Chaînes de Markov

Léo Gayral

Ces notes sont basées sur le cours de Randal Douc et Éric Moulines. On pourra se référer à l'ouvrage *Markov Chains* (Douc, Moulines, Priouret, Soulier, 2018) pour une lecture complémentaire.

Table des matières

1 Introduction aux chaînes de Markov

1.1 Généralités

Définition 1 (Espace de probabilité filtré) **:**
Un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est dit filtré lorsqu'il est muni d'une filtration $(\mathcal{F}_k)_{k\in\mathbb{N}}$.

Définition 2 (Chaîne de Markov) **:**
On considère un espace mesurable $(\mathbb{X}, \mathcal{X})$. Un processus (X_k) sur $\mathbb{X}$ est une chaine de Markov si, pour toute fonction bornée mesurable $f \in L^\infty(\mathbb{X})$, on a $\mathbb{E}[f(X_{k+1})|\mathcal{F}_k] = \mathbb{E}[f(X_{k+1})|X_k]$.

De façon équivalente, pour toute variable bornée $Y \in \sigma(X_j,\, j \geq k)$, on a $\mathbb{E}[Y|\mathcal{F}_k] = \mathbb{E}[Y|X_k]$.

Définition 3 (Noyau markovien) **:**
Un noyau de $(\mathbb{X}, \mathcal{X})$ dans $(\mathbb{Y}, \mathcal{Y})$ est une application $N : \mathbb{X} \times \mathcal{Y} \to \overline{\mathbb{R}^+}$ telle que :

1. $\forall x \in \mathbb{X}$, $A \mapsto N_x(A)$ est une mesure sur $(\mathbb{Y}, \mathcal{Y})$,
2. $\forall A \in \mathcal{Y}$, $x \mapsto N_x(A)$ est mesurable.

On dit que ce noyau est borné lorsque $\sup\limits_{x\in\mathbb{X}} N_x(\mathbb{Y}) < \infty$.

On dit que ce noyau est markovien lorsque, pour tout $x \in \mathbb{X}$, on a $N_x(\mathbb{Y}) = 1$.

Remarque 4 (Opérateur sur les fonctions) **:**
Un noyau N induit un opérateur linéaire sur les fonctions mesurables positives sur $\mathbb{Y}$, qui à f associe $Nf(x) = \int\limits_{\mathbb{Y}} f(y)\,\mathrm{d}N_x(y)$.

On a alors $\|Nf\|_\infty \leq \sup\limits_{x\in\mathbb{X}} N_x(\mathbb{Y}) \times \|f\|_\infty$. Autrement dit, si le noyau est borné, on obtient un opérateur linéaire continu.

Remarque 5 (Opérateur sur les mesures) **:**
Un noyau N induit un opérateur linéaire sur les mesures sur $(\mathbb{X}, \mathcal{X})$. Pour $A \in \mathcal{Y}$ on pose $\mu N(A) = \int\limits_{\mathbb{X}} N_x(A)\,\mathrm{d}\mu(x)$. On peut vérifier que μN est bien une mesure sur $(\mathbb{Y}, \mathcal{Y})$.

Lorsque N est markovien et μ une probabilité sur $\mathbb{X}$ alors μN est une probabilité sur $\mathbb{Y}$.

Lemme 6 (Composition de noyaux) **:**

Si on a deux noyaux M de $\mathbb{X}$ vers $\mathbb{Y}$ et N de $\mathbb{Y}$ dans $\mathbb{Z}$, alors on peut définir le noyau MN de $\mathbb{X}$ dans $\mathbb{Z}$ via :

$$\forall (x, A) \in \mathbb{X} \times \mathcal{Z},\ MN_x(A) = \int_{\mathbb{Y}} N_y(A)\,\mathrm{d}M_x(y)\,.$$

En particulier, on peut itérer un noyau de $\mathbb{X}$ dans lui-même.

Définition 7 (Produit tensoriel) **:**

Si on a des noyaux M de $\mathbb{X}$ dans $\mathbb{Y}$ et N de $\mathbb{X}$ dans $\mathbb{Z}$, on peut alors définir le noyau tensorisé $M \otimes N$ de $\mathbb{X}$ dans $\mathbb{Y} \times \mathbb{Z}$ muni de la tribu produit, via :

$$\forall x \in \mathbb{X},\ \forall A \in \mathcal{X} \otimes \mathcal{Y},\ M \otimes N_x(A) = \int_{\mathbb{Y}} \int_{\mathbb{Z}} \mathbb{1}_A(y, z)\,\mathrm{d}N_x(z)\,\mathrm{d}M_x(y)\,.$$

Cette opération est associative, autrement dit $(M \otimes N) \otimes P = M \otimes (N \otimes P)$.

Définition 8 (Chaîne de Markov homogène) **:**

On considère un espace probabilisé filtré $\big(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k \in \mathbb{N}}, \mathbb{P}\big)$ et un processus (X_k) adapté à la filtration, à valeurs dans $(\mathbb{X}, \mathcal{X})$.

On dit que (X_k) est une chaîne de Markov homogène, de noyau P, lorsque pour tout $k \in \mathbb{N}$ et $A \in \mathcal{X}$, on a $\mathbb{P}(X_{k+1} \in A | \mathcal{F}_k) = P_{X_k}(A)$ ($\mathbb{P}$-presque-sûrement).

Remarque 9 :

Notons que, si X est une chaîne de Markov pour une filtration $(\mathcal{F}_k)$ quelconque, alors c'est une chaîne de Markov pour sa filtration canonique.

Remarque 10 (Modèle auto-régressif fonctionnel) **:**

On peut par exemple considérer $\mathbb{X} = \mathbb{R}$ muni de la tribu borélienne, $h : \mathbb{R} \to \mathbb{R}$ mesurable, et les variables réelles $(Z_k) \overset{\text{iid}}{\sim} \mu$. En partant de X_0 indépendant de Z, on obtient une marche aléatoire via $X_{k+1} = h(X_k) + Z_{k+1}$.

On considère f positive mesurable, et la filtration $\mathcal{F}_k = \sigma(X_0, Z_1, \ldots, Z_k)$. Dans ce cas :

$$\begin{aligned} \mathbb{E}[f(X_{k+1})|\mathcal{F}_k] &= \mathbb{E}[f(h(X_k) + Z_{k+1})|\mathcal{F}_k] \\ &= \int_{\mathbb{R}} f(h(X_k) + z)\,\mathrm{d}\mu(z) \\ &= \mathbb{E}[f(X_{k+1})|X_k] \end{aligned}$$

donc (X_k) est bien une chaîne de Markov. Son noyau est $P_x(A) = \int_{\mathbb{R}} \mathbb{1}_A(h(x) + z)\,\mathrm{d}\mu(z)$.

Remarque 11 (Loi invariante) **:**

De façon générale, si X_0 a la loi μ, alors μN est la loi de X_1, et plus généralement μN^k est la loi de X_k pour tout $k \in \mathbb{N}$.

On s'intéresse souvent à l'existence et l'unicité d'une loi invariante sous P, une mesure π telle que $\pi P = \pi$, auquel cas on peut étudier la convergence éventuelle de μP^k vers π.

Remarque 12 (Loi jointe) **:**

On s'intéresse désormais à la loi jointe du vecteur $(X_0, \ldots, X_k)$.

Il suffit pour cela d'étudier les fonctions du type $f(x) = \prod_{i=0}^{k} f_i(x_i)$. On a :

$$\begin{aligned}
\mathbb{E}[f(X_0, \ldots, X_k)] &= \mathbb{E}\left[\prod_{i=1}^{k-1} f_i(X_i)\mathbb{E}[f_k(X_k)|X_{k-1}]\right] \\
&= \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(X_i) \times Pf_k(X_{k-1})\right] \\
&= \int_{\mathbb{X}^k} \prod_{i=0}^{k} f_i(x_i)\, \mathrm{d}P_{x_{k-1}}(x_k) \ldots \mathrm{d}P_{x_0}(x_1)\, \mathrm{d}\mu(x_0)\,.
\end{aligned}$$

Pour une loi initiale μ, on notera alors $\mu \otimes P^{\otimes k}$ la mesure de la famille $(X_0, \ldots, X_k)$.

Définition 13 (Réversibilité) **:**

Une mesure ξ sur $(\mathbb{X}, \mathcal{X})$ est P-réversible si la mesure tensorisée $\xi \otimes P$ est symétrique.

Autrement dit, pour tous évènements A et B, $\xi \otimes P(A \times B) = \xi \otimes P(B \times A)$.

Dans ce cas, on a :

$$\mathbb{E}_\xi[f(X_0, X_1)] = \int_{\mathbb{X}^2} f(x_0, x_1)\, \mathrm{d}P_{x_0}(x_1) d\xi(x_0) = \int_{\mathbb{X}^2} f(x_0, x_1)\, \mathrm{d}P_{x_1}(x_0)\, \mathrm{d}\xi(x_1) = \mathbb{E}_\xi[f(X_1, X_0)]\,.$$

Si ξ est une mesure réversible par rapport à P, alors c'est une mesure invariante :

$$\xi P(A) = \mathbb{E}_\xi[\mathbb{1}_A(X_1)] = \mathbb{E}_\xi[\mathbb{1}_A(X_0)] = \xi(A)\,.$$

1.2 Algorithme de Monte Carlo par Chaînes de Markov

On s'intéresse à une densité $\pi > 0$ pour une mesure λ sur $\mathbb{R}^n$, généralement connue à une constante de normalisation multiplicative près. L'idée est de construire un noyau P pour lequel la mesure $\pi\, \mathrm{d}\lambda$ est réversible.

L'algorithme de Metropolis découle du processus suivant. On part de $X_0 \sim \mu$ de loi relativement arbitraire. Pour la transition à partir de X_k, on propose un mouvement $Y_{k+1} = X_k + Z_{k+1}$, avec les $(Z_i) \overset{\text{iid}}{\sim} q\, \mathrm{d}\lambda$ à densité symétrique, de sorte que $q(-x) = q(x)$. On accepte la proposition (autrement dit $X_{k+1} = Y_{k+1}$) avec probabilité $\alpha(X_k, Y_{k+1})$, où $\alpha(x,y) := \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$, et sinon on rejette la proposition (auquel cas $X_{k+1} = X_k$).

Lemme 14 :

Ce processus est une chaîne de Markov par construction. Son noyau est :

$$P_x(A) = \int_A \alpha(x,y) q(y-x) \, \mathrm{d}\lambda(y) + \mathbb{1}_A(x) \overline{\alpha}(x)$$

où $\overline{\alpha}(x) = \int_{\mathbb{R}^n} (1 - \alpha(x,y)) q(y-x) \, \mathrm{d}\lambda(y)$ est la probabilité de rejet.

Lemme 15 :

La mesure $\pi \, \mathrm{d}\lambda$ est P-réversible.

Démonstration. On veut obtenir $\pi(x) dP_x(y) d\lambda(x) = \pi(y) dP_y(x) d\lambda(y)$. Cette propriété découle du fait que :

$$\begin{aligned} \pi(x)\alpha(x,y)q(y-x) &= \max(\pi(x), \pi(y)) \times q(y-x)\,, \\ &= \max(\pi(x), \pi(y)) \times q(x-y)\,, \\ &= \pi(y)\alpha(y,x)q(y-x)\,. \end{aligned}$$

□

On aimerait désormais obtenir la convergence de μP^k vers π. Pour cela, il faut choisir quelle topologie utiliser. On va montrer que, sous certaines conditions, $d_{\mathrm{VT}}\left(\mu P^k, \pi\right) = O\left(\rho^k\right)$ converge géométriquement vers 0, pour un certain $0 < \rho < 1$.

Définition 16 (Distance en variation totale) **:**

Soit ξ une mesure signée finie. Par Hahn-Jordan on a la décomposition $\xi = \xi^+ - \xi^-$, avec ξ^+ et ξ^- des mesures positives singulières. Si ξ est à densité h par rapport à la mesure de Lebesgue, alors ξ^+ est à densité h^+, ξ^- à densité h^- et $|\xi|$ à densité $|h|$.

On pose $\|\xi\|_{\mathrm{VT}} = |\xi|(\mathbb{X})$.

Pour deux mesures de probabilités ξ et ξ', on a alors $d_{\mathrm{VT}}(\xi, \xi') = \frac{1}{2}\|\xi - \xi'\|_{\mathrm{VT}}$.

Lemme 17 :

On a les propriétés suivantes :

1. $\|\xi\|_{\mathrm{VT}} = \sup\{\xi(f),\ f \in L^\infty(\mathbb{X}),\ \|f\|_\infty \leq 1\}$,
2. Si $\xi(\mathbb{X}) = 0$, alors $\|\xi\|_{\mathrm{VT}} = 2\sup\{\xi(f),\ f \in L^\infty(\mathbb{X}),\ \mathrm{osc}(f) \leq 1\}$ où on définit l'amplitude de f par $\mathrm{osc}(f) = \sup\limits_{x,x' \in \mathbb{X}} |f(x) - f(x')|$.

Corollaire 18 :
Si $f \in L^\infty$ et ξ, ξ' des mesures de probabilité, alors $|\xi(f) - \xi'(f)| \leq \frac{1}{2} d_{\mathrm{VT}}(\xi, \xi') \mathrm{osc}(f)$.

Remarque 19 (Couplage de variables) **:**
Si on considère deux variables $X \sim f \,\mathrm{d}\lambda$ et $X' \sim f' \,\mathrm{d}\lambda$, alors on peut établir un couplage entre X et X' de sorte que $\mathbb{P}(X = X') = d_{\mathrm{VT}}(X, X')$.

Pour ce faire, posons $\varepsilon = \int\limits_{\mathbb{X}} \min(f, f') \,\mathrm{d}\lambda$. On a $f = (1-\varepsilon) \times \frac{f - \min(f,f')}{1-\varepsilon} + \varepsilon \times \frac{\min(f,f')}{\varepsilon}$, et de même pour f'.

Définissons Y la variable à densité $\frac{\min(f,f')}{\varepsilon}$, ainsi que Z (resp. Z') à densité $\frac{f-\min(f,f')}{1-\varepsilon}$ (resp. $\frac{f-\min(f,f')}{1-\varepsilon}$), ainsi qu'une variable $B \sim \mathcal{B}(\varepsilon)$ indépendante de Y, Z et Z'. Dans ce cas, les densités de Z et Z' ne sont jamais simultanément positives donc $\mathbb{P}(Z = Z') = 0$. Remarquons en outre que $X \overset{\mathrm{d}}{=} \overline{X} = BY + (1-B)Z$ et $X' \overset{\mathrm{d}}{=} \overline{X'} = BY + (1-B)Z'$, d'où $d_{\mathrm{VT}}(X, X') = \varepsilon = \mathbb{P}\big(\overline{X} = \overline{X'}\big)$.

Définition 20 (Coefficient de Dobrushin) **:**
Soit P un noyau sur $(\mathbb{X}, \mathcal{X})$. Son coefficient de Dobrushin $\Delta(P)$ est sa constante de Lipschitz pour la distance en variation totale sur les mesures de probabilité. Autrement dit :

$$\Delta(P) = \sup_{\xi \neq \xi'} \frac{d_{\mathrm{VT}}(\xi P, \xi' P)}{d_{\mathrm{VT}}(\xi, \xi')} .$$

Lorsque P est Markovien, on a toujours $\Delta(P) \leq 1$.

