# Étude de la vitesse de convergence d'une chaîne de Markov ergodique vers sa mesure stationnaire

Charles Bouillaguet
Stéphane Glondu

Mercredi 12 janvier 2005

## 1 Résultats théoriques et première simulation

1. On tire aléatoirement de manière uniforme une matrice stochastique $P$[1] :

$$P \approx \begin{bmatrix} 0.408251041846551977 & 0.129787409172955336 & 0.368142728656067952 & 0.0938188204175708501 \\ 0.318456037710306062 & 0.110769387848808881 & 0.538344141845045087 & 0.0324304322777909729 \\ 0.0745981423184967768 & 0.227008399192856591 & 0.574213960092339382 & 0.124179498525375437 \\ 0.407231489199482711 & 0.00476508262148211178 & 0.167950739034889274 & 0.420052689037131444 \end{bmatrix}$$

   La probabilité qu'un des coefficients de $P$ soit nul est nulle. En effet, soit $X$ une variable aléatoire distribuée uniformément dans $[0, 1]$. Notons $A = \{X = 0\}$ et, pour tout $n \geqslant 1$, posons $A_n = \left[0, \frac{1}{n}\right]$. On a $A = \bigcap_{n \in \mathbb{N}} A_n$, donc pour tout $n \geqslant 1$, on a $A \subset A_n$, d'où :

$$\mathbb{P}\left[A\right] \leqslant \mathbb{P}\left[A_n\right] = \frac{1}{n} \xrightarrow[n \to +\infty]{} 0,$$

   on a donc $\mathbb{P}\left[X = 0\right] = 0$. Ainsi, avec probabilité 1, tous les coefficients de la matrice $P$ sont strictement positifs, donc $P$ est ergodique.

2. L'existence de la loi invariante $\pi$ est assurée par le cours. Ici, on a :

$$\pi \approx \begin{bmatrix} 0.240937340739176248 \\ 0.153316801390551038 \\ 0.459752547028839664 \\ 0.145993310751226485 \end{bmatrix}$$

   Il s'agit d'un vecteur propre à gauche de $P$ pour la valeur propre 1, normalisé.

[1] Les valeurs numériques données ici ont été calculées avec Maple.

3. On tire aléatoirement de manière uniforme une loi $\mu_0$ :

$$\mu_0 \approx \begin{bmatrix} 0.381024491592579728 \\ 0.243105554887164260 \\ 0.146030456012167381 \\ 0.229839497487497630 \end{bmatrix}$$

Remarque : le cours nous assure que $\mu_0$ importe peu.

4. On a :

$$\mu_t^* = \mu_0^* P^t,$$

où $\mu^*$ désigne la matrice ligne du vecteur $\mu$. Pour calculer les $\mu_t$, il suffit donc d'itérer $t$ fois la fonction $\mu \mapsto \mu^* P$ sur $\mu_0$. On constate expérimentalement que pour $t \geqslant t_{\max} = 8$, on est déjà très proche de la loi limite (voir la figure 1).

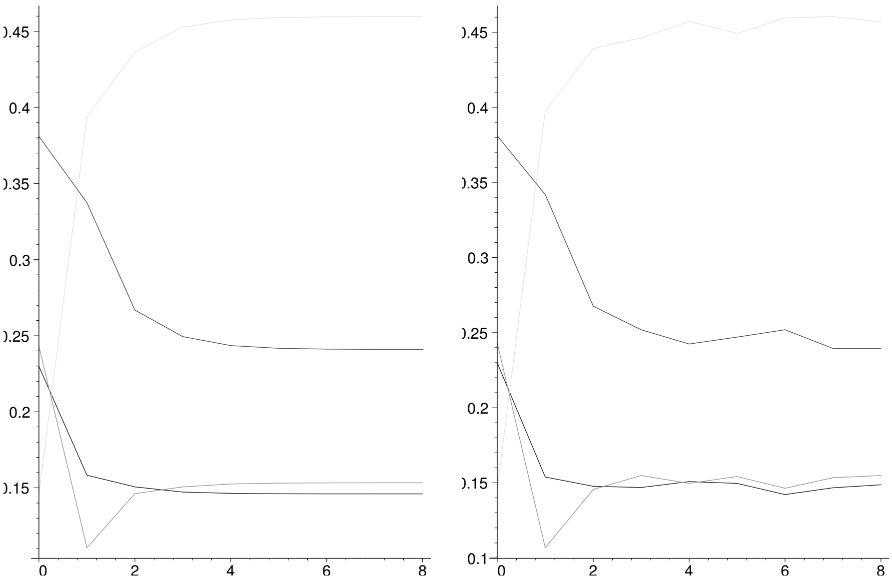


Fig. 1 – Convergence vers la loi limite
Chaque courbe représente l'évolution d'un certain $\mu_t(i)$ en fonction de $t$.
À gauche, on a représenté les valeurs théoriques ;
à droite, les valeurs expérimentales pour $N = 10\,000$.

