

# Quelques études de convergence de marches aléatoires sur un groupe fini

Exposé de maîtrise par Cédric Bounya et Rémi Tardivo

Encadrés par Pierre Nolin

Sujet proposé par Christophe Garban

14 septembre 2007

## Résumé

Ces notes d'exposé de M1 traitent de quelques aspects des processus de Markov sur un groupe fini, et présentent des arguments typiques dans l'étude de la vitesse de convergence de tels processus aléatoires, à travers quelques exemples simples sur des groupes classiques, ainsi qu'une rapide présentation des applications de ces processus.

## Table des matières

<b>1</b>	<b>Outils</b>	<b>2</b>
<b>2</b>	<b>Étude de l'urne d'Ehrenfest</b>	<b>5</b>
<b>3</b>	<b>Marches <i>modulo</i> <math>n</math></b>	<b>9</b>
<b>4</b>	<b>Modèles de battage de cartes</b>	<b>10</b>

## Notations et conventions

Dans toute la suite, on identifiera une distribution de probabilité  $\mathbb{P}$  sur un univers fini  $\Omega$  à la fonction positive associée  $\mathbb{P}(x) = \mathbb{P}(\{x\})$ .

La notation  $X \sim \mu$  signifiera que la variable aléatoire  $X$  suit la loi de probabilité  $\mu$ .

On dira qu'un couple de variables aléatoires  $(X, Y)$  admet pour marginales les lois  $\mu$  et  $\nu$  et on notera  $(X, Y) \sim (\mu, \nu)$  si  $X \sim \mu$  et  $Y \sim \nu$ .

Pour une famille de variables aléatoires, l'abréviation *iid* (*indépendantes identiquement distribuées*) signifiera qu'elles sont indépendantes et de même loi.

## 1 Outils

**Définition 1.1.** On définit sur l'espace des mesures de probabilité d'un univers fini  $\Omega$  la distance  $d$  dite distance en variation totale :

$$d(\mu, \lambda) = \sup_{A \in \mathcal{P}(\Omega)} |\mu(A) - \lambda(A)| \in [0; 1].$$

**Proposition 1.1.** 1. La distance en variation totale peut se calculer également par la somme :

$$d(\mu, \lambda) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \lambda(x)|.$$

2. Elle peut aussi être vue comme « couplage optimal » :

$$d(\mu, \lambda) = \inf\{\mathbb{P}\{X \neq Y\} : X \sim \lambda, Y \sim \mu\}.$$

*Démonstration.* 1. Il s'agit de considérer l'ensemble  $B$  suivant qui réalise le sup :  
 $B = \{x \in \Omega : \mu(x) > \lambda(x)\}$ .

Si  $A \in \mathcal{P}(\Omega)$ , on écrit  $A = (A \cap B) \cup (A \cap \bar{B})$ , et on obtient l'inégalité souhaitée. La preuve complète figure en [2].

2. Soit  $(X, Y)$  un couple de variables de marginales  $(\mu, \nu)$ . On a :

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbb{P}\{X \in A\} - \mathbb{P}\{X \notin A, Y \in A\} - \mathbb{P}\{X \in A, Y \in A\} \\ &\leq \mathbb{P}\{X \in A, Y \notin A\} \\ &\leq \mathbb{P}\{X \neq Y\} \end{aligned}$$

On a donc bien  $d(\mu, \nu) \leq \inf\{\mathbb{P}\{X \neq Y\} : X \sim \mu, Y \sim \nu\}$ .

On va maintenant exhiber un couple  $(X, Y)$  pour lequel l'égalité est réalisée. Si  $d(\mu, \lambda) = 1$  alors, d'après la définition de  $d(\mu, \lambda)$ , il existe  $A$  tel que

$$|\mu(A) - \lambda(A)| = 1.$$

Quitte à passer au complémentaire, on suppose :  $\mu(A) = 1$  et  $\lambda(A) = 0$ . Soit  $X$  et  $Y$  deux variables aléatoires réelles de lois respectives  $\mu$  et  $\lambda$ . Alors :

$$\begin{aligned} \mathbb{P}\{X = Y\} &= \mathbb{P}\left(\{X = Y, X \in A\} \cup \{X = Y, X \notin A\}\right) \\ &\leq \mathbb{P}\left(\{Y \in A\} \cup \{X \notin A\}\right) \\ &\leq \mathbb{P}\{Y \in A\} + \mathbb{P}\{X \notin A\} \\ &= 0. \end{aligned}$$

En passant à la borne inférieure sur toutes les  $X$  et  $Y$  correspondantes, on obtient :  $\inf\{\mathbb{P}\{X \neq Y\} : X \sim \mu, Y \sim \nu\} = 1 = d(\mu, \lambda)$ .

On suppose maintenant :  $d(\mu, \lambda) < 1$ .

$$\begin{aligned} 2d(\mu, \nu) &= \sum_{x \in \Omega} \mathbb{1}_{\{\mu(x) \geq \nu(x)\}} (\mu(x) - \nu(x)) + \sum_{x \in \Omega} \mathbb{1}_{\{\mu(x) < \nu(x)\}} (\nu(x) - \mu(x)) \\ &= \sum_{x \in \Omega} \mathbb{1}_{\{\mu(x) \geq \nu(x)\}} (\mu(x) - \mu(x) \wedge \nu(x)) + \sum_{x \in \Omega} \mathbb{1}_{\{\mu(x) < \nu(x)\}} (\nu(x) - \mu(x) \wedge \nu(x)) \\ &= \sum_{x \in \Omega} \underbrace{(\mu(x) - \mu(x) \wedge \nu(x))}_{=0 \text{ si } \mu(x) < \nu(x)} + \sum_{x \in \Omega} \underbrace{(\nu(x) - \mu(x) \wedge \nu(x))}_{=0 \text{ si } \mu(x) \geq \nu(x)} \\ &= 2 - 2 \sum_{x \in \Omega} \mu(x) \wedge \nu(x). \end{aligned}$$

On peut donc définir par  $p := \frac{\mu \wedge \nu}{1 - d(\mu, \nu)}$  une probabilité sur  $\Omega$ .

