



HAL
open science

Estimation non paramétrique pour les processus markoviens déterministes par morceaux

Romain Azaïs

► **To cite this version:**

Romain Azaïs. Estimation non paramétrique pour les processus markoviens déterministes par morceaux. Mathématiques générales [math.GM]. Université Sciences et Technologies - Bordeaux I, 2013. Français. NNT : 2013BOR14796 . tel-00844395

HAL Id: tel-00844395

<https://theses.hal.science/tel-00844395v1>

Submitted on 15 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée à

L'UNIVERSITÉ BORDEAUX 1

École Doctorale de Mathématiques et Informatique de Bordeaux

pour obtenir le grade de DOCTEUR

Spécialité : MATHÉMATIQUES APPLIQUÉES

Romain AZAÏS

ESTIMATION NON PARAMÉTRIQUE POUR
LES PROCESSUS MARKOVIENS
DÉTERMINISTES PAR MORCEAUX

Directeurs de Recherche : François DUFOUR et Anne GÉGOUT-PETIT

Soutenue publiquement le 1^{er} Juillet 2013 devant la commission d'examen composée de

Mme Fabienne COMTE	<i>Université Paris Descartes</i>	Présidente du jury
M. François DUFOUR	<i>Institut Polytechnique de Bordeaux</i>	Directeur de thèse
Mme Anne GÉGOUT-PETIT	<i>Université Bordeaux Segalen</i>	Co-directrice de thèse
M. Davy PAINDAVEINE	<i>Université Libre de Bruxelles</i>	Rapporteur
Mme Patricia REYNAUD-BOURET	<i>CNRS & Université de Nice</i>	Rapporteuse
M. Jérôme SARACCO	<i>Institut Polytechnique de Bordeaux</i>	Examinateur

Numéro d'ordre : 4796

ESTIMATION NON PARAMÉTRIQUE POUR LES PROCESSUS MARKOVIENS DÉTERMINISTES PAR MORCEAUX

MOTS-CLÉS : Processus markoviens déterministes par morceaux ; Chaînes de Markov ergodiques ; Estimation non paramétrique ; Estimation de taux de saut ; Estimation de noyau de transition ; Régression semi-paramétrique.

RÉSUMÉ : M.H.A. Davis a introduit les processus markoviens déterministes par morceaux (PDMP) comme une classe générale de modèles stochastiques non diffusifs, donnant lieu à des trajectoires déterministes ponctuées, à des instants aléatoires, par des sauts aléatoires. Dans cette thèse, nous présentons et analysons des estimateurs non paramétriques des lois conditionnelles des deux aléas intervenant dans la dynamique de tels processus. Plus précisément, dans le cadre d'une observation en temps long de la trajectoire d'un PDMP, nous présentons des estimateurs de la densité conditionnelle des temps inter-sauts et du noyau de Markov qui gouverne la loi des sauts. Nous établissons des résultats de convergence pour nos estimateurs. Des simulations numériques pour différentes applications illustrent nos résultats. Nous proposons également un estimateur du taux de saut pour des processus de renouvellement, ainsi qu'une méthode d'approximation numérique pour un modèle de régression semi-paramétrique.

NONPARAMETRIC ESTIMATION FOR PIECEWISE-DETERMINISTIC MARKOV PROCESSES

KEYWORDS: Piecewise-deterministic Markov processes; Ergodic Markov chains; Nonparametric estimation; Jump rate estimation; Transition kernel estimation; Semiparametric regression.

ABSTRACT: Piecewise-deterministic Markov processes (PDMP's) have been introduced by M.H.A. Davis as a general family of non-diffusion stochastic models, involving deterministic motion punctuated by random jumps at random times. In this thesis, we propose and analyze nonparametric estimation methods for both the features governing the randomness of such a process. More precisely, we present estimators of the conditional density of the inter-jumping times and of the transition kernel for a PDMP observed within a long time interval. We establish some convergence results for both the proposed estimators. In addition, numerical simulations illustrate our theoretical results. Furthermore, we propose an estimator for the jump rate of a nonhomogeneous renewal process and a numerical approximation method based on optimal quantization for a semiparametric regression model.

CENTRE DE RECHERCHE INRIA BORDEAUX – SUD-OUEST
200, Avenue de la Vieille Tour,
33 405 TALENCE

INSTITUT DE MATHÉMATIQUES DE BORDEAUX
Université Bordeaux 1
351, Cours de la Libération,
33 405 TALENCE Cedex

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE DE BORDEAUX
U.F.R. de Mathématiques et Informatique
Bâtiment A33 – 1^{er} étage – Porte 112
351, Cours de la Libération
33 405 TALENCE Cedex

Conserva tus sueños, nunca sabes cuando te harán falta.¹

Carlos Ruiz Zafón, *La Sombra del Viento*

1. Garde tes rêves, tu ne sais jamais quand tu en auras besoin.

À mes parents, pour leur soutien sans faille.

REMERCIEMENTS

Je souhaite ici adresser mes remerciements sincères à mes directeurs de thèse, François Dufour et Anne Gégout-Petit, pour leur encadrement et leur soutien, et aussi pour la confiance qu'ils m'ont témoignée au cours des trois dernières années.

Je tiens également à exprimer ma reconnaissance à mes rapporteurs, Patricia Reynaud-Bouret et Davy Paindaveine, pour avoir rapporté ma thèse et pour l'intérêt qu'ils ont porté à mon travail. Je remercie aussi Fabienne Comte et Jérôme Saracco d'avoir accepté de participer à mon jury.

J'ai une pensée particulière pour la ville qui m'a accueilli il y a déjà cinq ans et que je m'apprête tristement à quitter. Chère Bordeaux, ton Cours du Chapeau rouge, ton Grand théâtre et ton Palais Gallien vont me manquer.

Enfin, tous mes proches qui estiment avoir leur place ici n'ont nul besoin d'être cités pour savoir que sans eux, ça n'aurait pas été pareil.

TABLE DES MATIÈRES

1	INTRODUCTION GÉNÉRALE	1
I	ESTIMATION NON PARAMÉTRIQUE D'UN TAUX DE SAUT	5
2	ESTIMATION POUR DES PROCESSUS DE RENOUVELLEMENT	7
2.1	Introduction	9
2.2	Définition et premiers résultats	12
2.3	Espace d'état discret	15
2.4	Espace d'état continu	17
2.4.1	Estimation de Λ	17
2.4.2	Estimation de λ	32
2.5	Illustration numérique	38
2.6	Conclusion du chapitre	40
2.7	Preuves de la section 2.2	41
2.7.1	Preuve de la proposition 2.1	41
2.7.2	Preuve du lemme 2.2	42
2.8	Preuves de la section 2.4	44
2.8.1	Preuve de la proposition 2.5	44
2.8.2	Preuve du lemme 2.7	44
2.8.3	Preuve du lemme 2.10	45
2.9	Application à des données biologiques	45
3	ESTIMATION POUR LES PDMP	51
3.1	Introduction	53
3.2	Formulation du problème	55
3.2.1	Définition du processus	55
3.2.2	Hypothèses	56
3.2.3	Résultats principaux	59

3.3	Considérations préliminaires	62
3.3.1	Loi conditionnelle de S_{n+1} sachant Z_n et Z_{n+1}	62
3.3.2	Ergodicité	67
3.4	Démonstrations des résultats principaux	69
3.4.1	Estimation de \tilde{l}	70
3.4.2	Approximation de f	75
3.4.3	Estimation de \tilde{H}	79
3.5	Simulations	83
3.5.1	Mouvement d'une bactérie	83
3.5.2	Un modèle de croissance-fragmentation	86
3.6	Conclusion du chapitre	88
3.7	Preuve de la proposition 3.29	89
II ESTIMATION D'UNE TRANSITION		93
4	QUANTIFICATION APPLIQUÉE À LA MÉTHODE SIR	95
4.1	Introduction	97
4.2	Quantification optimale en norme \mathbb{L}^p	98
4.2.1	Quantification d'un vecteur aléatoire	99
4.2.2	Quantification d'une chaîne de Markov	104
4.3	Étude d'un modèle non paramétrique	107
4.4	Étude du modèle semi-paramétrique	109
4.4.1	Approximation de la direction de β	109
4.4.2	Prédiction dans le cadre semi-paramétrique	111
4.5	Simulations	112
4.5.1	Approximation du sous-espace engendré par β	113
4.5.2	Approximation de la loi de Y sachant $\beta'X$	114
4.6	Conclusion du chapitre	115
4.7	Preuves des résultats des sections 4.3 et 4.4	120
4.7.1	Preuve de la proposition 4.9	120
4.7.2	Preuve de la proposition 4.12	121
4.7.3	Preuve de la proposition 4.11	123
4.7.4	Preuve du théorème 4.13	123
5	ESTIMATION DU NOYAU D'UN PDMP	129
5.1	Introduction	131

5.2	Considérations préliminaires	133
5.3	Estimation de la loi invariante de (Z_n^-)	135
5.3.1	Quelques propriétés de (Z_n^-)	136
5.3.2	Estimation de p	143
5.4	Estimation de la loi invariante de (Z_n^-, Z_n)	146
5.4.1	Quelques propriétés de (Z_n^-, Z_n)	146
5.4.2	Estimation de h	148
5.5	Illustration numérique	155
5.5.1	Premières simulations	155
5.5.2	Le processus TCP window size	157
5.6	Conclusion du chapitre	158
6	CONCLUSION ET PERSPECTIVES	159
	BIBLIOGRAPHIE	160

INTRODUCTION GÉNÉRALE

Les modèles probabilistes sont de plus en plus répandus et utilisés dans de nombreuses applications scientifiques, notamment en biologie, en fiabilité ou encore en finance. En effet, une modélisation stochastique permet de prendre en compte à la fois des phénomènes intrinsèquement aléatoires et notre manque de connaissance de systèmes déterministes.

Dans la grande famille des processus stochastiques à temps continu, on peut distinguer les propagations diffusives – régies par des équations différentielles stochastiques et étudiées à travers le calcul stochastique d'Itô, – et un groupement plus hétérogène de processus non diffusifs. Davis a proposé dans les années 1980 dans [31] d'unifier la classe des applications non diffusives sous le nom de processus markoviens déterministes par morceaux (piecewise-deterministic Markov processes, PDMP, en anglais). Il s'agit de dynamiques déterministes ponctuées à des instants aléatoires par des sauts aléatoires. Ils ont parallèlement été introduits par Jacobsen dans [56] mais sous une définition un peu différente.

On peut trouver dans la littérature récente de nombreux papiers qui traitent des propriétés des PDMP. Certaines recherches s'intéressent aux propriétés de stabilité et d'ergodicité des PDMP [25, 42, 47], au contrôle optimal [24, 26, 46], à des résultats asymptotiques de type loi des grands nombres ou grandes déviations [44, 45], ou encore à des méthodes d'approximation numérique [15, 16]. Les PDMP sont déjà présents dans de nombreux domaines comme la fiabilité [6, 34], la biologie [29, 54] ou les neurosciences [17, 49]. Leur caractère hybride permet de coupler des évolutions déterministes régies par des équations différentielles ou des EDP issues de la physique ou de la biologie à des perturbations aléatoires.

Dans ce contexte, il apparaît important de proposer des méthodes statistiques d'estimation pour cette classe de processus stochastiques. Les recherches dans cette direction sont encore peu conséquentes. On peut notamment citer les travaux très récents de Doumic *et al.* [38, 39] sur des méthodes non paramétriques d'estimation pour des modèles de type croissance-fragmentation. L'objectif de cette thèse est de construire des estimateurs pour les deux sources d'aléa intervenant dans la dynamique des PDMP, c'est-à-dire la loi

des durées inter-sauts et la loi des perturbations aléatoires. Le flot étant souvent donné par des équations déterministes de la physique ou de la biologie, nous ne nous sommes pas intéressés à son estimation. Sous le contrôle de François Dufour et Anne Gégout-Petit, j'ai choisi de travailler dans un cadre non paramétrique et pour des processus observés en temps long. Dans la suite de cette introduction, nous présentons les PDMP tels que définis dans [31], puis nous donnons l'organisation de ce manuscrit.

Un PDMP (X_t) est défini comme un processus hybride de la forme $X_t = (m_t, \xi_t)$, où le mode m_t est une variable discrète alors que ξ_t est une variable euclidienne. La quantité euclidienne ξ_t modélise la ou les variables d'intérêt qui peuvent être la position, la vitesse ou encore la taille de la cellule. Le mode correspond lui à l'ambiance, l'environnement ou le régime dans lequel évoluent les variables euclidiennes. L'espace d'état du mode M est fini ou dénombrable. Pour $m \in M$, les variables euclidiennes évoluent dans E_m un ouvert de \mathbb{R}^d . L'espace d'état du processus est donc

$$E = \{(m, \xi) : m \in M, \xi \in E_m\}.$$

La dynamique du PDMP est régie par trois caractéristiques locales dépendant du mode $m \in M$: le flot déterministe Φ_m , le taux de saut λ_m qui gouverne la loi des temps inter-sauts et le noyau de transition Q_m qui régit la loi des sauts.

- Pour tout $m \in M$, $\Phi_m : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$ est le flot dans le mode m . Φ_m peut être défini comme solution d'une équation différentielle ; par conséquent, $\Phi_m(\cdot, t)$ est un homéomorphisme, et l'application $t \mapsto \Phi_m(\cdot, t)$ est un semi-groupe :

$$\forall \xi \in \mathbb{R}^d, \Phi_m(\xi, t + s) = \Phi_m(\Phi_m(\xi, s), t).$$

Pour tout $x = (m, \xi) \in E$, le temps de sortie déterministe de E est défini comme

$$t^*(x) = \inf \{t > 0 : \Phi_m(\xi, t) \notin E\},$$

avec la convention habituelle $\inf \emptyset = +\infty$.

- Pour tout $m \in M$, le taux de saut $\lambda_m : \bar{E}_m \rightarrow \mathbb{R}_+$ est une fonction mesurable qui vérifie,

$$\forall (m, \xi) \in E, \exists \varepsilon > 0, \int_0^\varepsilon \lambda_m(\Phi_m(\xi, t)) dt < +\infty.$$

Grâce à cette condition, on évite que le processus ne saute immédiatement.

- Pour tout $m \in M$, Q_m est un noyau de Markov sur $(\bar{E}_m, \mathcal{B}(\bar{E}_m))$ qui vérifie,

$$\forall \xi \in \bar{E}_m, Q_m(\xi, E \setminus \{(m, \xi)\}) = 1.$$

Cela impose que le processus ne saute pas sur place : le mode et/ou la variable euclidienne sont modifiés lors d'un saut.

A partir de ces caractéristiques, on définit la dynamique de (X_t) sur un espace de probabilité $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}_x)$ de la façon suivante. Partant de $x = (m, \xi) \in E$, on définit le premier temps de saut T_1 à partir de sa fonction de survie,

$$\forall t \geq 0, \mathbb{P}_x(T_1 > t) = \exp\left(-\int_0^t \lambda_m(\Phi_m(\xi, s)) ds\right) \mathbb{1}_{\{0 \leq t < t^*(x)\}}.$$

T_1 intervient soit spontanément avec le taux λ_m pris en le flot $\Phi_m(\xi, \cdot)$, soit lorsque le processus atteint la frontière de son espace de définition. On génère alors une variable aléatoire Z_1 à valeurs dans E selon le noyau de Markov $Q_m(\Phi_m(\xi, T_1), \cdot)$. La trajectoire de (X_t) entre les instants 0 et T_1 est donnée par

$$X_t = \begin{cases} (m, \Phi_m(\xi, t)) & \text{pour } t < T_1, \\ Z_1 & \text{pour } t = T_1. \end{cases}$$

Partant de $X_{T_1} = Z_1$, on détermine de manière similaire la durée $S_2 = T_2 - T_1$, ainsi que la position après le deuxième saut Z_2 , et ainsi de suite. On obtient ainsi un processus de Markov fort à temps continu (voir [31, Section 25]), continu à droite et sautant aux instants T_n , avec $T_0 = 0$. La suite des instants inter-sauts (S_n) vérifie, pour tout n , $T_{n+1} = T_n + S_{n+1}$ et $S_0 = 0$. Les positions lors des sauts sont les $Z_n = X_{T_n}$. Dans la suite de nos investigations, nous allons beaucoup travailler à partir des chaînes de Markov (Z_n) et surtout (Z_n, S_n) qui contient tout l'aléa du processus (X_t) .

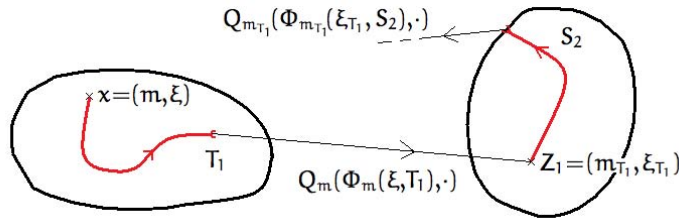


FIGURE 1.1 – Dynamique d'un processus markovien déterministe par morceaux.

J'ai fait le choix de rappeler la définition de la dynamique des PDMP dans le contexte spécifique des chapitres concernés : dans le chapitre 3, on s'intéresse à l'estimation de la densité conditionnelle des temps inter-sauts pour un PDMP défini sur un espace métrique, alors que dans le chapitre 5, on étudie l'estimation du noyau de Markov Q pour un PDMP défini sur \mathbb{R}^d . En particulier, cela permet la lecture indépendante de chaque chapitre de ce manuscrit.

Une partie de mes recherches concerne l'estimation non paramétrique de la loi conditionnelle de la durée inter-sauts S_n sachant la position $Z_{n-1} = x$ du processus à l'instant T_{n-1} . Cette loi a une masse en $t^*(x)$: elle correspond au saut forcé lorsque le processus atteint la frontière de son espace d'état. Entre les instants 0 et $t^*(x)$, cette loi admet une densité conditionnelle $f(x, t)$,

$$f(x, t) = \lambda_m(\Phi_m(\xi, t)) \exp\left(-\int_0^t \lambda_m(\Phi_m(\xi, s)) ds\right),$$

si $x = (m, \xi)$. Dans le chapitre 3 de ce manuscrit, on s'intéresse à l'estimation non paramétrique de la densité conditionnelle f à partir de l'observation d'une trajectoire du processus en temps long. Ces recherches utilisent en partie la méthodologie mise en place dans le chapitre 2 dans lequel on étudie l'estimation non paramétrique du taux de saut pour une certaine classe de processus de renouvellement qui partage un certain nombre de caractéristiques avec les PDMP. On a regroupé ces deux chapitres au sein d'une même partie intitulée *Estimation non paramétrique d'un taux de saut*. Enfin, les travaux présentés ont fait l'objet des articles [10] à paraître dans Annales de l'Institut Henri Poincaré et [9] soumis pour publication.

Dans la seconde partie du manuscrit *Estimation d'une transition*, j'ai regroupé deux sujets de recherche assez différents : l'application d'une méthode d'approximation numérique adaptée aux lois de probabilité à un problème de régression semi-paramétrique dans le chapitre 4, et l'estimation non paramétrique du noyau de Markov Q d'un PDMP dans le chapitre 5. Dans les deux cas, il s'agit d'étudier comme estimer ou approcher numériquement la loi d'une transition. Ces travaux ont fait l'objet des articles [11] publié dans Journal of Statistical Planning and Inference et [7] soumis pour publication.

La conclusion de cette thèse et des perspectives à l'ensemble des résultats et méthodes présentés dans ce manuscrit sont données dans le chapitre 6.