5. On simule $N$ trajectoires $M^{(1)}, \ldots, M^{(N)}$ et on calcule les mesures empiriques $\hat{\mu}_t$ :

$$\hat{\mu}_t(j) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{M_t^{(i)} = j}.$$

D'après la loi des grands nombres, lorsque $N \to +\infty$, les mesures empiriques $\hat{\mu}_t$ se rapprochent des mesures théoriques $\mu_t$. Il faut l'ajuster

— en faisant appel au théorème limite central, voir section 2 — selon la précision souhaitée pour les $\hat{\mu}_t$. Sur la figure 1, on a pris $N = 10\,000$.

6. On constate que les courbes $t \mapsto \log \|\pi - \mu_t\|_1$ et $t \mapsto \log \|\pi - \hat{\mu}_t\|_1$ (voir figure 2) sont des droites (au moins pour des petites valeurs de $t$), ce qui veut dire que la convergence vers la loi limite est exponentielle. L'ordonnée à l'origine est donnée par $\mu_0$, et la pente $\rho$ peut être déterminée par régression linéaire. On constate que les points s'éloignent plus rapidement de la droite dans le cas de la simulation ; cela est dû à la faible précision sur les $\hat{\mu}_t$ (par rapport aux $\mu_t$).

   Dans le cours, on a montré la convergence (au moins) exponentielle de $\mu_t$ vers $\pi$. On a même la majoration suivante :

   $$\|\pi - \mu_t\|_1 \leqslant (1-a)^t \|\pi - \mu_0\|,$$

   où $a$ est le plus petit coefficient de $P$. On a donc (théoriquement) $\rho \leqslant \log(1-a)$. En pratique,

   $$\begin{aligned} \rho &\approx -1.205560132 \\ e^\rho &\approx 0.2995241794 \\ a &\approx 0.004765082621482111 78, \end{aligned}$$

   donc l'inégalité est largement vérifiée.

7. Dans le cas présent, le spectre de $P$ est :

   $$Sp(P) \approx \left\{ \begin{array}{c} 1.00000000001739520 + 0.0\,i \\ 0.299784914391212520 + 0.0286611166718473697\,i \\ 0.299784914391212520 - 0.0286611166718473697\,i \\ -0.0862827499749896504 + 0.0\,i \end{array} \right\}.$$

   Même les approximations de Maple sur les valeurs théoriques se font légèrement ressentir !

   Nous avons procédé à quelques expériences. Dans un certain nombre d'entre elles, nous avons eu l'impression que $e^\rho$ majorait les valeurs absolues des parties réelles des valeurs propres de $P$. Il semblerait même que ce soit l'une d'entre elles. On a en fait :

   $$|Sp(P)| \approx \left\{ \begin{array}{c} 1.0 \\ 0.3011518795 \\ 0.3011518795 \\ 0.08628274997 \end{array} \right\},$$

   et on vérifie en pratique que $e^\rho$ est le plus grand module d'une valeur propre de $P$ (plus petite que 1).

   Les inévitables erreurs d'arrondi de Maple — la précision des flottants utilisés pour les calculs courants est limitée — ne favorisent pas les devinettes.

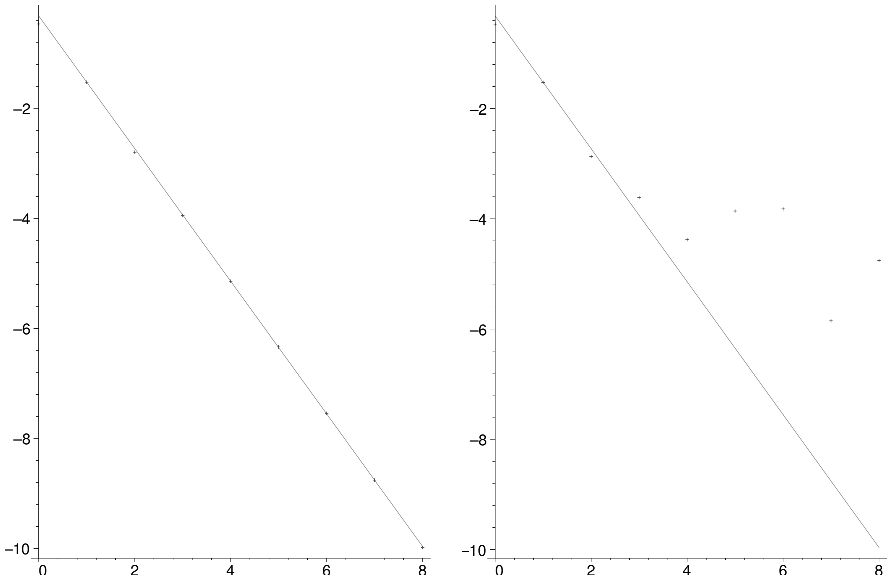


FIG. 2 – Convergence exponentielle vers la loi limite
Les nuages de points représentent $\log \|\pi - \mu_t\|_1$ (à gauche) et $\log \|\pi - \hat{\mu}_t\|_1$ (à droite) en fonction de $t$. La droite est une regression linéaire sur les valeurs théoriques.

