[M. Gubinelli - Processus discrets - M1 MMD 2009/2010 - 20091102 - v.3]

III Théorèmes limites pour les chaînes de Markov

1 Le théorème ergodique

Corollaire 1. *Soit* $N_x = \sum_{n=0}^{\infty} 1_{[X_n = x]}$. *Alors si* x *est transient*

$$\mathbb{P}_x(N_x = k) = (1-a)a^{k-1}, \quad k \geq 1, \quad a = \mathbb{P}_x(T_x < \infty) < 1 \tag{1}$$

et si x *est récurrent on a* $\mathbb{P}_x(N_x = \infty) = 1$.

Démonstration.

$$\mathbb{P}_x(N_x = k) = \mathbb{P}_x\Big(T_x^{i+1} - T_x^i < \infty, i = 0, \ldots, k-1; T_x^{k+1} - T_x^k = \infty\Big) = (1-a)a^{k-1}$$

□

Théorème 2. *Soit* P *une matrice irréductible récurrente positive et* π *sa probabilité stationnaire. Alors pour tout* $x, y \in M$

$$\mathbb{P}_x\left(\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{[X_k = y]} = \pi(y)\right) = 1. \tag{2}$$

Démonstration. Par la récurrence on a que $T_y^k < +\infty$ pour tout $k \geqslant 1$, donc les v.a. $\tau_y^k = T_y^{k+1} - T_y^k$ sont bien définies pour tout $k \geqslant 1$ et par le théorème de régénération on a que la suite $(\tau_y^k)_{k \geqslant 1}$ est iid et tel que $\mathbb{P}_x(\tau_y^k = m) = \mathbb{P}_y(T_y = m)$. D'après la loi des grands nombres

$$\frac{1}{k} \sum_{i=1}^{k-1} \tau_y^i \underset{k \to \infty}{\longrightarrow} \mathbb{E}_y(T_y) = \frac{1}{\pi(y)} \qquad \mathbb{P}_x - p.s.$$

et $\mathbb{P}_x(T_y < +\infty) = 1$ donc

$$\frac{1}{k} T_y^k = \frac{1}{k}\left(\sum_{i=1}^{k-1} \tau_y^i + T_y\right) \underset{k \to \infty}{\longrightarrow} \mathbb{E}_y(T_y^1) = \pi(y)^{-1} \qquad \mathbb{P}_x - p.s.$$

Par ailleurs, si pour tout $n \geq 1$ on pose $N_y^n = \sum_{k=0}^{n-1} 1_{[X_k = y]}$ alors on a que

$$T_y^{N_y^n} \leqslant n \leqslant T_y^{N_y^n + 1}$$

et donc

$$\frac{N_y^n}{T_y^{N_y^n + 1}} \leqslant \frac{1}{n} \sum_{k=0}^{n-1} 1_{[X_k = y]} \leqslant \frac{N_y^n}{T_x^{N_y^n}}.$$

D'où le résultat. □

Remarque 3. En modifiant légèrement cette preuve, on obtient le résultat plus général suivant.

Corollaire 4. *Soit* P *une matrice irréductible récurrente positive et* π *sa probabilité stationnaire. Soit* f *une fonction dans* $L^1(\pi)$, *i.e.* $\sum_{x \in M} |f(x)| \pi(x) < \infty$. *Alors*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{x \in M} f(x) \pi(x) \qquad \mathbb{P}_x - p.s.$$

2 Convergence vers l'équilibre

Un corollaire du théorème ergodique est que lorsque P est une matrice de transition irréductible récurrente positive de probabilité invariante π

$$\frac{1}{n}\sum_{j=0}^{n-1} \mu P^j \xrightarrow[n\to\infty]{\text{loi}} \pi,$$

pour toute probabilité $\mu\in\Pi(M)$. Ceci n'implique pas en général que $\mu P^n \to \pi$:

Exemple 5. La matrice

$$P = \begin{pmatrix} 0 & 1\\ 1 & 0 \end{pmatrix}$$

a pour loi stationnaire $\pi = (\frac{1}{2}, \frac{1}{2})$ et comme puissances

$$P^{2n} = \begin{pmatrix} 1 & 0\\ 0 & 1 \end{pmatrix} \quad \text{et} \quad P^{2n+1} = \begin{pmatrix} 0 & 1\\ 1 & 0 \end{pmatrix}.$$

Donc $\mu P^n = \mu$ si n est pair et $\mu P^n = \mu P$ si n est impair et on voit que μP^n ne converge pas vers π.

On cherche maintenant des conditions sur P pour que $\mu P^n \to \pi$. On cherche également à estimer la vitesse de cette convergence.

