plus loin : pour le collectionneur de coupons, en plus du livre de Feller [Fel71, Fel68], le livre [MR95], et l'article [Hol01]. Pour la convergence abrupte (cutoff) des battages de cartes, les articles de Diaconis et al [AD86, DMP95, DFP92, BD92]. Plusieurs autres manières de mélanger un paquet de cartes possèdent cette propriété de convergence abrupte. Le phénomène de convergence abrupte pour les chaînes de Markov a été caractérisé par Chen et Saloff-Coste pour les distances L^p avec $p > 1$ [CSC10, CSC08], le cas $p = 1$ de la variation totale restant ouvert.

Références

[AD86] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5) :333–348, 1986.

[BD92] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2) :294–313, 1992.

[CSC08] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, 13(3) :26–78, 2008.

[CSC10] G.-Y. Chen and L. Saloff-Coste. The L^2-cutoff for reversible Markov processes, *J. Funct. Anal.*, 258(7) :2246–2315, 2010.

[DFP92] Persi Diaconis, James Allen Fill, and Jim Pitman. Analysis of top to random shuffles. *Combin. Probab. Comput.*, 1(2) :135–155, 1992.

[DMP95] Persi Diaconis, Michael McGrath, and Jim Pitman. Riffle shuffles, cycles, and descents. *Combinatorica*, 15(1) :11–29, 1995.

[Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[Hol01] Lars Holst. Extreme value distributions for random coupon collector and birthday problems. *Extremes*, 4(2) :129–145 (2002), 2001.

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms.* Cambridge University Press, Cambridge, 1995.

[Ste94] J. Michael Steele. Le Cam's inequality and Poisson approximations. *Amer. Math. Monthly*, 101(1) :48–54, 1994.