# 2 Ajustement de $N$

Nous avons :

$$\hat{\mu}_t(j) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{M_t^{(i)}=j}.$$

On voudrait calculer $N$ tel que $\hat{\mu}_k(j)$ soit proche de $\mu_k(j)$ à $10^{-d}$ près. On va utiliser pour cela le théorème limite centrale (TLC dans la suite). Rappelons son énoncé :

**Théorème 2.1** (Théorème limite centrale)**.** *Soit $(X_j)_{j\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi $\mu$ et de variance $\sigma^2$. On note :*

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

*On a :*

1. $\overline{X}_n \xrightarrow[n\to+\infty]{} m = \mathbb{E}[X_1] = \sum_x x\,\mu(x),$
2. $\mathbb{P}\left[a \leqslant \sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right) \leqslant b\right] \xrightarrow[n\to+\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \mathrm{d}x.$

Comment ce résultat s'applique-t-il à notre problème ? Le TLC nous donne en fait :

$$\begin{aligned}\hat{\mu}_t(j) \longrightarrow \mathbb{E}\left[\mathbf{1}_{\{M_t^{(i)}=j\}}\right] &= \mathbb{P}\left[M_t^{(i)} = j\right] \\ &= \mu_t(j)\end{aligned}$$

d'une part (c'est en fait la loi des grands nombres), mais surtout :

$$\mathbb{P}\left[a \leqslant \sqrt{n}\left(\frac{\hat{\mu}_t(j) - \mu_t(j)}{\sigma}\right) \leqslant b\right] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{x^2}{2}} \mathrm{d}x.$$

Imposons que $\hat{\mu}_t(j)$ soit proche de $\mu_t(j)$ à $10^{-d}$ près avec une probabilité supérieure à 95 %. On a :

$$\frac{1}{\sqrt{2\pi}} \int_{-2}^{2} e^{\frac{x^2}{2}} dx \geqslant 0,95,$$

donc poser $a = -2$ et $b = 2$ est suffisant pour atteindre notre but.

Cherchons ensuite $N$. Nous avons :

$$\begin{gathered}-2 \leqslant \sqrt{N}\left(\frac{\hat{\mu}_t(j) - \mu_t(j)}{\sigma}\right) \leqslant 2 \\ \frac{-2\sigma}{\sqrt{N}} \leqslant \hat{\mu}_t(j) - \mu_t(j) \leqslant \frac{2\sigma}{\sqrt{N}},\end{gathered}$$

et on voudrait imposer :

$$-10^{-d} \leqslant \hat{\mu}_t(j) - \mu_t(j) \leqslant 10^{-d}.$$

Il suffit donc de poser :

$$\frac{2\sigma}{\sqrt{N}} \leqslant 10^{-d}.$$

Ici,

$$\begin{aligned}\sigma^2 &= \operatorname{Var} \mathbf{1}_{\{M_t^{(i)}=j\}} \\ &= \mathbb{E}\left[\left(\mathbf{1}_{\{M_t^{(i)}=j\}} - \mu_t(j)\right)^2\right] \\ &= \mathbb{E}\left[1^2_{\{M_t^{(i)}=j\}}\right] - 2\mu_t(j)\,\mathbb{E}\left[1_{\{M_t^{(i)}=j\}}\right] + \mu_t(j)^2 \\ &= \mathbb{E}\left[1_{\{M_t^{(i)}=j\}}\right] - 2\mu_t(j)^2 + \mu_t(j)^2 \\ &= \mu_t(j) - \mu_t(j)^2 \\ &\leqslant \frac{1}{4},\end{aligned}$$

d'où :

$$\begin{aligned}\sqrt{N} &\geqslant 2\,\sigma\,10^d\\ N &\geqslant 4\,\sigma^2 10^{2d}.\end{aligned}$$

On pourrait calculer

$$N = \max_{\substack{1\leqslant t\leqslant t_{\max}\\ 1\leqslant j\leqslant 4}} \mu_t(j) - \mu_t(j)^2,$$

mais on obtiendra notre objectif en prenant $N$ un peu plus grand que nécessaire (et en s'épargant des calculs) :

$$N \geqslant 10^{2d}.$$

Quelques expériences montrent qu'en calculant le max, on trouve une valeur de $N$ environ 10 % inférieure à $10^{2d}$ .

Si on veut une précision de deux chiffres, on aura $N \geqslant 10\,000$, ce qui fait déjà beaucoup ! À titre d'indication, pour $N = 10\,000$, le calcul prend environ trois minutes. La précision qu'on peut atteindre est donc forcément très vite limitée par les moyens matériels dont nous disposons.