Première partie

ESTIMATION NON PARAMÉTRIQUE D'UN
TAUX DE SAUT

ESTIMATION DU TAUX DE SAUT D'UN PROCESSUS DE RENOUVELLEMENT MARQUÉ

Dans ce chapitre, on s'intéresse à une méthode d'estimation non paramétrique du taux de saut d'un processus de renouvellement marqué. Dans le cadre d'une observation en temps long du processus, on propose un estimateur de type Nelson-Aalen du taux de saut cumulé, ainsi qu'un estimateur à noyau du taux de saut. Les estimateurs et les résultats asymptotiques donnés ici ont fait l'objet du papier *Nonparametric estimation of the jump rate for nonhomogeneous marked renewal processes* [10] avec François Dufour et Anne Gégout-Petit, à paraître dans Annales de l'Institut Henri Poincaré. Enfin, cette étude est une étape clé dans la procédure d'estimation de la loi conditionnelle des temps inter-sauts d'un PDMP présentée dans le chapitre 3.

Ce chapitre est organisé de la façon suivante. Dans la section 2.1, on présente le problème et son contexte. Ensuite, on définit précisément la classe de processus que nous considérons et nous donnons un exemple d'application dans la section 2.2. La section 2.3 est consacrée au cas discret pour lequel on montre qu'on retrouve l'hypothèse d'intensité multiplicative proposée par Aalen dans les années soixante-dix. Dans la section 2.4, on se propose d'étudier le cas où l'espace d'état est continu et on propose un estimateur du taux de saut cumulé dans la sous-section 2.4.1 et un estimateur à noyau du taux de saut dans la sous-section 2.4.2. Une étude de simulations est donnée dans la section 2.5, alors qu'une conclusion et des perspectives à ce travail sont présentées dans la section 2.6. Enfin, des preuves techniques sont reportées dans les sections 2.7 et 2.8. On présente brièvement une application à des données biologiques réelles, réalisée avec Raphaël Coudret et Gilles Durrieu, dans la section 2.9.

SOMMAIRE DU CHAPITRE

2.1	Introduction	9
2.2	Définition et premiers résultats	12
2.3	Espace d'état discret	15
2.4	Espace d'état continu	17
2.4.1	Estimation de Λ	17
2.4.2	Estimation de λ	32
2.5	Illustration numérique	38
2.6	Conclusion du chapitre	40
2.7	Preuves de la section 2.2	41
2.7.1	Preuve de la proposition 2.1	41
2.7.2	Preuve du lemme 2.2	42
2.8	Preuves de la section 2.4	44
2.8.1	Preuve de la proposition 2.5	44
2.8.2	Preuve du lemme 2.7	44
2.8.3	Preuve du lemme 2.10	45
2.9	Application à des données biologiques	45

2.1 Introduction

L'objet de ce chapitre est de proposer une méthode non paramétrique d'estimation du taux de saut d'un processus de renouvellement marqué non homogène, quand une seule observation du processus en temps long est disponible. Notre stratégie est de développer une généralisation du célèbre modèle à intensité multiplicative introduit par Aalen dans les années soixante-dix dans [2, 3].

On considère une classe générale de processus de renouvellement marqués, définis sur un ouvert d'un espace métrique. Il s'agit de processus constants par morceaux dont la dynamique dépend de trois caractéristiques : le taux de saut λ qui gouverne la loi des temps inter-sauts, le noyau de transition Q et une fonction t^* qui joue le rôle d'une censure déterministe fonction de l'état du processus. Dans ce cadre, le taux de saut λ est une fonction de deux variables : la marque spatiale et le temps. Notre but est d'estimer cette fonction, ainsi que sa version intégrée – appelée taux de saut cumulé, – à partir d'une observation du processus en temps long.

On ne peut pas directement relier cette classe de modèles stochastiques aux PDMP, et ce, pour deux raisons. D'une part, un PDMP constant par morceaux a un taux de saut homogène en temps, ce qui n'est pas nécessairement le cas ici. D'autre part, lorsque le flot du PDMP est constant au cours du temps, le temps d'atteinte de la frontière est toujours infini, et il n'y a donc aucun saut forcé. Ils sont pourtant autorisés pour les processus de renouvellement que nous considérons dans ce chapitre. Toutefois, les PDMP et les processus de renouvellement marqués ont des dynamiques similaires. La différence majeure réside dans la loi des sauts : pour un PDMP, la nouvelle position lors d'un saut dépend – à travers le flot bien sûr – du temps inter-saut précédent, ce qui n'est pas le cas pour les modèles de renouvellement.

Aalen a introduit dans le milieu des années soixante-dix le célèbre modèle à intensité multiplicative (voir sa thèse [1], ou ses articles [2, 3]) : l'intensité d'un processus de comptage sous-jacent N est supposée s'écrire comme le produit d'un processus prévisible Y et d'une fonction déterministe λ , appelée taux de saut ou taux de risque suivant le contexte. Dans ce cadre, l'estimateur de Nelson-Aalen permet d'approcher le taux cumulé $\Lambda(t) = \int_0^t \lambda(s) ds$. Ramlau-Hansen a construit quelques années plus tard, dans [74], un estimateur à noyau de λ à partir de l'estimateur de Nelson-Aalen.

Un grand nombre de problèmes d'estimation en analyse de survie ou en statistique des processus sont reliés à des processus de comptage dépendant d'une variable spatiale. Celle-

ci peut être vue comme une marque ou une covariable suivant le contexte. L'estimateur de Nelson-Aalen est une méthode efficace et populaire dans ce type de problèmes puisque l'hypothèse multiplicative est vérifiée dans une grande variété d'applications (voir par exemple [5]). En particulier, on peut appliquer l'approche de Nelson-Aalen pour estimer le taux de saut d'un processus de comptage marqué dont l'espace d'état est fini, à partir d'un grand nombre d'observations indépendantes. Plus récemment, en 2011, Comte *et al.* ont proposé dans [21] une méthode adaptative pour estimer le taux de saut d'un processus marqué sous l'hypothèse multiplicative.

De nombreux papiers traitent de méthodes non paramétriques et semi-paramétriques lorsque la marque spatiale est à valeurs dans un espace continu. Sans prétendre être exhaustif, une liste significative de références sur ce champ de recherche peut être trouvée dans [4, 5, 55, 64]. En particulier, McKeague et Utikal se sont intéressés dans [66] à l'estimation du taux de saut lorsque la covariable est dans $[0, 1]$. Leur approche est basée sur un lissage de l'estimateur de Nelson-Aalen dans les directions temporelle et spatiale. Li et Doss ont choisi une autre approche, basée sur un ajustement linéaire local dans la direction spatiale (voir [62]). Ils ont étendu le travail de McKeague et Utikal au cas multidimensionnel et prouvé des résultats de convergence faible. On peut aussi citer les papiers de Utikal [83, 84] portant sur l'estimation du taux de saut pour deux classes très spécifiques de processus marqués sous des hypothèses de martingale à temps continu. La structure euclidienne de l'espace d'état de la covariable est à chaque fois un point clé dans les approches précédemment citées. Pendant la même période, des approches non paramétriques ont aussi été étudiées par Beran dans [14], Stute dans [81] et Dabrowska dans [30], mais pour des observations indépendantes. Des méthodes semi-paramétriques ont également été envisagées, notamment par Cox dans [28]. Le lecteur intéressé peut consulter [5] ainsi que les références qui y sont données pour une revue complète de la littérature sur ces modèles.

Pour de nombreuses raisons, notre approche et les résultats mentionnés plus haut sont à la fois différents et complémentaires. Dans ce chapitre, nous proposons d'adapter le modèle à intensité multiplicative lorsqu'on fait intervenir une discrétisation de l'espace d'état et, par conséquent, une approximation des fonctions d'intérêt. Puisque l'espace d'état que nous considérons est très général, les stratégies employées par McKeague et Utikal [66], Li et Doss [62], ou Utikal [83, 84] ne sont pas envisageables. De plus, ces auteurs considèrent des hypothèses portant à la fois sur des propriétés de martingale à temps continu et le comportement asymptotique du processus Y . D'un point de vue pratique, ces hypothèses ne sont pas complètement satisfaisantes car difficiles à vérifier, comme c'est le cas dans notre problème. Dans ce travail, nous assurons la convergence de notre estimateur par des conditions portant directement sur les caractéristiques du processus.

La suite de ce chapitre est divisée en deux grandes parties. Dans la première, on considère que le noyau de transition du processus charge seulement un nombre fini de points. Cela revient à considérer que l'espace d'état du processus est discret. Dans ce contexte, le théorème 2.3 établit que le modèle à intensité multiplicative est vérifié. Dans la seconde partie, on suppose que le noyau Q est une mesure diffuse, *id est* qui ne charge pas des singletons. Si les Z_i désignent les marques du processus sous-jacent, pour tout x et pour tout entier i , l'indicatrice $\mathbb{1}_{\{Z_i=x\}}$ est alors presque sûrement nulle ; cela exclut l'approche explorée dans le cas discret. Notre méthode s'appuie alors sur une partition de l'espace d'état, notée (A_k) . Dans ce cadre, il apparaît intuitif de considérer le processus de comptage

$$N_n(A_k, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A_k\}} \mathbb{1}_{\{S_{i+1} \leq t\}},$$

où les S_i sont les temps inter-sauts du processus. Bien que le modèle multiplicatif d'Aalen ne soit pas vérifié pour ce processus de comptage, l'intensité stochastique de $N_n(A_k, t)$ est presque sûrement équivalente (quand n tend vers l'infini) au produit $Y_n(A_k, t)l(A_k, t)$ (voir la proposition 2.15), où $l(A_k, t)$ est une approximation du taux de saut $\lambda(x, t)$ pour $x \in A_k$, et

$$Y_n(A_k, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A_k\}} \mathbb{1}_{\{S_{i+1} \geq t\}}.$$

Il est alors naturel d'introduire les processus suivants,

$$\widehat{L}_n(A_k, t) = \int_0^t Y_n(A_k, s)^+ dN_n(A_k, s),$$

où $Y_n(A_k, t)^+$ est l'inverse généralisé de $Y_n(A_k, t)$, et

$$L_n^*(A_k, t) = \int_0^t l(A_k, s) \mathbb{1}_{\{Y_n(A_k, s) > 0\}} ds.$$

Dans les papiers [2, 3], la différence $\widehat{L}_n(A_k, t) - L_n^*(A_k, t)$ est une martingale à temps continu alors qu'ici, un terme supplémentaire noté $\alpha_n(t)$ apparaît. Celui-ci tend vers 0 quand n tend vers l'infini. Intuitivement, cela signifie que le modèle à intensité multiplicative est vérifié asymptotiquement. En utilisant à la fois l'inégalité de Lengart et le comportement asymptotique du terme $\alpha_n(t)$, on montre que $\widehat{L}_n(A_k, t)$ est un estimateur consistant de $L(A_k, t) = \int_0^t l(A_k, s) ds$ (voir la proposition 2.19). On en déduit un estimateur consistant de $\Lambda(x, t) = \int_0^t \lambda(x, s) ds$ (voir le théorème 2.25), puisque $L(A_k, t)$ et $\Lambda(x, t)$ sont proches pour $x \in A_k$ (voir le lemme 2.22). Ensuite, on s'intéresse au lissage de cet estimateur par des méthodes à noyau afin d'estimer $l(A_k, t)$ (voir la proposition 2.28) et donc la fonction

d'intérêt $\lambda(x, t)$ (voir le théorème 2.30). La présence des sauts forcés est une difficulté supplémentaire tout au long de cette étude.

La suite du chapitre est organisée de la façon suivante. On donne, dans la section 2.2, la définition précise de la classe des processus de renouvellement marqués non homogènes qui nous intéresse, ainsi qu'une application en fiabilité. On établit aussi des résultats techniques à propos d'indépendances conditionnelles et de martingales. La section 2.3 est consacrée au cas discret, lorsque Q charge seulement un nombre fini de points. Dans la section 2.4, nous n'imposons aucune contrainte sur l'espace d'état et nous donnons des estimateurs consistants du taux cumulé (voir le théorème 2.25) et du taux de saut (voir le théorème 2.30). Finalement, dans la section 2.5, on présente une étude de simulations qui illustre le bon comportement de nos estimateurs. Une conclusion et des perspectives à ce travail sont données dans la section 2.6. Les démonstrations techniques de certains résultats sont reportées dans les sections 2.7 et 2.8. Une application à des données biologiques réelles réalisée avec Raphaël Coudret et Gilles Durrieu est présentée dans la section 2.9.

2.2 Définition et premiers résultats

On considère un processus à temps continu et constant par morceaux noté (X_t) . Pour tout $t \geq 0$, X_t est à valeurs dans E , un ouvert d'un espace métrique séparable (\mathcal{E}, d) . L'évolution de (X_t) peut être décrite comme suit. Pour tout $t \geq 0$,

$$X_t = Z_n \quad \text{si} \quad S_0 + \dots + S_n \leq t < S_0 + \dots + S_{n+1}.$$

Les Z_n correspondent aux marques du processus (X_t) tandis que les S_n sont les temps inter-sauts. On suppose ici que la suite (Z_n) est une chaîne de Markov sur E muni de la tribu borélienne $\mathcal{B}(E)$ et définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P}_{\nu_0})$. Son noyau de transition est noté Q , alors que la loi de la position initiale Z_0 est ν_0 . Une formulation équivalente (voir par exemple le théorème 2.4.3 de [13]) est la suivante,

$$\forall n \geq 1, Z_n = \psi(Z_{n-1}, \varepsilon_{n-1}), \tag{2.1}$$

où ψ est une fonction mesurable et (ε_n) est une suite de variables aléatoires indépendantes et de même loi. On définit la loi de la suite (S_n) sur \mathbb{R}_+ à partir de deux fonctions : $\lambda : E \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est le taux de saut du processus et $t^* : E \rightarrow]0, +\infty]$ joue le rôle d'une censure déterministe. Pour chaque entier $n \geq 1$, la fonction de survie du temps S_n est

donnée, pour tout $t \geq 0$, par

$$\begin{aligned} \mathbb{P}_{\nu_0}(S_n > t | \{Z_i : i \geq 0\}, S_0, \dots, S_{n-1}) &= \mathbb{P}_{\nu_0}(S_n > t | Z_{n-1}) \\ &= \exp\left(-\int_0^t \lambda(Z_{n-1}, s) ds\right) \mathbb{1}_{\{0 \leq t < t^*(Z_{n-1})\}}. \end{aligned} \quad (2.2)$$

On suppose de plus que $S_0 = 0$. Par conséquent, il existe une fonction φ et une suite de variables aléatoires (δ_n) , indépendantes et de même loi, indépendantes de la suite (ε_n) , telles que,

$$\forall n \geq 1, S_n = \varphi(Z_{n-1}, \delta_{n-1}).$$

Les deux suites (ε_n) et (δ_n) sont supposées être indépendantes de la marque initiale Z_0 .

L'objet de ce chapitre est l'estimation non paramétrique du taux de saut λ à partir de l'observation de la chaîne (Z_n, S_n) . La méthode développée est une première étape dans notre approche de l'estimation de la loi conditionnelle des temps inter-sauts d'un PDMP. Néanmoins, proposer un estimateur non paramétrique du taux de saut pour cette classe de processus de renouvellement a aussi un intérêt intrinsèque. L'exemple qui suit satisfait les hypothèses mentionnées plus haut.

On considère une machine dont le régime de production prend ses valeurs dans un ouvert de \mathbb{R}^d . La dynamique du régime est supposée être un processus de renouvellement non homogène : l'état est constant par morceaux jusqu'à ce qu'une défaillance ait lieu. On peut naturellement considérer que le taux de panne dépend de la configuration de la machine. Quand la machine tombe en panne, la réparation a lieu instantanément, et la configuration est changée aléatoirement selon un noyau de transition Q dépendant seulement de l'état de marche précédent. On suppose de plus qu'il existe une période déterministe d'inspection, dépendant également du régime de production de la machine. Lors d'une inspection, le nouveau régime est choisi à nouveau selon le noyau Q . Dans cette application, il est intéressant d'estimer le taux de panne de la machine à partir de l'observation en temps long de son comportement. On peut alors, par exemple, comparer la fiabilité de deux machines concurrentes.

On peut associer au taux de saut λ son intégrale appelée taux de saut cumulé, ainsi que la fonction de survie et la densité de probabilité qui lui correspondent. La densité conditionnelle f vérifie,

$$\forall \xi \in E, \quad \forall t \geq 0, \quad f(\xi, t) = \lambda(\xi, t) \exp\left(-\int_0^t \lambda(\xi, s) ds\right). \quad (2.3)$$

Le taux de saut cumulé Λ est donné par,

$$\forall \xi \in E, \forall t \geq 0, \Lambda(\xi, t) = \int_0^t \lambda(\xi, s) ds. \quad (2.4)$$

Finalement, la fonction de survie conditionnelle est notée G ,

$$\forall \xi \in E, \forall t \geq 0, G(\xi, t) = \exp\left(-\int_0^t \lambda(\xi, s) ds\right). \quad (2.5)$$

On a la relation immédiate $\lambda = f/G$. Pour chaque entier n , \mathcal{G}_n désigne la tribu engendrée par les n premiers Z_i ,

$$\mathcal{G}_n = \sigma(Z_0, \dots, Z_{n-1}). \quad (2.6)$$

Pour chaque entier i , le processus de comptage à un seul saut N^{i+1} est donné, pour tout $t \geq 0$, par

$$N^{i+1}(t) = \mathbb{1}_{\{S_{i+1} \leq t\}},$$

et on note (\mathcal{F}_t^{i+1}) la filtration qui lui est associée. Dans cette partie, nous allons démontrer deux résultats : le premier concerne deux indépendances conditionnelles, alors que dans le second, on s'intéresse au compensateur du processus de comptage N^{i+1} dans une certaine filtration. Le symbole $\perp_{\mathcal{T}}$ désigne l'indépendance conditionnellement à la tribu \mathcal{T} .

PROPOSITION 2.1 *Soient n un entier et $1 \leq i \leq n$. Pour chaque entier $j \neq i$, soient $t_j \geq 0$ et $t \geq 0$. Alors, on a*

$$\bigvee_{j \neq i} \mathcal{F}_{t_j}^j \perp_{\mathcal{G}_n} \mathcal{F}_t^i \quad \text{et} \quad \mathcal{F}_t^i \perp_{\sigma(Z_{i-1})} \mathcal{G}_n.$$

On peut alors déduire de la proposition 6.8 de [58] que, pour tous $s < t$, $0 \leq i \leq n-1$,

$$\bigvee_{j \neq i+1} \mathcal{F}_s^j \perp_{\mathcal{G}_n \vee \mathcal{F}_s^{i+1}} \mathcal{F}_t^{i+1} \quad \text{et} \quad \mathcal{F}_t^{i+1} \perp_{\sigma(Z_i) \vee \mathcal{F}_s^{i+1}} \mathcal{G}_n.$$

PREUVE. Elle est donnée dans la section 2.7. □

On en déduit la martingale associée au processus de comptage N^{i+1} .

LEMME 2.2 Pour chaque entier i , le processus M^{i+1} donné par,

$$\forall 0 \leq t < t^*(Z_i), M^{i+1}(t) = N^{i+1}(t) - \int_0^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du, \quad (2.7)$$

est une martingale à temps continu dans la filtration $(\sigma(Z_i) \vee \mathcal{F}_s^{i+1})$.

PREUVE. Elle est donnée dans la section 2.7. □

La proposition 2.1 et le lemme 2.2 ont une importance cruciale dans la suite de l'étude. De plus, la référence à la probabilité \mathbb{P}_{ν_0} est implicite dans le texte. Enfin, dans un souci de lisibilité, on écrira \mathbb{P}_x au lieu de $\mathbb{P}_{\delta_{\{x\}}}$.