Définition 6. *On definit une distance d sur $\Pi(M)$ par*

$$d(\mu,\nu) = \sup_{C\subseteq M} |\mu(C) - \nu(C)| = \frac{1}{2}\sum_{x\in M} |\mu(x) - \nu(x)|.$$

On a $d(\mu,\nu) \leqslant 1$ et $d(\mu,\nu) = 0 \Leftrightarrow \mu = \nu$.

Démonstration. Il est clair que $d(\mu,\nu) \leqslant 1$ (car si $\mu(A) > \nu(A)$ alors $|\mu(A) - \nu(A)| \leqslant \mu(A) \leqslant 1$ pour des probabilités) et que $d(\mu,\nu) = 0 \Rightarrow \mu = \nu$. Pour montrer l'equivalence des deux expressions de $d(\mu,\nu)$ on observe que

$$d(\mu,\nu) = \sup_{C\subseteq M} \left|\sum_{x\in C} [\mu(x) - \nu(x)]\right| \leqslant \sup_{C\subseteq M}\sum_{x\in C} |\mu(x) - \nu(x)| \leqslant \sum_{x\in A} |\mu(x) - \nu(x)|$$

et si on pose $A = \{x\in M \colon \mu(x) > \nu(x)\}$ on obtient l'inegalité opposée:

$$\begin{aligned}\frac{1}{2}\sum_{x\in M} |\mu(x) - \nu(x)| &= \frac{1}{2}\Big(\sum_{x\in A} [\mu(x) - \nu(x)] + \sum_{x\notin A} [\nu(x) - \mu(x)]\Big)\\ &= \frac{1}{2}(\mu(A) - \nu(A) + \nu(A^c) - \mu(A^c)) \leqslant d(\mu,\nu).\end{aligned}$$

□

Dans la suite on notera $P^n(x, A) = \sum_{y\in A} P^n(x, y) = \mathbb{P}_x(X_n \in A)$ et on utilisera le lemme suivante

Lemme 7. *Soit $f\colon M \to \mathbb{R}$ bornée alors*

$$\inf_{c\in\mathbb{R}} \sup_{x\in M} |f(x) - c| = \frac{1}{2}\sup_{x,y\in M} |f(x) - f(y)|$$

Démonstration. D'une part, pour tout $c\in\mathbb{R}$

$$\sup_{x,y\in M} |f(x) - f(y)| \leqslant \sup_{x,y\in M} (|f(x) - c| + |f(y) - c|) = 2\sup_{x\in M} |f(x) - c|$$

et donc

$$\inf_{c\in\mathbb{R}} \sup_{x\in M} |f(x) - c| \geqslant \frac{1}{2}\sup_{x,y\in M} |f(x) - f(y)|.$$

D'autre part on a que

$$\sup_{x,y\in M}|f(x)-f(y)|=\sup_{x\in M}f(x)-\inf_{y\in M}f(y),$$

soit alors $\hat{c}=(\sup_{x\in M}f(x)+\inf_{y\in M}f(y))/2$:

$$\hat{c}-\frac{1}{2}\sup_{x,y\in M}|f(x)-f(y)|=\inf_{x\in M}f(x)\leqslant f(x)\leqslant\sup_{x\in M}f(x)=\hat{c}+\frac{1}{2}\sup_{x,y\in M}|f(x)-f(y)|$$

ce qui donne

$$\inf_{c\in\mathbb{R}}\sup_{x\in M}|f(x)-c|\leqslant|f(x)-\hat{c}\,|\leqslant\frac{1}{2}\sup_{x,y\in M}|f(x)-f(y)|.$$

□

Proposition 8. *Soit P une matrice de transition sur M. On pose*

$$\rho_n=\sup_{x,y\in M}d(P^n(x,\cdot),P^n(y,\cdot))\tag{3}$$

Alors,

1. $\rho_{n+m}\leq\rho_n\rho_m$
2. *Soit $\rho_n=1$ pour tout $n\geq 1$, soit $\exists C<\infty$ et $\theta<1$ tels que $\rho_n\leq C\theta^n$ $\forall n\geq 1$.*
3. *Dans le cas $\rho_n\leq C\theta^n$ avec $\theta<1$, il existe une unique probabilité stationnaire π et de plus,*

$$|P^n(x,y)-\pi(y)|\leq C\theta^n,\forall n\geq 1,x,y\in M.\tag{4}$$

Donc, lorsqu'il existe un indice n_0 tel que $\rho_{n_0}<1$, on a (4), ce qui implique que $\mu P^n\to\pi$, $\forall\mu\in\Pi(M)$, et que ces convergences ont lieu à des vitesses décroissant exponentiellement vite.

