THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *28 Juin 2017* **par :**

CLAIRE DELPLANCKE

Méthodes quantitatives pour l'étude asymptotique de processus de Markov homogènes et non-homogènes

---

JURY

BERNARD BERCU	Professeur d'Université	Président du jury
LAURENT DECREUSEFOND	Professeur d'Université	Rapporteur
GERSENDE FORT	Directrice de recherche	Membre du jury
ARNAUD GUILLIN	Professeur d'Université	Rapporteur
ALDÉRIC JOULIN	Maître de Conférences	Directeur de thèse
LAURENT MICLO	Directeur de recherche	Directeur de thèse
CLÉMENTINE PRIEUR	Professeur d'Université	Membre du jury
ANTHONY RÉVEILLAC	Professeur d'Université	Membre du jury

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse
Directeur(s) de Thèse :
Aldéric Joulin et *Laurent Miclo*
Rapporteurs :
Laurent Decreusefond et *Arnaud Guillin*

Remerciements

Je voudrais tout d'abord remercier mes directeurs de thèse, Aldéric pour le temps et l'attention qu'il m'a consacré, dans le domaine strictement mathématique et, au-delà, au sujet de mon avenir professionnel, et Laurent qui m'a initiée à de très belles mathématiques et m'a appris à y naviguer pendant ces quatre années.

J'adresse mes remerciements à mes rapporteurs Arnaud et Laurent, pour le temps et l'énergie qu'ils ont consacrés à la lecture de ce manuscrit et leurs suggestions constructives, ainsi qu'à Anthony, Bernard, Clémentine et Gersende qui ont accepté de faire partie du jury. Je remercie aussi Laurent de son invitation à Paris et de nos discussions mathématiques, et Anthony de m'avoir éclairée sur le calcul de Malliavin lors de plusieurs discussions.

Mes pensées vont à mes collaborateurs Bertrand et Sébastien. Merci Bertrand pour ces séances de travail dans la bonne humeur et les bières post-effort, et merci à toi et Sanae pour votre accueil chaleureux à Montpellier. Merci Sébastien pour ta motivation, pour m'avoir fait découvrir un domaine plus statistique que celui qui m'était habituel, et pour ton écoute et tes encouragements.

Merci à Laure Coutin, qui m'a prêté une oreille attentive et a su me conseiller lors des passages de doute qui semblent émailler toute thèse, et pour être quelqu'un dont j'admire l'énergie et la persévérance. Merci à Jean-Michel Roquejoffre et les membres de l'École Doctorale MITT qui ont suivi le déroulement de cette thèse et ont su apporter des suggestions constructives.

Je remercie Agnès Requis et Martine Labruyère, de l'Ecole Doctorale MITT, et Françoise Michel de l'IMT, pour leur constante efficacité qui a permis de fluidifier bon nombre de démarches administratives.

Ces quatre années de thèse ont aussi été l'occasion pour moi de me frotter à l'enseignement, et je remercie mes collègues des équipes enseignantes de l'INSA et de l'UT 1, avec lesquels cela a été un plaisir de travailler.

Je tiens également à remercier les personnes qui ont joué un rôle crucial dans ma formation : je remercie M. Bayard de m'avoir donné à découvrir ce que sont les mathématiques, dans toute leur rigueur et leur beauté, et m'avoir exhortée à être au niveau de l'exigence intellectuelle qu'elles appellent. Je remercie M. Horvilleur, dont les enseignements m'ont ouvert une vision plus haute du monde et de moi-même, et dont ma mémoire conserve le souvenir de la rigueur, de l'humilité, du sens aigü de l'analyse et de l'humour. Enfin, je remercie Frédéric Pascal pour son investissement dans la formation des étudiants de l'ENS Cachan.

Je garderai un souvenir vivace de la bonne ambiance et du soutien mutuel partagés entre co-doctorants, sans lesquels ces quatre années n'auraient pas le même visage. Merci à mes co-bureaux Ioana et Kevin avec qui j'ai partagé l'essentiel de cette aventure ; merci à Ioana pour les pauses clope et les pauses non-clope, et à Kevin pour avoir sans relâche tenté d'égayer mes fonds d'écran. Merci Anton pour ta présence attentive, Adil pour l'intérêt que tu portes au monde, Malika pour ta chaleur et ton invitation en Aveyron, Claire B. pour ton caractère franc et entier, tes invitations dans le Tarn et à Nantes, et tes conseils et tes réassurances qui m'ont énormément aidée à franchir la dernière ligne droite. Merci à Benoît pour ton engagement et tes t-shirts à motif, à mon grand frère de thèse Pierre aux lunettes rondes et à la logique carrée, et à Nil qui montre tant d'enthousiasme pour les imitations de bruits d'animaux. Merci à Fanny pour toutes ces conversations sur le féminisme, pour avoir co-organisé le premier café femmes de l'IMT, et m'avoir remonté le moral quand c'était dur. Antoine, ce sont bien tes généreux dons de café quand j'étais démunie qui m'ont permis d'achever cette thèse dans les temps. Merci Sylvain et William pour ces super soirées, et merci à Tristan pour avoir partagé les retours tardifs du labo et les conversations politiques, et m'avoir aidé à débugger mon tex quand je paniquais ! Et merci à tous ceux avec qui j'ai partagé les repas du midi et de bons moments, mon co-bureau Pierre, Mélanie, Claire B., Tatiana, Laurent D., Laure, Guillaume, Hugo et son poisson rouge- paix à son âme. Merci à tous ceux qui ont commencé leur thèse quelques annnées après moi et contribuent à l'ambiance chaleureuse du couloir du second étage du 1R3.

Pendant ces quelques années, l'aérien m'a servi d'exutoire, m'a changé les idées, et m'a ouvert les portes d'un autre monde. Merci Marion, Appo, Amandine, Fred, Lydiana, Isa, Marianne, Cindy, Anaïs, Pia, Morgane, César, Pamela et tous les aériens pour l'énergie, l'humour, l'écoute et la poésie que nous avons partagés.

J'aimerais saluer le soutien que m'ont apporté mes amis, et je remercie en particulier Benjamin pour nos soirées et nos discussions, Romain C. pour toutes ces danses, et Romain L. pour être là depuis que nous sommes enfants. À celles avec qui j'ai partagé tant de moments heureux depuis dix ans, qui ont toujours cru en moi, qui ont ri avec moi dans les moments de joie et m'ont soutenue sans faillir dans les moments de doute, j'ai nommé Harry et Alexandra : bêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêêê-rci du fond de mon cœur de biquette aimante !

Enfin, merci à ma famille de Corse pour leur aide spirituelle et pour me rappeler où sont mes racines. Au plus près de mon cœur, que mon père, ma mère, ma grand-mère Jacqueline, et mes grands-parents André, Angèle et Nini, reçoivent le témoignage de tout mon amour, de ma tendresse et ma reconnaissance, pour leur amour et leur soutien indéfectibles, pour m'avoir transmis leur force et leur bonté, et pour m'avoir donné le monde en héritage.

Table des matières

Chapitre 1

Introduction générale

Cette thèse est consacrée à l'étude des propriétés analytiques et asymptotiques des processus de Markov, et à leurs applications. La première partie de la thèse porte sur l'étude asymptotique de processus de Markov inhomogènes par le biais d'inégalités fonctionelles de type Poincaré. Deux situations sont examinées : la première concerne l'obtention de bornes à la Berry-Esseen pour le théorème central limite, et la deuxième se rapporte à l'étude asymptotique d'un processus de saut faiblement mélangeant. Dans la seconde partie, indépendante de la première, des égalités fonctionnelles entre semi-groupes markoviens homogènes connues sous le nom d'entrelacement sont établies, et utilisées pour établir des majorations de la solution de l'équation de Poisson cruciales dans la méthode de Stein.

Première partie : étude asymptotique des processus de Markov inhomogènes

Un processus de Markov inhomogène en temps est un processus stochastique $(X_t)_{t\in I}$ à valeurs dans un espace d'état polonais E, à temps discret si $I = \mathbb{N}$ ou à temps continu si $I = [0, +\infty[$, vérifiant la propriété de Markov : pour tout événement A mesurable de E,

$$\mathbb{P}(X_t \in A \,|\, \sigma(X_r,\, 0 \leq r \leq s)) = \mathbb{P}(X_t \in A | X_s), \qquad t \geq s, \qquad t, s \in I.$$

Le processus $(X_t)_{t\in I}$ est dit homogène si la probabilité ci-dessus ne dépend que de l'événement A et de l'écart $t - s$, et inhomogène dans le cas contraire.

Les processus de Markov homogènes forment un objet central de la théorie des probabilités développée depuis le début du siècle dernier. Un point focal est la question du comportement asymptotique de ces processus, pour laquelle de nombreuses méthodes d'étude existent, notamment les méthodes de martingale et le calcul stochastique, les méthodes spectrales et les inégalités fonctionnelles, les méthodes à la Meyn-Tweedie, les méthodes de couplage, la méthode de Stein, le calcul de Malliavin.

A l'inverse, les processus de Markov inhomogènes en temps n'ont pas fait l'objet d'une attention aussi approfondie. L'inhomogénéité peut traduire différents phénomènes. Le premier se rapporte à la variation d'environnement : un exemple évident est le changement de saison, par exemple si le processus modélise la croissance d'une espèce. L'inhomogénéité sera alors périodique. Dans le même esprit, en finance l'évolution d'un titre est modélisée par un processus de diffusion ; l'inhomogénéité des coefficients de diffusion et de dérive représente alors les variations du marché. L'inhomogénéité apparaît également lors de l'étude de systèmes de particules en interaction (Del Moral and Guionnet [2001]). Une étude approfondie de l'asymptotique des chaînes de Markov inhomogènes à espace d'état fini via l'analyse spectrale a été menée dans une récente série de travaux par Saloff-Coste et Zúñiga (par exemple, Saloff-Coste and Zúñiga [2007]; Saloff-Coste and Zúñiga [2011]).

D'autre part, une large classe de processus de Markov inhomogènes est fournie par les algorithmes stochastiques : par exemple, algorithmes à la Robbins et Monro (Robbins and Monro

[1951]), algorithmes de recuit simulé, méthodes de Monte-Carlo markoviennes. Citons Douc et al. [2004] pour une approche par couplage permettant de quantifier la convergence de chaînes de Markov inhomogènes, appliquée à un algorithme de recuit simulé. Dans le cadre de cette thèse, on s'intéresse à des processus tels que la convergence suivante soit vérifiée dans un certain sens :

$$L_t \underset{t\to+\infty}{\to} L, \tag{1.1}$$

où $(L_t)_{t\geq 0}$ désigne la suite des générateurs instantanés qui dirigent le processus inhomogène et L représente le générateur d'un processus réversible pour une mesure de probabilité μ. Dans ce cadre, l'inhomogénéité disparaît à la limite en temps grand et il faut voir le processus inhomogène comme une *approximation* d'une dynamique homogène.

Dans les deux cas examinés dans la thèse, la propriété (1.1) est vérifiée, avec L le générateur du processus d'Ornstein-Uhlenbeck, défini pour toute fonction assez régulière comme

$$L[f](z) = f''(z) - zf'(z), \qquad z \in \mathbb{R},$$

dont l'unique mesure invariante, et réversible, est la mesure gaussienne centrée réduite μ.

L'équation (1.1) traduit alors la normalité asymptotique des processus considérés, comme on le voit par l'approche classique tension/identification. Cette approche permet seulement d'obtenir la convergence en loi. Pour quantifier cette convergence dans le cas de chaînes de Markov inhomogènes correspondant à des algorithmes stochastiques à la Robbins-Monro, une approche analytique est proposée dans Benaïm et al. [2016], qui utilise la notion de pseudo-trajectoire asymptotique introduite par Benaïm (Benaïm [1999]). La propriété (1.1) est utilisée pour obtenir la vitesse de convergence de ces algorithmes, via une distance ad hoc définie sur une classe de fonctions qui doivent obéir à des conditions de régularité assez fortes.

Dans cette thèse, je propose une méthode pour obtenir des résultats quantitatifs sur la convergence asymptotique reposant sur les inégalités fonctionnelles, qui reprend et étend des idées de Fill [1991] et Arnaudon and Miclo [2016]. La propriété (1.1) n'est pas directement utilisée, mais fait figure d'heuristique pour la stratégie : le couple (L, μ), formé du générateur d'Ornstein-Uhlenbeck et de la mesure gaussienne, satisfait à une inégalité de Poincaré, qui permet de contrôler la vitesse de convergence du processus homogène. L'idée est d'obtenir des analogues de l'inégalité de Poincaré pour le couple (L_t, μ) à t fixé, et contrôler ainsi la vitesse de convergence du processus inhomogène pour des distances telles que la distance du χ_2 ou de Wasserstein.

Le premier processus considéré dans la thèse apparaît dans le cadre du théorème central limite. Pour une suite de variables aléatoires $(X_i)_{i\in\mathbb{N}^*}$ indépendantes identiquement distribuées et centrées réduites, considérons la suite des sommes renormalisées

$$Y_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i, \qquad n \in \mathbb{N}^*.$$

Alors,

$$Y_{n+1} = \sqrt{1 - \frac{1}{n+1}}Y_n + \frac{1}{\sqrt{n+1}}X_{n+1}, \qquad n \in \mathbb{N}^*,$$

faisant de $(Y_n)_{n\in\mathbb{N}^*}$ une chaîne de Markov inhomogène. Cette observation est la base de la méthode utilisée pour quantifier la convergence de la la suite $(Y_n)_{n\in\mathbb{N}^*}$ vers la mesure gaussienne. La stratégie présentée ci-dessus permet alors d'établir une borne à la Berry-Esseen pour la distance du χ_2.

Le second processus considéré est un processus de saut à temps continu $(Z_t)_{t\geq 0}$, de générateur instantané défini comme

$$L_t[f](z) = p_t(z)(f(z + a_t) - f(z)) + q_t(z)(f(z - a_t) - f(z)), \qquad z \in \mathbb{R}, \quad t \geq 0. \tag{1.2}$$

L'équation (1.2) se lit de la manière suivante : si le processus est dans l'état z au temps $t \geq 0$, il saute en $z + a_t$ avec une intensité $p_t(z)$ ou en $z - a_t$ avec une intensité $q_t(z)$. Pour des choix particuliers des intensités de saut, ce processus correspond à une version continue d'un algorithme stochastique de recherche de médiane à la Robbins-Monro. Cet algorithme est fortement consistant et normalement asymptotique, mais peu de résultats existent quant à la quantification de la convergence vers la mesure gaussienne μ, ce qui motive l'étude asymptotique de la version continue $(Z_t)_{t\geq 0}$. La stratégie développée dans la thèse donne lieu à une conjecture sur la décroissance de la distance de Wasserstein entre les marginales temporelles du processus et μ.

Deuxième partie : entrelacements et méthode de Stein

Cette partie s'articule entre deux thèmes de recherche : les entrelacements entre gradients et processus de Markov d'une part, la méthode de Stein d'autre part, dans le cadre discret.

Au premier ordre, la formule d'entrelacement entre un semi-groupe de Markov homogène $(P_t)_{t\geq 0}$ et un opérateur de gradient D sur l'espace d'état polonais E s'écrit :

$$DP_t = \widetilde{P}_t^{\widetilde{V}} D, \qquad t \geq 0, \tag{1.3}$$

où $(\widetilde{P}_t^{\widetilde{V}})_{t\geq 0}$ représente un semi-groupe de Feynman-Kac, formé d'un semi-groupe de Markov alternatif $(\widetilde{P}_t)_{t\geq 0}$ et d'une fonction minorée $\widetilde{V} : E \to \mathbb{R}$, appelée potentiel. Notons $(\widetilde{X}_t^x)_{t\geq 0}$ le processus markovien associé à $(\widetilde{P}_t)_{t\geq 0}$ avec pour valeur initiale $x \in E$. Pour des fonctions suffisamment régulières, le semi-groupe de Feynman-Kac admet la représentation suivante :

$$\widetilde{P}_t^{\widetilde{V}} f(x) = \mathbb{E}\left[f(\widetilde{X}_t^x) e^{-\int_0^t \widetilde{V}(\widetilde{X}_s^x) ds} \right], \qquad x \in E, \quad t \geq 0.$$

Considérons l'exemple d'une diffusion de générateur

$$Lf(x) = \sigma^2 f''(x) + b(x) f'(x)$$

sur $\mathbb{R}$, réversible pour la mesure π. La forme de Dirichlet correspondante s'écrit

$$\mathcal{E}_\pi(f,f) := \int \Gamma(f,f) d\mu = \sigma^2 \int (\nabla f)^2 d\pi, \qquad f \in \mathcal{D}(\mathcal{E}_\pi).$$

Ceci pousse à choisir pour l'opérateur D, plutôt que le gradient lui-même, le gradient à poids $D = \sigma\nabla$. La relation (1.3) entraîne alors pour tout $t \geq 0$:

$$\mathcal{E}_\pi(P_t f, P_t f) = \int (DP_t f)^2 d\pi \leq e^{-2ct} \int (\widetilde{P}_t(Df))^2 d\pi \leq e^{-2ct} \mathcal{E}_\pi(f,f)\,; \qquad c := \inf \widetilde{V},$$

résultat obtenu habituellement comme conséquence du critère de Bakry-Émery (Bakry and Émery [1985]). En fait, la relation (1.3) implique que $\Gamma_2(f,f) \geq c\Gamma(f,f)$ (Bonnefont and Joulin [2014]) : on peut ainsi penser à l'entrelacement (1.3) comme à un raffinement du critère de Bakry-Émery.

L'établissement systématique d'entrelacements de type (1.3) est dû dans le cas discret à Chafaï et Joulin (Chafaï and Joulin [2013]) pour les procesus de naissance-mort, et dans le cas continu à Bonnefont et Joulin (Bonnefont and Joulin [2014]) pour les diffusions réelles. Dans le cas discret, l'opérateur D correspond à un opérateur aux différences de la forme

$$D_u f(x) = \frac{1}{u(x)}(f(x+1) - f(x)), \qquad x \in \mathbb{N},$$

où $(u(x))_{x\in\mathbb{N}}$ est une suite strictement positive ; dans le cas continu, c'est un gradient à poids de la forme $D_u = u\nabla$, où u est une fonction strictement positive de $\mathbb{R}$ dans $\mathbb{R}$. L'équation (1.3) s'écrit alors plus précisément

$$D_u P_t = P_{u,t}^{V_u} D_u, \qquad t \geq 0. \tag{1.4}$$

L'introduction du poids u est cruciale, car elle permet de relier le potentiel apparaissant dans (1.4) à l'étude spectrale du processus considéré : si $\lambda_1(L,\pi)$ désigne le trou spectral de l'opérateur $(-L)$ dans $L^2(\pi)$, où L est réversible par rapport à la mesure de probabilité π, alors sous les hypothèses adéquates,

$$\lambda_1(L,\pi) \geq \sup_{u:E\to]0,+\infty[} \inf_{x\in\mathbb{R}} V_u(x).$$

Cette propriété est due à Chen et Wang pour les diffusions (Chen and Wang [1997]) et à Chen pour les processus de naissance-mort (Chen [1996]), grâce à des méthodes de couplage. L'entrelacement (1.4) en donne une nouvelle preuve.

La contribution aux formules d'entrelacements présentée dans cette thèse réside dans l'établissement d'une formule au second ordre,

$$\Delta P_t = \widehat{P}_t^{\widehat{V}} \Delta, \qquad t \geq 0. \tag{1.5}$$

Le semi-groupe $(P_t)_{t\geq 0}$ correspond à un processus de naissance-mort sur $\mathbb{N}$, et l'opérateur Δ désigne un laplacien discret à poids. L'application principale concerne les facteurs de Stein.

La méthode de Stein (Stein [1972]), ou méthode de Stein-Chen dans le cadre discret (Chen [1975]), désigne un ensemble de techniques visant à évaluer la distance entre deux mesures de probabilité sur l'espace E. Elle implique de désymétriser les rôles joués par chacune des mesures, dont l'une, notée ici π et appelée approximande, fait figure de référence. Une étape cruciale de la méthode consiste alors à évaluer des quantités associées à l'approximande π et à la distance utilisée, connues sous le nom de facteurs de Stein. Une littérature abondante existe à ce sujet, pour diverses mesures de référence π : citons, dans le cas discret, Barbour and Eagleson [1983] lorsque π est une loi de Poisson, [Barbour et al., 1992, Chapitre 9] pour la loi binomiale, et Peköz [1996] pour la loi géométrique. L'article Brown and Xia [2001] représente une avancée considérable, en ce qu'il propose un critère général sur l'approximande π permettant de majorer les facteurs de Stein associés à π et à la distance en variation totale, et permet ainsi d'automatiser une partie des calculs.

Dans une optique similaire, je présente dans cette thèse une méthode universelle d'évaluation des facteurs de Stein associés à des mesures de probabilité π sur $\mathbb{N}$, qui exploite les entrelacements au premier et second ordre (1.4) et (1.5) de manière novatrice et systématique. Cette méthode permet de retrouver certains résultats classiques liés à l'approximation en variation totale, et fournit également des résultats pour des distances moins étudiées comme la distance de Wasserstein et la distance de Kolmogorov.

Structure

La première partie est consacrée aux processus de Markov inhomogènes, et est composée de trois chapitres. Dans le chapitre 2, après des rappels sur la théorie des processus markoviens et sur la notion de pseudo-trajectoire asymptotique, je présente de manière heuristique la stratégie développée dans cette thèse pour l'étude des processus de Markov inhomogènes en la mettant en relation avec les techniques spectrales classiques. Les deux chapitres suivants présentent les résultats nouveaux obtenus en appliquant cette stratégie : le chapitre 3 présente des bornes à la Berry-Esseen dans le cadre du théorème central limite classique pour la distance du χ_2, tandis que le chapitre 4 concerne le comportement asymptotique d'un processus de saut faiblement mélangeant qui peut être vu comme l'analogue continu d'un algorithme stochastique de recherche de médiane.

La deuxième partie s'organise de la manière suivante : le chapitre 5 présente dans un premier temps un tour d'horizon des idées majeures de la méthode de Stein-Chen, afin de replacer l'étude consacrée aux facteurs de Stein au sein d'un contexte général. Les contributions apportées par mes travaux sont l'objet du chapitre 6.

Enfin, des pistes de recherche sont présentées dans le chapitre 7.

Principales contributions

Les contributions de cette thèse font l'objet des chapitres 3, 4 et 6.

Bornes à la Berry-Esseen pour la distance du χ_2 dans le théorème central limite (Chapitre 3).

Pour des variables aléatoires réelles, indépendantes et identiquement distribuées $(X_i)_{i\in\mathbb{N}^*}$, d'espérance nulle et de variance 1, considérons

$$Y_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i, \qquad n \in \mathbb{N}^*.$$

Le théorème central limite affirme la convergence en loi de la suite $(Y_n)_{n\geq 1}$ vers la mesure gaussienne centrée réduite μ. Berry et Esseen obtiennent la première quantification de cette convergence, sous une condition de moment, en distance de Kolmogorov (Berry [1941]; Esseen [1945]), ouvrant le champ à des travaux portant sur d'autres distances. Parmi les travaux récents, citons Bally and Caramellino [2016] pour la distance en variation totale, et Artstein et al. [2004]; Bobkov et al. [2013] pour l'entropie relative.

Je m'intéresse ici à la distance du χ_2 définie pour deux mesures de probabilité θ, μ par

$$\chi_2(\theta,\mu) := \left(\int \left(\frac{d\theta}{d\mu} - 1\right)^2 d\mu\right)^{\frac{1}{2}}$$

si θ est absolument continue par rapport à μ, et $+\infty$ sinon. La quantité $\chi_2(\theta,\mu)$, qui n'est pas symétrique en ses arguments, porte le nom de distance par abus de language. Elle majore la distance en variation totale et l'entropie relative.

Le résultat principal est le suivant : supposons que les moments de X_1 soient identiques à ceux de μ jusqu'à un entier $r \geq 2$; supposons de plus que X_1 admette une densité par rapport à μ, et que cette densité soit polynomiale. Alors, sous des conditions sur les coefficients de la densité,

$$\limsup_{n\to+\infty} \quad n^{\frac{r-1}{2}}\,\chi_2(\mathcal{L}(Y_n),\mu) < +\infty.$$

Ce taux de convergence est optimal. On retrouve un aspect observé pour la distance en variation totale et l'entropie, qui est le gain d'un facteur $\sqrt{n}$ pour chaque moment supplémentaire identique à celui de la gaussienne.

C'est la technique de preuve qui fait l'originalité de ce travail. L'observation fondamentale est que la suite $(Y_n)_{n\in\mathbb{N}^*}$ forme une chaîne de Markov inhomogène. Ceci entraîne une relation de récursion sur les densités successives : en notant f_n la densité de Y_n par rapport à la loi normale μ, il existe un opérateur K_n, borné dans $L^2(\mu)$, tel que

$$f_{n+1} = K_n(f_n), \qquad n \in \mathbb{N}^*.$$

Le choix de la distance du χ_2 est bien adapté à ce cadre, car

$$\chi_2^2(\mathcal{L}(Y_n),\mu) = \int (f_n - 1)^2 d\mu, \qquad n \in \mathbb{N}^*.$$

Le résultat principal découle d'une analyse spectrale fine de l'opérateur K_n, menée grâce à la décomposition de $L^2(\mu)$ sur la base des polynômes de Hermite. La majoration du rayon spectral de l'opérateur symétrisé $K_n^* K_n$, au cœur de la preuve, peut être vue comme un analogue de l'inégalité de Poincaré pour le semi-groupe d'Ornstein-Uhlenbeck.

Il s'agit d'un travail en collaboration avec Laurent Miclo, qui sera soumis pour publication très prochainement.

Bornes à la Berry-Esseen pour un processus de Markov inhomogène à saut (chapitre 4).

Le processus de Markov inhomogène $(Z_t)_{t\geq 0}$ considéré évolue par sauts de deux types : saut à droite de $+a_t$, saut à gauche de $-a_t$, où a_t, la taille du saut à l'instant $t \geq 0$, décroît en temps grand vers 0. Les intensités de saut dépendent du temps et de la position, de telle sorte que le générateur instantané du processus s'écrit :

$$L_t[f](z) = p_t(z)(f(z+a_t) - f(z)) + q_t(z)(f(z-a_t) - f(z)), \qquad z \in \mathbb{R}, \quad t \geq 0.$$

L'étude de ce processus est motivée par un algorithme stochastique à la Robbins-Monro de recherche de médiane. Soit une mesure de probabilité ν sur $\mathbb{R}$ et $(X_i)_{i\in\mathbb{N}^*}$ des variables indépendantes, identiquement distribuées de loi ν. Alors, l'algorithme $(Y_n)_{n\in\mathbb{N}^*}$, défini récursivement comme

$$Y_{n+1} = Y_n + \gamma_{n+1}\left(\mathbf{1}_{X_{n+1}>Y_n} - \frac{1}{2}\right), \qquad n \geq 1,$$

pour une suite de pas $(\gamma_n)_{n\geq 1}$ de la forme $\gamma_n = cn^{-\beta}$ pour $\beta \in]0,1]$, converge vers la médiane de la mesure ν. De plus, l'algorithme est normalement asymptotique, c'est-à-dire que la suite $(Y_n)_{n\in\mathbb{N}^*}$ renormalisée converge vers la gaussienne (au moins dans le cas $\beta < 1$; des précautions sur la taille des pas à l'initialisation doivent être prises dans le cas $\beta = 1$ (Duflo [1997])). Peu de résultats relatifs à la quantification de cette convergence existent, car les hypothèses de forte convexité de la fonction objectif, habituelles dans le cadre des algorithmes de Robbins-Monro, ne sont pas satisfaites. Ceci engage à essayer une approche par inégalités fonctionnelles, tout d'abord sur un modèle simplifié à temps continu, qui correspond au processus de Markov $(Z_t)_{t\geq 0}$ ci-dessus pour des choix particuliers des intensités de saut.

On commence l'analyse du processus de Markov $(Z_t)_{t\geq 0}$ engendré par $(L_t)_{t\geq 0}$ dans le cas où celui-ci est homogène et réversible par rapport à la loi normale centrée réduite μ. Sous une condition de borne inférieure sur les intensités de saut permettant au processus de revenir de l'infini, le résultat suivant est établi : il existe des constantes $c, d > 0$ telles que

$$W(\mathcal{L}(Z_t), \mu) \leq (\mathrm{Var}_\mu[f_0])^{1/2}\left(\exp\left(-c \cdot a^2 \cdot t\right) + d \cdot a\right), \qquad t \geq 0,$$

où W représente la distance de Wasserstein d'ordre 1, f_0 la densité de la loi de Z_0 par rapport à la loi normale μ et a la taille des sauts, constante dans le cas homogène. Pour comprendre ce résultat, il faut remarquer que le processus $(Z_t)_{t\geq 0}$ partant du point $z \in \mathbb{R}$ vit sur le réseau $\Gamma_{a,z} = z + a\mathbb{Z}$, ce qui explique qu'il ne puisse pas converger vers la mesure gaussienne μ et que la majoration ci-dessus ne tende pas vers 0 en temps grand. Par contre, le point crucial est que le processus conditionné à rester sur $\Gamma_{a,z}$ est ergodique par rapport à la mesure μ restreinte à $\Gamma_{a,z}$: le terme en $\exp(-c \cdot a^2 \cdot t)$ vient alors d'une inégalité de Poincaré pour le processus conditionnel. Le terme en $d \cdot a$ reflète l'erreur faite en intégrant sur le réseau $\Gamma_{a,z}$ plutôt que sur $\mathbb{R}$ tout entier.

Au vu de l'inégalité ci-dessus, on conjecture que dans le cas inhomogène, où la taille des sauts a est remplacée par une fonction $(a_t)_{t\geq 0}$ tendant vers 0, la distance de Wasserstein $W(\mathcal{L}(Z_t), \mu)$ tend vers 0. L'énoncé précis de cette conjecture, dans le cas où le processus inhomogène $(Z_t)_{t\geq 0}$ est encore réversible par rapport à μ et dans le cas où il est seulement normalement asymptotique, est donné dans le chapitre 4. Les preuves sont complètes, à défaut d'un point technique qui est en cours d'investigation.

Il s'agit d'un travail en cours, en collaboration avec Sébastien Gadat et Laurent Miclo.

Entrelacements et facteurs de Stein pour les processus de naissance-mort (chapitre 6).

On considère un processus de naissance-mort sur $\mathbb{N}$, c'est-à-dire un processus de Markov homogène, de générateur

$$L(f)(x) = \alpha(x)(f(x+1) - f(x)) + \beta(x)(f(x-1) - f(x)), \qquad x \in \mathbb{N},$$

et on note $(P_t)_{t\geq 0}$ le semi-groupe de Markov associé. On définit les gradients discrets ∂ et ∂^* comme :

$$\partial f(x) = f(x+1) - f(x), \qquad x \in \mathbb{N}; \qquad \partial^* f(x) = f(x-1) - f(x), \qquad x \in \mathbb{N}^*.$$

Il existe une relation d'entrelacement au premier ordre entre le semi-groupe de naissance-mort $(P_t)_{t\geq 0}$ et le gradient discret ∂ (Chafaï and Joulin [2013]) :

$$\partial P_t = \widetilde{P}_t^{\widetilde{V}} \partial, \qquad t \geq 0,$$

où $(\widetilde{P}_t^{\widetilde{V}})_{t\geq 0}$ est un semi-groupe de Feynman-Kac formé d'un semi-groupe de naissance-mort alternatif $(\widetilde{P}_t)_{t\geq 0}$ et d'un potentiel $\widetilde{V}$.

Le premier résultat de cette partie de la thèse consiste à établir une relation d'entrelacement au second ordre :

$$\partial^* \partial P_t = \widehat{P}_t^{\widehat{V}} \partial^* \partial, \qquad t \geq 0,$$

où $(\widehat{P}_t^{\widehat{V}})_{t\geq 0}$ est un nouveau semi-groupe de Feynman-Kac, formé d'un semi-groupe de Markov $(\widehat{P}_t)_{t\geq 0}$, et d'un potentiel $\widehat{V}$. Contrairement au semi-groupe $(\widetilde{P}_t)_{t\geq 0}$ obtenu au premier ordre, le semi-groupe $(\widehat{P}_t)_{t\geq 0}$ n'est pas de naissance-mort dans le cas général. L'opérateur $\partial^*\partial$, qui s'écrit

$$\partial^* \partial f(x) = -f(x+1) + 2f(x) - f(x-1), \qquad x \in \mathbb{N}^*,$$

peut être interprété comme un laplacien discret. La relation d'entrelacement est en fait valable pour des opérateurs aux différences plus généraux, faisant intervenir des gradients à poids.

La principale application de cette formule d'entrelacement concerne la méthode de Stein-Chen, qui consiste en un ensemble de techniques visant à majorer

$$d_{\mathcal{F}}(\mu, \pi) := \sup_{f\in\mathcal{F}} |\mu(f) - \pi(f)|,$$

la distance entre μ et π deux mesures de probabilité sur $\mathbb{N}$, définie sur une classe $\mathcal{F}$ de fonctions-tests de $\mathbb{N}$ dans $\mathbb{R}$. Une étape-clé de cette méthode, qui désymétrise complètement les rôles joués par μ et π, est l'évaluation des facteurs de Stein, relatifs à la mesure de référence π et la classe de fonctions $\mathcal{F}$. Plus précisément, soit L le générateur infinitésimal d'un processus de naissance-mort réversible pour la mesure π. L'équation de Poisson associée s'écrit, par rapport à la donnée $f : \mathbb{N} \to \mathbb{R}$,

$$L(h_f) = f - \pi(f).$$

Les facteurs de Stein du premier et second ordre sont alors définis comme

$$\sup_{f\in\mathcal{F}} \|\partial(h_f)\|_\infty, \qquad \sup_{f\in\mathcal{F}} \|\partial^*\partial(h_f)\|_\infty.$$

Or, la solution de l'équation de Poisson h_f admet une représentation intégrale, sous la forme

$$h_f = -\int_0^\infty P_t(f)dt,$$

où $(P_t)_{t\geq 0}$ est le semi-groupe de naissance-mort associé à L. Ainsi,

$$\partial(h_f) = -\int_0^\infty \partial P_t(f)dt, \qquad \partial^*\partial(h_f) = -\int_0^\infty \partial^*\partial P_t(f)dt,$$

ce qui fait apparaître le début d'un entrelacement du premier et du second ordre sous l'intégrale. En se basant sur cette observation, je présente une méthode d'évaluation des facteurs de Stein qui utilise systématiquement les relations d'entrelacement du premier et second ordre.

L'avantage de cette méthode est son caractère universel, c'est-à-dire qu'elle est valable pour toute mesure discrète π satisfaisant aux hypothèses des théorèmes. Elle permet ainsi d'automatiser une partie des calculs relatifs à la méthode de Stein, qui peuvent se révéler assez lourds. En cela, elle se compare aux résultats présentés dans Brown and Xia [2001], qui concernent la distance en variation totale. Notre méthode permet de considérer diverses distances, dont des distances moins étudiées comme la distance de Wasserstein d'ordre 1 W.

Par exemple, appelons $\mathrm{NB}(r,p)$ la loi binomiale négative de paramètres (r,p). Dans le régime $p \to 1$, $r \to \infty$ et $r(1-p)/p \to \lambda$, cette mesure tend en loi vers la loi de Poisson $\mathcal{P}_\lambda$. Nos résultats permettent d'obtenir la borne suivante :

$$W\left(\mathrm{NB}(r,p), \mathcal{P}_{\frac{r(1-p)}{p}}\right) \leq \frac{8}{3\sqrt{2e}}\sqrt{\frac{r(1-p)}{p}}\frac{(1-p)}{p},$$

qui semble être la première quantification de cette convergence bien connue.

Il s'agit d'un travail en collaboration avec Bertrand Cloez, soumis pour publication.

Première partie

Processus de Markov inhomogènes

Chapitre 2

Introduction

Dans ce chapitre, on commence par un bref rappel de la définition des processus de Markov inhomogènes et de quelques-unes de leurs propriétés utiles dans la suite. Les deux parties suivantes forment une introduction heuristique à la stratégie développée dans cette thèse pour l'étude asymptotique des processus de Markov inhomogènes. La partie 2.2 présente la notion de pseudo-trajectoire asymptotique. Quoique non utilisée directement, cette notion présente des similitudes avec les idées de la thèse. Dans la partie 2.3, on explique comment obtenir des inégalités de type Poincaré pour des dynamiques non réversibles.

2.1 Définition des objets

2.1.1 Généralités

Introduisons tout d'abord les notations utiles pour la suite. Soit $(E, \mathcal{B})$ un espace polonais, c'est-à-dire métrique, complet et séparable, muni de la tribu borélienne. Parmi les fonctions $f : E \to \mathbb{R}$, on note $\mathcal{S}$ l'ensemble des fonctions C^∞ à support compact, $\mathcal{F}_b(E)$ l'ensemble des fonctions mesurables bornées et $\mathcal{F}_+(E)$ l'ensemble des fonctions mesurables positives. Pour toute fonction $f : E \to \mathbb{R}$, on définit

$$\|f\|_\infty := \max_{x \in E} |f(x)|.$$

Le symbole $\mathbf{1}$ désigne la fonction de E dans $\mathbb{R}$ identiquement égale à 1, et Id l'opérateur identité agissant sur les fonctions $f : \mathbb{E} \to \mathbb{R}$. Le domaine d'un opérateur L agissant sur les fonctions $f : \mathbb{E} \to \mathbb{R}$ est noté $\mathcal{D}(L)$. Pour toute mesure de probabilité ν sur $(E, \mathcal{B})$ et pour toute fonction $f \in \mathcal{F}_+(E)$ ou $f \in L^1(\nu)$, l'intégrale de f contre ν est notée indifféremment

$$\nu[f] = \int f d\nu.$$

Enfin, pour toute $f \in L^2(\nu)$,

$$\mathrm{Var}_\nu(f) = \int (f - \nu[f])^2 \, d\nu.$$

La définition et quelques propriétés classiques des processus markoviens (Ethier and Kurtz [1986], Eberle [2009]) sont rappelées maintenant sans démonstration. Soit $(\Omega, \mathcal{U}, \mathbb{P})$ un espace de probabilités et $I = \mathbb{N}$ ou $I = [0, +\infty[$. Un processus stochastique est une collection de variables aléatoires $(X_t)_{t \in I}$, avec $X_t : \Omega \to E$ mesurable pour tout $t \in I$. En temps discret, c'est-à-dire lorsque $I = \mathbb{N}$, on qualifie $(X_t)_{t \in I}$ de chaîne. On note $(\mathcal{F}_t)_{t \in I}$ la filtration engendrée par $(X_t)_{t \in I}$,

$$\mathcal{F}_t := \sigma(X_s, \, s \in I, \, s \le t), \qquad t \ge 0.$$

On dit que le processus $(X_t)_{t\in I}$ est markovien si pour toute fonction $f \in \mathcal{F}_b(E)$ ou $f \in \mathcal{F}_+(E)$,

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \mathbb{E}[f(X_t)|X_s], \qquad t \geq s,\ s,t \in I.$$

On associe alors au processus $(X_t)_{t\in I}$ son semi-groupe $(P_{s,t})_{t\geq s,\, s,t\in I}$, défini pour toute fonction $f \in \mathcal{F}_b(E)$ comme

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = P_{s,t}[f](X_s), \qquad \mathbb{P}-\text{p.s.}, \qquad t \geq s,\ s,t \in I.$$

Le processus de Markov $(X_t)_{t\in I}$ est dit homogène en temps si le semi-groupe vérifie

$$P_{s,t} = P_{s+u,t+u}, \qquad t \geq s,\ s,t,u \in I,$$

et inhomogène sinon. La majorité de la littérature sur les processus markoviens est consacrée aux processus homogènes ; on peut transférer un certain nombre de théorèmes concernant les processus homogènes aux processus inhomogènes en remarquant que, si $(X_t)_{t\in I}$ est un processus de Markov inhomogène, le processus $(\hat{X}_t)_{t\in I}$ à valeurs dans $I\times E$, de valeur initiale $(s,x) \in I\times E$ défini par

$$\hat{X}_t := (s+t, X_{s+t}), \qquad t \geq 0, \tag{2.1}$$

est un processus de Markov homogène.

Le semi-groupe $(P_{s,t})_{t\geq s,\, s,t\in I}$ possède les propriétés suivantes, dont la première qui justifie l'appellation de semi-groupe :

Proposition 2.1.1 (Propriétés des semi-groupes markoviens)**.**

1. *Pour tous temps $r,s,t \in I$ avec $r \leq s \leq t$, on a $P_{r,s}P_{s,t} = P_{r,t}$.*
2. *Si $f \in \mathcal{F}_b(E)$ est positive, alors $P_{s,t}[f]$ est aussi positive pour tous $s,t \in I$, $s \leq t$.*
3. *Pour tous $s,t \in I$, $s \leq t$, $P_{s,t}[\mathbf{1}] = \mathbf{1}$.*
4. *Pour toute fonction $f \in \mathcal{F}_b(E)$,*

$$\|P_{s,t}[f]\|_\infty \leq \|f\|_\infty, \qquad s,t \in I, \quad s \leq t,$$

En particulier, le semi-groupe $(P_{s,t})_{t\geq s,\, s,t\in I}$ est formé d'opérateurs contractants sur l'espace de Banach $\mathcal{F}_b(E)$ muni de la norme $\|\cdot\|_\infty$.

Par définition, le semi-groupe $(P_{s,t})_{t\geq s,\, s,t\in I}$ agit sur les fonctions de E dans $\mathbb{R}$, et par dualité sur les mesures de probabilité sur $(E,\mathcal{B})$. Pour toute mesure de probabilité ν sur $(E,\mathcal{B})$, et pour tous temps $s,t \in I$ avec $s \leq t$, on note $\nu P_{s,t}$ la loi du processus au temps t, X_t, sachant que X_s est distribué selon ν : la mesure $\nu P_{s,t}$ est alors caractérisée par l'équation

$$(\nu P_{s,t})[f] = \int P_{s,t}[f](x)d\nu(x), \qquad f \in \mathcal{F}_b(E).$$

La propriété de semi-groupe prend alors la forme suivante :

$$\nu P_{r,t} = (\nu P_{r,s})P_{s,t}.$$

La mesure de probabilité ν est dite invariante pour le semi-groupe (ou de manière équivalente, invariante pour le processus, ou pour le générateur défini ci-dessous) si, pour tous temps $s,t \in I$ avec $s \leq t$, on a $\nu P_{s,t} = \nu$. Dans ce cas, on peut compléter la proposition 2.1.1 ci-dessus avec une propriété de contraction dans $L^p(\nu)$: pour tout réel $p \geq 1$ et pour toute fonction $f \in L^p(\nu)$, on peut définir $P_{s,t}[f]$ par complétion et

$$\|P_{s,t}[f]\|_{L^p(\nu)} \leq \|f\|_{L^p(\nu)}, \qquad s,t \in I, \quad s \leq t. \tag{2.2}$$

Enfin, on dira que la mesure de probabilité ν est réversible par rapport au semi-groupe si, pour toutes fonctions $f,g \in L^2(\nu)$,

$$\int fP_{s,t}[g]d\nu = \int gP_{s,t}[f]d\nu, \qquad s,t \in I, \quad s \leq t.$$

Une mesure réversible est invariante.

2.1.2 Cas discret

Un objet central dans la théorie markovienne est le générateur associé au processus, ou plutôt dans le cas inhomogène, une collection de générateurs. En temps discret, le générateur $(L_t)_{t\in\mathbb{N}}$ associé au processus est

$$L_t := P_{t,t+1} - \mathrm{Id}, \qquad t \in \mathbb{N}.$$

Avec ces notations et par la propriété de Markov, on voit que pour toute fonction $f \in \mathcal{F}_b(E)$,

$$P_{0,t+1}[f] - P_{0,t}[f] = P_{0,t}[L_t[f]], \qquad t \in \mathbb{N}.$$

Cette équation porte le nom d'équation de Kolmogorov *forward*. Soit ν une mesure de probabilité de référence sur $(E, \mathcal{B})$. Supposons que pour le temps $t \in \mathbb{N}$, la loi de la chaîne X_t soit absolument continue par rapport à ν, de densité notée f_t, et que f_t soit dans le domaine de $P^*_{t,t+1}$, l'adjoint de $P_{t,t+1}$ dans $L^2(\nu)$. Alors, pour toute fonction $f \in \mathcal{F}_b(E)$,

$$P_{0,t+1}[f] = P_{0,t}[[P_{t,t+1}[f]] = \int f_t P_{t,t+1}[f] d\nu = \int f P^*_{t,t+1}[f_t] d\nu,$$

donc la loi de X_{t+1} est absolument continue par rapport à ν, et sa densité f_{t+1} vérifie :

$$f_{t+1} = P^*_{t,t+1}[f_t]. \tag{2.3}$$

2.1.3 Cas continu

Dans le cas d'un processus de Markov à temps continu, $I = [0, +\infty[$, la définition du générateur est moins directe. On dira que le semi-groupe $(P_{s,t})_{t\geq s\geq 0}$ est fortement continu dans l'espace de Banach $(\mathcal{F}_b(E), \|\cdot\|_\infty)$ si pour toute fonction $f \in \mathcal{F}_b(E)$,

$$\lim_{t\downarrow s} \|P_{s,t}[f] - f\|_\infty = 0, \qquad s \geq 0.$$

On définit alors

$$L_s[f] := \lim_{t\downarrow s} \frac{P_{s,t}[f] - f}{t - s}, \qquad f \in \mathcal{D}(L_s), \qquad s \geq 0,$$

où le domaine de L_s, $\mathcal{D}(L_s)$, est l'ensemble des fonctions de $\mathcal{F}_b(E)$ telle que la limite ci-dessus existe. On peut alors énoncer la proposition suivante :

Proposition 2.1.2 (Equations de Kolmogorov). *Si $(P_{s,t})_{t\geq 0}$ est fortement continu sur $(\mathcal{F}_b(E), \|\cdot\|_\infty)$, alors pour toute fonction $f \in \mathcal{F}_b(E)$ et tout $s \geq 0$, la fonction*

$$t \mapsto P_{s,t}[f],$$

définie sur $[s, +\infty[$, est continue. De plus, pour toute fonction $f \in \cap_{s\geq 0}\mathcal{D}(L_s)$,

$$P_{s,t}[f] \in \bigcap_{s>0} \mathcal{D}(L_s), \qquad t \geq s \geq 0,$$

et

$$\partial_t P_{s,t}[f] = L_s[P_{s,t}[f]] = P_{s,t}[L_s[f]], \qquad t \geq s \geq 0. \tag{2.4}$$

Ces deux égalités forment les équations de Kolmogorov *forward* et *backward*. La proposition 2.1.2 correspond à [Eberle, 2009, Theorem 3.12] dans le cas homogène ; on trouve l'énoncé dans le cadre inhomogène en considérant le processus homogène $(\hat{X}_t)_{t\geq 0}$ défini par (2.1), et en observant que le générateur infinitésimal de ce processus s'écrit

$$L[f](s,x) = \partial_s f(s,x) + L_s[f(s,\cdot)](x),$$

pour toute fonction $f \in \mathcal{F}_b([0,+\infty[\times E)$ dérivable en sa première variable et telle que $\forall s \in [0,+\infty[,\ f(s,\cdot) \in \mathcal{D}(L_s)$.

Etant donné une mesure de probabilité ν de référence sur $(E,\mathcal{B})$, supposons que le processus de Markov $(X_t)_{t\geq 0}$ ait une marginale au temps $t \geq 0$ admettant une densité f_t par rapport à ν, et que $f_t \in \mathcal{D}(L_t^*)$, où L_t^* désigne l'adjoint de L_t dans $L^2(\nu)$. On a alors pour toute fonction $f \in \mathcal{S}$,

$$\partial_t P_{0,t}[f] = P_{0,t}[L_t[f]] = \int f_t L_t[f] d\nu = \int f L_t^*[f_t] d\nu,$$

et d'autre part

$$\partial_t P_{0,t}[f] = \int (\partial_t f_t) f d\nu.$$

Comme ceci est vrai pour toute fonction $f \in \mathcal{S}$, on en déduit que

$$\partial_t f_t = L_t^*[f_t]. \tag{2.5}$$

Par ailleurs, la proposition 2.1.2 s'étend à tous les espaces de Banach $(B, \|\cdot\|_B)$ inclus dans l'espace des fonctions mesurables sur lesquels le semi-groupe est contractant et fortement continu.

La donnée du générateur et de son domaine permet de reconstruire un unique semi-groupe markovien vérifiant les équations de Kolmogorov (2.4). Quand le domaine n'est pas connu, on se donne dans le cas homogène le générateur et un cœur, qui est une sous-algèbre $\mathcal{A}$ de l'espace de Banach B, telle que le graphe $\{(f, L[f]), f \in \mathcal{A}\}$ soit dense dans l'espace produit des fonctions mesurables pour la norme $\|f\|_B + \|L[f]\|_B$. En utilisant la définition (2.1), on pourrait alors donner une condition équivalente pour un processus de Markov inhomogène.

Cette formalisation n'est pas nécessaire pour les processus considérés dans cette thèse. Au chapitre 3, on considère en effet une chaîne de Markov. Au chapitre 4, le processus étudié admet pour générateur

$$L_t[f](z) = p_t(z)(f(z+a_t) - f(z)) + q_t(z)(f(z-a_t) - f(z)), \qquad f \in \mathcal{F}_b(\mathbb{R}), \quad z \in \mathbb{R}, \quad t \geq 0,$$

où p, q sont des fonctions positives dépendant du temps et de l'espace et $(a_t)_{t\geq 0}$ une fonction positive. On suppose que

$$\|p_t\|_\infty, \|q_t\|_\infty < +\infty, \qquad t \geq 0.$$

Pour tout temps $t \geq 0$, l'opérateur L_t est donc borné dans $(\mathcal{F}_b(E), \|\cdot\|_\infty)$, et le processus admet une construction explicite. En effet, supposons que $X_0 = x \in \mathbb{R}$, et soit U une variable aléatoire de loi exponentielle d'intensité 1, et $T > 0$ tel que

$$\int_0^T (p_t(x) + q_t(x))dt = U.$$

Indépendamment de U, tirons une variable aléatoire Y valant a_T avec probabilité $p_T(x)/(p_T(x) + q_T(x))$, et valant $-a_T$ avec probabilité $q_T(x)/(p_T(x)+q_T(x))$. On pose alors

$$\begin{cases} X_t = x, & t \in [0,T[, \\ X_T = x + Y. \end{cases}$$

On procède ensuite de même à partir du temps T.

2.2 Pseudo-trajectoires asymptotiques

L'inhomogénéité peut traduire différents phénomènes. On s'intéresse ici à des cas où l'inhomogénéité peut être vue comme venant de l'approximation d'une dynamique homogène.
Pour comprendre le phénomène, donnons d'abord une analogie déterministe, concernant le schéma d'Euler à pas décroissant associé à l'équation différentielle ordinaire

$$\dot{x}(t) = \phi(x(t)), \qquad t \geq 0, \tag{2.6}$$

où ϕ est une fonction de classe C^1 bornée de $\mathbb{R}$ dans $\mathbb{R}$. On considère le flot associé à (2.6) : pour tout $x \in \mathbb{R}$ et $t \geq 0$, notons $\Phi_t(x)$ la solution de (2.6) au temps $t \geq 0$ telle que $x(0) = x$, qui existe vu les hypothèses sur ϕ. Le flot Φ vérifie

$$\Phi_{t+s} = \Phi_t \Phi_s, \qquad t, s \geq 0. \tag{2.7}$$

Pour toute fonction $f : \mathbb{R} \to \mathbb{R}$, pour tout temps $t \geq 0$ et $x \in \mathbb{R}$, introduisons $P_t[f](x) = f(\Phi_t(x))$. On voit alors grâce à (2.7) que $(P_t)_{t\geq 0}$ est un semi-groupe (déterministe) homogène.

Soit $(\gamma_n)_{n\in\mathbb{N}^*}$ une suite de pas décroissante et tendant vers 0, telle que $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$. La suite de temps correspondante $(T_n)_{n\geq n}$ est définie par $T_0 = 0$ et $T_n = \sum_{k=1}^n \gamma_k$ pour $n \in \mathbb{N}^*$. Pour tout $n \in \mathbb{N}$ et pour tout $x \in \mathbb{R}$, le schéma d'Euler à pas décroissant associé à (2.6) est défini de manière récursive comme

$$\begin{cases} x_n = x, \\ x_{k+1} = x_k + \gamma_{k+1}\phi(x_k), \quad k \geq n. \end{cases} \tag{2.8}$$

La suite $(x_k)_{k\geq n}$ est bornée car on a supposé que la fonction ϕ l'est. Soit $(\bar{x}(t))_{t\geq T_n}$ la fonction dont le graphe interpole linéairement les points $(T_k, x_k)_{k\geq n}$ et définissonns $\bar{\Phi}_{T_n,t}(x) = \bar{x}(t)$. Pour toute fonction $f : \mathbb{R} \to \mathbb{R}$, définissons également $Q_{T_n,t}[f](x) = f(\bar{\Phi}_{T_n,t}(x))$. Le flot approché $\bar{\Phi}$ vérifie :

$$\bar{\Phi}_{T_m,T_p}\bar{\Phi}_{T_n,T_m} = \bar{\Phi}_{T_n,T_p}, \qquad p \geq m \geq n. \tag{2.9}$$

L'équation (2.9) se traduit par l'égalité $Q_{T_n,T_m}Q_{T_m,T_p} = Q_{T_n,T_p}$, ainsi $(Q_{T_n,T_m})_{n,m\in\mathbb{N},\, n\leq m}$ peut être vu comme un semi-groupe inhomogène.

Intuitivement, comme $\gamma_n \to 0$, on voudrait penser que le flot approché se rapproche du flot exact. Pour préciser cela, on regarde la solution approchée de (2.6) construite par le schéma d'Euler à pas décroissant ci-dessus dans une fenêtre de taille 1, et on fait glisser cette fenêtre vers l'infini : pour $x \in \mathbb{R}$ fixé, considérons la suite de fonctions $(x^{(n)})_{n\in\mathbb{N}}$ définie par :

$$x^{(n)} : [0,1] \mapsto \mathbb{R}, \quad t \to \bar{\Phi}_{0,T_n+t}.$$

Le théorème de Kushner-Clark vient alors préciser l'intuition ci-dessus :

Proposition 2.2.1 (Kushner-Clark, Kushner and Clark [1978])**.**
- *La suite de fonctions $(x^{(n)})_{n\in\mathbb{N}}$ est relativement compacte pour la topologie de la convergence uniforme sur $[0,1]$, et toute valeur d'adhérence est une solution de* (2.6) *sur* $[0,1]$ *;*
- *Si x^* est un zéro asymptotiquement stable de ϕ, c'est-à-dire qu'il existe une région d'attraction $\Gamma \subset \mathbb{R}$ telle que pour tout $x \in \Gamma$, on ait $\lim_{t\to+\infty} \Phi_t(x) = x^*$, et si la suite de points $(x_n)_{n\in\mathbb{N}}$ retourne infiniment souvent dans Γ, alors $\lim_{n\to+\infty} x_n = x^*$.*

La preuve repose sur le fait suivant : pour tout point de départ $x \in \mathbb{R}$,

$$\sup_{t\in[0,1]} \left|\bar{\Phi}_{0,t+T_n}(x) - \Phi_t\left(\bar{\Phi}_{0,T_n}(x)\right)\right| \underset{n\to+\infty}{\to} 0. \tag{2.10}$$

En d'autres termes, quand le temps est grand deux phénomènes ont lieu simultanément : d'une part, le flot approché se rapproche du flot exact ; d'autre part, le flot exact se rapproche de la solution stationnaire de (2.6) $x(t) = x^*$ pour $t \geq 0$. En conséquence, le flot approché emmène également le schéma vers la solution stationnaire. Au niveau des semi-groupes, la propriété (2.10) se traduit par la convergence

$$\sup_{t \in [0,1]} |Q_{0,t+T_n}[f](x) - Q_{0,T_n}[P_t[f]](x)| \underset{n \to +\infty}{\to} 0, \tag{2.11}$$

pour tout $x \in \mathbb{R}$ et toute fonction f continue.

Revenons au cadre stochastique, et considérons en particulier un algorithme stochastique de la forme

$$X_{n+1} = X_n + \gamma_{n+1}\Phi(X_n) + \gamma_{n+1}^2 U_{n+1}, \tag{2.12}$$

où $(\gamma_n)_{n \in \mathbb{N}^*}$ est une suite de pas décroissante et tendant vers 0, telle que $\sum_{n \in \mathbb{N}} \gamma_n = +\infty$, et $(U_n)_{n \in \mathbb{N}^*}$ une suite de variables indépendantes et identiquement distribuées, et centrées et réduites. L'algorithme (2.12) est une version stochastique de (2.8). On attend qu'il soit fortement consistant, c'est-à-dire que la suite $(X_n)_{n \in \mathbb{N}^*}$ tende presque sûrement vers un zéro attractif de Φ. On dit que la propriété de pseudo-trajectoire asymptotique est vérifiée, si la propriété (2.10) est vraie presque sûrement. Cette notion, introduite par Benaïm (Benaïm [1999]), est bien adaptée à l'étude des valeurs limites que peut atteindre l'algorithme (2.12) en temps grand.

On attend également de l'algorithme (2.12) qu'il soit asymptotiquement normal, c'est-à-dire que l'algorithme convenablement renormalisé $(\widetilde{X}_n)_{n \in \mathbb{N}}$ tende en loi vers la mesure gaussiene. D'autres exemples d'algorithmes qui convergent en loi vers une mesure de probabilité sont les schémas d'Euler à pas décroissant approchant des diffusions (Lamberton and Pagès [2002], Lemaire [2005]). Il faut alors adapter (2.11).

Considérons un processus de Markov inhomogène $(X_t)_{t \geq 0}$, à temps continu par mesure de simplicité, de semi-groupe $(Q_{s,t})_{t \geq s \geq 0}$, et un semi-groupe de Markov homogène $(P_t)_{t \geq 0}$. Pour une distance d entre mesures de probabilité sur $\mathbb{R}$, on dit que le processus $(X_t)_{t \geq 0}$ correspond à une pseudo-trajectoire asymptotique de $(P_t)_{t \geq 0}$ si pour toute loi initiale ν_0,

$$\sup_{t \in [0,1]} d(\nu_0 Q_{0,s+t}, \nu_0 Q_{0,s} P_t) \underset{s \to +\infty}{\to} 0. \tag{2.13}$$

Cette définition, qui admet un analogue pour les chaînes de Markov inhomogènes en temps discret, est introduite dans Benaïm et al. [2016]. Elle permet d'étudier la convergence du processus inhomogène $(X_t)_{t \geq 0}$ vers la mesure invariante μ du processus homogène de semi-groupe $(P_t)_{t \geq 0}$.

L'article Benaïm et al. [2016] établit des résultats quantitatifs, qui portent sur la distance entre la loi marginale temporelle du processus à l'instant $t \geq 0$, et la mesure de probabilité μ. La preuve repose sur un développement limité du générateur du processus inhomogène faisant apparaître à la limite en temps grand le générateur du processus homogène. La classe de fonctions qui intervient dans la définition de la distance utilisée,

$$d_{\mathcal{F}}(\theta, \nu) = \sup_{f \in \mathcal{F}} |\theta[f] - \nu[f]|,$$

doit être assez restreinte pour que le développement limité soit vrai.

2.3 Inégalités de Poincaré et réversibilisation

2.3.1 Idée générale

Dans les deux chapitres suivants, on s'intéresse à un processus de Markov inhomogène $(X_t)_{t \geq 0}$, qui tend en loi vers une mesure de probabilité ν, et le but est de quantifier cette convergence.

L'heuristique rejoint celle de la notion de pseudo-trajectoire asymptotique (2.13) : en temps grand, la dynamique du processus est similaire à la dynamique homogène engendrée par un générateur $\bar{L}$, de mesure invariante ν.

De plus, le couple $(\bar{L}, \nu)$ est supposé vérifier une inégalité fonctionnelle de type Poincaré : il existe une constante $\lambda > 0$, telle que pour toute fonction f assez régulière,

$$-\int f\bar{L}[f]d\nu \geq \lambda \operatorname{Var}_\nu[f]. \tag{2.14}$$

Cette inégalité est liée à la décroissance exponentielle de la variance sous l'action du semi-groupe homogène qui correspond à $\bar{L}$, comme on le rappelle plus bas. C'est donc un outil adapté à l'étude de la convergence asymptotique des processus homogènes.

Dans les paragraphes suivants, on présente la stratégie qui est au cœur des contributions de cette thèse à l'étude des processus de Markov inhomogènes, et qui consiste à obtenir un analogue de (2.14) lorsque $\bar{L}$ est remplacé par L^*, l'adjoint dans $L^2(\nu)$ d'un générateur L non symétrique par rapport à ν. Il faut imaginer que L joue le rôle de L_t, le générateur instantané du processus inhomogène à l'instant $t \geq 0$.

