

Convergence en temps long pour les processus de Markov

Pierre Monmarché
pierre.monmarche@sorbonne-universite.fr

(Version du 31 janvier 2022)

Bibliographie

Référence générale sur les probabilités et les processus de Markov :

- O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

Sur les diffusions :

- D.W. Stroock and S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag Berlin Heidelberg, 2006.

Sur les processus de saut :

- M. Davis. *Markov Models and Optimization*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1993.

Sur l'approche par fonctions de Lyapunov :

- R. Khasminskii. *Stochastic Stability of Differential Equations*. Springer-Verlag Berlin Heidelberg, 2012.
- S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.

Sur les inégalités fonctionnelles :

- D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

Sur les couplages :

- C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

1 Introduction

Commençons par une discussion informelle pour motiver le cours et en introduire les principaux objets en éludant les questions techniques. Pour se fixer les idées, prenons comme exemple la solution de l'équation différentielle stochastique

$$dX_t = b(X_t)dt + \sqrt{2}dB_t$$

où $b \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ et $(B_t)_{t \geq 0}$ est un mouvement brownien d -dimensionnel.

1.1 Les objets en jeux

On considère l'opérateur de Markov associé P_t défini par

$$P_t f(x) = \mathbb{E}_x(f(X_t))$$

pour tout f mesurable bornée, où l'indice x signifie que la condition initiale du processus est $X_0 = x$. Ici, $(X_t)_{t \geq 0}$ est un processus de Markov homogène (en temps)

$$\forall t \geq s \geq 0, \quad \mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s) = \mathbb{E}_{X_s}(f(X_{t-s}))$$

où $(\mathcal{F}_s)_{s \geq 0}$ la filtration associée au processus. Ceci se transcrit sur P_t par la propriété de semi-groupe :

$$\forall t, s \geq 0, \quad P_t P_s = P_{t+s}.$$

D'autre part, d'après le calcul d'Itô, pour tout $f \in \mathcal{C}^2(\mathbb{R}^d)$,

$$M_t = f(X_t) - f(X_0) - \int_0^t (b(X_s) \cdot \nabla f(X_s) + \Delta f(X_s)) ds$$

est une martingale, et donc, en prenant l'espérance,

$$P_t f(x) = f(x) + \int_0^t P_s(Lf)(x) ds \quad \text{avec} \quad Lf(x) = b(x) \cdot \nabla f(x) + \Delta f(x). \quad (1)$$

Combiné avec la propriété de semi-groupe homogène, ceci mène à

$$P_{t+s}f(x) - P_t f(x) = \int_0^s P_u P_t(Lf)(x) du.$$

On a d'autre part $P_u f \rightarrow f$ quand $u \rightarrow 0$ pour tout f continue et donc, pour $f \in \mathcal{C}^2(\mathbb{R}^d)$, $x \in \mathbb{R}^d$ et $t \geq 0$,

$$\frac{1}{s} (P_{t+s}f(x) - P_t f(x)) \xrightarrow{s \rightarrow 0} P_t Lf(x).$$

D'autre part, en appliquant l'égalité (1) en remplaçant t par s et f par $P_t f$, on obtient également

$$P_{t+s}f(x) - P_t f(x) = \int_0^s P_u(LP_t f)(x) du$$

et donc

$$\frac{1}{s} (P_{t+s}f(x) - P_t f(x)) \xrightarrow{s \rightarrow 0} LP_t f(x).$$

Conclusion, pour $f \in \mathcal{C}^2(\mathbb{R}^d)$, $x \in \mathbb{R}^d$ et $t \geq 0$,

$$\partial_t P_t f(x) = P_t Lf(x) = LP_t f(x).$$

En particulier, $f_t := P_t f$ est la solution de l'équation aux dérivées partielles (linéaire)

$$\forall t \geq 0, \quad \partial_t f_t = L f_t, \quad f_0 = f.$$

qu'on appelle l'équation de Kolmogorov *backward* associée au processus.

Notons ν_t la loi du processus au temps t , partant d'une condition initiale ν_0 . Cette loi est caractérisée par les valeurs de $\nu_t(f) = \int_{\mathbb{R}^d} f d\nu_t$ pour tout f mesurable bornée. En conditionnant par la condition initiale,

$$\nu_t(f) = \mathbb{E}(f(X_t)) = \int_{\mathbb{R}^d} \mathbb{E}_x(f(X_t)) \nu_0(dx) = \nu_0(P_t f),$$

ce qu'on note $\nu_t = \nu_0 P_t$. En particulier, pour tout $f \in \mathcal{C}^2(\mathbb{R}^d)$,

$$\partial_t \nu_t(f) = \partial_t \nu_0(P_t f) = \nu_0(P_t L f) = \nu_t L f = (L^\dagger \nu_t) f,$$

avec L^\dagger le dual de L pour la dualité mesures/fonctions, c'est-à-dire l'opérateur défini par $(L^\dagger \nu)(f) = \nu(L f)$ pour tout f régulière. Autrement dit, ν_t est solution (au moins au sens faible) de l'équation de Kolmogorov *forward* $\partial_t \nu_t = L^\dagger \nu_t$. En fait, dans le cas de l'opérateur L donné par (1), la théorie des EDP elliptiques nous indique que ν_t possède une densité régulière pour tout $t > 0$ quelque soit la condition initiale, et L^\dagger se calcule en intégrant par partie :