2.3 Espace d'état discret

On suppose ici que le noyau markovien Q charge seulement un ensemble fini $\{x_1, \dots, x_M\}$. On peut associer à chaque point x_k le temps déterministe de sortie $t_k^* = t^*(x_k)$. On considère maintenant que k est fixé. Dans cette section, nous allons démontrer dans le théorème 2.3 que le modèle à intensité multiplicative est vérifié lorsqu'on cherche à estimer $\Lambda(x_k, \cdot)$.

Pour chaque entier n , on introduit les processus $N_n(x_k, \cdot)$ et $Y_n(x_k, \cdot)$,

$$\forall t \geq 0, N_n(x_k, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i = x_k\}}, \quad (2.8)$$

$$\forall t \geq 0, Y_n(x_k, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq t\}} \mathbb{1}_{\{Z_i = x_k\}}. \quad (2.9)$$

On définit aussi le processus $M_n(x_k, \cdot)$ par,

$$\forall 0 \leq t < t_k^*, M_n(x_k, t) = N_n(x_k, t) - \int_0^t \lambda(x_k, s) Y_n(x_k, s) ds, \quad (2.10)$$

et la filtration $(\tilde{\mathcal{F}}_t^n)$ par,

$$\forall 0 \leq t < t_k^*, \tilde{\mathcal{F}}_t^n = \mathcal{G}_n \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1}.$$

THÉORÈME 2.3 Soit $n \geq 1$. Le processus $M_n(x_k, \cdot)$ est une $(\tilde{\mathcal{F}}_t^n)$ -martingale à temps continu sous \mathbb{P}_{v_0} .

PREUVE. Soient $0 \leq s < t < t_k^*$. Par (2.8), (2.9) et (2.10), on a

$$\begin{aligned} M_n(x_k, t) &= \sum_{i=0}^{n-1} \left\{ N^{i+1}(t) \mathbb{1}_{\{Z_i=x_k\}} - \int_0^t \lambda(x_k, u) \mathbb{1}_{\{S_{i+1} \geq u\}} \mathbb{1}_{\{Z_i=x_k\}} du \right\} \\ &= \sum_{i=0}^{n-1} M^{i+1}(t) \mathbb{1}_{\{Z_i=x_k\}}, \end{aligned} \quad (2.11)$$

par (2.7). On obtient alors

$$\mathbb{E}_{v_0} \left[M_n(x_k, t) \middle| \tilde{\mathcal{F}}_s^n \right] = \sum_{i=0}^{n-1} \left\{ \mathbb{E}_{v_0} \left[N^{i+1}(t) \middle| \tilde{\mathcal{F}}_s^n \right] - \mathbb{E}_{v_0} \left[\int_0^t \mathbb{1}_{\{S_{i+1} \geq u\}} \lambda(x_k, u) du \middle| \tilde{\mathcal{F}}_s^n \right] \right\} \mathbb{1}_{\{Z_i=x_k\}}.$$

D'après la proposition 2.1, on a

$$\bigvee_{j \neq i+1} \mathcal{F}_s^j \perp_{\mathcal{G}_n \vee \mathcal{F}_s^{i+1}} \mathcal{F}_t^{i+1}.$$

On en déduit par la proposition 6.6 de [58] que

$$\begin{aligned} &\mathbb{E}_{v_0} \left[M_n(x_k, t) \middle| \tilde{\mathcal{F}}_s^n \right] \\ &= \sum_{i=0}^{n-1} \left\{ \mathbb{E}_{v_0} \left[N^{i+1}(t) \middle| \mathcal{G}_n \vee \mathcal{F}_s^{i+1} \right] - \mathbb{E}_{v_0} \left[\int_0^t \mathbb{1}_{\{S_{i+1} \geq u\}} \lambda(x_k, u) du \middle| \mathcal{G}_n \vee \mathcal{F}_s^{i+1} \right] \right\} \mathbb{1}_{\{Z_i=x_k\}}. \end{aligned}$$

De plus, à nouveau d'après la proposition 2.1, on a

$$\mathcal{F}_t^{i+1} \perp_{\sigma(Z_i) \vee \mathcal{F}_s^{i+1}} \mathcal{G}_n.$$

Donc, toujours par la proposition 6.6 de [58], on a

$$\begin{aligned} \mathbb{E}_{v_0} \left[M_n(x_k, t) \middle| \tilde{\mathcal{F}}_s^n \right] &= \sum_{i=0}^{n-1} \left\{ \mathbb{E}_{v_0} \left[N^{i+1}(t) \middle| \sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right] \mathbb{1}_{\{Z_i=x_k\}} \right. \\ &\quad \left. - \mathbb{E}_{v_0} \left[\int_0^t \mathbb{1}_{\{S_{i+1} \geq u\}} \lambda(x_k, u) du \middle| \sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right] \mathbb{1}_{\{Z_i=x_k\}} \right\}. \end{aligned}$$

Grâce au lemme 2.2, cela mène à

$$\mathbb{E}_{\nu_0} \left[M_n(x_k, t) \middle| \tilde{\mathcal{F}}_s^n \right] = \sum_{i=0}^{n-1} M^{i+1}(s) \mathbb{1}_{\{Z_i=x_k\}}.$$

Finalement, avec (2.11), $M_n(x_k, \cdot)$ est bien une martingale. \square

Le théorème 2.3 montre qu'on peut estimer le taux de saut cumulé $\Lambda(x_k, t)$ avec l'estimateur de Nelson-Aalen $\hat{\Lambda}_n(x_k, t)$ donné par

$$\hat{\Lambda}_n(x_k, t) = \sum_{i=0}^{n-1} \frac{1}{Y_n(x_k, S_{i+1})} \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i=x_k\}}.$$

On dirige le lecteur intéressé vers [2, 3] ou [5] pour une étude complète de cet estimateur. En particulier, sous réserve d'obtenir des propriétés asymptotiques de la suite $(Y_n(x_k, t))$, l'inégalité de Lenglar (voir [5], II.5.2.1. Lenglar's inequality) et le théorème de Rebolledo (voir [5], Theorem II.5.1) permettent de montrer la convergence en probabilité et la normalité asymptotique de $\hat{\Lambda}_n(x_k, t)$.

2.4 Espace d'état continu

Cette section est divisée en deux parties. Dans la première, on donne un estimateur du taux cumulé Λ . La seconde traite de l'estimation du taux λ en lissant l'estimateur de Λ par des méthodes à noyau.

2.4.1 Estimation de Λ

On suppose que Q est diffus, c'est-à-dire que Q ne charge pas les singletons. On ne peut plus utiliser la procédure mise en œuvre dans la section précédente puisque, pour tout $x \in E$ et pour tout entier i , $\mathbb{1}_{\{Z_i=x\}} = 0$ presque sûrement. On peut naturellement approcher, sous des conditions de régularité, le taux de saut en x par le taux de saut partant d'un voisinage de x . Comme on l'a mentionné dans l'introduction, on considère le processus de comptage $N_n(A, t)$ défini, pour tout $A \in \mathcal{B}(E)$, par

$$N_n(A, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{S_{i+1} \leq t\}}.$$

On verra dans (2.22) que l'intensité stochastique de $N_n(A, t)$ dans une filtration bien choisie est

$$\sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{S_{i+1} \geq t\}} \lambda(Z_i, t).$$

Les premiers résultats de cette section traitent de l'ergodicité des chaînes de Markov sous-jacentes. Leurs propriétés sont très importantes pour étudier le comportement asymptotique de l'intensité stochastique de $N_n(A, t)$ (voir la proposition 2.15). En s'appuyant sur ces résultats, nous étudierons, dans la proposition 2.19 et le théorème 2.20, l'estimation de $L(A, t) = \int_0^t l(A, s) ds$, où $l(A, t)$ est une approximation du taux de saut $\lambda(x, t)$ lorsque x est dans A . Nous en déduisons un estimateur du taux cumulé $\Lambda(x, t)$ (voir le théorème 2.25). Une difficulté supplémentaire concerne la mesure invariante de l'ensemble A que nous devons estimer.

On impose des hypothèses sur l'ergodicité de la chaîne (Z_n) et les caractéristiques du processus. Dans la suite, ν_n désigne la loi de Z_n , pour chaque entier n .

HYPOTHÈSE 2.4 *Il existe une mesure ν telle que, pour toute loi initiale $\nu_0 = \delta_{\{x\}}$, $x \in E$,*

$$\lim_{n \rightarrow +\infty} \|\nu_n - \nu\|_{TV} = 0.$$

Cette hypothèse peut être directement reliée au noyau de transition Q de la chaîne (Z_n) (existence d'une fonction de Foster-Lyapunov ou condition de Doeblin par exemple). On peut diriger le lecteur vers [67] pour des résultats de ce type. Grâce à cette hypothèse, on peut obtenir les résultats suivants. En particulier, la chaîne (Z_n) est Harris-récurrente.

PROPOSITION 2.5 *On a les résultats suivants.*

1. (Z_n) est ν -irréductible.
2. (Z_n) est positive Harris-récurrente et apériodique.
3. ν est l'unique loi invariante de (Z_n) .

PREUVE. La preuve se trouve dans la section 2.8. □

On note η_n la loi du couple (Z_n, S_{n+1}) et $\mu_z(\cdot)$ la loi du temps S_1 sachant $Z_0 = z$. Par construction (voir (2.2)), la loi de S_{n+1} sachant $Z_n = z$ ne dépend pas de n et est donnée par $\mu_z(\cdot)$. Le résultat suivant donne la loi du couple (Z_n, S_{n+1}) en fonction de (μ_z) et ν_n .

LEMME 2.6 *Pour tout entier n , η_n vérifie, pour tout $A \times \Gamma \in \mathcal{B}(E) \otimes \mathcal{B}(\mathbb{R}_+)$,*

$$\eta_n(A \times \Gamma) = \int_{A \times \Gamma} \mu_z(ds) \nu_n(dz).$$

PREUVE. Il s'agit de la désintégration de la mesure η_n par rapport à sa marginale ν_n . \square

Nous allons voir que la suite (η_n) converge vers la mesure η donnée par,

$$\forall A \times \Gamma \in \mathcal{B}(E) \otimes \mathcal{B}(\mathbb{R}_+), \eta(A \times \Gamma) = \int_{A \times \Gamma} \mu_z(ds) \nu(dz).$$

LEMME 2.7 *Pour toute loi initiale $\nu_0 = \delta_{\{x\}}$,*

$$\lim_{n \rightarrow +\infty} \|\eta_n - \eta\|_{TV} = 0.$$

PREUVE. La preuve est donnée dans la section 2.8. \square

Par conséquent, la chaîne de Markov (Z_n, S_{n+1}) a des propriétés d'ergodicité similaires à celles de la chaîne (Z_n) . Celles-ci sont résumées dans la proposition qui suit, dont l'énoncé est évidemment proche de la proposition 2.5.

PROPOSITION 2.8 *On a les résultats suivants.*

1. (Z_n, S_{n+1}) est η -irréductible.
2. (Z_n, S_{n+1}) is positive Harris-récurrente et apériodique.
3. η est l'unique loi invariante de (Z_n, S_{n+1}) .

PREUVE. On peut établir ce résultat à partir du lemme 2.7, avec les arguments donnés dans la preuve de la proposition 2.5. \square

Nous pouvons donc appliquer le théorème ergodique presque sûr aux chaînes de Markov (Z_n) et (Z_n, S_{n+1}) . On impose certaines conditions sur les caractéristiques du processus.

HYPOTHÈSES 2.9

1. Le taux de saut λ est uniformément Lipschitz,

$$\exists[\lambda]_{\text{Lip}} > 0, \forall \xi, \xi' \in E, \forall s \geq 0, |\lambda(\xi, s) - \lambda(\xi', s)| \leq [\lambda]_{\text{Lip}} d(\xi, \xi').$$

2. Il existe une fonction localement intégrable $M : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telle que,

$$\forall \xi \in E, \forall s \geq 0, \lambda(\xi, s) \leq M(s).$$

3. La densité f est continue en temps.

4. f est bornée.

5. La fonction t^* est continue.

L'ensemble de ces conditions est vérifié par l'application présentée dans la section 2.5. Sous ces hypothèses, on peut établir des résultats intermédiaires sur λ , t^* et G .

LEMME 2.10 Soit $A \in \mathcal{B}(E)$ un ensemble relativement compact tel que $\bar{A} \cap \partial E = \emptyset$. Alors,

$$\inf_{\xi \in A} t^*(\xi) > 0.$$

Dans ce cas, on note $t^*(A) = \inf_{\xi \in A} t^*(\xi)$. De plus,

$$\forall t \geq 0, \inf_{\xi \in A} G(\xi, t) > 0.$$

PREUVE. La preuve est reportée dans la section 2.8. \square

Introduisons les notations suivantes,

$$\mathcal{B}_v^+ = \{A \in \mathcal{B}(E) : A \text{ relativement compact, } v(A) > 0 \text{ et } \bar{A} \cap \partial E = \emptyset\}.$$

Le lemme suivant établit que, pour tout $A \in \mathcal{B}_v^+$, λ est borné sur $A \times [0, t^*(A)[$.

LEMME 2.11 *Soient $A \in \mathcal{B}_v^+$, $\xi \in A$ et $0 \leq t < t^*(A)$. Alors,*

$$\lambda(\xi, t) \leq \frac{\|f\|_\infty}{\inf_{\xi \in A} G(\xi, t^*(A))}.$$

PREUVE. Puisque f est bornée et $G(\xi, \cdot)$ décroissante, on obtient par le lemme 2.10,

$$\lambda(\xi, t) \leq \frac{\|f\|_\infty}{G(\xi, t)} \leq \frac{\|f\|_\infty}{G(\xi, t^*(A))}.$$

Cela entraîne immédiatement le résultat voulu. \square

Soit $A \in \mathcal{B}_v^+$. Pour tout entier n , on considère le processus à temps continu $Y_n(A, \cdot)$, défini par,

$$\forall 0 \leq t < t^*(A), Y_n(A, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq t\}} \mathbb{1}_{\{Z_i \in A\}}. \quad (2.12)$$

Afin d'établir deux résultats sur le comportement asymptotique de Y_n , on s'intéresse au lien entre $G(z, \cdot)$ et $\mu_z(\cdot)$.

REMARQUE 2.12 *Soient $A \in \mathcal{B}_v^+$ et $z \in A$. Pour tout $0 \leq t < t^*(A)$, on a*

$$\begin{aligned} \mu_z([t, +\infty[) &= \mathbb{P}_{v_0}(S_1 \geq t | Z_0 = z) \\ &= \mathbb{P}_{v_0}(S_1 > t | Z_0 = z), \end{aligned}$$

car $t < t^*(A) \leq t^*(z)$. Alors,

$$\mu_z([t, +\infty[) = G(z, t).$$

LEMME 2.13 Soient $A \in \mathcal{B}_v^+$ et $x \in E$. Alors, pour tout $0 \leq t < t^*(A)$,

$$\frac{Y_n(A, t)}{n} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \int_A G(z, t) \nu(dz),$$

quand n tend vers l'infini. De plus, cette limite est strictement positive.

PREUVE. D'après la proposition 2.8, la chaîne de Markov (Z_n, S_{n+1}) est positive Harris-récurrente et admet η comme unique loi invariante. Donc, par le théorème ergodique presque sûr (voir par exemple le théorème 17.1.7 de [67]), on a, quand n tend vers l'infini,

$$\frac{Y_n(A, t)}{n} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \int_A \mu_z([t, +\infty[) \nu(dz).$$

De plus, pour tout $z \in A$, comme $t < t^*(A)$ et d'après la remarque 2.12,

$$\mu_z([t, +\infty[) = G(z, t).$$

Il s'agit d'une quantité strictement positive puisque $\nu(A) > 0$ et $\inf_{\xi \in A} G(\xi, t) > 0$ par le lemme 2.10. \square

Soient $A \in \mathcal{B}_v^+$ et $0 \leq t < t^*(A)$. On définit l'inverse généralisé $Y_n(A, t)^+$ de $Y_n(A, t)$ par

$$Y_n(A, t)^+ = \begin{cases} 0 & \text{si } Y_n(A, t) = 0, \\ \frac{1}{Y_n(A, t)} & \text{sinon.} \end{cases} \quad (2.13)$$

LEMME 2.14 Soient $A \in \mathcal{B}_v^+$, $0 \leq t < t^*(A)$ et $x \in E$. Alors, pour tout entier n ,

$$Y_n(A, t)^+ \leq 1 \text{ } \mathbb{P}_x\text{-p.s.} \quad (2.14)$$

De plus, quand n tend vers l'infini,

$$Y_n(A, t)^+ \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0, \quad (2.15)$$

$$\mathbb{1}_{\{Y_n(A, t) = 0\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0, \quad (2.16)$$

$$\int_0^t \mathbb{1}_{\{Y_n(A, s) = 0\}} ds \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0.$$

PREUVE. $Y_n(A, t)^+$ est presque sûrement borné par 1, puisque $Y_n(A, t)$ est à valeurs entières. On obtient immédiatement les limites (2.15) et (2.16), car $Y_n(A, t)/n$ converge presque sûrement vers une limite strictement positive d'après le lemme 2.13. Finalement,

$$\limsup_{n \rightarrow +\infty} \int_0^t \mathbb{1}_{\{Y_n(A, s)=0\}} ds \leq \int_0^t \limsup_{n \rightarrow +\infty} \mathbb{1}_{\{Y_n(A, s)=0\}} ds = 0,$$

par (2.16). □

Dans la proposition suivante, on applique le théorème ergodique dans le but de définir, pour tout $A \in \mathcal{B}_v^+$, la fonction $l(A, \cdot)$ qui est une approximation du taux de saut $\lambda(\xi, \cdot)$ pour $\xi \in A$ (voir le lemme 2.22). On peut aussi montrer la continuité de cette approximation.

PROPOSITION 2.15 Soient $A \in \mathcal{B}_v^+$, $0 \leq t < t^*(A)$ et $x \in E$. Alors, quand n tend vers l'infini,

$$Y_n(A, t)^+ \sum_{i=0}^{n-1} \lambda(Z_i, t) \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{S_{i+1} \geq t\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} l(A, t) = \frac{\int_A f(z, t) \nu(dz)}{\int_A G(z, t) \nu(dz)}. \quad (2.17)$$

La fonction $l(A, \cdot)$ est continue sur $[0, t^*(A)[$. En particulier, on a

$$K_t(A) = \sup_{0 \leq s \leq t} |l(A, s)| < +\infty. \quad (2.18)$$

PREUVE. Le théorème ergodique (théorème 17.1.7 de [67]) appliqué à la chaîne (Z_n, S_{n+1}) mène à

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=0}^{n-1} \lambda(Z_i, t) \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{S_{i+1} \geq t\}} &= \int_A \lambda(z, t) \mu_z([t, +\infty[) \nu(dz) \quad \mathbb{P}_x\text{-p.s.} \\ &= \int_A \lambda(z, t) G(z, t) \nu(dz) \quad \mathbb{P}_x\text{-p.s.}, \end{aligned}$$

car $t < t^*(A)$. En remarquant que $f(z, t) = \lambda(z, t)G(z, t)$, on obtient

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=0}^{n-1} \lambda(Z_i, t) \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{S_{i+1} \geq t\}} = \int_A f(z, t) \nu(dz) \quad \mathbb{P}_x\text{-p.s.}$$

De plus, d'après le lemme 2.13,

$$\frac{1}{n} Y_n(A, t) \xrightarrow{\mathbb{P}_x\text{-p.s.}} \int_A G(z, t) \nu(dz).$$

Finalement, en faisant le rapport de ces deux limites, on obtient la convergence recherchée. D'après le théorème de continuité sous le signe intégrale, puisque f et G sont continues en temps et bornées, $l(A, \cdot)$ est continue. \square

Maintenant qu'on a étudié le comportement asymptotique de Y_n , on s'intéresse à des structures de martingales à temps continu. Grâce à l'inégalité de Lenglart pour les martingales, on va pouvoir estimer, pour tout $A \in \mathcal{B}_v^+$, les fonctions $l(A, \cdot)$ et $L(A, \cdot)$, où $L(A, \cdot)$ est donnée par,

$$\forall 0 \leq t < t^*(A), L(A, t) = \int_0^t l(A, s) ds. \quad (2.19)$$

THÉORÈME 2.16 *Soit $A \in \mathcal{B}_v^+$. Pour tout n , le processus $M_n(A, \cdot)$ défini par,*

$$\forall 0 \leq t < t^*(A), M_n(A, t) = \sum_{i=0}^{n-1} M^{i+1}(t) \mathbb{1}_{\{Z_i \in A\}}, \quad (2.20)$$

est une martingale à temps continu dans la filtration $(\mathcal{G}_n \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1})$.