La quantité qu'on utilise pour quantifier la convergence asymptotique est la distance du χ_2. Pour une mesure de probabilité ν, on note $f \cdot \nu$ la mesure de densité f par rapport à ν. La distance du χ_2 entre $f \cdot \nu$ et ν est définie par

$$\chi_2(f \cdot \nu, \nu) := \left(\int (f-1)^2 d\nu\right)^{1/2}.$$

Elle se réécrit encore

$$\chi_2(f \cdot \nu, \nu) = (\operatorname{Var}_\nu(f))^{1/2} = \sup\left\{\int (f-1)g\, d\nu, \quad g \in L^2(\nu), \int g^2 d\nu \leq 1\right\}.$$

On voit l'intérêt d'obtenir une inégalité analogue à (2.14) : elle permet de traiter une distance dont la définition se base sur la boule unité de $L^2(\nu)$, un ensemble de fonctions relativement large.

2.3.2 Cas continu

Soit L un générateur markovien et $(P_t)_{t\geq 0}$ le semi-groupe associé. Supposons que la dynamique soit réversible par rapport à la mesure de probabilité ν, c'est-à-dire que pour toutes fonctions f, g assez régulières,

$$\int fL[g]d\nu = \int gL[f]d\nu.$$

Introduisons l'énergie de Dirichlet :

$$\mathcal{E}_\nu(f) := -\int fL[f]d\nu, \qquad f \in \mathcal{D}(\mathcal{E}_\nu).$$

Par exemple, la diffusion de générateur défini, pour toute fonction $f \in \mathcal{S}$, par la formule

$$L[f](x) = f''(x) - U'(x)f'(x), \qquad x \in \mathbb{R},$$

est réversible par rapport à la mesure de Gibbs ν, dont la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est proportionnelle à $x \mapsto e^{-U(x)}$. L'énergie de Dirichlet s'écrit alors :

$$\mathcal{E}_\nu(f) = \int f'^2 d\nu, \qquad f \in \mathcal{D}(\mathcal{E}_\nu). \tag{2.15}$$

Si le processus est un processus de naissance-mort, de générateur

$$L[f](x) = p(x)(f(x+1) - f(x)) + q(x)(f(x-1) - f(x)), \qquad x \in \mathbb{Z},$$

symétrique par rapport à la mesure de probabilité ν sur $\mathbb{Z}$ si $\nu(z)p(z) = \nu(z+1)q(z+1)$ pour tout $z \in \mathbb{Z}$, alors

$$\mathcal{E}_\nu(f) = \int p(x)(f(x+1) - f(x))^2 d\nu(x), \qquad f \in \mathcal{D}(\mathcal{E}_\nu). \tag{2.16}$$

On dit que le générateur L admet un trou spectral si

$$\lambda_1(L) := \inf_{f \in \mathcal{D}(\mathcal{E}_\nu) \setminus \mathrm{Vect}(\mathbf{1})} \frac{\mathcal{E}_\nu(f)}{\mathrm{Var}_\nu(f)} > 0,$$

et pour une constante $\lambda > 0$, on appelle inégalité de Poincaré l'inégalité

$$-\int f L[f] d\nu \geq \lambda \mathrm{Var}_\nu[f], \qquad f \in \mathcal{D}(\mathcal{E}_\nu). \tag{2.17}$$

L'inégalité (2.17) entraîne la décroissance exponentielle de la variance sous l'action du semi-groupe : pour toute fonction $f \in L^2(\nu)$,

$$\mathrm{Var}_\nu(P_t[f]) \leq e^{-2\lambda t} \mathrm{Var}_\nu(f), \qquad t \geq 0. \tag{2.18}$$

Voyons la preuve, qui est le prototype de beaucoup d'autres dans la suite de cette thèse :

$$\begin{aligned}\partial_t \mathrm{Var}_\nu(P_t[f]) &= 2\int (P_t[f] - 1) L P_t[f] d\nu = 2 \int (P_t[f] - \nu[f]) L P_t[f - \nu[f]] d\nu = -2\mathcal{E}_\nu(P_t[f]) \\ &\leq -2\lambda \mathrm{Var}_\nu(P_t[f]),\end{aligned}$$

et le lemme de Gronwall permet de conclure. Réciproquement, l'inégalité (2.18) entraîne l'inégalité (2.17) par dérivation au temps 0.

On s'intéresse en particulier à une version de (2.18) pour une fonction f précise. Supposons que loi initiale du processus $(X_t)_{t \geq 0}$ admette une densité, notée f_0 par rapport à ν. On note alors f_t la densité de la loi de X_t par rapport à ν pour tout $t \geq 0$. D'après l'équation (2.5), et par la réversibilité de L, $\partial_t f_t = L f_t$, et donc $f_t = P_t[f_0]$ pour tout temps $t \geq 0$. Ainsi, l'inégalité (2.18) devient :

$$\chi_2^2(f_t \cdot \nu, \nu) \leq e^{-2\lambda t} \chi_2^2(f_0 \cdot \nu, \nu). \tag{2.19}$$

En suivant des idées de Fill (Fill [1991]), passons au cas d'un processus de Markov de générateur L et une mesure ν supposée invariante, mais pas nécessairement réversible. Alors, $L^*[\mathbf{1}] = 0$ et l'adjoint L^* de L dans $L^2(\nu)$ est encore un générateur markovien. Par l'équation (2.5), pour tout $t \geq 0$,

$$\partial_t \chi_2(f_t \cdot \nu, \nu) = 2 \int (f_t - 1) L^*[f_t] d\nu = 2 \int (f_t - 1) L^*[f_t - 1] d\nu.$$

Considérons l'opérateur $(L + L^*)/2$, qui s'appelle le réversibilisé additif de L et correspond à un générateur markovien symétrique par rapport à ν. On a alors :

$$\partial_t \chi_2(f_t \cdot \nu, \nu) = 2 \int (f_t - 1) \frac{L + L^*}{2} [f_t - 1] d\nu.$$

Si le processus réversibilisé de générateur $(L + L^*)/2$ admet un trou spectral λ, on peut conclure comme en (2.19).

Examinons maintenant le cas où ν n'est pas supposée réversible ni invariante par rapport à L. Alors, pour tout $t \geq 0$,

$$\begin{aligned}\partial_t \chi_2^2(f_t \cdot \nu, \nu) &= 2\int (f_t - 1)L^*[f_t]d\nu = 2\int (f_t-1)L^*[f_t-1]d\nu + 2\int (f_t-1)L^*[\mathbf{1}]d\nu \\ &\leq \int (f_t-1)(L+L^*)[f_t-1]d\nu + 2\left(\int (L^*[\mathbf{1}])^2 d\nu\right)^{1/2} \chi_2(f_t\cdot\nu,\nu),\end{aligned}$$

où on a utilisé l'inégalité de Cauchy-Schwarz dans la dernière ligne. On évalue alors séparément le terme

$$\int (L^*[\mathbf{1}])^2 d\nu,$$

qu'on peut voir comme une mesure du défaut d'invariance, puisque la mesure ν est invariante si et seulement si $L^*[\mathbf{1}] = 0$, et le terme recentré

$$\int (f_t-1)(L+L^*)[f_t-1]d\nu.$$

Cette méthode est introduite dans Arnaudon and Miclo [2016] pour étudier un algorithme de recherche de moyenne sur une variété, qui utilise des idées venant de la théorie du recuit simulé, et qui correspond à un processus de Markov inhomogène. Cet algorithme est construit de telle sorte que

$$L := \widetilde{L} + R,$$

où $\widetilde{L}$ est le générateur d'un processus de diffusion réversible, et admettant une inégalité de Poincaré, par rapport à la mesure de Gibbs ν. Grâce à la formule (2.15) donnant l'énergie de Dirichlet pour les diffusions, on obtient

$$\begin{aligned}\int (f_t-1)L^*[f_t-1]d\nu &= \int (f_t-1)\widetilde{L}[f_t-1]d\nu + \int (f_t-1)R^*[f_t-1]d\nu \\ &\leq -\lambda \int (f_t')^2 d\nu + \int (f_t-1)R^*[f_t-1]d\nu.\end{aligned}$$

Une utilisation habile de la formule de Taylor avec reste intégral permet ensuite de majorer le terme de reste sous la forme :

$$\int (f_t-1)R^*[f_t-1]d\nu \leq \alpha \int (f_t')^2 d\nu + \beta\left(\int (f_t-1)^2 d\nu\right)^{1/2},$$

avec $\alpha < \lambda$. Ce qui permet de conclure dans Arnaudon and Miclo [2016], c'est que l'ordre du développement asymptotique nécessaire pour faire apparaître le générateur symétrique est précisément le même que celui de l'ordre de dérivation obtenu dans la forme de Dirichlet. Or cette combinaison n'est pas vérifiée pour les processus étudiés dans les chapitres 3 et 4. Il faut donc trouver une autre méthode.

Revenons à l'expression

$$\int (f_t-1)(L+L^*)[f_t-1]d\nu.$$

L'opérateur $L + L^*$ est symétrique dans $L^2(\nu)$, mais n'est pas le générateur d'un processus de Markov si ν n'est pas invariante par rapport à L, car alors

$$(L+L^*)[\mathbf{1}] = L^*[\mathbf{1}] \neq \mathbf{1}.$$

Soit $L^2_0(\nu)$ l'ensemble des fonctions de $L^2(\nu)$ telles que $\nu[g]=0$. Cet espace n'est pas stable par $L+L^*$, mais en notant π la projection sur $L^2_0(\nu)$,

$$\pi: \quad L^2(\nu) \to L^2_0(\nu), \quad f \mapsto f-\nu[f],$$

on voit que pour toute fonction $g \in L^2_0(\nu)$,

$$\int g\pi((L+L^*)[g])d\nu = \int g(L+L^*)[g]d\nu.$$

Ainsi,

$$-\int (f_t-1)(L+L^*)[f_t-1]d\nu \geq \lambda \int (f_t-1)^2 d\nu,$$

où λ est la plus petite valeur propre de l'opérateur $-\pi(L+L^*)$ dans $L^2_0(\nu)$.
Dans le chapitre 4, on évalue cette valeur propre en introduisant l'opérateur

$$A(L) = L + L^* - L^*[\mathbf{1}]\mathrm{Id}.$$

Alors,

$$\begin{aligned}\int (f_t-1)L^*[f_t]d\nu &= \int (f_t-1)(L+L^*-L^*[\mathbf{1}]\mathrm{Id})[f_t-1]d\nu + \int L^*[\mathbf{1}](f_t-1)^2 d\nu \\ &\leq \int (f_t-1)A(L)[f_t-1]d\nu + \|L^*[\mathbf{1}]\|_\infty \int (f_t-1)^2 d\nu.\end{aligned}$$

L'opérateur $A(L)$ est symétrique dans $L^2(\nu)$ et vérifie

$$A(L)[\mathbf{1}] = 0.$$

L'intérêt d'introduire cet opérateur est qu'il correspond au générateur d'un processus réversible par rapport à ν.
Démontrons ceci dans le cas où L est un le générateur d'un processus à saut, défini pour toute fonction suffisament régulière par

$$L[f](x) = p(x)\int (f(x+y)-f(x))d\theta(y), \qquad x \in \mathbb{R},$$

pour θ une mesure de probabilité sur $\mathbb{R}$ et une fonction $p:\mathbb{R}\to\mathbb{R}$ positive, représentant l'intensité de saut. Supposons de plus que la mesure ν admet une densité strictement positive, qu'on notera encore ν, par rapport à la mesure de Lebesgue. On voit alors que

$$L^*[f](x) = \int f(x-y)p(x-y)\frac{\nu(x-y)}{\nu(x)}d\theta(y) - f(x)p(x), \qquad x \in \mathbb{R},$$

de telle sorte que

$$A(L)[f](x) = \int p(x)(f(x+y)-f(x)) + p(x-y)\frac{\nu(x-y)}{\nu(x)}(f(x-y)-f(x))d\theta(y), \qquad x \in \mathbb{R},$$

qui est de nouveau le générateur d'un processus markovien à sauts, et de plus réversible par rapport à la mesure de probabilité ν.
Si le générateur $A(L)$ vérifie l'inégalité de Poincaré (2.17) avec une constante $\lambda_1((A(L))$, on trouve alors que

$$\int (f_t-1)L^*[f_t]d\nu \leq (-\lambda_1((A(L)) + \|L^*[\mathbf{1}]\|_\infty)\int (f_t-1)^2 d\nu.$$

Faire apparaître $A(L)$ est intéressant dans la mesure où on peut appliquer des résultats connus pour évaluer le trou spectral $-\lambda_1((A(L))$. C'est la stratégie qu'on utilise au chapitre 4, où on utilise des inégalités de Hardy relatives aux processus de naissance-mort (Miclo [1999]). Le prix à payer est le terme $\|L^*[\mathbf{1}]\|_\infty$, qui est plus difficile à contrôler.

2.3.3 Cas discret

Voyons maintenant comment les idées ci-dessus se déclinent dans le cadre des chaînes de Markov à temps discret. Supposons dans un premier temps que la chaîne de Markov homogène $(X_n)_{n\in\mathbb{N}}$ soit réversible par rapport à une mesure de probabilité ν, et notons $P := P_{1,2}$. On désigne P sous le nom d'opérateur de transition de la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$. Par l'inégalité (2.2) de contraction dans $L^2(\nu)$, le spectre de P est inclus dans $[-1,1]$, de plus grand élément $\lambda_0(P) = 1$. Notons les éléments du spectre discret de P

$$\cdots \leq \lambda_n(P) \leq \cdots \leq \lambda_1(P) \leq \lambda_0(P) = 1.$$

D'après le théorème du min-max (Reed and Simon [1978]),

$$\lambda_1(P) = \max_{g\in L^2(\nu)\setminus\{0\},\, \nu[g]=0} \frac{\int gP[g]d\nu}{\mathrm{Var}_\nu(g)}.$$

Supposons que pour tout temps $n \in \mathbb{N}$, la marginale temporelle X_n admette une densité f_n par rapport à ν. Par l'équation (2.3) et par la réversibilité, on obtient que pour tout entier $n \in \mathbb{N}$, $f_{n+1} = P[f_n]$. Ainsi,

$$\begin{aligned}\chi_2^2(f_{n+1}\cdot\nu,\nu) &= \int (P[f_n]-1)^2 d\nu = \int (P[f_n-1])^2 d\nu = \int (f_n-1)P^2[f_n-1]d\nu \\ &\leq \lambda_1(P^2)\,\chi_2^2(f_n\cdot\nu,\nu).\end{aligned}$$

Si par exemple la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ est à valeurs dans un espace d'état fini et est irréductible, alors par le théorème de Perron-Frobenius, $\lambda_1(P) < 1$. Si de plus la chaîne est apériodique, alors P^2 est encore un opérateur de transition associé à une chaîne irréductible, donc $\lambda_1(P^2) < 1$ et l'inégalité ci-dessus permet d'obtenir la décroissance exponentielle de $\chi_2(f_n\cdot\nu,\nu)$. Citons par exemple Diaconis and Stroock [1991] pour des stratégies d'évaluation de $\lambda_1(P^2)$.

Supposons maintenant, comme dans Fill [1991], que la mesure ν soit invariante mais pas nécessairement réversible par rapport à la chaîne de Markov d'opérateur de transition P. Alors, P^* l'adjoint de P dans $L^2(\nu)$ vérifie $P^*[\mathbf{1}] = \mathbf{1}$ et est encore un opérateur de transition markovien. Ecrivons :

$$\begin{aligned}\chi_2^2(f_{n+1}\cdot\nu,\nu) &= \int (P^*[f_n]-1)^2 d\nu = \int (P^*[f_n-1])^2 d\nu = \int (f_n-1)PP^*[f_n-1]d\nu \\ &\leq \lambda_1(PP^*)\chi_2^2(f_n\cdot\nu,\nu).\end{aligned}$$

On voit apparaître le trou spectral de l'opérateur $M(P) = PP^*$, la réversibilisation multiplicative de P, qui est l'opérateur de transition d'une chaîne de Markov réversible par rapport à ν.

Dans le chapitre 3, la mesure cible ν n'est en général ni réversible ni invariante par rapport à la dynamique. On écrit alors grâce à l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}\chi_2(f_{n+1}\cdot\nu,\nu) &= \left(\int (P^*[f_n]-1)^2 d\nu\right)^{1/2} = \left(\int (P^*[f_n-1] + P^*[\mathbf{1}]-\mathbf{1})^2 d\nu\right)^{1/2} \\ &\leq \left(\int (P^*[f_n-1])^2 d\nu\right)^{1/2} + \left(\int (P^*[\mathbf{1}]-\mathbf{1})^2 d\nu\right)^{1/2}.\end{aligned}$$

Comme dans le cas continu, on traite séparément le terme traduisant le défaut d'invariance

$$\int (P^*[\mathbf{1}]-\mathbf{1})^2 d\nu,$$

et le terme recentré

$$\int (P^*[f_n-1])^2 d\nu = \int (f_n-1)PP^*[f_n-1]d\nu.$$

L'opérateur PP^* est symétrique dans $L^2(\nu)$, mais n'est pas markovien en général, car la masse n'est pas conservée : $PP^*[\mathbf{1}] \neq 1$ si la mesure ν n'est pas invariante pour P.

Pour conclure, il faut évaluer le rayon specral de la restriction de PP^* à l'espace $L_0^2(\nu)$ formé des fonctions de $L^2(\nu)$ de moyenne nulle, comme on le fait au chapitre 3.

2.3.4 Cas où la mesure de référence évolue

Jusqu'ici, on s'est placé dans le cadre où la mesure-cible ν est fixe. Examinons le cas où la mesure-cible $(\nu_t)_{t\geq 0}$ dépend du temps. Soit $(X_t)_{t\geq 0}$ un processus de Markov inhomogène, de générateur $(L_t)_{t\geq 0}$. On s'intéresse à l'évolution de la distance du χ_2 entre X_t et ν_t, $\chi_2^2(f_t \cdot \nu_t, \nu_t)$, où f_t désigne la densité de X_t par rapport à ν_t pour un temps $t \geq 0$. Supposons que pour tout temps $t \geq 0$, la mesure ν_t soit absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$, de densité encore notée ν_t par abus de notation. Formellement, l'équation (2.5) devient

$$\partial_t[f_t] = L_t^*[f_t] - \partial_t(\log \nu_t) f_t, \qquad t \geq 0. \tag{2.20}$$

Si l'équation (2.20) est vérifiée, on obtient

$$\begin{aligned}
\partial_t \chi_2^2(f_t \cdot \nu_t, \nu_t) &= 2\int (f_t - 1)(\partial_t f_t) d\nu_t + \int (f_t - 1)^2 (\partial_t \log \nu_t) d\nu_t \\
&= 2\int (f_t - 1) L_t^*[f_t] d\nu_t - 2\int (f_t - 1) f_t\, \partial_t(\log \nu_t) d\nu_t + \int (f_t - 1)^2 \partial_t(\log \nu_t) d\nu_t \\
&= 2\int (f_t - 1) L_t^*[f_t] d\nu_t - \int (f_t - 1)^2 \partial_t(\log \nu_t) d\nu_t - 2\int (f_t - 1)\partial_t(\log \nu_t) d\nu_t.
\end{aligned} \tag{2.21}$$

On reconnaît le terme habituel

$$\int (f_t - 1) L_t^*[f_t] d\nu_t,$$

qu'on a vu comment gérer dans la partie 2.3.2, et les termes supplémentaires liés à l'évolution de la mesure cible,

$$-\int (f_t - 1)^2 \partial_t(\log \nu_t) d\nu_t - 2\int (f_t - 1)\partial_t(\log \nu_t) d\nu_t.$$

Le terme le plus délicat est le premier,

$$-\int (f_t - 1)^2 \partial_t(\log \nu_t) d\nu_t,$$

qu'il faut majorer en fonction de

$$\chi_2^2(f_t \cdot \nu_t, \nu_t) = \int (f_t - 1)^2 d\nu_t.$$

Si on sait contrôler $\|\partial_t(\log \nu_t)\|_\infty$, alors les choses se passent bien. Par exemple, dans Arnaudon and Miclo [2016], l'espace d'état sur lequel vit l'algorithme est une variété compacte, ce qui entraîne que $\|\partial_t(\log \nu_t)\|_\infty < +\infty$ et permet de gérer les termes venant de l'évolution de la mesure-cible.

Néanmoins, il existe des exemples simples où $\|\partial_t(\log \nu_t)\|_\infty = +\infty$, ce qui empêche de conclure. Voyons à présent un tel exemple, inspiré de l'algorithme de schéma d'Euler à pas décroissant associé à une diffusion (Lamberton and Pagès [2002]). Considérons une diffusion de générateur défini pour toute fonction f assez régulière par

$$L[g] = \frac{1}{2}\sigma^2 g'' + bg',$$

où σ^2 est une fonction positive. Le schéma d'Euler à pas décroissant qui lui est associé est l'algorithme défini récursivement par

$$X_{n+1} = X_n + \gamma_{n+1} b(X_n) + \sqrt{\gamma_{n+1}}\sigma(X_n)U_{n+1},$$

où la suite $(\gamma_n)_{n\geq 1}$ décroît vers 0 et $\sum_{n\geq 1}\gamma_n = +\infty$. La suite de temps implicitement associée au schéma est $(T_n)_{n\geq 0}$ avec $T_0 = 0$ et $T_n = \sum_{k=1}^n \gamma_n$ pour tout $n \in \mathbb{N}^*$. Les variables aléatoires $(U_n)_{n\geq 1}$ sont indépendantes, identiquement distribuées, avec $\mathbb{E}[U_1] = 0$ et $\mathbb{E}[U_1^2] = 0$. Sous des conditions adaptées sur le coefficient de diffusion σ^2 et le drift b, l'algorithme converge en loi vers la mesure invariante de la diffusion de générateur L (Lamberton and Pagès [2002]; Lemaire [2005]).

On s'intéresse ici au schéma d'Euler associé au processus d'Ornstein-Uhlenbeck, qui est une diffusion de générateur

$$L[g] = g''(x) - xg'(x), \qquad x \in \mathbb{R}, \qquad g \in \mathcal{S}.$$

L'algorithme s'écrit alors

$$X_{n+1} = X_n - \gamma_{n+1}X_n + 2\sqrt{\gamma_{n+1}}U_{n+1}, \tag{2.22}$$

et on choisit ici $U_1 \sim \mu$, où μ désigne la mesure gaussienne centrée réduite $\mathcal{N}(0,1)$. Pour toute fonction $f \in \mathcal{F}_b(\mathbb{R})$, et pour tout réel $\gamma \in [0,1]$, introduisons l'opérateur

$$Q_\gamma[f](x) = \mathbb{E}[f((1-\gamma)x + \sqrt{2\gamma}U)], \qquad U \sim \mu.$$

Pour tout $\gamma \in [0,1]$, l'opérateur Q_γ est un opérateur de transition markovien, qui est symétrique par rapport à la mesure gaussienne centrée $\mu_\gamma := \mathcal{N}(0, \sigma_\gamma^2)$, de variance

$$\sigma_\gamma^2 = \frac{1}{1-\frac{\gamma}{2}}. \tag{2.23}$$

De plus, l'algorithme $(X_n)_{n\in\mathbb{N}}$ défini par (2.22) est une chaîne de Markov inhomogène, de semi-groupe

$$Q_{n,n+1} = Q_{\gamma_{n+1}}, \qquad n \in \mathbb{N}.$$

Par mesure de simplicité, considérons une version continue de l'algorithme (2.22), définie comme le processus de Markov inhomogène $(X_t)_{t\geq 0}$ de générateur

$$L_t[f] = \frac{1}{\eta_t^2}(Q_{\eta_t} - \mathrm{Id}),$$

pour une fonction $(\eta_t)_{t\geq 0}$ à valeur dans $]0,1]$ décroissant vers 0. L'intensité de saut η_t^2 est choisie pour correspondre aux ordres de grandeur du processus discret, où un saut de taille moyenne $\sqrt{\gamma_{n+1}}$ est réalisé dans une échelle de temps d'ordre $T_{n+1} - T_n = \gamma_{n+1}$.

Il semble alors naturel de s'intéresser à l'évolution de

$$\chi_2(f_t \cdot \mu_{\eta_t}, \mu_{\eta_t}),$$

où, avec les notations habituelles f_t désigne la densité de X_t au temps $t \geq 0$ par rapport à la mesure de référence $\mu_{\eta_t} = \mathcal{N}(0, \sigma_{\eta_t}^2)$, où $\sigma_{\eta_t}^2$ est défini par (2.23) pour $\gamma = \eta_t$. On note encore $x \mapsto \mu_{\eta_t}(x)$ la densité de μ_{η_t} par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Pour tout $\eta \in]0,1]$,

$$\mu_\eta(x) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{x^2}{2\sigma_\eta^2}\right), \qquad x \in \mathbb{R}.$$

On trouve alors que

$$\partial_t \log(\mu_{\eta t}) = \frac{\partial_t \sigma_{\eta t}^2}{2\sigma_{\eta t}^2} \left(\frac{x^2}{\sigma_{\eta t}^2} - 1 \right). \tag{2.24}$$

D'après (2.21), on a :

$$\partial_t \chi_2(f_t \cdot \mu_{\eta t}, \mu_{\eta t}) = 2 \int (f_t - 1) L_t^*[f_t] d\mu_{\eta t} - \int (f_t - 1)^2 \partial_t (\log \mu_{\eta t}) d\mu_{\eta t} - 2 \int (f_t - 1) \partial_t (\log \mu_{\eta t}) d\mu_{\eta t}.$$

Le terme

$$\int (f_t - 1) L_t^*[f_t] d\mu_{\eta t}$$

ne pose pas de problème, car pour tout temps $t \geq 0$, le générateur L_t est symétrique par rapport à $\mu_{\eta t}$ et vérifie une inégalité de Poincaré, par renormalisation de l'inégalité de Poincaré pour le semi-groupe d'Ornstein-Uhlenbeck. Cependant,

$$- \int (f_t - 1)^2 \partial_t (\log \mu_{\eta t}) d\mu_{\eta t} = -\frac{1}{2} \partial_t (\log \sigma_{\mu_{\eta t}}^2) \int (f_t(x) - 1)^2 \left(\frac{x^2}{\sigma_{\eta t}^2} - 1 \right) d\mu_{\eta t}(x).$$

Comme $t \to \sigma_{\eta t}^2$ est décroissante, le facteur $-\partial_t (\log \sigma_{\mu_{\eta t}}^2)$ est positif. De plus, il n'existe pas de constante $c > 0$ telle que pour toute fonction $g \in L^2(\mu_\eta)$ centrée, c'est-à-dire telle que $\mu_\eta[g] = 0$, on ait

$$\left| \int g(x)^2 \left(\frac{x^2}{\sigma_\eta^2} - 1 \right) d\mu_\eta(x) \right| \leq c \int g(x)^2 d\mu_\eta(x).$$

La méthode ne peut donc pas fonctionner sans plus d'information. Il faut des estimées a priori sur $\int x^2 (f_t(x) - 1)^2 d\mu_\eta(x)$.

Cet exemple montre donc que le fait d'avoir une mesure de référence dépendant du temps rend l'analyse de l'évolution de la distance du χ_2 plus compliquée. Notons au passage que dans l'exemple ci-dessus, une solution consisterait à regarder l'entropie relative plutôt que la distance du χ_2. Soit ν une distance de probabilité ; l'entropie relative de la distribution $f \cdot \nu$ par rapport à ν est définie par

$$\mathrm{Ent}_\nu(f) := \int f \log f \, d\nu.$$

On trouve alors, d'après les équations (2.20) et (2.24),

$$\begin{aligned}
\partial_t \mathrm{Ent}_{\mu_{\eta t}}(f_t) &= \int (\log f_t + 1)(\partial_t f_t) d\mu_{\eta t} + \int f_t \log f_t (\partial_t \log \mu_{\eta t}) d\mu_{\eta t} \\
&= \int (\log f_t + 1) L_t^*[f_t] d\mu_{\eta t} - \int (\partial_t \log \mu_{\eta t}) f_t d\mu_{\eta t} \\
&= \int (\log f_t) L_t[f_t] d\mu_{\eta t} - \frac{\partial_t \sigma_{\eta t}^2}{\sigma_{\eta t}^2} \int \left(\frac{x^2}{\sigma_{\eta t}^2} - 1 \right) f_t d\mu_{\eta t}.
\end{aligned}$$

Le générateur L_t est symétrique par rapport à $\mu_{\eta t}$ et vérifie une inégalité de log-Sobolev (qui se déduit de l'inégalité de log-Sobolev usuelle pour le processus d'Ornstein-Uhlenbeck), ce qui permet de traiter le terme

$$\int (\log f_t) L_t[f_t] d\mu_{\eta t}.$$

Enfin, pour contrôler le terme

$$-\frac{\partial_t \sigma_{\eta_t}^2}{\sigma_{\eta_t}^2} \int \left(\frac{x^2}{\sigma_{\eta_t}^2} - 1 \right) f_t d\mu_{\eta_t}$$

en fonction de l'entropie relative, on peut utiliser l'inégalité de Fenchel-Young, qui assure que $\forall x > 0$ et $\forall y \in \mathbb{R}$,

$$xy \leq x \log x + e^{y-1}.$$

Au final, on voit que pour une mesure-cible $(\nu_t)_{t\geq 0}$ évoluant dans le temps, avec un espace d'état non compact, l'entropie est peut-être un choix plus approprié que la distance du χ_2 pour mesurer la distance entre la loi de X_t et ν_t.

Dans le chapitre 4, un phénomène relié à l'évolution de la mesure-cible apparaît de manière plus détournée. Rappelons que l'objet de ce chapitre est de quantifier la convergence, sous des hypothèses adéquates, du processus de Markov inhomogène de générateur instantané

$$L_t[f](z) = p_t(z)(f(z + a_t) - f(z)) + q_t(z)(f(z - a_t) - f(z)), \qquad z \in \mathbb{R}, \quad t \geq 0,$$

vers la mesure gaussienne centrée réduite μ. Pour des raisons qui sont développées dans le chapitre 4, plutôt que de considérer directement la quantité

$$\chi_2^2(f_t \cdot \mu, \mu) = \int (f_t - 1)^2 d\mu,$$

on s'intéresse à la quantité

$$\int \mu_{a_t}[(f_t - \mu_{a_t}[f_t])^2] d\mu, \tag{2.25}$$

où μ_{a_t} correspond à une espérance conditionnelle par rapport à une sous-tribu qui dépend du temps à travers la taille des sauts $(a_t)_{t\geq 0}$. La dérivée en temps de (2.25) fait alors apparaître un terme de la forme

$$\partial_t(\log a_t) \int (f - \mu_{a_t}[f]) \nabla \mu_{a_t}[f] \, \xi \, d\mu,$$

où ξ désigne la fonction identité de $\mathbb{R}$ dans $\mathbb{R}$. L'investigation de ce terme est en cours de finalisation ; on conjecture qu'il existe un réel $a_0 > 0$ tel que pour tout $0 < a < a_0$ et pour tout $f \in L^2(\mu)$,

$$\int (f - \mu_a[f]) \nabla \mu_a[f] \, \xi \, d\mu \geq 0.$$

Les résultats du chapitre 4 sont énoncés sous cette conjecture.

Chapter 3

Berry-Esseen bounds for the χ_2-distance in the Central Limit Theorem

This chapter consists in a joint work with Laurent Miclo shortly to be submitted for publication.

Abstract

The main result of this article is a Berry-Esseen-like bound, which states the convergence to the normal distribution of sums of independent identically distributed random variables in chi-square distance, defined as the variance of the density with respect to the normal distribution. Our main assumption is that the independent identically distributed random variables have polynomial density. The method consists of taking advantage of the underlying time-inhomogeneous Markovian structure and providing a Poincaré-like inequality for the non-reversible transition operator, which allows to find the optimal rate in the convergence above under matching moments assumptions.

3.1 Introduction

The present article is devoted to the study of the rate of convergence in Central Limit Theorem in χ_2-distance, defined for two probability distributions θ, μ as:

$$\chi_2(\theta,\mu) := \left(\int \left(\frac{d\theta}{d\mu} - 1\right)^2 d\mu\right)^{\frac{1}{2}}$$

if θ is absolutely continuous with respect to μ and $+\infty$ if not. In the following μ stands for the normal distribution. For a density function $f \in L^2(\mu)$ we use the shortened notation

$$\chi_2(f) := \chi_2(f \cdot \mu, \mu)$$

to refer to the χ_2-distance between the distribution with density f with respect to μ, denoted by $f \cdot \mu$, and μ itself. The χ_2-distance bounds by above usual quantities like total variation distance and relative entropy:

$$d_{TV}(f,\mu) := \int |f-1| d\mu \leq \chi_2(f), \qquad \mathrm{Ent}(f||\mu) := \int f \log f d\mu \leq \chi_2^2(f), \tag{3.1}$$

the first relation being a consequence of Cauchy-Schwarz inequality and the second of inequality $\log x \leq x - 1$ for $x > 0$.

The interest of this note is twofold. First, originality of the method: starting from the simple observation that the sequence of renormalized sums of i.i.d. random variables $(X_i)_{i\geq 1}$ is a non-homogeneous Markov chain, we import to the non-homogeneous framework spectral methods used for the asymptotic study of time-homogeneous and reversible Markov processes. Second, we show that the rate of convergence for the χ_2-distance between the renormalized sums and normal distribution μ improves by a factor $\sqrt{n}$ for each supplementary moment of X_1 that agrees with the corresponding moment of μ. This echoes similar behaviour for total variation distance and relative entropy (Bally and Caramellino [2016]; Bobkov et al. [2013]). With our method, the desired behaviour is naturally embedded in the proof, i.e. it is not harder to get the optimal rate of convergence for an arbitrary number of matching moments than for two moments.

For f, φ two density functions in $L^2(\mu)$ and a, $b \in \mathbb{R}$, we call $af * b\varphi$ the density of the random variable $aU + bV$ where U has density f and V density φ. At the heart of our study is a formula of the form:

$$\chi_2\left(af * \sqrt{1-a^2}\varphi\right) \leq a\big(1 + d_\varphi(a)\big)\,\chi_2(f) + \sqrt{1-a^2}\,\chi_2(\varphi), \qquad \lim_{a\to 1} d_\varphi(a) = 0,$$

which describes the evolution of the χ_2-distance under barycentric convolution. More precisely, set $(\bar{H}_n)$ the set of renormalized Hermite polynomials, forming an orthonormal basis of $L^2(\mu)$:

$$\bar{H}_n(x) := \frac{(-1)^n}{\sqrt{n!}} e^{\frac{x^2}{2}} D^n\left(e^{-\frac{x^2}{2}}\right), \qquad x \in \mathbb{R}, \quad n \in \mathbb{Z}^+,$$

where D denotes the derivation operator acting on smooth functions from $\mathbb{R}$ to $\mathbb{R}$. We show the following result:

Theorem 3.1.1 (Barycentric convolution and χ_2)**.** *Let r be a natural integer and f, φ be two densities in $L^2(\mu)$ whose moments match the moments of μ up to order $r \in \mathbb{Z}^+$, and assume moreover that the density φ is polynomial. In particular, φ admits a decomposition on the Hermite basis of the form:*

$$\varphi = 1 + \sum_{k=r+1}^{N} \varphi_k \bar{H}_k. \tag{3.2}$$

Set:

$$a_\varphi := \left(1 + \frac{N}{r+1}\right)^{-\frac{1}{4}} \in (0, 1],$$

and for all $a \in (0, 1)$,

$$d_\varphi(a) := \sum_{k=r+1}^{N} \frac{|\varphi_k|}{\sqrt{k!}} \left(-2(r+1+N)\left(1 + \frac{N}{r+1}\right)\log a\right)^{k/2}.$$

*If φ satisfies to Hypothesis **(H)** stated below, then for all $a \in (a_\varphi, 1)$, the following inequality stands:*

$$\chi_2\left(af * \sqrt{1-a^2}\varphi\right) \leq a^{r+1}\left(1 + d_\varphi(a)\right)\chi_2(f) + (1-a^2)^{\frac{r+1}{2}}\chi_2(\varphi). \tag{E_r}$$

Moreover,

$$\lim_{a\to 1} |\log a|^{-\frac{r+1}{2}} d_\varphi(a) < +\infty.$$

Equation (3.2), which involves Hermite coefficients $(\varphi_k)_{r+1\leq k\leq N}$, is explained in Section 3.2. Hypothesis **(H)** relates to theHermite coefficients and is stated in Section 3.2.

The respective roles of f and φ are not symmetric in inequality (E_r). However, setting $a = 1-h$, the term $d_\varphi(1-h)$ is of order $h^{(r+1)/2}$ while the prefactor a^{r+1} writes $a^{r+1} = 1-(r+1)h+o(h^2)$. Hence, for a polynomial φ, this asymmetry is negligible as $a \to 1$ if (and only if) $r \geq 2$. Let us also mention that there exists an alternative inequality, holding without the polynomial assumption, stated in Section 3.4.

Bound (E_r) bears a similarity to Shannon-Stam inequality for (absolute) entropy (Shannon and Weaver [1949]; Stam [1959]):

$$\mathrm{Ent}\left(a\,U + \sqrt{1-a^2}V\right) \leq a\,\mathrm{Ent}(U) + \sqrt{1-a^2}\,\mathrm{Ent}(V), \qquad a \in [0,1], \tag{3.3}$$

although, by the observation above, the coefficients in front of Ent are of different order than the coefficients in front of χ for all natural integer r.

Bound (E_r) also compares to a Poincaré inequality for the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$, defined for $f \in L^2(\mu)$ as

$$P_t[f](x) = \mathbb{E}\left[f\left(e^{-t}x + \sqrt{1-e^{-2t}}Z\right)\right], \quad Z \sim \mu; \qquad x \in \mathbb{R}, \quad t \geq 0. \tag{3.4}$$

Indeed, in the case where φ is the density of the normal distribution μ, that is to say $\varphi = 1$, we adopt the convention $r = +\infty$ and $N = 0$ in equality 3.2, and set $a_\varphi = 0$ and $d_\varphi(a) = 0$ for all $a \in [0,1]$. As we will see in Section 3.4, (E_1) then corresponds (up to a positivity assumption which can actually be discarded in the proof) to Poincaré inequality for $(P_t)_{t\geq 0}$: for $f \in L^2(\mu)$,

$$\mathrm{Var}_\mu(P_t f) \leq e^{-2t}\,\mathrm{Var}_\mu(f), \qquad t \geq 0. \tag{3.5}$$

Let us turn to the intended application of (E_r) to the framework of the Central Limit Theorem. Let $(X_i)_{i\in\mathbb{Z}^+}$ be a sequence of independent identically distributed variables with density $\varphi \in L^2(\gamma)$ such that $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$ and let f_n be the density of the renormalized sum:

$$Y_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i, \qquad n \geq 1.$$

By the centering and normalizing assumption $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$, the integer r above is greater than 2. Relation (E_r) implies then the following asymptotic bound on the χ_2-distance between $(Y_n)_{n\geq 1}$ and the normal distribution μ:

Theorem 3.1.2 (Asymptotic bound on χ_2)**.** *If the moments of X_1 and the moments of μ match up to order r for a given integer $r \geq 2$, and if the density φ of X_1 with respect to μ is polynomial and satisfies to Hypothesis **(H)** below, then*

$$\limsup_{n\to+\infty} \quad n^{\frac{r-1}{2}}\,\chi_2(f_n) < +\infty.$$

The rate in Theorem 3.1.2 is optimal, as is seen by comparison with the results of the literature. The first quantification result for the convergence of renormalized sums of i.i.d. variables was obtained independently by Berry and Esseen through Kolmogorov distance (Berry [1941]; Esseen [1945]). Rate of convergence in total variation distance is first addressed in Siraždinov and Mamatov [1962] and further studied in Bally and Caramellino [2016] using Malliavin's calculus. They show that under a regularity assumption, if X_1 has moments of order $r+1$ and if the moments up to order r agree with the moments of μ, then there exists a constant c depending on $\mathbb{E}[|X|^r]$ such that

$$d_{TV}(f_n, \mu) \quad \leq \quad c n^{-\frac{r-1}{2}},$$

and they prove that the rate $n^{-\frac{r-1}{2}}$ is optimal, implying in turn optimality of the rate in Theorem 3.1.2 by inequality (3.1).

The rate of convergence in Theorem 3.1.2 relies crucially on the fact that the inequality (E_r) incorporates the matching moments assumption through exponent r on the barycentric coefficients. By comparison, Shannon-Stam inequality (3.3) implies (jointly with a result of monotonicity of relative entropy and Fisher information under convolution) the convergence $\mathrm{Ent}(f_n) \to 0$ without rate (Brown [1982]; Barron [1986]). The optimal rate is derived in Artstein et al. [2004] when the random variable X_1 satisfies to a Poincaré inequality. Bobkov et al. [2013] provides an asymptotic expansion of entropy involving the moments of X_1. Similar developments occured for Fisher information (Brown [1982], Bobkov et al. [2014]).

Other distances include Sobolev (Goudon et al. [2002]) and Wasserstein (Ibragimov [1966]; Tanaka [1973]; Rio [2009, 2011]; Dedecker et al. [2009]) distances. A typical assumption in Berry-Esseen theorems is the existence of moments up to a certain order for the random variable X_1. In the present framework, the fact that φ is in $L^2(\mu)$ implies that moments of all order exist. This is consistent with the fact that the χ_2-distance bounds by above the usual quantities (as shown in (3.1); a similar inequality holds for Wasserstein distance of order 1).

In another direction of research, Stein's method and Malliavin calculus revealed to be powerful tools to study, including in a quantitative way, the asymptotic normality of multidimensional random variables living in Gaussian chaoses. Let us cite, among an increasingly rich literature, the reference book Nourdin and Peccati [2012]. It is interesting to remark the formal similarity between their objects and ours, although the results do not compare. Indeed, in the Stein-Malliavin framework, the typical random variable X_1 writes $X_1 = \varphi(Z)$, with Z a Gaussian random variable living in $\mathbb{R}^n$ with law μ_n and $\varphi \in L^2(\mu_n)$. To compare to our framework, let us take $n = 1$ and φ a density in $L^2(\mu)$. The random variable $X_1 = \varphi(Z)$, where $Z \sim \mu$, bears no relation with the random variable X_1 which has *density* φ with respect to μ; hence, the two approaches are not reducible one to another. The authors would like to thank Anthony Réveillac for interesting discussions on this subject.

The remaining of this paper is organized as follows. In Section 3.2, Hermite-Fourier decomposition (3.2) is detailed, and Hypothesis **(H)** is stated and commented. Section 3.3 is devoted to the explicit expression of the convolution operator and of its Hermite-Fourier decomposition. Theorem 3.1.1 is proved in Section 3.4 by taking advantage of the Markovian nature of the sequence of sums. Finally Section 3.5 is devoted to the proof of Theorem 3.1.2.

3.2 Hermite-Fourier decomposition of the density

First, let us introduce some notation. The symbol $\mathbf{1}$ stands for the function from $\mathbb{R}$ to $\mathbb{R}$ identically equal to 1, $\mathbb{Z}^+$ for the set of natural integers and $\binom{n}{k}$ for the binomial coefficient associated to natural integers $k \leq n$. For a fonction $f \in L^1(\mu)$, we denote indifferently

$$\mu(f) = \int f d\mu.$$

The space $L^2(\mu)$ is a Hilbert space, with scalar product and associated norm defined as

$$\langle f, g \rangle_{L^2(\mu)} := \int fg \, d\mu, \qquad \|f\|_{L^2(\mu)} := \sqrt{\mu(f^2)}, \qquad f, g \in L^2(\mu).$$

Set

$$\mathrm{Var}_\mu(f) := \int (f - \mu(f))^2 d\mu.$$

Hermite polynomials $(H_n)_{n\in\mathbb{Z}^+}$ are defined as:

$$H_n(x) := (-1)^n e^{\frac{x^2}{2}} D^n \left(e^{-\frac{x^2}{2}}\right), \qquad x \in \mathbb{R}, \quad n \in \mathbb{Z}^+,$$

where we recall that D stands for the derivation operator acting on smooth functions from $\mathbb{R}$ to $\mathbb{R}$. Hermite polynomials are also characterized by the following equation: for all smooth functions $f : \mathbb{R} \to \mathbb{R}$,

$$\int f H_n d\mu = \int D^n f d\mu. \tag{3.6}$$

They form an orthogonal basis of $L^2(\mu)$: for all $n, m \in \mathbb{Z}^+$,

$$\int H_m H_n d\mu = n!\, \delta_{n,m}.$$

In the paper it is more convenient to work with renormalized Hermite polynomials $(\bar{H}_n)_{n\in\mathbb{N}}$:

$$\bar{H}_n = \frac{1}{\sqrt{n!}} H_n, \qquad n \in \mathbb{Z}^+,$$

which form an orthonormal basis of $L^2(\mu)$. By convention set $\bar{H}_{-1} = 0$. One has:

$$D\bar{H}_n = \sqrt{n}\bar{H}_{n-1}, \qquad n \in \mathbb{Z}^+. \tag{3.7}$$

As $H_0 = 1$, for all natural integer $\bar{H}_n$ is of degree n. The basis $(\bar{H}_n)_{n\in\mathbb{Z}^+}$ is diagonal for the Ornstein-Uhlenbeck semigroup defined in (3.4):

$$P_t[\bar{H}_n] = e^{-nt}\bar{H}_n, \qquad n \in \mathbb{Z}^+, \quad t \geq 0.$$

For all functions $g \in L^2(\gamma)$, call $(g_k)_{k\in\mathbb{Z}^+}$ its coefficients on the orthonormal Hermite basis,

$$g = \sum_{k\in\mathbb{Z}^+} g_k \bar{H}_k,$$

where the equality stands in $L^2(\mu)$, and denote indifferently $\mathcal{F}(g) := \overrightarrow{g} = (g_k)_{k\in\mathbb{Z}^+}$ the sequence of its coefficients, which belongs to the Hilbert space l^2, defined as the set of real sequences $(u_k)_{k\in\mathbb{R}}$ such that $\sum_{k\in\mathbb{Z}^+} u_k^2 < +\infty$. The application

$$L^2(\mu) \to l^2, \quad g \mapsto \mathcal{F}g, \tag{3.8}$$

is an isometry of Hilbert spaces.

If $\varphi \in L^2(\mu)$ is a density, then $\varphi_0 = 1$. The matching moments assumption has a nice interpretation in terms of the coefficients: for all positive integer r,

$$\left(\forall k \in \{1, \dots r\}, \quad \int x^k \varphi(x) d\mu(x) = \int x^k d\mu(x)\right) \quad \Leftrightarrow \quad (\forall k \in \{1, \dots r\}, \quad \varphi_k = 0).$$

Indeed, Hermite polynomial $\bar{H}_n$ being of degree n for all natural integers, one has the equivalence

$$\begin{aligned} &\left(\forall k \in \{1, \dots r\}, \quad \int x^k \varphi(x) d\mu(x) = \int x^k d\mu(x)\right) \\ \Leftrightarrow \quad &\left(\forall k \in \{1, \dots r\}, \quad \int \bar{H}_k(x) \varphi(x) d\mu(x) = \int \bar{H}_k(x) d\mu(x)\right), \end{aligned}$$

and by orthogonality of $(\bar{H}_n)_{n\in\mathbb{Z}^+}$ it stands that for all $k \in \mathbb{Z}^+$,

$$\int \bar{H}_k d\mu = \int \bar{H}_k \bar{H}_0 d\mu = \delta_{0,k}.$$

The assumption that φ is a polynomial density whose moments agree with moments of μ up to r hence amounts to:

$$\exists N \in \mathbb{Z}^+, \quad N > r, \quad \varphi = 1 + \sum_{k=r+1}^{N} \varphi_k \bar{H}_k,$$

which corresponds to equality (3.2) above. When φ is the density of μ itself, that is to say $\varphi = \bar{H}_0 = 1$, we set $K = +\infty$ and $N = 0$ by convention.
Let us introduce the quantities

$$C_k := \left(1 + \frac{N}{K}\right)^{k/2}, \qquad \gamma_k = \frac{1}{\sqrt{k!}} C_k |\varphi_k|, \qquad k \in \mathbb{Z}^+.$$

We are now ready to state Hypothesis **(H)**, which is composed of two parts, **(H1)** and **(H2)** as follows.
(H1) If $K \leq N-2$, for all $K \leq k \leq N-2$,

$$(k+2)\gamma_{k+2} \leq \gamma_k.$$

For non-vanishing φ_k this relation is equivalent to

$$\left|\frac{\varphi_{k+2}}{\varphi_k}\right| \leq \frac{1}{1+\frac{N}{K}} \left(1 - \frac{1}{k+2}\right)^{1/2}$$

if $\varphi_k \neq 0$ and is implied by the simplest assumption
(H1') If $K \leq N-2$, for all $K \leq k \leq N$,

$$|\varphi_{k+1}| \leq r\,|\varphi_k|\,; \qquad r := \left(\left(1 - \frac{1}{K+2}\right)^{1/2} \frac{1}{1+\frac{N}{K}}\right)^{1/2}.$$

Remark that assumption **(H1)** implies that $\gamma_k = 0 \Rightarrow \gamma_{k+2} = 0$.
The second condition **(H2)** has two parts:
(H2a) If $K \leq N-1$,

$$\left(\frac{2(K+1)\gamma_{K+1}}{N}\right)^N \left(\frac{K-1}{2\gamma_N}\right)^{K-1} \vee \left(\frac{2K\gamma_K}{N-1}\right)^{N-1} \left(\frac{K-2}{2\gamma_{N-1}}\right)^{K-2} \leq \frac{1}{(N-K+1)^{N-K+1}}.$$

(H2b) If $K = N$,

$$\gamma_N \leq \frac{1}{2(N-2)^{(N-2)/2}}.$$

Loosely speaking, assumption **(H1)**, which is present only if $K \leq N-2$, amounts to ask a geometric decrease for the coefficients (φ_k). If $K = N$ or $K = N-1$, assumption **(H2)** requires the leading coefficients φ_N and φ_{N-1} to be not too big, which is fair enough; it is more painful to write when $K < N-1$, but it can be interpreted as the requirement that the coefficients $(\varphi_k)_{K\leq k\leq N}$ do not decay too fast.
We conclude this section by the following comment on the range of validity of Theorems 3.1.1 and 3.1.2.

Remark 3.2.1 (On the polynomial assumption)**.** *We conjecture that inequality (E_r) holds without Hypothesis **(H)**: indeed, the fact that φ, as a density, is nonnegative already implies restrictions on the coefficients, which are in fact sufficient to prove (E_r) in the case of densities of the form $\varphi = 1 + c\bar{H}_2$ and $\varphi = 1 + c'\bar{H}_4$.*
Whether the polynomial assumption is necessary is less clear. For comparison, the hypothesis of Gaussian chaos of finite orders is needed in Nourdin and Poly [2013].

3.3 Explicit expression of the convolution operator

3.3.1 Convolution as a Markovian transition

Set $(X_i)_{i\geq 1}$ random variable of density $\varphi \in L^2(\mu)$. Throughout the paper, notation f_n stands for the density of the renormalized sum

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i, \qquad n \geq 1,$$

which satisfy to the recursion equation

$$Y_{n+1} = \sqrt{1 - \frac{1}{n+1}} Y_n + \frac{1}{\sqrt{n+1}} X_{n+1}, \quad n \geq 1. \tag{3.9}$$

For a parameter $a \in [0, 1]$, introduce the bilinear barycentric convolution operator K_a defined as

$$\forall f, \varphi \in L^2(\mu), \qquad K_a(f\,, \varphi) := af * \sqrt{1-a^2}\,\varphi.$$

Equation (3.9) in turn yields the corresponding recursion relation for the successive densities:

$$f_{n+1} = K_{a_{n+1}}(f_n, \varphi), \qquad a_{n+1} := \sqrt{1 - \frac{1}{n+1}}, \quad n \geq 1. \tag{3.10}$$

On the other hand, relation (3.9) translates into the the fact that $(Y_n)_{n\geq 1}$ is an inhomogeneous Markov chain. The associated semigroup $(\mathcal{Q}_{p,q})_{q\geq p\geq 1}$ is defined for continuous bounded functions as:

$$\mathcal{Q}_{p,q}[f](x) := \mathbb{E}[f(Y_q)|Y_p = x], \qquad x \in \mathbb{R}, \quad q \geq p \geq 1.$$

By relation (3.9), the explicit expression of $\mathcal{Q}_{n,n+1}$ is straightforward: for all $f \in L^2(\mu)$,

$$\mathcal{Q}_{n,n+1}[f] = Q_{a_{n+1},\varphi}[f],$$

where the operator $Q_{a,\varphi}$ is defined for all $a \in [0, 1]$ and density $\varphi \in L^2(\mu)$ as:

$$Q_{a,\varphi}[f](x) := \int f(ax + \sqrt{1-a^2}y)\varphi(y)dy, \qquad x \in \mathbb{R}.$$

Denote $\mathcal{Q}^*_{n,n+1}$ (resp. $Q^*_{a,\varphi}$) the adjoint of $\mathcal{Q}_{n,n+1}$ (resp. $Q_{a,\varphi}$) in $L^2(\mu)$. The discrete version of Kolmogorov backward relation reads:

$$f_{n+1} := \mathcal{Q}^*_{n,n+1}[f_n], \qquad n \geq 1.$$

Hence $\mathcal{Q}^*_{n,n+1}[f_n] = K_{a_{n+1}}(f_n, \varphi)$. More generally, for f, φ densities in $L^2(\mu)$, the following identity holds:

$$K_a(f, \varphi) = Q^*_{a,\varphi}[f], \qquad a \in [0, 1].$$

The strategy used in the paper relies on this interpretion of convolution as action of a Markovian transition, as explained in Section 3.4.1.

Proposition 3.3.1 (Explicit expression of the operators)**.** *For all $a \in [0, 1]$, one has*

$$Q_{a,\varphi}[f](x) = \int f(ax + \sqrt{1-a^2}y)\varphi(y)d\mu(y), \tag{3.11}$$

$$Q^*_{a,\varphi}[f](x) = \int f(ax - \sqrt{1-a^2}y)\varphi(\sqrt{1-a^2}x + ay)d\mu(y) = K_a(f, \varphi)(x). \tag{3.12}$$

The formulas are to be understood in the following sense: if f is bounded and continuous, they stand for all $x \in \mathbb{R}$; if $f \in L^2(\mu)$, they stand in the almost everywhere sense. One sees that the operators $Q_{a,\varphi}$ and $Q^*_{a,\varphi}$ are actually defined for all $f, \varphi \in L^2(\gamma)$ independently from them being densities; in what follows, $Q_{a,\varphi}$ and $Q^*_{a,\varphi}$ refers to this extended definition when required. Furthermore, $Q_{a,\varphi}$ and $Q^*_{a,\varphi}$ are bounded in $L^2(\mu)$ for all $\varphi \in L^2(\mu)$: indeed by Cauchy-Schwarz inequality, for all $f \in L^2(\mu)$,

$$\int (Q_{a,\varphi}[f])^2 \, d\mu \leq \|f\|^2_{L^2(\mu)} \|\varphi\|^2_{L^2(\mu)}, \qquad \int \left(Q^*_{a,\varphi}[f]\right)^2 d\mu \leq \|f\|^2_{L^2(\mu)} \|\varphi\|^2_{L^2(\mu)}.$$

Proof of Proposition 3.3.1. Formula (3.11) has already been given; let us prove formula (3.12). As $a \in [0,1]$, there exists $\theta \in \mathbb{R}$ such that $\cos\theta = a$ and $\sin\theta = \sqrt{1-a^2}$. Denote R_θ the rotation of $\mathbb{R}^2$ with parameter θ and Γ the Gaussian distribution on $\mathbb{R}^2$ with identity as covariance matrix. Then, invariance of Γ under the action of R_θ implies that for all $f, g \in L^2(\gamma)$,

$$\begin{aligned}
\int f Q_{a,\varphi}[g] d\mu &= \iint f(x)\varphi(y) g((\cos\theta)x + (\sin\theta)y) d\mu(x) d\mu(y) \\
&= \iint f(\Re(X))\varphi(\Im(X)) g(\Re(R_\theta X)) d\Gamma(X) \\
&= \iint f(\Re(R_{-\theta}X))\varphi(\Im(R_{-\theta}X)) g(\Re(X)) d\Gamma(X) = \int g Q^*_{a,\varphi}[f] d\mu.
\end{aligned}$$

□

Remark 3.3.2 (Link with Ornstein-Uhlenbeck semigroup)**.** *Recall the definition of the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ given in* (3.4)*. Hence, for all $a \in [0,1]$ and $f, \varphi \in L^2(\mu)$, one has the two useful equalities:*

$$Q^*_{a,\mathbf{1}}[f] = K_a(f, \mathbf{1}) = P_{-\log a}[f], \qquad Q^*_{a,\varphi}[\mathbf{1}] = K_a(\mathbf{1}, \varphi) = P_{-\frac{1}{2}\log(1-a)^2}[\varphi].$$

Let us make a brief aside on the article Fill [1991], which focuses on homogeneous Markov chains living on $\{1, \dots, n\}$, for a positive integer n. The transition matrix is called Q and the chain is supposed invariant with respect to a probability measure θ on $\{1, \dots, n\}$. Calling Q^* the matrix representing the adjoint of Q in $L^2(\theta)$, Fill defines the *multiplicative reversibilization* of Q as $M(Q) = QQ^*$. The matrix $M(Q)$ is the transition matrix of a Markov homogeneous chain which is now reversible with respect to θ, and Fill [1991] shows how to control the asymptotic behaviour of the Markov chain associated to Q with spectral analysis of $M(Q)$.

In the present case, it reveals useful to introduce the multiplicative reversibilization of the Markov transition operator $Q_{a,\varphi}$, defined as

$$M_{a,\varphi} := Q_{a,\varphi} Q^*_{a,\varphi}. \tag{3.13}$$

This new operator is now symmetric in $L^2(\mu)$, though in the general case, it is not Markovian: as $Q_{a,\varphi}$ is not invariant, then $Q^*_{a,\varphi}[\mathbf{1}] \neq \mathbf{1}$ and the mass conservation property $M_{a,\varphi}[\mathbf{1}] = \mathbf{1}$ does not hold. Nonetheless, we will see in Section 3.4.3 that spectral analysis of $M_{a,\varphi}$ gives quantitative information on the action of Markovian transition $Q_{a,\varphi}$ and convolution operator $Q^*_{a,\varphi}$.

3.3.2 Hermite-Fourier decomposition of the convolution operator

Let us now determine how the operators $Q_{a,\varphi}$, $Q^*_{a,\varphi}$ and $M_{a,\varphi}$ act with respect to the Hermite-Fourier decomposition defined in (3.8). For a bounded operator Q of $L^2(\mu)$, we call $\overrightarrow{Q}$ the infinite matrix defined as:

$$\overrightarrow{Q} := (\overrightarrow{Q}(m,n))_{n,m\in\mathbb{Z}^+}; \qquad (\overrightarrow{Q})(m,n) := \int Q(\bar{H}_n)\bar{H}_m d\mu, \qquad n, m \in \mathbb{Z}^+.$$

The matrix $\overrightarrow{Q}$ is the unique bounded operator of l^2 such that

$$\mathcal{F}(Q[f]) = \overrightarrow{Q}\overrightarrow{f}, \qquad f \in L^2(\mu).$$

Set $\|R\|_{\mathrm{op}}$ the operator norm of a bounded operator R on a Hilbert space $\mathcal{H}$, defined with evident notation as

$$\|R\|_{\mathrm{op}} := \sup_{h\in\mathcal{H}\setminus\{0\}} \frac{\|Rh\|_{\mathcal{H}}}{\|h\|_{\mathcal{H}}}.$$

The following property reveals useful: for all bounded operator Q on $L^2(\mu)$, it stands that:

$$\|Q\|_{\mathrm{op}} = \|\overrightarrow{Q}\|_{\mathrm{op}}. \tag{3.14}$$

In the following proposition, we give the matrices $\overrightarrow{Q}_{a,\varphi}$, $\overrightarrow{Q^*}_{a,\varphi}$ and $\overrightarrow{M}_{a,\varphi}$ associated to the operators $Q_{a,\varphi}$, $Q^*_{a,\varphi}$ and $M_{a,\varphi}$.