$$(L^\dagger \nu) f = \int_{\mathbb{R}^d} \nu_t(x) L f(x) dx = \int_{\mathbb{R}^d} (\Delta \nu_t(x) - \nabla \cdot (b(x) \nu_t(x))) f(x) dx$$

où $\nabla \cdot$ est l'opérateur divergence. Autrement dit, ν_t est solution de l'EDP

$$\partial_t \nu_t = \Delta \nu_t(x) - \nabla \cdot (b(x) \nu_t(x)). \quad (2)$$

1.2 Convergence à l'équilibre

On dit que μ est une mesure invariante du processus (ou du semi-groupe) si $\mu P_t = \mu$ pour tout $t \geq 0$, autrement dit si μ est une solution stationnaire de l'équation forward ou, dit encore autrement, si

$$X_0 \sim \mu \quad \Rightarrow \quad X_t \sim \mu \quad \forall t \geq 0.$$

En dérivant l'équation $\mu P_t f = \mu f$ (par densité il suffit de considérer des fonctions f régulières), l'invariance de μ est équivalente à $\mu L P_t f = 0$ pour tout f et $t \geq 0$, autrement dit

$$\mu(L f) = 0$$

pour tout $f \in \mathcal{C}^2(\mathbb{R}^d)$.

Par exemple, dans le cas particulier du générateur défini en (1) avec $b(x) = -\nabla U(x)$ pour un $U \in \mathcal{C}^2(\mathbb{R}^d)$, on peut directement vérifier que μ la mesure de probabilité de densité proportionnelle à $\exp(-U)$ (qu'on suppose intégrable) est une solution constante de (2), autrement dit est une mesure invariante du processus associé.

La question centrale du cours est celle-ci : étant donné un processus de Markov, on veut typiquement montrer qu'il existe une unique mesure invariante et qu'elle est attractive au sens où $\nu P_t \rightarrow \mu$ (au sens faible) quand $t \rightarrow +\infty$ pour toute condition initiale ν (ou au moins toute condition initiale dans un sous-ensemble donné de $\mathcal{P}(\mathbb{R}^d)$, par exemple celles qui ont une densité régulière, ou un moment d'ordre 2). Plus précisément, l'objectif sera de quantifier cette convergence, c'est-à-dire d'obtenir des estimées de la forme, typiquement,

$$\forall t \geq 0, \quad d(\nu P_t, \mu) \leq C(\nu) e^{-\lambda t}$$

où d est une certaine distance sur les mesures de probabilités, $\lambda > 0$ est un taux de convergence et $C(\nu)$ une constante dépendant de la condition initiale. L'exemple le plus standard de distance est la variation totale,

$$\|\nu - \nu'\|_{TV} = 2 \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\nu(A) - \nu'(A)| = \sup_{\|f\|_\infty \leq 1} |\nu(f) - \nu'(f)|. \quad (3)$$

On verra d'autres distances (ou pseudo-distances) dans le cours, avec lesquelles il est plus ou moins facile selon les cas de travailler, et qui donnent une convergence en des sens plus ou moins forts.

Par dualité, la question est également de montrer que $P_t f \rightarrow \mu(f)$ pour tout f dans une classe donnée de fonctions (par exemple les fonctions bornées, ou Lipschitz), et l'on cherche alors des bornes du type

$$\forall t \geq 0, \quad d(P_t f, \mu(f)) \leq C(f) e^{-\lambda t}.$$

pour une distance d sur les fonctions.

Dans une formulation EDPiste, la question est donc de montrer que toute solution de l'équation (2) converge en temps long vers l'unique solution stationnaire, et de donner une vitesse de convergence. Parfois, on a au départ un problème probabiliste concernant un processus stochastique et on utilise des outils d'EDP pour montrer la convergence en temps long (cf. par exemple les méthodes d'entropie plus tard dans le cours). Parfois, au contraire, on veut montrer la relaxation à l'équilibre pour une EDP qu'on interprète comme l'équation forward d'un processus de Markov pour exploiter des arguments probabilistes (cf. par exemple les méthodes de couplage). La distinction n'est de toutes façons pas toujours très pertinente.

Encore quelque mot sur la dualité. Si μ est invariante pour $(P_t)_{t \geq 0}$, un cadre naturel est de travailler dans $L^2(\mu)$. Supposons que la loi $\nu_0 P_t$ de X_t (pour une condition initiale ν_0 donnée) admette pour tout $t \geq 0$ une densité h_t par rapport à μ , c'est-à-dire que $\nu P_t f = \mu(h_t f)$. Autrement dit, pour tout f ,

$$\int_{\mathbb{R}^d} h_t(x) f(x) \mu(dx) = \int_{\mathbb{R}^d} P_t f(x) h_0(x) \mu(dx) = \int_{\mathbb{R}^d} f(x) P_t^* h_0(x) \mu(dx)$$

où P_t^* est le dual de P_t mais, cette fois, pour le produit scalaire de $L^2(\mu)$. On a donc $h_t = P_t^* h_0$. On a

$$P_{t+s}^* = (P_t P_s)^* = P_s^* P_t^*$$

autrement dit $(P_t^*)_{t \geq 0}$ est aussi un semi-groupe, qui laisse μ invariant puisque $\mu(P_t^* f) = \mu((P_t \mathbb{1}) f) = \mu(f)$ avec $\mathbb{1}$ la fonction constante égale à 1. D'autre part, pour tout $f \in \mathcal{C}^2(\mathbb{R}^d)$,

$$\begin{aligned} \partial_t \int_{\mathbb{R}^d} P_t^* h(x) f(x) \mu(dx) &= \partial_t \int_{\mathbb{R}^d} h(x) P_t f(x) \mu(dx) = \int_{\mathbb{R}^d} h(x) P_t L f(x) \mu(dx) \\ &= \int_{\mathbb{R}^d} L^* P_t^* h(x) f(x) \mu(dx) \end{aligned}$$

avec L^* l'adjoint dans $L^2(\mu)$ de L . Autrement dit $\partial_t P_t^* h = L^* P_t^* h$, et en utilisant dans le calcul précédent que $\partial_t P_t f = L P_t f$ on obtient également que $\partial_t P_t^* h = P_t^* L^* h$.