Lemme 21 :
La distance d_{VT} se comporte comme la norme L^1 sur les probabilités. L'opérateur P laisse donc stable un convexe fermé dans un espace complet.

Si $\Delta(P) < 1$, par théorème de point fixe, il existe une unique mesure invariante π. En particulier, on a alors $d_{\mathrm{VT}}\big(\mu P^k, \pi\big) = O\big(\Delta(P)^k\big)$.

Proposition 22 :
On a l'égalité $\Delta(P) = \sup_{x,x' \in \mathbb{X}} d_{\mathrm{VT}}(P_x, P_{x'})$.

Démonstration. Comme $P_x = \delta_x P$ et pour $x \neq x'$, on a $d_{\mathrm{VT}}(\delta_x, \delta_{x'}) = 1$, on a naturellement l'inégalité $\Delta(P) \geq \sup_{x,x' \in \mathbb{X}} d_{\mathrm{VT}}(P_x, P_{x'})$ par restriction à un sous-ensemble.

Si on considère deux probabilités ξ et ξ', on peut définir la mesure signée $\eta = \xi - \xi'$ de masse nulle. La mesure ηP est également de masse nulle. En outre :

$$\begin{aligned}
\|\eta P\|_{\mathrm{VT}} &= 2 \sup_{\mathrm{osc}(f)\leq 1} (\eta P)f \\
&= 2 \sup_{\mathrm{osc}(f)\leq 1} \eta(Pf) \\
&\leq 2 \sup_{\mathrm{osc}(f)\leq 1} \tfrac{1}{2}\|\eta\|_{\mathrm{VT}}\mathrm{osc}(Pf) \\
&= \|\eta\|_{\mathrm{VT}} \sup_{\mathrm{osc}(f)\leq 1} \sup_{x,x'\in\mathbb{X}} |Pf(x) - Pf(x')| \\
&= \|\eta\|_{\mathrm{VT}} \sup_{x,x'\in\mathbb{X}} \sup_{\mathrm{osc}(f)\leq 1} |(P_x - P_{x'})f| \\
&= \|\eta\|_{\mathrm{VT}} \sup_{x,x'\in\mathbb{X}} d_{\mathrm{VT}}(P_x, P_{x'})\,,
\end{aligned}$$

d'où l'autre inégalité. □

Proposition 23 (Condition de Doeblin) **:**

On dit que P vérifie la condition de Doeblin s'il existe γ une mesure de probabilité et $\varepsilon > 0$ tels que $\forall x \in \mathbb{X}$, $P_x \geq \varepsilon\gamma$.

Dans cas, $P_x = (1-\varepsilon)\frac{P_x - \varepsilon\gamma}{1-\varepsilon} + \varepsilon\gamma = (1-\varepsilon)R_x + \varepsilon\gamma$. On a alors $P_x - P_{x'} = (1-\varepsilon)(R_x - R_{x'})$, donc $d_{\mathrm{VT}}(P_x, P_{x'}) \leq 1 - \varepsilon$.

En conséquence, $\Delta(P) \leq 1 - \varepsilon < 1$.

1.3 Construction canonique

On veut considérer une chaîne sur l'espace mesurable $(\mathbb{X}, \mathcal{X})$. On va se placer sur l'espace $\Omega = \mathbb{X}^{\mathbb{N}}$ muni de la tribu cylindrique $\mathcal{F} = \mathcal{X}^{\otimes \mathbb{N}}$. Les variables aléatoires X_k sont ici les projections canoniques, et les $\omega \in \mathbb{X}^{\mathbb{N}}$ sont des trajectoires. La filtration canonique est donnée par $\mathcal{F}_k = \sigma(X_i,\, i \leq k)$.

Théorème 24 (Iosifescu-Tulcea) **:**
Pour toute loi initiale μ sur $\mathbb{X}$ et tout noyau P sur $\mathbb{X}$, il existe une unique mesure de probabilités $\mathbb{P}_\mu$ sur $(\Omega, \mathcal{F})$ telle que (X_k) est une P-chaîne de Markov, de loi initiale μ.

Proposition 25 :
Pour tout évènement $A \in \mathcal{F}$, l'application $x \mapsto \mathbb{P}_x(A)$ est mesurable. Plus généralement, $\mathbb{P}_\mu(A) = \int\limits_{\mathbb{X}} \mathbb{P}_x(A)\,\mathrm{d}\mu(x)$.

Définition 26 (Opérateur de translation) **:**
L'opérateur de translation est l'application $\theta : (\omega_0, \omega_1, \omega_2, \ldots) \mapsto (\omega_1, \omega_2, \omega_3, \ldots)$. C'est une application mesurable sur $(\Omega, \mathcal{F})$ et $X_k \circ \theta_j = X_{k+j}$, avec θ_j l'opérateur de translation itéré.

Définition 27 (Temps d'arrêt) **:**
La variable $\tau : \Omega \to \overline{\mathbb{N}}$ est un temps d'arrêt lorsque, pour tout entier $n \in \mathbb{N}$, l'évènement $\{\tau = n\}$ est mesurable dans $\mathcal{F}_n$.

On définit $\mathcal{F}_\tau = \{A \in \mathcal{F},\, \forall n \in \mathbb{N},\, A \cap \{\tau = n\} \in \mathcal{F}_n\}$ la tribu des évènements antérieurs.

Proposition 28 (Composition de temps d'arrêt) **:**
Soient τ, σ deux temps d'arrêt. On définit $\rho = \sigma + \tau \circ \theta_\sigma$ lorsque $\sigma < \infty$ et $\rho = \infty$ sinon.

Alors ρ est un temps d'arrêt, et lorsque $\rho < \infty$, on a $X_\rho = X_\tau \circ \theta_\sigma$.

Démonstration. On a $\{\rho = n\} = \bigcup\limits_{j=0}^{n} \{\sigma = j\} \cap \{\tau \circ \theta_j = n - j\} = \bigcup\limits_{j=0}^{n} \{\sigma = j\} \cap \theta_j^{-1}\{\tau = n - j\}$. Or $\{\sigma = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$ et $\{\tau = n - j\} \in \mathcal{F}_{n-j}$ donc $\theta_j^{-1}\{\tau = n - j\} \in \mathcal{F}_n$. Donc $\{\rho = n\} \in \mathcal{F}_n$.

D'autre part, lorsque $\rho < \infty$, on a $[X_\rho](\omega) = X_{\rho(\omega)}(\omega) = X_{\tau \circ \theta_\sigma(\omega)}(\theta_{\sigma(\omega)}(\omega)) = X_\tau \circ \theta_\sigma$. $\square$

Proposition 29 (Propriété de Markov faible) **:**
Pour tout rang $k \in \mathbb{N}$ et toute variable Y positive ou bornée, on a $\mathbb{E}_{X_0}[Y \circ \theta_k | \mathcal{F}_k] = \mathbb{E}_{X_k}[Y]$ presque-sûrement.

Démonstration. Soient $A \in \mathcal{X}^{\otimes(k+1)}$ et $B \in \mathcal{X}^{\otimes(l+1)}$. En passant à une intégrale multiple pour la loi jointe, on obtient :

$$
\begin{aligned}
E[\mathbb{1}_A(X_0,\dots,X_k)\times\mathbb{1}_B(X_0,\dots,X_l)\circ\theta_k] &= \mathbb{E}[\mathbb{1}_A(X_0,\dots,X_k)\times\mathbb{1}_B(X_k,X_{k+1},\dots,X_{k+l})] \\
&= \dots \\
&= \mathbb{E}[\mathbb{1}_A(X_0,\dots,X_k)\mathbb{E}_{X_k}[\mathbb{1}_B(X_0,\dots,X_l)]]
\end{aligned}
$$

d'où le résultat désiré par densité de l'espace engendré par ces fonctions indicatrices. □

Proposition 30 (Propriété de Markov forte) **:**

Soit τ un temps d'arrêt. Pour tout variable Y positive ou bornée, on a :

$$\mathbb{E}[Y\circ\theta_\tau\times\mathbb{1}_{\tau<\infty}|\mathcal{F}_\tau]=\mathbb{E}_{X_\tau}[Y]\times\mathbb{1}_{\tau<\infty}\,.$$

Démonstration. Il suffit de décomposer $\mathbb{1}_{\tau<\infty}$ en $\sum\limits_{k\in\mathbb{N}}\mathbb{1}_{\tau=k}$, puis d'utiliser Markov faible. □

Définition 31 (Temps d'atteinte) **:**

Soit $C\in\mathcal{X}$ une partie mesurable. On pose $\tau_C=\min\{n\geq 0,\ X_n\in C\}$ le temps d'atteinte.

On peut aussi considérer $\sigma_C=\tau_C\circ\theta$ le temps de retour en C.

Sous l'évènement $\{X_0\notin C\}$ on a $\sigma_C=\tau_C$.

Proposition 32 (Partie récurrente) **:**

Si pour tout $x\in C$ on a $\mathbb{P}_x(\sigma_C<\infty)=1$ alors pour toute itérée $\sigma_C^{(n)}=\sigma_C+\sigma_C^{(n-1)}\circ\theta_{\sigma_C}$ on a $\mathbb{P}_x\left(\sigma_C^{(n)}<\infty\right)=1$.

Démonstration. Par Markov fort, on a :

$$
\begin{aligned}
P_x\left(\sigma_C^{(n)}<\infty\right) &= \mathbb{E}_x\left[\mathbb{1}_{\sigma_C<\infty}\times\mathbb{1}_{\sigma_C^{(n-1)}\circ\theta_{\sigma_C}<\infty}\right] \\
&= \mathbb{E}_x\left[\mathbb{1}_{\sigma_C<\infty}\times\mathbb{E}_{X_{\sigma_C}}\left[\mathbb{1}_{\sigma_C^{(n-1)}<\infty}\right]\right] \\
&= 1\,.
\end{aligned}
$$

□

Remarque 33 :

Lorsque $\sup\limits_C\mathbb{P}_x(\sigma_C<\infty)\leq\rho<1$, on peut de même établir $\mathbb{P}_x\left(\sigma_C^{(n)}<\infty\right)\leq\rho^n$.

Remarque 34 :

Si pour tout $x\notin C$ on a $\mathbb{P}_x(\sigma_C<\infty)=1$, alors c'est aussi le cas pour tout $x\in C$.

1.4 Fonctions harmoniques et martingales

Définition 35 (Fonctions harmoniques) **:**
Soit f positive ou bornée. On dit que f est harmonique (resp. sur, sous) sur $A \in \mathbb{X}$ lorsque, pour tout $x \in A$, on a $Pf(x) = f(x)$ (resp. $\leq$, $\geq$).

Proposition 36 :
La fonction f est surharmonique sur $A^C = \mathbb{X} \backslash A$ ssi pour toute loi initiale, le processus $(f(X_{n \wedge \tau_A}))$ est une sur-martingale.

Démonstration. Si f est surharmonique, alors :

$$\begin{aligned} \mathbb{E}\big[f\big(X_{(n+1)\wedge\tau_A}\big)|\mathcal{F}_n\big] &= \mathbb{E}\big[f\big(X_{(n+1)\wedge\tau_A}\big)\mathbb{1}_{X_n \in A}|X_n\big] \\ &= \mathbb{1}_{\tau_A > n} Pf(X_n) + \mathbb{1}_{\tau_A \leq n} f(X_{\tau_A}) \\ &\leq f(X_{\tau_A \wedge n}) \,. \end{aligned}$$

Réciproquement, si on a cette propriété pour tout point de départ $x \in A^C$, comme $\tau > 0$, on a $Pf(x) = \mathbb{E}_x[f(X_1)|X_0] \leq f(X_0) = f(x)$. $\square$

Remarque 37 :
Ainsi, la fonction $h : x \mapsto \mathbb{P}_x(\tau_A < \infty)$ est harmonique sur A^C, et sur-harmonique sur A car $Ph(x) \leq 1 = h(x)$ dans ce cas.

Définition 38 (Noyau potentiel) **:**
Soit $N_A = \sum\limits_{n \in \mathbb{N}} \mathbb{1}_{X_n \in A}$ le nombre de visites en A. Le noyau potentiel de A en partant de x est $U(x, A) = \mathbb{E}_x[N_A] \in \overline{\mathbb{R}^+}$.

Proposition 39 :
L'application $x \mapsto \mathbb{P}_x(N_A = \infty)$ est harmonique.

Définition 40 (Parties connectées) **:**
Soient $A, B \in \mathcal{X}$ tels que $\delta := \inf\limits_{x \in A} \mathbb{P}_x(\sigma_B < \infty) > 0$.

On dit que A conduit à B, et on note $A \to B$.

Proposition 41 :
Si $A \to B$, alors pour toute loi initiale, on a l'inclusion d'évènements $\{N_A = \infty\} \subset \{N_B = \infty\}$ (à une partie négligeable près).

Démonstration. Comme chaque passage en A conduit à B avec probabilité au moins δ, on a $\{N_A = \infty\} \subset \big\{\omega,\, \mathbb{P}_{X_n(\omega)}(\sigma_B < \infty) \geq \delta \text{ infiniment souvent}\big\}$.

Par un argument de martingale, on va montrer $\mathbb{P}_{X_n}(\sigma_B < \infty) \xrightarrow{\text{p.s.}} \mathbb{1}_{N_B=\infty}$. En supposant ce résultat, lorsque $N_A(\omega) = \infty$, alors par inclusion on a $\mathbb{1}_{N_B=\infty}(\omega) = \lim \mathbb{P}_{X_n}(\sigma_B < \infty)(\omega) \geq \delta > 0$ d'où $\mathbb{1}_{N_B=\infty}(\omega) = 1$.

Montrons maintenant cette propriété de limite. L'application $h : x \mapsto \mathbb{P}_x(\sigma_B < \infty)$ est positive, à valeurs dans $[0, 1]$. Cette application et sur-harmonique sur $\mathbb{X}$ tout entier, donc $M_n = h(X_n)$ est une sur-martingale positive, convergente vers M_∞ presque-sûrement et dans tout L^p fini.

On aimerait montrer désormais que $M_\infty = \mathbb{1}_{N_B=\infty}$. Soit $F \in \mathcal{F}_p$. Par convergence L^1 on peut intervertir le passage à la limite et l'intégrale, donc :

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_F \times M_\infty] &= \mathbb{E}[\mathbb{1}_F \times \lim h(X_n)] \\
&= \lim \mathbb{E}[\mathbb{1}_F \times h(X_n)] \\
&= \lim \mathbb{E}[\mathbb{1}_F \times \mathbb{1}_{\sigma_B<\infty} \circ \theta_n] \\
&= \mathbb{E}[\mathbb{1}_F \times \lim \mathbb{1}_{\sigma_B<\infty} \circ \theta_n] \,,
\end{aligned}$$

car $\{\sigma_B \circ \theta_n < \infty\} = \bigcup_{k \geq n} \{X_k \in B\}$ est une famille décroissante d'évènements en n. La limite de cette famille est l'évènement pour lequel on a des visites en B à des moments arbitrairement grands, autrement dit quand $N_B = \infty$.