Proposition 3.3.3 (Matrix form of operators). *For all $\varphi \in L^2(\mu)$ and $a \in [0,1]$, one has:*

$$\forall m,n \in \mathbb{Z}^+, \quad \overrightarrow{Q}_{a,\varphi}(m,n) = \begin{cases} \binom{n}{m}^{\frac{1}{2}} a^m \left(1-a^2\right)^{\frac{n-m}{2}} \varphi_{n-m}, & m \le n \\ 0, & m > n, \end{cases} \tag{3.15}$$

Denoting ${}^TN$ the transpose matrix of N, it stands that:

$$\overrightarrow{Q^*}_{a,\varphi} = {}^T\overrightarrow{Q}_{a,\varphi}.$$

Finally, the matrix $\overrightarrow{M}_{a,\varphi}$ is symmetric and

$$\forall l,i \in \mathbb{Z}^+,\ i \le l, \quad \overrightarrow{M}_{a,\varphi}(l,l-i) = a^{2l-i} \sum_{k\ge 0} \binom{k+l}{k}^{1/2} \binom{k+l}{k+i}^{1/2} (1-a^2)^{\frac{2k+i}{2}} \varphi_{k+i}\varphi_k. \tag{3.16}$$

Proof. To begin with, one needs to compute the Hermite-Fourier decomposition of $Q_{a,\bar{H}_m}[\bar{H}_n]$, for $a \in [0,1]$ and $n,m \in \mathbb{Z}^+$. Applying the properties of Hermite polynomials recalled in Section 3.2 yields:

$$Q_{a,\bar{H}_m}[\bar{H}_n] = \int \bar{H}_n(ax+\sqrt{1-a^2}y)\bar{H}_m(y)d\mu(y) = \frac{1}{\sqrt{n!}\sqrt{m!}} \int D^m(H_n(ax+\sqrt{1-a^2}y))d\mu(y).$$

By the degree property, the integral vanishes for $n < m$. For $n \ge m$,

$$\begin{aligned} Q_{a,\bar{H}_m}[\bar{H}_n] &= \left(1-a^2\right)^{\frac{m}{2}} \frac{n\cdots(n-m+1)}{\sqrt{n!}\sqrt{m!}} \int H_{n-m}(ax+\sqrt{1-a^2}y))d\mu(y) \\ &= \left(1-a^2\right)^{\frac{m}{2}} \frac{n\cdots(n-m+1)}{\sqrt{n!}\sqrt{m!}} P_{-\log a}[H_{n-m}] \\ &= a^{n-m}\left(1-a^2\right)^{\frac{m}{2}} \frac{n\cdots(n-m+1)}{\sqrt{n!}\sqrt{m!}} H_{n-m} \\ &= a^{n-m}\left(1-a^2\right)^{\frac{m}{2}} \binom{n}{m}^{\frac{1}{2}} \bar{H}_{n-m}. \end{aligned}$$

By bilinearity, write

$$\begin{aligned} Q_{a,\varphi}(\bar{H}_n) &= \sum_{m\in\mathbb{Z}^+} \varphi_m Q_{a,\bar{H}_m}[\bar{H}_n] = \sum_{m=0}^{n} \varphi_m a^{n-m}\left(1-a^2\right)^{\frac{m}{2}} \binom{n}{m}^{\frac{1}{2}} \bar{H}_{n-m} \\ &= \sum_{m=0}^{n} \varphi_{n-m} a^m \left(1-a^2\right)^{\frac{n-m}{2}} \binom{n}{m}^{\frac{1}{2}} \bar{H}_m = \sum_{m=0}^{n} \overrightarrow{Q}_{a,\varphi}(m,n)\bar{H}_m, \end{aligned}$$

which proves (3.15). Furthermore, definition (3.13) implies that $\overrightarrow{M}_{a,\varphi} = \overrightarrow{Q}_{a,\varphi}{}^T\overrightarrow{Q}_{a,\varphi}$, and formula (3.16) follows by simple computation. $\square$

Remark 3.3.4 (On Hermite decomposition of the convolution operator)**.**

— *It is to be noted that convolution with barycentric coefficients only admits a nice decomposition; contrarily to what happens with Fourier transform associated to Lebesgue measure, the usual convolution has no explicit Hermite-Fourier representation.*

— *Thanks to formula* (3.15) *above, one finds that for all* $\varphi \in L^2(\mu)$*,* $m \in \mathbb{Z}^+$ *and* $a \in [0,1]$*,*

$$Q^*_{a,\varphi}(\bar{H}_m) = a^m \sum_{n \geq 0} \binom{m+n}{n}^{1/2} (1-a^2)^{\frac{n}{2}} \varphi_n \bar{H}_{m+n},$$

which allows to better understand the behaviour of the barycentric convolution: each non-vanishing coefficient on $\bar{H}_m$ *and* $\bar{H}_n$ *in the respective decompositions of* f *and* φ *contribute to a coefficient on* $\bar{H}_{m+n}$ *in the decomposition of* $K_a(f,\varphi)$*.*

— *If* φ *is polynomial, then* $\overrightarrow{M}_{a,\varphi}$ *is a band matrix.*

We already noticed that $Q_{a,\varphi}$ and $Q^*_{a,\varphi}$, and by composition $M_{a,\varphi} = Q_{a,\varphi}Q^*_{a,\varphi}$, are bounded operators. In fact, they are Hilbert-Schmidt operators. By definition, a bounded operator R on the Hilbert space $\mathcal{H}$ is Hilbert-Schmidt, if, $(e_n)_{n\in\mathbb{Z}^+}$ standing for an orthonormal basis of $\mathcal{H}$, one has:

$$\sum_{n\in\mathbb{Z}^+} \|R(e_n)\|^2_{\mathcal{H}} < +\infty.$$

Proposition 3.3.5 (Hilbert-Schmidt operators)**.** *For all* $\varphi \in L^2(\mu)$ *and* $a \in [0,1)$*, the operators* $Q_{a,\varphi}, Q^*_{a,\varphi}, M_{a,\varphi}$ *are Hilbert-Schmidt, hence compact.*

Proof. Consider first $Q^*_{a,\varphi}$.

$$\sum_{n\in\mathbb{Z}^+} \|Q^*_{a,\varphi}(\bar{H}_n)\|^2_{L^2(\mu)} = \sum_{n\in\mathbb{Z}^+} \langle \bar{H}_n, M_{a,\varphi}\bar{H}_n\rangle_{L^2(\mu)} = \sum_{n\in\mathbb{Z}^+} \overrightarrow{M}_{a,\varphi}(n,n) = \sum_{n,k\in\mathbb{Z}^+} \binom{k+n}{k}(1-a^2)^k a^{2n}\varphi_k^2.$$

Now, by the equality

$$\sum_{n\in\mathbb{Z}^+} \binom{k+n}{k} u^n = 1/(1-u)^{k+1}, \qquad k \in \mathbb{Z}^+, \quad u \in [0,1),$$

we find that

$$\sum_{n\in\mathbb{Z}^+} \|Q^*_{a,\varphi}(\bar{H}_n)\|^2_{L^2(\mu)} = \sum_{k\in\mathbb{Z}^+} \frac{(1-a^2)^k}{(1-a^2)^{k+1}}\varphi_k^2 = \frac{1}{1-a^2}\|\varphi\|^2_{L^2(\mu)} < +\infty.$$

This implies that $Q_{a,\varphi}$ is Hilbert-Schmidt and in turn so is $M_{a,\varphi}$ by composition. □

3.4 Proof of Theorem 3.1.1

3.4.1 Strategy for inhomogeneous Markov chains

Let us now explain the strategy to exploit the Markovian framework. In Remark 3.3.2, we noticed that if $\varphi = \mathbf{1}$, *i.e.* the X_i's are normal, then $Q^*_{a,\mathbf{1}} = P_{-\log a}$. In this case, the renormalized sums $(Y_n)_{\geq 1}$ are also normal, which corresponds to the fact that the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ is invariant, and in fact reversible, with respect to μ. The semigroup $(P_t)_{t\geq 0}$ also enjoys a Poincaré inequality recalled in equation (3.5). If f is a density then $\mathrm{Var}_\mu(f) = \chi^2(f)$, hence Poincaré inequality for the Ornstein-Uhlenbeck semigroup reads:

$$\chi_2(P_t[f]) \leq e^{-t}\,\chi_2(f), \qquad t \geq 0.$$

Furthermore, $f_{n+1} = P_{-\log a_{n+1}}[f_n]$ by (3.10), hence:

$$\chi_2(f_{n+1}) = \chi_2\left(P_{-\log a_{n+1}}[f_n]\right) \leq a_n\,\chi_2(f_n), \qquad n \geq 1,$$

and by straighforward calculation one gets the decrease of $\chi_2(f_n)$.
The idea underlying our method consists in mimicking the reasoning above for the true operator $Q^*_{a,\varphi}$ acting on densities, which is neither reversible nor satisfies to the mass conservation property in the general case, as was explained above. For $a \in [0,1]$ and f a density in $L^2(\mu)$, let us write by triangular inequality:

$$\chi_2(Q^*_{a,\varphi}[f]) = \left\|Q^*_{a,\varphi}[f] - 1\right\|_{L^2(\mu)} \leq \left\|Q^*_{a,\varphi}[f-1]\right\|_{L^2(\mu)} + \left\|Q^*_{a,\varphi}[\mathbf{1}] - 1\right\|_{L^2(\mu)}.$$

The term

$$\left\|Q^*_{a,\varphi}[\mathbf{1}] - 1\right\|_{L^2(\mu)},$$

can be thought of as a measure of the divergence from invariance of the transition operator, an idea tracing back to Arnaudon and Miclo [2016].
Second, the centered term $\left\|Q^*_{a,\varphi}[f-1]\right\|_{L^2(\mu)}$ rewrites:

$$\left\|Q^*_{a,\varphi}[f-1]\right\|^2_{L^2(\mu)} = \int \left(Q^*_{a,\varphi}[f-1]\right)^2 d\mu = \int (f-1)Q_{a,\varphi}Q^*_{a,\varphi}[f-1]d\mu = \int (f-1)M_{a,\varphi}[f-1]d\mu,$$

making appear the multiplicative reversibilization $M_{a,\varphi}$ of $Q^*_{a,\varphi}$ introduced above.
In terms of convolution, this amounts to consider separately $K_a(f-1,\varphi)$ and $K_a(\mathbf{1},\varphi)$.
Following this roadmap, Proposition 3.4.1 below deals with the default of invariance $\|Q^*_{a,\varphi}[\mathbf{1}] - 1\|_{L^2(\mu)}$ and Proposition 3.4.2 with the centered quantity $\|Q^*_{a,\varphi}[f-1]\|_{L^2(\mu)}$. Theorem 3.1.1 is then proved in Section 3.4.4, and we conclude the part by stating an alternative bound to (E_r) in Section 3.4.5.

3.4.2 Improved Poincaré inequality for Ornstein-Uhlenbeck

The following result is an improvement of the usual Poincaré inequality for the Ornstein-Uhlenbeck semigroup (3.5) when more information is avalaible on the function $f \in L^2(\mu)$ at play.

Proposition 3.4.1 (Improved Poincaré)**.** *Let $r \in \mathbb{Z}^+$ and f be a function in $L^2(\mu)$ with Hermite decomposition of the form*

$$f = f_0 + \sum_{n \geq r+1} f_k \bar{H}_k,$$

Then for all $t \geq 0$,

$$Var_\mu(P_t f) \leq e^{-2(r+1)t}\,Var_\mu(f).$$

In particular, if φ is the density of a variable agreeing with the Gaussian moments up to r, then for all $a \in [0,1]$,

$$\|Q^*_{a,\varphi}[\mathbf{1}] - 1\|_{L^2(\mu)} \leq (1-a^2)^{\frac{r+1}{2}}\chi_2(\varphi).$$

Proof. Thanks to the properties of Hermite polynomials,

$$P_t f = \sum_{k=0}^{\infty} f_k P_t[\bar{H}_k] = h_0 + \sum_{k=r+1}^{\infty} f_k e^{-kt}\bar{H}_k;$$

$$\mathrm{Var}_\mu(P_t f) = \sum_{k=r+1}^{\infty} f_k^2 e^{-2kt} \leq e^{-2(r+1)t} \sum_{k=r+1}^{\infty} f_k^2 = e^{-2(r+1)t}\mathrm{Var}_\mu(f).$$

The second inequality of Proposition 3.4.1 follows from the first one by Remark 3.3.2. □

3.4.3 Poincaré-like inequality for the convolution operator

For $a \in [0,1]$ and a centered $g \in L^2(\mu)$ (that is $\mu(g) = 0$), let us consider the quantity $\|Q^*_{a,\varphi}[g]\|_{L^2(\mu)}$. By analogy with the Poincaré inquality for the Ornstein-Uhlenbeck semigroup, which is reversible with respect to μ, we call the following result a Poincaré-like inequality holding for the operator $Q^*_{a,\varphi}$, which in general is non-reversible.

Proposition 3.4.2 (Poincaré-like inequality)**.** *Assume that φ is a polynomial density in $L^2(\mu)$ whose moments match the moments of μ up to order $r \in \mathbb{Z}^+$, and which satisfies to Hypothesis **(H)** stated in Section 3.2. Set $a_\varphi \in [0,1)$ and $d_\varphi : (0,1) \to \mathbb{R}$ as in Theorem 3.1.1. Then, for all function $g \in L^2(\mu)$ which writes as:*

$$g = \sum_{k=r+1}^{\infty} g_k \bar{H}_k,$$

and for all $a \in (a_\varphi, 1)$, it stands that:

$$\int (Q^*_{a,\varphi}[g])^2 d\mu \leq a^{r+1} \left(1 + d_\varphi(a)\right)^2 \int g^2 d\mu.$$

The proof is cut out in a number of steps. We begin by the following lemma:

Lemma 3.4.3 (Gershgorin's theorem)**.** *Let φ and g be as in Proposition 3.4.2, and set $K = r+1$. Then, for all $a \in (0,1)$,*

$$\int (Q^*_{a,\varphi}[g])^2 d\mu \leq \sup_{l \geq K} \Sigma_{a,\varphi}(l) \int g^2 d\mu, \tag{3.17}$$

where

$$\Sigma_{a,\varphi}(l) := \sum_{j=K}^{+\infty} \left| \overrightarrow{M}_{a,\varphi}(l,j) \right|, \qquad l \geq K.$$

Proof. Let $r \in \mathbb{Z}^+$, $a \in (0,1)$ and φ as in the statement of the proposition, and let $K = r+1$. First, notice that V_K, the set of functions $g \in L^2(\mu)$ with Hermite decomposition

$$g = \sum_{k=K}^{\infty} g_k \bar{H}_k,$$

is stable under action of $Q^*_{a,\varphi}$ by Remark 3.3.4. The space V_K equipped with the $L^2(\mu)$ structure is again a Hilbert space. Let us call $Q^*_{a,\varphi|_{V_K}}$ the restriction of $Q^*_{a,\varphi}$ to V_K. It is again bounded, with operator norm

$$\left\| Q^*_{a,\varphi|_{V_K}} \right\|_{\mathrm{op}} := \sup_{g \in V_K \setminus \{0\}} \frac{\|Q^*_{a,\varphi}[g]\|_{L^2(\mu)}}{\|g\|_{L^2(\mu)}}.$$

Hence, the desired majoration (3.17) is equivalent to the following bound on the operator norm:

$$\left\| Q^*_{a,\varphi|_{V_K}} \right\|^2_{\mathrm{op}} \leq \sup_{l \geq K} \Sigma_{a,\varphi}(l). \tag{3.18}$$

The isometry between $L^2(\mu)$ and l^2 restricts to an isometry between V_K and l^2_K, defined as the space of real sequences $(u_n)_{n\geq K}$ with $\sum_{n\geq K} u_n^2 < +\infty$. By this isometry, if $N_K = (N_K(i,j))_{i,j\geq K}$ stands for the infinite matrix associated to $Q^*_{a,\varphi|_{V_K}}$, then:

$$\left\| Q^*_{a,\varphi|_{V_K}} \right\|_{\mathrm{op}} = \|N_K\|_{\mathrm{op}}.$$

Furthermore, by the properties of block matrix multiplication, one sees that the matrix ${}^T N_K N_K$ is nothing else but the matrix $\overrightarrow{M}_{a,\varphi}$ defined in (3.16) (Section 3.3) restricted to l^2_K, that is:

$${}^T N_K N_K = \left(\overrightarrow{M}_{a,\varphi}(i,j)\right)_{i,j\geq K}.$$

For a complex Banach space E, set $\mathcal{G}(E)$ the set of inversible operators on E. The spectral radius of a bounded operator M in E is then defined as

$$\rho(M) := \max\left\{|\lambda|,\ \lambda \mathrm{Id} - M \in \mathcal{G}(E)\right\}.$$

Moreover, if E is Hilbert and if T is a bounded operator of E with adjoint T^*, then

$$\|T\|_{\mathrm{op}} = \sqrt{\rho(T^*T)}.$$

The operator N_K being a bounded operator of l^2_K (by restriction of a bounded operator), the preceding equation applies:

$$\|N_K\|_{\mathrm{op}} = \sqrt{\rho({}^T N_K N_K)}.$$

Let us recall a theorem of Gershgorin (Gershgorin [1931]) related to finite complex auto-adjoint matrices A, where $A = (A_{i,j})_{1\leq i,j\leq n}$ for a positive integer n. Denoting $\mathcal{G}_n(\mathbb{C})$ the set of invertible matrices of size n and $B(x,r)$ the complex ball of center $x \in \mathbb{C}$ and $r > 0$, one has:

$$\{\lambda \in \mathbb{C},\ \lambda \mathrm{Id} - A \in \mathcal{G}_n(\mathbb{C})\} \subset \bigcup_{l=1}^{n} B\left(A(l,l), \left|\sum_{1\leq j\leq n,\, j\neq l} A(l,j)\right|\right).$$

As a consequence,

$$\rho(A) \leq \sup_{1\leq l\leq n} |A(l,l)| + \left|\sum_{1\leq j\leq n,\, j\neq l} A(l,j)\right| \leq \sup_{1\leq l\leq n} \sum_{j=1}^{n} |A(l,j)|.$$

Gershgorin's theorem is stated for finite matrices, but the proof extends without difficulty to eigenvalues of operators on l^2_K. The operator ${}^T N_K N_K$ being autoadjoint and compact by Proposition 3.3.5, its spectrum is included in the set of eigenvalues united with the singleton $\{0\}$, hence the formula above applies and yields majoration (3.18), which proves the lemma.

□

In order to derive an upper-bound of $\sup_{l\geq K} \Sigma_{a,\varphi}(l)$ from the explicit expression of $\overrightarrow{M}_{a,\varphi}$ stated in (3.16), we need two technical lemmas.

Lemma 3.4.4 (First technical lemma). *Let $N \geq K$ be positive integers, and recall that*

$$a_\varphi := \left(1 + \frac{N}{K}\right)^{-\frac{1}{4}}, \qquad C_k := \left(1 + \frac{N}{K}\right)^{k/2}, \qquad k \in \mathbb{Z}^+.$$

Let i, k be natural integers such that

$$0 \le k \le N-1, \qquad K \le i+k \le N, \qquad 1 \le i \le N.$$

Then, for all $a \in (a_\varphi, 1)$ and for all $l \ge K$,

$$\begin{aligned} a^{-i}\binom{k+l}{k}^{1/2}\binom{k+l}{k+i}^{1/2}\mathbf{1}_{l\ge i+K} + a^i\binom{k+l+i}{k}^{1/2}\binom{k+l+i}{k+i}^{1/2} \\ \le 2C_kC_{k+i}\binom{k+l}{k}^{1/2}\binom{k+l+i}{k+i}^{1/2}. \end{aligned}$$

Proof. Let $N \ge K$ be positive integers and i, k, l be natural integers such that

$$0 \le k \le N-1, \qquad K \le i+k \le N, \qquad 1 \le i \le N, \qquad l \ge K.$$

For two positive integers $m \ge n$, the notation $[m]_n$ stands for $[m]_n := m \cdots (m-n+1)$. One has:

$$\binom{k+i+l}{k}\binom{k+l}{k}^{-1} = \frac{[k+i+l]_k}{[k+l]_k}.$$

If $i \ge k$, then

$$\binom{k+i+l}{k}\binom{k+l}{k}^{-1} \le \left(\frac{l+N}{l+1}\right)^k \le \left(\frac{l+N}{l+1}\right)^i.$$

If $i < k$, then

$$\binom{k+i+l}{k}\binom{k+l}{k}^{-1} = \frac{[k+i+l]_i\,[k+l]_{k-i}}{[k+l]_{k-i}\,[l+i]_i} = \frac{[k+i+l]_i}{[i+l]_i} \le \left(\frac{l+N}{l+1}\right)^i.$$

In both cases,

$$\binom{k+i+l}{k}\binom{k+l}{k}^{-1} \le \left(\frac{l+N}{l+1}\right)^i.$$

Furthermore, in the case where $l \ge i$,

$$\binom{k+l}{k+i}\binom{k+i+l}{k+i}^{-1} = \frac{[k+l]_k\,[l]_i}{[k+i+l]_i\,[k+l]_k} = \frac{[l]_i}{[k+i+l]_i} \le \left(\frac{l}{l+1}\right)^i.$$

Hence, for all $a \in (0,1)$,

$$\begin{aligned} & a^{-i}\binom{k+l}{k}^{1/2}\binom{k+l}{k+i}^{1/2}\mathbf{1}_{l\ge i+K} + a^i\binom{k+l+i}{k}^{1/2}\binom{k+l+i}{k+i}^{1/2} \\ & \le \left(\frac{l}{l+1}\right)^{i/2}\binom{k+l}{k}^{1/2}\binom{k+l+i}{k+i}^{1/2}\left(a^{-i}\mathbf{1}_{l\ge i+K} + a^i\left(1+\frac{N}{l}\right)^{i/2}\right) \\ & \le \binom{k+l}{k}^{1/2}\binom{k+l+i}{k+i}^{1/2} f(a), \end{aligned} \tag{3.19}$$

where we defined

$$f(a) := a^{-i} + \left(1+\frac{N}{K}\right)^{i/2} a^i, \qquad a \in (0,1).$$

As one checks easily, the inequality (3.19) still holds true if $l < i$, and $f'(a) \ge 0$ if and only if $a \ge a_\varphi = (1+N/K)^{-1/4}$. This yields for all $a \in (a_\varphi, 1)$,

$$f(a) \le f(1) = 1 + \left(1+\frac{N}{K}\right)^{i/2} \le 2\left(1+\frac{N}{K}\right)^{i/2} \le 2C_kC_{k+i},$$

where C_k has been defined as $C_k = (1+N/K)^{k/2}$ for all $k \in \mathbb{Z}^+$. □

Lemma 3.4.5 (Second technical lemma). *Let $m > q$ be positive integers, and consider the polynomial $P = -\alpha X^m + \beta X^q - 1$ with $\alpha, \beta > 0$. Then $P \leq 0$ on $\mathbb{R}^+$ if and only if*

$$\left(\frac{\beta}{m}\right)^m \left(\frac{q}{\alpha}\right)^q \leq \frac{1}{(m-q)^{m-q}}.$$

Proof of Lemma 3.4.5. Let $P(x) = -\alpha x^m + \beta x^q - 1$ be as in the wording of the lemma. Then, for all $x \in \mathbb{R}$,

$$P'(x) \quad = \quad -m\alpha x^{m-1} + q\beta x^{q-1} = x^{q-1}(-m\alpha x^{m-q} + q\beta),$$

thus P attains its maximum on $[0, +\infty)$ at the point $x_0 = ((\beta q)/(\alpha m))^{1/(m-q)}$. Moreover,

$$\begin{aligned} P(x_0) &= x_0^q(-\alpha x_0^{m-q} + \beta) - 1 = \left(\frac{\beta q}{\alpha m}\right)^{\frac{q}{m-q}} \left(-\frac{\beta q}{m} + \beta\right) - 1 \\ &= \frac{\beta}{m}\left(\frac{\beta q}{\alpha m}\right)^{\frac{q}{m-q}} (m-q) - 1 = \left(\frac{\beta}{m}\right)^{\frac{m}{m-q}} \left(\frac{q}{\alpha}\right)^{\frac{q}{m-q}} (m-q) - 1, \end{aligned}$$

so that $P(x_0) \leq 0$ if and only if

$$\left(\frac{\beta}{m}\right)^m \left(\frac{q}{\alpha}\right)^q \leq \frac{1}{(m-q)^{m-q}}.$$

□

We are now ready to show Proposition 3.4.2.

Proof of Proposition 3.4.2. Set $K = r+1$, let $a \in (0,1)$ and l a positive integer such that $l \geq K$. Then,

$$\begin{aligned} \Sigma_{a,\varphi}(l) &= \sum_{j=K}^{+\infty} \left|\overrightarrow{M}_{a,\varphi}(l,j)\right| = \left|\overrightarrow{M}_{a,\varphi}(l,l)\right| + \sum_{i=1}^{N} \left(\left|\overrightarrow{M}_{a,\varphi}(l,l-i)\right| \mathbf{1}_{l-i\geq K} + \left|\overrightarrow{M}_{a,\varphi}(l,l+i)\right|\right) \\ &= a^{2l} \sum_{k\geq 0} \binom{k+l}{k} (1-a^2)^k \varphi_k^2 + \sum_{i=1}^{N} a^{2l-i} \sum_{k\geq 0} \binom{k+l}{k}^{1/2} \binom{k+l}{k+i}^{1/2} (1-a^2)^{\frac{2k+i}{2}} \varphi_{k+i}\varphi_k \mathbf{1}_{l-i\geq K} \\ &\quad + \sum_{i=1}^{N} a^{2l+i} \sum_{k\geq 0} \binom{k+l+i}{k}^{1/2} \binom{k+l+i}{k+i}^{1/2} (1-a^2)^{\frac{2k+i}{2}} \varphi_{k+i}\varphi_k \\ &= a^{2l} \left(1 + \sum_{k=K}^{N} \binom{k+l}{k} (1-a^2)^k \varphi_k^2 + \sum_{0\leq k<k+i\leq N} \left\{(1-a^2)^{\frac{2k+i}{2}} |\varphi_k||\varphi_{k+i}| \mathcal{C}_{a,\varphi}(l,i,k)\right\}\right), \end{aligned}$$

where

$$\mathcal{C}_{a,\varphi}(l,i,k) := a^{-i} \binom{k+l}{k}^{1/2} \binom{k+l}{k+i}^{1/2} \mathbf{1}_{l\geq i+K} + a^i \binom{k+l+i}{k}^{1/2} \binom{k+l+i}{k+i}^{1/2}$$

is precisely the quantity addressed in Lemma 3.4.4. If $\varphi_k\varphi_{k+i} \neq 0$ and $i \geq 1$ then $i + k \geq K$, hence the assumptions of Lemma 3.4.4 hold and we get for all $a \in (a_\varphi, 1)$:

$$\begin{aligned} \Sigma_{a,\varphi}(l) &\leq a^{2l} \left(1 + \sum_{k=K}^{N} \binom{k+l}{k} (1-a^2)^k \varphi_k^2\right) \\ &\quad + 2\,a^{2l} \left(\sum\sum_{0\leq k<k+i\leq N} \left\{(1-a^2)^{\frac{2k+i}{2}} |\varphi_k||\varphi_{k+i}| C_k C_{k+i} \binom{k+l}{k}^{1/2} \binom{k+l+i}{k+i}^{1/2}\right\}\right). \end{aligned}$$

Noticing that $C_k \geq 1$ for every positive integer K and that $C_0 = 1$ allows to recognize the development of a square:

$$\Sigma_{a,\varphi}(l) \leq a^{2l} \left(1 + \sum_{k=K}^{N} \binom{k+l}{k}^{1/2} (1-a^2)^{\frac{k}{2}} C_k |\varphi_k| \right)^2 \leq a^{2l} \left(1 + \sum_{k=K}^{N} \frac{C_k|\varphi_k|}{\sqrt{k!}} \left((N+l)(1-a^2)\right)^{\frac{k}{2}} \right)^2,$$

where we used that for all natural integer $k \leq N$,

$$\binom{k+l}{k} \leq \frac{(N+l)^k}{k!}.$$

We recognize the coefficient γ_k introduced in Section 3.2 to state Hypothesis **(H)**:

$$\gamma_k = \frac{C_k|\varphi_k|}{\sqrt{k!}}, \qquad k \geq K,$$

so that

$$\Sigma_{a,\varphi}(l) \leq a^{-2N} a^{2(N+l)} \left(1 + \sum_{k=K}^{N} \gamma_k \left((N+l)(1-a^2)\right)^{\frac{k}{2}} \right)^2.$$

For all $a \in (0,1)$ and $l \in \mathbb{Z}^+$, we perform the change of variables

$$u_{a,\varphi}(l) := -(l+N)\log a > 0 \quad \Rightarrow \quad (N+l)(1-a^2) = (N+l)\left(1 - \exp\left(-2\frac{u_{a,\varphi}(l)}{l+N}\right)\right) \leq 2u_{a,\varphi}(l),$$

and introduce the function

$$h(u) = \exp(-u)\left(1 + \sum_{k=K}^{N} \gamma_k (2u)^{k/2}\right), \qquad u \geq 0.$$

Then,

$$\Sigma_{a,\varphi}(l) \leq a^{-2N} h^2\left(u_{a,\varphi}(l)\right).$$

The last part of the proof is devoted to showing that the function h is non-increasing on $[0,+\infty)$; indeed in that case, we have for all $a \in (a_\varphi, 1)$ and $l \geq K$:

$$\Sigma_{a,\varphi}(l) \leq a^{-2N} h^2\left(u_{a,\varphi}(K)\right) = a^{2K} \left(1 + \sum_{k=K}^{N} \gamma_k \left(-2(K+N)\log a\right)^{k/2}\right)^2,$$

which, jointly with Lemma 3.4.3, proves Proposition 3.4.2. So let us study the variation of h. For all $u \geq 0$,

$$h'(u) = \exp(-u)\left(-1 - \sum_{k=K}^{N} \gamma_k (2u)^{k/2} + \sum_{k=K}^{N} \gamma_k k (2u)^{(k-2)/2}\right).$$

Let us consider separately the powers of $u^{\frac{1}{2}}$ ranging from K to $N-2$ (when existing) and the remaining powers:

$$\begin{aligned} -1 - \sum_{k=K}^{N} \gamma_k (2u)^{k/2} + \sum_{k=K}^{N} \gamma_k k (2u)^{(k-2)/2} &= \sum_{k=K}^{N-2} \left(-\gamma_k + (k+2)\gamma_{k+2}\right)(2u)^{k/2} \\ &\quad + K\gamma_K (2u)^{(K-2)/2} + \gamma_{K+1}(K+1)(2u)^{(K-1)/2} \\ &\quad - \left(1 + \gamma_N (2u)^{N/2} + \gamma_{N-1}(2u)^{(N-1)/2}\right). \end{aligned}$$

As Hypothesis **(H1)** holds, the sum $\sum_{k=K}^{N-2}$ is nonpositive. If $K \leq N-1$, the remaining term writes

$$\frac{1}{2}\left(-1-2\gamma_N(2u)^{N/2}+2\gamma_{K+1}(K+1)(2u)^{(K-1)/2}\right)+\frac{1}{2}\left(-1-2\gamma_{N-1}(2u)^{(N-1)/2}+2K\gamma_K(2u)^{(K-2)/2}\right),$$

which is nonpositive thanks to Hypothesis **(H2a)** and Lemma 3.4.5. If $K = N$, the same arguments provide the nonpositivity of the remaining term, which reduces to

$$-1-\gamma_N(2u)^{N/2}+N\gamma_N(2u)^{(N-2)/2}.$$

Finally, under Hypothesis **(H)**, we find that h has a nonpositive derivative on $[0,+\infty[$ hence is non-increasing, which completes the proof. □

3.4.4 Proof of Theorem 3.1.1

Let us turn to the proof Theorem 3.1.1.

Proof of Theorem 3.1.1. Let φ, $f \in L^2(\mu)$ be as in the wording of the theorem. For all $a \in [0,1]$,

$$\chi_2\left(af * \sqrt{1-a^2}\varphi\right) = \|K_a(f,\varphi)-1\|_{L^2(\mu)} \quad \leq \|K_a(f-1,\varphi)\|_{L^2(\mu)} + \|K_a(\mathbf{1},\varphi)-1\|_{L^2(\mu)}.$$

According to Proposition 3.4.1, for all $a \in [0,1]$,

$$\|K_a(\mathbf{1},\varphi)-1\|_{L^2(\mu)} = \|Q^*_{a,\varphi}[\mathbf{1}]-1\|_{L^2(\mu)} \leq (1-a^2)^{\frac{r+1}{2}}\chi_2(\varphi),$$

while by Proposition 3.4.2, for all $a \in (a_\varphi, 1)$,

$$\|K_a(f-1,\varphi)\|_{L^2(\mu)} = \|Q^*_{a,\varphi}[f-1]\|_{L^2(\mu)} \leq a^{r+1}\left(1+d_\varphi(a)\right)\chi_2(f),$$

which proves the theorem. □

3.4.5 Alternative bound

For the sake of completeness, let us conclude this section with a bound alternative to inequality (E_r).

Proposition 3.4.6 (Alternative bound on χ_2 under convolution)**.** *Let $f, \varphi \in L^2(\mu)$ be density with moments matching the Gaussian moments up to order $r \in \mathbb{Z}^+$, and moreover assume that f is $(r+1)$-times derivable, with $D^{r+1}f \in L^2(\mu)$. Then, there exists a universal constant $c_r > 0$ such that $\forall a \in (0,1)$,*

$$\begin{aligned}\chi_2\left(af * \sqrt{1-a^2}\varphi\right) \leq{}& a^{r+1}\chi_2(f) + (1-a^2)^{\frac{r+1}{2}}\chi_2(\varphi) \\ &+ c_r(1-a^2)^{\frac{r+1}{2}}\left(\chi_2(f)\chi_2(\varphi) + \left\|D^{r+1}f\right\|_{L^2(\mu)}\left\|D^{-(r+1)}\varphi\right\|_{L^2(\mu)}\right), \quad (3.20)\end{aligned}$$

where D^{-1} stands for the operator which maps a function $\varphi \in L^2(\mu)$ onto its primitive with vanishing mean.

Contrarily to what happens for bound (E_r), (3.20) stands for all $a \in (0,1)$ and the polynomial assumption on φ is not required, making the relative roles of f and φ more symmetric. The drawback of bound (3.20) is that it involves the norm of the $(r+1)$-th derivative of f, which we fail to control in the framework of the Central Limit Theorem.

Proof of Proposition 3.4.6. Set $r \in \mathbb{Z}^+$, $K = r+1$ and f, φ two densities as in the wording of the remark, so that

$$f = 1 + \sum_{k=K}^{+\infty} f_k \bar{H}_k, \qquad \varphi = 1 + \sum_{k=K}^{+\infty} \varphi_k \bar{H}_k.$$

For all $a \in (0,1)$, one has:

$$\begin{aligned}\chi_2\left(af * \sqrt{1-a^2}\,\varphi\right) &= \|K_a(f,\varphi) - 1\|_{L^2(\mu)} \\ &\leq \|K_a(f-1,\mathbf{1})\|_{L^2(\mu)} + \|K_a(\mathbf{1},\varphi-1)\|_{L^2(\mu)} + \|K_a(f-1,\varphi-1)\|_{L^2(\mu)}.\end{aligned}$$

Now,

$$K_a(f-1,\mathbf{1}) = P_{-\log a}[f-1], \qquad K_a(\mathbf{1},\varphi-1) = P_{-\frac{1}{2}\log(1-a^2)}[\varphi-1],$$

hence by the improved Poincaré inequality from Proposition 3.4.1 we get the two first terms of the bound. It remains to consider $\|K_a(f-1,\varphi-1)\|_{L^2(\mu)}$. For all $a \in (0,1)$, one has by Remark 3.3.4:

$$K_a(f-1,\varphi-1) = \sum_{m=K}^{+\infty}\sum_{n=K}^{+\infty} \binom{m+n}{m}^{1/2} a^m (1-a^2)^{n/2} f_m \varphi_n \bar{H}_{n+m},$$

hence

$$\begin{aligned}\|K_a(f-1,\varphi-1)\|^2_{L^2(\mu)} &= \sum_{l=2K}^{+\infty}\left(\sum_{\substack{m+n=l\\ m,n\geq K}} \binom{l}{m}^{1/2} a^m (1-a^2)^{n/2} f_m\varphi_n\right)^2 \\ &= (1-a^2)^{K/2}\sum_{l=2K}^{+\infty}\left(\sum_{\substack{m+n=l-K\\ m\geq K, n\geq 0}} \binom{l}{m}^{1/2} a^m (1-a^2)^{n/2} f_m\varphi_{n+K}\right)^2.\end{aligned}$$

By Cauchy-Schwarz inequality and Pascal formula, this rewrites again

$$\begin{aligned}&(1-a^2)^{K/2}\sum_{l=2K}^{+\infty}\left(\sum_{\substack{m+n=l-K\\ m\geq K, n\geq 0}} \binom{l-K}{m}^{1/2}\left(\frac{l\cdots(l-K+1)}{(n+K)\cdots(n+1)}\right)^{1/2} a^m(1-a^2)^{n/2} f_m\varphi_{n+K}\right)^2 \\ &\qquad \leq (1-a^2)^{K/2}\sum_{l=2K}^{+\infty}\sum_{\substack{m+n=l-K\\ m\geq K, n\geq 0}} \frac{l\cdots(l-K+1)}{(n+K)\cdots(n+1)} f_m^2\varphi_{n+K}^2.\end{aligned}$$

Now, there exists $c_K > 0$ such that $\forall x \geq K, \forall y \geq 0$,

$$(x+y+K)\cdots(x+y+1) \leq c_K\left(x\cdots(x-K+1) + (y+K)\cdots(y+1)\right).$$

Applying this to $x = m$ and $y = n$, we find that

$$\begin{aligned}\|K_a(f-1,\varphi-1)\|^2_{L^2(\mu)} &\leq c_K(1-a^2)^{K/2}\sum_{l=2K}^{+\infty}\sum_{\substack{m+n=l-K\\ m\geq K, n\geq 0}}\left(1+\frac{m\cdots(m-K+1)}{(n+K)\cdots(n+1)}\right) f_m^2\varphi_{n+K}^2 \\ &= c_K(1-a^2)^{K/2}\left(\sum_{m\geq K} f_m^2\right)\left(\sum_{n\geq K}\varphi_n^2\right) \\ &\quad + c_K(1-a^2)^{K/2}\left(\sum_{m\geq K} m\cdots(m-K+1) f_m^2\right)\left(\sum_{n\geq K}\frac{\varphi_n^2}{(n+K)\cdots(n+1)}\right),\end{aligned}$$

which is the Hermite representation of the expected quantity by formula (3.7). □

3.5 Proof of Theorem 3.1.2

Finally, we conclude the article with the proof of our main theorem, Theorem 3.1.2, which follows on from the recursion formula (3.10) and barycentric convolution inequality for χ_2-distance (E_r).

Proof of Theorem 3.1.2. In the framework of the theorem, denote

$$n_0 := \left\lceil \frac{1}{1-a_\varphi^2} \right\rceil \vee 2.$$

By the two aforementioned relations, we have for all integer $n \geq n_0$:

$$\chi_2(f_n) \leq \left(1-\frac{1}{n}\right)^{\frac{r+1}{2}} \left(1+d_\varphi\left(\sqrt{1-1/n}\right)\right)\chi_2(f_{n-1}) + \frac{1}{n^{\frac{r+1}{2}}}\chi_2(\varphi).$$

Remembering that $r \geq 2$, let us call for all $n \geq n_0$,

$$c_n := \left(1-\frac{1}{n}\right)^{\frac{r+1}{2}} \left(1+d_\varphi\left(\sqrt{1-1/n}\right)\right) = 1 - \frac{r+1}{2n} + \mathcal{O}\left(\frac{1}{n^{\frac{3}{2}}}\right), \qquad d_n := \frac{1}{n^{\frac{r+1}{2}}}\chi_2(\varphi),$$

where we denote $v_n = \mathcal{O}(u_n)$ if $\limsup_{n\to+\infty} |v_n/u_n| < +\infty$. The preceding recursive inequality yields

$$\chi_2(f_n) \leq \left(\prod_{k=n_0}^{n} c_k\right)\chi_2(f_{n_0-1}) + \sum_{k=n_0}^{n}\left(\prod_{j=k+1}^{n} c_k\right) d_k.$$

Now,

$$\log\left(\prod_{k=n_0}^{n} c_k\right) = \sum_{k=n_0}^{n}\log c_k = -\sum_{k=n_0}^{n}\left(\frac{r+1}{2n} + \mathcal{O}\left(\frac{1}{n^{\frac{3}{2}}}\right)\right) = -\frac{r+1}{2}\log n + \mathcal{O}(1).$$

This leads to

$$\prod_{k=n_0}^{n} c_k = \mathcal{O}\left(\frac{1}{n^{\frac{r+1}{2}}}\right); \qquad \sum_{k=n_0}^{n}\left(\prod_{j=k+1}^{n} c_k\right) d_k = \left(\prod_{j=n_0}^{n} c_j\right)\sum_{k=n_0}^{n}\frac{d_k}{\prod_{j=n_0}^{k} c_j} = \mathcal{O}\left(\frac{1}{n^{\frac{r+1}{2}}}\right)\sum_{k=n_0}^{n}\mathcal{O}(1)$$
$$= \mathcal{O}\left(\frac{1}{n^{\frac{r-1}{2}}}\right).$$

Finally,

$$\chi_2(f_n) := \mathcal{O}\left(\frac{1}{n^{\frac{r+1}{2}}}\right) + \mathcal{O}\left(\frac{1}{n^{\frac{r-1}{2}}}\right) = \mathcal{O}\left(\frac{1}{n^{\frac{r-1}{2}}}\right),$$

which proves the theorem. □

Chapter 4

Berry-Esseen bounds for a time-inhomogeneneous Markov jump process

This chapter consists in a joint work with Sébastien Gadat and Laurent Miclo under finalisation.

Abstract

We are interested in a time-inhomogeneous Markov process that evolves with jumps of two types, either $+a_t$ or $-a_t$ for a time $t \geq 0$, and which is inspired by a stochastic algorithm finding the median of a probability distribution. The goal of this note is to present a method allowing to quantify the convergence of the process to the normal distribution. The jump size, as well as the jump rates are time-dependent, hence the usual results for time-homogeneous processes do not apply. On the other hand, the usual quantitative methods for stochastic algorithms do not work either for the median algorithm which motivates the study. Thanks to a careful analysis of the dynamic, which relies on spectral methods, we propose a conjecture on the rate of convergence of the process to the normal distribution in Wasserstein distance. The result is proved up to a technical point which we are currently finishing to investigate.

Introduction

In this article, we are interested in the real, continuous-time Markov process $(Z_t)_{t\geq 0}$ with instantaneous generator acting on bounded functions $f : \mathbb{R} \to \mathbb{R}$ as:

$$L_t[f](z) = p_t(z)(f(z+a_t) - f(z)) + q_t(z)(f(z-a_t) - f(z)), \qquad z \in \mathbb{R}, \quad t \geq 0, \tag{4.1}$$

where p, q are non-negative, space and time-dependent rate functions and $(a_t)_{t\geq 0}$ a positive function. The process $(Z_t)_{t\geq 0}$ evolves by jumps of two types: either from z to $z + a_t$ with rate $p_t(z)$, or by jumps from z to $z - a_t$ with rates $q_t(z)$. Notation $\mathcal{L}(Z_t)$ stands for its marginal distribution at time $t \geq 0$.

Let μ denote the normal distribution and by extension its density with respect to the Lebesgue measure on $\mathbb{R}$. If we assume that the equilibrium relation $(\mathbf{H_{Rev}})$ below holds,

$$p_t(z - a_t)\mu(z - a_t) = q_t(z)\mu(z), \qquad z \in \mathbb{R}, \quad t \geq 0, \tag{$\mathbf{H_{Rev}}$}$$

the Gaussian distribution μ is invariant and in fact reversible for the process $(Z_t)_{t\geq 0}$. We will see below that there exist time-dependent distributions $(\mu_{a_t})_{t\geq 0}$ which are also reversible for the process.

The purpose of this note is to quantify the ergodicity of the process. Its distinctive feature, to evolve with jumps of only two types—either $+a_t$ or $-a_t$— affects the asymptotic behaviour: as will be explained below, the lack of diffusitivity slows down the convergence of $(Z_t)_{t\geq 0}$ to the asymptotic measure μ. As such, the process is said to be *weakly mixing*. We claim that the Wasserstein distance of order 1 between the marginal distribution of $(Z_t)_{t\geq 0}$ and the distribution μ decreases with a rate depending explicitely on rate and jump size parameters p, q, a, and that this result extends to some examples where the process is not supposed to be reversible with respect to μ but only asymptotically normal. The proof of these results requires a careful analysis of the dynamic and is completed up to a point which we are still investigating, and to which we have given the status of conjecture in this note. We also state a theorem relative to the homogeneous case, Theorem 4.0.1 below, which is fully proved.

Apart from its own interest, the investigation of dynamics of type (4.1) is motivated by a problem arising in statistics with the so-called recursive quantile estimation algorithm. Consider $(X_i)_{i\geq 1}$ a sequence of independent, identically distributed real variable with probability distribution ν. We are interested in approximating the α-quantile of the distribution ν. The computational cost of empirical and kernel estimators (David [1970], Nadaraja [1964]) being prohibitive for data streams, the preferred solution is a recursive stochastic algorithm of Robbins-Monro type (Robbins and Monro [1951]):

$$Y_{n+1} = Y_n + \gamma_{n+1}(\mathbf{1}_{X_{n+1}>Y_n} - \alpha), \qquad n \geq 1, \tag{4.2}$$

where $(\gamma_n)_{n\geq 1}$ is a sequence of steps of the form $\gamma_n = cn^{-\beta}$ for $\beta \in (0,1]$. Under regularity assumptions on the distribution ν, it is strongly consistent and asymptotically normal (Blum [1954], Duflo [1997]), and satisfies to large deviations and a law of the iterated logarithm (Woodroofe [1972], Gapoškin and Krasulina [1974]). However, the assumptions required in most studies relative to non-asymptotic bounds for recursive stochastic algorithm do not hold for algorithm (4.2) (strong convexity assumption in Bach and Moulines [2011], Frikha and Menozzi [2012]; dimensionality assumption in Cardot et al. [2015], Godichon-Baggioni [2016]). A bound on the quadratic risk was obtained in Labopin-Richard [2016] providing that the density is bounded from below, which implicitly amounts to a strong convexity condition. To push further the analysis, one wants to investigate a potential Berry-Esseen bound. The martingale methods commonly used in the field of Robbins-Monro algorithms do not lend themselves easily to such a study. On the other hand, the time-inhomogeneous Markov process led by (4.1) can be seen as a continuous-time version of the renormalized algorithm (4.2) in the median case ($\alpha = 1/2$), which evolves by innovations of the form $\pm\sqrt{\gamma_{n+1}}/2$. On the basis of our results, qualitative considerations on the median algorithm are further developed in Section 4.2.2.

Let us take a closer look at the dynamic, in the homogeneous case to begin with. The generator then reads:

$$L[f](z) = p(z)(f(z+a) - f(z)) + q(z)(f(z-a) - f(z)), \qquad z \in \mathbb{R}, \tag{4.3}$$

and the equilibrium relation ensuring the reversibility of the process with respect to the normal distribution μ is

$$p(z-a)\mu(z-a) = q(z)\mu(z), \qquad z \in \mathbb{R}. \tag{$\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Rev}}$}$$

The size of the jumps being constant, the process with initial value x stays on the grid $\Gamma_{a,x} = x + a\mathbb{Z}$. As a consequence, it is not ergodic: its marginal distribution, having support on a grid, cannot converge towards the invariant distribution μ, which is absolutely continuous on $\mathbb{R}$. On the other hand, for any fixed $x \in \mathbb{R}$, the process conditioned to stay on the grid $\Gamma_{a,x}$, that is to say conditioned to take its initial value on the grid, behaves as a birth-death process. The name birth-death process usually refers to a Markovian process living on $\mathbb{Z}$, which evolves either by a jump to its right (a birth), or to its left (a death), with rates depending on its position, and can be thought of as the continuous-time equivalent of a space-inhomogeneous random walk on $\mathbb{Z}$. By extension, we call again birth-death process with birth rates $(p(x+an))_{n\in\mathbb{Z}}$ and death rates $(q(x+an))_{n\in\mathbb{Z}}$ the conditional process living on $\Gamma_{a,x} = x + a\mathbb{Z}$, which, if it is at point $x+an$ for $n \in \mathbb{Z}$, jumps to the next step $z + a(n+1)$ with rate $p(z+an)$, or to the preceding step $z + a(n-1)$ with rate $q(z+an)$.

This observation is at the heart of our method, as it turns out that, unlike the unconditional process, the conditional birth-death process has good ergodicity properties under the following assumption on the jump rates:

$$\exists \kappa > 0, \qquad \forall z \le 0, \quad p(z) \ge \kappa\,; \qquad \forall z \ge 0, \quad q(z) \ge \kappa. \tag{$\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Ergo}}$}$$

Let us call $\mu_{a,x}$ the distribution μ restricted to the above a-graduated grid $\Gamma_{a,x} = x + a\mathbb{Z}$, which is reversible for the conditional process. We refer to Hypothesis ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Ergo}}$) as an ergodicity assumption, because it guarantees that the conditional birth-death process spends enough time in a domain supporting most of the mass of the reversible distribution $\mu_{a,x}$: indeed, if the conditional process has gone very far to the right of the grid $\Gamma_{a,x}$, that is to say to a domain with small weight under $\mu_{a,x}$, it is brought back towards the bulk of the distribution by jumps to the left, whose probability of occurence is controlled by the death rate q, and the same comment goes for the left-hand side of the grid and the birth rate. Hypothesis ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Ergo}}$) actually allows to quantify the convergence of the conditional birth-death process on the subset $\Gamma_{a,x}$.

To quantify the distance between the marginal distribution of the process and the unconditional invariant distribution μ, we use the order 1 Wasserstein distance, defined for two probability distributions ν, $\widetilde{\nu}$ as:

$$W(\nu, \widetilde{\nu}) = \sup_{f \in \mathrm{Lip}_1} |\nu[f] - \widetilde{\nu}[f]|\,,$$

with Lip_1 denoting the set of 1-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$. Let us now state the result relative to the homogeneous, reversible process :

Theorem 4.0.1 (Evolution of Wasserstein distance, homogeneous case)**.** *Let us assume that Hypothesis ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Rev}}$) and ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Ergo}}$) hold, and that the marginal of the process at time* 0 *admits a square-integrable density f_0 with respect to the normal distribution μ.*

Then, it stands that for all $t \ge 0$,

$$W\left(\mathcal{L}(Z_t), \mu\right) \le \left(Var_\mu[f_0]\right)^{1/2} \left(\exp\left(-\frac{\kappa a^2 t}{8(1+a)(1+2a)} \right) + \frac{a}{\pi}\left(1 + \frac{a}{\sqrt{2\pi}}\right)^{1/2} \right). \tag{4.4}$$

In particular,

$$\limsup_{t\to+\infty} W\left(\mathcal{L}(Z_t), \mu\right) \le \frac{a}{\pi}\left(1 + \frac{a}{\sqrt{2\pi}}\right)^{1/2} \left(Var_\mu[f_0]\right)^{1/2}.$$

Let us comment right now on this result.

Remark 4.0.2 (Comments on Theorem 4.0.1)**.**

— *The Wasserstein distance between the marginal distribution of the process and the normal distribution μ do not decrease towards* 0*: indeed, the homogeneous process is not ergodic, as explained above.*

— *The exponential term in the bound* (4.4) *comes from the convergence of the conditional distribution. It decreases if the constant κ introduced in* ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Ergo}}$) *becomes bigger, which makes sense because then the process comes back from infinity faster; and increases for smaller jump size a, as in this case the process has to make a lot of jumps to come back from big values. These two effects appear to balance if κ is of the same magnitude as a^{-2}, a feature which is preserved in the non-homogeneous case, see Remark 4.1.5 below.*

— *The second term in* (4.4)*, which is proportional to the jump size a, amounts to a bound on the error made when integrating on the grid $\Gamma_{a,x}$ rather than on the whole line. To get the convergence of the time-inhomogeneous process to the distribution μ, it is then natural to require that the jump size $(a_t)_{t\geq 0}$ decreases to* 0 *as t goes to ∞, as we do in Theorems* (4.1.4) *and* (4.2.1)*.*

We state a similar bound for a reversible, inhomogeneous process in Theorem 4.1.4, and for a non reversible inhomogeneous process in Theorem 4.2.1. Both theorems hold up to Conjecture 4.1.3, which is stated in the following section.

This article is organized as follows. Section 4.1 deals with the reversible case: in this part, we state Theorem 4.1.4 and we prove it, alongside with Theorem 4.0.1 stated above. In Section 4.2, we present a result for non-reversible dynamics, which is obtained through a reversibilization procedure inspired from Fill [1991] and Arnaudon and Miclo [2016], and we develop the link with the median algorithm.

4.1 Reversible case

4.1.1 Notation and generalities

Let us begin by some notation. For every function $f : \mathbb{R} \to \mathbb{R}$, we set

$$\|f\|_\infty = \max_{z\in\mathbb{R}} |f(z)|.$$

The symbol ξ stands for the identity function, $\xi : \mathbb{R} \to \mathbb{R}, x \mapsto x$.

We denote ∂_t the derivative in time, and f' the derivative of a differentiable function $f : \mathbb{R} \to \mathbb{R}$. We also use the notation ∇f, including in the case where ∇f stands for the derivative in the sense of distributions.

For the sake of simplicity, it is assumed throughout the article that for every $t \geq 0$ the rates $(p_t(z))_{z\in\mathbb{R}}$ and $(q_t(z))_{z\in\mathbb{R}}$ are bounded, that is

$$\|p_t\|_\infty < +\infty, \qquad \|q_t\|_\infty < +\infty, \qquad t \geq 0,$$

and the corresponding assumption holds for the time-homogeneous case. As a consequence, for all $t \geq 0$, the operator L_t is bounded in the space of bounded functions from $\mathbb{R}$ to $\mathbb{R}$, and there exists an explicit construction ([Ethier and Kurtz, 1986, Chap. 4]) of a process $(Z_t)_{t\geq 0}$, with corresponding semigroup $(P_{s,t})_{t\geq s\geq 0}$ defined for bounded measurable functions as

$$P_{s,t}[f](z) := \mathbb{E}\left[f(Z_t)|Z_s = z\right], \qquad z \in \mathbb{R}, \quad t \geq s \geq 0,$$

which satisfies to

$$P_{s,t}[f] = P_{s,u}\left[P_{u,t}[f]\right], \qquad L_t[f](z) = \lim_{s\to t} \frac{P_s[f](z) - P_t[f](z)}{s-t}, \qquad z \in \mathbb{R}, \quad t \geq u \geq s \geq 0.$$

For a distribution probability ν and fonction $f \in L^1(\nu)$, we denote indifferently

$$\nu(f) = \int f \, d\nu,$$

as well as

$$\mathrm{Var}_\nu[f] := \int (f - \nu(f))^2 d\nu.$$

Here and in what follows, when the integration domain is not precised, it is set to be $\mathbb{R}$. The space $L^2(\mu)$ is a Hilbert space, with scalar product and associated norm defined as

$$\langle f, g\rangle_{L^2(\mu)} := \int fg\, d\mu, \qquad \|f\|_{L^2(\mu)} := \sqrt{\mu(f^2)}, \qquad f, g \in L^2(\mu).$$

For any $a > 0$ and $x \in \mathbb{R}$, set $\Gamma_{a,x} := \{x + an,\, n \in \mathbb{Z}\}$ and define the probability distribution $\mu_{a,x}$ as follows:

$$Z_{x,a} := \sum_{n\in\mathbb{Z}} \mu(x + an), \qquad \mu_{a,x} := \frac{1}{Z_{x,a}} \sum_{n\in\mathbb{Z}} \delta_{x+an}\mu(x + an).$$

If $f : \mathbb{R} \to \mathbb{R}$ is bounded and continuous, let us denote by $\mu_a[f]$ the a-periodic function

$$\mu_a[f] : \quad \mathbb{R} \to \mathbb{R}, \quad x \quad \mapsto \quad \mu_{a,x}[f] = \frac{1}{Z_{x,a}} \sum_{n\in\mathbb{Z}} f(x + an)\mu(x + an).$$

We call ζ_a the distribution with density $Z_{x,a}$ with respect to Lebesgue measure on $[0, a)$; then:

$$\int f d\mu = \int_0^a \mu_a[f]d\zeta_a = \int \mu_a[f]d\mu. \tag{4.5}$$

Indeed,

$$\int f d\mu = \sum_{n\in\mathbb{N}} \left(\int_{x+an}^{x+a(n+1)} f(z)\mu(z)dz \right) = \int_0^a \left(\sum_{n\in\mathbb{N}} f(x + an)\mu(x + an) \right) dx = \int_0^a \mu_a[f]d\zeta_a, \tag{4.6}$$

which gives the first equality in (4.5). To get the second one, replace f by $\mu_a[f]$ in equation (4.6):

$$\int \mu_a[f]d\mu = \int_0^a \mu_a[\mu_a[f]]d\zeta_a = \int_0^a \mu_a[f]d\zeta_a.$$

More generally, equation (4.5) holds for all $f \in L^2(\mu)$ by defining $\mu_a[f]$ as the conditional expectation of f with respect to Σ_a, the σ-algebra generated by the collection of sets $\{\Gamma_{a,x},\, x \in \mathbb{R}\}$. If f is bounded and continuous, the mapping $x \to \mu_{a,x}[f]$ is then the continuous version of this conditional expectation.

Finally, for all $f \in L^2(\mu)$, it stands that:

$$\int f d\mu = \int \mu_a[f]d\mu = \int_0^a \mu_a[f]d\zeta_a. \tag{4.7}$$

Pythagoras theorem reads as:

$$\mathrm{Var}_\mu[f] = I_a[f] + J_a[f], \tag{4.8}$$

where

$$I_a[f] := \int (f - \mu_a[f])^2 d\mu, \qquad J_a[f] := \int (\mu_a[f] - \mu[f])^2 d\mu.$$

For a reversible Markov process, it is well-known that the existence of a spectral gap for the generator corresponds to the exponential decrease of the variance $\mathrm{Var}_\mu[f]$ under the action of the corresponding semigroup (Ané et al. [2000]). In our framework, the analogous property holds instead for the the conditional variance,

$$I_a[f] = \int (f - \mu_a[f])^2 d\mu.$$

More precisely, in the homogeneous case the following lemma holds:

Lemma 4.1.1 (Conditional ergodicity, homogeneous case)**.** *Let us assume that Hypothesis* ($\mathbf{H}_{\mathbf{Rev}}^{\mathbf{Hom}}$) *and Hypothesis* ($\mathbf{H}_{\mathbf{Ergo}}^{\mathbf{Hom}}$) *hold, and that the marginal distribution at time* 0 *admits a density with respect to* μ *denoted by* f_0, *such that* $f_0 \in L^2(\mu)$.

Then, if we denote by f_t *the density at time* $t \geq 0$ *of the process* $(Z_t)_{t\geq 0}$ *with respect to* μ, *it stands that*

$$I_a[f_t] \leq I_a[f_0] \exp\left(-\frac{\kappa a^2 t}{4(1+a)(1+2a)}\right), \qquad t \geq 0.$$

Lemma 4.1.1, which is proved in Section 4.1.5, is the cornerstone of Theorem 4.0.1.

4.1.2 Conjectured result

We now go back to the object of primary interest, the time-inhomogeneous dynamic. As for the homogeneous case, we require an ergodicity condition, which is analogous to ($\mathbf{H}_{\mathbf{Ergo}}^{\mathbf{Hom}}$):

$$\forall t \geq 0, \quad \exists \lambda_t > 0, \qquad \forall z \leq 0, \quad p_t(z) \geq \lambda_t\,; \qquad \forall z \geq 0, \quad q_t(z) \geq \lambda_t. \tag{$\mathbf{H}_{\mathbf{Ergo}}^{\mathbf{1}}$}$$

When trying to get the counterpart of Lemma 4.1.1 for the conditional variation $(I_{a_t}[f_t])_{t\geq 0}$, we deal with an additional term, coming from the evolution of $(a_t)_{t\geq 0}$. Recall that ξ stands for the identity function on $\mathbb{R}$ and let $\mathcal{H}_a(\mu)$ be the space of functions $f \in L^2(\mu)$, such that $\xi\nabla\mu_a[f]$ is in $L^2(\mu)$, where $\nabla\mu_a[f]$ is the derivative of $\mu_a[f]$ in the sense of distributions.

Lemma 4.1.2 (Conditional expectation derivative)**.** *Let us assume that the size jump function* $(a_t)_{t\geq 0}$ *is differentiable. Then, for all* $t \geq 0$, $f \in \mathcal{H}_{a_t}(\mu)$, *one has:*

$$\partial_t \int (f - \mu_{a_t}[f])^2 d\mu = 2\partial_t(\log a_t) \int (f - \mu_{a_t}[f])\xi\nabla\mu_{a_t}[f] d\mu.$$

Lemma 4.1.2 is proved in Section 4.1.5. We will make use of the following conjecture:

Conjecture 4.1.3.