Autrement dit, travailler avec $P_t f$ pour des fonctions tests f ou avec $P_t^* h$ pour h des densités relatives à l'équilibre μ revient essentiellement au même. Dans le cours on travaillera principalement avec $P_t f$, ce qui a l'avantage de ne pas demander de connaître l'expression explicite de μ .

D'ailleurs, dans l'exemple (1) avec $b(x) = -\nabla U(x)$ et $\mu \propto \exp(-U)$, en intégrant par partie on constate que

$$\int_{\mathbb{R}^d} g(x) (-\nabla U(x) \cdot \nabla f(x) + \Delta f(x)) e^{-U(x)} dx = \int_{\mathbb{R}^d} f(x) (-\nabla U(x) \cdot \nabla g(x) + \Delta g(x)) e^{-U(x)} dx,$$

et donc $L^* = L$ est auto-adjoint dans $L^2(\mu)$. On dit que le processus est réversible pour la mesure μ , et on a alors $P_t^* = P_t$ pour tout $t \geq 0$.

1.3 Quelques motivations

Obtenir des vitesses de convergence explicites peut avoir plusieurs motivations : quand le processus modélise un système physique (système moléculaire, population, etc...) la vitesse de relaxation à l'équilibre donne des informations sur les temps caractéristiques de certains phénomènes. Des transitions de phase peuvent être observées : par exemple, dans certains modèles de particules en interaction (en physique statistiques ou en dynamique des populations), dans certaines situations, le taux de convergence peut être indépendant du nombre N de particules, ou au contraire peut être équivalent e^{-aN} pour un $a > 0$. Dans le second cas, pour une grande population, la convergence à l'équilibre ne sera en fait jamais observée en pratique. Plus généralement, une vitesse de convergence extrêmement faible est une façon de quantifier la *métastabilité* de certains processus qui restent bloqués pendant des temps très long dans certaines zones (on y reviendra).

Une autre motivation vient des algorithmes stochastiques, tout particulièrement les algorithmes MCMC (Monte Carlo avec Chaîne de Markov). On utilisera cette motivation comme fil conducteur tout au long du cours. L'objectif des méthodes MCMC est d'estimer des intégrales de la forme $\int_{\mathbb{R}^d} f(x) \mu(dx)$ en grande dimension, où f est une fonction test donnée et μ est une mesure de probabilité cible. L'algorithme de Monte-Carlo de base consiste à utiliser la loi des grands nombres

$$\frac{1}{N} \sum_{i=1}^N f(X_i) \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \mu(f)$$

avec $(X_i)_{i \in \mathbb{N}}$ une suite i.i.d. de loi μ . Néanmoins dans de nombreuses situations (notamment quand on connaît μ à constante près, c'est-à-dire sans connaître la constante de normalisation), on ne peut pas générer des variables de loi μ , mais on sait générer un processus de Markov dont la mesure invariante est μ . On dit qu'un processus de Markov $(X_t)_{t \geq 0}$ est ergodique pour la mesure μ si la loi des grands nombres ergodique suivante a lieu :

$$\frac{1}{t} \int_0^t f(X_s) ds \xrightarrow[t \rightarrow +\infty]{\mathbb{P}} \mu(f).$$

Si c'est le cas, la moyenne empirique le long de la trajectoire (c'est-à-dire le terme de gauche) permet d'estimer la quantité d'intérêt. Évidemment en pratique on discrétise en temps l'intégrale et le processus, mais il y a déjà bien des choses à apprendre du processus continu donc on n'abordera pas cette (intéressante) question dans ce cours. Remarquons déjà que

$$\mathbb{E}_x \left(\frac{1}{t} \int_0^t f(X_s) ds \right) = \frac{1}{t} \int_0^t P_t f(x) ds.$$

Supposons qu'on sache que

$$|P_t f(x) - \mu(f)| \leq C(x) e^{-\lambda t} \tag{4}$$

pour la fonction f qui nous intéresse (qu'on va supposer bornée ici), pour certains $\lambda, C(x)$. Alors le biais de l'estimateur peut être borné par

$$\left| \mathbb{E}_x \left(\frac{1}{t} \int_0^t f(X_s) ds \right) - \mu(f) \right| \leq \frac{1}{t} \int_0^t C(x) e^{-\lambda s} ds \leq \frac{C(x)}{\lambda t}.$$