PREUVE. Soit $0 \leq s < t < t^*(A)$. Alors,

$$\mathbb{E}_{v_0} \left[M_n(A, t) \middle| \mathcal{G}_n \vee \bigvee_{i=0}^{n-1} \mathcal{F}_s^{i+1} \right] = \sum_{i=0}^{n-1} \mathbb{E}_{v_0} \left[M^{i+1}(t) \mathbb{1}_{\{Z_i \in A\}} \middle| \mathcal{G}_n \vee \bigvee_{j=0}^{n-1} \mathcal{F}_s^{j+1} \right].$$

D'après la proposition 2.1, on a

$$\bigvee_{j \neq i+1} \mathcal{F}_s^j \perp_{\mathcal{G}_n \vee \mathcal{F}_s^{i+1}} \mathcal{F}_t^{i+1} \quad \text{et} \quad \mathcal{F}_t^{i+1} \perp_{\sigma(Z_i) \vee \mathcal{F}_s^{i+1}} \mathcal{G}_n.$$

Par conséquent, puisque $\sigma(Z_i)$ est une sous-tribu de \mathcal{G}_n , on a par le corollaire 6.8 de [58],

$$\bigvee_{j \neq i+1} \mathcal{F}_s^j \perp_{\mathcal{G}_n \vee \mathcal{F}_s^{i+1}} \mathcal{F}_t^{i+1} \vee \sigma(Z_i) \quad \text{et} \quad \sigma(Z_i) \vee \mathcal{F}_t^{i+1} \perp_{\sigma(Z_i) \vee \mathcal{F}_s^{i+1}} \mathcal{G}_n.$$

Donc, comme pour tout i , $M^{i+1}(t)\mathbb{1}_{\{Z_i \in A\}}$ est $\sigma(Z_i) \vee \mathcal{F}_t^{i+1}$ -mesurable,

$$\begin{aligned} \mathbb{E}_{\nu_0} \left[M_n(A, t) \middle| \mathcal{G}_n \vee \bigvee_{i=0}^{n-1} \mathcal{F}_s^{i+1} \right] &= \sum_{i=0}^{n-1} \mathbb{E}_{\nu_0} [M^{i+1}(t)\mathbb{1}_{\{Z_i \in A\}} | \mathcal{G}_n \vee \mathcal{F}_s^{i+1}] \\ &= \sum_{i=0}^{n-1} \mathbb{E}_{\nu_0} [M^{i+1}(t)\mathbb{1}_{\{Z_i \in A\}} | \sigma(Z_i) \vee \mathcal{F}_s^{i+1}]. \end{aligned}$$

Par le lemme 2.2,

$$\mathbb{E}_{\nu_0} [M^{i+1}(t)\mathbb{1}_{\{Z_i \in A\}} | \sigma(Z_i) \vee \mathcal{F}_s^{i+1}] = M^{i+1}(s)\mathbb{1}_{\{Z_i \in A\}}.$$

Donc,

$$\mathbb{E}_{\nu_0} \left[M_n(A, t) \middle| \mathcal{G}_n \vee \bigvee_{i=0}^{n-1} \mathcal{F}_s^{i+1} \right] = \sum_{i=0}^{n-1} M^{i+1}(s)\mathbb{1}_{\{Z_i \in A\}}.$$

Par (2.20), cela assure que $M_n(A, \cdot)$ est une martingale. \square

Pour tout $A \in \mathcal{B}_V^+$, on considère le processus de comptage $N_n(A, \cdot)$, défini pour tout $0 \leq t < t^*(A)$, par

$$N_n(A, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i \in A\}}. \quad (2.21)$$

Par (2.7), (2.20) et (2.21), on peut directement relier la martingale $M_n(A, \cdot)$ au processus de comptage $N_n(A, \cdot)$. Pour tout $0 \leq t < t^*(A)$,

$$M_n(A, t) = N_n(A, t) - \int_0^t \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq u\}} \mathbb{1}_{\{Z_i \in A\}} \lambda(Z_i, u) du. \quad (2.22)$$

LEMME 2.17 Soit $A \in \mathcal{B}_V^+$. Pour tout entier n , le processus donné, pour tout $0 \leq s < t^*(A)$, par

$$\widetilde{M}_n(A, s) = \int_0^s Y_n(A, u)^+ dM_n(A, u), \quad (2.23)$$

est une martingale dont le crochet $\langle \widetilde{M}_n(A) \rangle$ vérifie, pour tout $x \in E$,

$$\forall 0 \leq s < t^*(A), \langle \widetilde{M}_n(A) \rangle(s) \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0,$$

quand n tend vers l'infini.

PREUVE. D'après le théorème 2.16, le processus $M_n(\mathbf{A}, \cdot)$ donné par (2.22) est une martingale à temps continu. Autrement dit, le processus $\mathcal{A}_n(\mathbf{A}, \cdot)$ donné par,

$$\forall 0 \leq s < t^*(\mathbf{A}), \mathcal{A}_n(\mathbf{A}, s) = \int_0^s \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq u\}} \mathbb{1}_{\{Z_i \in \mathbf{A}\}} \lambda(Z_i, u) du,$$

est le compensateur du processus de comptage $N_n(\mathbf{A}, \cdot)$. Afin de prouver que $\widetilde{M}_n(\mathbf{A}, \cdot)$ est aussi une martingale, on doit seulement vérifier la condition d'intégrabilité,

$$\mathbb{E}_{\nu_0} \left[\int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 d\mathcal{A}_n(\mathbf{A}, s) \right] < +\infty.$$

On rappelle que λ est borné sur $\mathbf{A} \times [0, t^*(\mathbf{A})[$ d'après le lemme 2.11. C désigne un majorant de λ sur cet ensemble. Alors,

$$\begin{aligned} \int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 d\mathcal{A}_n(\mathbf{A}, s) &= \sum_{i=0}^{n-1} \int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 \mathbb{1}_{\{S_{i+1} \geq s\}} \mathbb{1}_{\{Z_i \in \mathbf{A}\}} \lambda(Z_i, s) ds \\ &\leq \int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 C \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq s\}} \mathbb{1}_{\{Z_i \in \mathbf{A}\}} ds. \end{aligned}$$

De plus, par (2.12) et (2.13),

$$\int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 d\mathcal{A}_n(\mathbf{A}, s) \leq C \int_0^t Y_n(\mathbf{A}, s)^+ ds. \quad (2.24)$$

Comme $Y_n(\mathbf{A}, \cdot)^+$ est borné par 1 d'après (2.14), on a

$$\int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 d\mathcal{A}_n(\mathbf{A}, s) \leq Ct.$$

Ainsi,

$$\mathbb{E}_{\nu_0} \left[\int_0^t \left(Y_n(\mathbf{A}, s)^+ \right)^2 d\mathcal{A}_n(\mathbf{A}, s) \right] < +\infty.$$

Cela montre que $\widetilde{M}_n(\mathbf{A}, \cdot)$ est une martingale. Son crochet $\langle \widetilde{M}_n(\mathbf{A}) \rangle$ est donné par,

$$\forall 0 \leq s < t^*(\mathbf{A}), \langle \widetilde{M}_n(\mathbf{A}) \rangle(s) = \int_0^s \left(Y_n(\mathbf{A}, u)^+ \right)^2 d\mathcal{A}_n(\mathbf{A}, u).$$

Donc, par (2.24),

$$\forall 0 \leq s < t^*(\mathbf{A}), \langle \widetilde{M}_n(\mathbf{A}) \rangle(s) \leq C \int_0^s Y_n(\mathbf{A}, u)^+ du.$$

D'après le lemme 2.14, $Y_n(A, \cdot)^+$ est borné par 1 et pour tout u , $Y_n(A, u)^+$ tend presque sûrement vers 0. Donc, par le théorème de convergence dominée,

$$\langle \widetilde{M}_n(A) \rangle(s) \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0,$$

quand n tend vers l'infini. Cela termine la preuve. \square

On a établi le résultat précédent dans le but d'appliquer l'inégalité de Lengart à la martingale $\widetilde{M}_n(A, \cdot)$. Une référence à ce sujet peut être trouvée dans [5], II.5.2.1. Lengart's inequality.

REMARQUE 2.18 *D'après l'inégalité de Lengart et le résultat précédent, pour tous $A \in \mathcal{B}_v^+$, $0 \leq t < t^*(A)$, et $x \in E$,*

$$\sup_{0 \leq s \leq t} \left| \widetilde{M}_n(A, s) \right| \xrightarrow{\mathbb{P}_x} 0.$$

Soit $A \in \mathcal{B}_v^+$. On propose $\widehat{L}_n(A, \cdot)$ comme estimateur de la fonction $L(A, \cdot)$ définie par (2.19). Il est donné par,

$$\forall 0 \leq t < t^*(A), \widehat{L}_n(A, t) = \int_0^t Y_n(A, s)^+ dN_n(A, s). \quad (2.25)$$

$\widehat{L}_n(A, \cdot)$ est un estimateur de type Nelson-Aalen de $L(A, \cdot)$. Les résultats qui suivent concernent son comportement asymptotique. On verra aussi qu'on peut lisser cet estimateur afin d'obtenir un estimateur de sa dérivée $l(A, \cdot)$. Avant, introduisons la notation suivante, pour tout $n \geq 0$,

$$\forall 0 \leq t < t^*(A), L_n^*(A, t) = \int_0^t l(A, s) \mathbb{1}_{\{Y_n(A, s) > 0\}} ds. \quad (2.26)$$

On rappelle que $Y_n(A, \cdot)$ ainsi que son inverse ont été définis par (2.12) et (2.13).

PROPOSITION 2.19 *Soient $A \in \mathcal{B}_v^+$, $0 < t < t^*(A)$ et $x \in E$. Alors,*

$$\sup_{0 \leq s \leq t} \left| \widehat{L}_n(A, s) - L_n^*(A, s) \right| \xrightarrow{\mathbb{P}_x} 0,$$

quand n tend vers l'infini.

PREUVE. La définition de $\widetilde{M}_n(\mathbf{A}, s)$ (2.23), l'expression de $M_n(\mathbf{A}, s)$ (2.22) et la définition de $\widehat{L}_n(\mathbf{A}, s)$ (2.25) mènent à

$$\widetilde{M}_n(\mathbf{A}, s) = \widehat{L}_n(\mathbf{A}, s) - \int_0^s Y_n(\mathbf{A}, \mathbf{u})^+ \lambda(Z_i, \mathbf{u}) \mathbb{1}_{\{Z_i \in \mathbf{A}\}} \mathbb{1}_{\{S_{i+1} \geq \mathbf{u}\}} d\mathbf{u}.$$

Alors, $\widetilde{M}_n(\mathbf{A}, s)$ peut s'écrire de la façon suivante,

$$\widetilde{M}_n(\mathbf{A}, s) = \widehat{L}_n(\mathbf{A}, s) - L_n^*(\mathbf{A}, s) - \mathbf{a}_n(s), \quad (2.27)$$

où le terme additionnel $\mathbf{a}_n(s)$ est donné par

$$\mathbf{a}_n(s) = \int_0^s Y_n(\mathbf{A}, \mathbf{u})^+ \sum_{i=0}^{n-1} [\lambda(Z_i, \mathbf{u}) - l(\mathbf{A}, \mathbf{u})] \mathbb{1}_{\{Z_i \in \mathbf{A}\}} \mathbb{1}_{\{S_{i+1} \geq \mathbf{u}\}} d\mathbf{u}. \quad (2.28)$$

Alors,

$$\sup_{0 \leq s \leq t} \left| \widehat{L}_n(\mathbf{A}, s) - L_n^*(\mathbf{A}, s) \right| \leq \sup_{0 \leq s \leq t} \left| \widetilde{M}_n(\mathbf{A}, s) \right| + \sup_{0 \leq s \leq t} |\mathbf{a}_n(s)|.$$

On a déjà établi dans la remarque 2.18 que $\sup_{0 \leq s \leq t} \left| \widetilde{M}_n(\mathbf{A}, s) \right|$ tend en probabilité vers 0. Par conséquent, on a juste à étudier $\sup_{0 \leq s \leq t} |\mathbf{a}_n(s)|$. Avec (2.12), on a

$$|\mathbf{a}_n(s)| \leq \int_0^t \mathbb{1}_{\{Y_n(\mathbf{A}, \mathbf{u}) > 0\}} \left| Y_n(\mathbf{A}, \mathbf{u})^+ \sum_{i=0}^{n-1} \lambda(Z_i, \mathbf{u}) \mathbb{1}_{\{Z_i \in \mathbf{A}\}} \mathbb{1}_{\{S_{i+1} \geq \mathbf{u}\}} - l(\mathbf{A}, \mathbf{u}) \right| d\mathbf{u}.$$

L'intégrale converge presque sûrement vers 0 d'après la proposition 2.15. De plus, d'après le lemme 2.11, il existe une constante $C > 0$, qui est un majorant de λ sur $\mathbf{A} \times [0, t^*(\mathbf{A})]$. Ainsi, pour $\mathbf{u} \leq t$,

$$\begin{aligned} \left| Y_n(\mathbf{A}, \mathbf{u})^+ \sum_{i=0}^{n-1} \lambda(Z_i, \mathbf{u}) \mathbb{1}_{\{Z_i \in \mathbf{A}\}} \mathbb{1}_{\{S_{i+1} \geq \mathbf{u}\}} - l(\mathbf{A}, \mathbf{u}) \right| &\leq |l(\mathbf{A}, \mathbf{u})| + C \\ &\leq K_t(\mathbf{A}) + C, \end{aligned}$$

où $K_t(\mathbf{A})$ a déjà été défini par (2.18). Par conséquent, on peut appliquer le théorème de convergence dominée. On obtient

$$\lim_{n \rightarrow +\infty} \sup_{0 \leq s \leq t} |\mathbf{a}_n(s)| \xrightarrow{\mathbb{P}_{x\text{-p.s.}}} 0,$$

ce qui montre le résultat recherché. □

THÉORÈME 2.20 Soient $A \in \mathcal{B}_v^+$, $0 < t < t^*(A)$ et $x \in E$. Alors,

$$\sup_{0 \leq s \leq t} |\widehat{L}_n(A, s) - L(A, s)| \xrightarrow{\mathbb{P}_x} 0,$$

quand n tend vers l'infini.

PREUVE. Par l'inégalité triangulaire, on a

$$|\widehat{L}_n(A, s) - L(A, s)| \leq |\widehat{L}_n(A, s) - L_n^*(A, s)| + |L_n^*(A, s) - L(A, s)|.$$

D'après la proposition 2.19, il suffit de montrer que $\sup_{0 \leq s \leq t} |L_n^*(A, s) - L(A, s)|$ tend en probabilité vers 0. Par (2.18), (2.19) et (2.26),

$$\sup_{0 \leq s \leq t} |L_n^*(A, s) - L(A, s)| \leq K_t(A) \int_0^t \mathbb{1}_{\{Y_n(A, r) = 0\}} dr,$$

et le majorant tend vers 0 par le lemme 2.14. \square

REMARQUE 2.21 D'après l'équation (2.27), la différence $\widehat{L}_n(A, t) - L_n^*(A, t)$ se décompose comme la somme d'une martingale $\widetilde{M}_n(A, t)$ et d'un terme de reste $\alpha_n(t)$. Certes, des résultats comme le théorème de Rebolledo (voir [5], Theorem II.5.1) permettent d'obtenir la vitesse de convergence pour la partie martingale, mais le terme additionnel $\alpha_n(t)$ – défini par (2.28), – semble beaucoup plus difficile à traiter.

Soit $A \in \mathcal{B}_v^+$. On a donné un estimateur de $L(A, \cdot)$ sur l'intervalle $[0, t^*(A)[$ (voir le théorème 2.20). On va montrer que $L(A, \cdot)$ et le taux cumulé $\Lambda(\xi, \cdot)$, donné par (2.4), sont proches pour tout $\xi \in A$ si A est suffisamment petit. On peut aussi voir que $l(A, \cdot)$, donné par (2.17), et le taux de saut $\lambda(\xi, \cdot)$ sont proches pour tout $\xi \in A$.

LEMME 2.22 Soient $A \in \mathcal{B}_v^+$, $z \in A$ et $0 \leq s < t^*(A)$. On a

$$\begin{aligned} |\lambda(z, s) - l(A, s)| &\leq [\lambda]_{\text{Lip}} \text{diam } A, \\ |\Lambda(z, s) - L(A, s)| &\leq s [\lambda]_{\text{Lip}} \text{diam } A. \end{aligned}$$

PREUVE. Tout d'abord, on a

$$\begin{aligned}
|\lambda(z, s) - l(A, s)| &= \left| \lambda(z, s) - \frac{\int_A f(\xi, s) \nu(d\xi)}{\int_A G(\xi, s) \nu(d\xi)} \right| \\
&\leq \frac{1}{\int_A G(\xi, s) \nu(d\xi)} \int_A |\lambda(z, s) - \lambda(\xi, s)| G(\xi, s) \nu(d\xi) \\
&\leq [\lambda]_{\text{Lip}} \text{diam } A.
\end{aligned}$$

De plus, en intégrant le résultat obtenu, on a $|\Lambda(z, s) - L(A, s)| \leq s [\lambda]_{\text{Lip}} \text{diam } A$. \square

Remarquons qu'on ne peut définir $L(A, \cdot)$ que si $\nu(A)$ est strictement positif. Par conséquent, nous devons estimer l'indicatrice $\mathbb{1}_{\{\nu(A) > 0\}}$. Pour cela, on estime la mesure invariante $\nu(A)$ par sa version empirique,

$$\widehat{\nu}_n(A) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}}.$$

LEMME 2.23 *Par le théorème ergodique, $\widehat{\nu}_n(A) \rightarrow \nu(A)$ presque sûrement. De plus,*

$$\mathbb{1}_{\{\widehat{\nu}_n(A) > n^{-1/2}\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \mathbb{1}_{\{\nu(A) > 0\}},$$

quand n tend vers l'infini, pour tout $x \in E$.

PREUVE. On distingue les cas $\nu(A) > 0$ et $\nu(A) = 0$.

- (i) Si $\nu(A) > 0$, $\widehat{\nu}_n(A)$ a une limite presque sûre qui est strictement positive. Alors, pour n assez grand, $\widehat{\nu}_n(A) > n^{-1/2}$ presque sûrement.
- (ii) Si $\nu(A) = 0$, le nombre de visites dans A de la chaîne de Markov (Z_n) est presque sûrement fini, car cette chaîne de Markov est positive Harris récurrente. Dans ce cas, la somme $\sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}}$ converge presque sûrement vers une somme finie.

Cela montre le résultat voulu. \square

Dans la remarque suivante, on se focalise sur le comportement asymptotique de $\sup_{0 \leq s \leq t} \widehat{L}_n(A, s)$ lorsque $\nu(A) = 0$. C'est important car le théorème 2.20 n'est valable que lorsque $\nu(A) > 0$.