— *There exists a constant* $A > 0$, *such that for all* $a \in]0, A[$ *and for all* $f \in \mathcal{H}_a(\mu)$,

$$\int (f - \mu_a[f])\, \xi\, \nabla\mu_a[f] d\mu \geq 0. \tag{4.9}$$

— *If for all* $t \geq 0$, *one has* $0 < a_t < A$, *and if the density of the process at time* 0 *with respect to* μ, *denoted by* f_0, *is in* $\mathcal{H}_{a_0}(\mu)$, *then the density* f_t *of the process at time* $t \geq 0$ *with respect to* μ *is in* $\mathcal{H}_{a_t}(\mu)$.

Manipulations, which are yet to be formalized, on the extrema of the mapping

$$f \in \mathcal{H}(\mu) \to \int (f - \mu_a[f])\, \xi\, \nabla\mu_a[f] d\mu,$$

lead us to believe that Conjecture 4.1.3 is reasonable.

We are now ready to state our claim relative to the convergence of the process towards the normal distribution μ in Wasserstein distance, which is analogous to Theorem 4.0.1:

Theorem 4.1.4 (Ergodicity in Wasserstein distance)**.** *Let us assume that:*

— *Equilibrium relation* ($\mathbf{H}_{\mathbf{Rev}}$) *is satisfied;*

— *Ergodicity condition* ($\mathbf{H}_{\mathbf{Ergo}}^{\mathbf{1}}$) *holds.*

Moreover assume that the jump size function, $(a_t)_{t\geq 0}$, is differentiable and decreasing, with $a_0 < A$, and that the marginal of the process at time 0 admits a density $f_0 \in \mathcal{H}_{a_0}(\mu)$ with respect to the normal distribution μ.
Then, if Conjecture 4.1.3 is true, one has for all $t \geq 0$,

$$W\left(\mathcal{L}(Z_t), \mu\right) \leq \left(Var_\mu[f_0]\right)^{1/2} \left(\exp\left(-\int_0^t \frac{\lambda_u a_u^2}{8(1+a_u)(1+2a_u)} du\right) + \frac{a_t}{\pi}\left(a_t + \frac{1}{\sqrt{2\pi}}\right)^{1/2}\right). \tag{4.10}$$

Let us comment right now this result.

Remark 4.1.5 (Comments on Theorem 4.1.4)**.**

— *The exponential term in (4.10) comes from the ergodicity of the conditional process and goes to 0 if the integral diverges, for example if $\lambda_t \sim \lambda/a_t^2$: roughly speaking, the jump intensity needs to be of the same magnitude as the inverse of the variance of the jump size. This relative scale appears naturally when one considers the continuous-time renormalized version of a stochastic algorithm of the form $X_{n+1} = X_n + \gamma_{n+1} f(X_n) + \gamma_{n+1}^2 M_{n+1}$ satisfying to a CLT with speed $\sqrt{\gamma_n}$, as we will see in the case of the median algorithm in Part 4.2.2. This motivates the introduction of a simplified alternative to Hypothesis $(\mathbf{H^1_{Ergo}})$, which reads*

$$\exists \lambda > 0, \quad \forall t \geq 0, \qquad \forall z \leq 0, \quad p_t(z) \geq \frac{\lambda}{a_t^2}; \qquad \forall z \geq 0, \quad q_t(z) \geq \frac{\lambda}{a_t^2}. \tag{$\mathbf{H^2_{Ergo}}$}$$

— *Again in the stochastic algorithm setting, the jump size $(a_t)_{t\geq 0}$ is most often a power function, hence the term in a_t in bound (4.10) prevails over the exponential term.*

The remaining of this section is devoted to the proof of Theorems 4.0.1 and 4.1.4. Section 4.1.3 investigates the spectral gap of the conditional birth-death process presented in the introduction, while Section 4.1.4 deals with the evaluation of $J_a[f]$ for fixed a and $f \in L^2(\mu)$. Finally, those elements are gathered in Section 4.1.5 to prove Theorems 4.0.1 and 4.1.4.

4.1.3 Spectral gap of the conditional process

We begin by a lemma that makes precise the idea, introduced above, according to which the instantaneous generator (4.1) has a spectral gap with respect to the conditional distribution. Let us define the Dirichlet energy for a fixed time $t \geq 0$ as:

$$\mathcal{E}_\mu^t[f] := -\int f L_t[f] d\mu, \qquad f \in L^2(\mu).$$

Lemma 4.1.6 (Spectral gap)**.** *Under equilibrium relation $(\mathbf{H_{Rev}})$ and ergodicity condition $(\mathbf{H^1_{Ergo}})$, the following relation stands for all $t \geq 0$:*

$$\mathcal{E}_\mu^t[f] \geq \frac{1}{8}\frac{1}{(1+a_t)(1+2a_t)} \lambda_t \, a_t^2 \, I_{a_t}[f], \qquad f \in L^2(\mu). \tag{4.11}$$

Remark 4.1.7 (Comparison with the standard case)**.** *In the usual case, the inequality above stands with the variance $Var_\mu[f]$ instead of the conditional variance $I_{a_t}[f]$ present here. In that case, the inequality is known as a Poincaré inequality, and it implies the decreasing of the variance under the action of the corresponding Markov semigroup. The bound (4.11) is weaker than the usual Poincaré inquality, because for all $f \in L^2(\mu)$ and $a > 0$,*

$$Var_\mu[f] \geq I_a[f].$$

Remark 4.1.8 (Homogeneous case). *Lemma 4.1.6 holds for a fixed time $t \geq 0$, so that there is no difference between the homogeneous and the inhomogeneous setting. If ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Rev}}$) and ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Ergo}}$) are satisfied, one then has:*

$$\mathcal{E}_\mu[f] \geq \frac{\kappa a^2}{8(1+a)(1+2a)} I_a[f], \qquad f \in L^2(\mu), \tag{4.12}$$

where $\mathcal{E}_\mu[f]$ stands for the Dirichlet form associated to the homogeneous generator L defined in (4.3).

Proof. Let us fix $t \geq 0$ and denote $a_t = a$ for the sake of simplicity. By a density argument, it is enough to establish the inequality for a continuous function f, for which $\mu_a[f]$ admits an explicit expression. The idea consists in deriving a Poincaré inequality for the conditional process. Indeed, by equation (4.7),

$$I_{a_t}[f] = \int \mu_{a,x}\left[(f - \mu_a(x)[f])^2\right] d\mu(x), \qquad \mathcal{E}^t_\mu[f] = -\int \mu_{a,x}[f L_t[f]] d\mu(x).$$

Hence, Lemma 4.1.6 is established as soon as we know that for all $x \in \mathbb{R}$,

$$\mu_{a,x}\left[(f - \mu_a(x)[f])^2\right] \leq C_a\, \mu_{a,x}[-f L_t[f]], \tag{4.13}$$

with the correct constant $C_a = \lambda_t a^2/(8(1+a)(1+2a))$. Without loss of generality, one can assume that $x \in [0,a)$. Now, we observe that the generator L_t leaves invariant the set of functions $f : \Gamma_{a,x} \to \mathbb{R}$, and by an abuse of notation we call again L_t the operator restricted to this set. It corresponds to the generator of a birth-death process on $\Gamma_{a,x}$, reversible with respect to the measure $\mu_{a,x}$ thanks to equilibrium relation ($\mathbf{H_{Rev}}$), with spectral gap:

$$\lambda_1(L_t, x) := \inf \quad \frac{\mu_{a,x}[-f L_t[f]]}{\mathrm{Var}_{\mu_{a,x}[f]}},$$

where the infimum is taken on the set of non-constant functions $f : \Gamma_{a,x} \to \mathbb{R}$ in $L^2(\mu_{a,x})$. By Hardy's inequalities for birth-death processes (Miclo [1999]),

$$\lambda_1(L_t, x) \geq \frac{1}{4 \max\left\{B^+_{a,x}, B^-_{a,x}\right\}},$$

where the constants $B^+_{a,x}$, $B^-_{a,x}$ are defined as follows:

$$B^+_{a,x} := \sup_{n \geq 1} \left(\sum_{k=1}^{n} \frac{1}{q_t(x+ak)\mu_{a,x}(x+ak)}\right) \sum_{k=n}^{+\infty} \mu_{a,x}(x+ak),$$
$$B^-_{a,x} := \sup_{n \geq 1} \left(\sum_{k=1}^{n} \frac{1}{p_t(x-ak)\mu_{a,x}(x-ak)}\right) \sum_{k=n}^{+\infty} \mu_{a,x}(x-ak).$$

By simplification of the renormalizing constant $Z_{a,x}$, this is equivalent to

$$B^+_{a,x} = \sup_{n \geq 1} \left(\sum_{k=1}^{n} \frac{1}{q_t(x+ak)\mu(x+ak)}\right) \sum_{k=n}^{+\infty} \mu(x+ak),$$
$$B^-_{a,x} = \sup_{n \geq 1} \left(\sum_{k=1}^{n} \frac{1}{p_t(x-ak)\mu(x-ak)}\right) \sum_{k=n}^{+\infty} \mu(x-ak).$$

Now, for all positive integer k, $x + ak \geq 0$ and $x - ak \leq 0$, so that Hypothesis ($\mathbf{H^1_{Ergo}}$) implies that

$$q_t(x+ak) \geq \lambda_t\,; \qquad p_t(x-ak) \geq \lambda_t.$$

Hence,

$$B^+_{a,x} \leq \frac{1}{\lambda_t} \sup_{n\geq 1} \left(\sum_{k=1}^{n} \frac{1}{\mu(x+ak)} \right) \sum_{k=n}^{+\infty} \mu(x+ak), \quad B^-_{a,x} \leq \frac{1}{\lambda_t} \sup_{n\geq 1} \left(\sum_{k=1}^{n} \frac{1}{\mu(x-ak)} \right) \sum_{k=n}^{+\infty} \mu(x-ak).$$

By symmetry of μ, notice that

$$\left(\sum_{k=1}^{n} \frac{1}{\mu(x-ak)} \right) \sum_{k=n}^{+\infty} \mu(x-ak) = \left(\sum_{k=0}^{n-1} \frac{1}{\mu((a-x)+ak)} \right) \sum_{k=n-1}^{+\infty} \mu((a-x)+ak).$$

Let us assume for a moment that the following majoration holds:

$$\sup_{n\geq 0} \left(\sum_{k=0}^{n} \frac{1}{\mu(x+ak)} \right) \sum_{k=n}^{+\infty} \mu(x+ak) \leq \frac{2}{a^2}(1+a)(1+2a), \qquad x \in [0,a]. \tag{4.14}$$

Then,

$$\max\left\{B^+_{a,x}, B^-_{a,x}\right\} \leq \frac{1}{\lambda_t}\frac{2}{a^2}(1+a)(1+2a),$$

and in turn

$$\lambda_1(L_t, x) \geq \frac{a^2\lambda_t}{(8(1+a)(1+2a))} = C_a,$$

which ends the proof.

It remains to prove inequality (4.14). We use the fact that the function

$$y \mapsto \mu(y) = \frac{\exp(-y^2/2)}{\sqrt{2\pi}}$$

is decreasing on $[0,+\infty)$. This means in particular that for all $y \geq a$ and $a > 0$,

$$\mu(y) \leq \frac{1}{a} \int_{y-a}^{y} \mu(s)ds.$$

Thus, for all $n \in \mathbb{N}$,

$$\sum_{k=n+1}^{\infty} \mu(x+ak) \leq \frac{1}{a} \int_{x+an}^{+\infty} \mu(y)dy \leq \frac{1}{a(x+an)} \int_{x+an}^{+\infty} y\mu(y)dy \leq \frac{\mu(x+an)}{a(x+an)},$$

where we used that for all $y \geq x+an$, one has $y/(x+an) \geq 1$ to make the primitive of μ appear. On the other hand, the function $x \mapsto 1/\mu(y)$ is increasing on $[0,+\infty)$, hence by the same line of arguments, for all integers $N, n \in \mathbb{N}$ with $N \leq n-1$, one has

$$\sum_{k=N}^{n-1} \frac{1}{\mu(x+ak)} \leq \frac{1}{a} \int_{x+aN}^{x+an} \frac{dy}{\mu(y)} \leq \frac{1}{a(x+aN)} \int_{x+aN}^{x+an} y\frac{dy}{\mu(y)} \leq \frac{1}{a(x+aN)\mu(x+an)}.$$

Set $N_a := \lceil \frac{1}{a} \rceil$, so that $N_a < 1 + 1/a$ and $x + aN_a \geq 1$, and assume $n \geq N_a + 1$. Then,

$$\begin{aligned}\sum_{k=0}^{n-1} \frac{1}{\mu(x+ak)} &= \sum_{k=0}^{N_a-1} \frac{1}{\mu(x+ak)} + \sum_{k=N_a}^{n-1} \frac{1}{\mu(x+ak)} \leq \frac{N_a}{\mu(x+an)} + \frac{1}{a(x+aN_a)\mu(x+an)} \\ &\leq \frac{1}{\mu(x+an)}\left(1+\frac{2}{a}\right).\end{aligned}$$

On the other hand,

$$\sum_{k=n+1}^{+\infty} \mu(x+ak) \leq \frac{1}{a(x+an)}\mu(x+an) \leq \frac{1}{a}\mu(x+an).$$

Hence, for all $n \geq N_a + 1$,

$$\left(\sum_{k=0}^{n} \frac{1}{\mu(x+ak)}\right) \sum_{k=n}^{+\infty} \mu(x+ak) \leq 2\left(1+\frac{1}{a}\right)^2 \leq \frac{2}{a^2}(1+a)(1+2a).$$

In the case where $0 \leq n \leq N_a$,

$$\sum_{k=0}^{n} \frac{1}{\mu(x+ak)} \leq (N_a+1)\frac{1}{\mu(x+an)} \leq \left(2+\frac{1}{a}\right)\frac{1}{\mu(x+an)},$$

$$\begin{aligned}\sum_{k=n}^{+\infty} \mu(x+ak) &= \sum_{k=n}^{N_a} \mu(x+ak) + \sum_{k=N_a+1}^{+\infty} \mu(x+ak) \leq (N_a+1)\mu(x+an) + \frac{1}{a}\mu(x+aN_a) \\ &\leq 2\left(1+\frac{1}{a}\right)\mu(x+an).\end{aligned}$$

It is now easy to conclude:

$$\left(\sum_{k=0}^{n} \frac{1}{\mu(x+ak)}\right) \sum_{k=n}^{+\infty} \mu(x+ak) \leq 2\left(1+\frac{1}{a}\right)\left(2+\frac{1}{a}\right) = \frac{2}{a^2}(1+a)(1+2a).$$

□

4.1.4 Bound on the approximation error

The goal of this section is to derive an upper-bound on the quantity

$$J_a[g] = \int (\mu_a[g] - \mu[g])^2 d\mu,$$

which can be interpreted as the error made in integrating on the a-graduated grid rather than on the whole real line. We show that:

Lemma 4.1.9 (Error bound)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be a function with bounded derivative on $\mathbb{R}$. For all $a > 0$, the following relation holds:*

$$J_a[g] \ \leq \ \|g'\|_\infty^2 \frac{a^2}{\pi^2}\left(1+\frac{a}{\sqrt{2\pi}}\right).$$

Proof. Thanks to equation (4.7), one sees that $J_a[g] = \mathrm{Var}_{\zeta_a}[\mu_a[g]]$, thus we write:

$$J_a[g] = \frac{1}{2}\int_0^a\int_0^a (\mu_{a,x}[g] - \mu_{a,y}[g])^2 d\zeta_a(x)d\zeta_a(y) = \frac{1}{2}\int_0^a\int_0^a (\mu_{a,x}[g - \tau_{y-x}\cdot g])^2 d\zeta_a(x)d\zeta_a(y),$$

where for all x, $c \in \mathbb{R}$, $\tau_c \cdot g(x) = g(x+c)$. The hypothesis on g yields $\|\tau_c \cdot g - g\|_\infty \leq |c|\|g\|_\infty$, hence

$$J_a[g] \leq \|g\|_\infty^2 * \frac{1}{2}\int_0^a\int_0^a (x-y)^2 d\xi_a(x)d\xi_a(y) = \|g\|_\infty^2 \mathrm{Var}_{\xi_a}[\xi].$$

Denote U_a the uniform probability distribution on $[0, a)$, which satisfies to a Poincaré inequality with constant a^2/π^2.

$$\mathrm{Var}_{\xi_a}[\xi] \leq \int_0^a (\xi - U_a[\xi])^2 d\xi_a \leq a \max_{x\in[0,a)} Z_{x,a}\mathrm{Var}_{U_a}[\xi] \leq a \max_{x\in[0,a)} Z_{x,a}\frac{a^2}{\pi^2},$$

so that $J_a[g] \leq a^2/\pi^2 \cdot a \max_{x\in[0,a)} Z_{x,a} \cdot \|g'\|_\infty^2$. Lemma 4.1.10 below yields the desired result. □

We finish this section by a technical result needed in the previous proof.

Lemma 4.1.10 (Bound on the renormalizing constant)**.** *For all $a > 0$ and $x \in \mathbb{R}$,*

$$Z_{x,a} := \sum_{n\in\mathbb{Z}} \mu(x+an) \leq \frac{1}{a} + \frac{1}{\sqrt{2\pi}}.$$

Proof. Observe that:

$$\forall y \geq a, \quad \mu(y) \leq \frac{1}{a}\int_{y-a}^{y} \mu(s)ds, \qquad\qquad \forall y \leq -a, \quad \mu(y) \leq \frac{1}{a}\int_{y}^{y+a} \mu(s)ds.$$

Without loss of generality, assume that $x \in [0, a)$, hence:

$$Z_{x,a} \leq \sum_{n=0,-1} \mu(x+an) + \frac{1}{a}\int_{x}^{+\infty} \mu(s)ds + \frac{1}{a}\int_{-\infty}^{x-a} \mu(s)ds.$$

Depending on the respective positions of x and $a/2$, either $\mu(x)$ or $\mu(x-a)$ is bounded by $a^{-1}\int_{x-a}^{x} \mu(s)ds$ and the opposite term is bounded by $(2\pi)^{-1/2}$, which concludes the proof. □

4.1.5 Proof of Theorems 4.0.1 and Theorem 4.1.4

We begin by a result on the variance of the density.

Lemma 4.1.11 (A priori bound on the variance of the density)**.** *Assume that the initial distribution of the process at time* 0 *has a square-integrable density f_0 with respect to μ, and denote by f_t the density of the marginal distribution of the process at time $t \geq 0$. In the reversible setting, either with* ($\mathbf{H}^{\mathbf{Hom}}_{\mathbf{Rev}}$) *or* ($\mathbf{H}_{\mathbf{Rev}}$)*, one has*

$$Var_\mu[f_t] \leq Var_\mu[f_0], \qquad t \geq 0.$$

Proof. In the inhomogeneous setting, Kolmogorov's backward equation states that $\partial_t f_t = L^*[f_t]$, where for all $t \geq 0$, L_t^* is defined as the adjoint of L_t in $L^2(\mu)$. Equilibrium relation ($\mathbf{H}_{\mathbf{Rev}}$) is equivalent to the identity $L_t^* = L_t$ for $t \geq 0$, hence:

$$\partial_t \mathrm{Var}_\mu[f_t] = 2\int (f_t - 1)L[f_t]d\mu = -2\mathcal{E}_\mu^t[f] \geq 0.$$

The proof is similar in the homogeneous setting. □

Let us now focus on the homogeneous case, and see the proof of Lemma 4.1.1, which was stated above.

Proof of Lemma 4.1.1. For all $t \geq 0$, one has:

$$\begin{aligned}
\partial_t I_a[f_t] &= 2\int (f_t - \mu_a[f_t])\partial_t(f_t - \mu_a[f_t])d\mu \\
&= 2\int (f_t - \mu_a[f_t])L[f_t]d\mu - 2\int (f_t - \mu_a[f_t])\mu_a[L[f_t]]d\mu \\
&= 2\int (f_t - \mu_a[f_t])L[f_t]d\mu,
\end{aligned}$$

where the second inequality comes from the fact that

$$\int (f_t - \mu_a[f_t])\mu_a[L[f_t]]d\mu = \int \mu_a\left[(f_t - \mu_a[f_t])\mu_a[L[f_t]]\right]d\mu = \int \mu_a[L[f_t]]\mu_a\left[(f_t - \mu_a[f_t])\right]d\mu = 0.$$

By almost sure a-periodicity of the function $\mu_a[f]$ for all $f \in L^2(\gamma)$, one sees that $L[\mu_a[f_t]] = 0$, hence by inequality (4.12)

$$\partial_t I_a[f_t] = -2\mathcal{E}_\mu[f_t] \leq -\frac{\kappa a^2}{4(1+a)(1+2a)}\, I_a[f_t],$$

and the expected result follows by Gronwall's lemma. □

We are now ready to prove Theorem 4.0.1 stated in the introduction.

Proof of Theorem 4.0.1. Let $g : \mathbb{R} \to \mathbb{R}$ be a function such that $\|g'\|_\infty \leq 1$. By polarization of relation (4.8) and Cauchy-Schwarz inequality,

$$\mathbb{E}\left[g(Z_t)\right] - \mu[g] = \int g(f_t - 1)d\mu = \int (g - \mu_a[g])(f_t - \mu_a[f_t])d\mu + \int (\mu_a[g] - \mu[g])(\mu_a[f_t] - 1)d\mu,$$

so that

$$|\mathbb{E}\left[g(Z_t)\right] - \mu[g]| \leq \left(I_a[g]\, I_a[f_t]\right)^{1/2} + \left(J_a[g]\, J_a[f_t]\right)^{1/2} \leq \left(\mathrm{Var}_\mu[g]\, I_a[f_t]\right)^{1/2} + \left(J_a[g]\, \mathrm{Var}_\mu[f_t]\right)^{1/2}.$$

The Poincaré inequality for the normal distribution yields $\mathrm{Var}_\mu[g] \leq \int g'^2 d\mu \leq 1$, and by Lemma 4.1.11 we know that $\mathrm{Var}_\mu[f_t] \leq \mathrm{Var}_\mu[f_0]$. Accordingly,

$$|\mathbb{E}\left[g(Z_t)\right] - \mu[g]| \leq \left(I_a[f_t]\right)^{1/2} + \left(\mathrm{Var}_\mu[f_0]\, J_a[g]\right)^{1/2}. \tag{4.15}$$

Applying Lemmas 4.1.1 and 4.1.9 above, this leads to

$$|\mathbb{E}\left[g(Z_t)\right] - \mu[g]| \leq \mathrm{Var}_\mu[f_0]^{1/2} \left(\exp\left(-\frac{\kappa a^2}{8(1+a)(1+2a)} \right) + \frac{a}{\pi}\left(1 + \frac{a}{\sqrt{2\pi}}\right)^{1/2} \right).$$

An argument of density allows to extend the majoration to all functions g which are 1-Lipschitz, which concludes the proof according to the definition of Wasserstein distance $W\left(\mathcal{L}(Z_t), \mu\right)$. □

Let us now turn to the time-inhomogeneous case. We begin by showing Lemma 4.1.2.

Proof of Lemma 4.1.2. Consider a function $f \in \mathcal{C}_b^1(\mathbb{R})$, the state of continuously differentiable functions with bounded derivative. Remember that notation ξ stands for the identity function on $\mathbb{R}$, $\xi(x) = x$. First, let us show that for all $x \in \mathbb{R}$ and $a > 0$,

$$\nabla_x \mu_{a,x}[f] = \mu_{a,x}[\nabla f - \xi f] + \mu_{a,x}[f]\mu_{a,x}[\xi], \tag{4.16}$$

$$\nabla_a \mu_{a,x}[f] = \frac{1}{a}\left(\mu_{a,x}[\xi(\nabla f - \xi f)] + \mu_{a,x}[f]\mu_{a,x}[\xi^2]\right) - \frac{1}{a}\xi(x)\left(\mu_{a,x}[\nabla f - \xi f] + \mu_{a,x}[f]\mu_{a,x}[\xi]\right). \tag{4.17}$$

We first show (4.16). For the sake of simplicity, let us denote $\square_n = x + an$. One has:

$$\nabla_x \left(\sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right) = \sum_{n\in\mathbb{Z}} \left(\nabla f(\square_n) - \square_n f(\square_n)\right)\mu(\square_n).$$

By taking $f = 1$, one has also

$$\nabla_x Z_{a,x} = \nabla_x \left(\sum_{n\in\mathbb{Z}} \mu(x+an) \right) = -\sum_{n\in\mathbb{Z}} \square_n \mu(\square_n).$$

Hence,

$$\begin{aligned}\nabla_x \mu_{a,x}[f] &= \nabla_x \left(\frac{1}{Z_{a,x}} \sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right) \\ &= \frac{1}{Z_{a,x}} \nabla_x \left(\sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right) - \frac{\nabla_x Z_{a,x}}{Z_{a,x}} \left(\frac{1}{Z_{a,x}} \sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right) \\ &= \mu_{a,x}[\nabla f - \xi f] + \mu_{a,x}[\xi]\mu_{a,x}[f],\end{aligned}$$

which proves (4.16). The proof of (4.17) is similar. First observe that

$$\begin{aligned}\nabla_a \left(\sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right) &= \sum_{n\in\mathbb{Z}} n \left(\nabla f(\square_n) - \square_n f(\square_n) \right) \mu(\square_n) \\ &= \frac{1}{a} \sum_{n\in\mathbb{Z}} (\square_n - x) \left(\nabla f(\square_n) - \square_n f(\square_n) \right) \mu(\square_n),\end{aligned}$$

which in turn implies by taking $f = 1$ that

$$\nabla_a Z_{a,x} = -\frac{1}{a} \sum_{n\in\mathbb{Z}} \square_n (\square_n - x) \mu(\square_n).$$

Writing

$$\nabla_a \mu_{a,x}[f] = \frac{1}{Z_{a,x}} \nabla_a \left(\sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right) - \frac{\nabla_a Z_{a,x}}{Z_{a,x}} \left(\frac{1}{Z_{a,x}} \sum_{n\in\mathbb{Z}} f(x+an)\mu(x+an) \right)$$

then leads to (4.17).
As a consequence, one has for all $f \in \mathcal{C}_b^1(\mathbb{R})$:

$$\begin{aligned}\int (f - \mu_a[f]) \nabla_a \mu_a[f] d\mu &= \frac{1}{a} \int (f - \mu_a[f]) \left(\mu_a[\xi(\nabla f - \xi f)] + \mu_a[f]\mu_a[\xi^2] \right) d\mu - \frac{1}{a} \int (f - \mu_a[f]) \xi \nabla \mu_a[f] d\mu \\ &= -\frac{1}{a} \int (f - \mu_a[f]) \xi \nabla \mu_a[f] d\mu.\end{aligned}$$

Hence, if the jump size functions $(a_t)_{t\geq 0}$ is differentiable, one has for all $t \geq 0$

$$\partial_t \int (f - \mu_{a_t}[f])^2 d\mu = -2\partial_t a_t \int (f - \mu_{a_t}[f]) \nabla_a \mu_{a_t}[f] d\mu = 2\partial_t (\log a_t) \int (f - \mu_{a_t}[f]) \xi \nabla \mu_{a_t}[f] d\mu.$$

The previous equality extends to all function $f \in \mathcal{H}_{a_t}(\mu)$ by a density argument, which proves Lemma 4.1.2. $\square$

Let us finish this section by the proof of Theorem 4.1.4 under Conjecture 4.1.3.

Proof of Theorem 4.1.4. In the framework of the theorem, one has for all $t \geq 0$,

$$\begin{aligned}\partial_t I_{a_t}[f_t] &= 2 \int (f_t - \mu_{a_t}[f_t]) \partial_t (f_t - \mu_{a_t}[f_t]) d\mu \\ &= 2 \int (f_t - \mu_{a_t}[f_t]) L_t[f_t] d\mu - 2 \int (f_t - \mu_{a_t}[f_t]) \mu_{a_t}[L_t[f_t]] d\mu + \\ &\quad 2\partial_t (\log a_t) \int (f_t - \mu_{a_t}[f_t]) \xi \nabla \mu_{a_t}[f_t] d\mu,\end{aligned}$$

where the second line is a consequence of Lemma 4.1.2. By the same reasoning than in the proof of Lemma 4.1.1, it stands that

$$2\int (f_t - \mu_{a_t}[f_t])L_t[f_t]d\mu - 2\int (f_t - \mu_{a_t}[f_t])\mu_{a_t}[L_t[f_t]]d\mu = -2\mathcal{E}_\mu^t[f_t].$$

The function $(a_t)_{t\geq 0}$ being non-increasing, one has $(\partial_t \log a_t) \leq 0$, hence by Conjecture 4.1.3,

$$\partial_t(\log a_t)\int (f_t - \mu_{a_t}[f_t])\xi\nabla\mu_{a_t}[f_t]d\mu \leq 0.$$

Hence, one finds that

$$\partial_t I_{a_t}[f_t] \leq -2\,\mathcal{E}_\mu^t[f_t].$$

By Gronwall's lemma and Lemma 4.1.6, one gets the following result on $I_{a_t}[f_t]$ for all $t \geq 0$, which is analogous to Lemma 4.1.1 relative to the homogeneous case:

$$I_{a_t}[f_t] \leq I_{a_0}[f_0]\exp\left(-\frac{1}{4}\int_0^t \frac{\lambda_u a_u^2}{(1+a_u)(1+2a_u)}\right).$$

The same reasoning than in Theorem 4.0.1 then allows to conclude. □

4.2 Extension to asymptotically normal processes

4.2.1 Conjectured result and examples

This section is devoted to the case where the process is no longer reversible but only asymptotically normal, as is the case for the quantile algorithm which is the root motivation of this work. By *asymptotically normal*, we mean that the instantenous generator converges as time goes to $+\infty$ to the generator of a diffusion with invariant measure μ, that is the generator of the Ornstein-Uhlenbeck process. Assuming that the jump size $(a_t)_{t\geq 0}$ goes to 0, we have for smooth functions f:

$$L_t[f](z) \underset{t\to 0}{\sim} \frac{a_t^2}{2}(p_t(z)+q_t(z))f''(z) + a_t(p_t(z)-q_t(z))f'(z) + o_z(a_t^2), \qquad z \in \mathbb{R},$$

where $o_z(a_t^2)$ depends of the point z (and of derivatives of f). Recall that if $L[f] = \sigma^2 f'' + bf'$ is the generator of a diffusion process with drift b and diffusion function σ^2, it admits an invariant measure proportionnal to

$$\frac{1}{\sigma^2(z)}\exp\left(-\int^z \frac{b(u)}{\sigma^2(u)}du\right)dz.$$

Accordingly, if the process is expected to converge to μ, it makes sense to assume that there exists a positive function $(C_t)_{t\geq 0}$ such that

$$p_t(z) + q_t(z) = C_t(1 + o_z(1)), \qquad p_t(z) - q_t(z) = C_t a_t(-z + o_z(1)), \qquad z \in \mathbb{R}.$$

This is equivalent to the following hypothesis: there exists a positive function $(C_t)_{t\geq 0}$, and two functions $(h_t(z))_{z\in\mathbb{R},t\geq 0}$, $(g_t(z))_{z\in\mathbb{R},t\geq 0}$ satisfying to the pointwise convergence $h_t(z) \to 0$, $g_t(z) \to 0$ for all $z \in \mathbb{R}$, such that

$$p_t(z) = C_t(1 - \frac{a_t}{2}z + h_t(z) + a_t g_t(z)), \quad q_t(z) = C_t(1 + \frac{a_t}{2}z + h_t(z) - a_t g_t(z)), \qquad z \in \mathbb{R}. \tag{$\mathbf{H_N}$}$$

The notation $(\mathbf{H_N})$ refers to the asymptotic normality of the process. It implies that

$$\epsilon_t(z) := q_t(z)\mu(z) - p_t(z-a_t)\mu(z-a_t) \underset{t\to+\infty}{\to} 0, \qquad z \in \mathbb{R}, \tag{4.18}$$

which can be seen as an asymptotic form of the reversibility relation $(\mathbf{H_{Rev}})$.
In fact, Hypothesis $(\mathbf{H_N})$ gives information on the magnitude of the asymptotic invariance. Denote by L_t^* the adjoint in $L^2(\mu)$ of the operator L_t and recall that the invariance is equivalent to the identity $L_t^*[\mathbf{1}] = 0$. We observe that:

$$L_t^*[\mathbf{1}](z) = \frac{1}{\mu(z)}(\epsilon_t(z+a_t) - \epsilon_t(z)), \qquad z \in \mathbb{R}, \quad t \geq 0,$$

hence relation $(\mathbf{H_N})$ implies the pointwise asymptotic invariance $L_t^*[\mathbf{1}](z) \to 0$, $z \in \mathbb{R}$. Assuming that the functions h_t and g_t coming into play in equation $(\mathbf{H_N})$ are smooth, with derivatives converging towards 0 at each point, it even stands that

$$L_t^*[\mathbf{1}](z) \underset{t\to 0}{\ll} C_t a_t^2, \qquad z \in \mathbb{R}. \tag{4.19}$$

Let us now explain how to adapt the ideas of the preceding section to the non-reversible case: on one hand, we introduce a *reversibilization* of the dynamic, which consists in an alternative process, again with generator of type (4.1), reversible with respect to μ. This extends an idea of Fill (Fill [1991]) relative to Markov processes which we detail now. Fill considers the case of a homogeneous Markov process $(Q_t)_{t\geq 0}$, with generator L, invariant with respect to the probability distribution ν, and proposes to write for all functions g in the domain of the generator

$$\int gL[g]d\nu = \frac{1}{2}\int g(L+L^*)[g]d\nu,$$

where L^* stands for the adjoint of L in $L^2(\nu)$. This seemingly innocuous equation reveals indeed quite powerful, as it allows to transform the generator of a non-reversible Markov process into the generator of a reversible one. In our case, for all fixed time $t \geq 0$ the generator L_t is not even invariant in the general case. Taking over an idea introduced for stochastic algorithms finding means on a manifold or on a graph (Arnaudon and Miclo [2016], Gadat et al. [2016]), we adapt the preceding idea to the non-invariant case by introducing the operator

$$A(L)[g](z) = L[g](z) + L^*[g](z) - L^*[\mathbf{1}](z)g(z), \qquad z \in \mathbb{R}.$$

Then,

$$\int gL[g]d\nu = \frac{1}{2}\left(\int gA(L)[g]d\nu + \int g^2 L^*[\mathbf{1}]d\nu\right).$$

As we will see below, in our case the operator $A(L_t)$ is again a Markovian generator of type (4.1) and corresponds to a reversible process with respect to μ. This reversibilized process has good ergodicity properties provided that $(\mathbf{H^1_{Ergo}})$ holds (with $\lambda_t = C_t$ to be consistent with $\mathbf{H_N}$): we will see that the spectral gap on a restricted domain is of order $C_t a_t^2$.
The function $L_t^*[\mathbf{1}]$ coming up in the remaining term

$$\int g^2 L_t^*[\mathbf{1}]d\mu$$

can be thought as a way to quantify the amount by which the process at plays differs from its reversibilized counterpart. Observe that equation (4.19) gives the relative order of magnitude of the quantities at play: the quantity measuring the difference between the reversible and the non-reversible processes is required to be negligible with respect to the spectral gap.

Actually, a stronger, uniform version of (4.19), which is a consequence of the asymptotic normality assumption ($\mathbf{H_N}$), is needed. In the light of Remark 4.1.5 and for the sake of simplicity, let us restrict the analysis to the case where the simplified assumption ($\mathbf{H^2_{Ergo}}$) introduced in Remark 4.1.5 holds. In this framework, the required condition relative to the asymptotic invariance reads:

$$\exists C > 0, \quad \max_{z\in\mathbb{R}} |L_t^*[\mathbf{1}](z)| \le C\, a_t, \qquad t \ge 0. \tag{$\mathbf{H^1_{SN}}$}$$

The notation ($\mathbf{H^1_{SN}}$) refers to the strong asymptotic normality of the process.

Under Hypotheses ($\mathbf{H^2_{Ergo}}$) and ($\mathbf{H^1_{SN}}$), and if Conjecture 4.1.3 holds, the following result allows to quantify in Wasserstein distance the convergence of the process towards its asymptotic distribution μ:

Theorem 4.2.1 (Ergodicity in Wasserstein distance)**.** *We suppose that the ergodicity condition* ($\mathbf{H^2_{Ergo}}$) *and the strong asymptotic invariance condition* ($\mathbf{H^1_{SN}}$) *stand. Furthermore, we assume that the size of the jumps is of the form:*

$$a_t = \frac{a_0}{(t+1)^r}, \qquad t \ge 0; \quad a_0 > 0, \quad r > 1,$$

with $a_0 < A$, and that and that the marginal of the process at time 0 admits a density $f_0 \in \mathcal{H}_{a_0}(\mu)$ with respect to the normal distribution μ.

If Conjecture 4.1.3 is true, then there exists a time $T > 0$ and explicit constants c_1, c_2, c_3, c_4 > 0 such that for all $t \ge T$,

$$W\left(Law(Z_t), \mu\right) \le \frac{c_1}{(t/2+1)^{r-1}} + \frac{c_2}{(t+1)^r} + c_3 \exp\left(-\frac{\lambda}{32}t\right) + c_4 \exp\left(-\frac{\lambda}{64}t\right).$$

To streamline the paper, the proof of this result is postponed to section 4.2.3.

Although the process at play is weakly mixing, we are able to provide a decreasing upper-bound on the Wasserstein distance between its marginals and the asymptotic distribution μ if Conjecture 4.1.3 is true. This comes at the cost of little explicit assumptions: Hypothesis ($\mathbf{H^1_{SN}}$) amounts to a uniform Taylor expansion which needs to be checked on a case-by-case basis.

Let us give instances of processes and check wether the assumptions of Theorem 4.2.1 are satisfied. The process with rates

$$\begin{aligned}
&\forall z > \frac{a_t}{2}, \qquad q_t(z) = \frac{1}{a_t^2}\,; \qquad \forall z \le \frac{a_t}{2}, \qquad q_t(z) = \frac{1}{a_t^2}\exp\left(a_t z - \frac{a_t^2}{2}\right), \\
&\forall z \le -\frac{a_t}{2}, \qquad p_t(z) = \frac{1}{a_t^2}\,; \qquad \forall z > -\frac{a_t}{2}, \qquad p_t(z) = \frac{1}{a_t^2}\exp\left(-a_t z - \frac{a_t^2}{2}\right), \qquad t \ge 0,
\end{aligned}$$

satisfies to the reversibility condition ($\mathbf{H_{Rev}}$). One can perturbate this dynamic in the following manner: for all $t \ge 0$, define

$$\begin{aligned}
&\forall z > \frac{a_t}{2}, \quad q_t(z) = \frac{1}{a_t^2}\,; \;\; \forall z \le \frac{a_t}{2}, \quad q_t(z) = \frac{1}{a_t^2}\exp\left(a_t z - \frac{a_t^2}{2}\right) - C a_t \exp\left(-\left(\frac{1}{a_t^\rho} - 1\right)\frac{z^2}{2}\right), \\
&\forall z \le -\frac{a_t}{2}, \;\; p_t(z) = \frac{1}{a_t^2}\,; \;\; \forall z > -\frac{a_t}{2}, \;\; p_t(z) = \frac{1}{a_t^2}\exp\left(-a_t z - \frac{a_t^2}{2}\right) + C a_t \exp\left(-\frac{1}{2}\left(\frac{1}{a_t^\rho}(z+a_t)^2 - z^2\right)\right).
\end{aligned}$$

With this choice, assumptions ($\mathbf{H^2_{Ergo}}$) and ($\mathbf{H^1_{SN}}$) stand for every $\rho > 0$ and $C \in \mathbb{R}$ (chosen small enough to ensure the positivity of p, q).

On the other hand, the process with rates defined for all $t \ge 0$ as:

$$\forall z \ge 0, \qquad p_t(z) = \frac{1}{a_t^2}, \quad q_t(z) = \frac{1}{a_t^2}(1 + a_t z)\,; \quad \forall z \le 0, \qquad p_t(z) = \frac{1}{a_t^2}(1 - a_t z), \quad q_t(z) = \frac{1}{a_t^2}.$$

satisfies to the weak normality relation ($\mathbf{H_N}$) (with $h_t(z) = a_t|z|/2$ and $g_t(z) = 0$) and to the ergodicity relation ($\mathbf{H^2_{Ergo}}$) with $\lambda = 1$. However, one sees that

$$L_t^*[\mathbf{1}](z) = \frac{1}{a_t^2}\left(e^{a_t z - a_t^2/2} - (2 + a_t z) + (1 + a_t z + 2a_t^2)e^{-a_t z - a_t^2/2}\right), \qquad z \geq a_t, \quad t \geq 0,$$

hence for all fixed $t \geq 0$, $\max L_t^*[\mathbf{1}] = +\infty$ and Hypothesis ($\mathbf{H^1_{SN}}$) fails.

Actually, the theorem holds under a weaker but more intricate form of ($\mathbf{H^1_{SN}}$):

Remark 4.2.2 (On Hypothesis ($\mathbf{H^1_{SN}}$))**.** *In the light of Lemma 4.2.6 below, Hypothesis* ($\mathbf{H^1_{SN}}$) *can be weakened in*

$$\exists C_1, C_2 > 0, \quad \forall t \geq 0, \quad \max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z) \leq C_1 a_t\,, \quad \min_{z\in\mathbb{R}} L_t^*[\mathbf{1}] > -\infty\,, \quad \left(\int (L_t^*[\mathbf{1}])^2 d\mu\right)^{1/2} \leq C_2 a_t. \tag{$\mathbf{H^2_{SN}}$}$$

4.2.2 Median algorithm

The inhomogeneous jump process which this article investigates is related to a recursive stochastic algorithm finding the median of a probability distribution. This stochastic algorithm is a special case of the quantile algorithm, which we recall now. Set $\alpha \in [1/2, 1)$ and $(X_n)_{n\geq 1}$ identically distributed, indepedent real random variables, with density φ with respect to the Lebesgue measure on $\mathbb{R}$ and repartition function F. The quantile of level α of the distribution φ is defined as:

$$q_\alpha := \inf\left\{x \in \mathbb{R}, \int_x^{+\infty} \varphi(y)dy \geq 1 - \alpha\right\}.$$

If $\varphi(q_\alpha) > 0$, the following recursive algorithm is a strongly consistent estimator of q_α ([Duflo, 1997, Theorem 1.4.26], Blum [1954]):

$$Y_{n+1} = Y_n + \gamma_{n+1}(\mathbf{1}_{X_{n+1}>Y_n} - \alpha), \qquad n \geq 1, \tag{4.2}$$

where $(\gamma_n)_{n\geq 1}$ is a sequence of steps of the form $\gamma_n = \frac{c}{n^\beta}$, $n \geq 1$ with $c > 0$ and $0 < \beta \leq 1$. In the following development, it is assumed that $\beta < 1$. The algorithm (4.2) rewrites:

$$Y_{n+1} = Y_n - \gamma_{n+1}H(X_{n+1}, Y_n), \qquad n \geq 1,$$

where $H(x, y) = -(\mathbf{1}_{x>y} - \alpha)$ for x, $y \in \mathbb{R}$. We see that $h(y) := \mathbb{E}\left[H(X, y)\right] = F(y) - (1 - \alpha)$, hence:

$$\forall y \geq q_\alpha, \quad \mathbb{E}\left[H(X, y)\right] \geq 0\,, \qquad \forall y \leq q_\alpha, \quad \mathbb{E}\left[H(X, y)\right] \leq 0. \tag{4.20}$$

The relation (4.20) intuitively explains the convergence of the algorithm: if the quantile is overestimated, the correction is negative; if it is underestimated, it is positive. Accordingly, in the usual stochastic algorithm approach, one uses the formulation $Y_{n+1} = Y_n - \gamma_{n+1}\left(h(Y_n) + \Delta M_{n+1}\right)$, where ΔM_{n+1} is the delta of a martingale.

The algorithm (4.2) satisfies to a Central Limit Theorem:

Theorem 4.2.3 (CLT for the quantile algorithm, Benveniste et al. [1990])**.** *Assume that the common distribution of the X_i's has a continuously differentiable density φ with respect to the Lebesgue distribution on $\mathbb{R}$, with $\varphi(q_\alpha) > 0$. If the sequence of steps is of the form $\gamma_n = cn^{-\beta}$ with $0 < \beta < 1$, one has the following convergence in distribution:*

$$\widetilde{Y}_n := \frac{1}{\sqrt{\gamma_n}}(Y_n - q_\alpha) \quad \hookrightarrow \quad \mathcal{N}(0, \sigma^2), \qquad \sigma^2 = \frac{\alpha(1-\alpha)}{2\varphi(q_\alpha)}.$$

Remark 4.2.4 (Case $\beta = 1$). *If $\gamma_n = cn^{-1}$, Theorem 4.2.3 still holds if $\varphi(q_\alpha) > c/2$. If $\varphi(q_\alpha) \leq c/2$, one should choose a different renormalization as explained in [Duflo, 1997, Theorem 2.2.12]. This comes from the fact that the quantity $(\gamma_n/\gamma_{n+1})^{1/2}$ is negligible with respect to γ_{n+1} if and only if $\beta < 1$.*

The goal is to go further in the analysis of the quantile algorithm through a Berry-Esseen bound for the algorithm, that is to say a quantification of the convergence stated in Theorem 4.2.3 for an adequate distance. In the present note, we focus on a simplified model, and consider a process which plays the continuous-time counterpart to the renormalized algorithm $(\widetilde{Y}_n)_{n\geq 1}$ defined above. To lighten the notation, we suppose that $q_\alpha = 0$. Observe that

$$\widetilde{Y}_{n+1} = \theta_{n+1}\widetilde{Y}_n + \sqrt{\gamma_{n+1}}(\mathbf{1}_{X_{n+1}>\theta_{n+1}\sqrt{\gamma_{n+1}}\widetilde{Y}_n} - \alpha), \qquad n \geq 1,$$

with $\theta_{n+1} := (\gamma_n/\gamma_{n+1})^{1/2}$. In the case $0 < \beta < 1$, the factor $|1-\theta_n|$ is negligible with respect to γ_n, which justifies that as a first simplification of the model, θ_n is supposed to be 1. This writes

$$\widetilde{Y}_{n+1} = \widetilde{Y}_n + \sqrt{\gamma_{n+1}}(\mathbf{1}_{X_{n+1}>\sqrt{\gamma_{n+1}}\widetilde{Y}_n} - \alpha), \qquad n \geq 1.$$

We are left with an inhomogeneous Markov chain with transition semigroup $(Q_{p,q})_{q\geq p\geq 1}$, where for all regular functions f and parameter $\gamma > 0$ and for all $y \in \mathbb{R}$,

$$Q_{n,n+1}[f] = Q_{\gamma_n}[f], \quad n \geq 1\,; \quad Q_\gamma[f](y) = (1 - F(\sqrt{\gamma}y))f(y + \sqrt{\gamma}(1-\alpha)) + F(\sqrt{\gamma}y)f(y - \sqrt{\gamma}\alpha).$$

The natural sequence of times associated to the recursive algorithm (4.2) is $T_n = \sum_{k=1}^n \gamma_k$. Taking this into account, the Poissonization of the Markov chain leads to an inhomogeneous Markov process with instantaneous generator $(L_{\eta_t})_{t\geq 0}$, where we defined for all regular functions and $z \in \mathbb{R}$:

$$\begin{aligned} L_\eta[f](z) &= \frac{1}{\eta^2}\left((1 - F(\eta z))(f(z + (1-\alpha)\eta) - f(z)) + F(\eta z)(f(z - \alpha\eta) - f(z))\right), \\ \eta_t &= C(c,\beta)\frac{1}{t^{\beta/(2(1-\beta))}}, \qquad C(c,\beta) := c^{-1/(2(1-\beta))}\beta^{-\beta/(2(1-\beta))}. \end{aligned}$$

To obtain these formulas, we first consider the continuous-time process with generator $(L_{\gamma_u})_{u\geq 0}$, where $L_\gamma = Q_\gamma - \mathrm{Id}$, and the function $\gamma_u = \gamma_0(u+1)^{-\beta}$ serves as the natural counterpart of the sequence $(\gamma_n)_{n\in\mathbb{N}}$. We get the final form above by the change of time $t(u) = \int_0^u \gamma_s ds$ corresponding to the discrete change of time $T_n = \sum_{k=1}^n \gamma_k$.
The median algorithm corresponds to the case $\alpha = \frac{1}{2}$. The associated continuous-time dynamic is indeed of type (4.1), with

$$\alpha_t = \frac{\eta_t}{2}, \qquad p_t(z) = \frac{1 - F(\eta_t z)}{\eta_t^2}, \qquad q_t(z) = \frac{F(\eta_t z)}{\eta_t^2}; \qquad z \in \mathbb{R}, \qquad t \geq 0.$$

Let us examine whether Theorem 4.2.1 holds. For $2/3 < \beta < 1$, the assumption on the jump size $(a_t)_{t\geq 0}$ is satisfied. Furthermore, one sees that for all $t \geq 0$,

$$\forall z \leq 0, \quad p_t(z) \geq \frac{1 - F(0)}{\eta_t^2} = \frac{1}{8a_t^2}\,, \qquad \forall z \geq 0, \quad q_t(z) \geq \frac{F(0)}{\eta_t^2} = \frac{1}{8a_t^2}\,,$$

hence Hypothesis $(\mathbf{H^2_{Ergo}})$ holds with $\lambda = 1/8$. If the density φ is smooth (differentiable three times on $\mathbb{R}$ for example), the asymptotic normality $(\mathbf{H_N})$ holds, and

$$L_t^*[\mathbf{1}](z) \underset{t\to 0}{=} O_z(a_t^3), \qquad z \in \mathbb{R}.$$

Do the continuous-time algorithm satisfy to Assumption ($\mathbf{H^1_{SN}}$)? The explicit expression of $L_t^*[\mathbf{1}]$ reads:

$$L_t^*[\mathbf{1}](z) = \frac{1}{4a_t^2}\left((1-F(2a_t(z-a_t))e^{a_tz-\frac{a_t^2}{2}} + F(2a_t(z+a_t))e^{-a_tz-\frac{a_t^2}{2}} - 1\right), \qquad z\in\mathbb{R}, \quad t\geq 0.$$

Hence, if the distribution of the $X_i's$ is compactly supported, for fixed $t\geq 0$ we have $L_t^*[\mathbf{1}](z)\to -1/(4a_t^2)$ as z goes to ∞ and Assumption ($\mathbf{H^1_{SN}}$) fails. It could actually be expected: when the distribution of the $X_i's$ is compactly supported, the conditional birth-death process at the core of our analysis lives on a finite set, while its reversibilization lives on the whole grid. Hence, it makes sense that the function $L_t^*[\mathbf{1}]$, that can be thought of as a measure of the difference between the process and its reversibilization, would not go to 0 on the edges of the domain. However, the weakened form ($\mathbf{H^2_{SN}}$) stated in Remark 4.2.2 may stand.

Unfortunately, there do not seem to exist any simple assumptions on the distribution of the $X_i's$ ensuring that either ($\mathbf{H^1_{SN}}$) or ($\mathbf{H^2_{SN}}$) holds. The information which one can take away is that the continuous-time median process, and by extension the quantile algorithm, may simply be *too* weakly mixing. Indeed, observe that in the algorithm (4.2), the size of the jumps does not depend from the distance between the current approximation Y_n and the new information X_{n+1}, but only from its sign. This makes the mean reverting effect quite weak, and prevents other quantification results relying on a Lipschitz property for h (Bach and Moulines [2011]; Frikha and Menozzi [2012]) to be applied.

To improve this situation, we would need a function H satisfying to the relation (4.20), and additionaly such that $|H(x,y)|$ is stricly increasing with $|x-y|$, but up to the authors' knowledge, such a function does not exist. Extending a work of Holst (Holst [1987]), Amiri and Thiam (Amiri and Thiam [2014]) have proposed to use smoothing kernels to produce a sequence of functions $(H_n)_{n\geq 1}$ with the desired increasing property and such that relation (4.20) holds asymptotically. They derived the almost sure convergence and asymptotic normality of the corresponding stochastic algorithm, opening the door to future investigations relative to Berry-Esseen bounds for this algorithm.

4.2.3 Proof of Theorem 4.2.1

Taking into account the non-reversibility of the process, the goal of this section is to adapt the ideas of the reversible case developed in the preceding part. While the bound on the error approximation detailed in Lemma 4.1.9 is still valid, we need to find counterparts for the spectral gap property (Lemma 4.1.6) and the *a priori* bound on the variance (Lemma 4.1.11). Let us first introduce the reversibilization of the process. By simple calculations, one shows that for functions $f\in L^2(\mu)$, for almost all $z\in\mathbb{R}$, and for all $t\geq 0$,

$$L_t^*[f](z) = p_t(z-a_t)\frac{\mu(z-a_t)}{\mu(z)}f(z-a_t) - p_t(z)f(z) + q_t(z+a_t)\frac{\mu(z+a_t)}{\mu(z)}f(z+a_t) - q_t(z)f(z).$$

For a Markov generator L and its adjoint L^* with respect to a probability measure μ, we define the reversibilization $A(L)$ for sufficiently regular functions as:

$$A(L)[f](z) = L[f](z) + L^*[f](z) - L^*[\mathbf{1}](z)f(z), \qquad z\in\mathbb{R}.$$

The operator $A(L)$ is reversible with respect to the measure μ, and represents an extension to the non-invariant case to the additive reversibilization for invariant but non-reversible continuous-time processes of [Fill, 1991, Section 2.7]. $A(L)$ is Markovian, as is straightforward to check on the explicit expression in the present case: for a regular function f and for all $t\geq 0$,

$$\begin{aligned} A(L_t)[f](z) = &\left(p_t(z) + q_t(z+a_t)\frac{\mu(z+a_t)}{\mu(z)}\right)(f(z+a_t)-f(z)) \\ &+ \left(q_t(z) + p_t(z-a_t)\frac{\mu(z-a_t)}{\mu(z)}\right)(f(z-a_t)-f(z)), \qquad f\in L^2(\mu), \quad \text{a.a. } z\in\mathbb{R}. \end{aligned}$$

The reversibilized process admits similar ergodicity properties than the reversible process of the preceding section, as shows the following lemma:

Lemma 4.2.5 (Spectral gap). *For all $t \geq 0$, denote $\mathcal{E}^t_\mu[f]$ the Dirichlet energy associated to $A(L_t)$ and μ:*

$$\mathcal{E}^t_\mu[f] = -\int f A(L_t)[f] d\mu, \qquad f \in L^2(\mu).$$

Under ergodicity condition ($\mathbf{H^1_{Ergo}}$), the following relation stands:

$$\mathcal{E}^t_\mu[f] \geq \frac{1}{8} \frac{1}{(1+a_t)(1+2a_t)} \lambda_t \, a_t^2 \, I_{a_t}[f], \qquad f \in L^2(\mu), \qquad t \geq 0.$$

Proof. Apply Lemma 4.1.6 to the reversible process generated by $A(L_t)$, whose rates satisfy under ($\mathbf{H^1_{Ergo}}$) to:

$$p_t(z) + q_t(z+a_t)\frac{\mu(z+a_t)}{\mu(z)} \geq p_t(z) \geq \lambda_t, \quad z \leq 0; \quad q_t(z) + p_t(z-a_t)\frac{\mu(z-a_t)}{\mu(z)} \geq q_t(z) \geq \lambda_t, \quad z \geq 0.$$

□

Let us now focus on the conditional ergodicity,

$$I_{a_t}[f_t] = \int (f - \mu_{a_t}[f])^2 d\mu, \qquad t \geq 0,$$

when the marginal of the process at time 0 admits a density $f_0 \in \mathcal{H}_{a_0}(\mu)$ with respect to the normal distribution μ and Conjecture 4.1.3 holds, and if $\mathbf{H^1_{Ergo}}$ is satisfied.
By the identity $\partial_t f_t = L^*_t[f_t]$, $t \geq 0$ and by Lemma 4.1.2, we write:

$$\begin{aligned}
\partial_t I_{a_t}[f_t] &= 2\int (f_t - \mu_{a_t}[f_t]) \partial_t (f_t - \mu_{a_t}[f_t]) d\mu \\
&= 2\int (f_t - \mu_{a_t}[f_t])(L^*_t[f_t] - \mu_{a_t}[L^*_t[f_t]]) d\mu + 2\partial_t(\log a_t) \int (f_t - \mu_{a_t}[f_t]) \xi \nabla \mu_{a_t}[f_t] d\mu \\
&\leq 2\int (f_t - \mu_{a_t}[f_t]) L^*_t[f_t] d\mu,
\end{aligned}$$

where the majoration comes from Conjecture 4.1.3 and the fact that for a fixed $t \geq 0$, the function $x \mapsto \mu_{a_t,x}[L^*_t[f_t]]$ is measurable with respect to the same σ-algebra as μ_{a_t}. Going back to the first equation,

$$\begin{aligned}
\partial_t I_{a_t}[f_t] &\leq 2\int (f_t - \mu_{a_t}[f_t]) L^*_t[f_t - \mu_{a_t}[f_t]] d\mu + 2\int (f_t - \mu_{a_t}[f_t]) L^*_t[\mu_{a_t}[f_t]] d\mu \\
&= \int (f_t - \mu_{a_t}[f_t]) A(L_t)[(f_t - \mu_{a_t}[f_t])] d\mu + \int (f_t - \mu_{a_t}[f_t])^2 L^*_t[\mathbf{1}] d\mu \\
&+ 2\int (f_t - \mu_{a_t}[f_t]) L^*_t[\mu_{a_t}[f_t]] d\mu,
\end{aligned}$$

Thanks to Lemma 4.2.5, we are left with:

$$\partial_t I_{a_t}[f_t] \leq -\frac{\lambda_t \, a_t^2}{8(1+a_t)(1+2a_t)} I_{a_t}[f_t] + \int (f_t - \mu_{a_t}[f_t])^2 L^*_t[\mathbf{1}] d\mu + 2\int (f_t - \mu_{a_t}[f_t]) L^*_t[\mu_{a_t}[f_t]] d\mu.$$

Let us focus on the term on the first line. A rough majoration would consist in writing

$$\int (f_t - \mu_{a_t}[f_t])^2 L_t^*[\mathbf{1}] d\mu \leq \max_{z\in\mathbb{R}} \{|L_t^*[\mathbf{1}](z)|\}\ I_{a_t}[f_t].$$

Let us observe that, for a nonnegative function $g \in L^2(\mu)$ and a bounded function h,

$$\int gh\, d\mu \leq \max\{h(z),\, z\in\mathbb{R}\} \int g\, d\mu$$

(optimize in $c \in \mathbb{R}$ the sum $\int g(h-c)d\mu + c\int g d\mu$), which is better than the majoration

$$\int gh\, d\mu \leq \max\{|h(z)|,\, z\in\mathbb{R}\} \int g\, d\mu.$$

Hence, if for all $t \geq 0$ the function $L_t^*[\mathbf{1}]$ is bounded, we have:

$$\partial_t I_{a_t}[f_t] \leq -\eta_t I_{a_t}[f_t] + 2\int (f_t - \mu_{a_t}[f_t]) L_t^*[\mu_{a_t}[f_t]] d\mu,$$

where the constant η_t stands for:

$$\eta_t := \frac{\lambda_t\, a_t^2}{8(1+a_t)(1+2a_t)} - \max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z).$$

Thanks to the almost sure a_t-periodicity of the function $x \mapsto \mu_{a_t,x}[f]$ for all $f \in L^2(\gamma)$ and to the explicit expression of L_t^*, one sees that $L_t^*[\mu_{a_t}[f_t]] = \mu_{a_t}[f_t] L_t^*[\mathbf{1}]$. By successive applications of Cauchy-Schwarz inequality, the remaining term writes

$$\begin{aligned}
\int (f_t - \mu_{a_t}[f_t]) L_t^*[\mu_{a_t}[f_t]] d\mu &= \int (f_t - \mu_{a_t}[f_t]) \mu_{a_t}[f_t] L_t^*[\mathbf{1}] d\mu \\
&= \int (f_t - \mu_{a_t}[f_t])(\mu_{a_t}[f_t]-1) L_t^*[\mathbf{1}] d\mu + \int (f_t - \mu_{a_t}[f_t]) L_t^*[\mathbf{1}] d\mu \\
&\leq \max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z) \left(1 + J_{a_t}[f_t]^{1/2}\right) I_{a_t}[f_t]^{1/2}.
\end{aligned}$$

Finally, we get that:

$$\partial_t I_{a_t}[f_t] \leq -\eta_t I_{a_t}[f_t] + 2\max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z) \left(1 + \mathrm{Var}_\mu[f_t]^{1/2}\right) I_{a_t}[f_t]^{1/2}.$$

Let us now turn to the evolution of $\mathrm{Var}_\mu[f_t]$, to provide an alternative to the *a priori bound* stated in Lemma 4.1.11. One has

$$\begin{aligned}
\partial_t \mathrm{Var}_\mu[f_t] &= 2\int (f_t-1) L_t^*[f_t] d\mu = 2\int (f_t-1) L_t^*[f_t-1] d\mu + 2\int (f_t-1) L_t^*[\mathbf{1}] d\mu \\
&= \int (f_t-1) A(L_t^*)[f_t-1] d\mu + \int L_t^*[\mathbf{1}](f_t-1)^2 d\mu + 2\int (f_t-1) L_t^*[\mathbf{1}] d\mu \\
&\leq -\frac{\lambda_t\, a_t^2}{8(1+a_t)(1+2a_t)} I_{a_t}[f_t] + \left(\max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z)\right) \mathrm{Var}_\mu[f_t] + 2\mu\left[L_t^*[\mathbf{1}]^2\right]^{1/2} \mathrm{Var}_\mu[f_t]^{1/2} \\
&\leq \left(\max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z)\right) \mathrm{Var}_\mu[f_t] + 2\mu\left[L_t^*[\mathbf{1}]^2\right]^{1/2} \mathrm{Var}_\mu[f_t]^{1/2},
\end{aligned}$$

as soon as the function $L_t^*[\mathbf{1}]$ is bounded. The following lemma sums up the situation.