Mieux, en notant $\bar{f} = f - \mu(f)$,

$$\begin{aligned} \mathbb{E}_x \left(\left| \frac{1}{t} \int_0^t f(X_s) ds - \mu(f) \right|^2 \right) &= \frac{1}{t^2} \int_0^t \int_0^t \mathbb{E}_x (\bar{f}(X_s) \bar{f}(X_u)) ds du \\ &= \frac{2}{t^2} \int_0^t \int_0^s \mathbb{E}_x (\bar{f}(X_s) \bar{f}(X_u)) du ds \\ &= \frac{2}{t^2} \int_0^t \int_0^s \mathbb{E}_x (P_{s-u} \bar{f}(X_u) \bar{f}(X_u)) du ds \\ &\leq \frac{4\|f\|_\infty}{t^2} \int_0^t \int_0^s \mathbb{E}_x (C(X_u)) e^{-\lambda(s-u)} du ds. \end{aligned}$$

Supposons qu'on ait également des moments uniformes en temps au sens où

$$\sup_{u \geq 0} \mathbb{E}_x (C(X_u)) = K(x) < +\infty$$

(on verra comment obtenir ce genre d'information). On en arrive à

$$\mathbb{E}_x \left(\left| \frac{1}{t} \int_0^t f(X_s) ds - \mu(f) \right|^2 \right) \leq \frac{4\|f\|_\infty K(x)}{\lambda t}.$$

Ceci implique la convergence en probabilité de l'estimateur vers $\mu(f)$, donne une borne sur le risque quadratique et donc, par l'inégalité de Markov, un intervalle de confiance non-asymptotique de niveau $\varepsilon > 0$ avec

$$\mathbb{P}_x \left(\left| \frac{1}{t} \int_0^t f(X_s) ds - \mu(f) \right| > \delta_t \right) \leq \frac{2\|f\|_\infty K(x)}{\lambda t \delta_t^2} = \varepsilon \quad \text{avec} \quad \delta_t^2 = \frac{2\|f\|_\infty K(x)}{\lambda t \varepsilon}.$$

(intervalle de confiance dont la longueur est d'ordre $1/\sqrt{t}$ quand $t \rightarrow +\infty$).

On voit donc que la convergence en temps long (4), qui n'était a priori qu'une convergence de la marginale en temps (c'est-à-dire la loi au temps t pour un t fixé) permet de récupérer (et quantifier) l'ergodicité, c'est-à-dire la convergence de la moyenne empirique de long de la trajectoire (qui est la quantité accessible en pratique).

De plus, la question de la vitesse de convergence devient crucial quand plusieurs processus différents sont disponibles. Par exemple, pour échantillonner la mesure μ de densité proportionnelle à $\exp(-U)$ pour $U \in \mathcal{C}^2(\mathbb{R}^d)$, on peut utiliser le processus de Langevin suramorti

$$dX_t = -\nabla U(X_t) dt + \sqrt{2} dB_t$$

ou bien le processus de Langevin cinétique $X = (Y, V) \in \mathbb{R}^{2d}$ solution de

$$\begin{cases} dY_t = V_t dt \\ dV_t = -\nabla U(Y_t) dt - \gamma V_t dt + \sqrt{2\gamma} dB_t \end{cases}$$

avec $\gamma > 0$ qu'on appelle le paramètre de friction (ou d'amortissement), encore le processus de Monte Carlo Hamiltonien (HMC) qui satisfait

$$dY_t = V_t dt, \quad dV_t = -\nabla U(Y_t) dt$$

entre des temps de sauts $(T_k)_{k \in \mathbb{N}}$ qui sont soit déterministes (avec $T_{k+1} = T_k + \tau$ avec $\tau > 0$ la période, fixée), soit aléatoire (avec $T_{k+1} = T_k + E_k$ avec E_k une variable de loi exponentielle de paramètre donné $\theta > 0$) et, aux temps de sauts, $V_{T_k} = W_k$ avec $(W_k)_{k \in \mathbb{N}}$ une suite i.i.d. de gaussienne standard d -dimensionnelle (autrement dit, périodiquement ou selon un processus de Poisson, la vitesse est rafraîchie à une nouvelle valeur gaussienne). À quoi on pourrait rajouter un algorithme de Metropolis-Hastings avec une proposition donnée par un noyau gaussien et bien d'autres choses encore.

La question qui se pose en pratique est donc celle-ci : quel processus utiliser ? Un critère de comparaison est de prendre celui qui converge le plus vite à l'équilibre (soit au sens de la moyenne ergodique, soit au sens de la loi au temps t). De plus, parmi les processus ci-dessus, certains dépendent de paramètres (γ, τ, θ) , la variance de la proposition gaussienne dans un algorithme de Metropolis-Hastings), et là encore le but est de maximiser la vitesse de convergence.

En résumé, concernant les méthodes MCMC, obtenir des bornes quantitatives sur la vitesse de convergence des processus de Markov a les motivations suivantes :

- Obtenir des bornes d'erreurs et des intervalles de confiance non-asymptotique pour les estimateurs.
- Comparer différents processus ayant la même mesure invariante.
- Choisir les valeurs des paramètres.

Ce n'est pas une liste exhaustive, par exemple, avoir des estimations de convergence à l'équilibre permet d'obtenir des bornes sur les solutions des équations de Poisson de la forme $Lg = f - \mu(f)$ (d'inconnue g) qui interviennent dans l'étude de l'erreur numérique induite par un schéma de discrétisation (Euler-Maruyama ou autre), la stabilité de la mesure invariante en fonction du drift (méthode de Stein), les expansions de Talay-Tubaro et les interpolations de Romberg, etc.