Démonstration.

$$\begin{aligned}
\left|P^{n+m}(x,A)-P^{n+m}(y,A)\right| &= \left|\sum_{z\in M}P^m(z,A)[P^n(x,z)-P^n(y,z)]\right|\\
&= \inf_{c\in\mathbb{R}}\left|\sum_{z\in M}[P^m(z,A)-c][P^n(x,z)-P^n(y,z)]\right|\\
&\leqslant \inf_{c\in\mathbb{R}}\sum_{z\in M}|P^m(z,A)-c||P^n(x,z)-P^n(y,z)|\\
&\leqslant \sum_{z\in M}|P^n(x,z)-P^n(y,z)|\inf_{c\in\mathbb{R}}\sup_{z\in M}|P^m(z,A)-c|\\
&\leqslant \frac{1}{2}\sum_{z\in M}|P^n(x,z)-P^n(y,z)|\sup_{z,z'\in M}|P^m(z,A)-P^m(z',A)|\\
&\leqslant \rho_n\rho_m
\end{aligned}$$

et (1) s'ensuit.

Pour montrer (2), supposons qu'il existe n_0 tel que $\rho_{n_0}<1$. Soit alors $m<n_0$ tel que $n=\left[\frac{n}{n_0}\right]n_0+m$. Alors, par (1), on a

$$\rho_n\leq\rho_{\left[\frac{n}{n_0}\right]n_0}\rho_m\leq\rho_{n_0}^{\left[\frac{n}{n_0}\right]}=(\rho_{n_0}^{n_0^{-1}})^{\left[\frac{n}{n_0}\right]n_0}\leq(\rho_{n_0}^{-n_0^{-1}})^m(\rho_{n_0}^{n_0^{-1}})^n.$$

Donc, si on pose $C=(\rho_{n_0}^{-n_0^{-1}})^m$ et $\theta=\rho_{n_0}^{n_0^{-1}}$, on obtient bien $\rho_n\leq C\theta^n$.

Pour le dernier point, on observe que

$$|P^{n+m}(x,y)-P^n(x,y)|=|\sum_z P^m(x,z)(P^n(z,y)-P^n(x,y)|\leq\rho_n\leq C\theta^n,$$

donc $P^n(x,\ y)$ est une suite de Cauchy. Soit $\pi_x(y)$ sa limite. On voit facilement qu'elle ne dépend pas de x, puisque

$$|P^n(x,y)-P^n(x',y)|\leq\rho_n\to 0.$$

On note π la limite de $P^n(x,\cdot)$. On voit que π est stationnaire, donc que

$$\begin{aligned}|P^n(x,y)-\pi(y)| &= |P^n(x,y)-\pi P^n(y)|\\ &= \left|\sum_z \pi(z)(P^n(x,y)-P^n(z,y))\right|\\ &\leq \rho_n.\end{aligned}$$

□

2.1 Chaînes fortement irréductibles

Définition 9. *Une matrice de transition P est dite fortement irréductible s'il existe $n_0\in\mathbb{N}$ tel que pour tous $x,y\in M$ on ait $P^{n_0}(x,y)>0$.*

Proposition 10. *Si $|M|<\infty$ et si P est fortement irréductible, alors il existe n_0 tel que $\rho_{n_0}<1$.*

Démonstration. Par irréductibilité forte, il existe n_0 tel que $P^{n_0}(x,y)>0$ $\forall x,y\in M$. Soit $A=\{z\in M: P^{n_0}(x,z)\geqslant P^{n_0}(y,z)\}$ on a que

$$\rho_{n_0}(x,y)=\frac{1}{2}\sum_z |P^{n_0}(x,z)-P^{n_0}(y,z)|$$

$$=\frac{1}{2}\sum_{z\in A}(P^{n_0}(x,z)-P^{n_0}(y,z))+\frac{1}{2}\sum_{z\notin A}(P^{n_0}(y,z)-P^{n_0}(x,z))$$

$$\leqslant 1-\frac{1}{2}\sum_{z\in A}P^{n_0}(y,z)-\frac{1}{2}\sum_{z\notin A}P^{n_0}(x,z)<1$$

et puisque M est fini, on en déduit que $\rho_{n_0}=\sup_{x,y}\rho_{n_0}(x,y)<1$. □

Et donc, dans ce cas, d'après la formule (4), μP^n converge exponentiellement vite vers l'(unique) probabilité stationnaire.

Exemple 11.