Au final, $\mathbb{E}[\mathbb{1}_F \lim h(X_n)] = \mathbb{E}[\mathbb{1}_F \mathbb{1}_{N_B=\infty}]$ pour des évènements F qui engendrent la tribu cylindrique $\mathcal{F}$, d'où $M_\infty = \mathbb{1}_{N_B=\infty}$ presque-sûrement, le résultat souhaité. □

1.5 Théorème de comparaison

Théorème 42 :

Soient V, Y et Z des processus sur un espace filtré, à valeurs dans $\overline{\mathbb{R}^+}$. Si pour tout $n \in \mathbb{N}$ on a $\mathbb{E}[V_{n+1}|\mathcal{F}_n] + Z_n \leq V_n + Y_n$ alors pour tout temps d'arrêt T on a :

$$\mathbb{E}\left[V_T \mathbb{1}_{T<\infty} + \sum_{k=0}^{T-1} Z_k\right] \leq \mathbb{E}\left[V_0 + \sum_{k=0}^{T-1} Y_k\right].$$

Démonstration. Montrons d'abord le résultat un rang $T = n$ fixé. Pour $n = 1$, il suffit de passer à l'espérance dans l'inégalité initiale. On étend le résultat à tout rang $n \in \mathbb{N}$ par induction.

On a $\{T > n\} \in \mathcal{F}_n$. Remarquons que $\mathbb{E}\big[V_{(n+1)\wedge T}|\mathcal{F}_n\big] = \mathbb{1}_{T>n}\mathbb{E}[V_{n+1}|\mathcal{F}_n] + \mathbb{1}_{T\leq n}V_T$. On en déduit l'inégalité :

$$\mathbb{E}\big[V_{(n+1)\wedge T}|\mathcal{F}_n\big] + \mathbb{1}_{T>n}Z_n \leq V_{n\wedge T} + \mathbb{1}_{T>n}Y_n\,.$$

En conséquence, en utilisant le résultat au rang n fini pour ces processus, on obtient :

$$\begin{aligned}
\mathbb{E}\left[V_{n\wedge T}\mathbb{1}_{T<\infty} + \sum_{k=0}^{n\wedge T-1} Z_k\right] &= \mathbb{E}\left[V_{n\wedge T} + \sum_{k=0}^{n-1} \mathbb{1}_{T>k}Z_k\right] \\
&\leq \mathbb{E}\left[V_0 + \sum_{k=0}^{n-1} \mathbb{1}_{T>k}Y_k\right] \\
&\leq \mathbb{E}\left[V_0 + \sum_{k=0}^{n\wedge T-1} Y_k\right] \\
&\leq \mathbb{E}\left[V_0 + \sum_{k=0}^{T-1} Y_k\right].
\end{aligned}$$

Comme $V_{n\wedge T}\mathbb{1}_{T<\infty}$ est presque-sûrement constante à partir d'un rang, et que $\sum_{k=0}^{n\wedge T-1} Z_k$ converge vers la somme jusqu'au rang $T-1$ (éventuellement infini) de façon croissante, dans $\overline{\mathbb{R}^+}$, on a bien $\lim_{n\to\infty} V_{n\wedge T}\mathbb{1}_{T<\infty} + \sum_{k=0}^{n\wedge T-1} Z_k = V_T + \sum_{k=0}^{T-1} Z_k$. En appliquant le lemme de Fatou, on en déduit le résultat voulu. □

Proposition 43 :

Soient P un noyau sur $\mathbb{X}$, et $V, f : \mathbb{X} \to \overline{\mathbb{R}^+}$ mesurables.

Supposons que pour $x \notin C \in \mathcal{X}$, on a $PV(x) + f(x) \leq V(x)$. Alors pour tout $x \in \mathbb{X}$, on a :

$$\mathbb{E}_x\left[V(X_{\sigma_C})\mathbb{1}_{\sigma_C<\infty} + \sum_{k=0}^{\sigma_C-1} f(X_k)\right] \leq \mathbb{1}_C(x)(PV(x) + f(x)) + \mathbb{1}_{C^c}(x)V(x)\,.$$

Démonstration. Définissons les variables suivantes :

— $V_0 = V(X_0)\mathbb{1}_{C^c}(X_0)$,
— $V_k = V(X_k)$ pour $k > 0$,
— $Z_k = f(X_k)$,
— $Y_0 = (PV(X_0) + f(X_0)) \times \mathbb{1}_C(X_0)$,
— $Y_k = d \times \mathbb{1}_C(X_k)$,

avec $d = \|(PV + f)\mathbb{1}_C\|_\infty \in \overline{\mathbb{R}^+}$.

En admettant qu'on peut appliquer le théorème précédent, avec $T = \sigma_C$, on a :

$$\mathbb{E}_x\left[V_{\sigma_C}\mathbb{1}_{\sigma_C<\infty} + \sum_{k=0}^{\sigma_C-1} f(X_k)\right] \leq V(x)\mathbb{1}_{C^c}(x) + (PV(x) + f(x))\mathbb{1}_C(x)\,,$$

c'est-à-dire le résultat souhaité.

Il reste donc à vérifier les hypothèses du théorème, à savoir $\mathbb{E}[V_{k+1}|\mathcal{F}_k] + Z_k \leq V_k + Y_k$.

Compte tenu de l'asymétrie dans les définitions, faisons le cas $k = 0$ à part. Au rang $k = 0$, on a directement $\mathbb{E}_x[V_1|\mathcal{F}_0] + Z_0 = PV(x) + f(x) \leq \mathbb{1}_{C^c}(x)V(x) + \mathbb{1}_C(x)(PV(x) + f(x)) = V_0 + Y_0$.

Pour les rangs suivants, il faut un peu plus majorer :

$$\begin{aligned}
E_x[V_{k+1}|\mathcal{F}_k] + Z_k &= PV(X_k) + f(X_k)\\
&\leq \mathbb{1}_{C^c}(X_k)V(X_k) + \mathbb{1}_C(X_k)(PV(X_k) + f(X_k))\\
&\leq V(X_k) + \mathbb{1}_C(X_k) \times d\\
&= V_k + Y_k\,.
\end{aligned}$$

□

Corollaire 44 :
Si $d < \infty$, alors $\sup\limits_{x\in C} \mathbb{E}_x\left[\sum\limits_{k=0}^{\sigma_C-1} f(X_k)\right] \leq d$.

Corollaire 45 :
Si $d < \infty$ et P a une mesure invariante π telle que $\pi(\{V = \infty\}) = 0$ alors on a $\pi(f) \leq d$.

Démonstration. Par invariance de la mesure π, on a :

$$\begin{aligned}
\pi(f_m) &= \mathbb{E}_\pi[f(X_0)]\\
&= \tfrac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_\pi[f_m(X_k)]\\
&= \tfrac{1}{n}\sum_{k=0}^{n-1} \pi\left(P^k f_m\right)\\
&\leq \pi\left(\tfrac{1}{n}\sum_{k=0}^{n-1}\left(P^k f\right) \wedge m\right)\\
&\leq \pi\left(\left(\tfrac{1}{n}\sum_{k=0}^{n-1}\left(P^k f\right)\right) \wedge m\right).
\end{aligned}$$

avec les dernières étapes découlant d'inégalités de Jensen, notamment $P(f \wedge m) \leq (Pf) \wedge m$.

On a $PV + f \leq V + d$, donc plus généralement $P^nV + P^{n-1}f \leq P^{n-1}V + d$ pour $n > 0$. En passant à la somme :

$$\sum_{k=1}^{n} P^kV(x) + \sum_{k=0}^{n-1} P^kf(x) \leq \sum_{k=0}^{n-1} P^kV(x) + nd$$

Il en découle $\frac{1}{n}\sum_{k=0}^{n-1} P^kf \leq d + \frac{V}{n}$. On a donc $\pi(f_m) \leq \pi\big((d+\frac{V}{n}) \wedge m\big)$. Comme V est presque sûrement fini sous la loi π, on en déduit $(d+\frac{V}{n}) \wedge m \xrightarrow[n\to\infty]{\pi\text{-p.s.}} d \wedge m$, Les applications sont bornées par m, donc par convergence dominée, $\pi(f_m) \leq \pi(d \wedge m) = d \wedge m$.

Finalement, par convergence monotone, $\pi(f) \leq d$ pour $m \to \infty$. □

Théorème 46 (Condition de drift de Foster) **:**
Si $b := \sup_{x\in C} \mathbb{E}_x[\beta^{\sigma_C}] < \infty$ pour un certain $\beta > 1$, alors $V(x) = \mathbb{E}_x[\beta^{\tau_C}]$ vérifie la condition de Foster-Lyapunov : $\beta PV \leq V + b\mathbb{1}_C$.

Réciproquement, si il existe $V : \mathbb{X} \to [1,\infty]$, $\lambda \in [0,1[$ et $b < \infty$ tels que $PV \leq \lambda V + b\mathbb{1}_C$, alors pour tout $x \in \mathbb{X}$ on a $\mathbb{E}_x[\lambda^{-\sigma_C}] \leq V(x) + \frac{b}{\lambda}$.

Démonstration. Commençons par l'implication directe. On a $V(x) = \mathbb{E}_x[\beta^{\tau_C}]$. Alors :

$$PV(x) = \mathbb{E}_x[V(X_1)] = \frac{1}{\beta}\mathbb{E}_x[\beta^{\sigma_C}]$$

d'où le résultat, car $\sigma_C = \tau_C$ lorsque $x \notin C$ et $\mathbb{E}_x[\beta^{\sigma_C}] \leq b$ lorsque $x \in C$.

Pour le sens réciproque, si $V(x) = \infty$ alors on a de toute façon l'inégalité. Supposons donc $V(x) < \infty$, et posons $f = (1-\lambda)V$. On a $PV + f \leq V + b\mathbb{1}_C$, donc le théorème précédent s'applique :

$$(1-\lambda)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} V(X_k)\right] \leq (PV(x) + (1-\lambda)V(x))\mathbb{1}_C(x) + V(x)\mathbb{1}_{C^c}(x) \leq V(x) + b\mathbb{1}_C(x)$$

or $V \geq 1$ donc $\mathbb{E}_x[\sigma_C] \leq \frac{V(x)+b\mathbb{1}_C(x)}{1-\lambda} < \infty$ et en particulier, σ_C est fini presque-sûrement. Soient désormais $V_n = \frac{V(X_n)}{\lambda^n}$. On a $\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq V_n + \frac{b}{\lambda^{n+1}}\mathbb{1}_C(X_n)$.

On définit alors $Z_n = 0$ et $Y_n = \frac{b}{\lambda^{n+1}}\mathbb{1}_C(X_n)$. Le théorème donne :

$$\mathbb{E}_x\left[\lambda^{-\sigma_C}\right] \leq \mathbb{E}_x[V_{\sigma_C}] \leq \mathbb{E}_x[V_0] + b\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \frac{1}{\lambda^{k+1}}\mathbb{1}_C(X_k)\right] = V(x) + \frac{b}{\lambda}\mathbb{1}_C(x)\,.$$

□

2 Chaînes atomiques

2.1 Atomes

Définition 47 (Atome) **:**

Un atome du noyau P est un ensemble α pour lequel il existe une mesure μ de sorte que pour tout $x \in \alpha$ on a $\mathbb{P}_x = \mu$. On note alors P_α cette mesure.

Remarque 48 :

Dans le cas des marches aléatoires discrètes, tous les états sont atomiques.

Définition 49 (Ensemble accessible) **:**

L'ensemble $A \in \mathcal{X}$ est accessible si, pour tout $x \in \mathbb{X}$, on a $\mathbb{P}_x(\sigma_A < \infty) > 0$. On note $\mathcal{X}^+$ l'ensemble des parties accessibles.

Remarque 50 (Exemple) **:**

Lorsque $X_{n+1} = (X_n + W_{n+1})^+$, pour des déplacements $W_i \sim \mathcal{N}(0,1)$, $\{0\}$ est un atome accessible.

Lemme 51 :

Soit α un atome accessible. Alors :

1. $A \in \mathcal{X}$ est accessible ssi $\mathbb{P}_\alpha(\sigma_A < \infty) > 0$,
2. si A n'est pas accessible, alors A^c l'est.

Démonstration. Pour le premier point, le sens direct est évident. Le sens réciproque découle d'une propriété de Markov forte, via :

$$\mathbb{P}_x(\sigma_A < \infty) \geq \mathbb{P}_x(\sigma_A \circ \sigma_\alpha < \infty) = \mathbb{P}_x(\sigma_\alpha < \infty)\mathbb{P}_\alpha(\sigma_A < \infty) > 0\,.$$

Pour le second point, si A n'est pas accessible, alors $P_\alpha(A) = 0$ donc $P_\alpha(A^c) = 1$. En conséquence, $\mathbb{P}_\alpha(\sigma_{A^c} < \infty) = 1$. D'après la première équivalence, A^c est donc accessible. □

2.2 Récurrence et transience

Définition 52 (Atome récurrent) **:**

Rappelons que $U(x, A) = \mathbb{E}_x[N_A]$ est le nombre moyen de visites en A en partant de x.

L'atome α est récurrent si $U(\alpha, \alpha) = \infty$. L'atome α est transient sinon.

Théorème 53 :

Les propriétés suivantes sur l'atome α sont équivalentes :

1. $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$,
2. $\mathbb{P}_\alpha(N_\alpha = \infty) = 1$,
3. α est récurrent.

Démonstration. L'implication $(2 \Rightarrow 3)$ est directe : si N_α est infini p.s., il n'est pas intégrable.

L'équivalence $(1 \Leftrightarrow 2)$ découle de $\mathbb{P}_\alpha\left(\sigma_\alpha^{(n)} < \infty\right) = \mathbb{P}_\alpha(\sigma_\alpha < \infty)^n \underset{n\to\infty}{\longrightarrow} \mathbb{P}_\alpha(N_\alpha = \infty)$.

L'équivalence $(1 \Leftrightarrow 3)$ découle de $\mathbb{E}[N_\alpha] = \sum_{n\geq 1} \mathbb{P}_\alpha(N_\alpha \geq n) = \sum_{n\geq 1} \mathbb{P}_\alpha(\sigma_\alpha < \infty)^n$. □

Définition 54 (Ensemble absorbant) **:**

Un ensemble $A \in \mathcal{X}$ est absorbant si pour tout $x \in A$, on a $P_x(A) = 1$.

Lemme 55 :

Soit α un atome accessible. Alors le domaine d'attraction de l'atome α, défini par l'ensemble $\alpha_\infty = \{x \in \mathbb{X},\, \mathbb{P}_x(N_\alpha = \infty) = 1\}$, est absorbant.

Démonstration. Soit $h(x) = \mathbb{P}_x(N_\alpha = \infty)$ harmonique sur $\mathbb{X}$. Lorsque $x \in \alpha_\infty$, on a :

$$h(x) = 1 = P_x(\alpha_\infty) + \mathbb{E}_x\left[\mathbb{1}_{\alpha_\infty^c}(X_1)\mathbb{P}_{X^1}(N_\alpha = \infty)\right],$$

donc $P_x(\alpha_\infty^c) = 0$, sans quoi $1 < 1$. □

Proposition 56 :

Soit α un atome accessible récurrent. Alors tout atome β tel que $\mathbb{P}_\alpha(\sigma_\beta < \infty) > 0$, l'atome β est en fait accessible et récurrent, et on a $\mathbb{P}_\alpha(N_\beta = \infty) = \mathbb{P}_\beta(N_\alpha = \infty) = 1$.