2 Le cadre formel

Si l'on regarde la section 1.1, on peut voir qu'il y a quelques zones d'ombre. Déjà, est-ce clair que la solution de l'EDS est définie pour tout temps ? Y a-t-il existence et/ou unicité des solutions pour les EDP backward et forward ? Pour quelles conditions initiales, et pour des solutions en quel sens ? Pourquoi a-t-on $P_t f \rightarrow f$ quand $t \rightarrow 0$ (pour quelles f ?), et en quel sens a lieu cette convergence (simple, uniforme ?) ? Dans la suite du cours un certain nombre d'arguments demanderont que telle ou telle fonction ait une régularité suffisante ou soit intégrable pour telle mesure (par exemple pour dériver sous le signe intégrale), etc. Dans le but de se concentrer sur les points principaux du raisonnements et de limiter la technicité (tout en écrivant des choses correctes...) on travaillera toujours dans les conditions que nous allons détailler dans cette section. Dans beaucoup de cas pratiques, ces conditions sont satisfaites (on donne quelques références ci-dessous) ; ou alors on peut approcher le processus par un processus régularisé pour lesquels ces conditions sont satisfaites, appliquer le résultat et obtenir le résultat pour le processus initial en passant à la limite ; ou bien encore reprendre la preuve et l'adapter en vérifiant que tous les objets en jeux sont bien définis et que les calculs sont justifiés, quitte à utiliser là-encore des arguments de troncature ou de densité. Ces discussions ne sont pas l'objet de ce cours.

Voici les conditions qui seront implicitement supposée dans tous le reste du cours.

On s'intéresse à un processus de Markov non-exposif (c'est-à-dire défini pour tout temps) $(X_t)_{t \geq 0}$ sur \mathbb{R}^d . Autrement dit, on se donne un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ (qu'on n'explicitera en fait jamais) et, pour toute condition initiale $x \in \mathbb{R}^d$, une variable aléatoire $X^x : \Omega \rightarrow$

$\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$, où $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ est l'espace de Skorokhod des fonctions càdlàg de \mathbb{R}_+ dans \mathbb{R}^d . On note \mathbb{P}_x la loi de X^x , et on note simplement $X_t = X_t^x$, sans expliciter la condition initiale. Le processus satisfait la propriété de Markov

$$\forall t \geq s \geq 0, \quad \mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s)$$

où $(\mathcal{F}_s)_{s \geq 0}$ est la filtration associée au processus, et également la propriété de Markov forte

$$\mathbb{E}(f(X_{\tau+t})|\mathcal{F}_\tau) = \mathbb{E}(f(X_{\tau+t})|X_\tau)$$

pour tout temps d'arrêt τ . Sauf mention contraire, le processus est homogène en temps, autrement dit

$$\mathbb{E}(f(X_t)|X_s) = \mathbb{E}_{X_s}(f(X_{t-s})) .$$

On suppose que pour tout $T > 0$ il existe $\alpha_T > 0$ tel que pour tout compact $K \subset \mathbb{R}^d$,

$$\sup_{x \in K} \sup_{t \in [0, T]} \mathbb{E}_x \left(e^{\alpha_T |X_t|^2} \right) < \infty$$

On note \mathcal{A} l'ensemble des fonctions \mathcal{C}^∞ dont toutes les dérivées croissent au plus polynomialement à l'infini. Autrement dit, une fonction \mathcal{C}^∞ f est dans \mathcal{A} si pour tout $n \in \mathbb{N}$ et tout multi-indice $\alpha \in \llbracket 1, d \rrbracket^n$, il existe C, k telle que $|\partial_\alpha f(x)| \leq C(1 + |x|^k)$ pour tout $x \in \mathbb{R}^d$. Pour tout $f \in \mathcal{A}$, on note

$$P_t f(x) = \mathbb{E}_x(f(X_t)) .$$

Pour tout $f \in \mathcal{A}$, $(P_t f - f)/t$ converge simplement quand $t \rightarrow 0$ vers une fonction qu'on note Lf . Pour tout $f \in \mathcal{A}$ et $t \geq 0$, $Lf \in \mathcal{A}$, $P_t f \in \mathcal{A}$. De plus, pour tout $T > 0$ et tout $f \in \mathcal{A}$, il existe $C, k > 0$ tels que $|LP_t f(x)| \leq C(1 + |x|^k)$ pour tout $t \in [0, T]$, et de même pour toutes les dérivées de $LP_t f$.

Donnons tout de suite une caractérisation de l'invariance de μ sous ces conditions.

Lemma 1. Soit $\mu \in \mathcal{P}(\mathbb{R}^d)$ telle que $\int_{\mathbb{R}^d} |x|^k \mu(dx) < +\infty$ pour tout $k \in \mathbb{N}$. Alors μ est invariante pour $(P_t)_{t \geq 0}$ si et seulement si $\mu(Lf) = 0$ pour tout $f \in \mathcal{A}$.