Les fonctions  $\tilde{\mu} = \frac{1}{d(\mu, \nu)} \mathbb{1}_{\{\mu > \nu\}} (\mu - \nu)$  et  $\tilde{\nu} = \frac{1}{d(\mu, \nu)} \mathbb{1}_{\{\nu > \mu\}} (\nu - \mu)$  sont des distributions de probabilité, en tant que barycentres de  $\mu, \nu$  et  $p$ . Leurs supports sont clairement disjoints.

Prenons  $P, \tilde{X}, \tilde{Y}, \xi$  des variables indépendantes de loi respectives  $p, \tilde{\mu}, \tilde{\nu}, \mathcal{B}(d(\mu, \nu))$  et définissons  $X, Y$  par  $X = Y = P$  sur  $\{\xi = 0\}$ ,  $X = \tilde{X}, Y = \tilde{Y}$  sur  $\{\xi = 1\}$ . Alors la variable  $X$  (respectivement  $Y$ ) a pour loi  $\mu$  (respectivement  $\nu$ ), et  $\{X = Y\} = \{\xi = 0\}$ . Donc  $\mathbb{P}\{X \neq Y\} = d(\mu, \nu)$ . □

On énonce maintenant des résultats généraux dont les preuves figurent en [2].

**Proposition 1.2.** *Si  $(X_i)_{i \geq 0}$  est une chaîne de Markov irréductible sur un ensemble fini  $\Omega$ , on note  $\tau_a^+ := \inf\{n \geq 1 : X_n = a\}$  le premier temps  $> 0$  où la chaîne atteint l'état  $a$ . Il existe une unique distribution stationnaire, donnée par*

$$\pi(x) := \frac{1}{\mathbb{E}_a[\tau_a^+]} \mathbb{E}_a \left[ \sum_{n=0}^{\tau_a^+ - 1} \mathbb{1}_{\{X_n = x\}} \right].$$

**Théorème 1.1.** *Supposons  $X$  irréductible aperiodique, de matrice de transition  $P$ , de distribution stationnaire  $\pi$ , alors  $\forall x \in \Omega, (d(P^n(x, \cdot), \pi))_{n \geq 0}$  tend vers 0 à vitesse géométrique.*

**Théorème 1.2.** *Soit une famille  $\{(X_t, Y_t)^{x,y}, (x, y) \in \Omega^2\}$  de couples de chaînes de Markov telle que*

$$\forall (x, y) \in \Omega^2, \begin{cases} (X_0, Y_0)^{x,y} = (x, y) \\ (X_t^{x,y}) \text{ est une chaîne de Markov de matrice } P \\ (Y_t^{x,y}) \text{ est une chaîne de Markov de matrice } \pi \end{cases}$$

et soit  $\tau$  le temps où  $X$  et  $Y$  se rejoignent pour la première fois :  $\tau := \min\{t : X_t = Y_t\}$ . Alors

$$\max_{\mu} d(\mu P^t, \pi) \leq \max_{(x,y) \in \Omega^2} \mathbb{P}^{x,y}\{\tau > t\}.$$

Ce théorème explique que l'on s'intéresse à l'espérance du temps de couplage puisque l'on peut alors majorer  $\mathbb{P}\{\tau > t\}$  par l'inégalité de Markov. On obtient :

$$\max_{\mu} d(\mu P^t, \pi) \leq \max_{(x,y) \in \Omega^2} \frac{\mathbb{E}[\tau]}{t}.$$

Pour prouver le théorème, on utilise la proposition qui suit, avec  $\nu = \pi$  : dans ce cas,  $\pi P^t = \pi$ .

**Proposition 1.3.** *Si  $X$  (resp.  $Y$ ) est une chaîne de Markov de matrice  $P$ , de distribution initiale  $\mu$  (resp.  $\nu$ ),*

$$\max_{\mu, \nu} d(\mu P^t, \nu P^t) \leq \max_{(x,y) \in \Omega^2} \mathbb{P}^{x,y}\{\tau > t\}.$$

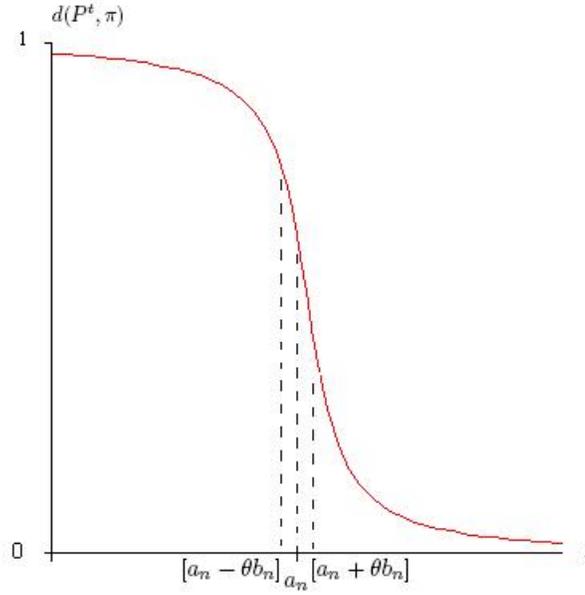
Cette proposition se démontre en utilisant le fait que  $X_t = Y_t$  dès que  $\tau \leq t$  : en effet, la propriété de Markov entraîne qu'une fois que les deux chaînes se sont rencontrées, elles suivent la même trajectoire. La preuve complète figure en [2].

Définissons  $\bar{d}(t) := \max_{(x,y) \in \Omega^2} d(P^t(x, \cdot), P^t(y, \cdot))$ . On a le lemme suivant, démontré en [2].