Démonstration. Comme $\mathbb{P}_\alpha(\sigma_\beta < \infty) > 0$, d'après une des lemmes précédentes, l'atome β est accessible. Comme $\alpha \to \beta$, on a l'inclusion d'évènements $\{N_\alpha = \infty\} \subset \{N_\beta = \infty\}$ presque-sûre. Par récurrence de α, $\mathbb{P}_\alpha(N_\beta = \infty) \geq \mathbb{P}_\alpha(N_\alpha = \infty) = 1$. En outre, par propriété de Markov forte, $\mathbb{P}_\beta(N_\beta = \infty) \geq \mathbb{P}_\beta(\sigma_\alpha < \infty)\mathbb{P}_\alpha(N_\beta = \infty) = 1$, donc β est récurrent.

Comme β est accessible et récurrent, et α accessible, quitte à intervertir α et β dans la preuve précédente, on obtient $\mathbb{P}_\beta(N_\alpha = \infty) = 1$. □

Définition 57 (Ensemble et noyau récurrent) **:**

Un ensemble A est récurrent si $U(x, A) = \infty$ pour tout $x \in A$.

Un noyau P est récurrent si tout ensemble accessible est récurrent.

Définition 58 (Ensemble et noyau transient) **:**

Un ensemble A est uniformément transient si $\sup_{x \in A} U(x, A) < \infty$.

Un ensemble A est transient si il est union d'ensembles uniformément transients.

Un noyau P est transient si $\mathbb{X}$ est transient.

Théorème 59 :

Soit α un atome accessible. Alors P est récurrent ssi α est récurrent.

Inversement, P est transient ssi α est transient.

Démonstration. Si P est récurrent, en particulier α est récurrent car il est accessible.

Supposons donc α récurrent. Soit maintenant $A \in \mathcal{X}$ accessible quelconque. En particulier $\mathbb{P}_\alpha(\sigma_A < \infty) > 0$. On a l'inclusion $\{N_\alpha = \infty\} \subset \{N_A = \infty\}$ p.s. pour toute loi initiale. En particulier, on a $\mathbb{P}_\alpha(N_A = \infty) = 1$. Pour tout $x \in A$ on a donc :

$$U(x, A) = \mathbb{E}_x[N_A] \geq \mathbb{P}_x(\sigma_\alpha < \infty)\mathbb{E}_\alpha[N_A] = \infty\,.$$

L'autre équivalence se montre par des raisonnements analogues. $\square$

2.3 Période d'un atome

Définition 60 (Période d'un atome) **:**
Soit α un atome de P. La période de α est le pgcd de $E_\alpha = \{n \in \mathbb{N},\ P^n_\alpha(\alpha) > 0\}$, noté $d(\alpha)$.

Lemme 61 :
Si $m, n \in E_\alpha$, alors $m + n \in E_\alpha$.

Démonstration. On a :

$$P^{n+m}_\alpha(\alpha) = \int \mathbb{P}_x(X_m \in \alpha)\,\mathrm{d}P^n_\alpha(x) \geq \int_\alpha \mathbb{P}_\alpha(X_m \in \alpha)\,\mathrm{d}P^n_\alpha(x) = P^m_\alpha(\alpha)P^n_\alpha(\alpha) > 0\,.$$

□

Lemme 62 :
A partir d'un rang n_0, pour tout $n \geq n_0$, on a $nd(\alpha) \in E_\alpha$.

Démonstration. Comme d est un pgcd, par relation de Bezout, on a des entiers $n_1, \ldots, n_s \in E_\alpha$, et $n_1, \ldots, n_s \in \mathbb{Z}$ tels que $\sum\limits_{i=1}^{s} a_i n_i = d$.

Quitte à séparer les parties positives et négatives, on a $q = \sum\limits_{i=1}^{s} a_i^+ n_i \in E_\alpha$ et $p = \sum\limits_{i=1}^{s} a_i^- n_i \in E_\alpha$ par le lemme précédent, et $d = q - p$. Comme $p \in E_\alpha$, d divise p donc $p = kd$, et naturellement $q = (k+1)d$. On a $nd = n(q-p) = n(q-kd)$.

Quitte à faire une division euclidienne par k, on a $n = lk + r$, donc $nd = lkd + rq - rp = (l-r)p + rq$. Autrement dit, dès que $l > r$, on a $nd \in E_\alpha$. Comme r est un reste modulo k, dès que $l \geq k$, dès que $n \geq k^2$, on a le résultat souhaité. □

Lemme 63 :
Soient α et β deux atomes accessibles. Alors $d(\alpha) = d(\beta)$.

Démonstration. Soit $n \in E_\alpha$. Montrons que $d(\beta)$ divise n. Comme α et β sont accessibles, on a u et v tels que $P^u_\alpha(\beta) > 0$ et $P^v_\beta(\alpha) > 0$. On a alors $P^{u+v}_\beta(\beta) \geq P^u_\beta(\alpha)P^v_\beta(\alpha) > 0$ et $P^{u+n+v}_\beta(\beta) \geq P^u_\beta(\alpha)P^n_\alpha(\alpha)P^v_\beta(\alpha) > 0$, donc $d(\beta)$ divise $(u+v+n) - (u+v) = n$.

En passant au pgcd, $d(\beta)$ divise donc $d(\alpha)$. Par symétrie du raisonnement, on a l'égalité. □

Définition 64 (Période d'un noyau) **:**

Si P admet un atome accessible, alors $d(P)$ est la période de ses atomes accessibles.

Si $d(P) = 1$, on dit que P est apériodique.

Proposition 65 :

Le noyau P est apériodique ssi pour tout atome accessible, $P^n_\beta(\beta) > 0$ à partir d'un rang.

Si P est apériodique, alors pour tous atomes accessibles α et β, à partir d'un rang, $P^n_\alpha(\beta) > 0$.

2.4 Mesures invariantes

Définition 66 :

Soit α un atome de P. L'atome est :

— positif-récurrent si $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$,

— nul-récurrent si $\mathbb{E}_\alpha[\sigma_\alpha] = \infty$ mais $U(\alpha, \alpha) = \mathbb{E}[N_\alpha] = \infty$.

Proposition 67 (Formule de Kac) **:**

Soient π une mesure P-invariante et $C \in \mathcal{X}$ tel que $\mathbb{P}_x(\sigma_C < \infty) \overset{\pi\text{-p.s.}}{>} 0$.

Pour toute fonction positive f, on pose d'une part $\pi^0_C(f) = \int\limits_C \mathbb{E}_x\left[\sum\limits_{k=0}^{\sigma_C - 1} f(X_k)\right] \mathrm{d}\pi(x)$ et d'autre part $\pi^1_C(f) = \int\limits_C \mathbb{E}_x\left[\sum\limits_{k=1}^{\sigma_C} f(X_k)\right] \mathrm{d}\pi(x)$. Alors $\pi = \pi^0_C = \pi^1_C$.

Démonstration. En premier lieu, il est clair que π^0_C et π^1_C sont des mesures de probabilité sur $\mathbb{X}$, et que $P\pi^0_C = \pi^1_C$ par propriété de Markov faible.

Remarquons alors que :

$$\begin{aligned}
\pi(f) &= P^n\pi(f) \\
&= \mathbb{E}_\pi[f(X_n)] \\
&= \mathbb{E}_\pi\left[f(X_n)\left(\mathbb{1}_{\sigma_C > n} + \sum_{k=0}^{n-1} \mathbb{1}_{X_{n-k} \in C,\, X_{n-k+1}, \dots, X_n \notin C}\right)\right] \\
&= \mathbb{E}_\pi[f(X_n)\mathbb{1}_{\sigma_C > n}] + \sum_{k=0}^{n-1} \mathbb{E}_\pi\left[\mathbb{E}_{X_{n-k}}[\mathbb{1}_{X_0 \in C,\, X_1, \dots, X_k \notin C} f(X_k)]\right] \\
&= \mathbb{E}_\pi[f(X_n)\mathbb{1}_{\sigma_C > n}] + \sum_{k=0}^{n-1} \mathbb{E}_\pi[\mathbb{E}_\pi[\mathbb{1}_{X_0 \in C}\mathbb{1}_{\sigma_C > k} f(X_k)]] \\
&= \mathbb{E}_\pi[f(X_n)\mathbb{1}_{\sigma_C > n}] + \mathbb{E}_\pi\left[\mathbb{1}_C(X_0) \sum_{k=0}^{\sigma_C \wedge n-1} f(X_k)\right].
\end{aligned}$$

Par convergence monotone, le terme de droite a pour limite $\mathbb{E}_\pi\left[\mathbb{1}_C(X_0) \sum\limits_{k=0}^{\sigma_C - 1} f(X_k)\right] = \pi^0_C(f)$ pour $n \to \infty$. On en déduit en particulier que $\pi^0_C(f) \le \pi(f)$. On peut donc définir une mesure $\Delta\pi = \pi - \pi^0_C$ positive, absolument continue par rapport à π.

Soit $f(x) = \mathbb{P}_x(\sigma_C < \infty)$. Par propriété de Markov forte :

$$\mathbb{E}_\pi[f(X_n)\mathbb{1}_{\sigma_C > n}] = \mathbb{P}_\pi(n < \sigma_C < \infty) \underset{n\to\infty}{\longrightarrow} 0\,.$$

Il en découle $\Delta\pi(f) = \pi(f) - \pi_C^0(f) = \lim \mathbb{P}_\pi(n < \sigma_C < \infty) = 0$. Or f est strictement positive π-p.s. par hypothèse, et $\Delta\pi \ll \pi$, donc nécessairement $\Delta\pi = 0$.

Au final, on a bien $\pi_C^0 = \pi = \pi P = \pi_C^0 P = \pi_C^1$. □

Proposition 68 :

Soit α un atome accessible de P. Alors α est positif ssi il existe une mesure P-invariante π.

De plus, dans ce cas, π est unique, et $\pi(A) = \frac{\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} \mathbb{1}_A(X_k)\right]}{\mathbb{E}_\alpha[\sigma_\alpha]}$.

Démonstration. Si α est positif, alors $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$, donc l'application π ci-dessus est bien définie, et c'est une mesure. En outre, par propriété de Markov :

$$\begin{aligned}
\pi P(A) &= \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} P_{X_k}(A)\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}_\alpha[P_{X_k}(A)\mathbb{1}_{\sigma_\alpha > k}] \\
&= \sum_{k=1}^{\infty} \mathbb{E}_\alpha[\mathbb{1}_A(X_{k+1})\mathbb{1}_{\sigma_\alpha > k}] \\
&= \mathbb{E}_\alpha\left[\sum_{k=2}^{\sigma_\alpha+1} \mathbb{1}_A(X_k)\right] \\
&= \mathbb{E}_\alpha\left[\sum_{k=2}^{\sigma_\alpha} \mathbb{1}_A(X_k)\right] + \mathbb{E}_\alpha\left[\mathbb{E}_{X_{\sigma_\alpha}}[\mathbb{1}_A(X_1)]\right] \\
&= \mathbb{E}_\alpha\left[\sum_{k=2}^{\sigma_\alpha} \mathbb{1}_A(X_k)\right] + \mathbb{E}_\alpha[\mathbb{1}_A(X_1)] \\
&= \pi(A)\,.
\end{aligned}$$

Pour le sens réciproque, on utilise la formule de Kac ci-dessus. Comme α est accessible, on a $\pi = \pi_\alpha^1$. Pour la fonction constante $f = 1$, on a :

$$1 = \pi(f) = \pi_\alpha^1(f) = \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right] = \pi(\alpha)\mathbb{E}_\alpha[\sigma_\alpha]$$

d'où $\mathbb{E}_\alpha[\sigma_\alpha] = \frac{1}{\pi(\alpha)} < \infty$, α est positif et on a l'unicité sous la forme souhaitée. □

Remarque 69 :

Si T est un temps d'arrêt, et θ l'opérateur de shift, on rappelle que $[\theta_T](\omega) = \theta_{T(\omega)}(\omega)$.

Proposition 70 :

Soient α un atome P-récurrent. On considère des variables $Z_0, \dots, Z_k$ qui sont $\mathcal{F}_{\sigma_\alpha}$-mesurables, telle que pour $0 \leq i \leq k$, $\mathbb{E}_x[Z_i]$ ne dépend pas de $x \in \alpha$.

Alors pour toute mesure initiale λ, si $\mathbb{P}_\lambda(\sigma_\alpha < \infty)$, on a $\mathbb{E}_\lambda\left[\prod_{i=0}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i)}}\right] = \mathbb{E}_\lambda[Z_0] \prod_{i=1}^{k} \mathbb{E}_\alpha[Z_i]$.

Démonstration. Si $k = 0$ le résultat est évident. Pour $k \geq 1$, en utilisant la propriété de Markov forte :

$$\begin{aligned}
\mathbb{E}_\lambda\left[\prod_{i=0}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i)}}\right] &= \mathbb{E}_\lambda\left[\mathbb{1}_{\sigma_\alpha<\infty} \prod_{i=0}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i)}}\right] \\
&= \mathbb{E}_\lambda\left[\mathbb{1}_{\sigma_\alpha<\infty} Z_0 \times \left(\prod_{i=0}^{k-1} Z_{i+1} \circ \theta_{\sigma_\alpha^{(i)}}\right) \circ \theta_{\sigma_\alpha}\right] \\
&= \mathbb{E}_\lambda\left[\mathbb{1}_{\sigma_\alpha<\infty} Z_0 \times \mathbb{E}_\alpha\left[\prod_{i=0}^{k-1} Z_{i+1} \circ \theta_{\sigma_\alpha^{(i)}}\right]\right] \\
&= \prod_{i=1}^{k} \mathbb{E}_\alpha[Z_i] \mathbb{E}_\lambda[\mathbb{1}_{\sigma_\alpha<\infty} Z_0]\,,
\end{aligned}$$

d'où le résultat voulu, car $\mathbb{P}_\lambda(\sigma_\alpha < \infty) = 1$. □

3 Espaces discrets

Définition 71 :

Si $\mathbb{X}$ est discret, on dit que :

1. P est irréductible si il existe $x \in \mathbb{X}$ tel que l'atome $\{x\}$ est accessible.
2. P est fortement irréductible si pour tout $x \in \mathbb{X}$, l'atome $\{x\}$ est accessible.

Remarque 72 (Notation) **:**

Ici, comme l'espace est discret, il suffit d'étudier P sur les singletons de $\mathbb{X}$, et on écrira $P(x, y) = P_x(\{y\})$.

D'autre part, on posera $\overline{P}((x, x'), (y, y')) = P(x, y)P(x', y')$, le noyau qui fait évoluer deux P-chaînes de Markov indépendamment.

Remarque 73 (Rappels sur la variation totale) **:**

Pour deux mesures de probabilité ξ et ξ' on a :

$$\|\xi - \xi'\|_{\mathrm{VT}} := \sup_{|f|\leq 1} |\xi(f) - \xi'(f)| = 2 \inf_{\mu \in C(\xi,\xi')} \int \mathbb{1}_{x \neq x'} \,\mathrm{d}\mu(x, x') = 2 \inf_{\substack{X \sim \xi \\ X' \sim \xi'}} \mathbb{P}(X \neq X')$$

où $C(\xi, \xi')$ est l'ensemble des couplages entre ξ et ξ', des mesures sur $\mathbb{X}^2$ dont la loi marginale gauche (resp. droite) est ξ (resp. ξ'). Lorsque $\xi = \varphi\mu$ et $\xi' = \varphi'\mu$ sont absolument continues pour une même mesure μ, alors $\|\xi - \xi'\|_{\mathrm{VT}} = \int |\varphi - \varphi'| \,\mathrm{d}\mu$.

En outre, $d_{\mathrm{VT}}(\xi,\xi') = \frac{1}{2}\|\xi - \xi'\|_{\mathrm{VT}}$.