REMARQUE 2.24 Soit $A \in \mathcal{B}(E)$ tel que $\bar{A} \cap \partial E = \emptyset$ et $\nu(A) = 0$. Soient $0 \leq s \leq t < t^*(A)$ et $x \in E$. Par (2.25) et (2.14), on a

$$\sup_{0 \leq s \leq t} \widehat{L}_n(A, s) \leq \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}} \mathbb{P}_x\text{-p.s.}$$

De plus, le nombre de visites dans A est presque sûrement fini puisque $\nu(A) = 0$. Donc, $\sup_{0 \leq s \leq t} \widehat{L}_n(A, s)$ converge presque sûrement vers une somme finie.

Pour estimer Λ sur $\mathcal{K} \times [0, t]$ où \mathcal{K} est un compact inclus dans E , on considère une partition assez fine (A_k) de \mathcal{K} et on peut alors estimer $\Lambda(\xi, s)$ par $\widehat{L}_n(A_j, s)$, pour $\xi \in A_j$.

THÉORÈME 2.25 Soient \mathcal{K} un sous-ensemble compact de E et $\xi \in E$. Pour tous $\varepsilon, \eta > 0$, il existe un entier N et une partition finie $P = (A_k)$ de \mathcal{K} , tels que, pour tous $n \geq N$ et $0 < t < \min_k t^*(A_k)$,

$$\mathbb{P}_\xi \left(\sup_{x \in \mathcal{K}} \sup_{0 \leq s \leq t} \left| \sum_{k=1}^{|\mathcal{P}|} \widehat{L}_n(A_k, s) \mathbb{1}_{\{\widehat{\nu}_n(A_k) > \frac{1}{\sqrt{n}}\}} \mathbb{1}_{\{x \in A_k\}} - \Lambda(x, s) \mathbb{1}_{\{x \in \mathcal{K}'\}} \right| > \eta \right) < \varepsilon,$$

où \mathcal{K}' est défini par

$$\mathcal{K}' = \bigcup_{\nu(A_k) > 0} A_k.$$

PREUVE. On fixe s et x . On a

$$\begin{aligned} & \sum_{k=1}^{|\mathcal{P}|} \widehat{L}_n(A_k, s) \mathbb{1}_{\{\widehat{\nu}_n(A_k) > \frac{1}{\sqrt{n}}\}} \mathbb{1}_{\{x \in A_k\}} - \Lambda(x, s) \mathbb{1}_{\{x \in \mathcal{K}'\}} \\ &= \sum_{k=1}^{|\mathcal{P}|} \widehat{L}_n(A_k, s) \mathbb{1}_{\{\widehat{\nu}_n(A_k) > \frac{1}{\sqrt{n}}\}} \mathbb{1}_{\{x \in A_k\}} - \sum_{k=1}^{|\mathcal{P}|} \Lambda(x, s) \mathbb{1}_{\{x \in A_k\}} \mathbb{1}_{\{\nu(A_k) > 0\}} \\ &= \sum_{k=1}^{|\mathcal{P}|} \widehat{L}_n(A_k, s) \mathbb{1}_{\{\widehat{\nu}_n(A_k) > \frac{1}{\sqrt{n}}\}} \mathbb{1}_{\{x \in A_k\}} \mathbb{1}_{\{\nu(A_k) = 0\}} \\ &+ \sum_{k=1}^{|\mathcal{P}|} \left(\widehat{L}_n(A_k, s) \mathbb{1}_{\{\widehat{\nu}_n(A_k) > \frac{1}{\sqrt{n}}\}} - \Lambda(x, s) \right) \mathbb{1}_{\{x \in A_k\}} \mathbb{1}_{\{\nu(A_k) > 0\}}. \end{aligned}$$

On note l l'unique entier entre 1 et $|P|$ tel que x est dans A_l . On distingue les deux cas $\nu(A_l) > 0$ et $\nu(A_l) = 0$. Par hypothèse sur \mathcal{K} , A_l est relativement compact et tel que $\overline{A_l} \cap \partial E = \emptyset$.

- (i) Si $\nu(A_l) = 0$, $\widehat{L}_n(A_l, s)$ tend uniformément presque sûrement vers une somme finie, par la remarque 2.24. Alors, par le lemme 2.23,

$$\sup_{0 \leq s \leq t} \widehat{L}_n(A_l, s) \mathbb{1}_{\{\widehat{\nu}_n(A_l) > \frac{1}{\sqrt{n}}\}} \xrightarrow{\mathbb{P}_x} 0.$$

- (ii) Si $\nu(A_l) > 0$. On a, par l'inégalité triangulaire,

$$\begin{aligned} \left| \widehat{L}_n(A_l, s) \mathbb{1}_{\{\widehat{\nu}_n(A_l) > \frac{1}{\sqrt{n}}\}} - \Lambda(x, s) \right| &\leq \widehat{L}_n(A_l, s) \left| \mathbb{1}_{\{\widehat{\nu}_n(A_l) > n^{-1/2}\}} - 1 \right| \\ &\quad + \left| \widehat{L}_n(A_l, s) - L(A_l, s) \right| \\ &\quad + \left| L(A_l, s) - \Lambda(x, s) \right|. \end{aligned}$$

- (a) Le premier terme tend uniformément vers 0 en probabilité, puisque la différence $|\mathbb{1}_{\{\widehat{\nu}_n(A_l) > n^{-1/2}\}} - 1|$ tend presque sûrement vers 0 par le lemme 2.23, et car $\widehat{L}_n(A_l, \cdot)$ tend uniformément en probabilité vers $L(A_l, \cdot)$, selon le théorème 2.20.
- (b) Le second terme tend uniformément vers 0 en probabilité d'après le théorème 2.20.
- (c) Le troisième terme est borné par $t[\lambda]_{\text{Lip}} \max_k \text{diam } A_k$, selon le lemme 2.22, ce qui ne dépend pas de x et est arbitrairement petit.

Puisque la somme est finie, ceci termine la démonstration. \square

REMARQUE 2.26 *Il apparaît dans la démonstration du théorème 2.25 que l'erreur entre le taux de saut cumulé $\Lambda(x, t)$ et son estimateur se décompose comme la somme d'une erreur statistique et d'une erreur déterministe. D'après le lemme 2.22, cette dernière est linéaire en le diamètre des éléments de la partition (A_k) considérée.*

2.4.2 Estimation de λ

Dans cette partie, on s'intéresse au lissage de l'estimateur $\widehat{L}_n(A, \cdot)$ de $L(A, \cdot)$ afin de déterminer un estimateur à noyau consistant de $l(A, \cdot)$ (voir la proposition 2.28), et, par conséquent de λ (voir le théorème 2.30).

Soit K un noyau continu dont le support est $[-1, 1]$. On introduit les notations suivantes. Pour tout réel $b > 0$ et pour tout $0 < t < t^*(A)$, on note,

$$\forall 0 \leq u \leq t, \widehat{l}_{n,b,t}(A, u) = \frac{1}{b} \int_0^t K\left(\frac{u-s}{b}\right) d\widehat{L}_n(A, s). \quad (2.29)$$

Il s'agit d'un estimateur à noyau de $l(A, \cdot)$ sur l'intervalle $[0, t]$. On note aussi,

$$\forall 0 \leq u \leq t, l_{n,b,t}^*(A, u) = \frac{1}{b} \int_0^t K\left(\frac{u-s}{b}\right) dL_n^*(A, s).$$

On commence par démontrer le lemme suivant.

LEMME 2.27 Soient $A \in \mathcal{B}_v^+$, $b > 0$ et $0 < r < t < t^*(A)$. On a

$$\sup_{0 \leq s \leq r} \left| \widehat{l}_{n,b,t}(A, s) - l_{n,b,t}^*(A, s) \right| \leq \frac{2}{b} V(K) \sup_{0 \leq s \leq t} \left| \widehat{L}_n(A, s) - L_n^*(A, s) \right|,$$

où $V(K)$ désigne la variation totale du noyau K .

PREUVE. On note,

$$\forall 0 \leq s \leq t, g(s) = \widehat{L}_n(A, s) - L_n^*(A, s).$$

Alors, par une intégration par parties, puisque $g(0) = 0$,

$$K((s-t)/b)g(t) = \int_0^t K((s-u)/b) dg(u) - \int_0^t g(u-) dK((s-u)/b).$$

Par l'inégalité triangulaire,

$$b \sup_{0 \leq s \leq r} \left| \widehat{l}_{n,b,t}(A, s) - l_{n,b,t}^*(A, s) \right| \leq \sup_{0 \leq s \leq r} |K((s-t)/b)g(t)| + \sup_{0 \leq s \leq r} \int_0^t g(u-) dK((s-u)/b).$$

Pour tout $v \notin [-1, 1]$, $K(v) = 0$. Cela induit que

$$\begin{aligned} |K((s-t)/b)g(t)| &\leq \sup_{0 \leq u \leq t} |g(u)| K((s-t)/b) \\ &= \sup_{0 \leq u \leq t} |g(u)| \left\{ K((s-t)/b) - K(v) \right\} \\ &\leq \sup_{0 \leq u \leq t} |g(u)| V(K). \end{aligned}$$

De plus,

$$\begin{aligned} \int_0^t g(u-) dK((s-u)/b) &= \lim_{\max\{s_{i+1}-s_i\} \rightarrow 0} \sum_{i=1}^p g(s_i) \left\{ K((s-s_{i+1})/b) - K((s-s_i)/b) \right\} \\ &\leq \sup_{0 \leq u \leq t} |g(u)| V(K), \end{aligned}$$

ce qui permet de conclure. □

PROPOSITION 2.28 Soient $A \in \mathcal{B}_v^+$ et $0 < r_1 < r_2 < t < t^*(A)$. Il existe une suite (β_n) , qui converge presque sûrement vers 0, telle que

$$\sup_{r_1 \leq s \leq r_2} \left| \widehat{l}_{n,\beta_n,t}(A, s) - l(A, s) \right| \xrightarrow{\mathbb{P}_x} 0,$$

quand n tend vers l'infini, pour tout $x \in E$.

PREUVE. Par l'inégalité triangulaire, on a pour toute suite (b_n) ,

$$\left| \widehat{l}_{n,b_n,t}(A, s) - l(A, s) \right| \leq \left| \widehat{l}_{n,b_n,t}(A, s) - l_{n,b_n,t}^*(A, s) \right| + \left| l_{n,b_n,t}^*(A, s) - l(A, s) \right|. \quad (2.30)$$

On considère la suite (b_n) définie par,

$$\forall n \geq 0, b_n = \sqrt{\sup_{0 \leq s \leq t} \left| \widehat{L}_n(A, s) - L_n^*(A, s) \right|}.$$

D'après la proposition 2.19, cette suite tend en probabilité vers 0, et on a

$$\sup_{0 \leq s \leq t} \left| \widehat{L}_n(A, s) - L_n^*(A, s) \right| \stackrel{\mathbb{P}_x}{=} o(b_n), \quad (2.31)$$

quand n tend vers l'infini. Alors, par le lemme 2.27,

$$\sup_{r_1 \leq s \leq r_2} \left| \widehat{l}_{n,b_n,t}(A, s) - l_{n,b_n,t}^*(A, s) \right| \xrightarrow{\mathbb{P}_x} 0, \quad (2.32)$$

quand n tend vers l'infini. Maintenant, on peut regarder le comportement asymptotique de $\left| l_{n,b_n,t}^*(A, s) - l(A, s) \right|$. Par le changement de variable $s - b_n v = u$, on a

$$l_{n,b_n,t}^*(A, s) = \int_{(s-t)/b_n}^{s/b_n} K(v) \mathbb{1}_{\{Y_n(A, s-b_n v) > 0\}} l(A, s - b_n v) dv.$$

Comme (b_n) tend en probabilité vers 0, il existe une sous-suite

$$(\beta_n) = (b_{\alpha(n)}),$$

qui tend presque sûrement vers 0. Soit

$$\Omega_1 = \left\{ \omega \in \Omega : \lim_{n \rightarrow +\infty} \beta_n(\omega) = 0 \text{ et } \lim_{n \rightarrow +\infty} \mathbb{1}_{\{Y_n(A, r_2, \omega) = 0\}} = 0 \right\}. \quad (2.33)$$

D'après ce qui précède et (2.16), $\mathbb{P}_x(\Omega_1) = 1$. Soient $\omega \in \Omega_1$ et N entier tel que pour tout $n \geq N$,

$$r_1/\beta_n(\omega) \geq 1 \quad \text{et} \quad (r_2 - t)/\beta_n(\omega) \leq -1.$$

Pour tout $n \geq N$, on a

$$\frac{s}{\beta_n(\omega)} \geq \frac{r_1}{\beta_n(\omega)} \geq 1 \quad \text{et} \quad \frac{s - t}{\beta_n(\omega)} \leq \frac{r_2 - t}{\beta_n(\omega)} \leq -1.$$

Donc,

$$l_{n,\beta_n(\omega),t}^*(A, s, \omega) = \int_{-1}^1 K(v) \mathbb{1}_{\{Y_n(A, s - \beta_n(\omega)v, \omega) > 0\}} l(A, s - \beta_n(\omega)v) dv,$$

car le support de K est $[-1, 1]$. Ainsi,

$$\begin{aligned} & \left| l_{n,\beta_n(\omega),t}^*(A, s, \omega) - l(A, s) \right| \\ & \leq \left| \int_{-1}^1 K(u) \left(l(A, s - \beta_n(\omega)u) - l(A, s) \right) du \right| \\ & \quad + \left| \int_{-1}^1 K(v) \left(\mathbb{1}_{\{Y_n(A, s - \beta_n(\omega)v, \omega) > 0\}} - 1 \right) l(A, s - \beta_n(\omega)v) dv \right|. \end{aligned} \quad (2.34)$$

On va prouver que les termes

$$\sup_{r_1 \leq s \leq r_2} \left| \int_{-1}^1 K(u) \left(l(A, s - \beta_n(\omega)u) - l(A, s) \right) du \right|$$

et

$$\sup_{r_1 \leq s \leq r_2} \left| \int_{-1}^1 K(v) \left(\mathbb{1}_{\{Y_n(A, s - \beta_n(\omega)v, \omega) > 0\}} - 1 \right) l(A, s - \beta_n(\omega)v) dv \right|$$

convergent vers 0 pour tout $\omega \in \Omega_1$. On commence par le second terme.

(i) On rappelle que K et $l(A, \cdot)$ sont deux fonctions continues. Soit C défini par

$$C = K_t(A) \sup_{-1 \leq v \leq 1} K(v),$$

où $K_t(A)$ a déjà été défini par (2.18). Alors, on a

$$\begin{aligned} & \left| \int_{-1}^1 K(v) \left(\mathbb{1}_{\{Y_n(A, s - \beta_n(\omega)v, \omega) > 0\}} - 1 \right) l(A, s - \beta_n(\omega)v) dv \right| \\ & \leq C \int_{-1}^1 \left| \mathbb{1}_{\{Y_n(A, s - \beta_n(\omega)v, \omega) > 0\}} - 1 \right| dv \\ & \leq C \int_{-1}^1 \mathbb{1}_{\{Y_n(A, s - \beta_n(\omega)v, \omega) = 0\}} dv. \end{aligned} \quad (2.35)$$

Le changement de variable $u = s - \beta_n(\omega)v$ mène à

$$\begin{aligned} \int_{-1}^1 \mathbb{1}_{\{Y_n(\mathcal{A}, s - \beta_n(\omega)v, \omega) = 0\}} dv &\leq \frac{1}{\beta_n(\omega)} \int_{s - \beta_n(\omega)}^{s + \beta_n(\omega)} \mathbb{1}_{\{Y_n(\mathcal{A}, u, \omega) = 0\}} du \\ &\leq \frac{1}{\beta_n(\omega)} \int_{r_2 - \beta_n(\omega)}^{r_2 + \beta_n(\omega)} \mathbb{1}_{\{Y_n(\mathcal{A}, u, \omega) = 0\}} du, \end{aligned} \quad (2.36)$$

car $u \mapsto \mathbb{1}_{\{Y_n(\mathcal{A}, u, \omega) = 0\}}$ est croissante. De plus, par la définition de l'intégrale de Riemann pour les fonctions continues par morceaux, on a quand n tend vers l'infini,

$$\int_{r_2 - \beta_n(\omega)}^{r_2 + \beta_n(\omega)} \mathbb{1}_{\{Y_n(\mathcal{A}, u, \omega) = 0\}} du \sim \beta_n(\omega) \left(\mathbb{1}_{\{Y_n(\mathcal{A}, r_2^-, \omega) = 0\}} + \mathbb{1}_{\{Y_n(\mathcal{A}, r_2^+, \omega) = 0\}} \right).$$

Donc, toujours quand n tend vers l'infini,

$$\frac{1}{\beta_n(\omega)} \int_{r_2 - \beta_n(\omega)}^{r_2 + \beta_n(\omega)} \mathbb{1}_{\{Y_n(\mathcal{A}, u, \omega) = 0\}} du \sim \mathbb{1}_{\{Y_n(\mathcal{A}, r_2^-, \omega) = 0\}} + \mathbb{1}_{\{Y_n(\mathcal{A}, r_2^+, \omega) = 0\}}. \quad (2.37)$$

Par définition de Ω_1 (voir (2.33)), quand n tend vers l'infini,

$$\mathbb{1}_{\{Y_n(\mathcal{A}, r_2, \omega) = 0\}} \rightarrow 0. \quad (2.38)$$

Finalement, (2.35), (2.36), (2.37) et (2.38) montrent que

$$\sup_{r_1 \leq s \leq r_2} \left| \int_{-1}^1 K(v) \left(\mathbb{1}_{\{Y_n(\mathcal{A}, s - \beta_n(\omega)v, \omega) > 0\}} - 1 \right) l(\mathcal{A}, s - \beta_n(\omega)v) dv \right|$$

tend vers 0 pour presque tout ω , puisque $\mathbb{P}_x(\Omega_1) = 1$.

(ii) Par le théorème de convergence dominée, pour tout $\omega \in \Omega_1$,

$$\sup_{r_1 \leq s \leq r_2} \left| \int_{-1}^1 K(u) \left(l(\mathcal{A}, s - \beta_n(\omega)u) - l(\mathcal{A}, s) \right) du \right| \rightarrow 0,$$

quand n tend vers l'infini, car $l(\mathcal{A}, \cdot)$ est continue.

Finalement, par (2.34), lorsque n tend vers l'infini,

$$\sup_{r_1 \leq s \leq r_2} |l_{n, \beta_n, t}^*(\mathcal{A}, s) - l(\mathcal{A}, s)| \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0.$$

La convergence a donc aussi lieu en probabilité. Avec (2.30) et (2.32), on a le résultat recherché. \square

REMARQUE 2.29 La proposition 2.28 établit l'existence d'une suite (aléatoire) (β_n) telle que l'estimateur $\widehat{l}_{n,\beta_n,t}(A, \cdot)$ tend vers $l(A, \cdot)$, sur tout compact $[r_1, r_2]$, avec $0 < r_1 < r_2 < t$ et $t < t^*(A)$. Néanmoins, ce résultat ne donne aucun moyen d'écrire explicitement une telle suite (en particulier, c'est une suite extraite). Toutefois, il apparaît dans la démonstration que toute suite déterministe (b_n) satisfaisant l'équation (2.31) est une suite adéquate. Déterminer une suite de fenêtres revient donc à étudier la vitesse de convergence de la différence $\widehat{L}_n(A, t) - L_n^*(A, t)$. Ce point délicat a été discuté dans la remarque 2.21.

Finalement, pour tout $A \in \mathcal{B}_v^+$, on a donné un estimateur $\widehat{l}_{n,\beta_n,t}(A, \cdot)$ de $l(A, \cdot)$ sur $[0, t]$ pour tout $0 < t < t^*(A)$. Cependant, c'est le taux de saut λ qui nous intéresse. Puisque λ est Lipschitz, si A est assez petit, alors $l(A, \cdot)$ et $\lambda(\xi, \cdot)$ sont proches pour $\xi \in A$ (voir le lemme 2.22). Cela mène à notre principal résultat de convergence.