**Lemme 1.1.** *La fonction  $\bar{d}$  vérifie les propriétés suivantes :*

1.  $\max_{\mu} d(\mu P^t, \pi) = \max_x d(\delta_x P^t, \pi) \leq \bar{d}(t)$ .
2.  $\ln(\bar{d})$  est sous-additive.

Un aspect intéressant des marches aléatoires sur les groupes finis est le phénomène de *cut-off*, en français *convergence abrupte*.

FIG. 1 – Un exemple de *cut-off* au temps  $a_n$ 

**Définition 1.2** (*cut-off*). On dit qu'une suite de chaînes de Markov  $((P_n^{*t})_{n \in \mathbb{N}})$  présente un *cut-off* (convergence abrupte en français) dans sa convergence vers la suite de probabilités  $(\pi_n)$  s'il existe deux suites  $(a_n), (b_n)$ , avec  $a_n \rightarrow \infty$ ,  $(b_n) = o(a_n)$ , et une fonction  $c(\theta)$  vérifiant  $\lim_{-\infty} c = 1, \lim_{+\infty} c = 0$ , telle que si  $t_n = [a_n + \theta b_n]$ , alors

$$d(P_n^{*t_n}, \pi_n) \rightarrow_n c(\theta).$$

Le sens de cette notion est que la convergence vers la probabilité d'équilibre se fait d'une manière très brusque (voir Figure 1) autour d'un temps  $t_n$ , lorsque le paramètre pertinent  $n$  devient grand. Ce phénomène de *cut-off* apparaît dans de nombreuses convergences de processus de Markov, par exemple dans le cas d'un mélange d'un paquet de  $n$  cartes lorsque  $n$  devient grand.

## 2 Étude de l'urne d'Ehrenfest

### Marche sur l'hypercube

On considère ici  $\Omega := (\mathbb{Z}/2\mathbb{Z})^d$ . Chaque pas de la marche consiste à choisir uniformément l'une des  $d$  coordonnées et à la modifier avec probabilité  $1/2$ .

On souhaite connaître le temps nécessaire pour avoir touché toutes les coordonnées. L'usage a consacré une heuristique pour ce résultat : le problème du

collectionneur de coupons (*coupon collector* en anglais) qui doit rassembler chacun des  $d$  coupons dans les boîtes de ses céréales favorites. On montre comme dans la suite qu'en moyenne, il lui faudra acheter  $d \log d$  paquets de céréales.

Soit  $\tau^{(l)}$  le premier instant où  $l$  coordonnées distinctes ont été touchées.  $\tau := \tau^{(d)}$  est donc le temps qu'il faut attendre pour que la distribution sur l'hypercube soit uniforme.  $X_0 = 0, X_{t+1} = X_t + \bar{\epsilon}_t(\bar{\delta}_{\xi_t, i})_{i=1\dots d}$ , où  $(\xi_t)_{t \geq 0}$  est une suite de variables *iid* uniformes sur  $[1 \dots d]$ , et  $(\epsilon_t)_{t \geq 0}$  une suite de lancers de pièces équilibrées.

On va s'intéresser à l'ensemble  $A = \{\tau^{(l+1)} - \tau^{(l)} \geq k\}$ . Un événement appartient à  $A$  si et seulement si :

$$\forall t \in \{\tau^{(l)} + 1, \dots, \tau^{(l)} + k - 1\}, \text{ soit } \epsilon_t = 0, \text{ soit } \xi_t \in \{\xi_0, \dots, \xi_{\tau^{(l)}-1}\}.$$

La probabilité que  $\epsilon_t = 1$  et  $\xi_t \notin \{\xi_0, \dots, \xi_{\tau^{(l)}-1}\}$  est égale à  $\frac{1}{2}(1 - \frac{l}{d})$ .  
On obtient finalement :

$$\mathbb{P}\{\tau^{(l+1)} - \tau^{(l)} \geq k\} = \left(1 - \frac{1}{2}\left(1 - \frac{l}{d}\right)\right)^{k-1} = \left(\frac{l+d}{2d}\right)^{k-1}.$$

En sommant sur  $k$ , on déduit l'espérance :  $\mathbb{E}[\tau^{(l+1)} - \tau^{(l)}] = \frac{2d}{d-l}$  d'où :

$$\mathbb{E}[\tau] \sim 2d \log d.$$

## Étude du cut-off de l'hypercube

On reprend ici  $\Omega := (\mathbb{Z}/2\mathbb{Z})^d \simeq \{-1; 1\}^d$ , vu multiplicativement pour des raisons de convenance calculatoire. Nous allons montrer que si l'on part de  $(1, \dots, 1)$ , la distance entre la loi de  $X_{t(d)}$  et la distribution uniforme tend vers 1 pour  $t(d) = \frac{1-\epsilon}{2}d \log d$ .

**Définition 2.1.** *Le temps d'arrêt  $\tau_1(\epsilon) := \max_{x \in \Omega} \min\{t : d(\delta_x P^t, \pi) \leq \epsilon\}$  est appelé temps de mélange à  $\epsilon$  près.*

On va ici chercher à minorer ce temps de mélange par :  $\tau_1(\epsilon) \leq \min\{t : |\delta_x P^t(A) - \pi(A)| \leq \epsilon\}$  pour un certain  $x$  et un événement  $A$  bien choisis. Il se trouve que la variable  $S_t = \sum_{i=1}^d X_t^i$  décrit bien la qualité du mélange. En effet, connaissant la loi de  $S_t$ , on peut calculer  $\#\{i/X_i \text{ n'a pas été touché}\} = \mathbb{E}_{\mu_t}[S_t]$ , et l'on a :  $\mathbb{E}_\pi[S_t] = 0, \text{Var}_\pi(S) = d$ . Nous allons montrer que si l'on part de  $(1, \dots, 1)$ , alors  $S_t$  reste loin de 0 au début.