Proposition 74 :

Soient ξ et ξ' deux lois sur $\mathbb{X}$, $n \in \mathbb{N}$ un entier et $a \in \mathbb{X}$ quelconques. Alors :

$$d_{\mathrm{VT}}(\xi P^n, \xi' P^n) \leq \mathbb{P}_{\xi\otimes\xi'}(T > n)$$

où $T = \inf\{n \geq 0,\, (X_n, X'_n) = (a,a)\}$ est le temps de premier croisement.

Démonstration. On utilise ici la définition via le sup sur les fonctions f pour lesquelles $|f| \leq 1$. À f fixée, on a :

$$\begin{aligned}
\mathbb{E}_{\xi\otimes\xi'}[f(X_n) - f(X'_n)] &= \mathbb{E}_{\xi\otimes\xi'}[(f(X_n) - f(X'_n))\mathbb{1}_{T>n}] \\
&+ \sum_{k=0}^{n} \mathbb{E}_{\xi\otimes\xi'}\left[\mathbb{1}_{T=k}\mathbb{E}_{(a,a)}\left[f(X_{n-k}) - f\left(X'_{n-k}\right)\right]\right] \\
&= \mathbb{E}_{\xi\otimes\xi'}[(f(X_n) - f(X'_n))\mathbb{1}_{T>n}]
\end{aligned}$$

or chaque terme de la somme sur k est nul. Comme $|f(x) - f(x')| \leq 2$, en passant au supremum sur les fonctions f, on en déduit que $\|\xi P^n - \xi' P^n\|_{\mathrm{VT}} \leq 2\mathbb{P}_{\xi\otimes\xi'}(T > n)$, et donc le résultat souhaité du d_{VT}. $\square$

Lemme 75 :

Si P est apériodique et fortement irréductible, alors $\overline{P}$ aussi.

Démonstration. On a $\overline{P}^n((x,x'),(y,y')) = P^n(x,y)P^n(x',y')$. Or à partir d'un certain rang, $P^n(x,y) > 0$ et $P^n(x',y') > 0$, d'où le résultat sur $\overline{P}$. $\square$

Proposition 76 :
Si P est fortement irréductible, apériodique, positif, alors $d_{\mathrm{VT}}(\xi P^n, \pi) \underset{n\to\infty}{\longrightarrow} 0$.

Démonstration. Ici, π est l'unique probabilité invariante du noyau positif P. On a $\overline{\pi} = \pi \otimes \pi$ la probabilité invariante de $\overline{P}$. En outre, par le lemme précédent, $\overline{P}$ est fortement irréductible.

On veut montrer que $\mathbb{P}_{\xi\otimes\pi}(T > n) \to 0$. Comme $\mathbb{P}_{\xi\otimes\pi}(T > n) = \int \mathbb{P}_{x,x'}(T > n) \,\mathrm{d}\xi(x) \,\mathrm{d}\pi(x')$, il suffit de montrer que $\mathbb{P}_{x,x'}(T > n) \to 0$ puis conclure par convergence dominée.

Soient $\overline{x} = (x, x')$ et $\overline{a} = (a, a) \in \mathbb{X}^2$. Il suffit de montrer que $\mathbb{P}_{\overline{x}}(\tau_{\overline{a}} < \infty) = 1$ pour obtenir la limite souhaitée. Comme $\mathbb{P}_{\overline{x}}(\sigma_{\overline{a}} < \infty) > 0$, par le lemme utile, on a la chaîne d'inclusions $\{N_{\overline{x}} = \infty\} \subset \{N_{\overline{a}} = \infty\} \subset \{\sigma_{\overline{a}} < \infty\}$ presque-sûrement sous $\mathbb{P}_{\overline{x}}$. Or l'état $\overline{x}$ est récurrent pour $\overline{P}$, d'où $\mathbb{P}_{\overline{x}}(N_{\overline{x}=\infty}) = 1$, ce qui conclut la preuve. □

4 Théorie de renouvellement

Définition 77 (Processus de renouvellement) **:**
Soient $(Y_i)_{i\geq 1} \overset{\text{iid}}{\sim} b$ des temps d'attente sur $\mathbb{N}^*$, indépendants de $Y_0 \sim a$ à valeurs dans $\mathbb{N}$.

Alors $S_n = \sum_{i=0}^{n} Y_i$ est un processus de renouvellement, une chaîne de Markov discrète pour le noyau $P(i,j) = b(j-i)\mathbb{1}_{j>i}$, et on note alors $\mathbb{P}_a$ la probabilité sur $\mathbb{N}^{\mathbb{N}}$ induite par le processus.

Définition 78 (Processus de renouvellement apériodique) **:**
On dit que b est apériodique si $\mathrm{pgcd}(\{n \in \mathbb{N}^*,\, b(n) > 0\}) = 1$.

Définition 79 (Processus de renouvellement sans délai) **:**
On dit que S est sans délai si $a = \delta_0$, si $S_0 = 0$.

Définition 80 (Probabilités de renouvellement) **:**
On pose $V_n = \sum_{j\in\mathbb{N}} \mathbb{1}_{S_j = n}$ la variable qui indique un renouvellement à l'instant n. On note alors $v_a(k) = \mathbb{P}_a(V_k = 1)$, et $u(k) = v_{\delta_0}(k)$. On a $u(k) = \delta_0(k) + \sum_{j=1}^{k} \mathbb{P}(S_m = k) = \sum_{j\in\mathbb{N}} b^{*j}(k)$ où la fonction $b^{*j}(k) = \sum_{l_1+\cdots+l_j=k} \left(\prod_{m=1}^{j} b(l_m)\right)$ est la convolution itérée de b. Avec le même argument, $v_a = a * u$.

Théorème 81 :

Pour toute mesure a sur $\mathbb{N}$, v_a est l'unique solution de $v = a + b * v$. En particulier, u est l'unique solution de $u = \delta_0 + b * u$.

Démonstration. On a bien u solution du second système d'après ce qui précède. En conséquence, $a * u = a * \delta_0 + a * b * u = a + b * (a * u)$, donc v_a est solution du premier système.

Pour l'unicité, soit v solution du système. Quitte à itérer, on a $v = a * \sum_{j=0}^{n} b^{*j} + b^{*(n+1)} * v$. Sur tout intervalle $[\![0, k]\!]$, dès que $n > k$, on a $v(k) = a * \sum_{j=0}^{n} b^{*j}(k) + b^{*(n+1)} * v(k) = a * \sum_{j\in\mathbb{N}} b^{*j}(k)$, d'où l'unicité de v_a. □

On pose désormais $U(z) = \sum_{n\in\mathbb{N}} u(n)z^n$ la série génératrice de u, de rayon de convergence 1 au moins. On peut vérifier que $U(z) \to \infty$ pour $z \to 1$ dans $\mathbb{R}$, donc le rayon de convergence est précisément égal à 1. On pose de même les séries génératrice A de a, B de b et V_a de v_a.

Corollaire 82 :

En passant aux produits de Cauchy, lorsque $|z| < 1$, on obtient $U(z) = 1 + B(z)U(z)$ et $V_a(z) = A(z) + B(z)V_a(z)$.

Proposition 83 :

Supposons que le temps moyen d'attente $m := \sum_{j\in\mathbb{N}^*} jb(j) < \infty$ est fini. Alors v_a est constante ssi $a(k) = a_S(k) := \frac{1}{m} \sum_{j=k+1}^{\infty} b(j)$, et alors on a $v_a = \frac{1}{m}$.

Démonstration. Supposons $v_a = c$ constante. Alors $V_a = \frac{c}{1-z}$, donc :

$$A = c\frac{1-B}{1-z} = c(1-B)\sum_{k\in\mathbb{N}} z^k .$$

En particulier, $a(l) = c\left(1 - \sum_{j=1}^{l} b(j)\right) = c \sum_{j=l+1}^{\infty} b(j)$. On a alors $\sum_{l\in\mathbb{N}} a(l) = c \sum_{j\in\mathbb{N}} jb(j) = cm = 1$, d'où $c = \frac{1}{m}$. Réciproquement, si $a = a_S$, alors en remontant les calculs on trouve $V_a = \frac{1}{m} \times \frac{1}{1-z}$ d'où $v_a = \frac{1}{m}$ constante. □

4.1 Temps de vie résiduel, *Forward Recurrence Time Chain*

On définit les variables $\rho_k = \inf\{n \in \mathbb{N}, S_n > k\}$, $A_k = S_{\rho_k} - k$ pour $k \geq 1$ et $A_0 = Y_0 + \mathbb{1}_{Y_0=0}Y_1$.

Dans ce cas, $\mathbb{P}(A_0 = k) = a(0)b(k) + a(k)$.

Proposition 84 :

A est une chaîne de Markov sur $\mathbb{X} = [\![1, \sup\{n \in \mathbb{N},\, b(n) \neq 0\}]\!]$ de noyau Q défini par $Q(1,j) = b(j)$ pour $j \in \mathbb{X}$ et $Q(j, j-1) = 1$ lorsque $j \geq 2$. De plus, le noyau Markovien Q est fortement irréductible, récurrent.

Si $m < \infty$, alors Q est positif, et la mesure $\mu_S(k) = \frac{1}{m} \sum\limits_{j=k}^{\infty} b(j)$ est Q-invariante.

Démonstration. On peut vérifier aisément que A est une Q-chaîne de Markov.

L'irréductibilité de Q vient du fait que $Q_1 = b$, donc on va dépasser tout point de $\mathbb{X}$ en partant de 1, puis revenir en arrière jusqu'à l'atteindre, avant de revenir finalement en 1. En conséquence, Q est fortement irréductible et $\{1\}$ est un atome récurrent.

En outre, comme le temps de retour en 1 est égal à la longueur du saut initial, de loi b, on a $\mathbb{E}_1[\sigma_1] = \sum\limits_{j \in \mathbb{N}^*} j b(j) = m$. Le noyau Q est donc positif ssi $m < \infty$.

Dans ce cas, le corollaire de la formule de Kac nous donne la probabilité invariante :

$$\pi(k) = \frac{1}{\mathbb{E}_1[\sigma_1]} \mathbb{E}_1\left[\sum_{j=1}^{\sigma_1} \mathbb{1}_k(A_j)\right] = \frac{1}{m} \mathbb{E}_1[\mathbb{1}_{A_1 \geq k}] = \frac{1}{m} \sum_{j=k}^{\infty} b(j)\,.$$

□

Théorème 85 (Blackwell) **:**

Si b est apériodique et $m < \infty$, alors pour toute loi initiale a sur $\mathbb{N}$, $v_a(n) \underset{n \to \infty}{\longrightarrow} \frac{1}{m}$.

Démonstration. On a $v_a(n) = \mathbb{P}_a(\exists m \in \mathbb{N},\, S_m = n) = \mathbb{P}_a(A_{n-1} = 1)$.

Or Q est fortement irréductible, positif-récurrent car $m < \infty$. Comme $b(k) \leq Q^k(1,1)$, la période de Q divise la période de b, donc que Q est apériodique.

alors $d_{\mathrm{VT}}(A_{n-1}, \mu_S) \to 0$, donc en particulier $\mathbb{P}_a(A_{n-1} = 1) \to \mu_S(1) = \frac{1}{m}$. □

Théorème 86 (Théorème de Kendall) **:**

Par la suite, on note $\mathbb{D} = \{z \in \mathbb{C},\, |z| \leq 1\}$ le disque unité fermé.

Supposons que b est apériodique. Dans ce cas, B est holomorphe sur un voisinage ouvert de $\mathbb{D}$ ssi il existe $\lambda > 0$ tel que $\sum\limits_{n \in \mathbb{N}} (u(n) - \lambda) z^n$ est holomorphe sur un voisinage de $\mathbb{D}$.

Dans ce cas, on a $m < \infty$ et $\lambda = \frac{1}{m}$.

Démonstration. Si les deux points sont vérifiés, alors par holomorphie de B sur un disque rayon $\beta > 1$, on a $b(n) = O\left(\frac{1}{\beta}^n\right)$ qui converge vers 0 à une vitesse géométrique. En particulier, $(jb(j))$ est sommable donc $m < \infty$. Par holomorphie de l'autre fonction sur un voisinage de $\mathbb{D}$, on a en particulier $\sum_{n\in\mathbb{N}} |u(n)-\lambda|1^n < \infty$ donc $u(n) \to \lambda$. On a précédemment établi que, comme b est apériodique et $m < \infty$, on a $u(n) \to \frac{1}{m}$. Par unicité de la limite, $\lambda = \frac{1}{m}$.

Comme la suite $(u(n))$ est bornée, la série entière $F = \sum_{n\in\mathbb{N}} (u(n)-\lambda)z^n$ définit toujours une fonction holomorphe à l'intérieure de $\mathbb{D}$. Il en va de même pour $W = \sum_{n\in\mathbb{N}} (u(n+1)-u(n))z^{n+1}$. Dans l'intersection de leurs domaines de définitions, on a $W = (1-z)F+\lambda$. En outre, à l'intérieur de $\mathbb{D}$, on a $|B| < 1$ donc $W = (1-z)U = \frac{1-z}{1-B}$.

Supposons le premier point vérifié, et posons $\lambda = \frac{1}{m}$. Pour $r > 1$, on a :

$$r^n \left| u(n) - \frac{1}{m} \right| \leq r^n \sum_{k=n}^{\infty} |u(k+1)-u(k)| \leq \sum_{k=n}^{\infty} |u(k+1)-u(k)| r^{k+1} .$$

Si on arrive arrive à majorer le terme de gauche indépendamment de n pour un certain $r > 1$, alors F sera holomorphe dans le disque ouvert de rayon r, sur un voisinage ouverte de $\mathbb{D}$. Pour ce faire, il suffit de montrer que W est holomorphe sur un voisinage du disque.

Par hypothèse, la série B est bien définie sur un voisinage de $\mathbb{D}$. Comme $|B| < 1$ à l'intérieur de $\mathbb{D}$, $1-B$ ne s'y annule pas.

Sur $\partial\mathbb{D} = \mathbb{U}$, $B\big(e^{i\theta}\big)$ est bien défini. On peut alors effectuer une majoration terme à terme $\mathrm{Re}\Big(B\big(e^{i\theta}\big)\Big) = \sum_{k\in\mathbb{N}} b(k)\cos(k\theta) \leq \sum_{k\in\mathbb{N}} b(k) = 1$. En conséquence, si $B\big(e^{i\theta}\big) = 1$, alors pour tout $k \in \mathrm{Supp}(b)$, on doit avoir $\cos(k\theta) = 1$ donc $k\theta \equiv 0[2\pi]$. Comme b est apériodique, par relation de Bezout, on a $k_1, \ldots, k_j \in \mathrm{Supp}(b)$ et $l_1, \ldots, l_j \in \mathbb{Z}$ tels que $\sum_{m=1}^{j} k_m l_m = 1$. En conséquence, $\theta = \sum_{m=1}^{j} l_m \times k_m\theta \equiv 0[2\pi]$. Autrement dit, le seul point où $1-B$ s'annule sur $\partial\mathbb{D}$ est 1.

Par principe des zéros isolés appliqué à $1-B$, $\frac{1-z}{1-B}$ est un méromorphisme sur un voisinage de $\mathbb{D}$, dont la seule singularité est en 1. En outre, $B'(1) = m \geq 1 > 0$ donc la singularité en 1 est effaçable. En conséquence, $W = \frac{1-z}{1-B}$ est bien holomorphe sur un voisinage de $\mathbb{D}$.