THÉORÈME 2.30 Soient \mathcal{K} un sous-ensemble compact de E et $\xi \in E$. Pour tous $\varepsilon, \eta > 0$, il existe un entier N et une partition $P = (A_k)$ de \mathcal{K} tels que, pour tout $0 < t < \min_k t^*(A_k)$, il existe pour chaque k une suite $(\beta_n(A_k))$ (elle dépend de t) qui converge presque sûrement vers 0, telle que, pour tous $n \geq N$ et $0 < r_1 < r_2 < t$,

$$\mathbb{P}_\xi \left(\sup_{x \in \mathcal{K}} \sup_{r_1 \leq s \leq r_2} \left| \sum_{k=1}^{|\mathcal{P}|} \widehat{l}_{n,\beta_n(A_k),t}(A_k, s) \mathbb{1}_{\{\widehat{\nu}_n(A_k) > \frac{1}{\sqrt{n}}\}} \mathbb{1}_{\{x \in A_k\}} - \lambda(x, s) \mathbb{1}_{\{x \in \mathcal{K}'\}} \right| > \eta \right) < \varepsilon,$$

où \mathcal{K}' est défini par

$$\mathcal{K}' = \bigcup_{\nu(A_k) > 0} A_k.$$

PREUVE. C'est une conséquence de la proposition 2.28 et du lemme 2.22. La preuve est similaire à celle du théorème 2.25. \square

REMARQUE 2.31 Si le sous-ensemble compact \mathcal{K} est « proche » de E , alors, pour toute partition (A_k) de \mathcal{K} , le minimum des $t^*(A_k)$ est petit. Dans ce cas, on peut estimer le taux de saut sur une grande partie de l'espace d'état mais sur un intervalle de temps court. A contrario, si on choisit un ensemble \mathcal{K} petit et centré dans E , on estime le taux de saut sur une petite partie de l'espace mais sur un intervalle de temps grand.

2.5 Illustration numérique

Cette section est consacrée à une étude de simulation, dans le but d'illustrer les résultats de convergence établis dans les théorèmes 2.25 et 2.30. L'exemple que nous considérons peut être directement relié avec l'application pratique donnée dans la section 2.2.

On considère un processus de renouvellement marqué (X_t) défini sur l'intervalle $E =]0, 60[$ et partant du point 30. Au temps t , X_t modélise la configuration de température en degré Celsius d'une machine de production. La température est constante par morceaux. Quand une défaillance a lieu, un technicien répare instantanément la machine et change la configuration de température : il tente d'obtenir le régime normal à 20°C. Cependant, il fait une erreur aléatoire dont la variance dépend de la différence entre la température précédente et la température recherchée. On s'intéresse à l'estimation du taux de panne dans la configuration normale de température. Les trois caractéristiques λ , Q et t^* de (X_t) sont données, pour tout $x \in E$, par

- pour tout $A \in \mathcal{B}(E)$,

$$Q(x, A) = \frac{1}{K_x} \int_A \mathbb{1}_E(y) \exp\left(-\frac{(y-20)^2}{2\sigma_x^2}\right),$$

avec $\sigma_x = 0.5 + |x - 20|$,

- pour tout $t \geq 0$, $\lambda(x, t) = 3 + 0.05x$,
- $t^*(x) = 1$,

où K_x est la constante de normalisation. Ce noyau de transition satisfait de manière immédiate la condition de Doeblin. Par conséquent, l'hypothèse 2.4 est vérifiée pour la chaîne immergée (Z_n) . De plus, le taux cumulé Λ associé à λ est donné par $\Lambda(x, t) = (3 + 0.05x)t$. Une trajectoire simulée d'un tel processus est donnée dans la figure 2.1.

On simule des longues trajectoires du processus : l'observation de 200, 300 ou 400 sauts est disponible pour estimer le taux de défaillance λ et son intégrale Λ . Comme on l'a mentionné précédemment, on s'intéresse à l'estimation de $\Lambda(x, t)$ et $\lambda(x, t)$ pour $x = 20$. On choisit d'approcher $\lambda(x, t)$ par la fonction $l(A, t)$, avec $A = \{y \in E : |y - x| \leq 2\}$. Comme $t^*(A) = 1$, on estime $l(A, t)$ par (2.29) qui est une intégrale entre les temps 0 et 0.9. Donc, le résultat de convergence établi dans le théorème 2.30 est valide sur tout compact inclus dans $]0, 0.9[$. On présente notre étude sur l'intervalle $[0, 0.8]$ pour l'estimation de Λ et sur le compact $[0.2, 0.8] \subset]0, 0.9[$ pour λ .

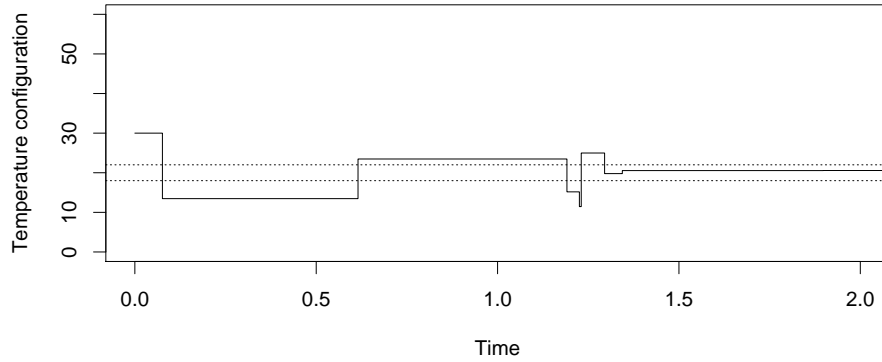


FIGURE 2.1 – Une trajectoire du processus de renouvellement simulée jusqu’à l’instant 2. Les temps pris en compte pour estimer le taux de défaillance à 20°C sont compris entre les deux lignes en pointillés.

Pour chaque trajectoire simulée, le nombre de visites dans A est proche de 73.5% du nombre de sauts observés. De plus, la fenêtre de lissage $\beta_n(A)$ que nous avons choisie peut s’écrire

$$\beta_n(A) = \frac{1}{h_n(A)^\alpha},$$

où $h_n(A)$ désigne le nombre (aléatoire) de visites dans A , et $\alpha = 1/4$. La figure 2.2 présente l’estimation des deux fonctions d’intérêt à partir de l’observation de 400 sauts. De plus, on s’intéresse à l’évaluation de l’erreur pour différentes tailles d’échantillon : 200, 300 et 400 sauts observés. Pour chaque taille d’échantillon, on génère 100 répliques du modèle. On donne dans la figure 2.3 les diagrammes en boîte de l’erreur quadratique intégrée (ISE pour integrated square error en anglais) entre les fonctions d’intérêt et leur estimation. Sans surprise, l’erreur est décroissante quand le nombre de sauts observés augmente.

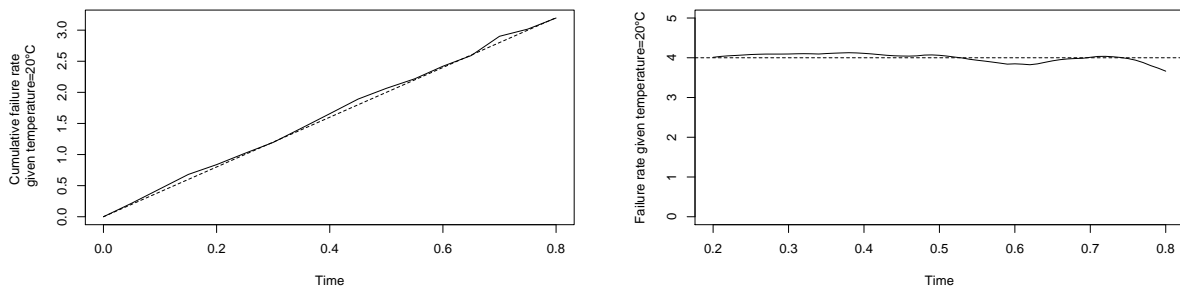


FIGURE 2.2 – Estimation du taux de défaillance cumulé (à gauche) et du taux de défaillance (à droite) sachant que la température est de 20°C à partir de l’observation de 400 sauts. Les estimations sont en ligne pleine, alors que les vrais taux sont en pointillés.

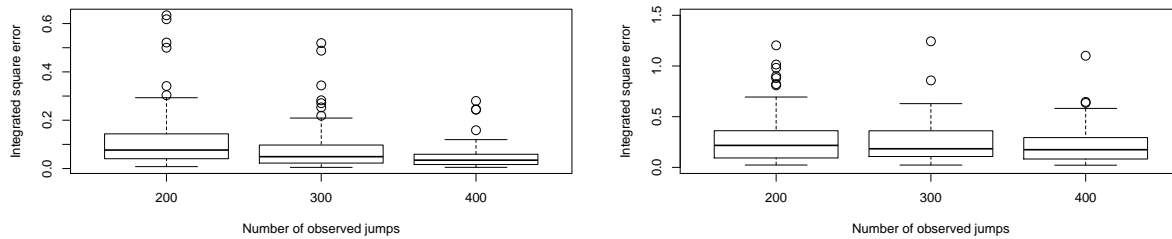


FIGURE 2.3 – Diagrammes en boîte de l’erreur quadratique intégrée pour l’estimation du taux cumulé (à gauche) et du taux de défaillance (à droite) sachant que la température est de 20°C .

2.6 Conclusion du chapitre

Dans ce chapitre, on a développé une méthode non paramétrique d’estimation du taux cumulé et du taux de saut pour une classe générale de processus de renouvellement marqués. Les deux difficultés majeures dans ce travail sont la censure déterministe dépendant de l’état, ainsi que l’absence de conditions sur l’espace d’état du processus.

Sous des hypothèses directement reliées aux caractéristiques du processus, on a montré la convergence uniforme en probabilité sur tout compact de nos estimateurs. Les résultats numériques apparaissent satisfaisants pour des tailles d’échantillon raisonnables, et la méthode est facile à implémenter. Toutefois, ce travail a deux limites importantes : le choix de la discrétisation de l’espace, ainsi que celui de la fenêtre de lissage. Pour d’éventuelles applications, il semble important d’étudier plus en avant – d’un point de vue théorique mais aussi numérique, – l’effet de ces paramètres sur la qualité de l’estimation.

Enfin, l’étude présentée ici est un point clé pour estimer la loi conditionnelle des temps inter-arrivées d’un processus markovien déterministe par morceaux dans le chapitre suivant.

2.7 Preuves de la section 2.2

2.7.1 Preuve de la proposition 2.1

Nous commençons par établir la première indépendance conditionnelle. Soient h_1, \dots, h_n des fonctions mesurables bornées de \mathbb{R}_+ dans \mathbb{R} . De (2.2), on a

$$\begin{aligned}
 \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] &= \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) \mathbb{E}_{\nu_0} [h_n(S_n) | \mathcal{G}_n \vee \sigma(S_1, \dots, S_{n-1})] | \mathcal{G}_n] \\
 &= \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) \mathbb{E}_{\nu_0} [h_n(S_n) | \mathcal{G}_n] | \mathcal{G}_n] \\
 &= \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_n] \mathbb{E}_{\nu_0} [h_n(S_n) | \mathcal{G}_n].
 \end{aligned} \tag{2.39}$$

De plus, par (2.1),

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_{n-1} \vee \sigma(\varepsilon_{n-2})] | \mathcal{G}_n].$$

Le produit $h_1(S_1) \dots h_{n-1}(S_{n-1})$ est $\mathcal{G}_{n-1} \vee \sigma(\delta_0, \dots, \delta_{n-2})$ -mesurable et

$$\sigma(\varepsilon_{n-2}) \perp \mathcal{G}_{n-1} \vee \sigma(\delta_0, \dots, \delta_{n-2}).$$

Avec l'équation (3) donnée dans [19, page 308], on obtient

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_{n-1}) | \mathcal{G}_{n-1} \vee \sigma(\varepsilon_{n-2})] = \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_{n-1}].$$

Finalement, on a

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_{n-1}].$$

A la lumière de (2.39), on a

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_{n-1}] \mathbb{E}_{\nu_0} [h_n(S_n) | \mathcal{G}_n].$$

Une récurrence immédiate montre que

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] = \prod_{i=1}^n \mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_i].$$

De plus, en prenant pour $j \neq i$, $h_j = \mathbb{1}$, on a

$$\mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_i].$$

Ainsi,

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] = \prod_{i=1}^n \mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_n].$$

Donc, on a

$$\bigvee_{j \neq i} \sigma(S_j) \perp_{\mathcal{G}_n} \sigma(S_i),$$

ce qui entraîne immédiatement le résultat voulu. La seconde indépendance conditionnelle est immédiate à partir de (2.2).

2.7.2 Preuve du lemme 2.2

Soit $0 \leq s < t < t^*(Z_i)$. Afin de prouver que M^{i+1} est une martingale, on doit montrer que

$$\begin{aligned} & \mathbb{E}_{\nu_0} \left[N^{i+1}(t) - \int_0^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du \mid \sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right] = M^{i+1}(s) \\ \Leftrightarrow & \mathbb{E}_{\nu_0} \left[\mathbb{1}_{\{s < S_{i+1} \leq t\}} - \int_s^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du \mid \sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right] = 0 \\ \Leftrightarrow & \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} \mid \sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right] \\ & = \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\int_s^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du \mid \sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right]. \end{aligned} \quad (2.40)$$

On commence par montrer que l'égalité (2.40) est équivalente à

$$\begin{aligned} & \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} \mid \sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right] \\ & = \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\int_s^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du \mid \sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right]. \end{aligned} \quad (2.41)$$

D'après le lemme 6.2 de [58], si les conditions suivantes sont vérifiées,

$$\{S_{i+1} > s\} \in \left(\sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right) \cap \left(\sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right)$$

et

$$\{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right) = \{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right),$$

alors, (2.40) et (2.41) sont équivalentes. La première condition est évidemment vérifiée. Pour la seconde, puisque $\sigma(N^{i+1}(s))$ est une sous-tribu de \mathcal{F}_s^{i+1} , on a

$$\{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right) \supset \{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right).$$

Par conséquent, on a seulement à prouver l'inclusion réciproque,

$$\{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \mathcal{F}_s^{i+1} \right) \subset \{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right).$$

On considère le λ -système (voir par exemple la définition 1.10 de [59]) \mathcal{D} donné par

$$\mathcal{D} = \left\{ C \in \sigma(Z_i) \vee \mathcal{F}_s^{i+1} : \{S_{i+1} > s\} \cap C \in \{S_{i+1} > s\} \cap \left(\sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right) \right\},$$

et le π -système (voir la définition 1.1 de [59]) \mathcal{C} défini par

$$\mathcal{C} = \left\{ \Omega, \{S_{i+1} \leq u\} \cap \{Z_i \in B\}, \{Z_i \in B\} : 0 \leq u \leq s, B \in \mathcal{B}(E) \right\}.$$

On a immédiatement que $\mathcal{C} \subset \mathcal{D}$ et $\sigma(\mathcal{C}) = \sigma(Z_i) \vee \mathcal{F}_s^{i+1}$. De plus, $\sigma(Z_i) \vee \mathcal{F}_s^{i+1} \subset \mathcal{D}$ d'après le théorème des classes monotones (voir par exemple le théorème 1.19 de [59]). De la définition de \mathcal{D} , on a l'inclusion réciproque recherchée. Par conséquent, les égalités (2.40) and (2.41) sont équivalentes. Maintenant, nous allons vérifier que (2.41) est vraie. D'un côté, on a

$$\begin{aligned} & \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} \mid \sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right] \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{P}_{\nu_0} (S_{i+1} \leq t \mid Z_i, S_{i+1} > s) \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \frac{\mathbb{P}_{\nu_0} (s < S_{i+1} \leq t \mid Z_i)}{\mathbb{P}_{\nu_0} (S_{i+1} > s \mid Z_i)}. \end{aligned}$$

Des définitions de f (2.3) et de G (2.5), on déduit

$$\begin{aligned} & \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} \mid \sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right] \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \frac{\int_s^t f(Z_i, u) du}{G(Z_i, s)}. \end{aligned}$$

De l'autre côté,

$$\begin{aligned} & \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\int_s^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du \mid \sigma(Z_i) \vee \sigma(N^{i+1}(s)) \right] \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \int_s^t \lambda(Z_i, u) \mathbb{P}_{\nu_0} (S_{i+1} \geq u \mid Z_i, S_{i+1} > s) du \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \int_s^t \lambda(Z_i, u) \frac{\mathbb{P}_{\nu_0} (S_{i+1} \geq u \mid Z_i)}{\mathbb{P}_{\nu_0} (S_{i+1} > s \mid Z_i)} du. \end{aligned}$$

Donc, puisque $u \leq t < t^*(Z_i)$,

$$\begin{aligned} & \mathbb{1}_{\{S_{i+1} > s\}} \mathbb{E}_{\nu_0} \left[\int_s^t \lambda(Z_i, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du \mid \sigma(Z_i) \vee \sigma(\mathbf{N}^{i+1}(s)) \right] \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \frac{\int_s^t \lambda(Z_i, u) G(Z_i, u) du}{G(Z_i, s)} \\ &= \mathbb{1}_{\{S_{i+1} > s\}} \frac{\int_s^t f(Z_i, u) du}{G(Z_i, s)}. \end{aligned}$$

On a donc montré (2.41), ainsi que le résultat recherché.

2.8 Preuves de la section 2.4

2.8.1 Preuve de la proposition 2.5

1. Soient $x \in E$ et $\nu_0 = \delta_{\{x\}}$. Alors, d'après l'hypothèse 2.4,

$$\lim_{n \rightarrow +\infty} \|Q^n(x, \cdot) - \nu\|_{TV} = 0.$$

On considère $A \in \mathcal{B}(E)$ tel que $\nu(A) > 0$. Alors, $Q^n(x, A) > 0$ pour n suffisamment grand. D'après la proposition 4.2.1 de [67], la chaîne de Markov (Z_n) est, par conséquent, ν -irréductible.

2. De plus, par le théorème 4.3.3 de [53], (Z_n) est positive Harris-récurrente et apériodique.
3. Une chaîne positive Harris-récurrente admet une unique (à une constante multiplicative près) loi invariante (voir, par exemple, l'introduction de [52]).

2.8.2 Preuve du lemme 2.7

Soit g une fonction mesurable bornée par 1. D'après le lemme 2.6,

$$\left| \int_{E \times \mathbb{R}_+} g(z, s) [\eta_n(dz \times ds) - \eta(dz \times ds)] \right| \leq \|h\|_\infty \|\nu_n - \nu\|_{TV},$$

par le théorème de Fubini, où la fonction h est donnée par,

$$\forall z \in E, \quad h(z) = \int_{\mathbb{R}_+} g(z, s) \mu_z(ds).$$

De plus, puisque $|h(z)| \leq 1$ pour tout z , on a

$$\left| \int_{E \times \mathbb{R}_+} g(z, s) (\eta_n(dz \times ds) - \eta(dz \times ds)) \right| \leq \|\nu_n - \nu\|_{TV}.$$

Donc, par l'hypothèse 2.4, $\|\eta_n - \eta\|_{TV}$ tend vers 0.