Notons  $A_t^{(i)} = \{\text{La } i^{\text{ème}} \text{ coordonnée n'a pas été touchée au temps } t\}$ . On voit facilement que  $\forall i \in \{1 \dots d\}, t \in \mathbb{N}, \mathbb{P}(A_t^{(i)}) = (1 - \frac{1}{d})^t$ , ce qui permet de calculer :

$$\begin{aligned} \mathbb{E}_{\mu_t}[X_t^{(i)}] &= \underbrace{\mathbb{E}[X_t^{(i)} | A_t^{(i)}]}_{=1} \mathbb{P}(A_t^{(i)}) + \underbrace{\mathbb{E}[X_t^{(i)} | \overline{A_t^{(i)}}]}_{=0} \mathbb{P}(\overline{A_t^{(i)}}) \\ &= \left(1 - \frac{1}{d}\right)^t. \end{aligned}$$

Donc  $\mathbb{E}_{\mu_t}[S_t] = d\left(1 - \frac{1}{d}\right)^t$ .

De la même manière, pour  $i \neq j$ ,  $\mathbb{E}_{\mu_t}[X_t^{(i)}X_t^{(j)}] = (1 - \frac{2}{d})^t$ , donc  $\text{cov}(X_t^{(i)}, X_t^{(j)}) = (1 - \frac{2}{d})^t - (1 - \frac{1}{d})^{2t} < 0$ , d'où  $\text{Var}_{\mu_t}(S_t) < d$ .

On choisit  $t = \frac{1-\epsilon}{2}d \log d$  pour tenter de prouver le résultat conjecturé précédemment. On sait que  $\mathbb{E}_{\mu_t}[S_t] = d(1 - \frac{1}{d})^t \sim d^{\frac{1+\epsilon}{2}}$ . On choisit donc  $A_t = \{x : S \geq \frac{d^{\frac{1+\epsilon}{2}}}{3}\}$  (et non  $\{x : S \geq \frac{d^{\frac{1+\epsilon}{2}}}{2}\}$  car on a seulement un équivalent de  $\mathbb{E}_{\mu_t}[S_t]$ ).

L'inégalité de Tchebychev fournit :

$$\pi(A_t) \leq \pi\left\{x : |S_t| \geq \frac{d^{\frac{1+\epsilon}{2}}}{3}\right\} \leq 9d^{-\epsilon}$$

et

$$\mu_t(A_t^c) \leq \mu_t\left\{|S - \mathbb{E}_{\mu_t}[S_t]| \geq \frac{d^{\frac{1+\epsilon}{2}}}{3}\right\} \leq 9d^{-\epsilon}$$

et on obtient finalement :

$$d(\mu_t, \pi) \geq \mu_t(A_t) - \pi(A_t) \geq 1 - 18d^{-\epsilon}.$$

Ceci signifie que, quand on fait tendre  $d$  vers  $+\infty$ , si  $t$  est asymptotiquement inférieur à  $\frac{d}{2} \log d$ , la distance en variation totale du mélange en temps  $t$  à la mesure d'équilibre tend rapidement vers 1 :

$$\lim_{d \rightarrow +\infty} d(\mu_{t(d)}, \pi) = 1 \text{ si } t(d) = \left\lfloor \frac{1-\epsilon}{2}d \log d \right\rfloor.$$

Pour montrer l'existence du *cut-off*, il resterait à montrer qu'à l'inverse

$$\lim_{d \rightarrow +\infty} d(\mu_{t(d)}, \pi) = 0 \text{ si } t(d) = \left\lfloor \frac{1+\epsilon}{2}d \log d \right\rfloor.$$

Pour ce faire, une bonne idée consiste à chercher à comprendre la répartition du spectre de la matrice de transition dans le segment  $[-1, 1]$ .

## Étude spectrale

Dans le cas qui nous intéresse, la matrice de transition  $P$  est symétrique, et irréductible aperiodique, ses valeurs propres vérifient donc  $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{2d} > -1$ . Il est usuel de noter  $|\lambda_\star| = \max_{j>1} |\lambda_j|$ , et on appelle *trou spectral* (*spectral gap*) la distance  $1 - |\lambda_\star|$ . On a le théorème général suivant, démontré en [2].

**Théorème 2.1.** *Pour une chaîne de Markov réversible (c'est-à-dire, vérifiant  $\pi(x)P(x, y) = \pi(y)P(y, x)$ ), on a :*

$$\tau_1(\epsilon) \leq \frac{\log(\epsilon \pi_{\min})}{\log(|\lambda_\star|)}$$

où  $\pi_{\min} = \inf_{x \in \Omega} \pi(x)$ .

Dans le cas de l'hypercube, les fonctions propres sont les  $(f_S)_{S \in \mathcal{P}(\Omega)}$  définies par  $\forall x = (x_i)_{i=1 \dots d} \in \{-1, 1\}^d$ ,  $f_S(x) = \prod_{i \in S} x_i$ . Par définition,

$$\begin{aligned} P f_S(x) &= \mathbb{E}[f_S(\delta_x P)] \\ &= \mathbb{P}\left(\bigcap_{i \in S} A_1^{(i)}\right) \underbrace{\mathbb{E}\left[f_S(\delta_x P) \middle| \bigcap_{i \in S} A_1^{(i)}\right]}_{=f_S(x)} + \mathbb{P}\left(\bigcup_{i \in S} \overline{A_1^{(i)}}\right) \underbrace{\mathbb{E}\left[f_S(\delta_x P) \middle| \bigcup_{i \in S} \overline{A_1^{(i)}}\right]}_{=\frac{1}{2}f_S(x) - \frac{1}{2}f_S(x)=0} \\ &= \left(1 - \frac{|S|}{d}\right) f_S(x). \end{aligned}$$

Ceci nous fournit  $|\lambda_*| = \lambda_2 = 1 - \frac{1}{d}$ . Le théorème nous donne alors :  $\tau_1(\epsilon) \leq K d^2$  pour un certain  $K$  réel et pour  $d$  suffisamment grand. Cette méthode s'avère donc décevante puisqu'elle ne nous permet pas de prouver le *cut-off* en  $d \log(d)$ .