Lemma 4.2.6 (Evolution system). *Let us assume that Hypothesis* ($\mathbf{H^1_{Ergo}}$) *holds and that the function* $L_t^*[\mathbf{1}]$ *is bounded. Furthermore, we suppose that the marginal of the process at time* 0 *admits a density* $f_0 \in \mathcal{H}_{a_0}(\mu)$ *with respect to the normal distribution* μ*, and denote by* f_t *the density of the marginal distribution of the process at time* $t \geq 0$*. For all* $t \geq 0$*, let us call:*

$$I_t := I_{a_t}[f_t], \qquad K_t := Var_\mu[f_t],$$

and introduce the quantities:

$$\eta_t := \frac{\lambda_t\, a_t^2}{8(1+a_t)(1+2a_t)} - \max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z), \qquad \iota_t := \mu\left[L_t^*[\mathbf{1}]^2\right]^{1/2}, \qquad \kappa_t = \max_{z\in\mathbb{R}} L_t^*[\mathbf{1}](z).$$

If Conjecture 4.1.3 is true, the couple (I_t, K_t) *satisfies to the following system:*

$$\partial_t I_t \leq -\eta_t I_t + 2\kappa_t \left(1 + K_t^{1/2}\right) I_t^{1/2}, \qquad \partial_t K_t \leq \kappa_t K_t + 2\iota_t K_t^{1/2}, \qquad t \geq 0.$$

We are now ready to prove Theorem 4.2.1.

Proof of Theorem 4.2.1. With the notation of the theorem and of Lemma 4.2.6 and under the hypothesis of the theorem, one has:

$$\eta_t \geq \frac{\lambda}{8(1+a_t)(1+2a_t)} - Ca_t, \qquad \kappa_t,\, \iota_t \leq Ca_t, \qquad t \geq 0.$$

Let us first show that $\mathrm{Var}_\mu[f_t]$ is bounded. By Lemma 4.2.6, for all $t \geq 0$,

$$\frac{\partial_t K_t}{2K_t^{1/2}} \leq \frac{C}{2} a_t K_t^{1/2} + Ca_t,$$

hence by Gronwall lemma,

$$\begin{aligned} K_t^{1/2} &\leq \exp\left(\frac{C}{2}\int_0^t a_s ds\right) K_0^{1/2} + C\exp\left(\frac{C}{2}\int_0^t a_s ds\right)\int_0^t a_u \exp\left(-\frac{C}{2}\int_0^u a_s ds\right) du \\ &\leq \exp\left(\frac{C}{2}\int_0^t a_s ds\right) K_0^{1/2} + 2\left(\exp\left(\frac{C}{2}\int_0^t a_s ds\right) - 1\right) \leq \widetilde{c}, \end{aligned} \tag{4.21}$$

where we defined

$$\widetilde{c} := (K_0^{1/2} + 2)\exp\left(\frac{C}{2}\int_0^{+\infty} a_s ds\right).$$

Now,

$$\eta_t \geq \frac{\lambda}{8(1+a_t)(1+2a_t)} - Ca_t \underset{t\to+\infty}{\to} \frac{\lambda}{8},$$

hence there exists a time $T \geq 0$ such that for all $t \geq T$, $\eta_t \geq \lambda/16$ for example. Unsing again Lemma 4.2.6 and the above upper-bound on K_t, one finds that for all $t \geq T$,

$$\partial_t I_t \leq -\frac{\lambda}{16} I_t + 2C\,(1+\widetilde{c})\, a_t I_t^{1/2},$$

hence by Gronwall's lemma,

$$I_t^{1/2} \leq \exp\left(-\frac{\lambda}{32}t\right) I_T^{1/2} + C(1+\widetilde{c})\exp\left(-\frac{\lambda}{32}t\right)\int_T^t \exp\left(\frac{\lambda}{32}s\right) a_s ds.$$

Let us remark that $\forall \rho > 0$ and $\forall r > 1$, and $\forall t \geq T$, one has:

$$\exp(-\rho t) \int_T^t \exp(\rho s)(s+1)^{-r} ds \leq \frac{1}{\rho} \exp(-\rho t/2) + \frac{(t/2+1)^{-(r-1)}}{r-1},$$

as is seen by cutting the integral on the intervals $(T, t/2)$ and $(t/2, t)$ and using straighforward majorations. Hence, we find that for all $t \geq T$,

$$I_t^{1/2} \leq \frac{c_1}{(t/2+1)^{r-1}} + c_2 \exp\left(-\frac{\lambda}{32} t\right) + c_3 \exp\left(-\frac{\lambda}{64} t\right).$$

Let us conclude as in the proof of Theorem 4.0.1. By inequality (4.15), with $\mathrm{Var}_\mu[f_0]$ replaced by $\widetilde{c}$, and majoration (4.21), one can write

$$W\left(\mathrm{Law}(Z_t), \mu\right) \leq I_t^{1/2} + \widetilde{c}(1 + a_t/\sqrt{2\pi})^{1/2} \frac{a_t}{\pi}.$$

Hence, for all $t \geq T$, there exists constants c_1, c_2, c_3, c_4 such that

$$W\left(\mathrm{Law}(Z_t), \mu\right) \leq \frac{c_1}{(t/2+1)^{r-1}} + c_2 \exp\left(-\frac{\lambda}{32} t\right) + c_3 \exp\left(-\frac{\lambda}{64} t\right) + \frac{c_4}{(t+1)^r},$$

which proves the theorem up to a renumbering of the constants. □

Deuxième partie

Entrelacements et méthode de Stein

Chapitre 5

La méthode de Stein-Chen : littérature et premiers développements

La méthode de Stein est introduite par Charles Stein dans Stein [1972] comme une méthode pour calculer la distance en variation totale entre une somme de variables aléatoires et la gaussienne. Elle est adaptée par Louis Chen dans Chen [1975] pour le cas de la loi de Poisson et de variables discrètes, et prend ainsi le nom de méthode de Stein-Chen dans le cadre discret. Tout en se révélant particulièrement efficace, la méthode de Stein demande des calculs spécifiques à chaque case et parfois assez lourds, qu'il convient d'automatiser au maximum. L'utilisation des facteurs de Stein répond à cette motivation. La contribution principale de cette thèse dans le domaine de la méthode de Stein consiste en une méthode d'évaluation universelle des facteurs de Stein, et fait l'objet du chapitre suivant. L'objectif du présent chapitre est de rappeler les résultats de la littérature concernant la méthode de Stein-Chen, afin de replacer les résultats de la thèse dans leur contexte au sens large.

5.1 Généralités

5.1.1 Les deux étapes

Notons E un espace polonais et $\mathcal{P}(E)$ l'ensemble des mesures de probabilités sur E. L'objectif est d'évaluer la distance entre une mesure de probabilité de référence, notée $\pi \in \mathcal{P}(E)$, et la mesure de probabilité μ, souvent la loi d'une variable aléatoire W issue de la modélisation du problème examiné. On désigne la mesure π sous le nom d'approximande et sous le nom d'approximée la variable aléatoire W, ou sa loi suivant le contexte. L'espace E peut être continu, par exemple si l'approximande est la loi gaussienne, ou discret, par exemple si on approche des variables aléatoires à valeurs entières ou des processus ponctuels. Dans cette présentation on s'intéresse en particulier au cas $E = \mathbb{N}$. La distance utilisée est caractérisée par une classe de fonctions $\mathcal{F} \subset \mathbb{R}^E$ et définie comme :

$$d_{\mathcal{F}}(\mu, \pi) := \sup_{f \in \mathcal{F}} |\mu(f) - \pi(f)|. \tag{5.1}$$

Cette définition recouvre les distances classiques, comme la distance en variation totale, de Wasserstein d'ordre 1, ou de Kolmogorov.
On se donne un opérateur $S : \mathrm{Dom}(S) \subset \mathbb{R}^E \to \mathbb{R}^E$, appelé opérateur de Stein ou opérateur caractéristique de la mesure de référence π, vérifiant la propriété suivante : pour toute fonction g dans une classe de fonctions-tests assez large, l'intégrale $\int (Sg)d\mu$ s'annule si et seulement si $\mu = \pi$. L'idée de la méthode de Stein consiste à caractériser l'écart entre μ et π par la quantité

$$\left| \int (Sg) d\mu \right|.$$

Pour faire apparaître cette quantité, appelée de manière générique "erreur" dans la suite de ce texte, on introduit l'équation de Stein avec donnée $f \in \mathcal{F}$ et inconnue g_f telle que :

$$Sg_f = f - \pi(f), \qquad f \in \mathcal{F}.$$

La distance $d_{\mathcal{F}}(\mu, \pi)$ se réécrit alors sous la forme

$$d_{\mathcal{F}}(\mu, \pi) = \sup_{f\in\mathcal{F}} |\mu(Sg_f)| = \sup_{f\in\mathcal{F}} |\mathbb{E}\left[Sg_f(W)\right]|, \qquad W \sim \mu.$$

La première étape de la méthode consiste à obtenir une majoration de cette espérance sous la forme

$$|\mathbb{E}\left[Sg(W)\right]| \leq C_1(\mu, S)\|g\|_\infty + C_2(\mu, S)\|Dg\|_\infty,$$

où D désigne un opérateur linéaire agissant sur les fonctions de E dans $\mathbb{R}$, qui correspond le plus souvent à un gradient. Dans le cas où $E = \mathbb{N}$, on utilise le gradient discret défini pour toute $f : \mathbb{N} \to \mathbb{R}$ par :

$$\partial f(x) = f(x+1) - f(x), \qquad x \in \mathbb{N}.$$

Par ailleurs, on appelle facteur de Stein d'ordre $k \in \mathbb{N}$ une borne supérieure [1] sur la quantité

$$\sup_{f\in\mathcal{F}} \|D^k g_f\|_\infty.$$

où k est un entier. Le plus souvent, seuls les facteurs d'ordre 0 et 1 interviennent.

Pour récapituler, la méthode de Stein se décompose en deux étapes indépendantes :

1. Majoration de l'erreur : transformation de l'expression $\|\mathbb{E}\left[Sg(W)\right]\|$ de façon à faire apparaître les quantités $\|D^k g\|_\infty$.

2. Evaluation des facteurs de Stein : majoration de leur supremum sur l'ensemble des solutions de l'équation de Stein de données $f \in \mathcal{F}$, $\sup_{f\in\mathcal{F}} \|D^k g_f\|_\infty$.

On trouve alors :

$$d_{\mathcal{F}}(\mu, \pi) \leq C_1(\mu, S) \sup_{f\in\mathcal{F}} \|g_f\|_\infty + C_2(\mu, S) \sup_{f\in\mathcal{F}} \|Dg_f\|_\infty. \tag{5.2}$$

On observe que l'approximée μ et l'approximande π jouent des rôles différents. La première étape de majoration de l'erreur -qui donne C_1, C_2 dans la formule ci-dessus- fait intervenir l'approximée μ ainsi que l'approximande π via son opérateur caractéristique, mais pas la classe de fonctions associée à la distance utilisée, tandis que la deuxième étape d'évaluation des facteurs de Stein dépend uniquement de π et de $\mathcal{F}$. Ainsi, une fois que les facteurs de Stein sont évalués pour une certaine mesure approximande et une certaine distance, la méthode peut être appliquée à une variété d'approximées W en modifiant seulement l'étape 1.

Ce point essentiel explique en quoi l'utilisation des facteurs de Stein participe de l'automatisation de la méthode, et montre qu'il est souhaitable de disposer de techniques universelles (c'est-à-dire non spécifiques à l'approximande π) permettant les évaluer. Le chapitre suivant propose des résultats dans ce sens.

Dans ce chapitre de rappels, on commence par présenter le prototype de la méthode de Stein-Chen, l'approximation Poisson-binomiale, puis on passe en revue l'état de l'art relatif à l'étape 1 et à l'étape 2, dans les parties 5.2 et 5.3 respectivement. On conclut la partie 5.3 par une piste de recherche explorée pendant la thèse.

1. Certains auteurs désignent sous le nom de facteurs de Stein les quantités $\sup_{f\in\mathcal{F}} \|\partial^k g_f\|_\infty$ elles-mêmes, et sous le nom d'estimées de régularité (*smoothness estimates*) les bornes supérieures.

5.1.2 L'approximation Poisson-binomiale

Le prototype de la méthode de Stein-Chen est l'approximation Poisson-binomiale, où l'approximande est une loi de Poisson et l'approximée suit une loi binomiale, ou plus généralement est une somme de variables de Bernoulli. Rappelons les grandes lignes de l'historique (détaillé dans l'introduction de Barbour et al. [1992]).

Le point de départ est la loi des petits nombres de Poisson, qui énonce que la loi binomiale $\mathcal{B}(n, \frac{\lambda}{n})$ tend en loi vers la loi de Poisson $\mathcal{P}_\lambda$ (résultat de Poisson en 1837). L'approximation de Poisson consiste à évaluer la distance entre une somme finie de variables de Bernoulli (X_i) de paramètres (p_i), $1 \le i \le n$ et la loi de Poisson $\mathcal{P}_\lambda$ où $\lambda = \sum_{i=1}^n p_i$, ce qui permet entre autres de quantifier la vitesse de convergence dans la loi des petits nombres. Notons d_{TV} la distance en variation totale entre les mesures de probabilité μ et π sur $\mathbb{N}$:

$$d_{TV}(\mu, \pi) := \sup_{A \subset \mathbb{N}} |\mu(A) - \pi(A)| = \sup_{f:\mathbb{N}\to[0,1]} |\mu(f) - \pi(f)| .$$

Le résultat d'approximation le plus basique s'énonce comme suit : si les Bernoulli sont indépendantes, posant $W = \sum_{i=1}^n X_i$,

$$d_{TV}(\mathcal{L}(W), \mathcal{P}_\lambda) \le \sum_{i=1}^n p_i^2. \tag{5.3}$$

En effet, d'après la propriété d'additivité par convolution de la distance en variation totale,

$$d_{TV}(\mathcal{L}(W), \mathcal{P}_\lambda) \le \sum_{i=1}^n d_{TV}(\mathrm{Ber}(p_i), \mathcal{P}_{p_i}),$$

et on vérifie facilement par le calcul que $d_{TV}(\mathrm{Ber}(p), \mathcal{P}_p) \le p^2$. Dans le cas de variables $(X_i)_{1\le i\le n}$ indépendantes, Le Cam (Le Cam [1960]) et Le Cam et Hodges (Hodges and Le Cam [1960]) obtiennent les raffinements suivants de l'inégalité (5.3), qui montrent que les erreurs successives, brutalement additionnées dans (5.3), se compensent en réalité :

$$d_{TV}(\mathcal{L}(W), \mathcal{P}(\lambda)) \le \frac{8}{\lambda} \sum_{i=1}^n p_i^2, \qquad\qquad d_{TV}(\mathcal{L}(W), \mathcal{P}(\lambda)) \le 4.5 \max p_i,$$

la première inégalité étant valide pour des p_i inférieurs à $1/4$.

Voyons maintenant comment la méthode de Stein-Chen permet de retouver les inégalités de Le Cam. On caractérise la loi de Poisson de paramètre λ par l'opérateur

$$Sg(x) = \lambda g(x+1) - xg(x), \qquad\qquad x \in \mathbb{N},$$

où g est une fonction de $\mathbb{N}$ dans $\mathbb{R}$. On note $W^i = \sum_{j\ne i} X_j$. Par indépendance des $(X_i)_{1\le i\le n}$, la première étape de majoration de l'erreur s'écrit :

$$\begin{aligned}\mathbb{E}[Sg(W)] = \mathbb{E}[\lambda g(W+1) - Wg(W)] &= \sum_{i=1}^n (p_i g(W+1) - X_i g(W)) = \sum_{i=1}^n p_i (g(W+1) - g(W^i+1)) \\ &= \sum_{i=1}^n p_i (g(W^i + X_i + 1) - g(W^i+1)) = \sum_{i=1}^n p_i^2 (g(W^i+2) - g(W^i+1)).\end{aligned}$$

Ainsi,

$$|\mathbb{E}[Sg(W)]| \le \|\partial g\|_\infty \sum_{i=1}^n p_i^2. \tag{5.4}$$

L'étape 2 fournit le facteur de Stein (Barbour and Eagleson [1983]) :

$$\sup_{0\le f\le 1} \|\partial g_f\|_\infty \le \frac{1-e^{-\lambda}}{\lambda}.$$

En combinant les deux étapes, on obtient l'approximation en variation totale suivante

$$d_{TV}(\mathcal{L}(W), \mathcal{P}_\lambda) \le \left(1 \wedge \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2. \tag{5.5}$$

Dans Barbour and Hall [1984], Barbour et Hall obtiennent la minoration

$$d_{TV}(\mathcal{L}(W), \mathcal{P}_\lambda) \ge \frac{1}{32}\left(1 \wedge \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2,$$

montrant que l'ordre de grandeur dans (5.5) est optimal, et justifiant le nom de *magique* qui est souvent donné au facteur de Stein. Ainsi, les deux inégalités de Le Cam sont retrouvées (pour la seconde, écrire que $\sum p_i^2 \le (\max p_i) \sum p_i$), avec une constante améliorée et sans condition sur les paramètres.

Mais le principal intérêt de la méthode de Stein-Chen réside dans l'extension au cas de variables non indépendantes. Les techniques de Le Cam reposent sur la convolution de plusieurs mesures et s'adaptent mal au cas non-indépendant. A l'inverse, la méthode de Stein-Chen s'y prête bien : on présente plusieurs stratégies de majoration de l'erreur dans la partie 5.2, retombant toutes sur l'inégalité (5.4) dans le cas indépendant.

5.1.3 L'opérateur de Stein

L'approximande π intervient uniquement via son opérateur caractéristique, qui n'est pas unique, comme on le verra plus bas. Ce choix influençant les deux étapes décrites ci-dessus, il est parfois lui-même considéré comme une étape à part entière.

Les deux approximandes historiques sont la loi gaussienne dans le cas continu et la loi de Poisson $\mathcal{P}_\lambda$ dans le cas discret, qui admettent respectivement comme opérateurs :

$$Sg(x) = g'(x) - xg(x), \qquad x \in \mathbb{R}, \tag{5.6}$$

$$Sg(x) = \lambda g(x+1) - xg(x), \qquad x \in \mathbb{N}. \tag{5.7}$$

La formule (5.6) peut s'interpréter comme une formule d'intégration par parties pour la mesure gaussienne, ou encore comme l'invariance de cette mesure sous l'action du processus de Ornstein-Uhlenbeck. De même, la formule (5.7) traduit l'invariance de la loi de Poisson pour un certain processus de naissance-mort. Ces deux opérateurs peuvent ainsi être vus comme des instances de la *méthode du générateur*.

Outre cette méthode, on peut citer l'approche par densité introduite dans Stein [1986] et qui connaît toujours de nombreux développements, par exemple l'article récent Ley et al. [2014] qui cherche à développer une stratégie universelle, ainsi que la méthode par polynômes orthogonaux de Diaconis and Zabell [1991].

La méthode du générateur a été introduite par Barbour [1990], Götze [1991] dans le cas continu, et par Barbour and Brown [1992] pour des processus ponctuels. Considérons une mesure discrète $(\pi(x))_{x\in\mathbb{N}}$ et supposons que l'on dispose de deux fonctions positives $(\alpha(x))_{x\in\mathbb{N}}\,(\beta(x))_{x\in\mathbb{N}}$ telles que

$$\alpha(x)\pi(x) = \beta(x+1)\pi(x+1), \qquad x \in \mathbb{N}. \tag{5.8}$$

On voit alors que l'opérateur défini pour les fonctions $g : \mathbb{N} \to \mathbb{R}$ par

$$Sg(x) = \alpha(x)g(x+1) - \beta(x)g(x), \qquad x \in \mathbb{N}, \tag{5.9}$$

est un opérateur caractéristique de la mesure π au sens défini précédemment. L'équation (5.8) se traduit par la réversibilité par rapport à la mesure π du processus de naissance-mort de taux de naissance $(\alpha(x))_{x\in\mathbb{N}}$ et taux de mort $(\beta(x))_{x\in\mathbb{N}}$. Ce processus de Markov a pour générateur l'opérateur défini pour les fonctions $g : \mathbb{N} \to \mathbb{R}$ par :

$$Lh(x) = \alpha(x)(h(x+1) - h(x)) + \beta(x)(h(x-1) - h(x)), \qquad x \in \mathbb{N}. \tag{5.10}$$

Le choix de S n'est pas unique, puisque toutes les intensités satisfaisant $\alpha(k)/\beta(k+1) = \pi(k+1)/\pi(k)$ conviennent. On choisit $\beta(0) = 0$ de manière à obtenir effectivement un processus à valeurs dans $\mathbb{N}$. Il convient également de se limiter aux intensités assurant que le processus de naissance-mort correspondant soit ergodique et non-explosif (Chen [2004]) :

$$\sum_{x=1}^{+\infty} \frac{\alpha(0)\alpha(1)\dots\alpha(x-1)}{\beta(1)\beta(2)\dots\beta(x)} < \infty, \quad \sum_{x=1}^{\infty} \left(\frac{1}{\alpha(x)} + \frac{\beta(x)}{\alpha(x)\alpha(x-1)} + \dots + \frac{\beta(x)\dots\beta(1)}{\alpha(x)\dots\alpha(0)} \right) = \infty.$$

Par exemple, la loi binomiale négative $\pi = \mathrm{BN}(r,p)$ de paramètres (r,p)

$$\pi(x) = \frac{\Gamma(r+x)}{\Gamma(r)x!}(1-p)^r p^x, \qquad x \in \mathbb{N},$$

est invariante pour un processus de Galton-Watson avec immigration d'opérateur de Stein défini, pour toute fonction $g : \mathbb{N} \to \mathbb{R}$, par

$$Sg(x) = p(r+x)g(x+1) - xg(x), \qquad x \in \mathbb{N}. \tag{5.11}$$

Pour la distribution géométrique $\mathcal{G}(\rho)$, telle que $p(x) = (1-\rho)\rho^x$ sur $\mathbb{N}$, on peut choisir (Peköz [1996],Eichelsbacher and Reinert [2008])

$$(Sg)(x) = \rho g(x+1) - \mathbf{1}_{x\in\mathbb{N}^*} g(x), \qquad x \in \mathbb{N}, \tag{5.12}$$

ce qui correspond à un processus de naissance-mort de taux de naissance et mort constants, connu sous le nom de processus $M/M/1$ dans la théorie des files d'attente (correspond au nombre d'individus en attente dans une file avec arrivées et départs exponentiels et un seul guichet). La distribution géométrique étant une la loi binomiale négative $\mathrm{BN}(1,\rho)$, un autre choix est (5.11) avec les coefficients correspondants (Phillips and Weinberg [2000]).

L'opérateur utilisé pour la loi de Poisson introduit dans (5.7) correspond à une file $M/M/\infty$ (arrivées et départs exponentiels, infinité de guichets).

Comme dans les cas (5.7) et (5.11), le taux de mort est souvent choisi proportionnel à la taille de la population, $\beta = \mathrm{id}$, ce qui en retour détermine le taux de naissance pour une mesure de probabilité cible fixée. C'est un choix naturel si l'on inteprète le processus de naissance-mort comme analogue continu d'un processus de branchement. En fonction du taux de naissance, il permet également d'obtenir des couplages entre processus de même dynamique différant au départ d'un individu. Par exemple, notons $(X_t^x)_{t\geq 0}$ la file $M/M/\infty$ de taille initiale $x \in \mathbb{N}$. En isolant un individu auquel on interdit de se reproduire pour garder le taux de naissance constant, on voit facilement que $X_t^{x+1} = X_t^x + Y_t$, où $Y_t \sim \mathrm{Ber}(e^{-t})$. Itérée, cette relation aboutit à la formule de Mehler (5.28) rappelée plus bas. Un raisonnement similaire montre que si $(X_t^x)_{t\geq 0}$ désigne maintenant le processus de Galton-Watson ave immigration, on a cette fois $X_t^{x+1} = X_t^x + Z_t^1$, où $(Z_t^1)_{t\geq 0}$ correspond à un processus de Galton-Watson sans immigration (paramètre $r = 0$) ayant initialement un individu. Pour des raisons de coût numérique, l'article Brown and Xia [2001] argumente de son côté pour le choix d'un taux de naissance constant et un taux de mort polynomial de degré deux dont les paramètres sont réglés de telle sorte que l'équation d'équilibre (5.8) soit vérifiée.

Pour toute fonction $h : \mathbb{N} \to \mathbb{R}$, on voit que $Lh(x) = S\partial h(x-1)$ sur $\mathbb{N}$. On peut donc penser à l'opérateur de Stein S comme au demi-générateur associé au processus invariant. Bien sûr, cela n'est possible qu'en dimension 1, où l'opérateur de gradient ne change pas l'espace de fonctions ; en dimension supérieure, l'opérateur caractéristique est le générateur lui-même.

5.2 Étape 1 : stratégies pour la majorer l'erreur

5.2.1 Principe

Etant donnés l'approximande π et l'approximée W, rappelons qu'il s'agit de majorer la quantité $\mathbb{E}[Sg(W)]$, où pour toute $g : \mathbb{N} \to \mathbb{R}$ dans une classe de fonctions-tests, $\int (Sg) d\pi = 0$.

La première idée consiste à introduire un opérateur T caractéristique de la loi de W et écrire

$$\mathbb{E}[Sg(W)] = \mathbb{E}[(S-T)g(W)], \tag{5.13}$$

ce qui ramène à la comparaison entre les opérateurs caractéristiques. Cette stratégie est employée avec profit dans le cas où W est un mélange de lois (voir Barbour et al. [1992] pour le mélange de loi de Poisson et Cloez and Delplancke [2016] (chapitre suivant) pour le cas quelconque). On voit également l'intérêt de choisir par défaut un taux de mort égal à l'identité, puisque la différence entre les deux opérateurs se lit alors uniquement sur les taux de naissance. Dans le cas continu multidimensionnel, une idée similaire est utilisée pour définir la notion de discrépance et obtenir des théorèmes centraux limite quantitatifs (Courtade et al. [2017]).

Une autre technique consiste à appliquer une transformation sur les mesures de probabilité en jeu, et à coupler les images obtenues. Comme on le voit ci-dessous, la transformation adéquate lorsque l'approximande est la loi de Poisson ou la loi binomiale négative est la transformation *size-bias*, tandis qu'on choisit pour la loi géométrique une transformation associée à la théorie du renouvellement. Des techniques analogues existent dans le cas continu : lorsque l'approximande est la loi gaussienne, on utilise par exemple la transformation du *zero-bias*.

En fait, cette méthode et les deux suivantes, la méthode locale et celle des paires échangeables, permettent de contourner le fait que la distribution de W n'est pas explicite, par exemple si W représente la somme de variables de Bernoulli dépendantes, et, partant, que l'opérateur T utilisé dans (5.13) est inconnu.

5.2.2 Méthode par couplage

Transformation *size-bias*

À une variable aléatoire Z à valeurs dans $\mathbb{N}$ d'espérance finie, la transformation du *size bias* fait correspondre une variable aléatoire Z^s caractérisée par la relation

$$\mathbb{E}[Zg(Z)] = \mathbb{E}[Z]\,\mathbb{E}[g(Z^s)], \qquad g : \mathbb{N} \to \mathbb{R}, \quad \mathbb{E}[Zg(Z)] < \infty.$$

Elle est utilisée de manière implicite dès les débuts du développement de la méthode de Stein-Chen, par exemple dans le livre de référence Barbour et al. [1992] qui consacre plusieurs chapitres à la méthode de couplage. Son utilisation explicite dans le cadre de la méthode de Stein apparaît dans Goldstein and Rinott [1996] et permet de mettre en évidence la logique sous-jacente à certaines manipulations calculatoires.

Transformées de lois usuelles

Dans le cas de la loi de Poisson, l'identité $\mathbb{E}[\lambda g(Z+1) - Zg(Z)] = 0$ où $Z \sim \mathcal{P}_\lambda$ indique que la transformée de Z suit la loi $\mathcal{L}(Z^s) = \mathcal{L}(Z+1)$.

Pour la loi binomiale négative $\mathrm{BN}(r,p)$ qui a pour espérance $rp/(1-p)$, on voit en utilisant l'opérateur caractéristique (5.11) que si $Z \sim \mathrm{BN}(r,p)$,

$$\mathbb{E}[Zg(Z)] = pr\mathbb{E}[g(Z+1)] + p\mathbb{E}[Zg(Z+1)].$$

Par une récurrence immédiate,

$$\mathbb{E}[Zg(Z)] = r\sum_{k\in\mathbb{N}^*} p^k\mathbb{E}[g(Z+k)] = \mathbb{E}[Z]\sum_{k\in\mathbb{N}}(1-p)p^k\mathbb{E}[g(Z+k+1)].$$

On trouve donc que $\mathcal{L}(Z^s) = \mathcal{L}(Z+Y+1)$, où Y suit la loi géométrique $\mathcal{G}(p)$.

Dans le cas où $W = \sum_{i\in I} X_i$ avec les X_i des Bernoulli de paramètre p_i, on peut construire W^s de la manière suivante : on tire indépendamment un indice J dans I avec distribution $\mathbb{P}(J=j) = p_j/\sum_{i\in I} p_i$ et des variables aléatoires $(W_j^*)_{j\in I}$, de lois respectives $\mathcal{L}(W|X_j=1)$. Alors W_J^* est la transformée par la transformation du *size-bias* de W. De plus, pour toute fonction $g:\mathbb{N}\to\mathbb{R}$ assez régulière,

$$\mathbb{E}[g(W_J^*)] = \frac{1}{\mathbb{E}[W]}\sum_{i\in I}\mathbb{E}[X_i]\,\mathbb{E}[g(W_i^*)]. \tag{5.14}$$

Approximation Poissonnienne

Pour une approximée W et l'approximande $\mathcal{P}_\lambda$ de paramètre $\lambda = \mathbb{E}[W]$, écrivons :

$$\mathbb{E}[Sg(W)] \;=\; \lambda\mathbb{E}[g(W+1)] - \mathbb{E}[W]\,\mathbb{E}[g(W^s)] \;=\; \lambda\left(\mathbb{E}[g(W+1)-g(W^s)]\right).$$

Ainsi, pour tout couplage entre $\mathcal{L}(W+1), \mathcal{L}(W^s)$,

$$|\mathbb{E}[Sg(W)]| \le 2\lambda\|g\|_\infty\mathbb{P}(W+1\neq W^s), \quad |\mathbb{E}[Sg(W)]| \le \lambda\|\partial g\|_\infty\mathbb{E}[|W+1-W^s|]. \tag{5.15}$$

(De manière plus formelle, ceci montre en fait que

$$|\mathbb{E}[Sg(W)]| \le 2\lambda\|g\|_\infty d_{TV}(\mathcal{L}(W+1),\mathcal{L}(W^s)), \quad |\mathbb{E}[Sg(W)]| \le \lambda\|\partial g\|_\infty d_W(\mathcal{L}(W+1),\mathcal{L}(W^s)).$$

Notons que comme $\|\partial g\|_\infty \le 2\|g\|_\infty$ et $d_{TV}(\mathcal{L}(W+1),\mathcal{L}(W^s)) \le d_W(\mathcal{L}(W+1),\mathcal{L}(W^s))$, les deux bornes ne sont pas comparables a priori.)

Dans le cas où $W = \sum_{i\in I} X_i$ avec les X_i des Bernoulli de paramètre p_i, l'équation (5.15) devient

$$|\mathbb{E}[Sg(W)]| \le \lambda\|\partial g\|_\infty\sum_{i\in I} p_i\mathbb{E}[|W+1-W_i^*|]. \tag{5.16}$$

Pour des Bernoulli indépendantes, on tire indépendamment $W_i = \sum_{j\neq i} X_j$ et X_i, et on pose $W_i^* = W_i + 1$ et $W = W_i + X_i$. On retrouve bien la majoration (5.4) :

$$|\mathbb{E}[Sg(W)]| \le \|\partial g\|_\infty\sum_{i\in I} p_i^2.$$

Généralement, l'équation (5.16) s'exploite bien quand la structure de dépendance des X_j, $j\neq i$ par rapport à X_i est identique pour tous les j. Le livre de référence Barbour et al. [1992] consacre plusieurs chapitres à cette situation. Un exemple classique est celui de l'urne de Pólya, où il s'agit de compter le nombres de boules d'une couleur donnée obtenues par tirage sans remise dans une urne contenant des boules de plusieurs couleurs. Un autre exemple concerne le comptage de sous-graphes dans un graphe d'Erdös-Rényi (Ross [2011]).

Approximation binomiale négative

Dans certaines situations, il peut être plus intéressant d'utiliser comme approximande la loi géométrique ou sa généralisation la loi binomiale négative $\pi = \mathrm{BN}(r,p)$, comme le montre l'étude de Brown and Phillips [1999] pour le modèle de l'urne de Pólya dans certains régimes. Pour comparer l'approximée W à la loi binomiale négative $\pi = \mathrm{BN}(r,p)$ de même espérance $rp/(1-p)$, on cherche à faire intervenir W^s et $W+Y+1$. Le calcul, moins direct que pour la loi de Poisson, s'écrit :

$$\begin{aligned}
\mathbb{E}\left[Sg(W)\right] &= pr\mathbb{E}\left[g(W+1)\right] + p\mathbb{E}\left[Wg(W+1)\right] - \mathbb{E}\left[Wg(W)\right] \\
&= (1-p)\mathbb{E}\left[W\right]\mathbb{E}\left[g(W+1)\right] + p\mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s+1)\right] - \mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s)\right] \\
&= \mathbb{E}\left[W\right]\mathbb{E}\left[g(W+1+Y)\mathbf{1}_{Y=0}\right] + p\mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s+1)\right] - \mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s)\right] \\
&= \mathbb{E}\left[W\right]\mathbb{E}\left[g(W+1+Y)\right] - \mathbb{E}\left[W\right]\mathbb{E}\left[g(W+1+Y)\mathbf{1}_{Y>0}\right] \\
&\quad + p\mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s+1)\right] - \mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s)\right].
\end{aligned}$$

Or, comme Y suit une loi géométrique, $\mathcal{L}(Y|Y>0) = \mathcal{L}(Y+1)$. Par conséquent,

$$\begin{aligned}
\mathbb{E}\left[Sg(W)\right] &= \mathbb{E}\left[W\right]\mathbb{E}\left[g(W+1+Y)\right] - p\mathbb{E}\left[W\right]\mathbb{E}\left[g(W+2+Y)\right] \\
&\quad + p\mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s+1)\right] - \mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s)\right] \\
&= \mathbb{E}\left[W\right]\mathbb{E}\left[g(W+1+Y) - g(W^s)\right] + p\mathbb{E}\left[W\right]\mathbb{E}\left[g(W^s+1) - g(W+2+Y)\right].
\end{aligned}$$

De ceci, on peut déduire la majoration utile

$$|\mathbb{E}\left[Sg(W)\right]| \leq \|\partial g\|_\infty \mathbb{E}\left[W\right](p+1)\mathbb{E}\left[|W+1+Y-W^s|\right], \tag{5.17}$$

qui est prouvée dans Brown and Phillips [1999] par des calculs identiques mais sans mention de leur lien avec la transformation du *size bias*.

Dans le cas où $W = \sum_{i\in I} X_i$ est une somme de variables aléatoires de Bernoulli, grâce à la formule (5.14), la majoration (5.17) devient ([Brown and Phillips, 1999, Théorème 1, équation (5)])

$$|\mathbb{E}\left[Sg(W)\right]| \leq \|\partial g\|_\infty (p+1)\sum_{i\in I}\mathbb{E}\left[X_i\right]\mathbb{E}\left[|W+1+Y-W_i^*|\right] \tag{5.18}$$

où $\mathcal{L}(W_i^*) = \mathcal{L}(W|X_i=1)$. (Dans Brown and Phillips [1999], les notations sont différentes : p est remplacé par $1-p$ et le symbole X_i^* désigne une variable aléatoire de loi $\mathcal{L}(W-1|X_i=1)$.)

Théorie du renouvellement

La loi géométrique correspondant à une binomiale négative avec $r=1$, on peut reprendre les résultats du paragraphe précédent si l'on choisit un opérateur de Stein avec taux de mort égal à l'identité. Si l'on fait plutôt le choix de taux de naissance et mort constants (équation (5.12)), la transformation utilisée est associée à la théorie du renouvellement dans le cadre discret, comme expliqué dans (Peköz et al. [2013]). Elle envoie la loi de W sur la loi d'une variable aléatoire W' caractérisée par l'équation

$$\mathbb{E}\left[g(W) - g(0)\right] = \mathbb{E}\left[W\right]\mathbb{E}\left[\partial g(W')\right], \tag{5.19}$$

pour toute fonction $g : \mathbb{N} \to \mathbb{R}$ assez régulière. La distribution de W' est appelée distribution d'équilibre associée à la loi de W.

La loi géométrique est caractérisée par l'égalité $\mathcal{L}(W') = \mathcal{L}(W)$. Ainsi, pour majorer l'erreur faite en approchant l'approximée W par loi $\mathcal{G}(\rho)$ de même espérance, c'est-à-dire $\mathbb{E}\left[W\right] = \rho/(1-\rho)$, on essaie de faire apparaître un couplage entre W et W'. Comme on le voit dans la partie 5.3, l'équation de Stein (5.20) admet une solution telle que $g(0)=0$. Supposons donc que $g(0)=0$ dans le reste de ce paragraphe. L'équation (5.19) devient

$$\rho\mathbb{E}\left[W\right] = (1-\rho)\mathbb{E}\left[(\partial g)(W')\right].$$

Il vient alors

$$\mathbb{E}\left[(Sg)(W)\right] = \mathbb{E}\left[\rho g(W+1) - g(W)\right] = \mathbb{E}\left[\rho\partial g(W) - (1-\rho)g(W)\right] = \rho\mathbb{E}\left[(\partial g)(W) - (\partial g)(W')\right],$$

et donc, pour tout couplage entre $\mathcal{L}(W)$ et $\mathcal{L}(W')$,

$$|\mathbb{E}\left[(Sg)(W)\right]| \leq 2\rho\|\partial g\|_\infty \mathbb{P}(W \neq W').$$

Ce résultat est obtenu dans Peköz et al. [2013] et appliqué à quatre exemples d'approximandes. (Le lecteur intéresssé prendra garde au changement de notation : la distribution $\mathcal{G}(\rho)$ y est notée $\mathrm{Ge}^0(p)$ avec $p = 1-\rho$.)

5.2.3 Méthode locale

Comme on vient de le voir, la méthode par couplage s'adapte bien au cas où l'approximande est la loi de Poisson et l'approximée est de la forme $W = \sum_{i\in I} X_i$ où la structure de dépendance des $(X_i)_{i\in I}$ est assez symétrique. La méthode locale, quant à elle, est conçue pour le cas où la dépendance entre les variables est locale, c'est-à-dire que pour chaque $i \in I$, on sépare les variables $(X_j)_{j\neq i}$ en deux groupes : celles qui dépendent "faiblement" de X_i et celles qui en dépendent "fortement". C'est le cadre de l'article fondateur de Chen (Chen [1975]), qui est repris dans Arratia et al. [1989], Arratia et al. [1990]. La présentation ci-dessous est inspirée de l'introduction de Barbour et al. [1992].

Soit Γ_i^s l'ensemble des indices j pour lesquels X_j dépend "beaucoup" de X_i et Γ_i^w l'ensemble des indices j pour lesquels X_j dépend "peu "de X_i, avec $\Gamma_i^s \cup \Gamma_i^w = \{j \in I, j \neq i\}$. Par exemple, si les variables aléatoires X_i sont indépendantes on pose $\Gamma_i^s = \emptyset$ et $\Gamma_i^w = \{j \neq i,\ 1 \leq j \leq n\}$. On sépare la variable $W_i = W - X_i$ en la somme de deux variables $W_i = W_i^s + W_i^w$, avec

$$W_i^s = \sum_{j\in\Gamma_i^s} X_j, \qquad W_i^w = \sum_{j\in\Gamma_i^w} X_j.$$

On choisit le paramètre de l'approximande $\mathcal{P}_\lambda$ de telle sorte qu'approximande et approximée aient même espérance, c'est-à-dire $\lambda = \sum_{i\in I} p_i$. L'erreur s'écrit

$$\begin{aligned}
\mathbb{E}\left[Sf(W)\right] &= \mathbb{E}\left[\lambda f(W+1) - Wf(W)\right] = \sum_{i\in I} p_i\mathbb{E}\left[f(W+1)\right] - \mathbb{E}\left[X_i f(W)\right] \\
&= \sum_{i\in I} p_i\mathbb{E}\left[f(W+1)\right] - \mathbb{E}\left[X_i f(W_i+1)\right] \\
&= \sum_{i\in I} p_i\mathbb{E}\left[f(W+1)\right] - \mathbb{E}\left[X_i f(W_i^w+1)\right] - \sum_{i\in I}\mathbb{E}\left[X_i(f(W_i+1) - f(W_i^w+1))\right] \\
&= \sum_{i\in I} p_i\mathbb{E}\left[f(W+1) - f(W_i^w+1)\right] + \sum_{i\in I}\mathbb{E}\left[(p_i - X_i)f(W_i^w+1)\right] \\
&- \sum_{i\in I}\mathbb{E}\left[X_i(f(W_i+1) - f(W_i^w+1))\right].
\end{aligned}$$

On pose

$$\begin{aligned}
A_i &= p_i\mathbb{E}\left[|W - W_i^w|\right] = p_i(p_i + \mathbb{E}\left[W_i^s\right]) \\
B_i &= \mathbb{E}\left[X_i|W_i - W_i^w|\right] = p_i\mathbb{E}\left[W_i^s|X_i = 1\right] \\
C_i &= \mathbb{E}\left[|\mathbb{E}\left[p_i - X_i|W_i^w\right]|\right].
\end{aligned}$$

Alors,

$$|\mathbb{E}\left[Sf(W)\right]| \leq \|\partial f\|_\infty\left(\sum_{i\in I} A_i + \sum_{i\in I} B_i\right) + \|f\|_\infty\sum_{i\in I} C_i.$$

Le terme $A = \sum_{i\in I} A_i$ reflète la taille des voisinages de forte dépendance Γ_i^s ; le terme $B = \sum_{i\in I} B_i$ mesure le nombre moyen d'occurences de 1 parmi les voisins de i sachant que $X_i = 1$, autrement dit mesure la tendance aux séries de 1 ; enfin, le terme $C = \sum_{i\in I} C_i$, nul si chaque variable X_i est indépendante des variables d'indice dans Γ_i^w, pénalise le choix d'un voisinage Γ_i^s trop petit. On peut préciser ces idées en introduisant des conditions de mélange sur les X_i et exprimer A, B et C en fonction de coefficients de mélange (Smith [1988]).

Si les variables aléatoires $(X_i)_{i\in I}$ sont indépendantes, on voit que $A = \sum_{i\in I} p_i^2$ et $B = C = 0$, et on retrouve la majoration (5.4).

Si, pour tout indice de I, la variable X_i et les variables d'indice appartenant à Γ_i^w sont indépendantes, la majoration de l'erreur se simplifie en

$$|\mathbb{E}\left[Sf(W)\right]| \leq \|\partial f\|_\infty \left(\sum_{i\in I} A_i + \sum_{i\in I} B_i\right).$$

On parle alors de voisinage (ou graphe) de dépendance. Un exemple typique apparaît si l'on s'intéresse à la longueur d'une série de "piles" lors d'un lancer de pièces. Cette variable, ainsi que de nombreux raffinements, apparaît lors de l'étude de brins d'ADN (cf Ross [2011] pour une brève introduction et les références associées). Les graphes de dépendance s'avèrent également utiles dans l'étude des tessellations (Chenavier [2014]).

5.2.4 Méthode des paires échangeables

La méthode des paires échangeables, introduite dans le cas continu par Stein (Stein [1972]), revient à construire implicitement un opérateur caractéristique de l'approximée W en se donnant une *paire échangeable.*

Supposons qu'on dispose d'une paire échangeable (W, W'), c'est-à-dire d'un couple de variables aléatoires tel que $\mathcal{L}(W, W') = \mathcal{L}(W', W)$ et d'un opérateur U qui envoie une fonction de $\mathbb{N}$ dans $\mathbb{R}$ sur une fonction antisymétrique de $\mathbb{N}^2$ dans $\mathbb{R}$: on a alors $\mathbb{E}\left[Ug(W, W')\right] = 0$. On voit ainsi que l'opérateur T défini pour toute fonction assez régulière $g : \mathbb{N} \to \mathbb{R}$ par la relation

$$Tg(w) = \mathbb{E}\left[Ug(W, W')|W = w\right], \qquad w \in \mathbb{N},$$

est un opérateur caractéristique de la loi de W, ce qui invite à écrire l'erreur sous la forme

$$\mathbb{E}\left[Sg(W)\right] = \mathbb{E}\left[Sg(W) - \mathbb{E}\left[Ug(W, W')|W\right]\right].$$

L'idée est de construire la paire (W, W') et l'opérateur U de telle sorte que la variable aléatoire $Tg(W, W')|W$ "ressemble" à la variable aléatoire $Sg(W)$.

L'article Holmes [2004] est consacré au développement de cette méthode pour une approximande discrète quelconque, et Chatterjee et al. [2005] passe en revue des exemples où l'approximande est la loi de Poisson. Ici, on se limitera à voir comment obtenir avec cette technique notre majoration usuelle (5.4). Si $W = \sum_{i\in I} X_i$ est la somme de Bernoulli indépendantes, on obtient W' en sélectionnant uniformément dans I un indice et en remplaçant la variable X_i correspondante par un nouveau tirage. L'opérateur T considéré est

$$Tg(x, x') := |I| f(x')\mathbf{1}_{x'=x+1} - |I| f(x)\mathbf{1}_{x=x'+1}, \qquad x, x' \in \mathbb{N}.$$

On trouve alors

$$\begin{aligned}
\mathbb{E}\left[\alpha g(W, W')|(X_i)_{i\in I}\right] &= |I|\mathbb{E}\left[g(W+1)\mathbf{1}_{W'=W+1} - g(W)\mathbf{1}_{W'=W-1}|(X_i)_{i\in I}\right] \\
&= \sum_{i\in I} \mathbb{E}\left[g(W+1)\mathbf{1}_{X_i=0, X_i'=1} - g(W)\mathbf{1}_{X_i=1, X_i'=0}|(X_i)_{i\in I}\right] \\
&= \sum_{i\in I} p_i g(W+1)(1 - X_i) - g(W)X_i(1 - p_i).
\end{aligned}$$

Par ailleurs,

$$Sg(W) = \lambda g(W+1) - Wg(W) = \sum_{i \in I} p_i g(W+1) - X_i g(W)$$

donc on retrouve l'inégalité (5.4) en écrivant

$$\mathbb{E}\left[Sg(W)\right] = \mathbb{E}\left[Sg(W) - \mathbb{E}\left[\alpha g(W, W')|(X_i)_{i \in I}\right]\right] = \sum_{i \in I} \mathbb{E}\left[p_i X_i (g(W+1) - g(W))\right].$$

5.2.5 Méthode de Stein-Malliavin

Mentionnons pour finir la méthode de Stein-Malliavin. Dans l'article fondateur Nourdin and Peccati [2009], les auteurs combinent la méthode de Stein et le calcul de Malliavin sur un espace gaussien pour obtenir des résultats de convergence de variables d'un espace gaussien vers la loi normale. Ces résultats ont ouvert un champ de recherche très actif, citons par exemple le livre de référence Nourdin and Peccati [2012].

Mettant à profit une version du calcul de Malliavin sur l'espace de Poisson, Peccati et al. [2010] s'intéresse à la convergence de fonctionnelles de mesures de Poisson vers la loi normale centrée réduite. Ce travail est poursuivi dans Peccati [2011], qui traite de l'approximation de fonctionnelles de mesures de Poisson vers une loi de Poisson de même espérance. Une des applications principales de ces travaux concerne l'étude des graphes aléatoires.

Reproduisons le résultat principal de Peccati [2011], qui correspond à l'étape 1 de majoration de l'erreur. On renvoie à Nualart and Vives [1990] et Peccati [2011] pour la définition des objets utilisés. Soit $(Z, \mathcal{Z}, \mu)$ un espace borélien, où $\mathcal{Z}$ est la tribu borélienne associée et μ une mesure borélienne σ-finie et non-atomique. Soit ν une mesure de Poisson contrôlée par μ, et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, où la tribu $\mathcal{F}$ est engendrée par ν. On note D l'opérateur de dérivation de Malliavin et L^{-1} le pseudo-inverse du générateur d'Ornstein-Uhlenbeck sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Soit W une variable aléatoire de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{N}$ et dans le domaine de l'opérateur de dérivation D, telle que $\mathbb{E}[W] = \lambda$, et soit $f : \mathbb{N} \to \mathbb{R}$ une fonction assez régulière (par exemple bornée). Alors, en notant g_f la solution de l'équation de Stein associée à f, la majoration suivante est vérifiée :

$$\begin{aligned}|\mathbb{E}[f(W)] - \mathcal{P}_\lambda(f)| \leq\ & \mathbb{E}\left[\left|\lambda - \langle DW, -DL^{-1}W\rangle_{L^2(\mu)}\right|\right] \|\partial g_f\|_\infty \\ & + \mathbb{E}\left[\int_Z \left|D_z W(D_z W - 1) D_z L^{-1} W\right| \mu(dz)\right] \|\partial^2 g_f\|_\infty.\end{aligned}$$

Conjointement avec les résultats existants sur les facteurs de Stein, cette majoration donne le résultat principal de Peccati [2011], le théorème 3.1., concernant l'approximation en distance en variation totale entre la loi de W et la mesure $\mathcal{P}_\lambda$.

5.3 Étape 2 : facteurs de Stein-Chen

5.3.1 Littérature

Rappelons que pour l'approximande π d'opérateur de Stein S défini par (5.9), et la donnée $f : \mathbb{N} \to \mathbb{R}$, l'équation de Stein s'écrit

$$S(g_f) = f - \int f d\pi. \tag{5.20}$$

Délivrons-nous dès à présent d'un point technique. La valeur de $g_f(0)$ n'intervenant pas dans (5.20), vu $\beta(0) = 0$, on suppose sauf indication contraire que $\partial^n g_f(0) = 0$ lorsqu'on évalue le facteur de Stein d'ordre n (il faut cependant être prudent si la valeur de $g(0)$ a déjà été fixée pour d'autres raisons, comme dans l'approximation géométrique par la méthode de couplage présentée plus haut). Ainsi, toutes les normes $\|\cdot\|_\infty$ peuvent en réalité être considérées comme des supremum sur $\mathbb{N}^*$.

Sans prétendre du tout à l'exhaustivité -ce qui serait presque impossible vu l'abondance de la littérature relative à la méthode de Stein-Chen- donnons quelques travaux relatifs aux facteurs de Stein qui nous semblent représentatifs, en insistant sur les tentatives faites pour parvenir à une méthode générale. Les références données dans le chapitre suivant viendront compléter ce tableau.

Dans l'article fondateur Chen [1975], Chen se place dans le cas où l'approximande est $\pi = \mathcal{P}_\lambda$ et l'opérateur de Stein défini par (5.7). Il montre que pour toute fonction f bornée, la solution de l'équation de Stein $Sg_f = f - \int f\mathcal{P}_\lambda$ s'écrit, pour tout entier $x \in \mathbb{N}^*$,

$$\begin{aligned} g_f(x) &= (x-1)!\lambda^x \sum_{k=x}^{\infty}(h(k) - \mathcal{P}_\lambda(k))\frac{\lambda^k}{k!} \\ &= \frac{1}{x\mathcal{P}_\lambda(x)}\left(\mathcal{P}_\lambda(f\mathbf{1}_{[0,x-1]}) - \mathcal{P}_\lambda([0,x-1])\mathcal{P}_\lambda(f)\right). \end{aligned} \tag{5.21}$$

De cette formulation, Chen déduit des facteurs de Stein

$$\sup_{0\le f\le 1}\|g_f\|_\infty \le c\left(1 \wedge \frac{1}{\sqrt{\lambda}}\right), \qquad \sup_{0\le f\le 1}\|\partial g_f\|_\infty \le c'\left(1 \wedge \frac{1}{\sqrt{\lambda}}\right),$$

avec les constantes $c = 4$ et $c' = 6$, qui sont améliorées dans Barbour and Eagleson [1983] en $c = 1.4$ et $c' = 1$. Des fomules analogues pour d'autres approximandes sont proposées, par exemple dans [Barbour et al., 1992, Chapitre 9] pour la loi binomiale et dans Peköz [1996] pour la loi géométrique.

L'article Barbour and Brown [1992] introduit une représentation intégrale de la solution de Stein dans le cas d'une loi de Poisson, que l'on présente ici dans le cas général. Introduisons pour le comfort des notations le gradient discret à gauche défini pour $h : \mathbb{N} \to \mathbb{R}$ et $x \in \mathbb{N}^*$ par $\partial^* h(x) = h(x-1) - h(x)$. Comme on l'a déjà remarqué, l'opérateur caractéristique correspond au "demi-générateur" (5.10) : $Lh = S(-\partial^* g)$ (la convention utilisée pour $\partial^* h(0)$ est indifférente par la remarque technique ci-dessus). Or, pour une fonction $f : \mathbb{N} \to \mathbb{R}$ assez régulière, on connaît la solution de l'équation de Poisson :

$$Sh_f = f - \int f d\pi\,, \qquad h_f = -\int_0^\infty P_t(f - \pi(f))d\pi,$$

où $(P_t)_{t\ge 0}$ est le semi-groupe de générateur L. Ainsi, une solution de l'équation de Stein est, si l'on peut intervertir l'opérateur d'intégration et le gradient

$$g_f = \int_0^\infty -\partial^* P_t(f)d\pi. \tag{5.22}$$

Une avancée cruciale est réalisée par Brown et Xia dans Brown and Xia [2001], où la formule (5.21) est généralisée à toute approximande de support $\mathbb{N}$. En partant de l'équation (5.22) et en utilisant des arguments probabilistes sur les temps de retour du procesus de naissance-mort en un point, ils montrent que pour tout $i, j \in \mathbb{N}^*$,

$$g_j(i) = -\frac{\pi(j)}{\pi(i)\beta(i)}\pi([0,i-1])\mathbf{1}_{i\le j} + \frac{\pi(j)}{\pi(i)\beta(i)}\pi([i,\infty))\mathbf{1}_{i>j}. \tag{5.23}$$

Ils établissent alors le résultat suivant, qui donne une borne universelle sur le facteur de Stein d'ordre 1 relatif à la distance en variation totale :

$$V(x) = \alpha(x) - \alpha(x+1) + \beta(x+1) - \beta(x), \qquad x \in \mathbb{N} \tag{5.24}$$

est positive, alors

$$\sup_{0\le f\le 1} \|\partial g_f\|_\infty \le \sup_{i\in\mathbb{N}^*} \frac{1}{\alpha(i)} \wedge \frac{1}{\beta(i)}.$$

Dans une optique comparable, les articles Eichelsbacher and Reinert [2008], Daly [2008], et Döbler et al. [2015] pour le cas continu, s'attachent à une étude valable dans un cadre général.

Enfin, remarquons que les représentations (5.23) et (5.22) gagnent à être utilisées *conjointement*, la forme fermée permettant trouver une fonction réalisant le supremum du facteur de Stein, qui est ensuite évalué grâce à la forme intégrale, comme dans Barbour and Xia [2006], Barbour et al. [2015], Cloez and Delplancke [2016] (chapitre suivant).

5.3.2 Une piste de recherche explorée pendant la thèse

L'observation qui motive cette étude est que la quantité (5.24) introduite dans Brown and Xia [2001] correspond au potentiel obtenu par Chafaï et Joulin dans leur étude des entrelacements pour les processus de naissance-mort (Chafaï and Joulin [2013]). Il semble dès lors naturel de chercher à faire le lien. Dans le cadre de l'approximation poissonienne, c'est l'objet du présent paragraphe.

Notons donc $\pi = \mathcal{P}_\lambda$, $(P_t)_{t\ge 0}$ le semi-groupe de naissance-mort de taux $(\lambda, x)_{x\in\mathbb{N}}$ invariant pour la mesure $\mathcal{P}_\lambda$, et $(X_t^x)_{t\ge 0}$ le processus de naissance-mort associé tel que $X_0^x = x$. La formule d'entrelacement de Chafaï and Joulin [2013] s'écrit ici très simplement

$$\partial P_t f = e^{-t} P_t \partial f, \qquad t \ge 0, \tag{5.25}$$

où l'on suppose dans un premier temps la fonction $f : \mathbb{N} \to \mathbb{R}$ bornée. Le potentiel présent dans l'exponentiel est ici constant égal à 1, tout comme la quantité (5.24) de Brown et Xia. Avec la notation $\overrightarrow{g}(x) = g(x+1)$, la formule (5.22) se réécrit

$$\overrightarrow{g_f} = -\int_0^\infty \partial P_t f dt.$$

Ainsi, en itérant l'entrelacement (5.25), il vient

$$\|\partial^n \overrightarrow{g_f}\|_\infty \le \int_0^\infty e^{-(n+1)t} \|P_t(\partial^{n+1} f)\|_\infty dt. \tag{5.26}$$

Il faut maintenant majorer $\|P_t(\partial^{n+1} f)\|_\infty$. D'une part, pour tout entier m,

$$\|P_t(\partial^m) f\|_\infty \le \|\partial^m f\|_\infty \le 2^m \|f\|_\infty, \qquad t \ge 0. \tag{5.27}$$

D'autre part, rappelons la formule de Mehler qui s'écrit, pour tout $x \in \mathbb{N}$,

$$X_t^x \sim B_t^x + Z_t, , \qquad t \ge 0, \tag{5.28}$$

où B_t suit la loi binomiale de paramètre x et e^{-t}, et Z_t désigne une variable aléatoire indépendante de loi de Poisson de paramètre $\lambda_t = \lambda(1 - e^{-t})$. Soit $\widetilde{f_x}$ la fonction définie sur $\mathbb{N}$ par l'équation

$$\widetilde{f_x}(Z_t) = \mathbb{E}\left[f(B_t^x + Z_t) | Z_t\right], \qquad t \ge 0.$$

Comme les variables aléatoires B_t^x et Z_t sont indépendantes pour tout $t \ge 0$, la fonction $\widetilde{f_x}$ s'écrit, pour tout z entier, $\widetilde{f_x}(z) = \mathbb{E}\left[f(B_t^x + z)\right]$, et en particulier $\widetilde{(\partial f)_x} = \partial \widetilde{f_x}$. Ainsi,

$$\partial^m P_t f(x) = \mathbb{E}\left[\partial^m f(X_t^x)\right] = \mathbb{E}\left[\partial^m \widetilde{f_x}(Z_t)\right].$$

L'idée consiste à obtenir une majoration moins triviale que (5.27) par des résultats d'analyse dans $L^2(\mathcal{P}_\lambda)$. L'adjoint de l'opérateur ∂ dans $L^2(\mathcal{P}_\lambda)$ s'écrit en effet

$$\partial_\lambda^\star g(y) = \frac{y}{\lambda} g(y-1) - g(y), \qquad y \in \mathbb{N},$$

avec la convention $g(-1) = 0$. Le polynôme de Charlier de degré n est défini par $P_{n,\lambda} = (\partial_\lambda^\star)^n \mathbf{1}$. Il vérifie, pour toute fonction f bornée

$$\int \partial^n f d\mathcal{P}_\lambda = \int f P_{n,\lambda} d\mathcal{P}_\lambda, \qquad \int P_{n,\lambda}^2 d\mathcal{P}_\lambda = \frac{n!}{\lambda^n}.$$

Ainsi, pour tous entiers $0 \leq k \leq m$ et $t \geq 0$,

$$P_t \partial^m f(x) = \mathbb{E}\left[\partial^m \widetilde{f}_x(Z_t)\right] = \mathbb{E}\left[P_{m-k,\lambda_t}(Z_t)\partial^k \widetilde{f}_x(Z_t)\right] = \mathbb{E}\left[P_{m-k,\lambda_t}(Z_t)\partial^k f(Z_t + B_t^x)\right].$$

D'après l'inégalité de Cauchy-Schwarz,

$$\|P_t(\partial^m f)\|_\infty \leq \|\partial^k f\|_\infty \left(\frac{m!}{\lambda_t^m}\right)^{1/2}, \qquad t \geq 0. \tag{5.29}$$

En combinant la majoration (5.27) et la majoration (5.29), on obtient le résultat suivant :

Theorem 5.3.1 (Bornes uniformes pour la solution de l'équation de Stein-Poisson)**.** *Pour tout entier $n \in \mathbb{N}$ et tout entier $k \leq n$, pour toute fonction $f : \mathbb{N} \to \mathbb{R}$ telle que $\|\partial^k f\|_\infty < \infty$, on a :*

$$\|\partial^n \overrightarrow{g_f}\|_\infty \leq \|\partial^k f\|_\infty \int_0^\infty e^{-(n+1)t} \left(2^{n+1-k} \wedge \left(\frac{(n+1-k)!}{\lambda(1-e^{-t})^{n+1-k}}\right)^{1/2}\right) dt.$$

En effet, le raisonnement des lignes précédentes où on avait supposé f bornée s'étend sans peine au cas $\|\partial^k f\|_\infty < \infty$. Examinons maintenant comment se décline le théorème.