2.8.3 Preuve du lemme 2.10

Par hypothèse, t^* est continue. De plus, pour tout $\xi \notin \partial E$, $t^*(\xi) > 0$. Puisque \bar{A} est compact et tel que $\bar{A} \cap \partial E = \emptyset$, on a

$$\inf_{\xi \in \bar{A}} t^*(\xi) > 0.$$

Ainsi,

$$\inf_{\xi \in A} t^*(\xi) \geq \inf_{\xi \in \bar{A}} t^*(\xi) > 0.$$

Pour tout s , $\lambda(\cdot, s)$ est continue car Lipschitz. De plus, $\lambda(\xi, \cdot)$ est bornée par la fonction M qui ne dépend pas de ξ , et est localement intégrable. On rappelle qu'on a

$$G(\xi, t) = \exp\left(-\int_0^t \lambda(\xi, s) ds\right).$$

Donc, par le théorème de Lebesgue de continuité sous le signe intégrale, $G(\cdot, t)$ est continue, et vérifie $G(\xi, t) > 0$ pour tout ξ . Ainsi,

$$\inf_{\xi \in A} G(\xi, t) \geq \inf_{\xi \in \bar{A}} G(\xi, t) > 0.$$

Les deux inégalités sont donc démontrées.

2.9 Application à des données biologiques

Dans cette section, on présente brièvement une étude statistique menée sur des données biologiques réelles avec Raphaël Coudret et Gilles Durrieu. Ce travail a été soumis pour publication sous le titre *A hidden renewal model for monitoring aquatic systems biosensors* [8]. Il s'agit de modéliser le comportement d'un mollusque bivalve par un processus de renouvellement et d'estimer ses caractéristiques afin d'inférer l'état de santé de l'animal.

Le laboratoire EPOC (Environnements et Paléoenvironnements Océaniques et Continentaux), situé à Arcachon (Gironde), est impliqué dans divers projets d'observation et d'étude de l'environnement marin. C'est dans cet esprit que des chercheurs ont mis au point un système embarqué d'électrodes pour analyser en temps réel le comportement de mollusques bivalves (des huîtres notamment) placés dans leur milieu naturel. Les biologistes obtiennent ainsi des signaux comme celui présenté dans la figure 2.4, donnant à chaque instant la distance entre les deux parties de la coquille de l'animal. L'huître est un animal très simple, seulement capable de s'ouvrir et de se fermer. L'enregistrement de telles données permet donc de connaître son comportement global, et donc l'état du milieu dans lequel il vit. Citons notamment les articles [75, 82] pour plus de détails sur ces problématiques, et [57, 79, 80] pour des analyses statistiques de ce type de données.

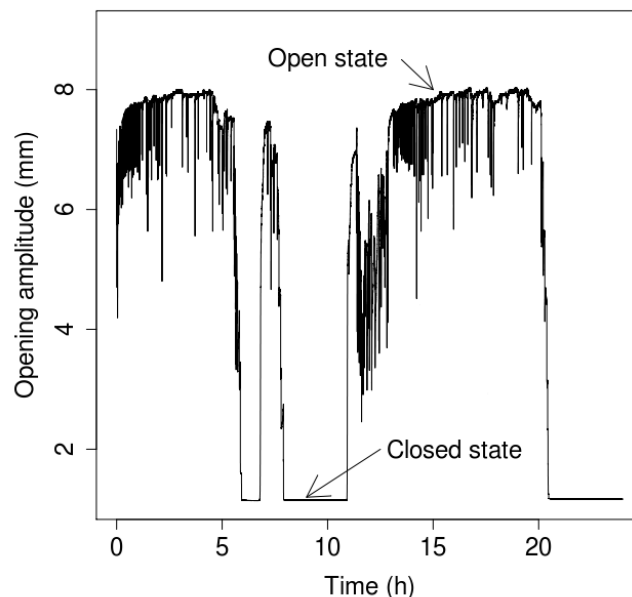


FIGURE 2.4 – Enregistrement valvométrique d'une huître placée dans le bassin d'Arcachon à Eyrac (Gironde) le 27 février 2007.

Depuis plusieurs années, le comportement de différents animaux – implantés dans plusieurs sites en France et à l'étranger, – est ainsi enregistré à la fréquence de 0.625 Hertz (c'est-à-dire toutes les 1.6 secondes) et est transféré, *via* le réseau GSM, dans une base de données accessible depuis Internet. Les données obtenues sont très nombreuses, bruitées, et par conséquent, difficiles à analyser.

On observe sur la figure 2.4 que le signal valvométrique peut se découper en deux parties : une partie haute (lorsque la coquille est ouverte) et une partie basse (lorsque la coquille est fermée). Des biologistes ont proposé de dire qu'un mollusque est ouvert lorsque la distance entre les deux parties de sa coquille est supérieure à un certain pourcentage (faible) de l'amplitude du signal. Le choix de ce seuil est important, notamment dans la gestion des micro-fermetures (lorsque l'animal se ferme sur une période très brève avant de se réouvrir). Raphaël Coudret a proposé au cours de sa thèse d'estimer ce seuil de manière moins arbitraire comme l'unique anti-mode de la densité du signal valvométrique (voir notamment [27]). Cela nous permet de décrire la dynamique au cours du temps de l'animal comme une succession d'états *ouvert*, *fermé* et *en transition*. Ainsi, nous avons proposé de modéliser le comportement d'une huître par un processus de renouvellement sur l'espace $E = \{\mathcal{O}, \mathcal{C}, \mathcal{T}_{\mathcal{O} \rightarrow \mathcal{C}}, \mathcal{T}_{\mathcal{C} \rightarrow \mathcal{O}}\}$, où \mathcal{O} (respectivement \mathcal{C}) désigne l'état ouvert (respectivement fermé) de l'animal alors que \mathcal{T} correspond à l'un des deux états de transition (voir la figure 2.5).

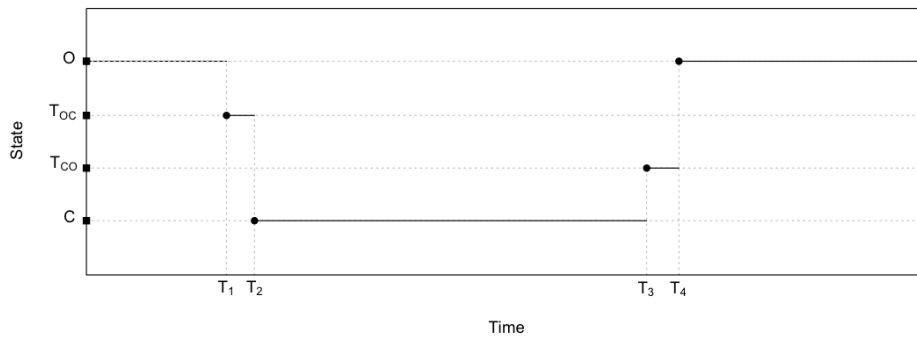


FIGURE 2.5 – Exemple de trajectoire du processus de renouvellement sous-jacent modélisant le comportement d'une huître.

Les temps de transition sont très courts et considérés comme peu informatifs. Nous nous sommes intéressés à l'estimation non paramétrique de la loi des temps passés dans les états ouvert et fermé, ainsi qu'à l'estimation de la loi des temps de premier retour dans l'état de transition $\mathcal{T}_{\mathcal{O} \rightarrow \mathcal{C}}$, pour différents animaux à partir de l'enregistrement de leur comportement pendant une semaine. Nous avons montré dans ce chapitre que l'hypothèse d'intensité multiplicative est vérifiée pour ce type de modèle; nous avons donc estimé le taux de saut cumulé $\Lambda(x, t)$, pour $x = \mathcal{O}$ ou $x = \mathcal{C}$, par la méthode de Nelson-Aalen. Nous avons préféré représenter l'estimateur de Fleming-Harrington de la fonction de survie $G(x, t)$ associée à $\Lambda(x, t)$ (voir la figure 2.6). Il est donné par,

$$\forall t \geq 0, \widehat{G}_n(x, t) = \exp\left(-\widehat{\Lambda}_n(x, t)\right),$$

où $\widehat{\Lambda}_n(x, t)$ est l'estimateur de Nelson-Aalen de $\Lambda(x, t)$. La fonction de survie estimée des temps de premier retour dans l'état $\mathcal{T}_{\mathcal{O} \rightarrow \mathcal{C}}$ est représentée dans la figure 2.7.

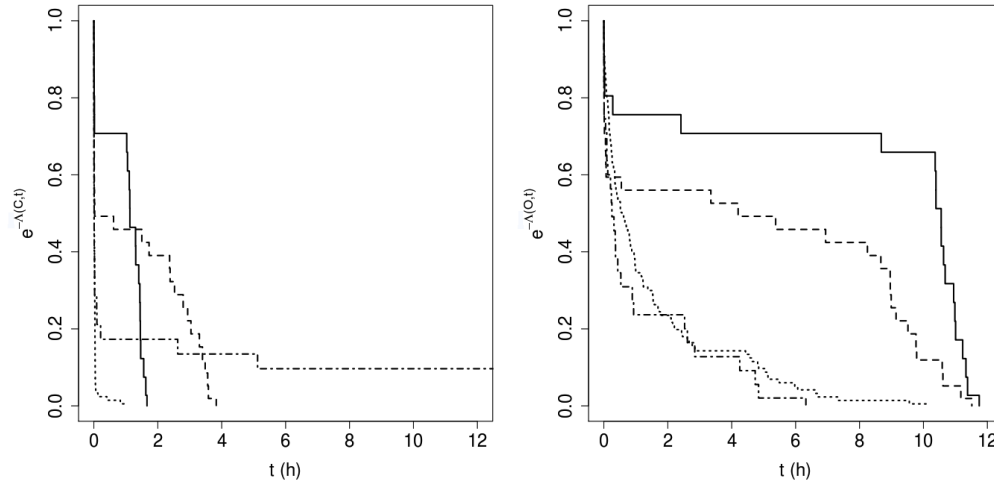


FIGURE 2.6 – Estimation non paramétrique de la fonction de survie $G(x, t)$ pour $x = \mathcal{C}$ (à gauche) et $x = \mathcal{O}$ (à droite) pour différents animaux.

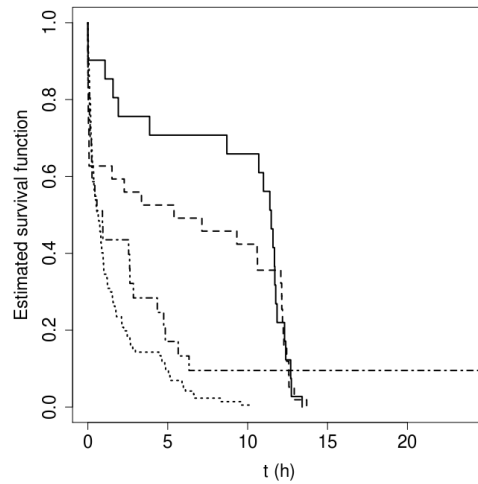


FIGURE 2.7 – Estimation non paramétrique de la fonction de survie des temps de premier retour dans l'état $\mathcal{T}_{\mathcal{O} \rightarrow \mathcal{C}}$ pour différents animaux.

On observe sur les figures 2.6 et 2.7 des comportements variés des fonctions de survie estimées. Ils correspondent à des états de santé différents des animaux ; ceux-ci sont déterminés *a priori* puisqu'on connaît la suite de la vie de l'animal grâce aux enregistrements

des données. Sur la figure 2.7, les deux fonctions de survie estimées présentant une forte décroissance autour de 12 heures sont celles d’huîtres en bonne santé ; ces animaux ont un comportement très lié à celui de la marée dont la périodicité est de 12 heures environ (voir notamment [80]). Les autres fonctions de survie sont celles d’animaux dont l’état de santé est mauvais ou vivant dans un environnement inhospitalier.

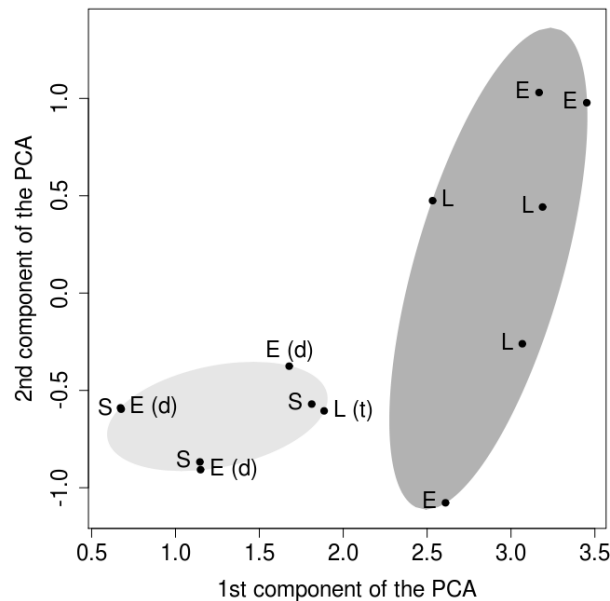


FIGURE 2.8 – Projection des fonctions de survie estimées sur le premier plan principal de l’analyse en composantes principales pour 15 animaux dont l’état de santé est déterminé *a priori*. Les huîtres sont désignées par leur position géographique : E (Eyrac), L (Locmariaquer), S (Santander), L(t) (Locmariaquer, animal affecté par des variations de température) et E(d) (Eyrac, animal mourant).

Nous avons proposé de classer les courbes obtenues sur un lot de 15 animaux, dont l’état de santé est connu *a priori*. Une analyse en composante principale est représentée sur la figure 2.8. Le premier axe explique 88% de la variabilité. Les deux groupes présentés ont été obtenus par l’algorithme de classification des k-means. Le groupe de droite est composé uniquement d’animaux supposés sains : E, huîtres vivant dans la baie d’Eyrac (Gironde) ou L, à Locmariaquer (Morbihan). Le groupe de gauche est formé d’animaux mourants E(d) situés à Eyrac, d’un animal L(t) vivant à Locmariaquer et affecté par de fortes variations de température, et d’huîtres S placées dans le port industriel de Santander (Espagne).

Ainsi, une procédure simple de classification permet de retrouver les états de santé des animaux étudiés à partir des estimations de la fonction de survie conditionnelle. Il semble donc que les temps passés dans les états ouvert et fermé des huîtres soient de bons indicateurs de leur santé, et donc de la qualité de leur environnement. Bien sûr, ce travail n'est qu'une étape et demande à être affiné et complété. L'étape de classification, notamment, devra être validée par de nouvelles études.

ESTIMATION DE LA LOI CONDITIONNELLE DES TEMPS INTER-SAUTS D'UN PDMP

Ce chapitre est consacré à l'analyse d'une méthode d'estimation non paramétrique de la loi conditionnelle des temps inter-sauts d'un processus markovien déterministe par morceaux, observé en temps long. La procédure d'estimation, les résultats et les simulations font l'objet de l'article *Nonparametric estimation of the conditional distribution of the inter-jumping times for piecewise-deterministic Markov processes* [9], avec François Dufour et Anne Gégout-Petit, soumis pour publication. Cette étude s'appuie en partie sur des résultats présentés dans le chapitre précédent sur l'estimation du taux de saut pour des processus de renouvellement.

Le chapitre est organisé de la façon suivante. Une introduction et les liens avec le chapitre 2 sont donnés dans la section 3.1. On présente le problème ainsi que les résultats principaux dans la section 3.2. Dans la section 3.3, on donne des résultats préliminaires techniques, alors que la section 3.4 est dédiée aux démonstrations des principaux résultats de ce chapitre. Une étude de simulations à travers deux applications en biologie est présentée dans la section 3.5. Enfin, la conclusion et des perspectives à ces investigations sont données dans la section 3.6. La preuve technique d'un résultat d'indépendance conditionnelle est reportée dans la section 3.7.

SOMMAIRE DU CHAPITRE

3.1	Introduction	53
3.2	Formulation du problème	55
3.2.1	Définition du processus	55
3.2.2	Hypothèses	56
3.2.3	Résultats principaux	59
3.3	Considérations préliminaires	62
3.3.1	Loi conditionnelle de S_{n+1} sachant Z_n et Z_{n+1}	62
3.3.2	Ergodicité	67
3.4	Démonstrations des résultats principaux	69
3.4.1	Estimation de \tilde{l}	70
3.4.2	Approximation de f	75
3.4.3	Estimation de \tilde{H}	79
3.5	Simulations	83
3.5.1	Mouvement d'une bactérie	83
3.5.2	Un modèle de croissance-fragmentation	86
3.6	Conclusion du chapitre	88
3.7	Preuve de la proposition 3.29	89

3.1 Introduction

Dans ce chapitre, on s'intéresse à l'estimation non paramétrique de la loi conditionnelle des temps inter-sauts d'un PDMP, à partir d'une observation du processus en temps long. Une partie de l'approche développée ici est basée sur les méthodes mises en œuvre dans le chapitre 2 de ce manuscrit. Nous allons notamment utiliser l'ergodicité des chaînes de Markov sous-jacentes et étudier des compensateurs de processus de comptage.

Une idée intéressante pour estimer le taux de saut λ d'un PDMP pourrait être de trouver un processus de comptage dont l'intensité stochastique ferait intervenir la composée $\lambda \circ \Phi$ comme dans l'étude menée pour les processus de renouvellement dans le chapitre 2 (voir le théorème 2.3). Malheureusement, si le lemme 2.2 reste vrai pour les PDMP, on ne peut plus sommer ces martingales de telle sorte que la somme reste une martingale dans une filtration bien choisie. La raison est la suivante : la nouvelle position Z_n du PDMP lors du $n^{\text{ème}}$ saut dépend – à travers le flot – du temps inter-saut précédent S_n . Par conséquent, la nouvelle position peut apporter de l'information sur la durée inter-saut précédente ; c'est l'indépendance conditionnelle établie dans la proposition 2.1 qui n'est plus valable pour les PDMP. On a alors l'intuition que c'est la loi des durées S_n sachant les positions précédente Z_{n-1} et suivante Z_n qu'il faut étudier pour conserver des propriétés de martingale à temps continu. C'est pourquoi une partie de ce chapitre est consacrée à l'analyse de la loi conditionnelle des temps inter-sauts sachant la position passée et la position future du PDMP.

Tout au long de cette étude, la présence des sauts forcés lorsque le processus atteint la frontière de l'espace d'état entraîne de nombreuses difficultés techniques. Cette caractéristique, introduite par Davis dans [31], fait la richesse des PDMP et se retrouve dans de nombreuses applications. On peut citer le modèle de capacité-expansion proposé par Davis dans [32]. On peut aussi mentionner l'article [54] : les auteurs développent un modèle de type PDMP pour décrire le mécanisme de production d'un antibiotique par une bactérie. Les sauts forcés sont utilisés pour modéliser les changements déterministes intervenant lorsque la concentration en nutriments atteint un certain seuil. Dans le domaine de la fiabilité, citons [6] dans lequel les auteurs proposent un modèle de dégradation avec des défaillances de type seuil. Une autre application peut être trouvée dans [36]. Dans ce papier, les auteurs s'intéressent au problème d'arrêt optimal pour un PDMP modélisant l'état d'une structure soumise à corrosion. Dans le livre [56], on peut trouver des formules de vraisemblance pour l'observation d'un PDMP sans sauts déterministes. Pour cette classe de modèles sans sauts à la frontière, on pourrait en déduire une procédure d'estimation

dans le cas paramétrique. Notre approche est différente et adaptée à la fois au cadre non paramétrique et aux PDMP avec sauts déterministes.