## Aspect physique

L'urne d'Ehrenfest est un modèle permettant de comprendre un paradoxe de la thermodynamique : « Comment le principe d'une entropie croissante peut-il être compatible avec un système auquel rien n'interdit de retourner à son état initial ? ». On dispose de 2 urnes (A et B) et de  $d$  particules. À chaque instant, on choisit une particule avec probabilité uniforme, puis, avec probabilité égale à  $1/2$ , on la déplace vers l'autre urne, *ou bien* on la laisse là où elle est. Au temps initial, toutes les particules sont dans l'urne A. Ce modèle est exactement celui de l'hypercube déjà étudié. En effet, attribuons le nombre 1 à l'urne A et le nombre -1 à l'urne B. On numérote ensuite les  $d$  particules :  $X : t \mapsto (x_1, x_2, \dots, x_d)(t)$ . La suite de variables aléatoires  $(X_t)$  est alors une chaîne de Markov sur l'hypercube  $\{-1, 1\}^d$ .

Introduisons l'entropie, dite de Shannon, définie par  $S = - \sum_{j \in \omega} P(j) \log P(j)$ ,

où  $\omega$  correspond à l'ensemble des états macroscopiques observables. Ici  $\omega = \{\{S_d = k\}, k = 0 \dots d\}$ .

On sait alors qu'il existe un phénomène de *cut-off* en  $\frac{1}{2} d \log d$ . Le paramètre  $d$  étant très grand (typiquement  $\mathcal{N}_A$ , le nombre d'Avogadro), le temps de *cut-off* est largement inférieur à  $2^d$ , temps nécessaire pour que l'on ait une chance raisonnable de revenir à l'état initial. Ainsi, on n'observera pas de retour à l'état initial et il n'y a donc plus de paradoxe apparent ([1]). L'entropie maximale, correspondant à l'état d'équilibre, est atteinte lorsque  $P$  est uniforme. L'entropie du mélange est croissante avec le temps et après le *cut-off* elle est pratiquement stationnaire.

À titre indicatif, on peut noter que si l'on choisit pour situation initiale  $d/2$  particules dans chaque urne (c'est-à-dire que l'on est, en un sens, proche de l'équilibre) on n'observe pas de phénomène de *cut-off*, la chaîne se mélange de manière progressive. De manière assez étrange, un éventuel phénomène de *cut-off* dépend donc de l'état initial de la chaîne de Markov considérée.

### 3 Marches *modulo* $n$

#### Marche paresseuse sur un cycle

Ici, on considère  $\Omega := \mathbb{Z}/n\mathbb{Z}$ , deux éléments étant dits *voisins* si leur différence est  $-1$  ou  $1$ . On dira qu'une marche aléatoire telle qu'à chaque étape, elle reste immobile avec probabilité  $1/2$  et visite chacun de ses voisins avec probabilité  $1/4$  est « *paresseuse* ». On considère la chaîne  $(X_t, Y_t)$  initialement en  $(x, y)$ , qui, à chaque étape, change une de ses coordonnées avec probabilité uniforme en un de ses voisins, jusqu'à ce que les deux coordonnées se rencontrent, après quoi elles demeurent égales, et effectuent la même marche paresseuse. Les marches aléatoires  $(X_t)$  et  $(Y_t)$  sont paresseuses. On s'intéresse à l'espérance du temps de rencontre de  $X$  et  $Y$ , noté  $\tau$ .

Le processus  $(X_t - Y_t)$  est une marche aléatoire simple sur le cycle, image d'une marche aléatoire simple  $(Z_t)$  sur  $\mathbb{Z}$  *modulo*  $n$ , initiée en un point  $k \in 0 \dots n - 1$ . Un petit calcul montre que le processus  $(Z_t^2 - t)$  est une martingale, et le théorème d'arrêt donne :

$$\mathbb{E}[Z_{\tau \wedge t}^2] - Z_0^2 = \mathbb{E}[\tau \wedge t]$$

Le membre de gauche converge par bornitude de l'intégrande vers

$$\mathbb{P}\{Z_\tau = 0\} \times 0 + \mathbb{P}\{Z_\tau = n\} \times n^2 - k^2$$

et le membre de droite converge par monotonie de l'intégrande vers  $\mathbb{E}[\tau]$ . Comme  $\mathbb{P}\{Z_\tau = n\} = \frac{k}{n}$  (ruine du joueur), il vient  $\mathbb{E}[\tau] = k(n - k) \leq (\frac{n}{2})^2$ .

#### Étude spectrale de la marche paresseuse sur le cycle

La matrice de transition est ici  $\frac{1}{4} \begin{pmatrix} 2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & \ddots & \ddots & \ddots & & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & \ddots & \ddots & 1 \\ 1 & 0 & \cdots & 0 & 1 & 2 \end{pmatrix} = \frac{1}{4}A.$

La matrice  $A$  se diagonalise dans la base naturelle des  $(V_i = (\zeta^{ij})_{j=1 \dots n})_{i=1 \dots n}$ , où  $\zeta$  est une racine primitive d'ordre  $n$  de l'unité, car on a :  $(AV_i)_j = \zeta^{ij}(2 + \zeta^i + \zeta^{-i})$ , donc  $AV_i = 2(1 + \Re(\zeta^i))V_i$ . Le spectre de la matrice de transition est donc  $\{\frac{1}{2}(1 + \cos(\frac{2i\pi}{n})), i = 0 \dots [\frac{n}{2}]\}$ , et le trou spectral est de l'ordre de  $\frac{\pi^2}{n^2}$ .