Facteurs de Stein associés à la distance en variation totale

En prenant $k = 0$ et $n = 0$, on trouve que

$$\sup_{0 \leq f \leq 1} \|\overrightarrow{g_f}\|_\infty = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} \|\overrightarrow{g_f}\|_\infty \leq 1 \wedge \frac{1}{\sqrt{\lambda}}.$$

De même, en prenant $k = 0$ et $n = 1$, on obtient l'existence d'une constance $d > 0$ telle que

$$\sup_{0 \leq f \leq 1} \|\partial \overrightarrow{g_f}\|_\infty \leq 1 \wedge \frac{d \log \lambda}{\lambda}.$$

Ces résultats sont à comparer avec ceux obtenus par Barbour and Eagleson [1983] : on retrouve l'ordre de grandeur connu pour le premier facteur de Stein, mais pas le second, pour lequel un facteur supplémentaire $\log \lambda$ apparaît.

Facteurs de Stein associés à la distance en Wasserstein

Rappelons que la classe de fonction associée à la distance en Wasserstein est $\{f : \mathbb{N} \to \mathbb{R}, \|\partial f\|_\infty \leq 1\}$. En prenant $k = 1$ et $n = 0, 1, 2$, on trouve que

$$\sup_{\|\partial f\|_\infty \leq 1} \|\overrightarrow{g_f}\|_\infty \leq 1, \quad \sup_{\|\partial f\|_\infty \leq 1} \|\partial \overrightarrow{g_f}\|_\infty \leq 2\left(1 \wedge \frac{1}{\sqrt{\lambda}}\right), \quad \sup_{\|\partial f\|_\infty \leq 1} \|\partial^2 \overrightarrow{g_f}\|_\infty \leq \frac{4}{3} \wedge \left(\frac{\sqrt{2}\log\lambda}{\lambda} + \frac{d}{\lambda}\right).$$

Ces résultats sont à comparer à ceux de Barbour et al. [2015] : on retrouve l'ordre de grandeur connu pour le premier et le second facteur de Stein, mais on perd un facteur $\log \lambda$ sur le troisième.

Une majoration plus générale

Les deux cas ci-dessus sont des instances d'une situation plus générale : d'après Daly ([Daly, 2008, Théorème 1.3]), pour tout entier $n \geq 2$

$$\|\partial^n \overrightarrow{g_f}\|_\infty \leq \frac{2}{\lambda}\|\partial^{n-1} f\|_\infty, \qquad \|\partial^n \overrightarrow{g_f}\|_\infty \leq 2\left(1 \wedge \frac{1}{\sqrt{\lambda}}\right)\|\partial^n f\|_\infty.$$

Notre méthode permet d'obtenir la première majoration à un multiplicaitf facteur $\log \lambda$ près, et de retrouver le bon ordre de grandeur dans la deuxième.

Finissons par quelques perspectives. Il est important de remarquer que dans l'inégalité (5.26), majorer la quantité $|\partial^m P_t f|$ par $\|\partial^m f\|_\infty$ ne permettrait pas d'obtenir le théorème 5.3.1. La majoration plus fine qui est utilisée ici repose sur les propriétés particulières associées à l'approximande poissonienne, qu'on ne peut guère espérer étendre à des distributions quelconques.

Dans le chapitre suivant, cet obstacle est contourné avec succès en déterminant d'abord l'argmax des facteurs de Stein d'ordre $n \in \{0, 1\}$ ponctuels

$$\sup_{f \in \mathcal{F}} |\partial^n g_f(i)|, \qquad i \in \mathbb{N}.$$

L'utilisation systématique des formules d'entrelacement fait alors apparaître des quantités du type

$$\sup_{i \in \mathbb{N}} \mathbb{P}(\widetilde{X}_t^i = i), \qquad \sup_{i \in \mathbb{N}^*} \mathbb{P}(\widetilde{X}_t^i = i) - \mathbb{P}(\widetilde{X}_t^i = i-1),$$

pour un processus de Markov alternatif $(\widetilde{X}_t^i)_{t \geq 0}$, comme on le voit maintenant.

Chapter 6

Intertwinings and Stein's magic factors for birth-death processes

This chapter consists of the article Cloez and Delplancke [2016] written in collaboration with Bertrand Cloez (INRA Montpellier).

Abstract

This article investigates second order intertwinings between semigroups of birth-death processes and discrete gradients on $\mathbb{N}$. It goes one step beyond a recent work of Chafaï and Joulin which establishes and applies to the analysis of birth-death semigroups a first order intertwining. Similarly to the first order relation, the second order intertwining involves birth-death and Feynman-Kac semigroups and weighted gradients on $\mathbb{N}$, and can be seen as a second derivative relation. As our main application, we provide new quantitative bounds on the Stein factors of discrete distributions. To illustrate the relevance of this approach, we also derive approximation results for the mixture of Poisson and geometric laws.

6.1 Introduction

A birth-death process is a continuous-time Markov process with values in $\mathbb{N} = \{0, 1, \ldots\}$ which evolves by jumps of two types: onto the integer just above (birth) or just below (death). We denote by $\mathrm{BDP}(\alpha, \beta)$ the birth-death process with positive birth rate $\alpha = (\alpha(x))_{x\in\mathbb{N}}$ and non-negative death rate $\beta = (\beta(x))_{x\in\mathbb{N}}$ satisfying to $\beta(0) = 0$. Its generator is defined for every function $f : \mathbb{N} \to \mathbb{R}$ as

$$Lf(x) = \alpha(x)(f(x+1) - f(x)) + \beta(x)(f(x-1) - f(x)), \qquad x \in \mathbb{N}.$$

For a generator L, associated to a semigroup $(P_t)_{t\geq 0}$ and a Markov process $(X_t)_{t\geq 0}$ on $\mathbb{N}$, and a function V on $\mathbb{N}$ (usually called a potential), the Schrödinger operator $L - V$ is defined for every function f as

$$(L - V)f(x) = (Lf)(x) - V(x)f(x), \qquad x \in \mathbb{N},$$

and is associated to the Feynman-Kac semigroup $(P_t^V)_{t\geq 0}$ defined for all bounded or non-negative functions f on $\mathbb{N}$ as

$$(P_t^V f)(x) = \mathbb{E}\left[f(X_t^x) e^{-\int_0^t V(X_s^x)ds}\right], \qquad x \in \mathbb{N}, \quad t \geq 0.$$

The starting point of our work is the recent article Chafaï and Joulin [2013] which establishes a first order intertwining relation involving birth-death and Feynman-Kac semigroups, and discrete gradients on $\mathbb{N}$. For example, it reads as

$$\partial P_t = \widetilde{P}_t^{\widetilde{V}} \partial, \tag{6.1}$$

where ∂ is the discrete gradient defined by $\partial f(x) = f(x+1) - f(x)$, the notation $(\widetilde{P}_t^{\widetilde{V}})_{t \geq 0}$ standing for an alternative Feynman-Kac semigroup. Actually, the precise result holds more generally for weighted gradients and allows to derive known as well as new results on the analysis of birth-death semigroups.

According to this observation, the aim of the present article is to extend this work by stating a second order intertwining relation. More precisely, let us define the backward gradient ∂^* by

$$\partial^* f(x) = f(x-1) - f(x), \qquad x \in \mathbb{N}^* = \{1, 2, \ldots\}; \qquad \partial^* f(0) = -f(0).$$

Under some appropriate conditions on the potential $\widetilde{V}$, we derive a formula of the type

$$\partial^* \partial P_t = \widehat{P}_t^{\widehat{V}} \partial^* \partial, \tag{6.2}$$

where $(\widehat{P}_t^{\widehat{V}})_{t \geq 0}$ is a new Feynman-Kac semigroup. The operator $\partial^* \partial$, which writes

$$\partial^* \partial f(x) = -f(x+1) + 2f(x) - f(x-1), \qquad x \in \mathbb{N}^*,$$

can be seen as a discrete Laplacian. Similarly to the first order, this second order intertwining relation, which is our main result, is given in the more general case of weighted gradients.
Once our second order relation is established, it reveals to have many interesting consequences. In particular, we derive results on the estimation of the so-called Stein factors. Stein's factors, also known as Stein's magic factors, are upper bounds on derivatives of the solution to Stein's equation and a key point in Stein's method, introduced by Stein in Stein [1972], which consists in evaluating from above distances between probability distributions. Among the important results appearing more or less recently in this very active field of research, let us cite some references within the framework of discrete probabilities distributions. Stein's factors related to the Poisson approximation in total variation and Wasserstein distances are studied in the seminal paper Chen [1975], in the reference book Barbour et al. [1992] and in the recent article Barbour et al. [2015] for example. For the binomial negative approximation, one can cite Brown and Phillips [1999] for the total variation distance and Barbour et al. [2015] for the Wasserstein distance; for the geometric approximation in total variation distance, see Peköz [1996] and Peköz et al. [2013]. An important advance is made in Brown and Xia [2001], where a universal approach to evaluate Stein's factors for the total variation distance is developed. The work Eichelsbacher and Reinert [2008] provides Stein's factors for the total variation distance when the target distribution is a Gibbs distribution.

In the present article, we propose a universal technique to evaluate Stein's factors related to the approximation in total variation, Wasserstein and Kolmogorov distances. On the basis of some results derived in Brown and Xia [2001], the main ingredients are the so-called method of the generator and the intertwining relations presented above. To the authors' knowledge, the systematic use of this last ingredient, which comes from the functional analysis, seems to be new within the context of Stein's method. It allows to construct a unified framework for the derivation of Stein's factors, which applies to a wide range of discrete probability distributions, and might be developed similarly in the continous setting of diffusion processes in a forthcoming paper.
As a result, a case-by-case examination has to be done in each situation and example of interest, revealing that our upper bounds sometimes improve on the ones already known, and sometimes

are not as sharp. For example, we improve the first Stein factor related to the negative binomial approximation in total variation distance and we derive new Stein's factors for the geometric approximation in Wasserstein distance.
As an additional part of independent interest, we study the approximation of mixture of discrete distributions in the spirit of the Stein method. Combined with the Stein bounds, the obtained results have potential applications of which we give a flavour through the following example. Denote $\mathrm{NB}(r,p)$ the negative binomial distribution of parameters (r,p). It is a mixed Poisson distribution, converging in law towards the Poisson distribution $\mathcal{P}_\lambda$ in the regime $p \to 1$, $r \to \infty$ and $r(1-p)/p \to \lambda$. The following bound in Wasserstein distance W seems to be the first attempt to quantify this well-known convergence:

$$W\left(\mathrm{NB}(r,p), \mathcal{P}_{\frac{r(1-p)}{p}}\right) \leq \frac{8}{3\sqrt{2e}}\sqrt{\frac{r(1-p)}{p}}\frac{(1-p)}{p}.$$

To conclude this introduction, let us announce the structure of the article. In Section 6.2, we state in Theorem 6.2.2 our main result about the second order intertwining, after having recalled the first order intertwining; we follow with an application to the ergodicity of birth-death semigroups. In Section 6.3, we firstly present theoretical bounds on Stein's factors derived from the intertwinings, and secondly we investigate the approximation of mixture of distributions. In Section 6.4, our results are applied to a wide range of examples, including M/M/∞ process and Poisson approximation, Galton-Watson process with immigration and negative binomial approximation, and M/M/1 process and geometric approximation. The three last sections are devoted to the various proofs of the results previously stated: Section 6.5 deals with the preparation and proof of our main result Theorem 6.2.2, Section 6.6 gathers the proofs of the bounds on Stein's factors and finally, a useful upper bound related to the pointwise probabilities of the M/M/∞ process is proved in Section 6.7.

Acknowledgement:
The authors thank A. Joulin for the interest he took in this research through many fruitful discussions and for sharing his insights on the subject.
This work was partially supported by the CIMI (Centre International de Mathématiques et d'Informatique) Excellence program, by the ANR PIECE (ANR-12-JS01-0006-01) and STAB (ANR-12-BS01-0019) and by the Chaire Modélisation Mathématique et Biodiversité.

6.2 Main result

Before stating our main result Theorem 6.2.2, let us introduce some notation. The set of positive integers $\{1,2,\dots\}$ is denoted $\mathbb{N}^*$. For all real-valued functions f on $\mathbb{N}$ and sets $A \subset \mathbb{N}$, we define $\|f\|_{\infty,A} = \sup\{|f(x)|, x \in A\}$ and $\|f\|_\infty = \|f\|_{\infty,\mathbb{N}}$. For all sequences u on $\mathbb{N}$, the shift-forward and shift-backward of u are defined as:

$$\overrightarrow{u}(x) = u(x+1), \qquad x \in \mathbb{N}; \qquad \overleftarrow{u}(x) = u(x-1), \qquad x \in \mathbb{N}^*; \qquad \overleftarrow{u}(0) = 0.$$

The symbol $\mathcal{P}$ stands for the set of probability measures on $\mathbb{N}$ and we denote by $\mathcal{L}(W)$ the distribution of the random variable W. For all real-valued functions f on $\mathbb{N}$ and $\mu \in \mathcal{P}$, we use indifferently the notation

$$\int f d\mu = \mu(f) = \sum_{x\in\mathbb{N}} f(x)\mu(x).$$

Recall that the discrete forward and backward gradients are defined for all real-valued functions f on $\mathbb{N}$ by

$$\partial f(x) = f(x+1) - f(x), \quad x \in \mathbb{N}; \quad \partial^* f(x) = f(x-1) - f(x), \quad x \in \mathbb{N}^*, \quad \partial^* f(0) = -f(0),$$

the convention chosen for ∂^* in 0 being interpreted as a Dirichlet-type condition (implicitly we set $f(-1) = 0$). Letting u be a positive sequence, we define the weighted gradients ∂_u and ∂_u^* respectively by

$$\partial_u = \frac{1}{u}\partial, \qquad\qquad \partial_u^* = \frac{1}{u}\partial^*.$$

With this notation, the generator of the BDP(α, β) reads for every function $f : \mathbb{N} \to \mathbb{R}$ as

$$Lf = \alpha\,\partial f + \beta\,\partial^* f.$$

Let us assume that the birth rate α is positive on $\mathbb{N}$ and that the death rate β is positive on $\mathbb{N}^*$ with moreover $\beta(0) = 0$. Hence the process is irreducible; to ensure that the process is ergodic and non-explosive we further assume respectively that (Chen [2004])

$$\sum_{x=1}^{+\infty} \frac{\alpha(0)\alpha(1)\ldots\alpha(x-1)}{\beta(1)\beta(2)\ldots\beta(x)} < \infty, \quad \sum_{x=1}^{\infty}\left(\frac{1}{\alpha(x)} + \frac{\beta(x)}{\alpha(x)\alpha(x-1)} + \cdots + \frac{\beta(x)\ldots\beta(1)}{\alpha(x)\ldots\alpha(0)}\right) = \infty.$$

The measure π defined on $\mathbb{N}$ as

$$\pi(0) = \left(1 + \sum_{x\geq 1}\prod_{y=1}^{x}\frac{\alpha(y-1)}{\beta(y)}\right)^{-1}, \qquad \pi(x) = \pi(0)\prod_{y=1}^{x}\frac{\alpha(y-1)}{\beta(y)}, \qquad x \in \mathbb{N}, \tag{6.3}$$

is then the invariant, and symmetric, probability measure for the associated semigroup.
Recall that if $(P_t)_{t\geq 0}$ is a Markov semigroup on $\mathbb{N}$ associated to the process $(X_t)_{t\geq 0}$ and if the potential $V : \mathbb{N} \to \mathbb{R}$ is bounded from below, the Feynman-Kac semigroup $(P_t^V)_{t\geq 0}$ is defined for all bounded or non-negative functions f on $\mathbb{N}$ as

$$(P_t^V f)(x) = \mathbb{E}\left[f(X_t^x)e^{-\int_0^t V(X_s^x)ds}\right], \qquad x \in \mathbb{N}, \quad t \geq 0. \tag{6.4}$$

When V is positive, the formula (6.4) admits an interpretation involving a killed, or extended, Markov process. Add a new state a to $\mathbb{N}$ and extend functions f on $\mathbb{N}$ to $\mathbb{N} \cup \{a\}$ by $f(a) = 0$. Then, we have:

$$P_t^V f(x) = \mathbb{E}\left[f(Y_t^x)\mathbf{1}_{\{Y_t^x \neq a\}}\right],$$

where the process $(Y_t^x)_{t\geq 0}$ is absorbed in a with rate $V(Y_t^x)$. The generator of the process $(Y_t^x)_{t\geq 0}$ acts on real-valued functions on $\mathbb{N} \cup \{a\}$ by the formula

$$(Kf)(x) = (Lf|_{\mathbb{N}})(x) + V(x)(f(a) - f(x)). \tag{6.5}$$

This interpretation can be extended to the case where V is bounded from below by adding and subtracting a constant to V inside the exponential.
The Kolmogorov equations associated to the Schrödinger operator $L - V$ and the Feynman-Kac semigroup defined in the introduction read for all functions f in the domain of L as

$$\partial_t P_t^V f = (L - V)P_t^V f = P_t^V (L - V)f, \qquad t \geq 0. \tag{6.6}$$

Here ∂_t denotes the derivative in time. In the following, when using a Feynman-Kac semigroup, we will always assume that the equation (6.6) stands for all bounded real-valued functions on $\mathbb{N}$. It is the case for example when L is the generator of a birth-death process with rates (α, β), and α, β, V are P_t-integrable for all $t \geq 0$.

In order to state the first intertwining relation, we associate to any positive sequence u a modified birth-death process on $\mathbb{N}$ with semigroup $(P_{u,t})_{t\geq 0}$, generator L_u, and potential V_u. For all functions $f:\mathbb{N}\to\mathbb{R}$ set

$$L_u f = \alpha_u\,\partial f + \beta_u\,\partial^* f, \qquad V_u = \alpha - \alpha_u + \overrightarrow{\beta} - \beta_u,$$
$$\alpha_u(x) = \frac{u(x+1)}{u(x)}\alpha(x+1), \qquad \beta_u(x) = \frac{u(x-1)}{u(x)}\beta(x)\mathbf{1}_{x\in\mathbb{N}^*}, \qquad x\in\mathbb{N}.$$

Under the compact form $V_u = \partial_u\left(\overleftarrow{u}\beta - u\alpha\right)$ one can see the parallel with the analogous formulas in the diffusion setting (Bakry and Émery [1985]; Bonnefont and Joulin [2014]).
We recall now the first order intertwining relation, due to Chafaï and Joulin [2013].

Theorem 6.2.1 (First order intertwining relation)**.** *If V_u is bounded from below, it holds for every real-valued function on $\mathbb{N}$ such that $\|\partial_u f\|_\infty < +\infty$ that:*

$$\partial_u P_t f = P_{u,t}^{V_u}\,\partial_u f, \qquad t\geq 0. \tag{6.7}$$

Although we will not prove this result in full generality, a new proof is proposed in Section 6.5.2 when the weight is $u=1$, the birth rates α are non-increasing and the death rates β are non-decreasing. This proof is based on a coupling argument and gives a probabilistic interpretation of the semigroup (and its jump rates) in the right-hand side of (6.7).

We now turn to the main theorem of this article. Let u and v be positive sequences and assume that the potential V_u defined above is non-increasing on $\mathbb{N}$. We define a modified process on $\mathbb{N}$ with semigroup $(P_{u,*v,t})_{t\geq 0}$ and generator $L_{u,*v}$ as follows: for all real-valued functions f on $\mathbb{N}$, set

$$\begin{aligned}
(L_{u,*v}f)(x) &= \alpha_{u,*v}(x)\partial f(x) + \beta_{u,*v}(x)\partial^* f(x)\\
&\quad + (\partial_v^* V_u)(x)\left(\sum_{j=0}^{x-2} v(j)\right)\sum_{k=0}^{x-2}\frac{v(k)}{\left(\sum_{j=0}^{x-2} v(j)\right)}(f(k)-f(x)), && x\geq 2,\\
(L_{u,*v}f)(x) &= \alpha_{u,*v}(x)\partial f(x) + \beta_{u,*v}(x)\partial^* f(x), && x=0,1,\\
\alpha_{u,*v}(x) &= \frac{v(x+1)}{v(x)}\frac{u(x+1)}{u(x)}\alpha(x+1), && x\in\mathbb{N},\\
\beta_{u,*v}(x) &= \frac{v(x-1)}{v(x)}\frac{u(x-2)}{u(x-1)}\beta(x-1) + v(x-1)\,\partial_v^* V_u(x), && x\geq 2,\\
\beta_{u,*v}(1) &= v(0)\,\partial_v^* V_u(1), \quad \beta_{u,*v}(0) = 0.
\end{aligned}$$

In contrast with the previous semigroups, this modified process is not a birth-death process in general. Indeed, if the process starts at a point $x\geq 2$, it can jump on the set $\{0,\ldots,x-2\}$ with rate $(\partial_v^* V_u)(x)\left(\sum_{j=0}^{x-2} v(j)\right)$. Remark that both this quantity and the death rate in 1, $\beta_{u,*v}(1) = (\partial^* V_u)(1)$, are non-negative thanks to the hypothesis V_u non-increasing on $\mathbb{N}$. We also define the potential $V_{u,*v}$ as

$$\begin{aligned}
V_{u,*v}(x) &= \left(1+\frac{u(x)}{u(x-1)}\right)\alpha(x) - \left(1+\frac{v(x+1)}{v(x)}\right)\frac{u(x+1)}{u(x)}\alpha(x+1)\\
&\quad + \beta(x+1) - \frac{v(x-1)}{v(x)}\frac{u(x-2)}{u(x-1)}\beta(x-1) - \left(\sum_{j=0}^{x-1} v(j)\right)\partial_v^* V_u(x), \qquad x\geq 1,\\
V_{u,*v}(0) &= \alpha(0) - \left(1+\frac{v(1)}{v(0)}\right)\frac{u(1)}{u(0)}\alpha(1) + \beta(1).
\end{aligned}$$

We are ready to state our main result.

Theorem 6.2.2 (Second order intertwining relation). *Assume that V_u is non-increasing, bounded from below, that $\inf_{x\in\mathbb{N}} v(x) > 0$ and that $V_{u,*v}$ is bounded from below. Then for every real-valued function on $\mathbb{N}$ such that $\|\partial_u f\|_\infty < +\infty$, we have*

$$\partial_v^*\partial_u(P_t f) = P_{u,*v,t}^{V_{u,*v}}(\partial_v^*\partial_u f), \qquad t \geq 0.$$

Since some preparation is needed, the proof of Theorem 6.2.2 is postponed to Section 6.5.

Remark 6.2.3 (Propagation of convexity ?). *Under the assumptions of Theorem 6.2.2, if $\partial_v^*\partial_u f$ is non-negative, so is $\partial_v^*\partial_u P_t f$ for all $t \geq 0$. A similar property for the first order intertwining admits an interpretation in terms of propagation of monotonicity ([Chafaï and Joulin, 2013, Remark 2.4]): the intertwining relation* (6.7) *implies that if a function $f : \mathbb{N} \to \mathbb{R}$ is non-decreasing, then so is $P_t f$ for every $t \geq 0$. However, it is not clear whether there is an analogous nice interpretation for the second order intertwining because, in contrast to the continous space case, the condition $\partial_v^*\partial_u f \geq 0$ is not equivalent to the convexity of f (even for $u = v = 1$).*

Let us comment further on Theorem 6.2.2. The interpretation of a Feynman-Kac semigroup as an extended Markov semigroup sheds light on various aspects of Theorem 6.2.2. As the first-order potential V_u is bounded from below, recall that the Feynman-Kac semigroup $(P_{u,t}^{V_u})_{t\geq 0}$ appearing in the right-hand side of equation (6.7) can be represented as a Markov semigroup $(S_t)_{t\geq 0}$ related to the process $(Y_t)_{t\geq 0}$ on $\mathbb{N} \cup \{-1\}$ by adding a point $a = -1$. The Markov process $(Y_t)_{t\geq 0}$ is then non-irreducible and absorbed in -1. To differentiate again in the equation (6.7) amounts to differentiate the Markov semigroup $(S_t)_{t\geq 0}$.
Firstly, this explains intuitively the use of the backward weighted gradient ∂_u^* instead of the regular weighted gradient ∂_u. Indeed, to deal with the absorption of the Markov process in -1, additional information at the boundary is needed. The use of ∂^* gives the missing information, since the knowledge of $\partial^* g$ is equivalent to the knowledge of ∂g in addition with the knowledge of $g(0) = -\partial^* g(0)$.
Secondly, this allows to understand the hypotheses required for Theorem 6.2.2 to apply. The main assumption of this theorem is V_u to be non-increasing. As noticed before, this assumption is necessary in order to have well-defined objects. The following remark provides another justification.

Remark 6.2.4 (Around the monotonicity assumption). *On the one hand, the second intertwining relation is equivalent to a first intertwining relation for the extended Markov semigroup $(S_t)_{t\geq 0}$. On the other hand, if a first intertwining relation holds for $(S_t)_{t\geq 0}$, then $(S_t)_{t\geq 0}$ propagates the monotonicity. Set $f = \mathbf{1}_{\mathbb{N}} = 1 - \mathbf{1}_{\{-1\}}$. Then for all $x, y \in \mathbb{N}$, $S_0 f(x) = S_0 f(y) = 1$ and by formula (6.5),*

$$\begin{aligned}
\partial_t(S_t f)(x)|_{t=0} &= (L_u f|_{\mathbb{N}})(x) + V_u(x)(f(-1) - f(x)) = -V_u(x),\\
\partial_t\big(S_t f(x) - S_t f(y)\big)|_{t=0} &= V_u(y) - V_u(x).
\end{aligned}$$

The function f is non-decreasing on $\mathbb{N} \cup \{-1\}$ and a necessary condition for $S_t f$ to be non-decreasing for all $t \geq 0$ is, in the light of the preceding equation, that $V_u(y) - V_u(x) \leq 0$ whenever $x \leq y$, i.e. V_u is non-increasing on $\mathbb{N}$.

If V_u is constant, then Theorem 6.2.2 admits a variant involving the gradient $\partial_v\partial_u$ instead of $\partial_v^*\partial_u$, which is stated in Theorem 6.2.5 below for the sake of completeness. In the applications, when V_u is constant, we choose to invoke Theorem 6.2.5 in lieu of Theorem 6.2.2, because the underlying arguments are much simpler. Indeed, in this case the equation (6.7) reduces to

$$\partial_u P_t f = e^{-V_u t} P_{u,t}\partial_u f, \qquad t \geq 0,$$

and it is no longer required to extend artificially the Markov process, nor to add information at the boudary, in order to differentiate a second time. As a matter of fact, one can notice that if V_u is constant, then the BDP associated to the semigroup $(P_{u,*v,t})_{t\geq 0}$ of Theorem 6.2.2 do not visit the state 0 unless it starts there.
In order to state the theorem, a new birth-death semigroup $(P_{u,v,t})_{t\geq 0}$ with generator $L_{u,v}$ and a potential $V_{u,v}$ is introduced. Set for all real-valued functions on $\mathbb{N}$:

$$\begin{aligned} L_{u,v}f(x) &= \alpha_{u,v}\partial f(x) + \beta_{u,v}\partial^* f(x), && x\in\mathbb{N},\\ \alpha_{u,v}(x) &= \frac{v(x+1)}{v(x)}\frac{u(x+2)}{u(x+1)}\alpha(x+2), \quad \beta_{u,v}(x) = \frac{v(x-1)}{v(x)}\frac{u(x-1)}{u(x)}\beta(x), && x\in\mathbb{N},\\ V_{u,v}(x) &= \alpha(x) - \frac{v(x+1)}{v(x)}\frac{u_{x+2}}{u(x+1)}\alpha(x+2) + \left(\frac{u(x)}{u(x+1)}+1\right)\beta(x+1)\\ &\quad - \left(1+\frac{v(x-1)}{v(x)}\right)\frac{u(x-1)}{u(x)}\beta(x), && x\in\mathbb{N}. \end{aligned}$$

In contrast to the Markov semigroup $(P_{u,*v,t})_{t\geq 0}$, the semigroup $(P_{u,v,t})_{t\geq 0}$ is always a birth-death semigroup.

Theorem 6.2.5 (Alternative version of the second intertwining relation)**.** *Assume that V_u is constant on $\mathbb{N}$ and that $V_{u,v}$ is bounded from below. For all real-valued functions on $\mathbb{N}$ such that $\|\partial_u f\|_\infty < +\infty$ and $\|\partial_v\partial_u f\|_\infty < +\infty$, we have:*

$$\partial_v\partial_u(P_t f) = P^{V_{u,v}}_{u,v,t}(\partial_v\partial_u f), \qquad t\geq 0.$$

Remark 6.2.6 (Link between the two versions of the second intertwining)**.** *Surprisingly, it is only possible to deduce directly Theorem 6.2.5 from Theorem 6.2.2 in the case where the sequence v is constant. When $v=1$ for instance, one can write that $\partial^*\partial_u f(\cdot+1) = -\partial\partial_u f$, yielding under the appropriate assumptions on $f:\mathbb{N}\to\mathbb{R}$ that:*

$$P_{u,1,t}f(x) = P_{u,*1,t}\overleftarrow{f}(x+1), \qquad x\in\mathbb{N},\quad t\geq 0.$$

At the level of the processes, this equation can be reformulated into the equality in law:

$$X^x_{1,u,t} = X^{x+1}_{1,*u,t} - 1, \qquad x\in\mathbb{N},\quad t\geq 0,$$

where $(X^x_{1,*u,t})_{t\geq 0}$ and $(X^x_{1,u,t})_{t\geq 0}$ are the Markov processes corresponding respectively to the semigroups $(P_{u,*1,t})_{t\geq 0}$ and $(P_{u,1,t})_{t\geq 0}$. If v is not constant, no similar relation holds in general.

Remark 6.2.7 (Other versions)**.** *It is possible to derive similar theorems for other gradients. For example, if the gradient $\partial^\star$ is defined as $\partial^\star f = \partial^* f$ on $\mathbb{N}^*$ and with the Neumann-like boundary condition in 0, $\partial^\star f(0) = 0$, then the analogous theorem to Theorem 6.2.2 holds for $\partial_v\partial^\star_u$. It is also possible to derive intertwining relations in the case where the semigroup lives on $[\![0,n]\!]$, although the underlying structures are rather different: for instance, the condition V_u non-increasing is no longer necessary.*

Let us turn to our first application of Theorem 6.2.2 and its variant Theorem 6.2.5. The first order intertwining relation recalled in Theorem 6.2.1 yields a contraction property in Wasserstein distance. Precisely, under the assumptions of Theorem 6.2.1, by [Chafaï and Joulin, 2013, Corollary 3.1], we have for all $\mu,\nu\in\mathcal{P}$,

$$W_{d_u}(\mu P_t, \nu P_t) \leq e^{-\sigma(u)t} W_{d_u}(\mu,\nu), \tag{6.8}$$

where the distance d_u on $\mathbb{N}$ and the related Wasserstein distance W_{d_u} on $\mathcal{P}$ are defined in the forthcoming section, Section 6.3.1. Similarly, Theorems 6.2.2 and 6.2.5 lead to a contraction property for the distances $\zeta_{u,*v}$ and $\zeta_{u,v}$, defined respectively for two sequence of positive weights u and v by

$$\zeta_{u,*v} = \sup_{f\in\mathcal{F}_{u,*v}} |\mu(f)-\nu(f)|, \qquad \mathcal{F}_{u,*v} = \{f:\mathbb{N}\to\mathbb{R},\ \|\partial_u f\|_\infty<\infty,\ \|\partial_v^*\partial_u f\|_\infty\le 1\},$$
$$\zeta_{u,v} = \sup_{f\in\mathcal{F}_{u,v}} |\mu(f)-\nu(f)|, \qquad \mathcal{F}_{u,v} = \{f:\mathbb{N}\to\mathbb{R},\ \|\partial_u f\|_\infty<\infty,\ \|\partial_v\partial_u f\|_\infty\le 1\}.$$

We call $\zeta_{u,*v}$ and $\zeta_{u,v}$ second order Zolotarev-type distances since they are simple metric distances in the sense of Zolotarev (Zolotarev [1976]) and can be seen as the discrete counterparts of the distance ζ_2 defined on the set of real probability distributions (the distance ζ_2, introduced in Zolotarev [1976] and further studied in Rio [1998], is associated to the set of continuously differentiable functions on $\mathbb{R}$ whose derivative is Lipschitz). The contraction property reads as follows:

Theorem 6.2.8 (Contraction of the BDP in second order distances).

— *Under the same hypotheses as in Theorem 6.2.2, we set $\sigma(u,*v)=\inf V_{u,*v}$. Then, for all $\mu,\nu\in\mathcal{P}$, we have:*

$$\zeta_{u,*v}(\mu P_t,\nu P_t)\le e^{-\sigma(u,*v)t}\zeta_{u,*v}(\mu,\nu). \tag{6.9}$$

— *Under the assumptions of Theorem 6.2.5, define $\sigma(u,v)=\inf V_{u,v}$. Letting $\mu,\nu\in\mathcal{P}$, it stands that:*

$$\zeta_{u,v}(\mu P_t,\nu P_t)\le e^{-\sigma(u,v)t}\zeta_{u,v}(\mu,\nu). \tag{6.10}$$

Proof. The proof is done in the first case, the second one being similar. For all real-valued functions f on $\mathbb{N}$ such that $\|\partial_u f\|_\infty<\infty$ and $\|\partial_v^*\partial_u f\|_\infty\le 1$, Theorem 6.2.2 implies that

$$\|\partial_v^*\partial_u P_t f\|_\infty \le e^{-\sigma(u,*v)t}\|\partial_v^*\partial_u f\|_\infty \le e^{-\sigma(u,*v)t}, \qquad t\ge 0.$$

Hence,

$$\begin{aligned}\zeta_{u,*v}(\mu P_t,\nu P_t) &= \sup\left\{\left|\int P_t f\,d\mu-\int P_t f\,d\nu\right|,\ \|\partial_v^*\partial_u f\|_\infty\le 1\right\}\\ &\le \sup\left\{\left|\int g\,d\mu-\int g\,d\nu\right|,\ \|g\|_\infty\le e^{-\sigma(u,*v)t}\right\}\\ &= e^{-\sigma(u,*v)t}\zeta_{u,*v}(\mu,\nu).\end{aligned}$$

□

If the quantity $\sigma(u,*v)$ (resp. $\sigma(u,v)$) is positive, the first bound (resp. the second) is a contraction. In particular, if we take $\nu=\pi$ the invariant measure of the BDP, then Theorem 6.2.8 gives the rate of convergence of the BDP towards its invariant measure in a second order distance.

Remark 6.2.9 (Generalization and optimality).

— *The proof of Theorem 6.2.8 can be generalized to the Zolotarev-type distance associated to the set of functions $f:\mathbb{N}\to\mathbb{R}$ such that $\|Df\|_\infty\le 1$ as soon as we have an inequality of the type $\|DP_tf\|_\infty\le e^{-\sigma t}\|Df\|_\infty$ for every $t\ge 0$, some $\sigma>0$ and some finite difference operator D. In Section 6.4 below, we detail such convergences in higher order Zolotarev-type distances.*

— *By arguments similar to those developed in [Chafaï and Joulin, 2013, corollary 3.1], one can prove that the constants $\sigma(u,*v)$ and $\sigma(u,v)$ in the equations* (6.9) *and* (6.10) *are optimal. Indeed, the argument of Chafaï and Joulin [2013] relies on the propagation of the monotonicity and we have the analogous property at the second order (cf Remark 6.2.3).*

— *Using [Chen, 2004, Theorem 9.25], we see that, choosing a good sequence u, it is possible to obtain the contraction in the Wasserstein distance* (6.8) *at a rate corresponding to the spectral gap (even if there is no corresponding eigenvector). For the second order, we do not know if it is always possible to find sequences u, v such that $\sigma(*u, v)$ or $\sigma(u, v)$ is equal to the second smallest positive eigenvalue of $-L$.*

In the following section we focus our attention on our main application of intertwining relations, Stein's factors.

6.3 Application to Stein's magic factors

6.3.1 Distances between probability distributions

First of all, we introduce the distances between probability measures used to measure approximations in the sequel. They are of the form

$$\zeta_{\mathcal{F}}(\mu, \nu) = \sup \left\{ |\mu(f) - \nu(f)|, \, f \in \mathcal{F} \right\},$$

where $\mathcal{F}$ is a subset of the set of real-valued functions on $\mathbb{N}$. The distances $\zeta_{u,*v}$, $\zeta_{u,v}$ presented at the end of the preceding section were examples of such distances; we now recall the definition of three classical distances on $\mathcal{P}$.

Total variation distance. The total variation distance d_{TV} is the distance associated to the set $\mathcal{F}_{\mathrm{TV}}$ of real-valued functions on $\mathbb{N}$ such that $0 \le f \le 1$. In contrast to the continous space case, the topology induced by the total variation distance on $\mathbb{N}$ is exactly the convergence in law. Some authors prefer to define the total variation distance as the distance associated to the set $\mathcal{F} = \{f : \mathbb{N} \to \mathbb{R}, \|f\|_\infty \le 1\}$. The two definitions vary by a factor $\frac{1}{2}$:

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{0 \le f \le 1} |\mu(f) - \nu(f)| = \frac{1}{2} \sup_{\|f\|_\infty \le 1} |\mu(f) - \nu(f)| = \frac{1}{2} \sum_{x \in \mathbb{N}} |\mu(x) - \nu(x)|.$$

Wasserstein distance. For a distance d on $\mathbb{N}$ let us call $\mathrm{Lip}(d)$ the set of real-valued functions on $\mathbb{N}$ such that

$$|f(x) - f(y)| \le d(x, y), \qquad x, y \in \mathbb{N}.$$

The Wasserstein distance between two probability measures μ and ν of $\mathcal{P}$ is defined as

$$W_d(\mu, \nu) = \inf \int d(x, y) d\Pi(x, y),$$

where the infimum is taken over all probability measures Π on $\mathbb{N}^2$ whose first marginal is μ and second marginal is ν. By Kantorovich-Rubinstein theorem (see e.g. Szulga [1982]),

$$W_d(\mu, \nu) = \zeta_{\mathrm{Lip}(d)}(\mu, \nu).$$

For a positive sequence u, define the distance d_u on $\mathbb{N}$ as

$$d_u(x, y) = \sum_{k=x}^{y-1} u(k), \quad x < y; \qquad d_u(x, y) = d_u(y, x), \quad x > y; \qquad d_u(x, y) = 0, \quad x = y.$$

Let us observe that $\mathrm{Lip}(d_u) = \{f : \mathbb{N} \to \mathbb{R}, \|\partial_u f\|_\infty \le 1\}$. Hence

$$W_{d_u}(\mu, \nu) = \sup_{f \in \mathrm{Lip}(d_u)} |\mu(f) - \nu(f)| = \sup_{\|\partial_u f\|_\infty \le 1} |\mu(f) - \nu(f)|.$$

The distance associated to the constant sequence equal to 1 is the usual distance $d_1(x,y) = |x-y|$. We denote by $W = W_{d_1}$ the associated Wasserstein distance.

Kolmogorov distance. The Kolmogorov distance is defined as the metric distance associated to the set $\mathcal{F}_K$ of indicator functions of intervals $[0,x]$:

$$d_K(\mu,\nu) = \sup_{x\in\mathbb{N}} |\mu([0,x]) - \nu([0,x])|.$$

Comparison between distances. For all $\mu,\nu \in \mathcal{P}$,

$$d_K(\mu,\nu) \;\leq\; d_{\mathrm{TV}}(\mu,\nu) \;\leq\; \frac{1}{\inf_{\mathbb{N}} u} W_{d_u}(\mu,\nu).$$

Indeed, both inequalities are consequences of the inclusions

$$\mathcal{F}_K \subset \mathcal{F}_{\mathrm{TV}} \subset \frac{1}{\inf_{\mathbb{N}} u}\mathrm{Lip}(d_u).$$

The second inclusion follows from the implication

$$0 \leq f \leq 1 \Rightarrow \|\partial f\|_\infty \leq \frac{1}{\inf_{\mathbb{N}} u}.$$

The total variation distance is invariant by translation, whereas intuitively the Wasserstein distance gives more weight to the discrepancy between $\mu(x),\nu(x)$ if it occurs for a large integer x. The Kolmogorov distance may be used as an alternative to the total variation distance when the latter is too strong to measure the involved quantities.

6.3.2 Basic facts on Stein's method

Given a probability measure μ and a target probability measure π of $\mathcal{P}$, the Stein-Chen method provides a way to estimate the distances of the type $\zeta_{\mathcal{F}}(\mu,\pi)$. More precisely, consider a Stein's operator S:

$$S(g)(x) = \alpha(x)g(x+1) - \beta(x)g(x), \qquad x \in \mathbb{N}; \qquad \beta_0 = 0,$$

characterizing the probability measure π (meaning that $\int S(g)d\mu = 0$ for every function $g : \mathbb{N} \to \mathbb{R}$ in a sufficiently rich class of functions if and only if $\mu = \pi$) and the associated Stein equation

$$S(g_f) = f - \int f d\pi. \tag{6.11}$$

We call g_f a solution to Stein's equation. The interest of such solutions comes from the following error bound:

$$\zeta_{\mathcal{F}}(\mu,\pi) = \sup_{f\in\mathcal{F}} |\mu(f) - \pi(f)| = \sup_{f\in\mathcal{F}} \left| \int S(g_f)d\mu \right|. \tag{6.12}$$

As a consequence of equation (6.12), if it can be shown that

$$\left| \int S(g_f)d\mu \right| \leq \varepsilon_0 \|g_f\|_\infty + \varepsilon_1 \|\partial(g_f)\|_\infty, \tag{6.13}$$

then it follows that

$$\zeta_{\mathcal{F}}(\mu,\pi) \leq \varepsilon_0 \sup_{f\in\mathcal{F}} \|g_f\|_\infty + \varepsilon_1 \sup_{f\in\mathcal{F}} \|\partial(g_f)\|_\infty.$$

This strategy of proof is widely used, for example in the references about Stein's method provided in the introduction.

A key point of this approach consists then in evaluating the so-called first and second Stein factors, also known as *magic factors*:

$$\sup_{f\in\mathcal{F}} \|g_f\|_\infty, \qquad\qquad \sup_{f\in\mathcal{F}} \|\partial(g_f)\|_\infty.$$

Observe that the equation (6.11) does not determine the value of $g_f(0)$. When evaluating the first Stein factor $\sup_{f\in\mathcal{F}}\|g_f\|_\infty$, we pick for every $f \in \mathcal{F}$ the solution g_f of (6.11) such that $g_f(0) = 0$. Hence, it is sufficient to consider the quantity

$$\sup_{f\in\mathcal{F}} \|g_f\|_{\infty,\mathbb{N}^*} = \sup_{f\in\mathcal{F}} \|\overrightarrow{g_f}\|_\infty.$$

Similarly, for the second Stein factor, picking solutions g_f to (6.11) satisfying to $g_f(0) = g_f(1)$, i.e. $\partial g_f(0) = 0$, allows to consider only the quantity

$$\sup_{f\in\mathcal{F}} \|\partial g_f\|_{\infty,\mathbb{N}^*} = \sup_{f\in\mathcal{F}} \|\partial(\overrightarrow{g_f})\|_\infty.$$

To evaluate the above quantities, we use a method known as method of the generator and the semigroup representation deriving from it. Set L the generator and $(P_t)_{t\geq 0}$ the semigroup associated to the BDP(α,β) and assume that $(P_t)_{t\geq 0}$ is invariant with respect to the target probability distribution π. The operators S and L are linked by the relation

$$Lh = S(-\partial^* h).$$

The Poisson equation reads as

$$L(h_f) = f - \mu(f),$$

the centered solution h_f being given by the expression

$$h_f = -\int_0^\infty (P_t f - \mu(f))dt.$$

Then, we obtain a solution g_f to Stein's equation (6.11) under the so-called semigroup representation:

$$g_f = -\partial^*(h_f) = \int_0^\infty \partial^*(P_t f)dt. \tag{6.14}$$

Let us end this section by noticing that the generator L itself also characterizes π in the sense defined above. However, to make the step (6.13) easier, one often prefers working with Stein's operator S which has a simpler expression.

6.3.3 Bounds on Stein's magic factors

In this section, theoretical bounds on the first and second order Stein factors are proposed for the approximation in total variation, Wasserstein and Kolmogorov distances. Proofs are postponed to Section 6.6 in order to clarify the presentation. Before turning to the results, a few general comments are made.

1. Our method evaluates Stein factors by quantities of the form

$$\int_0^\infty e^{-\kappa t} \sup_{i\in\mathbb{N}} \mathbb{P}(\widetilde{X}_t^i = i)dt, \qquad \int_0^\infty e^{-\kappa t} \sup_{i\in\mathbb{N}^*} \Big(\mathbb{P}(\widetilde{X}_t^i = i) - \mathbb{P}(\widetilde{X}_t^i = i-1)\Big)\, dt.$$

The Markov process $(\widetilde{X}_t)_{t\geq 0}$ which occurs is an alternative process and is not necessarily the same as the BDP(α,β) with semigroup $(P_t)_{t\geq 0}$ appearing in the semigroup representation (6.14). To our knowledge, this is new and makes the originality of our work.

2. While the detailed proofs of the forthcoming results are given in Section 6.6, the scheme of proof is briefly explained here. First, the argmax f_i of the pointwise Stein factor

$$\sup_{f\in\mathcal{F}} |\partial^k g_f(i)|, \qquad i\in\mathbb{N}^*, \quad k\in\{0,1\},$$

is obtained by resuming and generalizing results from Brown and Xia [2001]. Secondly, the function f_i is plugged in the semigroup representation:

$$\partial^k g_{f_i}(i) = \int_0^\infty \partial^k \partial^* P_t f_i \, dt, \qquad k\in\{0,1\}.$$

The intertwining relations of Section 6.2 are then used to rewrite the term $\partial^k\partial^* P_t$. This technique is already employed for Poisson approximation in some works, Barbour and Brown [1992] and Barbour and Xia [2006] for example. In that context, the intertwining relation reads as:

$$\partial P_t = e^{-t} P_t \partial, \qquad t\geq 0,$$

where the semigroup $(P_t)_{t\geq 0}$ is the same on the left and on the right. The use of the intertwining relations permits to go beyond this case and to construct a universal method to derive Stein's factors.

3. For the sake of clarity, the present section only includes results on the uniform Stein factors. However, it can be seen in Section 6.6 that our upper bounds on the pointwise Stein factors are often sharp.

4. For the second order Stein factor, two sets of assumptions are used:

   Assumptions 6.3.1 (Assumptions)**.**

   $\mathbf{H_1}$*: The potential V_1 is non-increasing and non-negative, the potential $V_{1,*u}$ is bounded from below, and the sequence u is bounded from below by a positive constant. In this case, we define $\sigma(1,*u) = \inf_{\mathbb{N}} V_{1,*u}$ and denote by $(X^i_{1,*u,t})_{t\geq 0}$ the Markov process of generator $L_{1,*u}$ such that $X^i_{1,*u,0} = i$.*

   $\mathbf{H_2}$*: The potential V_1 is a non-negative constant and the potential $V_{1,u}$ is bounded from below. In this case, set $\sigma(1,u) = \inf_{\mathbb{N}} V_{1,u}$ and call $(X^i_{1,u,t})_{t\geq 0}$ the birth-death process of generator $L_{1,u}$ such that $X^i_{1,u,0} = i$.*

   This comes from the fact that the double intertwining relation is given by the main result Theorem 6.2.2 under $\mathbf{H_1}$ and by its analogous Theorem 6.2.5 under $\mathbf{H_2}$.

5. Stein's factors related to the different distances compare between each other through the inequalities:

$$\sup_{f=\mathbf{1}_{[0,m]},\, m\in\mathbb{N}} \|\partial^k g_f\|_\infty \leq \sup_{0\leq f\leq 1} \|\partial^k g_f\|_\infty \leq \frac{1}{\inf_{\mathbb{N}} u} \sup_{f\in \mathrm{Lip}(d_u)} \|\partial^k g_f\|_\infty, \qquad k\in\mathbb{N}.$$

We now state the main results of this section, formulated for each distance of interest.

Approximation in total variation distance.

Theorem 6.3.2 (First Stein's factor for bounded functions)**.** *Assume that V_u is bounded from below by some positive constant $\sigma(u)$. Then, we have:*

$$\sup_{0\leq f\leq 1} \|g_f\|_\infty \leq \int_0^\infty e^{-\sigma(u)t} \sup_{i\in\mathbb{N}} \mathbb{P}(X^i_{u,t} = i) dt.$$

This theorem is applied to the negative binomial approximation in Proposition 6.4.5.

Theorem 6.3.3 (Second Stein's factor for bounded functions I)**.** *Under* $\mathbf{H_1}$*,*

$$\sup_{0\le f\le 1} \|\partial g_f\|_\infty \;\le\; 2\int_0^\infty e^{-\sigma(1,*u)t} \sup_{i\in\mathbb{N}^*} \mathbb{P}(X^i_{1,*u,t} = i)dt.$$

If the sequence is chosen to be $u = 1$*, we have:*

$$\begin{aligned}\sup_{0\le f\le 1} \|\partial g_f\|_\infty \;\le\; & \int_0^\infty e^{-\sigma(1,*u)t} \sup_{i\in\mathbb{N}^*} \big(\mathbb{P}(X^i_{1,*u,t} = i) - \mathbb{P}(X^i_{1,*u,t} = i-1)\\ & +\mathbb{P}(X^i_{1,*u,t} = i) - \mathbb{P}(X^i_{1,*u,t} = i+1)\big)\, dt.\end{aligned}$$

The analogue of Theorem 6.3.3 under the alternative set of hypotheses reads as:

Theorem 6.3.4 (Second Stein's factor for bounded functions II)**.** *Under* $\mathbf{H_2}$*,*

$$\sup_{0\le f\le 1} \|\partial g_f\|_\infty \;\le\; 2\int_0^\infty e^{-\sigma(1,u)t} \sup_{i\in\mathbb{N}} \mathbb{P}(X^i_{1,u,t} = i)dt.$$

If the sequence is chosen to be $u = 1$*, we have:*

$$\begin{aligned}\sup_{0\le f\le 1} \|\partial g_f\|_\infty \;\le\; & \int_0^\infty e^{-\sigma(1,u)t} \sup_{i\in\mathbb{N}} \big(\mathbb{P}(X^i_{1,u,t} = i) - \mathbb{P}(X^i_{1,u,t} = i-1)\\ & +\mathbb{P}(X^i_{1,u,t} = i) - \mathbb{P}(X^i_{1,u,t} = i+1)\big)\, dt.\end{aligned}$$

Remark 6.3.5 (Alternative versions)**.** *By the same techniques, it is possible to upper bound the quantities*

$$\sup_{0\le f/u\le 1} \|g_f\|_\infty, \qquad\qquad \sup_{0\le f/u\le 1} \|\partial_u g_f\|_\infty.$$

It could be useful if one is interested in the approximation in V*-norm (Meyn and Tweedie [1993]) rather than in total variation distance.*

Approximation in Wasserstein distance.

Theorem 6.3.6 (First Stein's factor for Lipschitz functions)**.** *If* V_u *is bounded from below by some positive constant* $\sigma(u)$*, then we have:*

$$\sup_{f\in Lip(d_u)} \|\overrightarrow{g_f}/u\|_\infty \le \frac{1}{\sigma(u)}.$$

Moreover, if V_u *is constant, then the preceding inequality is in fact an equality.*

Theorem 6.3.7 (Second Stein's factor for Lipschitz functions I)**.** *Under* $\mathbf{H_1}$*,*

$$\sup_{f\in Lip(d_u)} \|\partial_u g_f\|_\infty \;\le\; \frac{1}{\sigma(1,*u)} \sup_{x\in\mathbb{N}^*} \left(1 + \frac{u(x-1)}{u(x)}\right).$$

If we assume that $u(x) = q^x$ *on* $\mathbb{N}$ *with* $q \ge 1$*, then it stands that:*

$$\sup_{f\in Lip(d_u)} \|\partial_u g_f\|_\infty \;\le\; \int_0^\infty e^{-\sigma(1,*u)t} \left(1 - \frac{1}{q} + 2\frac{1}{q} \sup_{i\in\mathbb{N}^*} \mathbb{P}(X^i_{1,*u,t} = i)\right) dt.$$

An instance of Theorems 6.3.6 and 6.3.7 in the context of geometric approximation is given by Proposition 6.4.7. The following theorem is the analogue of Theorem 6.3.7 under the alternative set of hypotheses.

Theorem 6.3.8 (Second Stein's factor for Lipschitz functions II). *Under* $\mathbf{H_2}$,

$$\sup_{f\in Lip(d_u)} \|\frac{1}{u}\overrightarrow{\partial g_f}\|_\infty \quad \leq \quad \frac{1}{\sigma(1,u)} \sup_{x\in\mathbb{N}} \left(1+\frac{u(x+1)}{u(x)}\right).$$

If the sequence is chosen to be $u(x)=q^x$ *on* $\mathbb{N}$ *with* $q\geq 1$*, then the following result holds:*

$$\sup_{f\in Lip(d_u)} \|\frac{1}{u}\overrightarrow{\partial g_f}\|_\infty \quad \leq \quad \int_0^\infty e^{-\sigma(1,u)t}\left(q-1+2\sup_{i\in\mathbb{N}}\mathbb{P}(X^i_{1,u,t}=i)\right)dt.$$

As an illustration of this theorem, we derive Proposition 6.4.4 in the case of negative binomial approximation.

Approximation in Kolmogorov distance.
The first theorem indicates that the inequality

$$\sup_{1_{[0,m]},\, m\in\mathbb{N}} \|g_f\|_\infty \leq \sup_{0\leq f\leq 1} \|g_f\|_\infty$$

is actually an equality. This comes from the fact that the function achieving the argmax of the pointwise factor for bounded functions is actually of the form $f=1_{[0,m]}$. As a consequence, our upper bounds for the first Stein factor are identical for the approximation in total variation and Kolmogorov distances.

Theorem 6.3.9 (First Stein's factor for indicator functions). *If* V_u *is bounded from below and* $\inf_{\mathbb{N}} V_u=\sigma(u)$*, it stands that:*

$$\sup_{1_{[0,m]},\, m\in\mathbb{N}} \|g_f\|_\infty = \sup_{0\leq f\leq 1} \|g_f\|_\infty \leq \int_0^\infty e^{-\sigma(u)t}\sup_{i\in\mathbb{N}}\mathbb{P}(X^i_{u,t}=i)dt.$$

The two following theorems deal with the second Stein factor under the two set of hypotheses.

Theorem 6.3.10 (Second Stein's factor for indicator functions I). *Under* $\mathbf{H_1}$,

$$\sup_{f=\mathbf{1}_{[0,m]},\, m\in\mathbb{N}} \|\partial g_f\|_\infty \quad \leq \quad \int_0^\infty e^{-\sigma(1,*u)t}\sup_{i\in\mathbb{N}}\mathbb{P}(X^i_{1,*u,t}=i)dt.$$

If the sequence is chosen to be $u=1$*, we have:*

$$\sup_{f=\mathbf{1}_{[0,m]},\, m\in\mathbb{N}} \|\partial g_f\|_\infty \quad \leq \quad \int_0^\infty e^{-\sigma(1,*u)t}\sup_{i\in\mathbb{N}^*}|\mathbb{P}(X^i_{1,*u,t}=i)-\mathbb{P}(X^i_{1,*u,t}=i-1)|dt.$$

Comparing this second bound with the second bound obtained in Theorem 6.3.3 for the total variation approximation, one notices the fact that the second Stein factor for the total variation approximation involves the second derivative of the function $f_i(x)=\mathbb{P}(X^i_{1,*u,t}=x)$, whereas the second Stein factor for the Kolmogorov approximation involves only its first derivative.
Finally, we state the analogous of Theorem 6.3.10 under the alternative set of hypotheses.

Theorem 6.3.11 (Second Stein's factor for indicator functions II). *Under* $\mathbf{H_2}$,

$$\sup_{f=\mathbf{1}_{[0,m]},\, m\in\mathbb{N}} \|\partial g_f\|_\infty \quad \leq \quad \int_0^\infty e^{-\sigma(1,u)t}\sup_{i\in\mathbb{N}}\mathbb{P}(X^i_{1,u,t}=i)dt.$$

If the sequence is chosen to be $u=1$*, we have:*

$$\sup_{0\leq f\leq 1} \|\partial g_f\|_\infty \quad \leq \quad \int_0^\infty e^{-\sigma(1,u)t}\sup_{i\in\mathbb{N}}|\mathbb{P}(X^i_{1,u,t}=i)-\mathbb{P}(X^i_{1,u,t}=i-1)|dt.$$

6.3.4 Stein's method and mixture of distributions

As another part of our work within the context of Stein's method, we present in the current section theoretical error bounds for the approximation of mixture of distributions. This section is independent from our study of Stein's factors contained in Section 6.3.3. Results from both sections are combined in Section 6.4 and applied to Poisson and geometric mixture approximation.

Let φ be a non-negative function on $\mathbb{N}$ such that $\varphi(0) = 0$. For $\lambda > 0$, we denote by $\mathcal{I}_\varphi(\lambda)$ the probability distribution on $\mathbb{N}$ whose Stein's operator is

$$S_\lambda g(x) = \lambda g(x+1) - \varphi(x)g(x), \qquad x \in \mathbb{N}.$$

By letting φ vary, one finds back for $\mathcal{I}_\varphi(\lambda)$ every probability distribution supported on $\mathbb{N}$. In particular, the choice $\varphi(x) = x$ gives the Poisson law and is studied in Barbour et al. [1992]. The choice $\varphi(x) = r + x$ and $\varphi(x) = 1$ leads respectively to the binomial negative and geometric laws. A less classical example is $\varphi(x) = x^2$, for which $\mathcal{I}_\varphi(\lambda)$ is a distribution with pointwise probabilities $C_\lambda \lambda^x/(x!)^2$ for $x \in \mathbb{N}$ (C_λ is the renormalizing constant).

The first theorem of this section reads as follows.

Theorem 6.3.12 (Closeness of two $\mathcal{I}_\varphi(\lambda)$ distributions)**.** *Set $\lambda, \lambda' > 0$. We have:*

$$d_{\mathcal{F}}(\mathcal{I}_\varphi(\lambda'), \mathcal{I}_\varphi(\lambda)) \le |\lambda - \lambda'| \sup_{f\in\mathcal{F}} \|g_{\lambda,f}\|_\infty,$$

where g_f is the solution of the Stein's equation $S_\lambda g_f = f - \int f d\mathcal{I}_\varphi(\lambda)$. More generally, for any positive sequence u, if $X \sim \mathcal{I}_\varphi(\lambda)$ and $X' \sim \mathcal{I}_\varphi(\lambda')$, then:

$$d_{\mathcal{F}}(\mathcal{I}_\varphi(\lambda'), \mathcal{I}_\varphi(\lambda)) \le |\lambda - \lambda'| \sup_{f\in\mathcal{F}} \|g_f/u\|_\infty \mathbb{E}[u(X'+1)].$$

Proof. By the usual Stein error bound (6.12),

$$d_{\mathcal{F}}(\mathcal{I}_\varphi(\lambda'), \mathcal{I}_\varphi(\lambda)) = \sup_{f\in\mathcal{F}} \left| \int f d\mathcal{I}_\varphi(\lambda') - \int f d\mathcal{I}_\varphi(\lambda) \right| = \sup_{f\in\mathcal{F}} \left| \mathbb{E}[S_\lambda g_f(X')] \right|,$$

where $X' \sim \mathcal{I}_\varphi(\lambda')$. We know that $\mathbb{E}[S_{\lambda'} g_f(X')] = 0$; this yields:

$$\begin{aligned} |\mathbb{E}[S_\lambda g_f(X')]| &= |\mathbb{E}\left[(\lambda-\lambda')g_f(X'+1)\right]| = |(\lambda-\lambda')\mathbb{E}\left[u(X'+1)g_f(X'+1)/u(X'+1)\right]| \\ &\le |\lambda-\lambda'| \|g_f/u\|_\infty \mathbb{E}\left[u(X'+1)\right]. \end{aligned}$$

□

Note that the right hand side of both inequalities stated in Theorem 6.3.12 is not symmetric in (λ, λ') due to the dependence of g_f on λ and one can slightly improve it by taking the minimum over the symmetrized form. The first inequality corresponds to the constant sequence $u = 1$.