Démonstration. Du fait du moment gaussien de P_t , par densité il suffit, pour tout $t \geq 0$, que $\mu P_t f = \mu f$ pour tout $f \in \mathcal{A}$ pour que cela soit vrai pour toute fonction f continue bornée, et donc pour que μ soit invariante pour $(P_t)_{t \geq 0}$. Par l'hypothèse sur μ et les bornes polynomiales uniformes en temps de $LP_t f$, la dérivation $\partial_t \mu P_t f = \mu LP_t f$ est justifiée si $f \in \mathcal{A}$, de sorte que μ est invariante si et seulement si $\mu Lf = 0$ pour tout $f \in \mathcal{A}$. \square

3 Exemples de processus

3.1 Processus de diffusion

Les diffusions sont les processus de Markov dont les trajectoires sont continues, et sont la plupart du temps obtenues comme solution d'équations différentielles stochastiques dirigées par des mouvements browniens,

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t .$$

avec, par exemple, $b \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ et $\sigma \in \mathcal{C}^1(\mathbb{R}^d, \mathcal{M}_{d,d}(\mathbb{R}))$ et $(B_t)_{t \geq 0}$ un mouvement brownien d -dimensionnel. Le générateur correspondant est

$$Lf(x) = b(x) \cdot \nabla f(x) + \frac{1}{2} \sum_{i,j=1}^d \Sigma_{i,j}(x) \partial_{x_i} \partial_{x_j} f(x)$$

avec $\Sigma(x) = \sigma(x)^T \sigma(x)$. On renvoie à [2] pour plus de détail sur ces processus, l'existence et l'unicité des solutions de l'EDS, le générateur infinitésimal, l'existence d'une densité régulière pour le noyau de transition, etc.

Lorsque la matrice Σ est définie positive, L est dit elliptique, l'EDP satisfaite par $P_t f$ (ou par la loi du processus) est une EDP parabolique et de nombreux résultats (régularité, bornes gaussiennes...) sont disponibles, cf [2, Chapitre 3].

Quelques cas particuliers d'intérêts :

Processus de Langevin suramorti

Parfois aussi appelé processus de Fokker-Planck, c'est la solution de

$$dX_t = -\nabla U(X_t)dt + \sqrt{2\beta^{-1}}dB_t \quad (5)$$

où, par rapport à l'introduction, on a ajouté un paramètre $\beta > 0$ qu'on appelle l'inverse de la température. L'unique mesure invariante est la mesure de Gibbs d'énergie U et de température inverse β , c'est-à-dire la mesure de probabilité sur \mathbb{R}^d de densité proportionnelle à $\exp(-\beta U)$ (en supposant que $\exp(-\beta U)$ est intégrable).

D'un point de vue théorique, ce processus est plutôt sympathique puisqu'il est réversible et son générateur est elliptique.

Processus de Langevin (cinétique)

C'est le processus $X = (Y, V)$ sur \mathbb{R}^{2d} solution de

$$\begin{cases} dY_t = V_t dt \\ dV_t = -\nabla U(Y_t)dt - \gamma V_t dt + \sqrt{2\gamma\beta^{-1}}dB_t \end{cases} \quad (6)$$

où $\gamma > 0$ est le paramètre dit de friction. On appelle Y_t la position et V_t la vitesse. Le générateur est donné par

$$L = v\nabla_y - (\nabla U(y) + \gamma v)\nabla_v + \frac{\gamma}{\beta}\Delta_v$$

qui n'est pas elliptique puisque la matrice de diffusion est

$$\Sigma = \begin{pmatrix} 0 & 0 \\ 0 & I_d \end{pmatrix}$$

qui est dégénérée (il n'y a du bruit directement que sur la vitesse). Cependant on peut montrer que le processus est hypoelliptique au sens où, partant d'une loi initiale donnée par une masse de Dirac, la loi du processus admet une densité régulière et strictement positive pour tout temps $t > 0$, cf [1, Section 2.3].

Le lien avec le processus précédent vient du fait que, quand $\gamma \rightarrow +\infty$, on peut montrer que $(Y_{\gamma t})_{t \geq 0}$ converge vers le processus de Langevin suramorti. En effet, (6) donne

$$\begin{aligned} Y_{\gamma t} - Y_0 &= \int_0^{\gamma t} V_s ds = \frac{1}{\gamma} \int_0^{\gamma t} \left(-\nabla U(Y_s) ds - dV_s + \sqrt{2\gamma\beta^{-1}}dB_s \right) \\ &= -\int_0^{\gamma t} \nabla U(Y_{\gamma s}) ds + \frac{V_{\gamma t} - V_0}{\gamma} + \sqrt{2\beta^{-1}}B_t^\gamma \end{aligned}$$

avec $(B_t^\gamma)_{t \geq 0} = (B_{\gamma t}/\sqrt{\gamma})_{t \geq 0}$ qui est un mouvement brownien standard quelque soit γ . En considérant \tilde{X}^γ le processus de Langevin suramorti solution de

$$\tilde{X}_t^\gamma - Y_0 = -\int_0^t \nabla U(X_s^\gamma) ds + \sqrt{2\beta^{-1}}B_t^\gamma,$$

en supposant que ∇U est L -Lipschitz et que $C = \sup_{s \geq 0} \mathbb{E}|V_s| < \infty$, on obtient

$$\mathbb{E}(|X_t^\gamma - Y_{\gamma t}|) \leq L \int_0^t \mathbb{E}(|X_s^\gamma - Y_{\gamma s}|) ds + \frac{2C}{\gamma}$$

et par le lemme de Grönwall, pour tout $t \geq 0$

$$\mathbb{E}(|X_t^\gamma - Y_{\gamma t}|) \leq \frac{2Ce^{Lt}}{\gamma} \xrightarrow{\gamma \rightarrow +\infty} 0.$$

Processus d'Ornstein-Uhlenbeck (généralisés)