#### Extension au tore discret

L'espace de probabilité choisi ici est  $\Omega := (\mathbb{Z}/n\mathbb{Z})^d$ . Deux points  $x$  et  $y$  sont voisins si toutes leurs coordonnées sont égales, sauf une suivant laquelle ils sont

voisins au sens précédent : il existe  $i$  tel que  $x_j = y_j$  si  $i \neq j$ , et  $x_i, y_i$  sont voisins dans  $\mathbb{Z}/n\mathbb{Z}$ .

On couple deux chaînes paresseuses  $(X_t, Y_t)$  initialement aux états déterministes  $(x, y)$  en définissant la transition suivante : l'une des  $d$  coordonnées est choisie uniformément. Si les deux chaînes sont identiques sur cette coordonnée, elles font toutes deux le même pas de marche paresseuse dans cette direction. Si les deux chaînes diffèrent par cette coordonnée, alors l'une des deux chaînes choisie au hasard se déplace en un de ses voisins par la coordonnée choisie.

Appelons  $\tau_i$  le temps de couplage de la  $i^{\text{ème}}$  coordonnée. On fixe  $i$  pour alléger les notations dans la preuve qui suit.

Désignons par  $t_k$  le temps où la  $i^{\text{ème}}$  coordonnée est choisie pour la  $k^{\text{ème}}$  fois :  $t_{k+1} = \inf\{t \geq t_k + 1 : \xi_t = i\}$  avec  $t_0 = 0$ . Les  $(t_{k+1} - t_k)_{k \geq 0}$  sont *iid* de loi géométrique, leur espérance vaut  $d$ . Soit  $(\tilde{X}_k^{(i)})_{k \geq 1} = (X_{t_k}(i))_{k \geq 1}$ ,  $(\tilde{Y}_k^{(i)})_{k \geq 1} = (Y_{t_k}(i))_{k \geq 1}$ . Si  $\tau$  désigne le temps d'arrêt défini pour la marche sur le cycle pour le couple  $(\tilde{X}^{(i)}, \tilde{Y}^{(i)})$ , on a l'égalité suivante :  $\tau_i = t_\tau$ , et l'on peut donc calculer :

$$\begin{aligned} \tau_i = t_\tau &= \sum_{n=0}^{+\infty} t_n \mathbb{1}_{\{\tau=n\}} \\ &= \sum_{n=0}^{+\infty} t_n \left( \mathbb{1}_{\{\tau \geq n\}} - \mathbb{1}_{\{\tau \geq n+1\}} \right) \\ &= \sum_{n=0}^{+\infty} (t_{n+1} - t_n) \mathbb{1}_{\{\tau \geq n+1\}} \end{aligned}$$

Les variables  $\tau$  et  $(t_{n+1} - t_n)$  sont indépendantes, donc

$$\mathbb{E}[(t_{n+1} - t_n) \mathbb{1}_{\{\tau \geq n+1\}}] = \mathbb{E}[(t_{n+1} - t_n)] \cdot \mathbb{E}[\mathbb{1}_{\{\tau \geq n+1\}}].$$

Il vient

$$\mathbb{E}[\tau_i] = \sum_{n=0}^{+\infty} \mathbb{E}[t_{n+1} - t_n] \cdot \mathbb{P}\{\tau \geq n+1\} = d \cdot \mathbb{E}[\tau]$$

ce qui n'a rien de surprenant.

Le temps de couplage de  $(X, Y)$  est  $\tau := \max_{1 \leq i \leq d} \tau_i \leq \sum_{i=1}^d \tau_i$ . Son espérance se laisse donc majorer brutalement :  $\mathbb{E}[\tau] \leq \frac{d^2 n^2}{4}$ .

## 4 Modèles de battage de cartes

On va ici décrire quelques processus modélisant des mélanges de tas de  $n$  cartes (typiquement  $n = 52$ ). Les mélanges décrits ici seront des suites de permutations aléatoires appliquées au tas de cartes successivement, chacune correspondant à une étape du battage. On supposera qu'à chaque étape,

c'est le même procédé qui est utilisé pour mélanger le tas, on va donc décrire le mélange comme un processus de Markov  $(X_t)$  de matrice de transition  $P \in \mathbb{R}^{(\mathfrak{S}_n^2)}$  telle que pour une certaine distribution de probabilité  $\mathbb{P}$  sur  $\mathfrak{S}_n$ , on ait  $\forall(\sigma, \sigma') \in \mathfrak{S}_n^2, P(\sigma, \sigma') = \mathbb{P}(\sigma' \sigma^{-1})$ .

**Proposition 4.1.** *Pour que la chaîne soit irréductible, il faut et il suffit que le support de  $\mathbb{P}$  soit une partie génératrice du groupe  $\mathfrak{S}_n$ .*

*Démonstration.* On utilise la caractérisation suivante :  $\mathbb{P}$  est irréductible si et seulement si, pour tous  $\sigma$  et  $\sigma'$ , il existe un entier  $n$  et  $\sigma_0, \dots, \sigma_n$  avec  $\sigma_0 = \sigma$  et  $\sigma_n = \sigma'$  tels que :

$$\forall i \in \{0, \dots, n-1\}, P(\sigma_i, \sigma_{i+1}) = \mathbb{P}\{\sigma_{i+1} \sigma_i^{-1}\} > 0, \text{ c'est-à-dire, } \sigma_{i+1} \sigma_i^{-1} \in \text{Supp } \mathbb{P}.$$

Si la chaîne est irréductible, alors, en choisissant  $\sigma' = Id$ , puis en notant  $\omega_i = \sigma_{i+1} \sigma_i^{-1}$  (avec les notations précédentes), on obtient  $\sigma = \left( \prod_{i=0}^{n-1} \omega_i \right)$ . La réunion des familles  $(\omega_i)_i$  sur l'ensemble des permutations forme donc une partie génératrice de  $\mathfrak{S}_n$  incluse dans le support de  $\mathbb{P}$ .