Réciproquement, supposons que F est holomorphe sur un voisinage de $\mathbb{D}$, pour un certain λ. Il en va de même pour $W = (1-z)F + \lambda$. On a donc le méromorphisme $B = 1 - \frac{1-z}{W}$ sur un voisinage de $\mathbb{D}$.

Comme B est continue sur $\mathbb{D}$, jusqu'au bord, on en déduit que W n'a aucune singularité sur $\mathbb{D}$ sauf peut-être en 1. En outre, comme $W = (1-z)F + \lambda$, on a $W(1) = \lambda \neq 0$. Finalement, par principe des zéros isolés, W ne s'annule pas sur un voisinage de $\mathbb{D}$, donc $B = 1 - \frac{1-z}{W}$ est en fait un holomorphisme sur ce domaine. □

4.2 Théorie de renouvellement et chaînes atomiques

Dans cette sous-section, on considère un noyau P avec un atome α récurrent. On pose ici $S_0 = \sigma_\alpha$ et $S_{k+1} = S_k + \sigma_\alpha \circ \theta_{S_k}$.

Si on a une mesure initiale ξ telle que $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, alors (S_k) est bien un processus de renouvellement, avec $a_\xi(k) = \mathbb{P}_\xi(\sigma_\alpha = k)$ la distribution du délai, et $b(k) = \mathbb{P}_\alpha(\sigma_\alpha = k)$ la distribution des temps d'attente.

On a $u(n) = \mathbb{P}_\alpha(X_n \in \alpha)$ et $v_{a_\xi}(n) = \mathbb{P}_\xi(X_n \in \alpha)$.

Corollaire 87 :
Si P a un atome accessible α apériodique positif-récurrent, alors :

$$P_\alpha^n(\alpha) = \mathbb{P}_\alpha(X_n \in \alpha) \underset{n\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]} .$$

Démonstration. Il nous faut satisfaire les hypothèses du théorème précédent. Autrement dit, on veut montrer que b est apériodique. On a vu que $P_\alpha^n(\alpha) = \sum_{k=0}^{\infty} b^{*k}(n)$. Si $P_\alpha^n(\alpha) > 0$, alors on a un certain k tel que $b^{*k}(n) > 0$, d'où $j_1, \ldots, j_k$ tels que $\sum_{r=1}^{k} j_r = n$ et $\prod_{r=1}^{k} b(j_r) > 0$. La période de b divise chaque j_r donc leur somme n. Elle divise donc le pgcd de tous ces n, la période de α égale à 1. On conclut en appliquant le théorème de Blackwell. $\square$

Proposition 88 :
Soient P avec un atome accessible α positif-récurrent, et π sa mesure invariante. On a alors :

$$\|\xi P^n - \pi\|_{\mathrm{VT}} \leq \mathbb{P}_\xi(\sigma_\alpha \geq n) + |a_\xi * u - \pi(\alpha)| * \psi(n) + \pi(\alpha) \sum_{k=n+1}^{\infty} \psi(k) ,$$

où $\psi(k) = \mathbb{P}_\alpha(\sigma_\alpha \geq k)$ pour $k \geq 1$, et $\psi(0) = 0$.

Démonstration. Soit h mesurable, telle que $|h| \leq 1$. En faisant une décomposition en fonction de la dernière visite en α, on a :

$$\begin{aligned} \xi P^n(h) &= \mathbb{E}_\xi[h(X_n)\mathbb{1}_{\sigma_\alpha \geq n}] + \sum_{l=0}^{n-1} \mathbb{E}_\xi\left[h(X_n)\mathbb{1}_\alpha(X_l) \prod_{k=l+1}^{n-1} \mathbb{1}_{\alpha^c}(X_k)\right] \\ &= \mathbb{E}_\xi[h(X_n)\mathbb{1}_{\sigma_\alpha \geq n}] + \sum_{l=0}^{n-1} \mathbb{P}_\xi(X_l \in \alpha)\mathbb{E}_\alpha[h(X_{n-l})\mathbb{1}_{\sigma_\alpha \geq n-l}] \\ &= \mathbb{E}_\xi[h(X_n)\mathbb{1}_{\sigma_\alpha \geq n}] + \sum_{l=0}^{n-1} (a_\xi * u)(l)\mathbb{E}_\alpha[h(X_{n-l})\mathbb{1}_{\sigma_\alpha \geq n-l}] . \end{aligned}$$

Par la formule de Kac, on a $\pi(h) = \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right] = \pi(\alpha) \sum_{k=1}^{\infty} \mathbb{E}_\alpha[h(X_k)\mathbb{1}_{\sigma_\alpha \geq k}]$.

En combinant les deux, on obtient :

$$
\begin{array}{rcl}
|\xi P^n(h) - \pi(h)| & \leq & |\mathbb{E}_\xi[h(X_n)\mathbb{1}_{\sigma_\alpha \geq n}]| \\
& + & \sum_{k=1}^{n} |a_\xi * u - \pi(\alpha)|(n-k) \times |\mathbb{E}_\alpha[(X_k)\mathbb{1}_{\sigma_\alpha \geq k}]| \\
& + & \pi(\alpha) \sum_{k=n+1}^{\infty} |\mathbb{E}_\alpha[h(X_k)\mathbb{1}_{\sigma_\alpha \geq k}]| \\
& \leq & \mathbb{P}_\xi(\sigma_\alpha \geq n) \\
& + & \sum_{k=1}^{n} |a_\xi * u - \pi(\alpha)|(n-k) \times \mathbb{P}_\alpha(\sigma_\alpha \geq k) \\
& + & \pi(\alpha) \sum_{k=n+1}^{\infty} \mathbb{P}_\alpha[\sigma_\alpha \geq k] \\
& = & \mathbb{P}_\xi(\sigma_\alpha \geq n) + |a_\xi * u - \pi(\alpha)| * \psi(n) + \pi(\alpha) \sum_{k=n+1}^{\infty} \psi(k)
\end{array}
$$

c'est-à-dire le résultat voulu en passant au supremum pour tout h. □

5 Notion d'ensemble *small*

5.1 Ensemble *small*

Définition 89 (Ensemble *small*) **:**

On dit que $C \subset \mathbb{X}$ est (m, μ)-*small* lorsqu'il existe $m > 0$ est une mesure de minoration μ non triviale ($0 < \mu(\mathbb{X}) < \infty$) tels que pour tout $x \in C$, $P_x^m \geq \mu$.

Remarque 90 (Exemple du modèle auto-régressif) **:**

On pose $X_{k+1} = \alpha X_k + Z_{k+1}$ où $Z_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$.

Plus généralement, les Z_k sont iid à densité $f > 0$. On a $P_x(A) = \int \mathbb{1}_A(y) f(y - \alpha x) \, \mathrm{d}y$.

Si C est compact, alors pour tout $x \in C$, $P_x(A) \geq P_x(A \cap C) \geq \lambda(A \cap C) \times \inf\limits_{x,y \in C} f(y - \alpha x)$.

Si f est semi-continue inférieurement, alors cet infimum est atteint, strictement positif.

Lemme 91 :

Si C est small et accessible, alors il existe une mesure η telle que C est η-small et $\eta(C) > 0$.

Démonstration. On définit le noyau de transition $K_\varepsilon = (1-\varepsilon) \sum\limits_{k=0}^{\infty} \varepsilon^k \times P^k$, qui effectue un nombre aléatoire de transitions (suivant une loi géométrique de paramètre $0 < \varepsilon < 1$) sous la loi P.

Comme C est (m, μ)-small, pour tout $x \in C$, on a $K_\varepsilon(x, A) \geq (1 - \varepsilon)\varepsilon^m \mu$.

Comme C est accessible, pour tout $x \in \mathbb{X}$, on a $K_\varepsilon(x, C) > 0$. En conséquence, en passant à l'espérance selon μ, on a $\mu K_\varepsilon(C) > 0$. Il existe donc $n \in \mathbb{N}$ tel que $\mu P^n(C) > 0$.

Pour tout $x \in C$, on a alors :

$$P_x^{n+m}(A) = \int_{\mathbb{X}} P_y^n(A) \, \mathrm{d}P_x^m(y) \geq \int_C P_y^n(A) \, \mathrm{d}\mu(y) = \mu P^n(A) \,,$$

donc C est $((n+m), \mu P^n)$-small, et $\mu P^n(C) > 0$. □

Lemme 92 :

Soit C un ensemble (m, μ)-small. Si A est accessible, alors il existe un entier $q \geq 1$ tel que $\inf\limits_{x \in C} P_x^q(A) > 0$.

Démonstration. Comme A est accessible, on a $\mu P^n(A) > 0$ pour un certain n. Alors pour tout

$x \in C$, on a $P_x^{n+m}(A) \geq \mu P^n(A) > 0$. □

Lemme 93 :

Supposons que D conduit uniformément à C un ensemble (m,μ)-small. Autrement dit, on suppose que $\inf\limits_{x\in D} P_x^n(C) = \delta > 0$ pour un certain n. Alors D est $(n+m, \delta\mu)$-small.

Démonstration. Pour $x \in D$, on a $P_x^{n+m}(A) = \int\limits_{\mathbb{X}} P_y^m(A)\,\mathrm{d}P_x^n(y) \geq \int\limits_C \mu(A)\,\mathrm{d}P_x^n(y) \geq \delta\mu(A)$. □

Corollaire 94 :

Si P admet un ensemble C small et accessible, alors on peut recouvrir $\mathbb{X}$ par une famille dénombrable d'ensembles small.

Démonstration. Posons $C_{m,n} = \left\{x \in \mathbb{X},\ P_x^n(C) \geq \frac{1}{m}\right\}$. Comme l'ensemble C est accessible, on a bien $\mathbb{X} = \bigcup\limits_{m,n\in\mathbb{N}^*} C_{m,n}$. Reste donc à vérifier que chaque ensemble $C_{m,n}$ est small.

Par définition, $\inf\limits_{x\in C_{m,n}} P_x^n(C) \geq \frac{1}{m}$ donc $C_{m,n}$ conduit uniformément à C. D'après le lemme précédent, c'est bien un ensemble small. □

Lemme 95 :

Soit C un ensemble (m,μ)-small et accessible. Si $\mu(A) > 0$ alors A est accessible.

Démonstration. Soit $x \in \mathbb{X}$. Par accessibilité de C on a $n > 0$ tel que $P_x^n(C) > 0$. Alors $P_x^{n+m}(A) \geq \int\limits_C P_y^m(A)\,\mathrm{d}P_x^m(y) \geq \mu(A) \times P_x^n(C) > 0$. Donc A est accessible. □

5.2 Irréductibilité

Définition 96 (Noyau irréductible) **:**

Dans ce contexte, P est irréductible lorsqu'il admet un ensemble small et accessible.

Définition 97 (Mesure d'irréductibilité) **:**

La mesure φ est une mesure d'irréductibilité lorsque $\varphi(A) > 0$ implique que A est accessible. On dit que cette mesure φ est maximale lorsque la réciproque est vraie, lorsque $\varphi(A) > 0$ ssi A accessible.

Remarque 98 :

La définition *classique* d'irréductibilité pour les chaînes de Markov est celle d'un noyau pour lequel il existe une mesure d'irréductibilité non triviale.

Théorème 99 :

Si φ est une mesure d'irréductibilité, alors φK_ε est une mesure d'irréductibilité maximale pour $0 < \varepsilon < 1$.

Toute mesure d'irréductibilité est absolument continue par rapport à toute mesure maximale. En particulier, toutes les mesures maximales sont équivalentes.

Démonstration. Pour la seconde partie, considérons φ irréductible et ψ maximale. Si $\psi(A) = 0$, alors A n'est pas accessible, donc $\varphi(A) = 0$, d'où $\varphi \ll \psi$.

Pour la première partie, considérons d'abord A accessible. Comme $K_\varepsilon(x, A) > 0$ pour tout x, on a donc $\varphi K_\varepsilon(A) > 0$. En admettant que la mesure est irréductibilité, elle sera maximale.

Reste à montrer que c'est bien une mesure d'irréductibilité. Pour un ensemble mesurable A quelconque, définissons l'ensemble $\overline{A} = \{x \in \mathbb{X},\, \mathbb{P}_x(\tau_A < \infty) > 0\} = \{x \in \mathbb{X},\, K_\varepsilon(x, A) > 0\}$, l'ensemble des configurations qui vont visiter A avec probabilité positive.

On a $K_\varepsilon(A) = \int_{\overline{A}} K_\varepsilon(x, A)\,\mathrm{d}\varphi(x)$. Si $K_\varepsilon(A) > 0$ alors en particulier $\varphi\big(\overline{A}\big) > 0$ d'où enfin $\overline{A}$ accessible. On a donc $\mathbb{P}_x(\sigma_A < \infty) \geq \mathbb{P}_x\big(\sigma_A \circ \theta_{\sigma_{\overline{A}}} < \infty\big) = \mathbb{E}_x\Big[\mathbb{1}_{\sigma_{\overline{A}} < \infty} \mathbb{P}_{X_{\sigma_{\overline{A}}}}(\tau_A < \infty)\Big] > 0$. □

Théorème 100 :

Soit P irréductible. Alors tout ensemble accessible contient un ensemble small accessible.

Démonstration. Posons D un ensemble (m, μ)-small accessible.

Soit un ensemble A accessible. Posons $A_{p,q} = \Big\{x \in A,\, \mathbb{P}_x(\sigma_D = p) \geq \frac{1}{q}\Big\}$. Ces ensembles conduisent uniformément à D donc sont small également.

En outre, en considérant une mesure d'irréductibilité maximale ψ, par exemple μK_ε, on a $\psi(A) = \psi\left(\bigcup_{p,q \in \mathbb{N}^*} A_{p,q}\right) > 0$, donc un des ensembles $A_{p,q}$ est small *et* accessible. □

Théorème 101 (Jain, Jamieson) **:**

Lorsque la tribu $\mathcal{X}$ est engendrée par une famille dénombrable, P est irréductible ssi il existe une mesure d'irréductibilité non triviale.

Démonstration. Admis. □

5.3 Lien entre irréductibilité et stationnarité

Définition 102 (Noyau positif) **:**
On dit que P est un noyau positif lorsque c'est un noyau irréductible, qui admet une mesure de probabilité invariante π.

Théorème 103 :
Si P est π-positif, alors π est une mesure d'irréductibilité maximale.

Démonstration. Si π est une mesure d'irréductibilité, comme $\pi K_\varepsilon = \pi$, alors π est maximale.

Supposons donc que $\pi(A) > 0$ et montrons que A est accessible. On définit l'ensemble $\overline{A} = \{x \in \mathbb{X},\, \mathbb{P}_x(\tau_A < \infty) > 0\}$ comme précédemment. En particulier, $A \subset \overline{A}$ ici.

Si $x \notin \overline{A}$, alors $0 = \mathbb{P}_x(\tau_A < \infty) \geq \mathbb{E}_x\left[\mathbb{1}_{\sigma_{\overline{A}} < \infty} \mathbb{P}_{X_{\sigma_{\overline{A}}}}(\tau_A < \infty)\right]$. Comme la probabilité dans l'espérance est toujours strictement positive, on a donc $\mathbb{P}_x(\sigma_{\overline{A}} < \infty) = 0$, et donc $K_\varepsilon\big(x, \overline{A}\big) = 0$. On en déduit que $\pi\big(\overline{A}\big) = \pi K_\varepsilon\big(\overline{A}\big) = \int_{\overline{A}} K_\varepsilon\big(x, \overline{A}\big)\, \mathrm{d}\pi(x)$. Comme $K_\varepsilon\big(x, \overline{A}\big) \leq 1$, la fonction doit être presque-sûrement égale à 1 sur $\overline{A}$.