Let W be a mixture of law $\mathcal{I}_\varphi(\lambda)$; namely there exists a random variable Λ on $\mathbb{R}^+$ such that

$$\mathcal{L}(W \mid \Lambda) = \mathcal{I}_\varphi(\Lambda).$$

(Recall that $\mathcal{L}(W)$ denotes the distribution of the random variable W.) A consequence of Theorem 6.3.12 is the following corollary.

Corollary 6.3.13 (Biased approximation of mixed $\mathcal{I}_\varphi(\lambda)$ laws)**.** *With the preceding notation, we have:*

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \le \mathbb{E}[|\lambda - \Lambda|] \sup_{f\in\mathcal{F}} \|g_f\|_\infty.$$

Proof. Indeed,

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \mathbb{E}[d_{\mathcal{F}}(\mathcal{L}(W|\Lambda), \mathcal{I}_\varphi(\lambda))] \leq \mathbb{E}[|\lambda - \Lambda|] \sup_{f\in\mathcal{F}} \|g_f\|_\infty.$$

□

However, one actually has the following better bound using the mixture property of W:

Theorem 6.3.14 (Unbiased approximation of mixed $\mathcal{I}_\varphi(\lambda)$ distributions)**.** *For every positive sequence u, letting $\lambda = E[\Lambda]$, we have:*

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \sup_{f\in\mathcal{F}} \|\partial_u g_f\|_\infty \sup_{f\in Lip(d_{\vec{u}})} \|g_f\|_\infty \, Var(\Lambda).$$

More generally, the following upper bound holds for all positive sequences u, v:

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \sup_{f\in\mathcal{F}} \|\partial_u g_f\|_\infty \sup_{r\in Lip(d_{\vec{u}})} \|g_r/v\|_\infty \, \mathbb{E}[|\lambda - \Lambda|^2 \mathbb{E}[v(W+1) \mid \Lambda]].$$

Proof of Theorem 6.3.14. For every real-valued function g on $\mathbb{N}$, $\mathbb{E}[S_\Lambda(g)(W)|\Lambda] = 0$. Hence, by taking $g = g_f$ the solution to Stein's equation associated with any fixed function $f : \mathbb{N} \to \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[S_\lambda g_f(W)] &= \mathbb{E}[\mathbb{E}[(S_\lambda - S_\Lambda)g_f(W)|\Lambda]] = \mathbb{E}[(\lambda - \Lambda)g_f(W+1)] \\
&= \mathbb{E}[(\lambda - \Lambda)(g_f(W+1) - g_f(Z+1))],
\end{aligned}$$

where $Z \sim \mathcal{I}_\varphi(\lambda)$. For two random variables Z, Z' on $\mathbb{N}$, by the Kantorovich-Rubinstein theorem recalled at the beginning of this section,

$$\begin{aligned}
\big|\mathbb{E}[g(Z'+1) - g(Z+1)]\big| &\leq \|\partial_u g\|_\infty W_{d_u}(\mathcal{L}(Z'+1), \mathcal{L}(Z+1)) = \|\partial_u g\|_\infty \inf \mathbb{E}[d_u(Z'+1, Z+1)] \\
&= \|\partial_u g\|_\infty W_{d_{\vec{u}}}(\mathcal{L}(Z'), \mathcal{L}(Z)),
\end{aligned}$$

where the infimum is taken on the set of couplings with first marginal $\mathcal{L}(Z)$ and second marginal $\mathcal{L}(Z')$. Now, by Theorem 6.3.12,

$$\begin{aligned}
\mathbb{E}[S_\lambda g_f(W)] &\leq \|\partial_u g\|_\infty \mathbb{E}[|\lambda - \Lambda| \mathbb{E}[W_{d_{\vec{u}}}(\mathcal{L}(W), \mathcal{L}(Z))|\Lambda] \\
&\leq \|\partial_u g\|_\infty \sup_{r\in \mathrm{Lip}(d_{\vec{u}})} \|g_r/v\|_\infty \mathbb{E}[(\lambda - \Lambda)^2 \mathbb{E}[v(W+1)|\Lambda]].
\end{aligned}$$

As in Theorem 6.3.12, the first inequality is an instance of the second one in the case $v = 1$. □

Remark 6.3.15 (Alternative bound via coupling)**.** *In the previous proof, we used Theorem 6.3.12 in order to bound $W_{d_{\vec{u}}}(\mathcal{I}_\varphi(\Lambda), \mathcal{I}_\varphi(\lambda))$. It is also possible to bound this distance via another method (for instance a coupling argument) instead of using a bound on Stein's solution.*

6.4 Examples

In this section we illustrate our results on some examples. The classical examples of the M/M/∞ and M/M/1 process come from the queueing theory. We also apply the results to the Galton-Watson process with immigration. Other explicit examples of birth-death processes for which a "good choice" of sequence u is known are given in [Chen, 2004, Table 9.1 p. 351] and in Chen [1996]. For the sake of conciseness we defer the proof of Lemma 6.7.1 about the pointwise probabilities of the M/M/∞ queue to Section 6.7.

6.4.1 The M/M/∞ process and the Poisson approximation

Let $(X_t)_{t\geq 0}$ be a BDP with constant birth death λ and linear death rate $x \mapsto x$. Its invariant measure is the Poisson law $\mathcal{P}_\lambda$. Let us set $u = v = 1$ on $\mathbb{N}$. By application of Theorem 6.2.1 we find that $V_1 = 1$ and that $(P_{1,t})_{t\geq 0} = (P_t)_{t\geq 0}$. Applying Theorem 6.2.5 (or re-applying Theorem 6.2.1) yields $V_{1,1} = 2$ and $(P_{1,1,t})_{t\geq 0} = (P_t)_{t\geq 0}$. By a straightforward induction, for all positive or bounded functions $f : \mathbb{N} \to \mathbb{R}$,

$$\partial^k P_t f = e^{-kt} P_t \partial^k f, \qquad t \geq 0, \quad k \in \mathbb{N}. \tag{6.15}$$

Combined with Theorem 6.2.8 and Remark 6.2.9, the equation (6.15) implies the following contraction in Zolotarev-type distance: for all $\mu \in \mathcal{P}$,

$$\sup_{\|\partial^k f\|_\infty \leq 1} |\mu(P_t f) - \mathcal{P}_\lambda(f)| \leq e^{-kt} \sup_{\|\partial^k f\|_\infty \leq 1} |\mu(f) - \mathcal{P}_\lambda(f)|, \qquad k \in \mathbb{N}^*.$$

Formula (6.15) is already known and often proved using Mehler's formula which reads, for any bounded function f, as:

$$P_t f(x) = \mathbb{E}[f(X_t^0 + B_t)], \qquad x \in \mathbb{N}, \quad t \geq 0, \tag{6.16}$$

where $(X_t^0)_{t\geq 0}$ is a M/M/∞ process starting from 0 and B_t is an independent random variable distributed as a binomial random variable with parameters (x, e^{-t}). It is also known that X_t^0 is distributed as a Poisson distribution with parameter $\lambda(1 - e^{-t})$ at every time $t \geq 0$. Conversely, the proof of the formula (6.16) can be deduced from Theorem 6.2.1 with similar (but simpler) arguments than those developed in Lemma 6.4.2 below.

We now turn to the subject of Poisson approximation and the associated Stein factors. Let g_f be the solution to Stein's equation (6.11) with Stein's operator $Sf(x) = \lambda f(x+1) - xf(x)$. The target measure is the Poisson distribution $\mathcal{P}_\lambda$. The following lemma allows to estimate from above the pointwise probabilities of the process $(X_t)_{t\geq 0}$.

Lemma 6.4.1 (Upper bounds of the instantaneous probabilites of the M/M/∞ queue)**.** *Let $(X_t^x)_{t\geq 0}$ be a BDP$(\lambda, x)_{x\in\mathbb{N}}$. For all $x \in \mathbb{N}$ and $t \geq 0$,*

$$\sup_{x\in\mathbb{N}} \mathbb{P}(X_t^x = x) \leq 1 \wedge \frac{c}{\sqrt{\lambda(1-e^{-t})}}, \qquad c = \frac{1}{\sqrt{2e}},$$
$$\sup_{x\in\mathbb{N}^*} |\mathbb{P}(X_t^x = x) - \mathbb{P}(X_t^x = x-1)| \leq 1 \wedge \frac{C}{\lambda(1-e^{-t})}, \qquad C = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{\sqrt{2}}} \leq 1.$$

The first upper bound is very classical, it derives from Mehler's formula (6.16) and an upper bound on the pointwise probabilitites of the Poisson distribution ([Barbour et al., 1992, Proposition A.2.7]). The second one is new and is proved in Section 6.7, since it is rather technical and can be omitted at first reading.

By applying Theorems 6.3.2, 6.3.6, 6.3.8 jointly with the first bound of Lemma 6.4.1, one finds back (and by the same techniques) the following upper bounds (Barbour and Brown [1992], Barbour and Xia [2006]):

$$\sup_{0\leq f\leq 1} \|g_f\|_\infty \leq 1 \wedge \sqrt{\frac{2}{\lambda e}}, \qquad \sup_{f\in \mathrm{Lip}(d_1)} \|g_f\|_\infty = 1, \qquad \sup_{f\in \mathrm{Lip}(d_1)} \|\partial g_f\|_\infty \leq 1 \wedge \frac{8}{3\sqrt{2e\lambda}}.$$

Of course, one may want to derive other known Stein's factors for Poisson approximation by our techniques, as for instance the second Stein factor for approximation in the total variation distance with rate $1 \wedge (1/\lambda)$ (Barbour et al. [1992]). However, when applying Theorem 6.3.4 with the second bound of Lemma 6.4.1, the non-integrability in 0 of the term $1/(1-e^{-t})$ leads to sub-optimal results (namely, after some careful computations, we recover the known rate, up to a multiplicative factor $\log \lambda$).

Let us now combine the Stein bounds with our results on the mixture of distributions. If $\varphi(x) = x$ then $\mathcal{I}_\varphi(\lambda) = \mathcal{P}_\lambda$. In particular, Theorem 6.3.12 and the preceding bounds give

$$d_{\mathrm{TV}}(\mathcal{P}(\lambda), \mathcal{P}(\lambda')) \leq \frac{1}{1 \wedge \sqrt{\lambda \vee \lambda'}} |\lambda - \lambda'|, \qquad W(\mathcal{P}_\lambda, \mathcal{P}_{\lambda'}) \leq |\lambda - \lambda'|.$$

The first bound is (almost) the result of [Barbour et al., 1992, Theorem 1.C p. 12]. The second one is in fact an equality and can also be proved via a coupling approach (Lindvall [2002]). Theorem 6.3.14 yields

$$W(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \left(1 \wedge \frac{8}{3\sqrt{2e\lambda}}\right) \mathrm{Var}(\Lambda), \qquad d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \frac{1}{\lambda} \mathrm{Var}(\Lambda).$$

While the second bound is exactly the same as in [Barbour et al., 1992, Theorem 1.C p. 12], the bound in Wasserstein distance seems to be new. Let us see an instance of it. We denote by $\mathrm{NB}(r,p)$ the negative binomial distribution of parameters (r,p), i.e.,

$$\mathrm{NB}(r,p)(x) = \frac{\Gamma(r+x)}{\Gamma(r)x!}(1-p)^r p^x, \qquad x \in \mathbb{N},$$

where Γ denotes the usual Γ function. The negative binomial law is a mixed Poisson distribution with Λ distributed as a Gamma law with parameters r and $\frac{1-p}{p}$. Consequently, we obtain the following result on the distance between the negative binomial distribution and the Poisson distribution, which serves as the reference measure:

$$W\left(\mathrm{NB}(r,p), \mathcal{P}_{r(1-p)/p}\right) \leq \frac{8}{3\sqrt{2e}} \sqrt{\frac{r(1-p)}{p}} \frac{(1-p)}{p},$$

which is the upper bound announced in the introduction. A similar approximation in total variation distance holds. Although the convergence of the binomial negative distribution towards a Poisson law in the regime $p \to 1$, $r \to \infty$ and $r(1-p)/p \to c$ for a positive constant c is a well-known fact, the preceding upper bound seems to be the first attempt to quantify this convergence.

6.4.2 The GWI process and the negative binomial approximation

We consider the BDP with rates $\alpha(x) = p(r+x)$, $\beta(x) = x$ on $\mathbb{N}$ with $r > 0$ and $0 < p < 1$. The coefficient pr can be interpreted as a rate of immigration, while the birth rate per capita is p and the death rate per capita is 1. Without the immigration procedure, this is a Galton-Watson process whose individuals have only one descendant (or simply a linear birth-death process). The invariant measure of this process is the negative binomial distribution $\mathrm{NB}(r,p)$ just defined. Remark that for the particular choice $r = 1$ it is nothing else than the geometric law of parameter p. If X is a $\mathrm{NB}(r,p)$ random variable then $X + r$ follows the so-called Pascal distribution; it represents the number of successes in a sequence of independent and identically distributed Bernoulli trials (with parameter p) before r failures when r is a positive integer.
Let us take $u = v = 1$ on $\mathbb{N}$. Theorem 6.2.1 shows that:

$$\partial P_t = P_{1,t}^{V_1}, \qquad t \geq 0,$$

where $(P_{1,t})_{t \geq 0}$ is a birth-death process with rates defined as

$$\alpha_1(x) = p(r+1+x), \qquad \beta_1(x) = x, \qquad x \in \mathbb{N}.$$

It is again a Galton-Watson process with immigration. The birth and death rates are unchanged and the immigration rate is increased by p. The potential V_1 is constant and takes the value $V_1 = 1-p$. By Theorem 6.2.5, we find that $(P_t)_{t\geq 0}$ and ∂^2 are intertwined via the Feynman-Kac semigroup composed of a birth-death semigroup with rates $(\alpha_{1,1}, \beta_{1,1})$ and of potential $V_{1,1}$, with:

$$\alpha_{1,1}(x) = p(r+2+x), \qquad \beta_{1,1}(x) = x, \qquad V_{1,1}(x) = 2(1-p), \qquad x \in \mathbb{N}.$$

Let us call $(P_{k,t})_{t\geq 0}$ the semigroup associated to a BDP with rates $(p(r+k+x), x)$ on $\mathbb{N}$. By a straightforward induction, for all positive or bounded functions $f : \mathbb{N} \to \mathbb{R}$, the following intertwining relation holds:

$$\partial^k P_t f = e^{-(1-p)kt} P_{k,t}\, \partial^k f, \qquad t \geq 0, \quad k \in \mathbb{N}. \tag{6.17}$$

As indicated in Remark 6.2.9, the previous equality gives the following improvement of Theorem 6.2.8: for every $\mu \in \mathcal{P}$,

$$\sup_{\|\partial^k f\|_\infty \leq 1} |\mu(P_t f) - \pi(f)| \leq e^{-(1-p)kt} \sup_{\|\partial^k f\|_\infty \leq 1} |\mu(f) - \pi(f)|\,.$$

Another consequence of the formula (6.17) is the invariance of polynomials under the action of $(P_t)_{t\geq 0}$: if Q is a polynomial of degree k, then for all $t \geq 0$, $\partial^k P_t Q$ is constant, hence $P_t Q$ is still a polynomial of degree k. This property also holds for the M/M/∞ process.

Intertwining relations can be seen in certain cases as consequences of Mehler-type formulas. Here, conversely, we are able to derive a Mehler-type formula from the first order intertwining relation. To our knowldedge, this formula is new, though another Mehler-type formula is proved in Barbour et al. [2015].

Lemma 6.4.2 (A Mehler's formula for the Galton-Watson process with immigration)**.** *Set $0 < p < 1$, $s > 0$ and $q = 1-p$. For all $x \in \mathbb{N}$ let $(Y_t^x)_{t\geq 0}$ be a birth-death process starting at x and with rates $(p(s+k), k)_{k\in\mathbb{N}}$. Let W_t be a random variable following the Poisson distribution $\mathcal{P}(p(1-e^{-t}))$ and define the sequence $(w(k))_{k\in\mathbb{N}}$ as*

$$w(0) = 1 - e^{-qt}\mathbb{P}(W_t = 0) \quad \textit{and} \quad \forall k \in \mathbb{N}^*, \quad w(k) = e^{-qt}(\mathbb{P}(W_t = k-1) - \mathbb{P}(W_t = k)).$$

For all $t \geq 0$, let the random variables $(Z_{i,t})_{i\in\mathbb{N}}$ be independent, identically distributed and independent of Y_t^0, with distribution given by the pointwise probabilities $(w(k))_{k\in\mathbb{N}}$. Then we have the equality in law

$$Y_t^x \;=\; Y_t^0 + \sum_{i=1}^{x} Z_{i,t}.$$

Proof. This proof is a corollary of the intertwining formula (6.17) for $k=1$. Indeed, for every bounded real-valued function on $\mathbb{N}$, Theorem 6.2.1 implies that

$$\mathbb{E}[f(Y_t^{x+1})] = \mathbb{E}[f(Y_t^x)] + e^{-qt}\mathbb{E}[f(\widetilde{Y}_t^x + 1) - f(\widetilde{Y}_t^x)], \qquad x \in \mathbb{N}, \quad t \geq 0,$$

where $(\widetilde{Y}_t^x)_{t\geq 0}$ is a BDP $(p(s+1+k), k)_{k\in\mathbb{N}}$. We notice that $(\widetilde{Y}_t^x)_{t\geq 0} = (Y_t^x + W_t)_{t\geq 0}$, where $(W_t)_{t\geq 0}$ is a birth-death process independent of $(Y_t^x)_{t\geq 0}$ with rates $(p, k)_{k\in\mathbb{N}}$ such that $W_0 = 0$. The process $(W_t)_{t\geq 0}$ is a M/M/∞ queue starting from 0 at time 0. It is distributed as a Poisson law $\mathcal{P}_{\lambda_t}$, $\lambda_t = p(1-e^{-t})$ at all times $t \geq 0$. We use below the observation that as $\lambda_t < 1$ for all

$t \geq 0$, the sequence $(\mathbb{P}(W_t = k))_{k\in\mathbb{N}}$ is non-increasing on $\mathbb{N}$. We have:

$$\begin{aligned}
\mathbb{E}[f(Y_t^{x+1})] &= \mathbb{E}[f(Y_t^x)] + e^{-qt}\sum_{k=0}^{\infty}\mathbb{P}(W_t = k)\mathbb{E}[f(Y_t^x + k + 1) - f(Y_t^x + k)] \\
&= (1 - e^{-qt}\mathbb{P}(W_t = 0))\mathbb{E}[f(Y_t^x)] + e^{-qt}\sum_{k=1}^{\infty}(\mathbb{P}(W_t = k-1) - \mathbb{P}(W_t = k))\mathbb{E}[f(Y_t^x + k)] \\
&= \sum_{k=0}^{\infty} w(k)\mathbb{E}[f(Y_t^x + k)],
\end{aligned}$$

where the sequence $(w(k))_{k\in\mathbb{N}}$ is defined in the statement of the lemma. It is easy to check that $\sum_{k=0}^{\infty} w(k) = 1$, and that the sequence $(w(k))_{k\in\mathbb{N}}$ is non-negative thanks to the observation above. For all $t \geq 0$, we define a random variable S_t such that S_t is independent of $(Y_t^x)_{t\geq 0}$ and that for all non-negative integer $\mathbb{P}(S_t = k) = w(k)$. This yields the equality in law $Y_t^{x+1} = Y_t^x + S_t$. The lemma follows by induction. □

Let us turn to the study of Stein's factors associated to the negative binomial approximation. We begin by a lemma on the instantaneous probabilities of a Galton-Watson process with immigration.

Lemma 6.4.3 (Upper bound of the instantaneous probabilities of a GWI process)**.** *Set $(X_t^x)_{t\geq 0}$ be a BDP$(p(r+k), k)_{k\in\mathbb{N}}$. We have:*

$$\sup_{x\in\mathbb{N}}\mathbb{P}(X_t^x = x) \leq e^{-(1-p)t}\left(1 \wedge \frac{1}{p(1-e^{-t})}\right), \qquad t \geq 0, \quad p \in (0,1), \quad r > 0,$$

$$\sup_{x\in\mathbb{N}}\mathbb{P}(X_t^x = x) \leq \frac{1}{\sqrt{2e}}\left(\frac{1-p}{p\,(1-e^{-(1-p)t})}\right)^{1/2}\frac{K(r)}{\sqrt{r}}, \qquad t \geq 0, \quad p \in (0,1), \quad r > \frac{1}{2},$$

with $K(r) = \sqrt{r}\,\Gamma(r-1/2)/\Gamma(r)$.

The first upper bound is proved now, as a consequence of Lemma 6.7.1 and Lemma 6.4.2. The second upper bound is already known (Barbour et al. [2015]) and is based on results of Kendall [1948] and Phillips [1996].

Proof of Lemma 6.4.3. By Lemma 6.4.2,

$$\sup_{x\in\mathbb{N}}\mathbb{P}(X_t^x = x) \leq \sup_{x\in\mathbb{N}}\mathbb{P}(Z_{1,t} = x) \wedge \sup_{x\in\mathbb{N}}\mathbb{P}(X_t^0 = x), \qquad t \geq 0,$$

where $Z_{1,t}$ is defined in Lemma 6.4.2. According to this lemma and to Lemma 6.7.1,

$$\begin{aligned}
\mathbb{P}(Z_{1,t} = x) &\leq e^{-(1-p)t}(\mathbb{P}(W_t = k-1) - \mathbb{P}(W_t = k)) \\
&\leq e^{-(1-p)t}\left(1 \wedge \frac{C}{p(1-e^{-t})}\right), \qquad x \in \mathbb{N}, \quad t \geq 0.
\end{aligned}$$

On the other hand, by a result of Kendall [1948], cited as Lemma 2.2 in Barbour et al. [2015], it is known that for all $t \geq 0$, X_t^0 is distributed as a negative binomial distribution of parameters $(r, \theta_t(p))$, with

$$\theta_t(p) = 1 - \frac{1-p}{1-pe^{-(1-p)t}}.$$

Now Phillips [1996] shows that when X is distributed as a negative binomial distribution with parameters (r, θ), and if $r > \frac{1}{2}$, then

$$\sup_{k\in\mathbb{N}}\mathbb{P}(X = k) \leq \frac{1}{\sqrt{2e}}\sqrt{\frac{1-\theta}{\theta}}\frac{K(r)}{\sqrt{r}}.$$

□

For the Stein factor associated with Lipschitz function, Theorem 6.3.6 and equation (6.17) yield

$$\sup_{f\in\mathrm{Lip}(d)} \|g_f\|_\infty = \frac{1}{\sigma(1)} = \frac{1}{1-p},$$

recovering [Barbour et al., 2015, Theorem 1.1, equation (1.3)].
The following proposition on the second Stein factor associated to Lipschitz function improves on the known upper bounds.

Proposition 6.4.4 (Estimation of the second Stein's factor for Lipschitz function and NB-approximation)**.** *Let $r > 0$ and $0 < p < 1$. For a real-valued function f on $\mathbb{N}$, let g_f be the (centered) solution to Stein's equation*

$$p(r+x)\,\partial g_f(x) \,+\, x\,\partial^* g_f(x) = f(x) - \int f dNB(r,p), \quad x\in\mathbb{N}.$$

Then,

$$\sup_{f\in Lip(d)} \|\partial g_f\|_\infty \;\le\; \min\left\{\frac{2}{3(1-p)}, \frac{D}{\sqrt{(r+2)p(1-p)}}\right\}, \qquad D = 2\frac{\sqrt{\pi}}{3\sqrt{e}} \simeq 0.72.$$

Proof. By application of Theorem 6.6.8 and formula (6.17), we find that

$$\sup_{f\in\mathrm{Lip}(d)} \|\partial g_f\|_\infty = 2\int_0^\infty e^{-2(1-p)t} \sup_{i\in\mathbb{N}} \mathbb{P}(X^i_{1,1,t} = i)dt,$$

where $(X^i_{1,1,t})_{t\ge 0}$ is a BDP $(p(r+2+x), x)_{x\in\mathbb{N}}$. Applying Lemma 6.4.3,

$$\sup_{f\in\mathrm{Lip}(d)} \|\partial g_f\|_\infty \le 2\int_0^\infty e^{-3(1-p)t}dt \wedge 2\sqrt{\frac{1p}{p}}\frac{K(r+2}{\sqrt{r+2}}\int_0^\infty \sqrt{e^{-2(1-p)t}}\sqrt{1-e^{-(1-p)t}}dt.$$

The function K is decreasing on $(\frac{1}{2},\infty)$ and $\int_0^\infty \frac{e^{-2(1-p)t}}{\sqrt{1-e^{-(1-p)t}}}dt = \frac{4}{3(1-p)}$. Hence

$$\sup_{f\in\mathrm{Lip}(d)} \|\partial g_f\|_\infty \;\le\; \min\left\{\frac{2}{3(1-p)}, \frac{D}{\sqrt{(r+2)p(1-p)}}\right\},$$

with $D = \frac{8K(2)}{3\sqrt{2e}} = \frac{4\Gamma(3/2)}{3\sqrt{e}} = \frac{2\sqrt{\pi}}{3\sqrt{e}}$.

□

The Proposition 6.4.4 might be compared to [Barbour et al., 2015, Theorem 1.1, equation (1.4)], which states the inequality

$$\sup_{f\in\mathrm{Lip}(d)} \|\partial g_f\|_\infty \le \min\left\{\frac{2}{1-p}, \frac{1+p}{(1-p)^2}, \frac{1.5}{\sqrt{rp(1-p)^3}}\right\}. \tag{6.18}$$

We observe that:
— The numerical constant in front of $1/(1-p)$ is improved.
— As $\frac{D}{\sqrt{(r+2)p(1-p)}} \le \frac{0.8}{\sqrt{rp(1-p)}}$ and $0.8 \le \frac{1.5}{1-p}$, we have:

$$\frac{D}{\sqrt{(r+2)p(1-p)}} \le \frac{1.5}{\sqrt{rp(1-p)^3}}.$$

Note that the proofs are similar up to the formula

$$\sup_{f\in \mathrm{Lip}(d)} |\partial g_f(i)| = -\int_0^\infty \partial\partial^* P_t \mathbf{1}_i dt.$$

We then apply the second order intertwining formula, whereas Barbour et al. [2015] use another technique. In both cases, a bound of the type $\sup_i \mathbb{P}(Y_t^i = i)$ is needed, but not for the same process $(Y_t)_{t\geq 0}$.
Let us make two short remarks on the proof of Proposition 6.4.4: firstly, we do not use the upper bound

$$\sup_{x\in\mathbb{N}} \mathbb{P}(X_t^x = x) \leq e^{-(1-p)t}\frac{1}{p(1-e^{-t})},$$

that would make appear the hypergeometric function ${}_2F_1$. Secondly, the function K is bounded from below by a positive constant on $(\frac{1}{2},\infty)$, hence by writing $K(r+2)\leq K(2)$ we do not lose the rate in r.

For the Stein factor associated to bounded functions, at the order 1 we find the following result.

Proposition 6.4.5 (Estimation of the first Stein factor for bounded functions and NB-approximation)**.** *With the same assumptions as in Theorem 6.4.4, we have:*

$$\sup_{0\leq f\leq 1} \|g_f\|_\infty \quad\leq\quad \frac{1}{2(1-p)} \wedge \frac{\sqrt{\pi}}{\sqrt{(r+1)p(1-p)}}.$$

We do not detail the proof which is very similar to the one of Proposition 6.4.4.
This result improves on a result of [Brown and Phillips, 1999, Lemma 3] which states

$$\sup_{0\leq f\leq 1} \|g_f\|_\infty \leq \frac{1}{p\vee(1-p)\mathbf{1}_{r\geq 1}}.$$

We do not develop the case of the second Stein factor of bounded functions, where the upper bound given by Theorem 6.3.3 recovers the simple inequality

$$\sup_{0\leq f\leq 1} \|\partial g_f\|_\infty \leq \sup_{f\in \mathrm{Lip}(d_1)} \|\partial g_f\|_\infty.$$

Results about this factor can be found in [Brown and Xia, 2001, Theorem 2.10], in [Eichelsbacher and Reinert, 2008, example 2.12] for the case $r=1$, and in [Brown and Phillips, 1999, Lemma 5].

If $\varphi: x\mapsto r+x$, $r\in\mathbb{N}$ and $\lambda\in(0,1)$ then $\mathcal{I}_\varphi(\lambda)=NB(r,\lambda)$. The variable $W+r$ then represents the number of trials that are necessary to obtain r successes in a Bernoulli experiment with a random probability of gain.

To conclude this section, we observe that the Stein operator associated to a probability measure is not unique, and that resulting Stein's factors depend on the choice of the operator. When $r=1$, we recover the geometric law as the invariant distribution, similarly to the forthcoming example. This is the choice of Eichelsbacher and Reinert [2008] to study the geometric distribution. In the next section we choose another Stein's operator.

6.4.3 The M/M/1 process and the geometric approximation

Let $(X_t^x)_{t\geq 0}$ be a BDP(α,β) with rates $\alpha(x)=\alpha$, $\beta(x)=\beta\mathbf{1}_{x\in\mathbb{N}^*}$ on $\mathbb{N}$. We suppose that $\rho:=\frac{\alpha}{\beta}<1$. We denote by $(P_t)_{t\geq 0}$ the associated semigroup. Its invariant distribution is the geometric law $\mathcal{G}(\rho)$ with pointwise probabilities $p(k)=(1-\rho)\rho^k$ for $k\in\mathbb{N}$. Notice that this is the definition of the geometric law with support $\mathbb{N}$ and not $\mathbb{N}^*$. Let us choose $u(x)=r^x$, $v(x)=q^x$ for $x\in\mathbb{N}$ with $r>0$, $q\geq 1$. Theorem 6.2.1 gives rise to a Feynman-Kac semigroup composed of a birth-death semigroup $(P_{u,t})_{t\geq 0}$ with rates (α_u,β_u) and a potential V_u, which are defined as

$$\alpha_u(x)=r\alpha,\qquad \beta_u(x)=\frac{1}{r}\beta,\qquad V_u(x)=(1-r)\alpha+\left(1-\frac{1}{r}\mathbf{1}_{x\in\mathbb{N}^*}\right)\beta,\qquad x\in\mathbb{N}.$$

The semigroup $(P_{u,t}^{V_u})_{t\geq 0}$ is still a semigroup associated to a M/M/1 queue, only with modified rates. The potential V_u, while non-constant, is non-increasing on $\mathbb{N}$. By Theorem 6.2.2, we find a Feynman-Kac semigroup $(P_{u,*v,t}^{V_{u,*v}})_{t\geq 0}$ where $(P_{u,*v,t})_{t\geq 0}$ is again a semigroup corresponding to a M/M/1 queue. The rates and potential are defined on $\mathbb{N}$ as

$$\begin{aligned}
&\alpha_{u,*v}(x)=qr\alpha,\quad \beta_{u,*v}(x)=\frac{1}{qr}\beta\mathbf{1}_{x\in\mathbb{N}^*},\qquad x\in\mathbb{N},\\
&V_{u,*v}(x)=(1-qr)\alpha+\left(1-\frac{1}{qr}\right)\beta,\qquad x\in\mathbb{N}^*,\qquad V_{u,*v}(0)=\alpha-(1+q)r\alpha+\beta.
\end{aligned}$$

Remark that, in contrast with the general case of Theorem 6.2.2, the semigroup $(P_{u,*v,t})_{t\geq 0}$ is again a birth-death semigroup. This is due to the fact that V_u is constant on $\mathbb{N}^*$. The potential V_u is not constant on $\mathbb{N}$, which prevents us to apply Theorem 6.2.5, but it is almost constant which explains heuristically why we find again a birth-death process when applying Theorem 6.2.2.

Set $\sigma(u,*v)=\inf_{x\in\mathbb{N}}V_{u,*v}(x)=\min(V_{u,*v}(0),V_{u,*v}(1))$. A few calculations show that

$$\max\{\sigma(u,*v)\,|\,u(x)=r^x,\ v(x)=q^x,\ r>0,\ q\geq 1\}=(\sqrt{\beta}-\sqrt{\alpha})^2,$$

and the arg max is realized for all $r\leq\sqrt{\beta/\alpha}=\rho^{-1}$ and $q=\rho^{-1}/r$. This means that there is a range of choice for the parameters (r,q) allowing to recover the spectral gap $(\sqrt{\beta}-\sqrt{\alpha})^2$ of the process in the convergence of Theorem 6.2.8. However, contrary to the two preceding examples, notice that the second order intertwining does not allow to improve on the spectral gap and that the rate of convergence in the distance $\zeta_{u,*v}$ is the same as the rate of convergence in the Wasserstein distance W_{d_u} for the best choices of u,v.

This example is maybe the most important because, in contrast with the two previous processes, the M/M/1 queue is not known to satisfy a Mehler formula of the type (6.16), which would make it rather difficult to differentiate directly. A Mehler-like formula can nevertheless be deduced from Theorem 6.2.1: choosing $u=1$ in this theorem, we derive

$$\mathbb{E}[f(X_t^{x+1})-f(X_t^x)]=\mathbb{E}\left[e^{-\int_0^t V(X_s^x)ds}(f(X_t^x+1)-f(X_t^x))\right],$$

where $(X_t^x)_{t\geq 0}$ is M/M/1 process starting from x and $V(x)=\beta\mathbf{1}_{x=0}$. As a consequence, if B_t is a Bernoulli random variable verifying

$$\mathbb{P}(B_t=1\mid(X_s^x)_{s\leq t})=e^{-\int_0^t V(X_s^x)ds},\qquad t\geq 0,$$

then,

$$\mathbb{E}[f(X_t^{x+1})]=\mathbb{E}[(f(X_t^x+B_t)],\qquad t\geq 0,$$

and by induction there exists a random variable Y_t^x such that

$$\mathbb{E}[f(X_t^x)] = \mathbb{E}[(f(X_t^0 + Y_t^x)], \qquad t \geq 0.$$

This formula seems to be new (even if the instantaneous distribution of the M/M/1 process is known, see Baccelli and Massey [1989]). Unfortunately, the random variable Y_t^x is not independent from X_t^0 and this makes this formula less powerful than (6.16). This approach is generalizable for every BDP with constant birth rate (so that the processes $(X_{1,t})_{t\geq 0}$ and $(X_t)_{t\geq 0}$ have the same law).

As in the preceding examples, we state a lemma related to the instantaneous probabilities of the modified process before turning to the Stein factors for geometric approximation.

Lemma 6.4.6 (Upper bound of the instantaneous probabilities of a M/M/1 queue)**.** *Let $(Y_t)_{t\geq 0}$ be a M/M/1 queue with rates $(\lambda, \lambda \mathbf{1}_{\mathbb{N}^*})$. Then for all $t \geq 0$,*

$$\sup_{i\in\mathbb{N}^*} \mathbb{P}(Y_t^i = i) \leq \frac{1}{\sqrt{\lambda t}}.$$

Proof. Let us consider the BDP $(\widetilde{Y}_t)_{t\geq 0}$ with rates $(1, \mathbf{1}_{x\in\mathbb{N}^*})$. Then for all $t \geq 0$, the equality in law $Y_t = \widetilde{Y}_{\lambda t}$ holds, hence it is enough to prove that $\sup_{i\in\mathbb{N}^*} \mathbb{P}(\widetilde{Y}_t^i = i) \leq \frac{1}{\sqrt{t}}$. By [Abate et al., 1991, Corollary 1 (d)], the sequence $(\mathbb{P}(\widetilde{Y}_t^i = i))_{i\geq 0}$ is non-increasing for every $t \geq 0$. Hence $\sup_{i\in\mathbb{N}} \mathbb{P}(\widetilde{Y}_t^i = i) = \mathbb{P}(\widetilde{Y}_t^0 = 0)$. By [Abate et al., 1991, formula (9) and Corollary 2 (a)],

$$\mathbb{P}(\widetilde{Y}_t^0 = 0) = \sum_{j=1}^{\infty} \frac{j}{t} \mathbb{P}(Z_t^0 = j) = \frac{1}{t}\mathbb{E}[Z_t^0 \mathbf{1}_{Z_t^0 > 0}],$$

where $(Z_t^0)_{t\geq 0}$ is a birth-death process with constant birth rate 1 and constant death rate 1 on the whole integer line $\mathbb{Z}$; namely this is the continuous-time simple random-walk. This process can be represented as

$$\forall t \geq 0, \quad Z_t^0 = N_t^+ - N_t^-,$$

where $(N_t^+)_{t\geq 0}$ and $(N_t^-)_{t\geq 0}$ are two independent Poisson processes with intensity 1. So, using that N^1 and N^2 have the same law and Cauchy-Schwarz's inequality

$$\begin{aligned}\mathbb{E}[Z_t^0 \mathbf{1}_{Z_t^0>0}] &= \mathbb{E}[(N_t^+ - N_t^-)\mathbf{1}_{N_t^+ > N_t^-}] = \mathbb{E}[(N_t^- - N_t^+)\mathbf{1}_{N_t^- > N_t^+}] \\ &= \frac{1}{2}\mathbb{E}[|N_t^+ - N_t^-|] \leq \frac{1}{2}\mathrm{Var}(N_t^+ - N_t^-)^{1/2} = \sqrt{\frac{t}{2}}.\end{aligned}$$

This yields

$$\sup_{i\in\mathbb{N}^*} \mathbb{P}(\widetilde{Y}_t^i = i) \leq \mathbb{P}(\widetilde{Y}_t^0 = 0) \leq \frac{1}{\sqrt{2t}},$$

which achieves the proof. □

Up to the knowledge of the authors, Stein's factors associated to the Wasserstein distance have not been studied yet. The following proposition provides upper bounds on these factors.

Proposition 6.4.7 (Estimation of the Stein's factors for Lipschitz function and geometric approximation)**.** *For all $0 < \alpha < \beta$, set $u(x) = q^x$ on $\mathbb{N}$ with $q = \sqrt{\frac{\beta}{\alpha}} = \rho^{-1/2}$. Then,*

$$\begin{aligned}\sup_{f\in Lip(d_u)} \left\| \frac{g_f}{u} \right\|_\infty &= \frac{1}{\sigma(u)} = \frac{1}{(\sqrt{\beta} - \sqrt{a})^2}, \\ \sup_{f\in Lip(d_u)} \|\partial_v g_f\|_\infty &\leq \frac{1}{(\sqrt{\beta} - \sqrt{a})^2}\left(1 + \sqrt{\frac{\alpha}{\beta}} \min\left\{1, \frac{2\sqrt{\pi}}{(\alpha\beta)^{1/4}}(\sqrt{\beta} - \sqrt{a}) - 1\right\}\right).\end{aligned}$$

Proof of Proposition 6.4.7. By application of Theorem 6.3.6, one has immediately the first equation. By Theorem 6.3.7 with $u(x) = q^x$, $q = \rho^{-1/2} = \sqrt{\frac{\beta}{\alpha}}$, we have:

$$\sup_{f\in \mathrm{Lip}(d_u)} \|\partial_v g_f\|_\infty \le \int_0^\infty e^{-(\sqrt{\beta}-\sqrt{a})^2 t}\left(1-\sqrt{\frac{\alpha}{\beta}}+2\sqrt{\frac{\alpha}{\beta}}\sup_{i\in\mathbb{N}^*}\mathbb{P}(X^i_{1,*u,t}=i)\right)dt,$$

where $(X^i_{1,*u,t})_{t\ge 0}$ is a M/M/1 queue with rates $(\sqrt{\alpha\beta}, \sqrt{\alpha\beta}\mathbf{1}_{\mathbb{N}^*})$. On the one hand, this yields directly

$$\sup_{f\in \mathrm{Lip}(d_u)} \|\partial_v g_f\|_\infty \le \frac{1}{(\sqrt{\beta}-\sqrt{a})^2}\left(1+\sqrt{\frac{\alpha}{\beta}}\right).$$

On the other hand, as a consequence of Lemma 6.4.6, one has

$$\begin{aligned}\sup_{f\in \mathrm{Lip}(d_u)} \|\partial_v g_f\|_\infty &\le \int_0^\infty e^{-(\sqrt{\beta}-\sqrt{a})^2 t}\left(1-\sqrt{\frac{\alpha}{\beta}}+2\frac{1}{(\alpha\beta)^{1/4}}\sqrt{\frac{\alpha}{\beta}}\frac{1}{\sqrt{t}}\right)dt\\ &= \frac{1}{(\sqrt{\beta}-\sqrt{a})^2}\left(1-\sqrt{\frac{\alpha}{\beta}}\right)+\frac{1}{(\sqrt{\beta}-\sqrt{\alpha})}\sqrt{\frac{\alpha}{\beta}}\frac{2}{(\alpha\beta)^{1/4}}\int_0^\infty e^{-t}\frac{dt}{\sqrt{t}}\\ &= \frac{1}{(\sqrt{\beta}-\sqrt{a})^2}\left(1-\sqrt{\frac{\alpha}{\beta}}\right)+\frac{1}{(\sqrt{\beta}-\sqrt{\alpha})}\sqrt{\frac{\alpha}{\beta}}\frac{2\sqrt{\pi}}{(\alpha\beta)^{1/4}}.\end{aligned}$$

□

Remark 6.4.8 (On the best upper bound)**.** *The expression* $\frac{2\sqrt{\pi}}{(\alpha\beta)^{1/4}}(\sqrt{\beta}-\sqrt{a})-1$ *is smaller than* 1 *as soon as*

$$\frac{\sqrt{\beta}-\sqrt{a}}{(\alpha\beta)^{1/4}} < \frac{1}{\sqrt{\pi}},$$

so there is a range of values of the parameters α *and* β*, for example if they are close to each other, for which the factor inside the* min *is actually a better upper bound than* 1*.*

We now turn to the subject of the mixture of geometric laws. Set $\varphi = 1$ and $\rho < 1$, then $\mathcal{I}_\varphi(\rho) = \mathcal{G}(\rho)$. We choose $u(k) = q^k$ on $\mathbb{N}$, hence $d_u(x,y) = |q^x - q^y|/|q-1|$. The preceding theorem put together with Theorem 6.3.12 gives for $q = \rho^{-1/2}$ and in the case where $\rho' < \sqrt{\rho}$,

$$W_{d_u}(\mathcal{G}(\rho),\mathcal{G}(\rho')) \le |\rho-\rho'| \times \frac{1}{(1-\sqrt{\rho})^2} \times \frac{1-\rho'}{\sqrt{\rho}-\rho'}.$$

The case $\rho' > \sqrt{\rho}$ is similar.

By the same reasoning as the one used in the proof of Theorem 6.3.14, for a random variable R such that $\mathbb{E}[R] = \rho$, and a random variable such that $\mathcal{L}(W|R) = \mathcal{G}(R)$, we have the inequality:

$$d_{\mathcal{F}}(\mathcal{L}(W),\mathcal{G}(\rho)) \le \sup_{f\in\mathcal{F}} \|\partial_u g_f\|_\infty \mathbb{E}[(\rho - R)d_u(W+1, G+1)],$$

where $G \sim \mathcal{G}(\rho)$. Let $G' \sim \mathcal{G}(\rho')$. With the interpretation of the geometric laws as the number of repetitions of a binary experiment before the first success, it is easy to find a coupling such that a.s. $G \le G'$ when $\rho \le \rho'$. This yields

$$\mathbb{E}[d_u(G,G')] = \frac{1}{|1-q|}\left|\frac{1-\rho}{1-q\rho} - \frac{1-\rho'}{1-q\rho'}\right| = \frac{|\rho-\rho'|}{|(1-q\rho)(1-q\rho')|}.$$

Hence, if a.s. $R < \frac{1}{q}$, by Remark 6.3.15:

$$d_{\mathcal{F}}(\mathcal{L}(W),\mathcal{G}(\rho)) \le \sup_{f\in\mathcal{F}} \|\partial_u g_f\|_\infty \frac{q}{1-q\rho}\mathbb{E}\left[\frac{|\rho-R|^2}{(1-qR)}\right].$$

Finally, by taking $q = \rho^{-1/2}$, one finds that for two random variables R, S such that $\mathbb{E}[R] = \rho$ and a.s. $R < \frac{1}{\sqrt{\rho}}$, and $\mathcal{L}(W|R) = \mathcal{G}(R)$, the following upper bound holds:

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{G}(\rho)) \le \frac{1 + \frac{1}{\sqrt{\rho}}}{(1-\sqrt{\rho})^3}\, \mathbb{E}\left[\frac{|\rho - R|^2}{(1 - \frac{R}{\sqrt{\rho}})}\right].$$

6.4.4 Another example

Let us consider the BDP(α, β) with $\alpha(x) = x + 2, \beta(x) = x^2$ on $\mathbb{N}$. Its invariant measure is a Poisson size-biased type distribution, defined as

$$\pi(x) = \frac{1}{2e}\frac{(x+1)}{x!}, \qquad x \in \mathbb{N}.$$

Here size-biased means that if $X \sim \pi$ and $Y \sim \mathcal{P}(1)$ then:

$$\mathbb{P}(X = x) = \frac{\mathbb{E}[(Y+1)\mathbf{1}_{Y=x}]}{\mathbb{E}[(Y+1)]} = \frac{(x+1)\mathbb{P}(Y=x)}{\sum_{j\ge 0}(j+1)\mathbb{P}(Y=j)}, \qquad x \in \mathbb{N}.$$

Choosing the weight u such that $u(x+1)/u(x) = (x+1)/(x+3)$ for all $x \in \mathbb{N}$, i.e. for example

$$u(x) = \frac{1}{(x+1)(x+2)}, \qquad x \in \mathbb{N},$$

we find that V_u is constant. By Theorem 6.2.5 with $v = 1$, we have an intertwining with potential $V_{u,v}(x) = 2x + 1$ on $\mathbb{N}$. Moreover, by Theorem 6.2.8, we have convergence of the semigroup towards π in the distance $\zeta_{u,1}$ at rate 1.

The three next sections are devoted to the omitted proofs of the previous results.

6.5 Proofs of Section 6.2

6.5.1 First order intertwining for the backward gradient ∂_u^*

First of all, let us state the analogous of Theorem 6.2.1 for the backward gradient ∂^*. Let $(P_{*u,t})_{t\ge 0}$ be the birth-death semigroup associated to the generator L_{*u}, where for all non-negative or bounded function $f : \mathbb{N} \to \mathbb{R}$ and $x \in \mathbb{N}$,

$$L_{*u}f = \alpha_{*u}\,\partial f + \beta_{*u}\,\partial^* f, \qquad V_{*u} = \overleftarrow{\alpha} - \alpha_{*u} + \beta - \beta_{*u},$$
$$\alpha_{*u}(x) = \frac{u(x+1)}{u(x)}\alpha(x), \qquad \beta_{*u}(x) = \frac{u(x-1)}{u(x)}\beta(x-1)\mathbf{1}_{x\in\mathbb{N}^*}.$$

The potential V_{*u} can be rewritten under the compacted form $V_{*u} = \partial_u^*\left(\overrightarrow{u}\alpha - u\beta\right)$. We can also notice that $V_{*u} = \overleftarrow{V_{\overrightarrow{u}}}$ on $\mathbb{N}^*$.

Theorem 6.5.1 (First-order intertwining relation for the backward gradient)**.** *If V_{*u} is bounded from below, then for every real-valued function on $\mathbb{N}$ such that $\|\partial_u^* f\|_\infty < +\infty$, and for all $t \ge 0$,*

$$\partial_u^* P_t f = P_{*u,t}^{V_{*u}}\,\partial_u^* f. \tag{6.19}$$

Let us call $(X_{*u,t}^x)_{t\ge 0}$ the birth-death process of generator L_{*u} such that $X_{*u,0}^x = x$. The process $(X_{*u,t}^x)_{t\ge 0}$ is not irreducible, although it is indecomposable, i.e. it possesses only one recurrent class. Indeed if $x \in \mathbb{N}^*$ then $(X_{*u,t}^x)_{t\ge 0}$ never visits the state 0 as $\beta_{*u}(1) = 0$ and if $x = 0$ the process $(X_{*u,t}^0)_{t\ge 0}$ leaves 0 almost surely.

Proof of Theorem 6.5.1. The core of the proof relies on the intertwining relation at the level of generators:

$$\partial_u^* Lf = L_{*u}\partial_u^* f - V_{*u}\partial_u^* f, \tag{6.20}$$

which is derived by easy computations. The intertwining at the level of the semigroups follows by the same arguments as in the proof of Theorem 2.1 of Chafaï and Joulin [2013]. We briefly recall these arguments. For all $s \in [0,t]$ let us set $J(s) = P_{*u,s}^{V_{*u}}(\partial_u^* P_{t-s} f)$. If the function $\partial_u^* P_{t-s} f$ is bounded on $\mathbb{N}$, then the Kolmogorov equations (6.6) for the Feynman-Kac semigroup $(P_{*u,t}^{V_{*u}})_{t\geq 0}$ hold and

$$J'(s) = P_{*u,s}^{V_{*u}}((L_{*u} - V_{*u})\partial_u^* P_{t-s} f - \partial_u^* L P_{t-s} f).$$

Thanks to the formula (6.20) this gives $J'(s) = 0$. Hence $J(0) = J(t)$ which is exactly the identity (6.19).

Let us show that $\partial_u^* P_{t-s} f$ is bounded on $\mathbb{N}$. Indeed, recall that $V_{*u}(x+1) = V_{\vec{u}}(x)$ on $\mathbb{N}$. Furthermore $\partial_u^* f(x+1) = \partial_{\vec{u}} f(x)$ on $\mathbb{N}$. Hence $V_{\vec{u}}$ and $\partial_{\vec{u}} f$ are bounded on $\mathbb{N}$, which implies that $\partial_{\vec{u}} P_{t-s} f$ is bounded (Chen [1996]). For all positive integer $\partial_{\vec{u}} P_{t-s} f(x) = \partial_u^* P_{t-s} f(x+1)$, so $\partial_u^* P_{t-s} f$ is bounded. □

6.5.2 Alternative proof of first order intertwining theorems

This section aims to give a sample path interpretation of the first order intertwining relations (6.7) and (6.19), at least in a particular case. It is independent of the other sections.

We focus on the case where the weight is $u = 1$ with non-increasing birth rates $(\alpha(x))_{x\in\mathbb{N}}$ and non-decreasing death rates $(\beta(x))_{x\in\mathbb{N}}$. When intertwining the birth-death semigroup with the forward gradient ∂, one obtains a new birth-death semigroup with shifted birth rate and unchanged death rate

$$\alpha_1 = \vec{\alpha}, \quad \beta_1 = \beta,$$

whereas when intertwining the birth-death semigroup with the backward gradient ∂^*, one obtains a new birth-death semigroup with shifted death rate and unchanged birth rate:

$$\alpha_{*1} = \alpha, \quad \beta_{*1} = \overleftarrow{\beta}.$$

In order to explain this fact, we will give a probabilistic proof of the formulae (6.7) and (6.19). Recall that for all real-valued bounded functions on $\mathbb{N}$ and $x \in \mathbb{N}$,

$$\begin{aligned} \partial P_t f(x) &= \mathbb{E}[f(X_t^{x+1}) - f(X_t^x)], \\ \partial^* P_t f(x+1) &= \mathbb{E}[f(X_t^x) - f(X_t^{x+1})]. \end{aligned}$$

At time $t = 0$, $X_t^{x+1} = X_t^x + 1$. We construct a process $(S_t)_{t\geq 0}$ such that for all $t \geq 0$, $X_t^{x+1} = X_t^x + S_t$ and $S_t \in \{0,1\}$. If for a time t, $S_t = 0$, then we choose the sticking coupling between $(X_{t+s}^x)_{s\geq 0}$ and $(X_{t+s}^{x+1})_{s\geq 0}$ (i.e. the process $(S_t)_{t\geq 0}$ is absorbed in 0). If $S_t = 1$, it is natural to construct the following coupling:

1. with rate $\alpha(X_t^x + 1) = \alpha(X_t^{x+1})$, X_t^x and X_t^{x+1} jump upwards together and S_t remains equal to 1,
2. with rate $\beta(X_t^x) = \beta(X_t^{x+1} - 1)$, X_t^x and X_t^{x+1} jump downwards together and S_t remains equal to 1,
3. with rate $\alpha(X_t^x) - \alpha(X_t^x + 1) = \alpha(X_t^{x+1} - 1) - \alpha(X_t^{x+1})$, X_t^x jumps upwards, X_t^{x+1} does not jump and S_t jumps from 1 to 0,
4. with rate $\beta(X_t^{x+1}) - \beta(X_t^x) = \beta(X_t^{x+1}) - \beta(X_t^{x+1} - 1)$, X_t^{x+1} jumps downwards, X_t^x does not jump and S_t jumps from 1 to 0.

This implies in particular that for all $t \geq 0$ the process S_t jumps from 1 to 0 with rate

$$\alpha(X_t^x) - \alpha(X_t^x + 1) + \beta(X_t^{x+1}) - \beta(X_t^x) \;=\; V_1(X_t^x) \;=\; V_{*1}(X_t^{x+1}).$$

Moreover, conditionally to $\{S_t = 1\}$, $(X_t^x)_{t\geq 0}$ evolves as a BDP$(\overrightarrow{\alpha}, \beta)$ and $(X_t^{x+1})_{t\geq 0}$ evolves as a BDP$(\alpha, \overleftarrow{\beta})$). Indeed, as long as $S_t = 1$, the steps (3) and (4) do not occur.

To exploit rigorously the preceding facts, let us introduce the BDP$(\overrightarrow{\alpha}, \beta)$ starting from x denoted by $(X_{1,t}^x)_{t\geq 0}$, whose standard filtration is $(\mathcal{F}_t)_{t\geq 0}$. The processes $(X_t^x)_{t\geq 0}$ and $(X_{1,t}^x)_{t\geq 0}$, as well as $(X_t^{x+1})_{t\geq 0}$ and $(X_{1,t}^x + 1)_{t\geq 0}$, can be coupled as follows :

1. Let E be an exponential with parameter 1 and T such that $T = \inf\{t \geq 0, \int_0^t V(X_{1,s}^x)ds > E\}$.
2. Set $S_t = 1$ if $t < T$ and $S_t = 0$ otherwise.
3. Set $X_t^x = X_{1,t}^x$ for $t \leq T$.
4. At time T, sample a random variable Z satisfying to

$$\mathbb{P}(Z = X_{1,T}^x + 1 \mid \mathcal{F}_T) = \frac{\alpha(X_{1,T}^x) - \alpha(X_{1,T}^x + 1)}{V(X_{1,T}^x)},$$
$$\mathbb{P}(Z = X_{1,T}^x \mid \mathcal{F}_T) = \frac{\beta(X_{1,T}^x + 1) - \beta(X_{1,T}^x)}{V_1(X_{1,T}^x)}.$$

5. Let evolve the process $(X_t^x)_{t\geq T}$ as a BDP(α, β) starting from Z.

The coupling $(X_t^x, X_{1,t}^x, S_t)_{t\geq 0}$ satisfy to

$$X_t^x \mathbf{1}_{S_t=1} = X_{1,t}^x \mathbf{1}_{S_t=1}, \quad X_t^{x+1}\mathbf{1}_{S_t=1} = (X_{1,t}^x + 1)\mathbf{1}_{S_t=1}, \quad \mathbb{P}(S_t = 1 | (X_{1,s}^x)_{0\leq s\leq t}) = e^{-\int_0^t V_1(X_{1,s}^x)ds}.$$

This allows to find back the formula (6.7):

$$\begin{aligned}\partial P_t f(x) &= \mathbb{E}[f(X_t^{x+1}) - f(X_t^x)] = \mathbb{E}\left[(f(X_t^{x+1}) - f(X_t^x))\mathbf{1}_{S_t=1}\right] \\ &= \mathbb{E}\left[(f(X_{1,t}^x + 1) - f(X_{1,t}^x))e^{-\int_0^t V_1(X_{1,s}^x)ds}\right] = P_{1,t}^{V_1}(\partial f)(x).\end{aligned}$$

Similarly it is possible to construct a coupling $(X_t^{x+1}, X_{*1,t}^{x+1}, S_t)_{t\geq 0}$ such that $(X_{*1,t}^{x+1})_{t\geq 0}$ is a BDP$(\alpha, \overleftarrow{\beta})$ starting from $x+1$ and satisfying to

$$X_t^{x+1}\mathbf{1}_{S_t=1} = X_{*1,t}^{x+1}\mathbf{1}_{S_t=1}, \quad X_t^x \mathbf{1}_{S_t=1} = (X_{*1,t}^{x+1} - 1)\mathbf{1}_{S_t=1}, \quad \mathbb{P}(S_t = 1 | (X_{*1,s}^{x+1})_{0\leq s\leq t}) = e^{-\int_0^t V_{*1}(X_{*1,s}^{x+1})ds},$$

leading to the formula (6.19):

$$\begin{aligned}\partial^* P_t f(x+1) &= \mathbb{E}[f(X_t^x) - f(X_t^{x+1})] = \mathbb{E}[(f(X_t^x) - f(X_t^{x+1}))\mathbf{1}_{S_t=1}] \\ &= \mathbb{E}\left[(f(X_{*1,t}^{x+1} - 1) - f(X_{*1,t}^{x+1}))e^{-\int_0^t V_{*1}(X_{*1,s}^{x+1})ds}\right] = P_{*1,t}^{V_{*1}}(\partial^* f)(x+1).\end{aligned}$$

It is interesting to remark that conversely, the intertwining formula (6.7) can in certain cases yield a coupling between $(X_t^x)_{t\geq 0}$ and $(X_t^{x+1})_{t\geq 0}$. The proof of Lemma 6.4.2 above is based on this idea.