La réciproque est laissée en exercice.  $\square$

## Mélange par transpositions

L'idée est ici d'utiliser les transpositions, ce qui correspond à l'échange de la position de deux cartes choisies au hasard dans le tas. Formellement, on considère  $\mathbb{U}$  la probabilité uniforme sur  $\{1, \dots, n\}^2$ , et

$$\varphi : \begin{cases} \{1, \dots, n\}^2 & \rightarrow \mathfrak{S}_n \\ (i, j) & \mapsto \tau_{i,j} \end{cases}$$

avec la convention :  $\tau_{i,i} = Id$ . On choisit la probabilité  $\mathbb{P} = \varphi(\mathbb{U})$ , et on obtient ainsi une chaîne irréductible et apériodique, qui a pour distribution stationnaire la probabilité uniforme sur  $\mathfrak{S}_n$ . Le mélange est cependant très lent, intuitivement, c'est une très mauvaise manière de procéder.

## Mélange en rafales

On désigne par le terme anglais de *riffle shuffle* (ou mélange en rafales), la technique de mélange souvent utilisée dans les casinos qui consiste à couper le paquet en deux à peu près au milieu et à les réunir en alternant les cartes des deux tas.

Gilbert et Shannon, dès 1955, ainsi que Reeds indépendamment en 1981, ont proposé un modèle pour ce type de mélanges. Les trois descriptions suivantes sont équivalentes.

**Première description.** On coupe le paquet de  $n$  cartes suivant une distribution binômiale. Si la coupe est effectuée au niveau de la  $p^{\text{ème}}$  carte, on choisit uniformément une rafale parmi les  $\binom{n}{p}$  rafales possibles entre les deux tas.

FIG. 2 – Un mélange par rafales, “*riffle shuffle*”.

**Deuxième description.** On coupe le paquet de  $n$  cartes suivant une distribution binômiale. On applique ensuite l’opération suivante jusqu’à ce que les deux tas soient épuisés : s’il reste dans le paquet de gauche (respectivement de droite)  $a$  cartes (respectivement  $b$  cartes), la carte qui va passer des deux tas au tas produit proviendra avec probabilité  $\frac{a}{a+b}$  (respectivement  $\frac{b}{a+b}$ ) du dessous du paquet de gauche (respectivement de droite).

**Troisième description.** On considère une famille de variables *iid*  $(\xi_k)_{k \in \{1 \dots n\}}$  de loi  $\mathcal{B}(\frac{1}{2})$ . Soit  $\sigma$  la permutation concaténée des deux suites croissantes  $(i)_{i=1 \dots n, \xi_i=0}$  et  $(i)_{i=1 \dots n, \xi_i=1}$ . La probabilité  $\mathbb{P}$  est la loi de la variable aléatoire  $\sigma^{-1}$ .

Les deux premières descriptions sont assez intuitives, et la troisième est celle utilisée dans l’étude mathématique de la vitesse de convergence de ce processus de Markov ([3], [4]).

Ce mécanisme de mélange a la réputation d’être le plus aléatoire parmi ceux qui sont sujets à une coupe de loi binômiale. C’est par exemple celui qui a la plus grande « entropie ». Diaconis a étudié expérimentalement si ce modèle décrivait bien le mélange en rafales. Un test du  $\chi^2$  montre que la concordance est très bonne, bien que les rafales soient légèrement plus « touffues » que les mélanges réels : pour les personnes exercées à la manipulation des cartes, deux cartes de la même main tombent successivement plus rarement que ce que prédit le modèle de Gilbert Shannon Reeds.

Les réalisations de mélanges inverses répétés de  $n$  cartes peuvent être décrites par une matrice binaire à  $n$  lignes, la  $i^{\text{ème}}$  ligne étant associée à la  $i^{\text{ème}}$  carte dans l’ordre de départ du paquet, la coordonnée en  $j$  décrivant le comportement de cette carte lors du  $j^{\text{ème}}$  mélange (par exemple, 1 si la carte provenait du tas de gauche, 0 si elle provenait du tas de droite). Notons  $\tau = \inf\{t : \text{la matrice binaire a des lignes distinctes}\}$ .

On va majorer la distance en variation totale en utilisant une propriété forte de ce temps d’arrêt. Nous allons introduire un nouvel outil permettant de déduire la vitesse de convergence de chaînes de Markov : les temps d’arrêt uniformes.

**Définition 4.1.** Un temps d'arrêt  $\tau$  est dit fortement uniforme si les variables  $\tau$  et  $X_\tau$  sont indépendantes et si  $X_\tau$  est uniforme.

**Lemme 4.1.** Le temps d'arrêt  $\tau$  est fortement uniforme si et seulement si

$$\forall t \in \mathbb{N}, \sigma \in \Omega, \mathbb{P}(X_t = \sigma | \tau = t) = \frac{1}{|\Omega|} = \pi(\sigma).$$

**Lemme 4.2.** Si le temps d'arrêt  $\tau$  est fortement uniforme et si  $\pi$  est la mesure uniforme ( $\pi(A) = \frac{|A|}{|\Omega|}$ ) alors :

$$d(P^t, \pi) \leq \mathbb{P}\{\tau > t\}.$$

*Démonstration.* Si  $A \in \mathcal{P}(\Omega)$ , alors

$$\begin{aligned} P^t(A) &= \mathbb{P}(X_t \in A) \\ &= \sum_{k=0}^t \mathbb{P}(X_t \in A | \tau = k) \mathbb{P}(\tau = k) + \mathbb{P}(X_t \in A | \tau > t) \mathbb{P}(\tau > t) \\ &= \left( \sum_{k=0}^t \frac{|A|}{|\Omega|} \mathbb{P}(\tau = k) \right) + \mathbb{P}(X_t \in A | \tau > t) \mathbb{P}(\tau > t) \\ &= \frac{|A|}{|\Omega|} + \left( \mathbb{P}(X_t \in A | \tau > t) - \frac{|A|}{|\Omega|} \right) \mathbb{P}(\tau > t). \end{aligned}$$

On a donc bien :

$$|P^t(A) - \pi(A)| \leq \mathbb{P}(\tau > t).$$

□

**Lemme 4.3.** Le temps d'arrêt  $\tau$  est fortement uniforme.

Les lemmes 4.1 et 4.3 sont démontrés en [5].