Étant donnés des matrices $A, B \in \mathcal{M}_{d,d}(\mathbb{R})$, le processus d'Ornstein-Uhlenbeck correspondant est la solution de

$$dX_t = AX_t dt + B dB_t,$$

le cas classique étant $A = -I_d$ et $B = I_d$. La particularité de ces processus est qu'ils sont gaussiens (si la condition initiale est gaussienne), autrement dit pour tout $0 \leq t_0 < \dots < t_n$ le vecteur $(X_{t_0}, \dots, X_{t_n})$ est un vecteur gaussien. En particulier la loi de X_t est caractérisée par sa moyenne $m_t \in \mathbb{R}^d$ et sa matrice de covariance Σ_t , qui sont les solutions d'équations linéaires

$$\partial_t m_t = A m_t, \quad \partial_t \Sigma_t = A \Sigma_t + \Sigma_t A + B^T B.$$

3.2 Processus déterministes par morceaux

[à compléter]

4 Fonctions de Lyapunov et conditions de couplage

4.1 Fonction de Lyapunov

Considérons la condition dite de Lyapunov suivante :

Hypothèse 1. *Il existe $\rho, C > 0$ et une fonction $V : \mathbb{R}^d \rightarrow [1, +\infty[$ avec $V \rightarrow +\infty$ à l'infini tels que*

$$\forall x \in \mathbb{R}^d, \quad LV(x) \leq -\rho V(x) + C.$$

De plus, $\partial_t P_t V = P_t LV$ pour tout $t \geq 0$

On appelle V une fonction de Lyapunov.

Exemple 1. *Considérons le générateur $L = b\nabla + \Delta$ avec $b \in C^1(\mathbb{R}^d)$. Supposons qu'il existe $R, \lambda > 0$ tel que $b(x) \cdot x \leq -\lambda|x|^2$ pour tout $x \in \mathbb{R}^d$ avec $|x| \geq R$. Prenons $V(x) = |x|^2$. Alors*

$$LV(x) = 2x \cdot b(x) + d \leq -2\lambda V(x) + 2\lambda R^2 + 2 \sup\{y \cdot b(y), |y| \leq R\} + d.$$

De façon alternative, prenons $V(x) = e^{\alpha|x|^2}$ pour $\alpha > 0$. Alors

$$LV(x) = (2\alpha x \cdot b(x) + 4\alpha^2|x|^2 + 2\alpha d) V(x).$$

Pour $\alpha < \lambda/2$, pour $|x|$ assez grand,

$$2\alpha x \cdot b(x) + 4\alpha^2|x|^2 + 2\alpha d \leq -2\alpha(\lambda - \alpha)|x|^2 + 2\alpha d \leq -1,$$

et on en déduit comme avant que $LV(x) \leq -V(x) + C$ pour tout $x \in \mathbb{R}^d$ pour un certain $C > 0$.

Exemple 2. *Considérons le générateur $Lf(x) = f'(x) + x(f(x/2) - f(x))$ du TCP sur \mathbb{R}_+ . À nouveau, prenons $V(x) = e^{\alpha x^2}$ pour $\alpha > 0$. Alors*

$$LV(x) = \left(2\alpha - 1 + e^{-3\alpha x^2/4}\right) xV(x).$$

Pour $\alpha < 1/2$, on a bien $LV \leq -V + C$ pour un $C > 0$.

Lemma 2. *Sous l'hypothèse 1, pour tout $t \geq 0$ et tout $x \in \mathbb{R}^d$,*

$$P_t V(x) = \mathbb{E}_x(V(X_t)) \leq e^{-\rho t} V(x) + (1 - e^{-\rho t}) \frac{C}{\rho}$$

Démonstration. On utilise simplement que

$$\partial_t (e^{\rho t} P_t V(x)) = e^{\rho t} (P_t LV(x) + \rho P_t V(x)) \leq C e^{\rho t}.$$

□

En particulier, on obtient des estimés uniformes en temps sur les moments de X_t , à savoir que, pour toute condition initiale intégrant V ,

$$\sup_{t \geq 0} \mathbb{E}(V(X_t)) \leq \mathbb{E}(V(X_0)) + \frac{C}{\rho}.$$

4.2 Condition de couplage locale

On rappelle la définition (3) de la distance en variation totale entre deux mesures de probabilités. Une autre caractérisation est

$$\|\nu - \nu'\|_{TV} = 2 \inf_{\pi \in \mathcal{C}(\nu, \nu')} \mathbb{P}_\pi(X \neq Y),$$

où $\mathcal{C}(\nu, \nu')$ est l'ensemble des mesures de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ dont les marginales sont ν et ν' , autrement dit qui sont telles que, si $(X, Y) \sim \pi$, alors $X \sim \nu$ et $Y \sim \nu'$. On dit que π (ou que (X, Y)) est un couplage de ν et ν' .

Cette formulation a ceci de pratique qu'elle permet de majorer facilement la variation totale en construisant un couplage. Dès qu'on dispose d'une variable (X, Y) avec $X \sim \nu$ et $Y \sim \nu'$, on obtient

$$\|\nu - \nu'\|_{TV} \leq 2\mathbb{P}(X \neq y).$$

On considère la condition suivante, dite de couplage locale :

Hypothèse 2. *Pour tout compact \mathcal{K} de \mathbb{R}^d , il existe $\alpha, t_* > 0$ tel que pour tout $x, y \in \mathcal{K}$,*

$$\|\delta_x P_{t_*} - \delta_y P_{t_*}\|_{TV} \leq 2(1 - \alpha). \quad (7)$$

Autrement dit, vue la discussion ci-dessus, sous cette condition, on peut coupler deux processus partant de x et y dans un compact de sorte qu'il soit égaux au temps t_* avec une probabilité au moins α (uniforme sur le compact).