6.5.3 Proof of Theorems 6.2.2 and 6.2.5

Proof of Theorem 6.2.2. Let us begin by showing the following intertwining relation at the level of the generators:

$$\partial_v^*\partial_u Lf = L_{u,*v}\partial_v^*\partial_u f - V_{u,*v}\partial_v^*\partial_u f.$$

By application of Theorem 6.2.1 and Theorem 6.5.1 we find that

$$\begin{aligned}\partial_v^*(\partial_u Lf) &= \partial_v^*(L_u(\partial_u f) - V_u\partial_u f)\\ &= (L_u)_{*v}\partial_v^*\partial_u f - (V_u)_{*v}\partial_v^*\partial_u f + \partial_v^*(-V_u\partial_u f),\end{aligned}$$

where $(L_u)_{*v}$ and $(V_u)_{*v}$ stand for the generator, respectively the potential, obtained by intertwining the BDP(α_u,β_u) and the ∂_{*v} gradient. The generator $(L_u)_{*v}$ is the generator of a BDP$((\alpha_u)_{*v},(\beta_u)_{*v})$ such that for all $x\in\mathbb{N}$,

$$\begin{aligned}(\alpha_u)_{*v}(x) &= \frac{v(x+1)}{v(x)}\alpha_u(x) = \frac{v(x+1)}{v(x)}\frac{u(x+1)}{u(x)}\alpha(x+1)\\ (\beta_u)_{*v}(x) &= \frac{v(x-1)}{v(x)}\beta_u(x-1) = \frac{v(x-1)}{v(x)}\frac{u(x-2)}{u(x-1)}\beta(x-1)\mathbf{1}_{x\in\mathbb{N}^*}.\end{aligned}$$

The potential $(V_u)_{*v}$ writes on $\mathbb{N}$

$$(V_u)_{*v}(x) = \alpha_u(x-1)\mathbf{1}_{x\in\mathbb{N}^*} - (\alpha_u)_{*v}(x) + \beta_u(x) - (\beta_u)_{*v}(x).$$

The next step is to rewrite the expression $\partial_v^*(-V_u\partial_u f)$ in terms of $\partial_v^*\partial_u f$. Let us denote $g=\partial_u f$ in the following lines. For every $x\in\mathbb{N}^*$, $\partial_u^*(fg)(x) = f(x)\partial_u^*g(x) + \partial_u^*f(x)g(x-1)$ and $f(x) = -\sum_{k=0}^{x} u(k)\partial_u^*f(k)$ so that

$$\begin{aligned}\partial_v^*(-V_u g)(x) &= -V_u(x)\partial_v^*g(x) - \partial_v^*V_u(x)g(x-1)\\ &= -V_u(x)\partial_v^*g(x) + \partial_v^*V_u(x)\sum_{k=0}^{x-1} v(k)\partial_v^*g(k)\\ &= \partial_v^*V_u(x)\sum_{k=0}^{x-1} v(k)(\partial_v^*g(k) - \partial_v^*g(x)) - \left(V_u(x) - \left(\sum_{k=0}^{x-1} v(k)\right)\partial_v^*V_u(x)\right)\partial_v^*g(x).\end{aligned}$$

Besides, $\partial_v^*(-V_u g)(0) = \frac{1}{v_0}V_u(0)g(0) = -V_u(0)\partial_v^*g(0)$. We do indeed find $\partial_v^*\partial_u L = (L_{u,*v} - V_{u,*v})\partial_v^*\partial_u$ with

$$\begin{aligned}L_{u,*v}f(x) &= (L_u)_{*v}f(x) + \partial_v^*V_u(x)v(x-1)(f(x-1)-f(x))\\ &\quad +\partial_v^*V_u(x)\left(\sum_{j=0}^{x-2} v(j)\right)\sum_{k=0}^{x-2}\frac{v(k)}{\left(\sum_{j=0}^{x-2}v(j)\right)}(f(k)-f(x))\\ V_{u,*v}(x) &= (V_u)_{*v}(x) + V_u(x) - \left(\sum_{k=0}^{x-1} v(k)\right)\partial_v^*V_u(x).\end{aligned}$$

The generator $L_{u,*v}$ has a birth-death component and a component making the process at point x jumping on the set $\{0,\ldots,x-2\}$. The birth rates are $\alpha_{u,*v} = (\alpha_u)_{*v}$. The death rates come from $(L_u)_{*v}$ and from the term $\partial_v^*V_u(x)v(x-1)(f(x-1)-f(x))$, so that

$$\beta_{u,*v}(x) = (\beta_u)_{*v} + \partial_v^*V_u(x)v(x-1)\mathbf{1}_{x\in\mathbb{N}^*}.$$

Remembering that $V_u(x) = \alpha(x) - \alpha_u(x) + \beta(x+1) - \beta_u(x)$ we get that for all positive integer x

$$\begin{aligned}(V_u)_{*v}(x) + V_u(x) &= \alpha(x) + \alpha_u(x-1) - (\alpha_u(x) + (\alpha_u)_{*v}(x)) + \beta(x+1) - (\beta_u)_{*v}(x)\\ &= \left(1+\frac{u(x)}{u(x-1)}\right)\alpha(x) - \left(1+\frac{v(x+1)}{v(x)}\right)\frac{u(x+1)}{u(x)}\alpha(x+1)\\ &\quad +\beta(x+1) - \frac{v(x-1)}{v(x)}\frac{u(x-2)}{u(x-1)}\beta(x-1),\end{aligned}$$

and

$$\begin{aligned}(V_u)_{*v}(0)+V_u(0) &= -(\alpha_u)_{*v}(0)+V_u(0)=\alpha(0)-(\alpha_u(0)+(\alpha_u)_{*v}(0))+\beta(1)\\ &= \alpha(0)-\left(1+\frac{v(1)}{v(0)}\right)\frac{u(1)}{u(0)}\alpha(1)+\beta(1).\end{aligned}$$

The same reasoning as in the proof of Theorem 6.5.1 allows to deduce the relation at the level of the semigroups from the relation at the level of the generators, provided that we can show that for all $t \geq 0$ the function $\partial_v^* \partial_u P_t f$ is bounded on $\mathbb{N}$. It is the case; indeed, by Theorem 6.2.1, $\partial_u P_t f = P_{u,t}^{V_u} \partial_u f$ is bounded and $\partial_v^* |\partial_u P_t f| \leq \frac{2}{\inf_{x\in\mathbb{N}} v(x)} |P_{u,t}^{V_u} \partial_u f|$. □

Proof of Theorem 6.2.5. Surprisingly, Theorem 6.2.5 cannot be deduced from Theorem 6.2.2 when $u \neq 1$. However, its proof goes along the same lines as the proof of Theorem 6.2.2, only easier because $\partial_v \partial_u (V_u \partial_u f) = V_u \partial_v \partial_u f$, so that the intertwining relation at the level of the generators follows directly. □

6.6 Proofs of Section 6.3

The semigroup representation (6.14) of the solution of Stein's equation g_f can be rewritten as:

$$\begin{aligned}\vec{g_f} &= -u\int_0^\infty \partial_u P_t f dt, &(6.21)\\ \partial g_f &= u\int_0^\infty \partial_u \partial^* P_t f dt, &(6.22)\\ \overrightarrow{\partial g_f} &= -u\int_0^\infty \partial_u \partial P_t f dt. &(6.23)\end{aligned}$$

The left-hand side of an intertwining relation between a weighted gradient and a birth-death semigroup appears under the integral. This fact suggests to apply the intertwining relations shown previously. However, it leads to sharper results to first identify the function $f \in \mathcal{F}$ that realizes the maximum in the pointwise Stein's factors

$$\max_{f\in\mathcal{F}} |g_f(i)|, \quad \max_{f\in\mathcal{F}} |\partial g_f(i)|,$$

for every $i \in \mathbb{N}$. This first step is based on Lemma 6.6.1 and Lemma 6.6.2 below. Indeed, Lemma 6.6.1 gives an alternative formulation of the solution of Stein's equation.

Lemma 6.6.1 ([Brown and Xia, 2001, Lemma 2.3]). *For all $i \in \mathbb{N}$, let us define $g_j := g_{\mathbf{1}_j}$ and*

$$e_i^+ = \frac{1}{\alpha(i)\pi(i)}\sum_{k=0}^{i}\pi(k), \quad i\in\mathbb{N} \qquad\qquad e_i^- = \frac{1}{\beta(i)\pi(i)}\sum_{k=i}^{\infty}\pi(k), \quad i\in\mathbb{N}^*.$$

Then, for all $i \in \mathbb{N}^*, j \in \mathbb{N}$,

$$g_j(i) = \pi(j)(-e_{i-1}^+ \mathbf{1}_{i\leq j} + e_i^- \mathbf{1}_{i\geq j+1}) \tag{6.24}$$
$$\partial g_j(i) = \pi(j)\left((e_{i-1}^+ - e_i^+)\mathbf{1}_{j\geq i+1} + (e_{i+1}^- + e_{i-1}^+)\mathbf{1}_{i=j} + (e_{i+1}^- - e_i^-)\mathbf{1}_{j\leq i-1}\right). \tag{6.25}$$

Lemma 6.6.2 ([Brown and Xia, 2001, Lemma 2.4]). *If $V_1 \geq 0$ then (e_i^+) is non-decreasing and (e_i^-) is non-increasing.*

6.6.1 Approximation in total variation distance

We begin by describing the argmax of the pointwise quantities. To the knowledge of the authors, equation (6.26) is not explicitly stated in preceding works. Equation (6.27) is proved in Brown and Xia [2001]. We briefly recall the arguments used for the sake of completeness.

Lemma 6.6.3 (Argmax of the pointwise Stein's factor)**.** *For all $i \in \mathbb{N}$,*

$$g_{\mathbf{1}_{[0,i]}}(i) = \sup_{0 \le f \le 1} \overrightarrow{g_f}(i). \tag{6.26}$$

Moreover if $V_1 \ge 0$, then for all $i \in \mathbb{N}^*$

$$\partial g_{\mathbf{1}_i}(i) \quad = \quad \max_{0 \le f \le 1} |\partial g_f(i)|. \tag{6.27}$$

Proof. By replacing f by $1-f$ if necessary

$$\sup_{0\le f\le 1} |g_f(i)| = \sup_{0\le f\le 1} g_f(i), \qquad \sup_{0\le f\le 1} |\partial g_f(i)| = \sup_{0\le f\le 1} \partial g_f(i).$$

By Lemma 6.6.1,

$$g_f(i+1) = e^-_{i+1} \sum_{j=0}^{i} \pi(j) f(j) - e^+_i \sum_{j=i+1}^{\infty} \pi(j) f(j) \le e^-_{i+1} \sum_{j=0}^{i} f(j),$$

with equality for $f = \mathbf{1}_{[0,i]}$ which proves (6.26). On the other hand,

$$\partial g_j(i) = \pi_j \left((e^+_{i-1} - e^+_i) \mathbf{1}_{i \le j-1} + (e^-_{i+1} + e^+_{i-1}) \mathbf{1}_{i=j} + (e^-_{i+1} - e^-_i) \mathbf{1}_{i \ge j+1} \right),$$

so by Lemma 6.6.2 the quantity $\partial g_j(i)$ is non-negative if and only if $i = j$. Hence, if f is a function on $\mathbb{N}$ with values in $[0,1]$,

$$\partial g_f(i) = \sum_{j=0}^{\infty} f(j) \partial g_j(i) \le \partial g_i(i),$$

and there is equality if $f = \mathbf{1}_i$. This shows (6.27). □

As a consequence, we have the following lemma of which Theorem 6.3.2 is a direct application.

Lemma 6.6.4 (Pointwise first Stein's factor for bounded functions)**.** *If V_u is bounded from below by $\sigma(u)$ then for all $i \in \mathbb{N}$,*

$$\sup_{0\le f\le 1} |g_f(i+1)| \le \int_0^\infty e^{-\sigma(u)t} \mathbb{P}(X^i_{u,t} = i) dt.$$

Moreover if V_u is constant then the preceding inequality is in fact an equality.

Proof. By the equation (6.21), Theorem 6.2.1 and Lemma 6.6.3, and because $\partial_u \mathbf{1}_{[0,i]} = -\frac{1}{u(i)} \mathbf{1}_i$, we have for all function f such that $0 \le f \le 1$

$$|g_f(i+1)| \le g_{\mathbf{1}_{[0,i]}}(i+1) \quad = -u(i) \int_0^\infty P^{V_u}_{u,t}(\partial_u \mathbf{1}_{[0,i]}) = \int_0^\infty P^{V_u}_{u,t}(\mathbf{1}_i) \le \int_0^\infty e^{-\sigma(u)t} \mathbb{P}(X^i_{u,t} = i) dt.$$

□

We now state results for the second pointwise Stein factor.

Lemma 6.6.5 (Pointwise second Stein's factor for bounded functions)**.**

— *Under* $\mathbf{H_1}$*, for all integer* $i \in \mathbb{N}^*$*, the quantity* $\sup_{0\leq f\leq 1} |\partial g_f(i)|$ *is bounded by*

$$\int_0^\infty e^{-\sigma(1,*u)t}\left(-\frac{u(i)}{u(i-1)}\mathbb{P}(X^i_{1,*u,t}=i-1)+2\mathbb{P}(X^i_{1,*u,t}=i)-\frac{u(i)}{u(i+1)}\mathbb{P}(X^i_{1,*u,t}=i+1)\right)dt.$$

— *Under* $\mathbf{H_2}$*, for all integer* $i \in \mathbb{N}$*, the quantity* $\sup_{0\leq f\leq 1} |\partial g_f(i+1)|$ *is bounded by*

$$\int_0^\infty e^{-\sigma(1,u)t}\left(-\frac{u(i)}{u(i-1)}\mathbb{P}(X^i_{1,u,t}=i-1)+2\mathbb{P}(X^i_{1,u,t}=i)-\frac{u(i)}{u(i+1)}\mathbb{P}(X^i_{1,u,t}=i+1)\right)dt.$$

Moreover, if the potential $V_{1,*u}$ *(respectively* $V_{1,u}$*) is constant, then the first (respectively the second) upper bound is in fact an equality.*

Proof. For every positive integer i, let $f_i = \mathbf{1}_i$. By the equation (6.22), Theorem 6.2.2 and Lemma 6.6.3, under $\mathbf{H_1}$,

$$\sup_{0\leq f\leq 1} |\partial g_f(i)| = \partial g_{f_i}(i) = u(i)\partial_u g_{f_i}(i) \leq u(i)\int_0^\infty e^{-\sigma(1,*u)t}\mathbb{E}\left[\partial_u^*\partial f_i(\widetilde{X}^i_t)\right]dt.$$

As $\partial_u^*\partial f_i = -\frac{1}{u(i-1)}\mathbf{1}_{i-1} + 2\frac{1}{u(i)}\mathbf{1}_i - \frac{1}{u(i+1)}\mathbf{1}_{i+1}$, we get the announced inequality.
Similarly the result under $\mathbf{H_2}$ derives from the equation (6.23), Theorem 6.2.5, Lemma 6.6.3 and the computation $-\partial_u\partial f_{i+1} = -\frac{1}{u(i-1)}\mathbf{1}_{i-1} + 2\frac{1}{u(i)}\mathbf{1}_i - \frac{1}{u(i+1)}\mathbf{1}_{i+1}$. □

Theorem 6.3.3 and 6.3.4 are direct consequences of Lemma 6.6.5.

6.6.2 Approximation in Wasserstein distance.

In contrast with the first order in total variation distance, the bound of Theorem 6.3.6 does not require a preliminary bound on pointwise Stein's factor.

Proof of Theorem 6.3.6. By Theorem 6.2.1,

$$\begin{aligned}\left|\frac{1}{u}g_f(\cdot+1)\right| &= |\partial_u h_f| = \left|\int_0^\infty \partial_u P_t f\,dt\right| \\ &= \left|\int_0^\infty \mathbb{E}\left[e^{-\int_0^t V_u(X_{u,s})ds}\partial_u f(X_{u,t})\right]dt\right| \leq \frac{1}{\sigma(u)}\|\partial_u f\|_\infty.\end{aligned}$$

Now to prove the sharpness if V_u is constant, it is enough to consider the map $f : x \mapsto -\sum_{k=1}^x u(x-1)$ for which the previous inequalities are in fact equalities. □

Remark 6.6.6 (Variant of Theorem 6.3.6)**.** *We can also derive an upper bound for*

$$\sup_{f\in Lip(d_u)} \|g_f/u\|_\infty,$$

under the condition that V_{*u} *is bounded by below, by using alternatively to equation* (6.21) *the equation*

$$g_f = -u\int_0^\infty \partial_u^* P_t f\,dt,$$

and Theorem 6.5.1 instead of Theorem 6.2.1.

For the second Stein factor, we begin by focusing on the pointwise quantity $\sup_{f\in\mathcal{F}} \partial_u g_f(i)$. For all $i \in \mathbb{N}$, let us introduce two functions ψ_i and Ψ_i defined for all $j \in \mathbb{N}$ as

$$\begin{aligned}\psi_i(j) &= \left(1 - \frac{u(j-1)}{u(j)}\right)\mathbf{1}_{j\le i-1} + \left(1 + \frac{u(j-1)}{u(j)}\right)\mathbf{1}_{j=i} + \left(\frac{u(j-1)}{u(j)} - 1\right)\mathbf{1}_{j\ge i+1}\\ \Psi_i(j) &= \left(1 - \frac{u(j+1)}{u(j)}\right)\mathbf{1}_{j\le i-1} + \left(1 + \frac{u(j+1)}{u(j)}\right)\mathbf{1}_{j=i} + \left(\frac{u(j+1)}{u(j)} - 1\right)\mathbf{1}_{j\ge i+1}.\end{aligned}$$

The following lemma allows to determine the functions that realize the supremum in the second pointwise Stein factor. This lemma is a generalization of a lemma of Barbour and Xia [2006], which addressed the case where $u = 1$ and $(\alpha(x), \beta(x))_{x\in\mathbb{N}} = (\lambda, x)_{x\in\mathbb{N}}$. Its proof depends on the already cited results of Brown and Xia [2001].

Lemma 6.6.7 (Argmax of the pointwise Stein's factor)**.** *If $V_1 \ge 0$, then for all $i \in \mathbb{N}^*$*

$$\partial g_{\varphi_i} = \max_{f\in Lip(d_u)} |\partial g_f(i)|, \qquad\qquad \varphi_i = -d_u(i,\cdot). \tag{6.28}$$

Proof. If f and $\widetilde{f}$ are two real-valued functions on $\mathbb{N}$, then $g_{f+\widetilde{f}} = g_f + g_{\widetilde{f}}$ and that if f is constant, then $g_f = 0$. As a consequence, by replacing f by $-f$ and $f - f(i)$ if necessary,

$$\sup_{f\in \mathrm{Lip}(d_u)} |\partial g_f(i)| = \sup_{\substack{f\in\mathrm{Lip}(d_u),\\ f(i)=0}} \partial g_f(i).$$

Recall that $g_j := g_{\mathbf{1}_j}$ for $j \in \mathbb{N}$. For all real-valued function f on $\mathbb{N}$,

$$g_f = \sum_{j\in\mathbb{N}} f(j) g_j, \qquad \partial g_f(i) = \sum_{j\in\mathbb{N}} f(j)\partial g_j(i), \quad i \in \mathbb{N}.$$

By Lemmas 6.6.1 and 6.6.2, if $f \in \mathrm{Lip}(d_u)$ and $f(i) = 0$ then

$$\begin{aligned}\partial g_f(i) &= \sum_{j=0}^{\infty} f(j)\partial g_j(i) = (e^+_{i-1} - e^+_i)\sum_{j\le i-1}\pi_j f(j) + (e^-_{i+1} - e^-_i)\sum_{j\ge i+1}\pi_j f(j)\\ &\le |\partial g_{\varphi_i}(i)| = (e^+_i - e^+_{i-1})\sum_{j\le i-1}\pi_j d_u(i,j) + (e^-_i - e^-_{i+1})\sum_{j\ge i+1}\pi_j d_u(i,j).\end{aligned}$$

□

Lemma 6.6.8 (Pointwise second Stein's factor for Lipschitz functions)**.**

— *Under $\mathbf{H_1}$, for all integer $i \in \mathbb{N}^*$,*

$$\sup_{f\in Lip(d_u)} |\partial_u g_f(i)| \le \int_0^\infty e^{-\sigma(1,*u)t}\mathbb{E}[\psi_i(X^i_{1,*u,t})]dt. \tag{6.29}$$

Moreover if $V_{1,*u}$ is constant then the preceding inequality is in fact an equality.

— *Under $\mathbf{H_2}$, for all integer $i \in \mathbb{N}$*

$$\sup_{f\in Lip(d_u)} \left|\frac{1}{u(i)}\partial g_f(i+1)\right| \le \int_0^\infty e^{-\sigma(1,u)t}\mathbb{E}[\Psi_i(X^i_{1,u,t})]dt. \tag{6.30}$$

Moreover if $V_{1,u}$ is constant then the preceding inequality is in fact an equality.

Proof of Lemma 6.6.8. Let us assume that $\mathbf{H_1}$ holds true. By the equation (6.22), Theorem 6.2.2 and Lemma 6.6.7, for every positive integer i,

$$\begin{aligned}\sup_{f\in \mathrm{Lip}(d_u)} |\partial_u g_f(i)| &= \frac{1}{u(i)} \sup_{f\in \mathrm{Lip}(d_u)} |\partial g_f(i)| = \frac{1}{u(i)}\partial g_{\varphi_i}(i) = \partial_u g_{\varphi_i}(i) = \int_0^\infty \partial_u^* \partial P_t \varphi_i dt \\ &= \int_0^\infty P^{V_{1,*u}}_{1,*u,t}(\partial_u^*\partial\varphi_i)dt \leq \int_0^\infty e^{-\sigma(1,*u)t}\mathbb{E}\left[\partial_u^*\partial\varphi_i(X^i_{1,*u,t})\right] dt.\end{aligned}$$

It is easy to check that $\psi_i = \partial_u^*\partial\varphi_i$, which proves (6.29).
Now, if $\mathbf{H_2}$ holds true, by the equation (6.23), Theorem 6.2.5 and Lemma 6.6.7, for all integer i,

$$\begin{aligned}\sup_{f\in \mathrm{Lip}(d_u)} \left|\frac{1}{u(i)}\partial g_f(i+1)\right| &= \frac{1}{u(i)} \sup_{f\in \mathrm{Lip}(d_u)} |\partial g_f(i+1)| = \frac{1}{u(i)}\partial g_{\varphi_{i+1}}(i+1) \\ &= -\int_0^\infty \partial_u\partial P_t\varphi_{i+1}(i)dt \\ &= -\int_0^\infty P^{V_{1,u}}_{1,u,t}\partial_u\partial\varphi_{i+1}(i)dt\end{aligned}$$

As $-\partial_u\partial\varphi_{i+1} = \Psi_i$, the equation (6.30) holds true. □

We deduce from Lemma 6.6.8 both Theorem 6.3.7 and Theorem 6.3.8. We only give the proof of Theorem 6.3.7 because Theorem 6.3.8 is similar.

Proof of Theorem 6.3.7. First of all let us notice that for all function $f : \mathbb{N} \to \mathbb{R}$,

$$\|\partial_u f\|_\infty \leq \sup_{x\in\mathbb{N}}\left(1+\frac{u(x+1)}{u(x)}\right)\|f/u\|_\infty, \quad \|\partial_u^* f\|_{\infty,\mathbb{N}^*} \leq \sup_{x\in\mathbb{N}^*}\left(1+\frac{u(x-1)}{u(x)}\right)\|f/u\|_\infty. \quad (6.31)$$

Under $\mathbf{H_1}$, as $\|\partial_u\varphi_i\|_\infty \leq 1$, it implies that

$$\sup_{x\in\mathbb{N}^*} |\partial_u\partial^*\varphi_i(x)| \leq \sup_{x\in\mathbb{N}^*}\left(1+\frac{u(x-1)}{u(x)}\right).$$

Plugging this in the equation (6.29) yields the first upper bound of the theorem.
On the other hand, if $u(x) = q^x$ on $\mathbb{N}$ with $q \geq 1$, then by using that $\mathbf{1}_{[0,i]} = 1 - \mathbf{1}_i - \mathbf{1}_{[i+1,\infty)}$, we write

$$\begin{aligned}\partial_u^*\partial\varphi_i(j) &= \left(1-\frac{u(j-1)}{u(j)}\right) + 2\frac{u(j-1)}{u(j)}\mathbf{1}_{j=i} + 2\left(\frac{u(j-1)}{u(j)}-1\right)\mathbf{1}_{j\geq i+1} \\ &\leq 1-\frac{1}{q}+2\frac{1}{q}\mathbf{1}_{j=i}\end{aligned}$$

which proves the second upper bound. □

6.6.3 Approximation in Kolmogorov distance

Proof of Theorem 6.3.9. As one can see in Lemma 6.6.3, the function that realizes the maximum in the first Stein factor associated to bounded functions, $f = \mathbf{1}_{[0,i]}$, is also an element of the class of the half-line indicator functions. Hence without further analysis the analogous of Lemma 6.6.4 and Theorem 6.3.2 hold by replacing $\mathcal{F} = \{0 \leq f \leq 1\}$ by $\mathcal{F} = \left\{\mathbf{1}_{[0,m]},\, m \in \mathbb{N}\right\}$. □

For the second Stein factor, we begin by determining the argmax of the pointwise factor, as we did previously.

Lemma 6.6.9 (Argmax of the pointwise Stein factor)**.** *For all $i \in \mathbb{N}$*

$$\max\left\{-\partial g_{\mathbf{1}_{[0,i-1]}}(i), \partial g_{\mathbf{1}_{[0,i]}}(i)\right\} = \sup_{f=\mathbf{1}_{[0,m]},\, m\in\mathbb{N}} |\partial g_f(i)|.$$

Proof. Let $f = \mathbf{1}_{[0,m]}$ for an integer m. By Lemma 6.6.1, if $m \leq i-1$,

$$\partial g_g(i) = \sum_{j=0}^{m} \pi(j)(e_{i+1}^- - e_i^-).$$

Hence by Lemma 6.6.2,

$$|\partial g_f(i)| = -\partial g_g(i) = (e_i^- - e_{i+1}^-) \sum_{j=0}^{m} \pi(j),$$

so the maximum when m browses the interval $[0, i-1]$ is attained in $m = i-1$.
Now, if $m \geq i$, let us call $F = 1 - f = \mathbf{1}_{[m+1,\infty)}$. By the same lemmas,

$$|g_f(i)| = |g_F(i)| = |e_i^+ - e_{i-1}^+| \sum_{j=m+1}^{\infty} \pi(j) = (e_i^+ - e_{i-1}^+) \sum_{j=m+1}^{\infty} \pi(j) = g_f(i),$$

so the maximum when m browses the interval $[i, +\infty)$ is attained in $m = i$. □

Lemma 6.6.10 (Second pointwise Stein's factor for indicator functions)**.**

— *Under* $\mathbf{H_1}$*, for all integer* $i \in \mathbb{N}^*$*, the quantity* $\sup_{f=\mathbf{1}_{[0,m]}, m\in\mathbb{N}} \partial g_f(i)$ *is bounded by the maximum of*

$$\int_0^{\infty} e^{-\sigma(1,*u)t} \left(\mathbb{P}(X_{1,*u,t}^i = i) - \frac{u(i)}{u(i-1)} \mathbb{P}(X_{1,*u,t}^i = i-1) \right) dt$$

and

$$\int_0^{\infty} e^{-\sigma(1,*u)t} \left(\mathbb{P}(X_{1,*u,t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1,*u,t}^i = i+1) \right) dt.$$

— *Under* $\mathbf{H_2}$*, for all integer* $i \in \mathbb{N}$*, the quantity* $\sup_{f=\mathbf{1}_{[0,m]}, m\in\mathbb{N}} \partial g_f(i+1)$ *is bounded by the maximum of*

$$\int_0^{\infty} e^{-\sigma(1,u)t} \left(\mathbb{P}(X_{1,u,t}^i = i) - \frac{u(i)}{u(i-1)} \mathbb{P}(X_{1,u,t}^i = i-1) \right) dt$$

and

$$\int_0^{\infty} e^{-\sigma(1,u)t} \left(\mathbb{P}(X_{1,u,t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1,u,t}^i = i+1) \right) dt.$$

Moreover, if the potential $V_{1,*u}$ *(respectively* $V_{1,u}$*) is constant, then the first (respectively the second) upper bound is in fact an equality.*

Proof. If $f = \mathbf{1}_{[0,m]}$ then $\partial_u^* \partial f_m = \frac{1}{u(m)} \mathbf{1}_m - \frac{1}{u(m+1)} \mathbf{1}_{m+1}$. Under $\mathbf{H_1}$, by equation (6.22) and Theorem 6.2.2,

$$\begin{aligned} -\partial g_{\mathbf{1}_{[0,i-1]}}(i) &= u(i) \int_0^{\infty} P_{1,*u,t}^{V_{1,*u}} \left(-\frac{1}{u(i-1)} \mathbf{1}_{i-1} + \frac{1}{u(i)} \mathbf{1}_i\right) dt \\ &\leq \int_0^{\infty} e^{-\sigma(1,*u)t} \left(\mathbb{P}(X_{1,*u,t}^i = i) - \frac{u(i)}{u(i-1)} \mathbb{P}(X_{1,*u,t}^i = i-1) \right) dt. \end{aligned}$$

Similarly,

$$\partial g_{\mathbf{1}_{[0,i]}}(i) \leq \int_0^{\infty} e^{-\sigma(1,*u)t} \left(\mathbb{P}(X_{1,*u,t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1,*u,t}^i = i+1) \right) dt.$$

We get the conclusion by Lemma 6.6.9. The proof is analogous under $\mathbf{H_2}$, using this time equation (6.23) and Theorem 6.2.5. □

Finally, Theorem 6.3.10 and 6.3.11 are simple consequences of the previous lemma.

6.7 Proof of Section 6.4

The second upper bound of Lemma 6.4.1 derives by classical arguments from Mehler's formula (6.16) and the following lemma.

Lemma 6.7.1 (Upper bound on differences of the pointwise probabilities of the Poisson distribution)**.**

$$\sup_{x\in\mathbb{N}} |\mathcal{P}_\lambda(x) - \mathcal{P}_\lambda(x-1)| \leq 1 \wedge \frac{C}{\lambda}, \qquad C := \frac{1}{\sqrt{2\pi}} e^{\frac{1}{\sqrt{2}}} \leq 1. \tag{6.32}$$

Proof of Lemma 6.7.1. Set

$$q(\lambda, x) = \lambda |\mathcal{P}_\lambda(x) - \mathcal{P}_\lambda(x-1)|, \qquad x \in \mathbb{N}, \quad \lambda > 0.$$

Let us show that

$$\sup_{x\in\mathbb{N},\,\lambda>0} q(\lambda, x) < +\infty.$$

Firstly,

$$q(\lambda, x) = \frac{\lambda^x e^{-\lambda}}{x!} |\lambda - x|, \qquad x \in \mathbb{N}, \quad \lambda > 0.$$

We first deal with the case where $x \in \mathbb{N}^*$. By a formula of Robbins (Robbins [1955]), we know that for all $x \in \mathbb{N}^*$,

$$x! > \sqrt{2\pi x}\, x^x\, e^{-x+\frac{1}{12x}} \geq \sqrt{2\pi}\, e^{\frac{1}{2}\log x + x\log x - x}.$$

Hence, $q(\lambda, x) \leq \frac{1}{\sqrt{2\pi}} e^{f(\lambda,x)}$ with

$$\begin{aligned} f(\lambda, x) &= x - \lambda + \log|x-\lambda| - \frac{1}{2}\log x + x\log\frac{\lambda}{x}, \\ \partial_\lambda f(\lambda, x) &= -1 + \frac{x}{\lambda} + \frac{1}{\lambda - x}. \end{aligned}$$

In the sequel we derive upper bounds of $f(\lambda, x)$ on relevant subsets of $(0,\infty)\times[1,\infty)$. One finds that

$$\partial_\lambda f(\lambda, x) = 0 \Leftrightarrow (\lambda - x)(x - \lambda + \sqrt{\lambda})(x - \lambda - \sqrt{\lambda}) = 0.$$

Let us call $\lambda_1(x)$ the solution of the equation $x = \lambda + \sqrt{\lambda}$ and $\lambda_2(x)$ the solution of the equation $x = \lambda - \sqrt{\lambda}$. We have $0 < \lambda_1(x) < x < \lambda_2(x)$.
If $\lambda \leq x$, then at x fixed the function $f(\lambda, x)$ is increasing on $(0, \lambda_1(x)]$ and decreasing on $[\lambda_1(x), x]$. Hence,

$$\sup_{x\geq1,\,0<\lambda\leq x} f(\lambda, x) = \sup_{x\geq 1} f(\lambda_1(x), x) = \sup_{\lambda>0} f(\lambda, \lambda+\sqrt{\lambda}).$$

Moreover, using that $\forall u \geq 0$, $\log(1+u) \geq u - u^2/2$, we find that

$$\begin{aligned} f(\lambda, \lambda+\sqrt{\lambda}) &= \sqrt{\lambda} + \frac{1}{2}\log\frac{\lambda}{\lambda+\sqrt{\lambda}} + (\lambda+\sqrt{\lambda})\log\frac{\lambda}{\lambda+\sqrt{\lambda}} \\ &= \sqrt{\lambda} - (\lambda + \sqrt{\lambda} + \frac{1}{2})\log\left(1 + \frac{1}{\sqrt{\lambda}}\right) \\ &\leq \sqrt{\lambda} - (\lambda+\sqrt{\lambda}+\frac{1}{2})\left(\frac{1}{\sqrt{\lambda}} - \frac{1}{2\lambda}\right) \\ &= -\frac{1}{2}\left(1 - \frac{1}{2\lambda}\right). \end{aligned} \tag{6.33}$$

Hence, if $\lambda \geq \frac{1}{2}$ then $f(\lambda, \lambda+\sqrt{\lambda}) \leq 0$. If $\lambda \leq \frac{1}{2}$, by going back up to the equation (6.33),

$$f(\lambda, \lambda+\sqrt{\lambda}) \leq \sqrt{\lambda} \leq \frac{1}{\sqrt{2}}.$$

At the end,

$$\sup_{x \in \mathbb{N}^*,\, 0<\lambda \leq x} f(\lambda, x) \leq \frac{1}{\sqrt{2}}.$$

Let us call $C_1 := \frac{1}{\sqrt{2\pi}} e^{\frac{1}{\sqrt{2}}} \sim 0,8$.
Now let us deal with the case where $\lambda \geq x$. We apply a different strategy for small integers x as for large integers x. First of all, for all $x \in \mathbb{N}^*$ and for all $\lambda \geq x$,

$$q(\lambda, x) = \frac{1}{x!} e^{-\lambda} \lambda^x (\lambda - x) \leq \frac{1}{x!} e^{-\lambda} \lambda^{x+1}$$

and it is easy to see that at x fixed the maximum of the right-hand expression is attained at $\lambda = x+1$. Hence

$$q(\lambda, x) \leq \frac{1}{x!} e^{-(x+1)} (x+1)^{x+1}.$$

Hence

$$\sup_{x \in \{1,2,3\},\, \lambda \geq x} q(\lambda, x) \leq C_2 := \max_{x \in \{1,2,3\}} \frac{1}{x!} e^{-(x+1)} (x+1)^{x+1} \sim 0.7.$$

On the other hand, by the same reasoning as below, we find that

$$\sup_{x \in [4,+\infty),\, \lambda \geq x} f(\lambda, x) = \sup_{x \in [4,+\infty)} f(\lambda_2(x), x) = \sup_{\lambda \geq 4} f(\lambda, \lambda - \sqrt{\lambda}).$$

Now, for all $\lambda > 1$,

$$f(\lambda, \lambda - \sqrt{\lambda}) = -\sqrt{\lambda} - \left(\frac{1}{2} + \lambda - \sqrt{\lambda}\right) \log\left(1 - \frac{1}{\sqrt{\lambda}}\right).$$

We use that $\forall u \in [0, \frac{1}{2}], -\log(1-u) \leq u + u^2$. As $\lambda \geq 4$ implies $\frac{1}{\sqrt{\lambda}} \leq \frac{1}{2}$,

$$\begin{aligned} f(\lambda, \lambda - \sqrt{\lambda}) &\leq -\sqrt{\lambda} + \left(\frac{1}{2} + \lambda - \sqrt{\lambda}\right)\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\lambda}\right) \\ &= -\frac{1}{2\sqrt{\lambda}}\left(1 - \frac{1}{\sqrt{\lambda}}\right) \\ &\leq 0. \end{aligned}$$

At the end,

$$\sup_{\lambda \geq x \geq 4} f(\lambda, x) \leq C_3 := \frac{1}{\sqrt{2\pi}} \sim 0.4.$$

It remains the case where $x = 0$, for which it is trivial to see that

$$q(\lambda, 0) = \lambda e^{-\lambda} \leq C_4 = e^{-1}, \qquad\qquad \lambda > 0.$$

The final result follows with $C = \max\{C_1, C_2, C_3, C_4\} = C_1$. $\square$

Chapitre 7

Quelques perspectives

Convergence asymptotique de processus de Markov inhomogènes

Autour du théorème central limite

Une perspective naturelle faisant suite aux résultats obtenus dans le chapitre 3 est d'affaiblir les hypothèses des théorèmes 3.1.1 et 3.1.2. Typiquement, on aimerait se passer de l'hypothèse selon laquelle la densité φ est polynomiale et de l'hypothèse **(H)**, qui intervient de manière cruciale dans la majoration du rayon spectral de l'opérateur de transition. Une piste à explorer vient de la remarque 3.2.1, d'après laquelle, pour des polynômes de faible degré, la condition de positivité est suffisante pour obtenir le résultat souhaité. Pour une densité quelconque φ, la positivité se traduit, dans le domaine d'Hermite-Fourier, par la positivité d'une certaine matrice symétrique faisant intervenir les coefficients $(\varphi_k)_{k\in\mathbb{N}}$. Il s'agit de voir si cette information supplémentaire permet de majorer le rayon spectral de la matrice associée à l'opérateur de transition.

Par ailleurs, il serait intéressant de voir si l'on peut adapter les techniques de la littérature concernant les bornes à la Berry-Esseen pour l'entropie, Bobkov et al. [2013] par exemple, à la distance du χ_2, et comparer avec les résultats obtenus dans le chapitre 3.

Enfin, on remarque que la suite $(Y_n)_{n\geq 1}$, constituée des sommes des $(X_i)_{i\in\mathbb{N}^*}$ renormalisées, est un processus auto-régressif inhomogène d'ordre 1, puisque

$$Y_{n+1} = \alpha_{n+1}Y_n + \beta_{n+1}X_{n+1}, \qquad n \in \mathbb{N}^*.$$

On peut envisager d'adapter la méthode du chapitre 3 pour donner des conditions sur $(\alpha_n)_{n\in\mathbb{N}^*}$ et $(\beta_n)_{n\in\mathbb{N}^*}$ permettat d'obtenir une majoration de la distance du χ_2 entre la loi de $(Y_n)_{n\in\mathbb{N}^*}$ et la loi normale μ. On aimerait également étendre les résultats du chapitre 3 à des variables aléatoires $(X_i)_{i\in\mathbb{N}^*}$ non indépendantes ; par exemple, supposer que la suite $(X_i)_{i\in\mathbb{N}^*}$ elle-même est un processus auto-régressif stationnaire centré réduit d'ordre 1, obéissant à la relation (Brockwell and Davis [1991])

$$X_{n+1} = c_1 + c_2X_n + \epsilon_n, \qquad n \in \mathbb{N}^*,$$

avec $(\epsilon_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes identiquement distribuées. Alors, le couple $(X_n, Y_n)_{n\in\mathbb{N}^*}$ est une chaîne de Markov inhomogène, à laquelle on peut envisager d'étendre la méthode développée dans le chapitre 3.

Autour de l'algorithme du quantile

La première chose à faire est bien sûr d'établir ou de réfuter la conjecture 4.1.3. Pour ce faire, on cherche à démontrer dans un premier temps l'inégalité (4.9). Soit $\mathcal{H}_a^N(\mu)$ le sous-espace de Hilbert formé des fonctions

$$f = \sum_{n=0}^{N} f_n\xi^n,$$

où les fonctions $(f_n)_{0\leq n\leq N}$ sont dans $L^2(\mu)$ et a-périodique. Soit $S_a^N(\mu)$ la boule unité de $\mathcal{H}_a^N(\mu)$. Une option est d'utiliser la méthode des multiplicateurs de Lagrange (Zeidler [1995]) pour déterminer les extrema de l'application

$$S_a^N(\mu) \to \mathbb{R}, \qquad f \mapsto \int (f - \mu_a[f])\, \xi\, \nabla \mu_a[f] d\mu.$$

Si ce point technique est réglé, les pistes de généralisation sont nombreuses :

— Passer de sauts symétriques, de la forme $\pm a_t$ à $t \geq 0$ donné, à des sauts asymétriques $+a_t, -b_t$. La difficulté est que le processus homogène correspondant, qui saute en $+a$, $-b$ pour des valeurs fixes de $a, b > 0$, ne peut plus être interprété comme un processus de naissance-mort sur un sous-domaine de $\mathbb{R}$; mais l'intérêt est que le processus inhomogène correspond alors à la version continue d'un algorithme stochastique de recherche de quantile de niveau α, et pas seulement de médiane.

— Le processus envisagé correspond, si l'on ne tient pas compte d'un terme négligeable, à une version continue de l'algorithme de la médiane renormalisé, pour des pas de la forme $\gamma_n = cn^{-\beta}$, avec $0 < \beta < 1$. Or, le cas $\beta = 1$ est précisément celui qui est employé en pratique. Dans ce cas, la version continue évolue sous l'action d'une transformation affine : du point z, le processus peut sauter au point $\theta_t z + a_t$ ou $\theta_t z - a_t$, avec $\theta_t > 1, a_t > 0$ au temps $t \geq 0$.

— Enfin, le processus considéré est une version continue de l'algorithme du quantile renormalisé, obtenue par poissonisation. Pour revenir à la version originale en temps discret, on pourra appliquer des idées similaires que dans le cas continu, en symétrisant de manière multiplicative l'opérateur de transition $Q_{n,n+1}$ comme expliqué dans la partie 2.

Par ailleurs, d'après un problème posé par Luc Pronzato, supposons que dans l'algorithme du quantile,

$$Y_{n+1} = Y_n + \gamma_{n+1}(\mathbf{1}_{X_{n+1}>Y_n} - \alpha), \qquad n \geq 1,$$

les variables aléatoires $(X_i)_{i\in\mathbb{N}^*}$ ne soient pas tirées de manière identiquement distribuées suivant une mesure de probabilité ν, mais qu'on sache seulement que la distribution des $(X_i)_{i\in\mathbb{N}^*}$ converge vers ν dans un sens à préciser. Cette situation arrive en pratique de manière naturelle si les données $(X_i)_{i\in\mathbb{N}^*}$ sont issues d'une simulation (Pronzato [2010]). Les expériences numériques semblent montrer que l'algorithme $(Y_n)_{n\in\mathbb{N}^*}$ converge toujours vers le quantile de ν. Un objectif de recherche naturel est de montrer cette convergence, de manière non quantitative dans un premier temps, via une approche tension/ identification.

Autour d'un algorithme stochastique d'estimation d'intégrale

Soit μ la mesure invariante d'une diffusion de générateur défini, pour toute fonction f assez régulière, par

$$L[g] = \frac{1}{2}\sigma^2 g'' + bg',$$

où σ^2 est une fonction positive. Le schéma d'Euler à pas décroissant associé s'écrit :

$$X_{n+1} = X_n + \gamma_{n+1} b(X_n) + \sqrt{\gamma_{n+1}}\sigma(X_n)U_{n+1},$$

avec $(U_n)_{n\geq 1}$ bruit blanc gaussien. Sous les hypothèses appropriées sur les fonctions b et σ et sur la séquence des pas $(\gamma_n)_{n\geq 1}$, on sait que cet algorithme converge vers la mesure invariante μ de $(Y_t)_{t\geq 0}$, ce qui fournit un moyen d'estimer numériquement l'intégrale $\mu(f)$ là où d'autres méthodes comme celles de Monte-Carlo peuvent s'avérer plus coûteuses (Lamberton and Pagès [2002]; Lemaire [2005]).

L'algorithme $(Z_n)_{n\geq 1}$ forme une chaîne de Markov inhomogène, dont la loi converge asymptotiquement vers μ. Une piste de recherche est d'appliquer la stratégie par inégalité fonctionnelle développée dans cette thèse dans le but de quantifier la convergence, sous l'hypothèse que le couple (L,μ) lui-même admet une inégalité de Poincaré ou de Log-Sobolev par exemple, où L désigne le générateur infinitésimal du processus $(Y_t)_{t\geq 0}$.

Dans le chapitre 2, les ingrédients nécessaires au calcul ont été donnés pour le processus d'Ornstein-Uhlenbeck et pour le choix de l'entropie relative comme quantité mesurant la convergence. La généralisation est un problème ouvert.

Entrelacements et méthode de Stein

Une première piste à explorer consiste à obtenir une formule d'entrelacement valable à tout ordre, ce qui permet de proposer des majorations sur les facteurs de Stein correspondants.

La seconde piste de recherche concerne l'adaptation des résultats principaux du chapitre 6 au cas continu, c'est-à-dire lorsque le semi-groupe de Markov considéré correspond à un semi-groupe de diffusion. L'entrelacement au premier ordre existe (Bonnefont and Joulin [2014]). L'entrelacement au second ordre reste à examiner. Soit Δ le laplacien sur $\mathbb{R}$, défini comme $\Delta f = f''$, L un générateur de diffusion et f une fonction régulière de $\mathbb{R}$ dans $\mathbb{R}$. Pour obtenir l'entrelacement au second ordre, le point crucial est de pouvoir interpréter l'expression

$$\Delta(Lf)$$

comme un générateur de Feynman-Kac $\widehat{L}-\widehat{V}$ appliqué à Δf. Or, par l'entrelacement au second ordre, on sait que

$$\Delta(Lf) = \nabla(\widetilde{L}-\widetilde{V})(\nabla f).$$

Il suffit donc d'obtenir un entrelacement au premier ordre, mais pour le générateur de Feynman-Kac $(\widetilde{L}-\widetilde{V})$. Ecrivons :

$$\nabla(\widetilde{L}-\widetilde{V})(g) = \nabla(\widetilde{L}g) - \widetilde{V}(\nabla g) - (\nabla\widetilde{V})g.$$

En réitérant l'entrelacement au premier ordre, le terme $\nabla(\widetilde{L}g)$ ne pose pas de problème, pas plus que le terme $\widetilde{V}(\nabla g)$. L'analyse doit se concentrer sur le terme $(\nabla\widetilde{V})g$, à exprimer en fonction de ∇g grâce à des conditions au bord bien choisies.

Enfin, une stratégie similaire à celle employée dans le chapitre 6 permettrait de dériver des facteurs de Stein relatifs à des mesures continues sur $\mathbb{R}$, de manière plus large que pour la mesure gaussienne qui est l'approximande classique.

Bibliographie

Joseph Abate, Masaaki Kijima, and Ward Whitt. Decompositions of the $M/M/1$ transition function. *Queueing Systems Theory Appl.*, 9(3) :323–336, 1991.

Aboubaca Amiri and Baba Thiam. A smoothing stochastic algorithm for quantile estimation. *Statistics and Probability Letters*, 93 :116–125, 2014.

Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses*. Société Mathématique de France, Paris, 2000.

Marc Arnaudon and Laurent Miclo. A stochastic algorithm finding p-means on the circle. *Bernoulli*, 22(4) :2237–2300, 2016.

R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations : the Chen-Stein method. *Ann. Probab.*, 17(1) :9–25, 1989.

Richard Arratia, Larry Goldstein, and Louis Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4) :403–434, 1990.

Shiri Artstein, Keith M Ball, Franck Barthe, and Assaf Naor. On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields*, 129(3) :381–390, 2004.

François Baccelli and William A. Massey. A sample path analysis of the $M/M/1$ queue. *J. Appl. Probab.*, 26(2) :418–422, 1989.

Francis Bach and Eric Moulines. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11, pages 451–459. Curran Associates Inc., 2011.

D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

Vlad Bally and Lucia Caramellino. Asymptotic development for the CLT in total variation distance. *Bernoulli*, 22(4) :2442–2485, 2016.

A. D. Barbour. Stein's method for diffusion approximations. *Probab. Theory Related Fields*, 84 (3) :297–322, 1990.

A. D. Barbour and T. C. Brown. Stein's method and point process approximation. *Stochastic Process. Appl.*, 43(1) :9–31, 1992.

A. D. Barbour and G. K. Eagleson. Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.*, 15(3) :585–600, 1983.

A. D. Barbour and Peter Hall. On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.*, 95(3) :473–480, 1984.

A. D. Barbour and Aihua Xia. On Stein's factors for Poisson approximation in Wasserstein distance. *Bernoulli*, 12(6) :943–954, 2006.

A. D. Barbour, Lars Holst, and Svante Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.

A. D. Barbour, H. L. Gan, and A. Xia. Stein factors for negative binomial approximation in Wasserstein distance. *Bernoulli*, 21(2) :1002–1013, 2015.

Andrew R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14(1) :336–342, 1986.

Michel Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.

Michel Benaïm, Florian Bouguet, and Bertrand Cloez. Ergodicity of inhomogeneous Markov chains through asymptotic pseudotrajectories. *Ann. Appl. Probab.*, 2016.

Albert Benveniste, Michel Métivier, and Pierre Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.

Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49 :122–136, 1941.

Julius R. Blum. Approximation methods which converge with probability one. *Ann. Math. Statistics*, 25 :382–386, 1954.

Sergey G. Bobkov, Gennadiy P. Chistyakov, and Friedrich Götze. Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.*, 41(4), 2013.

Sergey G. Bobkov, Gennadiy P. Chistyakov, and Friedrich Götze. Fisher information and the central limit theorem. *Probability Theory and Related Fields*, 159(1-2) :1–59, 2014.

Michel Bonnefont and Aldéric Joulin. Intertwining relations for one-dimensional diffusions and application to functional inequalities. *Potential Anal.*, 41(4) :1005–1031, 2014.

Peter J. Brockwell and Richard A. Davis. *Time series : theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

Lawrence D. Brown. A proof of the central limit theorem motivated by the Cramér-Rao inequality. In *Statistics and probability : essays in honor of C. R. Rao*, pages 141–148. North-Holland, Amsterdam-New York, 1982.

Timothy C. Brown and M. J. Phillips. Negative binomial approximation with Stein's method. *Methodol. Comput. Appl. Probab.*, 1(4) :407–421, 1999.

Timothy C. Brown and Aihua Xia. Stein's method and birth-death processes. *Ann. Probab.*, 29 (3) :1373–1403, 2001.

H. Cardot, P. Cénac, and A. Godichon. Online estimation of the geometric median in Hilbert spaces : non asymptotic confidence balls. *ArXiv e-prints*, 2015.

Djalil Chafaï and Aldéric Joulin. Intertwining and commutation relations for birth-death processes. *Bernoulli*, 19(5A) :1855–1879, 2013.

Sourav Chatterjee, Persi Diaconis, and Elizabeth Meckes. Exchangeable pairs and Poisson approximation. *Probab. Surv.*, 2 :64–106, 2005.

Louis H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability*, 3(3) :534–545, 1975.

Mu-Fa Chen. Estimation of spectral gap for Markov chains. *Acta Math. Sinica (N.S.)*, 12(4) : 337–360, 1996.

Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems.* World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.

Mu-Fa Chen and Feng-Yu Wang. Estimation of spectral gap for elliptic operators. *Trans. Amer. Math. Soc.*, 349(3) :1239–1267, 1997.

Nicolas Chenavier. A general study of extremes of stationary tessellations with examples. *Stochastic Process. Appl.*, 124(9) :2917–2953, 2014.

Bertrand Cloez and Claire Delplancke. Intertwinings and Stein's magic factors for birth-death processes. *arXiv preprint arXiv :1609.08390*, 2016.

T. A. Courtade, M. Fathi, and A. Pananjady. Existence of Stein kernels under a spectral gap, and discrepancy bound. *ArXiv e-print arXiv :1703.07707*, 2017.

Fraser Daly. Upper bounds for Stein-type operators. *Electron. J. Probab.*, 13 :no. 20, 566–587, 2008.

H. A. David. *Order statistics.* John Wiley & Sons, Inc., New York-London-Sydney, 1970.

Jérôme Dedecker, Florence Merlevède, and Emmanuel Rio. Rates of convergence for minimal distances in the central limit theorem under projective criteria. *Electron. J. Probab.*, 14 :no. 35, 978–1011, 2009.

Pierre Del Moral and Alice Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(2) :155–194, 2001.

Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1) :36–61, 1991.

Persi Diaconis and Sandy Zabell. Closed form summation for classical distributions : variations on a theme of de Moivre. *Statist. Sci.*, 6(3) :284–302, 1991.

Christian Döbler, Robert E. Gaunt, and Sebastian J. Vollmer. An iterative technique for bounding derivatives of solutions of Stein equations. *arXiv preprint arXiv :1510.02623*, 2015.

R. Douc, E. Moulines, and Jeffrey S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.*, 14(4) :1643–1665, 2004.

Marie Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1997. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

Andreas Eberle. Markov processes. *Lecture Notes at University of Bonn*, 2009.

Peter Eichelsbacher and Gesine Reinert. Stein's method for discrete Gibbs measures. *Ann. Appl. Probab.*, 18(4) :1588–1618, 2008.

Carl-Gustav Esseen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.*, 77 :1–125, 1945.

Stewart N. Ethier and Thomas G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1(1) :62–87, 1991.

Noufel Frikha and Stéphane Menozzi. Concentration bounds for stochastic approximations. *Electron. Commun. Probab.*, 17 :no. 47, 15, 2012. ISSN 1083-589X.

S. Gadat, I. Gavra, and L. Risser. How to compute the barycenter of a weighted graph. *ArXiv e-prints*, 2016.

V. F. Gapoškin and T. P. Krasulina. The law of the iterated logarithm in stochastic approximation processes. *Teor. Verojatnost. i Primenen.*, 19 :879–886, 1974.

Semyon Gershgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Bulletin de l'Académie des Sciences de l'URSS*, (6), 1931.

Antoine Godichon-Baggioni. Estimating the geometric median in Hilbert spaces with stochastic gradient algorithms : L^p and almost sure rates of convergence. *J. Multivariate Anal.*, 146 : 209–222, 2016.

Larry Goldstein and Yosef Rinott. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, 33(1) :1–17, 1996.

F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2) :724–739, 1991.

Thierry Goudon, Stéphane Junca, and Giuseppe Toscani. Fourier-based distances and Berry-Esseen like inequalities for smooth densities. *Monatshefte für Mathematik*, 135(2) :115–136, 2002.

J. L. Hodges, Jr. and Lucien Le Cam. The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.*, 31 :737–740, 1960.

Susan Holmes. Stein's method for birth and death chains. In *Stein's method : expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 45–67. Inst. Math. Statist., Beachwood, OH, 2004.

U. Holst. Recursive estimation of quantiles using recursive kernel density estimators. *Sequential Anal.*, 6(3) :219–237, 1987.

I. A. Ibragimov. On the accuracy of approximation by the normal distribution of distribution functions of sums of independent random variables. *Teor. Verojatnost. i Primenen*, 11 :632–655, 1966.

David G. Kendall. On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, 35 :6–15, 1948.

Harold J. Kushner and Dean S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences.* Springer-Verlag, New York-Berlin, 1978.

Tatiana Labopin-Richard. *Méthodes statistiques et d'apprentissage pour l'estimation de quantiles et de superquantiles dans des modèles de codes numériques ou stochastiques.* Thèse de doctorat, Université Toulouse 3 Paul Sabatier, 2016.

Damien Lamberton and Gilles Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3) :367–405, 2002.

Lucien Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.*, 10 :1181–1197, 1960.

Vincent Lemaire. *Estimation numérique de la mesure invariante d'un processus de diffusion.* Thèse de doctorat, Université Marne La Vallée, 2005.

Christophe Ley, Gesine Reinert, and Yvik Swan. Approximate computation of expectations : a canonical Stein operator. *arXiv preprint arXiv :1408.2998*, 2014.

Torgny Lindvall. *Lectures on the coupling method.* Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.

Laurent Miclo. An example of application of discrete Hardy's inequalities. *Markov Process. Related Fields*, 5(3) :319–330, 1999.

È. A. Nadaraja. Some new estimates for distribution functions. *Teor. Verojatnost. i Primenen.*, 9 :550–554, 1964.

Ivan Nourdin and Giovanni Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2) :75–118, 2009.

Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus : from Stein's method to universality*, volume 192. Cambridge University Press, 2012.

Ivan Nourdin and Guillaume Poly. Convergence in total variation on Wiener chaos. *Stochastic Process. Appl.*, 123(2) :651–674, 2013.

David Nualart and Josep Vives. Anticipative calculus for the Poisson process based on the Fock space. In *Séminaire de Probabilités, XXIV, 1988/89*, volume 1426 of *Lecture Notes in Math.*, pages 154–165. Springer, Berlin, 1990.

G. Peccati. The Chen-Stein method for Poisson functionals. *ArXiv e-prints*, 2011.

G. Peccati, J. L. Solé, M. S. Taqqu, and F. Utzet. Stein's method and normal approximation of Poisson functionals. *Ann. Probab.*, 38(2) :443–478, 2010.

Erol A. Peköz. Stein's method for geometric approximation. *J. Appl. Probab.*, 33(3) :707–713, 1996.

Erol A. Peköz, Adrian Röllin, and Nathan Ross. Total variation error bounds for geometric approximation. *Bernoulli*, 19(2) :610–632, 2013.

M. J. Phillips. *Stochastic process approximation and network applications.* PhD thesis, Univ. Melbourne, 1996.

M. J. Phillips and Graham V. Weinberg. Non-uniform bounds for geometric approximation. *Statist. Probab. Lett.*, 49(3) :305–311, 2000.

Luc Pronzato. Penalized optimal designs for drug finding. *Journal of Statistical Planning and Inference*, 140, 2010.

Michael Reed and Barry Simon. *Methods of modern mathematical physics. IV. Analysis of operators.* Academic Press, New York-London, 1978.

Emmanuel Rio. Distances minimales et distances idéales. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(9) :1127–1130, 1998.

Emmanuel Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3) :802–817, 2009.

Emmanuel Rio. Asymptotic constants for minimal distance in the central limit theorem. *Electron. Commun. Probab.*, 16 :96–103, 2011.

Herbert Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62 :26–29, 1955.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22, 1951.

Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8 :210–293, 2011.

L. Saloff-Coste and J. Zúñiga. Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic Processes and their Applications*, 117(8) :961 – 979, 2007.

Laurent Saloff-Coste and Jessica Zúñiga. Merging and stability for time inhomogeneous finite Markov chains. In *Surveys in stochastic processes*, EMS Ser. Congr. Rep., pages 127–151. Eur. Math. Soc., Zürich, 2011.

Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana, Ill., 1949.

S. H. Siraždinov and M. Mamatov. On mean convergence for densities. *Teor. Verojatnost. i Primenen.*, 7 :433–437, 1962.

Richard L. Smith. Extreme value theory for dependent sequences via the Stein-Chen method of Poisson approximation. *Stochastic Process. Appl.*, 30(2) :317–327, 1988.

A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2, 1959.

Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II : Probability theory*, pages 583–602. Univ. California Press, Berkeley, Calif., 1972.

Charles Stein. *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

A. Szulga. On minimal metrics in the space of random variables. *Teor. Veroyatnost. i Primenen.*, 27(2) :401–405, 1982.

Hiroshi Tanaka. An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 27 :47–52, 1973.

Michael Woodroofe. Normal approximation and large deviations for the Robbins-Monro process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 21 :329–338, 1972.

Eberhard Zeidler. *Applied functional analysis*, volume 109 of *Applied Mathematical Sciences.* Springer-Verlag, New York, 1995. Main principles and their applications.

V. M. Zolotarev. Metric distances in spaces of random variables and their distributions. *Mat. Sb. (N.S.)*, 101(143)(3) :416–454, 456, 1976.

Méthodes quantitatives pour l'étude asymptotique de processus de Markov homogènes et non-homogènes

L'objet de cette thèse est l'étude de certaines propriétés analytiques et asymptotiques des **processus de Markov**, et de leurs applications à la **méthode de Stein**. Le point de vue considéré consiste à déployer des **inégalités fonctionnelles** pour majorer la distance entre lois de probabilité.

La première partie porte sur l'étude asymptotique de **processus de Markov inhomogènes en temps** via des inégalités de type Poincaré. On se place d'abord dans le cadre du théorème central limite, qui affirme que la somme renormalisée de variables aléatoires converge vers la mesure gaussienne. L'étude est consacrée à l'obtention d'une **borne à la Berry-Esseen** en distance du χ^2 permettant de quantifier cette convergence, et complétant la littérature relative à d'autres distances.

Toujours dans le contexte non-homogène, on s'intéresse ensuite à un processus peu mélangeant relié à un **algorithme stochastique** de recherche de médiane, qui évolue par des sauts dont la taille et l'intensité dépendent du temps, et on présente une conjecture sur la distance de Wasserstein d'ordre 1 entre la loi du processus et la mesure gaussienne.

La seconde partie s'attache à l'étude des entrelacements entre processus de Markov (homogènes) et gradients, qu'on peut interpréter comme un raffinement du critère de Bakry-Emery, et leur application à la **méthode de Stein**, un ensemble de techniques permettant de majorer la distance entre mesures de probabilité. On prouve l'existence de relations d'entrelacement du second ordre pour les **processus de naissance-mort**, allant ainsi plus loin que les relations du premier ordre connues. Ces relations sont mises à profit pour construire une méthode originale et universelle d'évaluation des facteurs de Stein relatifs aux mesures de probabilité discrètes.

Mots-clefs : processus de Markov, processus de naissance-mort, inégalités fonctionnelles, inégalités de Berry-Esseen, algorithmes stochastiques, méthode de Stein.

Quantitative methods for the asymptotic study of homogeneous and non-homogeneous Markov processes

The object of this thesis is the study of some analytical and asymptotic properties of **Markov processes**, and their applications to **Stein's method**. The point of view consists in the development of **functional inequalities** in order to obtain upper-bounds on the distance between probability distributions.

The first part is devoted to the asymptotic study of **time-inhomogeneous Markov processes** through Poincaré-like inequalities. The first investigation takes place within the framework of the Central Limit Theorem, which states the convergence of the renormalized sum of random variables towards the normal distribution. It results in the statement of a **Berry-Esseen bound** allowing to quantify this convergence with respect to the χ^2-distance, which extends similar results relative to other distances.

Keeping with the non-homogeneous framework, we consider a weakly mixing process linked to a **stochastic algorithm** for median approximation, which evolves by jumps with time-dependent size and intensity, and present a conjecture on the Wasserstein distance of order 1 between the marginal distribution of the process and the normal distribution.

The second part concerns intertwining relations between (homogeneous) Markov processes and gradients, which can be seen as refinment of the Bakry-Emery criterion, and their application to **Stein's method**, a collection of techniques to estimate the distance between probability distributions. Second order intertwinings for **birth-death processes** are stated, going one step further than the existing first order relations. These relations are then exploited to construct an original and universal method of evaluation of discrete Stein's factors.

Keywords: Markov process, birth-death process, functional inequalities, Berry-Esseen inequalities, stochastic algorithm, Stein method.