1. La matrice
$$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$
n'est pas fortement irréductible.
2. La matrice
$$\begin{pmatrix}p & 1-p\\ 0 & 1\end{pmatrix}$$
n'est pas irréductible.
3. La matrice
$$\begin{pmatrix}p & 1-p\\ 1 & 0\end{pmatrix}$$
est fortement irréductible si $0<p<1$.

2.2 Aperiodicité

Définition 12. *Soit $x\in M$ et $R(x)=\{n\in\mathbb{N}: P^n(x,x)>0\}$. La période $p(x)$ de x est le plus grand commun diviseur de $R(x)$.*

Proposition 13. *Supposons que P est irréductible. Alors tous les points de M ont la même période.*

Démonstration. Soient x, $y \in M$ et $n_1 = n(x, y)$, $n_2 = n(y, x)$ tels que $P^{n_1}(x, y) > 0$, $P^{n_2}(y, x) > 0$. Alors

$$P^{n_1+n_2}(x,x) = \sum_{z\in M} P^{n_1}(x,z)P^{n_2}(z,x) \geq P^{n_1}(x,y)P^{n_2}(y,x) > 0$$

donc $n_1 + n_2 \in R(x)$. Si $r \in R(y)$, on a

$$P^{n_1+r+n_2}(x,x) \geq P^{n_1}(x,y)P^r(y,y)P^{n_2}(y,x) > 0$$

donc $n_1 + r + n_2 \in R(x)$. Par définition, la période $p(x)$ est un diviseur de $n_1 + n_2$ et de $n_1 + r + n_2$, donc $p(x)$ divise r, i.e. $p(x) \leq p(y)$. Si on répète l'argument en échangeant les rôles de x et y, on obtient $p(y) \leq p(x)$. Donc $p(x) = p(y)$, ceci $\forall x, y \in M$. □

Donc si P est irréductible, on peut parler de *période* de P.

Exemple 14.

1. La période de
$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
est 2.
2. La période de la marche aléatoire sur $\mathbb{Z}$ est également 2.

Lorsque P est irréductible de période 1, P est dite *apériodique*.

Proposition 15. *Supposons que P est irréductible. S'il existe $x \in M$ tel que $P(x,x) > 0$, alors P est apériodique.*

Démonstration. S'il existe $x \in M$ tel que $P(x,x) > 1$ alors $p(x) = 1$. Mais P est irréductible, donc tous les points ont période 1. □

Lemme 16. *Soit $A \subseteq \mathbb{N}$ un ensemble stable par addition et de p.g.c.d. égal à 1. Alors $A = \{N, N+1, \ldots\}$ pour un quelque $N \in \mathbb{N}$.*

Démonstration. L'ensemble $B = \{x - x' : x, x' \in A \cup \{0\}\}$ est un sous-groupe de $\mathbb{Z}$ et donc de la forme $d\mathbb{Z}$ où d est le plus petit élément non nul de B. De plus $A \subseteq B$ et état tout élément de A divisible par d on conclut que $d = 1$ et donc qu'il existe $a, b \in A \cup \{0\}$ tels que $a = b + 1$. Or $a \in A$ nécessairement. Si $b = 0$ alors $1 \in A$ et la preuve est terminé. Sinon $N = b^2$ convient puisque si $n \geqslant N$ on peut écrire $n = b^2 + b\,q + r$ avec $0 \leqslant r < b$ et donc $n = b(b + q - r) + r(b+1) \in A$. □

Proposition 17. *Une chaîne finie irréductible est fortement irréductible si et seulement si elle est apériodique.*

Démonstration. Il est claire que une chaîne fortement irréductible est apériodique car pour tout $x \in M$ $R(x) \subseteq \{N, N+1, \ldots\}$ pour un certain N. Considérons donc une chaîne apériodique et irréductible: l'ensemble $R(x)$ est stable par addition et son p.g.c.d. est 1. Par le lemme précèdent il existe $N(x)$ tel que $P^n(x,x) > 0$ pour tout $n \geqslant N(x)$. Par irréductibilité, pour tout x, y, $z \in M$, ils existent n_1 et n_2 tels que $P^{n_1}(x, z)P^{n_2}(z, y) > 0$ et donc $P^{n_1+n_2+n}(x, y) \geqslant P^{n_1}(x, z)P^n(z,z)P^{n_2}(z, y) > 0$ ce qui donne que $P^n(x,y) > 0$ pour tout $n \geqslant N(x,y) = n_1 + n_2 + N(z)$. Si on pose $N = \max_{x,y} N(x, y) < +\infty$ alors pour tout $x, y \in M$ et $n \geqslant N$ on a que $P^n(x, y) > 0$ ce qui signifie que la chaîne est fortement irréductible. □