On obtient facilement  $\mathbb{P}\{\tau \leq t\} = \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^t}\right)$ . On a donc, pour  $t = 2 \log_2(n/c)$ ,  $\mathbb{P}\{\tau \leq t\} \sim_n 1 - e^{-\frac{c^2}{2}} \sim_c \frac{c^2}{2}$ . Il y a un *cut off* en  $2 \log(n)$  sur cette borne. Le calcul exact du membre de droite de

$$d(P^t, \pi) \leq 1 - \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^t}\right) = M(t)$$

produit les valeurs suivantes, pour  $n = 52$  :

t	10	11	12	13	14
$M(t)$	.73	.48	.28	.15	.08

Ces chiffres sont assez concluants sur l'efficacité de la méthode de mélange : en 14 mélanges, on trouve en effet un très bon mélange avec cette majoration, qui est pourtant grossière. Une étude plus fine montrerait que 7 mélanges sont en fait suffisants pour atteindre pratiquement l'uniformité. On peut montrer en effet que ce procédé présente un *cut-off* autour du temps  $\frac{3}{2} \log_2 n$ , qui fournit environ 8,550... pour  $n = 52$ .

**Argument de minoration du temps de mélange.** Dans la suite, on identifiera une permutation  $\sigma \in \mathfrak{S}_n$  au  $n$ -uplet  $(\sigma(1), \dots, \sigma(n))$ .

**Définition 4.2.** *On appelle suite montante de la permutation  $\sigma$  une sous-suite croissante maximale.*

**Lemme 4.4.** *Si  $\sigma \neq Id$  peut être obtenue par une itération de riffle shuffle, alors elle contient exactement deux suites montantes.*

Le riffle shuffle a donc la propriété de ne pas trop altérer les suites montantes à chaque étape ([3]) :

**Proposition 4.2.** *Si  $\sigma$  peut être obtenue par  $t$  itérations de riffle shuffle, elle contient au maximum  $2^t$  suites montantes.*

Tant que  $t \leq \log_2 n$ , on ne peut donc pas espérer avoir bien mélangé le paquet, car on a une information déterministe forte sur le paquet : certaines configurations *ne peuvent pas être atteintes*, ainsi la permutation définie par l'équation  $\sigma(i) = n - i + 1$  ne pourra être atteinte que si  $t \geq \log_2 n$ .

Le temps de *cut-off*, en admettant qu'il existe est donc minoré par  $\frac{1}{\log 2} \log n$  et majoré par  $2 \log n$ .

## Note historique

Les suites montantes, invariants fondamentaux du mélange en rafales, ont été découvertes au début du siècle par deux prestidigitateurs, Williams et Jordan. Elles sont à l'origine d'un superbe tour de magie baptisé « Premo » par Jordan. L'opérateur sort un paquet de sa malette, le confie à un spectateur et en détourne son regard. La consigne est la suivante : « Coupez le paquet et mélangez-le en rafales, recommencez et coupez-le une troisième fois. Je suis sûr que vous ne me contredirez pas si je prétends que nul ne saurait prédire la carte qui est au dessus du paquet. Prenez connaissance de cette carte et insérez la n'importe où dans le paquet. Coupez à nouveau le paquet, mélangez-le une nouvelle fois et terminons par une coupe ».

Le prestidigitateur reprend alors le paquet, le répand en un grand arc sur la table et après une inspection attentive et concentrée du paquet, désigne la carte en question.

L'astuce réside dans le fait que les coupes respectent l'ordre cyclique des cartes, et qu'après trois mélanges, le paquet contiendra généralement 8 suites montantes, chacune d'une longueur de 6,5 cartes en moyenne. Le fait de déplacer une carte du haut vers le milieu engendrera donc généralement une aberration dans ces suites montantes, aisément repérable.

Persi Diaconis était, avant sa reconversion dans les mathématiques, un magicien professionnel, ce qui explique son intérêt pour ces questions de mélanges, qui lui valurent en 1992 une certaine médiatisation. Bayer et Diaconis ont simulé informatiquement ce tour de magie et déterminé qu'il avait empiriquement un taux de réussite de 84%, et de 94% si on autorise deux essais, ce qui est raisonnable pour la pratique de la prestidigitation. Une astuce plus sophistiquée consiste

à parier de retrouver la carte déplacée en 26 ( $= 52/2$ ) tentatives après 8 mélanges! Avec le bon algorithme, on peut cependant prétendre à une probabilité de réussite de 0,548.

## Références

- [1] P. Diaconis, The cutoff phenomenon in finite Markov chains, Proc. Natl. Acad. Sci. USA **93**, 1659-1664 (1996).
- [2] Y. Peres, Mixing for Markov Chains and Spin Systems, Lectures given at the 2005 PIMS Summer School in Probability held at the University of British Columbia (2005)<sup>1</sup>.
- [3] P. Biane, Combien de fois faut-il battre un jeu de cartes?, SMF, Gazette **91** (2002).
- [4] D. Bayer, P. Diaconis, Trailing the dovetail shuffle to its lair, The Annals of Applied Probability, vol.2 (1992).
- [5] P. Diaconis, Group representations in probability and statistics, Institute of Mathematical Statistics (1988).

---

<sup>1</sup>En particulier Lecture 7 : Cut-off and Eigenvalues of Reversible Chains.