Comme $\pi\big(\overline{A}\big) \geq \pi(A) > 0$, il existe en particulier $x \in \overline{A}$ tel que $K_\varepsilon\big(x, \overline{A}\big) = 1$. Finalement, $K_\varepsilon\big(x, \overline{A}^c\big) = 0$. Autrement dit, $\overline{A}^c$ n'est *pas* accessible, donc pour une mesure d'irréductibilité maximale, on a $\psi\big(\overline{A}^c\big) = 0$ donc naturellement $\psi\big(\overline{A}\big) > 0$.

En conclusion, $\overline{A}$ est accessible, et conduit avec probabilité positive à A, donc A est accessible. La mesure invariante est bien une mesure d'irréductibilité. $\square$

5.4 Unicité de la mesure invariante

Lemme 104 :
Soit λ une mesure signée finie P-invariante. Alors λ^+ est aussi P-invariante.

Démonstration. Soit S un ensemble de Jordan, tel que $\lambda^+(A) = \lambda(A \cap S)$. Dans ce cas, on a $\lambda^+ P(B) \geq \lambda^+ P(B \cap S) = \int P_x(B \cap S)\, \mathrm{d}\lambda^+(x) \geq \int P_x(B \cap S)\, \mathrm{d}\lambda(x) = \lambda(B \cap S) = \lambda^+(B)$.

La mesure $\lambda^+ P - \lambda^+$ est positive, de masse finie. En outre, $\lambda^+(\mathbb{X}) = \lambda^+ P(\mathbb{X})$, donc la mesure est nulle, λ^+ est P-invariante. $\square$

Corollaire 105 :
Si P est irréductible, alors P admet au plus une probabilité invariante.

Démonstration. Soient μ et γ deux probabilités invariantes. La mesure signée $\mu - \gamma$ est aussi invariante, donc $(\mu - \gamma)^\pm$ sont deux mesures invariantes, positives.

Comme $\mu(\mathbb{X}) - \gamma(\mathbb{X}) = 1 - 1 = 0$, alors les mesures $(\mu - \gamma)^\pm$ sont positives, P-invariantes, de même masse. En particulier, une des mesures est non-triviale ssi l'autre l'est.

Si ces mesures étaient non-triviales, alors à normalisation près, on obtiendrait deux mesures de probabilité invariantes, donc des mesures d'irréductibilité maximales, équivalentes par ce qui précède.

Cependant, ces mesures sont à supports de Jordan disjoints, donc elle ne peuvent pas décrire les mêmes ensembles accessibles. En conclusion, $(\mu - \gamma)^{\pm} = 0$, donc $\mu = \gamma$. $\square$

Proposition 106 :

Si P est irréductible, alors A accessible ssi $\overline{A} = \{x \in \mathbb{X},\, \mathbb{P}_x(\tau_A < \infty) > 0\}$ l'est.

Démonstration. On a en particulier $A \subset \overline{A}$ donc le sens direct est évident. Réciproquement, si A est inaccessible, soit ψ une mesure d'irréductibilité maximale, telle que $\psi(A) = 0$. On a donc $\psi K_\varepsilon(A) = 0 = \psi(\{x,\, \mathbb{P}_x(\tau_A < \infty) > 0\}) = \psi(\overline{A})$. □

Définition 107 (Ensemble *full*) **:**

On dit que F est full lorsque F^c est inaccessible.

Lemme 108 :

Si P est irréductible et F full, alors F accessible.

Si P est irréductible et A absorbant, alors A est full, donc accessible.

Définition 109 (Période d'un ensemble) **:**

Soit C small accessible. On pose $d(c) = \operatorname{pgcd}\left\{n \in \mathbb{N}^*,\, \inf_{x \in C} P_x^n(C) > 0\right\}$ la période de C.

Proposition 110 :

Tous les ensembles small accessibles ont la même période, appelée période de P.

Définition 111 :

On dit que P est apériodique lorsque $d = 1$.

Si il existe un ensemble C $(1, \mu)$-small tel que $\mu(C) > 0$, alors P est dit fortement apériodique.

Définition 112 :

Un ensemble C est *petite* lorsqu'il existe une probabilité a sur $\mathbb{N}$ (plus largement une famille sommable) et une mesure non nulle μ tels que $K_a(x, A) := \sum_{k \in \mathbb{N}} a(k) P_x^k(A) \geq \mu(A)$.

On dit alors que C est (a, μ)-petite.

Lemme 113 :

Si C est (n, μ)-small, alors C est (δ_n, μ)-petite.

6 Mesure invariante et *splitting construction*

Définition 114 (Splitting) **:**
Soit P un noyau irréductible, et C un ensemble $(1, \varepsilon\nu)$-small, où ν est une probabilité.

On pose $R_x(A) = \frac{P_x - \varepsilon\nu}{1-\varepsilon}$ lorsque $x \in C$ et $R_x = P_x$ lorsque $x \notin C$. Pour le dire autrement, on a $P_x = (1 - \varepsilon \mathbb{1}_C(x))R_x + \varepsilon \mathbb{1}_C(x)\nu$.

On définit alors noyau $Q_{x,d}$ de $\mathbb{X} \times \{0,1\}$ vers $\mathbb{X}$ via $Q_{x,d} = R_x$ lorsque $x \in C$ et $d = 0$, $Q_{x,d} = \nu$ lorsque $x \in C$ et $d = 1$, et $Q_{x,d} = P_x$ lorsque $x \notin C$. On pose en outre $b_\varepsilon = \varepsilon\delta_1 + (1-\varepsilon)\delta_0$ la loi de Bernouilli de paramètre ε, qu'on peut voir comme un noyau de $\mathbb{X} \times \{0,1\}$ vers $\{0,1\}$, qui retourne une variable $\mathcal{B}(\varepsilon)$ indépendante du passé.

On pose enfin $\check{P} = Q \otimes b_\varepsilon$. C'est un noyau Markovien sur $\mathbb{X} \times \{0,1\}$.

Lemme 115 :
Si le noyau P est irréductible et C est $(1, \varepsilon\nu)$-small, pour toute mesure positive ξ, on a $(\xi \otimes b_\varepsilon)\check{P}^n = (\xi P^n) \otimes b_\varepsilon$.

Démonstration. Il suffit de le montrer pour $n = 1$ pour conclure par induction directe. Or :

$$\begin{aligned}
\int \check{P}_{x,d}(A,B)\,\mathrm{d}\xi(x)\,\mathrm{d}b_\varepsilon(d) &= b_\varepsilon(B) \times \left(\int_C \varepsilon Q_{x,1}(A) + (1-\varepsilon)Q_{x,0}(A)\,\mathrm{d}\xi(x) + \int_{C^c} P_x(A)\,\mathrm{d}\xi(x) \right) \\
&= b_\varepsilon(B) \times \int_{\mathbb{X}} P_x(A)\,\mathrm{d}\xi(x) \\
&= b_\varepsilon(B) \times \xi P(A)\,,
\end{aligned}$$

d'où le résultat. □

Proposition 116 :
Soient P irréductible et C $(1, \varepsilon\nu)$-small. On considère (X, D) la $\check{P}$-chaîne de Markov initialisée en $\xi \otimes b_\varepsilon$. Alors X est une P-chaîne initialisée en ξ pour sa filtration $\left(\mathcal{F}_k^X\right)_{k\in\mathbb{N}}$ canonique.

Démonstration. Soit g une fonction positive mesurable sur $\mathbb{X}$. On étend naturellement g en fonction sur $\mathbb{X} \otimes \{0,1\}$, et en particulier $\check{P}g(x,d) = Qg(x,d)$. En conséquence :

$$\mathbb{E}\left[g(X_{n+1})\middle|\mathcal{F}_n^X\right] = \mathbb{E}\left[\mathbb{E}[g(X_{n+1})|\mathcal{F}_n]\middle|\mathcal{F}_n^X\right] = \mathbb{E}\left[Qg(X_n, d_n)\middle|\mathcal{F}_n^X\right].$$

On peut vérifier par calcul direct que $\mathbb{E}_{d\sim\mathcal{B}(\varepsilon)}[Qg(x,d)] = Pg(x)$. Comme ici $d_n \sim \mathcal{B}(\varepsilon)$ est indépendant de $\mathcal{F}_n^X$, on a bien :

$$\mathbb{E}\left[g(X_{n+1})\middle|\mathcal{F}_n^X\right] = \mathbb{E}\left[Qg(X_n, d_n)\middle|\mathcal{F}_n^X\right] = \mathbb{E}_{d\sim\mathcal{B}(\varepsilon)}[Qg(X_n, d)] = Pg(X_n)\,.$$

□

Proposition 117 :

Si P est irréductible et C est $(1, \varepsilon\nu)$-small, alors :

1. Si $\lambda P = \lambda$, alors $(\lambda \otimes b_\varepsilon)\check{P} = \lambda \otimes b_\varepsilon$.
2. Si $\check{\lambda}$ est $\check{P}$-invariante, alors en particulier on a la factorisation $\check{\lambda} = \check{\lambda}_0 \otimes b_\varepsilon$ qu'on définit par $\check{\lambda}_0(A) = \check{\lambda}(A \times \{0, 1\})$, et $\check{\lambda}_0$ est P-invariante.

Démonstration. Le premier point découle du lemme ci-dessus.

Pour le second point, vérifions d'abord l'égalité tensorisée. On a :

$$\check{\lambda}(A \times B) = \check{\lambda}\check{P}(A \times B) = \check{\lambda}Q(A) \times b_\varepsilon(B)$$

et en particulier pour $B = \{0, 1\}$ on vérifie bien $\check{\lambda}_0(A) = \check{\lambda}(A \times B) = \check{\lambda}Q(A) \times b_\varepsilon(B) = \check{\lambda}Q(A)$.

Avec cette factorisation, on a en particulier $\check{\lambda}_0 \otimes b_\varepsilon = \check{\lambda} = \check{\lambda}\check{P} = \big(\check{\lambda}_0 P\big) \otimes b_\varepsilon$ donc en particulier, $\check{\lambda}_0 = \check{\lambda}_0 P$ est bien P-invariante. □

Proposition 118 :

Supposons ici que P est irréductible, et que C est $(1, 2\varepsilon\nu)$-small, avec ν une mesure de probabilité sur C. En particulier, sur C, $R_x \geq \frac{\varepsilon}{1-\varepsilon}\nu$. Alors :

1. $\check{\alpha} = C \times \{1\}$ est un atome apériodique pour $\check{P}$.
2. $C \times \{0, 1\}$ est small pour $\check{P}$.
3. Si C est accessible, alors $\check{\alpha}$ est accessible pour $\check{P}$, donc $\check{P}$ est irréductible.
4. Pour $k \geq 1$, on a $\check{P}^k(\check{\alpha}, \check{\alpha}) = \varepsilon\nu P^{k-1}(C)$.
5. Si C est récurrent ($\forall x \in C$, $U(x, C) = \infty$) pour P alors $\check{\alpha}$ l'est pour $\check{P}$.
6. Si C est Harris-récurrent ($\forall x \in C$, $\mathbb{P}_x(\sigma_C < \infty) = 1$) pour P, alors pour toute loi initiale ξ sur $\mathbb{X}$, lorsque $\mathbb{P}_\xi(\sigma_C < \infty)$, on a $\mathbb{P}_{\xi \otimes b_\varepsilon}(\sigma_{\check{\alpha}} < \infty) = 1$.
7. Si C est accessible, et $\pi P = \pi$, alors $\check{\alpha}$ est un atome positif pour $\check{P}$, autrement dit $\mathbb{E}_{\check{\alpha}}[\sigma_{\check{\alpha}}] < \infty$.

Démonstration. Pour le premier point, il est clair que α est un atome par définition de $\check{P}$. En outre, $\check{P}(\check{\alpha}, \check{\alpha}) = \nu(C)b_\varepsilon(1) = \varepsilon > 0$, donc on a bien un atome apériodique.

Pour le second point, on a $\check{P}_{x,0} \geq \frac{\varepsilon}{1-\varepsilon}\nu \otimes b_\varepsilon$ et $\check{P}_{x,1}\nu \otimes b_\varepsilon$. Par construction, $2\varepsilon\nu(X) = 2\varepsilon \leq 1$ donc $\frac{\varepsilon}{1-\varepsilon} \leq 1$, d'où finalement $\check{P}_{x,d} \geq \frac{\varepsilon}{1-\varepsilon}\nu \otimes b_\varepsilon$. On a bien un ensemble $\big(1, \frac{\varepsilon}{1-\varepsilon}\big)$-small.

Pour le troisième point, si $n \geq 1$, on a :

$$\mathbb{P}_{x,d}(X_n \in C,\, d_n = 1) = \varepsilon\mathbb{P}_{x,d}(X_n \in C) = \varepsilon\mathbb{E}_{x,d}[\mathbb{P}_{X_1}(X_{n-1} \in C)] = \varepsilon\mathbb{P}_{Q_{x,d}}(X_{n-1} \in C)\,.$$

En conséquence, par accessibilité de C, pour tout $(x, d) \in \mathbb{X} \times \{0, 1\}$, on a $n \in \mathbb{N}$ tel que $\mathbb{P}_{Q_{x,d}}(X_{n-1} \in C) > 0$, donc $\check{\alpha}$ est bien un atome accessible pour $\check{P}$.

Le quatrième point se fait via la même propriété de Markov sur une étape, dans le cas où $x \in C$ (et $d = 1$) donc $X_1 \sim \nu$.

Pour le cinquième point, on a alors $U(\check{\alpha}, \check{\alpha}) = \sum_{k \in \mathbb{N}} \check{P}_{\check{\alpha}}^{k+1}(\check{\alpha}) = \varepsilon \sum_{k \in \mathbb{N}} \nu P^k(C) = \varepsilon \nu U(C)$. Comme $\nu(C) = 1$ et $U(x, C) = \infty$ pour tout $x \in C$, on en déduit bien que $U(\check{\alpha}, \check{\alpha}) = \infty$, $\check{\alpha}$ est récurrent.

Pour le sixième point, on va montrer $\mathbb{P}_{\xi \otimes b_\varepsilon}(N_{\check{\alpha}} = \infty) = 1$, ce qui implique directement le résultat désiré. On vérifie que $\inf_{z \in C \times \{0,1\}} \check{P}_z(\check{\alpha}) \geq \varepsilon^2$. Comme $\check{\alpha}$ est uniformément accessible depuis $C \times \{0, 1\}$, il nous suffit donc de montrer que $\mathbb{P}_{\xi \otimes b_\varepsilon}\big(N_{C \times \{0,1\}} = \infty\big) = 1$ sous les mêmes hypothèses. Si $\mathbb{P}_\xi(\sigma_C < \infty) = 1$, ce qui est ici en particulier le cas lorsque $\xi(C) = 1$, on a l'égalité $\mathbb{P}_{\xi \otimes b_\varepsilon}\big(\sigma_{C \times \{0,1\}} < \infty\big) = \mathbb{P}_\xi(\sigma_C < \infty) = 1$. En outre, $\mathbb{P}_{\xi \otimes b_\varepsilon} = (1 - \varepsilon)\mathbb{P}_{\xi \otimes \delta_0} + \mathbb{P}_{\xi \otimes \delta_1}$. On en déduit que l'évènement précédent est de probabilité 1 sous chacune des deux probabilités, et c'est en particulier le cas en partant de $(x, d) \in C \times \{0, 1\}$. Autrement dit, $C \times \{0, 1\}$ est récurrent, En conséquence, dès que $\mathbb{P}_\xi(\sigma_C < \infty) = 1$, on a $\mathbb{P}_{\xi \otimes b_\varepsilon}\big(N_{C \times \{0,1\}} = \infty\big) = 1$, ce qui conclut ce point.