Une façon classique d'obtenir l'hypothèse (2) est de passer par une condition dite de Doeblin (locale) :

Hypothèse 3. *Pour tout compact \mathcal{K} de \mathbb{R}^d , il existe $\alpha, t_* > 0$ et une mesure de probabilité ν sur \mathbb{R}^d tels que pour tout $x \in \mathcal{K}$,*

$$\delta_x P_{t_*} \geq \alpha \nu,$$

c'est-à-dire que $\mathbb{P}_x(X_{t_} \in A) \geq \alpha \nu(A)$ pour tout $A \in \mathcal{B}(\mathbb{R}^d)$.*

Lemma 3. *L'hypothèse 3 implique l'hypothèse 2.*

Démonstration. Soit \mathcal{K} un compact de \mathbb{R}^d . Soient α, t_*, ν donnés par l'hypothèse 3, et soient $x, y \in \mathbb{R}^d$. On peut décomposer

$$\delta_x P_{t_*} = \alpha \nu + (1 - \alpha) \frac{\delta_x P_{t_*} - \alpha \nu}{1 - \alpha} := \alpha \nu + (1 - \alpha) Q_x$$

avec Q_x qui est une mesure de probabilité. On décompose de même $\delta_y P_{t_*} = \alpha \nu + (1 - \alpha) Q_y$ avec Q_y une mesure de probabilité. Soient U, V_x, V_y, W des variables aléatoires indépendantes de lois respectives ν, Q_x, Q_y et uniforme sur $[0, 1]$. Posons

$$X = \mathbb{1}_{W \leq \alpha} U + \mathbb{1}_{W > \alpha} V_x, \quad Y = \mathbb{1}_{W \leq \alpha} U + \mathbb{1}_{W > \alpha} V_y.$$

Alors, pour tout f mesurable borné,

$$\mathbb{E}(f(X)) = \alpha \mathbb{E}(f(U)) + (1 - \alpha) \mathbb{E}(f(V_x)) = \alpha \nu(f) + (1 - \alpha) Q_x f = P_{t_*} f(x),$$

autrement dit X est de loi $\delta_x P_{t_*}$, et de même Y est de loi $\delta_y P_{t_*}$. On a donc

$$\|\delta_x P_{t_*} - \delta_y P_{t_*}\|_{TV} \leq 2\mathbb{P}(X \neq Y) \leq 2\mathbb{P}(W > \alpha) = 2(1 - \alpha).$$

□

Exemple 3. *Considérons la diffusion de générateur $L = b \cdot \nabla + \Delta$ avec $b \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ Lipschitz. Alors le processus admet une densité de transition continue et strictement positive pour tout temps strictement positif, cf. [2], autrement dit pour tout $t > 0$ il existe une fonction $x, y \mapsto p_t(x, y)$ continue et strictement positive telle que*

$$\mathbb{E}_x(f(X_t)) = \int_{\mathbb{R}^d} f(y) p_t(x, y) dy.$$

Étant donné un compact \mathcal{K} , on a donc $\alpha' := \inf_{\mathcal{K} \times \mathcal{K}} p_t > 0$, et on peut donc borner pour $x \in \mathcal{K}$ et $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}_x(X_t \in A) = \int_{\mathbb{R}^d} \mathbb{1}_A(y) p_t(x, y) dy \geq \alpha' \int_{\mathbb{R}^d} \mathbb{1}_{A \cap \mathcal{K}}(y) dy = \alpha \text{Leb}(\mathcal{K}) \nu(A)$$

avec ν la mesure uniforme sur \mathcal{K} .

Exemple 4. *Considérons le générateur $Lf(x) = f'(x) + \lambda \left(\int_0^\infty f(u) e^{-u} du - f(x) \right)$ sur \mathbb{R}_+ , avec un taux de saut $\lambda > 0$ constant. Étant donnée deux conditions initiales x, y , on considère le processus de Markov $(X_t, Y_t)_{t \geq 0}$ associé au générateur sur \mathbb{R}_+^2*

$$\mathcal{L}f(x, y) = \partial_x f(x, y) + \partial_y f(x, y) + \lambda \left(\int_0^\infty f(u, u) e^{-u} du - f(x, y) \right)$$

avec $(X_0, Y_0) = (x, y)$. Autrement dit, on a $dX_t = dt$ et $dY_t = dt$ et, à des temps aléatoires T_k suivant un processus de Poisson d'intensité, on pose $X_{T_k} = Y_{T_k} = V_k$ avec $(V_k)_{k \in \mathbb{N}}$ une suite i.i.d. de variables de loi exponentielle. En particulier $(X_t)_{t \geq 0}$ et $(Y_t)_{t \geq 0}$ sont bien des processus associés à L , et

$$\mathbb{P}(X_t \neq Y_t) \leq \mathbb{P}(T_1 > t) = e^{-\lambda t}.$$

Ainsi, la condition de couplage est globale (c'est-à-dire qu'elle n'est pas restreinte à des compacts \mathcal{K}) : pour tout $x, y \in \mathbb{R}_+$ et $t \geq 0$,

$$\|\delta_x P_t - \delta_y P_t\|_{TV} \leq 2e^{-\lambda t}.$$

Références

- [1] D. Nualart. *The Malliavin Calculus and Related Topics*. Springer-Verlag Berlin/Heidelberg, 2006.
- [2] D. W. Stroock and S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag Berlin Heidelberg, 2006.