Pour le septième point, comme π est P-invariante, $\check{\pi} = \pi \otimes b_\varepsilon$ est $\check{P}$-invariante. Le noyau $\check{P}$ est donc irréductible récurrent. Comme l'atome $\check{\alpha}$ est accessible, et $\pi(C) > 0$, il est positif. $\square$

7 Ergodicité V-uniforme

Théorème 119 :

Soit P un noyau sur $\mathbb{X}$. Soient $\mathbb{F}$ un sous-espace vectoriel de $\mathcal{M}^1(\mathbb{X})$, stable par P, et φ une distance sur $\mathbb{F}$ qui le rend complet.

Supposons qu'il existe des constantes A_r ($1 \leq r < m$) telles que $\forall \xi, \xi' \in \mathbb{F}$, on a uniformément $\varphi(\xi P^r, \xi' P^r) \leq A_r \varphi(\xi, \xi')$, et que $\xi \mapsto \xi P^m$ est α-contractante, pour $\alpha < 1$.

Alors il existe une unique mesure P-invariante $\pi \in \mathbb{F}$, et pour toute mesure $\xi \in \mathbb{F}$, on a :

$$\varphi(\xi P^n, \pi) \leq \max\left(1, \max_{1 \leq r < m} A_r\right) \times \varphi(\xi, \pi) \alpha^{\left\lfloor \frac{n}{m} \right\rfloor}.$$

Si de plus la convergence pour φ dans $\mathbb{F}$ implique la convergence *setwise* (pour tout mesurable A, on a $\pi_n(A) \to \mu(A)$) ou bien la convergence faible dans $\mathcal{M}^1(\mathbb{X})$, et que pour tout $x \in \mathbb{X}$ on a $\delta_x \in \mathbb{F}$, alors π est l'unique probabilité P-invariante sur $\mathcal{M}^1(\mathbb{X})$.

Démonstration. Si π est P-invariante, alors elle est P^m-invariante. Comme P^m est α-contractante, on en déduit l'unicité.

Par division euclidienne, on peut toujours décomposer $n = m \left\lfloor \frac{n}{m} \right\rfloor + r$. Quitte à utiliser les inégalités en hypothèse, on en déduit que pour toutes mesures $\xi, \xi' \in \mathbb{F}$, on a l'inégalité $\varphi(\xi P^n, \xi' P^n) \leq A_r \varphi\left(\xi P^{m \left\lfloor \frac{n}{m} \right\rfloor}, \xi' P^{m \left\lfloor \frac{n}{m} \right\rfloor}\right) \leq A_r \alpha^{\left\lfloor \frac{n}{m} \right\rfloor} \varphi(\xi, \xi') \leq \max\left(1, \max_{1 \leq r < m} A_r\right) \alpha^{\left\lfloor \frac{n}{m} \right\rfloor} \varphi(\xi, \xi')$. En particulier, (ξP^m) est une suite de Cauchy, converge vers π. En outre, on a alors la majoration $\varphi(\pi, \pi P) \leq \varphi(\pi, \xi P^{n+1}) + A_1 \varphi(\pi, \xi P^n) \to 0$, et donc $\pi = \pi P$.

Supposons désormais vérifiée la propriété de *setwise convergence* et considérons une mesure P-invariante μ. Avec $f = \mathbb{1}_A$, on a alors $\widetilde{\pi}(f) = \int P^n f(x) \, \mathrm{d}\widetilde{\pi}(x)$ pour tout n. En outre, par ce qui précède, $P^n f(x) = \delta_x P^n f \to \pi(f)$, car $\delta_x \in \mathbb{F}$. En conséquence, $P^n f$ converge simplement vers $\pi(f)$, et est bornée, donc $\widetilde{\pi}(f) = \pi(f)$ par convergence dominée, $\widetilde{\pi} = \pi$.

La preuve est analogue dans le cas de la convergence faible. □

Remarque 120 (Coefficient de Dobrushin) **:**

On rappelle que $\Delta(P) := \sup_{\xi, \xi' \in \mathcal{M}^1(\mathbb{X})} \frac{\|\xi P - \xi' P\|_{\mathrm{VT}}}{\|\xi - \xi'\|_{\mathrm{VT}}} = \sup_{x, x' \in \mathbb{X}} d_{\mathrm{VT}}(P_x, P_{x'})$.

Lemme 121 :

Pour tous noyaux P et Q on a $\Delta(PQ) \leq \Delta(P)\Delta(Q)$.

Démonstration. Il suffit de revenir à la définition. On a $\frac{\|\xi PQ-\xi' PQ\|_{\mathrm{VT}}}{\|\xi-\xi'\|_{\mathrm{VT}}} = \frac{\|\xi PQ-\xi' PQ\|_{\mathrm{VT}}}{\|\xi P-\xi' P\|_{\mathrm{VT}}} \frac{\|\xi P-\xi' P\|_{\mathrm{VT}}}{\|\xi-\xi'\|_{\mathrm{VT}}}$. On majore le supremum du produit par le produit des supremum, et pour le facteur de gauche, on a alors un supremum sur les mesures ξP, qu'on majore par un supremum sur toutes les probabilités, d'où le résultat. □

Définition 122 (Ensemble de Doeblin) **:**

L'ensemble $C \subset \mathbb{X}$ mesurable est (m, ε)-Doeblin lorsque, pour tous les points $x, x' \in C$, on a $d_{\mathrm{VT}}(P_x^m, P_{x'}^m) \leq 1 - \varepsilon$.

Si $\mathbb{X}$ est un Doeblin, on dit que P satisfait une condition de Doeblin uniforme.

Définition 123 :

Soit $V \geq 1$ une fonction mesurable sur $\mathbb{X}$.

Pour toute fonction f mesurable on note $\|f\|_V = \left\|\frac{f}{V}\right\|_\infty$, et :

$$osc_V(f) = \sup_{x,x' \in \mathbb{X}} \frac{|f(x) - f(x')|}{V(x) + V(x')}.$$

Si $\xi \in \mathcal{M}(\mathbb{X})$ est une mesure signée, $\|\xi\|_V = |\xi|(V)$ et $d_V(\xi, \xi') = \frac{1}{2}\|\xi - \xi'\|_V$.

Comme dans le cas de la distance en variation totale (qui correspond au cas $V = 1$), on a l'égalité $\|\xi\|_V = \sup\{|\xi(f)|, osc_V(f) \leq 1\}$.

On note alors $\mathcal{M}_V^1(\mathbb{X}) = \{\xi \in \mathcal{M}^1(\mathbb{X}), \xi(V) < \infty\}$. L'inclusion peut être stricte dans le cas général, et on a égalité ssi V est majoré. En effet, si V n'est *pas* majoré, on peut construire une mesure de probabilité ξ qui place sa masse là où V prend de grandes valeurs, de sorte que $\xi(V) = \infty$.

Proposition 124 :

L'espace $(\mathcal{M}_V^1(\mathbb{X}), d_V)$ est complet.

Démonstration. La démonstration est la même que pour la complétude de d_{VT}, en utilisant la probabilité $\lambda = \sum_{n=1}^{\infty} \frac{1}{2^n}\mu_n \in \mathcal{M}_V^1$ par rapport à laquelle toute la suite (μ_n) est absolument continue, car alors les densités forment une suite de Cauchy dans l'espace $L^1(V\lambda)$. □

Lemme 125 :

Pour tous $\xi, \xi' \in \mathcal{M}^1$, si $d_{\mathrm{VT}}(\xi, \xi') \leq 1 - \varepsilon$, alors $\|\xi - \xi'\|_V \leq \xi(V) + \xi'(V) - 2\varepsilon$.

Démonstration. Soit $\nu = \xi + \xi' - |\xi - \xi'| = 2\xi \wedge \xi'$ une mesure positive. On a alors l'égalité $\nu(\mathbb{X}) = \xi(\mathbb{X}) + \xi'(\mathbb{X}) - |\xi - \xi'|(\mathbb{X}) \geq 2\varepsilon$ car $|\xi - \xi'|(\mathbb{X}) = 2d_{\mathrm{VT}}(\xi, \xi')$. Comme $V \geq 1 = \mathbb{1}_{\mathbb{X}}$, on a alors $\nu(V) \geq 2\varepsilon$ dont on déduit le résultat. □

Définition 126 (Coefficient de V-Dobrushin) **:**

On pose $\Delta_V(P) = \sup_{\xi \neq \xi' \in \mathcal{M}_V^1} \frac{d_V(\xi P, \xi' P)}{d_V(\xi, \xi')} = \sup_{\xi \neq \xi' \in \mathcal{M}_V^1} \frac{\|\xi P - \xi' P\|_V}{\|\xi - \xi'\|_V}$.

Lemme 127 :

On a $\Delta_V(P) = \sup_{x \neq x'} \frac{\|P_x - P_{x'}\|_V}{V(x)+V(x')} = \sup_{osc_V(f) \leq 1} osc_V(Pf)$.

Démonstration. Pour obtenir l'égalité de droite, il suffit de remarquer qu'un supremum en $x \neq x'$ intervient dans la définition de $osc_V(Pf)$ et qu'inversement un supremum en f intervient pour $\|P_x - P_{x'}\|_V$. On a alors l'égalité en intervertissant les deux supremum.

En ce qui concerne l'égalité de gauche, le sens $\geq$ est direct, en restreignant le supremum sur $\xi \neq \xi' \in \mathcal{M}_V^1$ aux mesures de Dirac $\delta_x \neq \delta_{x'}$. Pour l'inégalité réciproque, on remarque que $\|\xi P - \xi' P\|_V = \sup_{osc_V(f) \leq 1} |\xi Pf - \xi' Pf|$. Or :

$$\begin{aligned}
|\xi Pf - \xi' Pf| &= \left| \int Pf(x)\,\mathrm{d}(\xi - \xi')^+(x) - \int Pf(x')\,\mathrm{d}(\xi' - \xi)^+(x') \right| \\
&= \left| \int \left(Pf(x) - \int Pf(x') \frac{\mathrm{d}(\xi' - \xi)^+(x')}{(\xi' - \xi)(\mathbb{X})} \right) \mathrm{d}(\xi - \xi')^+(x) \right| \\
&= \left| \int (Pf(x) - Pf(x')) \frac{\mathrm{d}(\xi' - \xi)^+(x')\,\mathrm{d}(\xi - \xi')^+(x)}{(\xi' - \xi)(\mathbb{X})} \right| \\
&\leq |\xi - \xi'|(V) \times \sup_{x \neq x'} \frac{|Pf(x) - Pf(x')|}{V(x)+V(x')} \\
&= \|\xi - \xi'\|_V \sup_{x \neq x'} \frac{|Pf(x) - Pf(x')|}{V(x)+V(x')},
\end{aligned}$$

pour toute fonction f, donc en passant au supremum, quitte à intervertir le supremum en f et en $x \neq x'$, on a finalement :

$$\|\xi P - \xi' P\|_V \leq \|\xi - \xi'\|_V \sup_{x \neq x'} \left(\sup_{osc_V(f) \leq 1} \frac{|Pf(x) - Pf(x')|}{V(x)+V(x')} \right),$$

d'où le résultat souhaité. □

Lemme 128 (Hairer-Mattingly) **:**

Soit $V \geq 1$. On suppose les hypothèses suivantes vérifiées :

1. P est un noyau tel que $PV \leq aV + b$ avec $0 < a < 1$ et $b \geq 0$,
2. $C = \{x,\, V(x) \leq d\}$ est un $(1, \varepsilon)$-Doeblin,
3. $\lambda + \frac{2b}{1+d} < 1$.

Alors avec $V_\beta = (1-\beta) + \beta V$, pour tout $\beta \in \left]0, 1 \wedge \frac{\varepsilon}{b+\lambda-1}\right[$, on a :

$$\Delta_{V_\beta}(P) \leq \gamma_1 \vee \gamma_2 < 1,$$

où $\gamma_1(\beta, b, \lambda, \varepsilon) = 1 - \varepsilon + \beta(b + \lambda - 1)$ et $\gamma_2(\beta, b, d, \lambda) = 1 - \beta \frac{(1-\lambda)(1+d) - 2b}{2(1-\beta) + \beta(1+d)}$.

Démonstration. On a par hypothèse $d_{\mathrm{VT}}(P_x, P_{x'}) \leq 1 - \varepsilon \mathbb{1}_{C\times C}(x,x')$. En conséquence, par un lemme précédent :

$$\begin{aligned}
\|P_x - P_{x'}\|_{V_\beta} &\leq PV_\beta(x) + PV_\beta(x') - 2\varepsilon \mathbb{1}_{C\times C}(x,x') \\
&\leq 2(1-\beta) + \beta(PV(x) + PV(x')) - 2\varepsilon \mathbb{1}_{C\times C}(x,x') \\
&\leq \lambda(V_\beta(x) + V_\beta(x')) + 2((1-\lambda)(1-\beta) + \beta b) - 2\varepsilon \mathbb{1}_{C\times C}(x,x') \\
&= \lambda(V_\beta(x) + V_\beta(x')) + 2b_\beta - 2\varepsilon \mathbb{1}_{C\times C}(x,x')\,,
\end{aligned}$$

et donc $\frac{\|P_x - P_{x'}\|_{V_\beta}}{V_\beta(x)+V_\beta(x')} \leq \lambda + 2\frac{b_\beta - \varepsilon \mathbb{1}_{C\times C}(x,x')}{V_\beta(x)+V_\beta(x')}$.

Si $(x,x') \notin C \times C$, alors on majore par $\lambda + 2\frac{(1-\lambda)(1-\lambda)+\beta b}{2(1-\beta)+\beta(1+d)} =: \gamma_2$. La fonction $\gamma_2(\beta)$ est strictement convexe, $\gamma_2(0) = 1$ et $\gamma_2(1) = \lambda + \frac{2b}{d+1} < 1$, donc sur $]0,1]$ on a $\gamma_2 < 1$.

Si $(x,x') \in C \times C$, on distingue plusieurs cas. Quand $b_\beta < \varepsilon$, on majore simplement par $\lambda \leq \gamma_2$. Dans le cas contraire, on majore par $\lambda + b_\beta - \varepsilon = \gamma_1$. Par un calcul direct, on vérifie que $\gamma_1 < 1$ ssi $\beta < \frac{\varepsilon}{b+\lambda-1}$.

Le seul point à vérifier est $b + \lambda - 1 > 0$. Pour ce point, quitte à itérer P, on a :

$$1 \leq P^n V \leq \lambda^n V + b\frac{1-\lambda^n}{1-\lambda} \underset{n\to\infty}{\longrightarrow} \frac{b}{1-\lambda}\,,$$

d'où le résultat. On a donc bien montré que $\Delta_{V_\beta}(P) \leq \gamma_1 \wedge \gamma_2$ en utilisant la caractérisation de $\Delta_{V_\beta}(P)$ donnée par lemme précédent. □