Notre objectif principal est de traiter le problème de l'estimation non paramétrique de la densité conditionnelle f des temps inter-sauts. Notre stratégie peut être divisée en quatre étapes détaillées dans la suite. La première étape consiste à donner une approximation spatiale de f faisant intervenir deux quantités d'intérêt : les fonctions $\tilde{l}(A, B, t)$ et $\tilde{H}(A, B, t)$, où A et B sont des ensembles mesurables. Sous nos hypothèses, on peut montrer que les chaînes de Markov immergées admettent une mesure invariante (voir les propositions 3.19 et 3.22), ce qui nous permet de donner une interprétation de $\tilde{l}(A, B, t)$ et $\tilde{H}(A, B, t)$. En effet, si ν désigne la loi invariante de la chaîne de Markov (Z_n) , la quantité $\tilde{l}(A, B, t)$ peut être interprétée comme le taux de saut de A vers B au temps t et sous la loi ν . De plus, $\tilde{H}(A, B, t)$ est exactement la probabilité conditionnelle $\mathbf{P}_\nu(S_1 > t, Z_1 \in B | Z_0 \in A)$. La qualité de cette approximation de f est établie dans la proposition 3.6. La seconde étape concerne l'estimation de la probabilité conditionnelle $\tilde{H}(A, B, t)$ par sa version empirique. La convergence uniforme est présentée dans la proposition 3.8. La troisième étape – plus compliquée – consiste à estimer la fonction $\tilde{l}(A, B, t)$. Nous montrons qu'il s'agit d'une approximation spatiale d'une fonction $\tilde{\lambda}(x, y, t)$ pour $x \in A$ et $y \in B$ (voir le lemme 3.37). Nous établissons dans la proposition 3.15 que $\tilde{\lambda}$ est en fait le taux de saut du temps S_{n+1} sachant les positions précédente et future Z_n et Z_{n+1} . En particulier, il est loin d'être évident que cette loi admet un taux de saut. Il apparaît alors naturel d'utiliser la méthodologie mise en œuvre dans le chapitre 2 pour estimer l'approximation $\tilde{l}(A, B, t)$. Cependant, cela n'est pas complètement évident pour deux raisons. Premièrement, le taux de saut $\tilde{\lambda}$ dépend d'une double marque spatiale. Deuxièmement, il est nécessaire de montrer un résultat supplémentaire d'indépendance conditionnelle pour appliquer certains résultats du chapitre 2 (voir la proposition 3.29). La convergence uniforme de notre estimateur non paramétrique de $\tilde{l}(A, B, t)$ est présentée dans la proposition 3.7. La dernière étape consiste à montrer dans le théorème 3.9 la convergence globale de notre estimateur de la densité conditionnelle d'intérêt.

La suite de ce chapitre est organisée comme suit. Dans la section 3.2, on présente le problème de manière détaillée. On rappelle dans la sous-section 3.2.1 la définition d'un PDMP sur un espace métrique. Dans la sous-section 3.2.2, on donne l'ensemble des hypothèses qui seront nécessaires dans la suite, ainsi que des commentaires et motivations sur ces hypothèses. Puis, dans la sous-section 3.2.3, on explique les grandes étapes de notre procédure d'estimation de la densité d'intérêt (voir les propositions 3.6, 3.7 et 3.8). Le résultat de consistance principal est donné dans le théorème 3.9. Dans la section 3.3, on donne des résultats préliminaires techniques. On introduit un nouveau processus dans la

sous-section 3.3.1 et on donne l'expression de son taux de saut $\tilde{\lambda}$ (voir la proposition 3.15). Dans la sous-section 3.3.2, on établit des résultats d'ergodicité des chaînes de Markov sous-jacentes. La section 3.4 est consacrée à la présentation des démonstrations des résultats de la sous-section 3.2.2. Finalement, une étude de simulations numériques est présentée dans la section 3.5 pour illustrer le bon comportement de notre estimateur. La conclusion et des perspectives à cette étude sont données dans la section 3.6. La démonstration de la proposition 3.29 est reportée dans la section 3.7.

3.2 Formulation du problème

3.2.1 Définition du processus

On présente ici la définition d'un PDMP sur un espace métrique. Le processus évolue dans un ouvert E d'un espace métrique séparable (\mathcal{E}, d) . La dynamique peut être décrite par les caractéristiques (λ, Q, Φ) .

- $\Phi : E \times \mathbb{R}_+ \rightarrow \mathcal{E}$ est le flot déterministe. Il vérifie,

$$\forall \xi \in E, \forall s, t \geq 0, \Phi(\xi, t + s) = \Phi(\Phi(\xi, t), s).$$

En particulier, $\Phi(\xi, 0) = \xi$. Pour tout $\xi \in E$, $t^*(\xi)$ désigne le temps de sortie déterministe de E :

$$t^*(\xi) = \inf\{t > 0 : \Phi(\xi, t) \in \partial E\},$$

avec la convention habituelle $\inf \emptyset = +\infty$.

- $\lambda : \bar{E} \rightarrow \mathbb{R}_+$ est le taux de saut. Il s'agit d'une fonction mesurable qui vérifie,

$$\forall \xi \in E, \exists \varepsilon > 0, \int_0^\varepsilon \lambda(\Phi(\xi, s)) ds < +\infty.$$

- Q est un noyau de Markov sur $(\bar{E}, \mathcal{B}(\bar{E}))$ qui satisfait,

$$\forall \xi \in \bar{E}, Q(\xi, E \setminus \{\xi\}) = 1.$$

La loi de la position initiale X_0 est ν_0 . Partant de $x \in E$, l'évolution du processus peut être décrite comme suit. T_1 est une variable aléatoire positive dont la fonction de survie est donnée par,

$$\forall t \geq 0, \mathbb{P}_{\nu_0}(T_1 > t | X_0 = x) = \exp\left(-\int_0^t \lambda(\Phi(x, s)) ds\right) \mathbb{1}_{\{0 \leq t < t^*(x)\}}.$$

Le temps de saut a lieu soit lorsque le flot atteint la frontière de l'espace d'état, soit avant avec le taux λ pris en la trajectoire du flot Φ . On choisit ensuite une variable aléatoire Z_1 sur E selon la loi $Q(\Phi(x, T_1), \cdot)$. La nouvelle position lors du saut dépend du temps T_1 *via* le flot déterministe partant de $X_0 = x$. La trajectoire entre les temps 0 et T_1 est donnée par

$$X_t = \begin{cases} \Phi(x, t) & \text{pour } 0 \leq t < T_1, \\ Z_1 & \text{pour } t = T_1. \end{cases}$$

Maintenant, partant de X_{T_1} , on génère le temps $S_2 = T_2 - T_1$ et la position Z_2 de la même manière que précédemment, et ainsi de suite. On obtient ainsi un processus markovien fort dont les temps de saut sont notés T_k (avec $T_0 = 0$). L'aléa d'un tel processus intervient seulement à travers le mécanisme des sauts. Les Z_n désignent les positions lors des sauts, alors que les S_n sont les temps inter-sauts : partant de $Z_n = X_{T_n}$, le nouveau saut aura lieu à l'instant $T_{n+1} = T_n + S_{n+1}$. La chaîne de Markov immergée associée au PDMP (X_t) est le processus à temps discret (Z_n, S_n) .

D'après [31] (Chapter 1, Section 24 Definition of the PDP), il existe deux fonctions mesurables φ et ψ , et deux suites indépendantes de variables aléatoires indépendantes (ε_n) et (δ_n) , telles que, pour tout $n \geq 1$,

$$\begin{cases} S_n = \varphi(Z_{n-1}, \delta_{n-1}), \\ Z_n = \psi(Z_{n-1}, S_n, \varepsilon_{n-1}). \end{cases} \quad (3.1)$$

On introduit les notations suivantes,

$$\begin{aligned} \forall \xi \in E, \quad \forall t \geq 0, \quad \bar{\lambda}(\xi, t) &= \lambda(\Phi(\xi, t)), \\ \forall \xi \in E, \quad \forall t \geq 0, \quad \forall A \in \mathcal{B}(E), \quad \bar{Q}(\xi, t, A) &= Q(\Phi(\xi, t), A). \end{aligned}$$

Dans la suite, on note f et G la densité et la fonction de survie associées au taux $\bar{\lambda}$.

$$\begin{aligned} \forall \xi \in E, \quad \forall t \geq 0, \quad G(\xi, t) &= \exp\left(-\int_0^t \bar{\lambda}(\xi, s) ds\right), \\ \forall \xi \in E, \quad \forall t \geq 0, \quad f(\xi, t) &= \bar{\lambda}(\xi, t)G(\xi, t). \end{aligned}$$

L'objet de ce chapitre est de donner un estimateur non paramétrique de la densité conditionnelle f , et d'établir la consistance de cet estimateur sous des hypothèses portant directement sur les caractéristiques du processus.

3.2.2 Hypothèses

Dans cette partie, on présente toutes les hypothèses nécessaires dans la suite, ainsi que des commentaires et motivations sur ces conditions. Elles peuvent être classées en deux

catégories : les hypothèses sur le noyau de transition et les hypothèses de régularité. Ces conditions sont peu restrictives, surtout quand l'espace d'état du PDMP est borné (voir les remarques 3.2 et 3.4). C'est le cas par exemple dans les applications étudiées dans la section 3.5 et dans les papiers [6, 32, 36, 54].

On note \mathcal{C}_1 la famille des ensembles mesurables dont l'intérieur est non vide,

$$\mathcal{C}_1 = \left\{ B \in \mathcal{B}(E) : \overset{\circ}{B} \neq \emptyset \right\}.$$

\mathcal{C}_2 désigne les ensembles relativement compacts de \mathcal{C}_1 dont l'intérieur est contenu dans E ,

$$\mathcal{C}_2 = \left\{ A \in \mathcal{C}_1 : A \text{ relativement compact tel que } \overline{A} \cap \partial E = \emptyset \right\}.$$

HYPOTHÈSES 3.1 *Hypothèses sur le noyau de transition.*

a) *Le noyau de transition \overline{Q} s'écrit de la façon suivante,*

$$\forall \xi \in E, \forall s \geq 0, \forall B \in \mathcal{B}(E), \overline{Q}(\xi, s, B) = \int_B \tilde{Q}(\xi, s, y) \mu(dy),$$

où μ est une mesure auxiliaire sur $(E, \mathcal{B}(E))$ telle que, pour tout ensemble mesurable B d'intérieur non-vide, $\mu(B) > 0$. De plus, $\tilde{Q}(x, t, y)$ est supposée continue en t et minorée uniformément, c'est-à-dire qu'il existe $m > 0$ tel que,

$$\forall x \in E, \forall s \geq 0, \forall y \in E, \tilde{Q}(x, s, y) \geq m. \quad (3.2)$$

b) *Pour tout $\varepsilon > 0$, il existe $p \geq 2$ et une famille d'ensembles disjoints $\{B_1, \dots, B_p\}$, $B_k \in \mathcal{C}_1$ pour tout k , tels que*

$$\text{diam } B_1 = \max_{1 \leq k \leq p} \text{diam } B_k < \varepsilon \quad \text{et} \quad \forall x \in E, \forall t \geq 0, \sum_{k=1}^p \overline{Q}(x, t, B_k) = 1.$$

REMARQUE 3.2 *Les hypothèses 3.1 peuvent paraître restrictives. Toutefois, l'existence d'une densité pour le noyau de Markov du PDMP est une condition naturelle, satisfaite dans de nombreuses applications. Par exemple, le PDMP qui modélise la croissance d'une cellule au cours du temps a des transitions gaussiennes (voir la sous-section 3.5.2). Un point plus sensible est l'hypothèse de minoration (3.2). Néanmoins, cette hypothèse est*

toujours vérifiée pour des noyaux gaussiens sur un espace borné par exemple. C'est le cas dans les deux applications développées dans la section 3.5. De plus, l'hypothèse 3.1 b) est également toujours vérifiée lorsque E est borné.

Par souci de clarté, on introduit les notations suivantes,

$$\begin{aligned}\forall \xi \in E, \forall s \geq 0, \forall y \in E, G\tilde{Q}(\xi, s, y) &= G(\xi, s)\tilde{Q}(\xi, s, y), \\ \forall \xi \in E, \forall s \geq 0, \forall y \in E, f\tilde{Q}(\xi, s, y) &= f(\xi, s)\tilde{Q}(\xi, s, y).\end{aligned}\quad (3.3)$$

De plus, on note aussi,

$$\forall \xi \in E, \forall s \geq 0, \forall B \in \mathcal{B}(E), G\bar{Q}(\xi, s, B) = G(\xi, s)\bar{Q}(\xi, s, B), \quad (3.4)$$

$$\forall \xi \in E, \forall s \geq 0, \forall B \in \mathcal{B}(E), f\bar{Q}(\xi, s, B) = f(\xi, s)\bar{Q}(\xi, s, B). \quad (3.5)$$

Dans la suite, on considère la distance d_2 sur \mathcal{E}^2 induite par la distance d et la norme de Manhattan sur \mathbb{R}^2 . Pour tous $(x, y), (u, v) \in \mathcal{E}^2$,

$$d_2((x, y), (u, v)) = d(x, u) + d(y, v).$$

$\text{diam}_2 A \times B_k$ désigne le diamètre de $A \times B_k \subset \mathcal{E}^2$ pour cette distance.

HYPOTHÈSES 3.3 Hypothèses de régularité.

a) $f(\xi, \cdot)$ est strictement positive et continue pour tout $\xi \in E$.

b) Il existe une fonction localement intégrable $M : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telle que,

$$\forall \xi \in E, \forall t \geq 0, \bar{\lambda}(\xi, t) \leq M(t).$$

c) $t^* : E \rightarrow \mathbb{R}_+$ est une fonction bornée.

d) Il existe une constante $[f\tilde{Q}]_{\text{Lip}} > 0$ telle que, pour tous $x, y, u, v \in E$ et pour tout $0 \leq t < t^*(x) \wedge t^*(u)$,

$$\left| f\tilde{Q}(x, t, y) - f\tilde{Q}(u, t, v) \right| \leq [f\tilde{Q}]_{\text{Lip}} d_2((x, y), (u, v)). \quad (3.6)$$

e) Il existe une constante $[f]_{\text{Lip}} > 0$ telle que,

$$\forall x, y \in E, \forall t \geq 0, |f(x, t) - f(y, t)| \leq [f]_{\text{Lip}} d(x, y).$$

f) Il existe une constante $[G\tilde{Q}]_{\text{Lip}} > 0$ telle que, pour tous $x, y, u, v \in E$,

$$\left| G\tilde{Q}(x, t^*(x), y) - G\tilde{Q}(u, t^*(u), v) \right| \leq [G\tilde{Q}]_{\text{Lip}} d_2((x, y), (u, v)). \quad (3.7)$$

REMARQUE 3.4 *Les conditions de régularité que nous imposons sont classiques et peu contraignantes, surtout lorsque l'espace d'état est borné. On peut citer l'article [15] dans lequel les auteurs font des hypothèses analogues.*

Avant de présenter les résultats principaux de ce chapitre, on donne un lemme technique mais utile dans la suite.

LEMME 3.5 *Soit $A \in \mathcal{C}_2$. On a alors $\inf_{\xi \in A} t^*(\xi) > 0$. On note cette quantité $t^*(A)$.*

PREUVE. La démonstration est similaire à celle du lemme 2.10. □

3.2.3 Résultats principaux

On considère un ensemble $A \in \mathcal{C}_2$ et $\epsilon > 0$. Cela implique l'existence de $p \geq 2$ et d'une famille $\{B_1, \dots, B_p\}$ (dépendant de ϵ) et satisfaisant l'hypothèse 3.1 b).

Notre stratégie pour estimer la densité conditionnelle f consiste à introduire deux nouvelles fonctions $\tilde{l}(A, B_k, t)$ et $\tilde{H}(A, B_k, t)$, pour $0 \leq t < t^*(A)$. Elles permettent d'approcher $f(\xi, t)$ pour $\xi \in A$, puisque $f(\xi, t)$ est proche en un certain sens de

$$\sum_{k=1}^p \tilde{l}(A, B_k, t) \tilde{H}(A, B_k, t)$$

(voir la proposition 3.6). La fonction \tilde{H} sera définie en (3.38), tandis que la définition de \tilde{l} sera donnée dans le lemme 3.28. On peut donner une interprétation de ces deux fonctions. On montrera que les chaînes immergées du PDMP admettent une loi invariante (voir les propositions 3.19 et 3.22). $\tilde{l}(A, B_k, t)$ est une approximation du taux de saut $\tilde{\lambda}(x, y, t)$ du temps inter-saut S_{n+1} sachant les positions $Z_n = x$ et $Z_{n+1} = y$ sous la loi invariante, au temps t , pour x dans A et y dans B_k . L'existence du taux de saut $\tilde{\lambda}$ sera établie dans la sous-section 3.3.1. Intuitivement, la fonction $\tilde{l}(A, B_k, t)$ peut être vue comme le taux de saut de A vers B_k au temps t et sous la loi invariante. De plus, on montrera que $\tilde{H}(A, B_k, t)$ correspond à la probabilité $\mathbb{P}_\nu(S_1 > t, Z_1 \in B_k | Z_0 \in A)$ où ν désigne la loi invariante de la chaîne de Markov (Z_n) . Dans ce cadre, un estimateur naturel de $f(\xi, t)$ est donné par

$$\hat{f}_n(A, t) = \sum_{k=1}^p \hat{l}_n(A, B_k, t) \hat{p}_n(A, B_k, t),$$

où $\widehat{l}_n(A, B_k, t)$ estime $\widetilde{l}(A, B_k, t)$, alors que $\widehat{p}_n(A, B_k, t)$ est un estimateur de $\widetilde{H}(A, B_k, t)$. Ces deux estimateurs seront donnés en (3.11) et (3.12). La procédure d'estimation que nous proposons dans ce chapitre ne nécessite pas la connaissance du flot déterministe Φ , mais seulement celle de la fonction t^* (ou d'une minoration suffisamment fine de t^*).

On donne maintenant les résultats majeurs de ce chapitre de manière plus précise. Les démonstrations se trouvent dans la section 3.4.

PROPOSITION 3.6 Soit $\xi \in A$. Pour tout $0 \leq t < t^*(A)$,

$$\left| f(\xi, t) - \sum_{k=1}^p \widetilde{l}(A, B_k, t) \widetilde{H}(A, B_k, t) \right| \leq C_1 \text{diam } A + C_2 \text{diam}_2 A \times B_1.$$

PREUVE. La sous-section 3.4.2 est consacrée à cette démonstration. \square

On s'intéresse maintenant à l'estimation de $\widetilde{l}(A, B_k, t)$. Pour cela, on considère les processus $N_n(A, B_k, \cdot)$ et $Y_n(A, B_k, \cdot)$ définis sur $[0, t^*(A)[$ par

$$N_n(A, B_k, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}}, \quad (3.8)$$

et

$$Y_n(A, B_k, t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq t\}} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}}. \quad (3.9)$$

Le processus $N_n(A, B_k, t)$ compte le nombre de temps inter-sauts inférieurs à t avec une position précédente dans A et une position future dans B_k . $Y_n(A, B_k, t)^+$ désigne l'inverse généralisé de $Y_n(A, B_k, t)$. Il est donné par

$$Y_n(A, B_k, t)^+ = \begin{cases} 0 & \text{si } Y_n(A, B_k, t) = 0, \\ \frac{1}{Y_n(A, B_k, t)} & \text{sinon.} \end{cases} \quad (3.10)$$

L'estimateur de la fonction $\widetilde{l}(A, B_k, \cdot)$ qu'on considère est obtenu par des méthodes à noyau. Soient K un noyau continu dont le support est $[-1, 1]$, $b > 0$ et $0 < t < t^*(A)$. Notre estimateur est défini, pour tout temps s entre 0 et t , par

$$\widehat{l}_{n,b,t}(A, B_k, s) = \frac{1}{b} \sum_{k=0}^{n-1} K\left(\frac{s - S_{i+1}}{b}\right) Y_n(A, B_k, S_{i+1})^+ \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}}. \quad (3.11)$$

On a le résultat suivant de convergence.

PROPOSITION 3.7 *Soient $0 < r_1 < r_2 < t < t^*(A)$ et $x \in E$. Il existe une suite $(\beta_n(A, B_k))$ (elle dépend de t) qui converge presque sûrement vers 0, telle que*

$$\sup_{r_1 \leq s \leq r_2} \left| \widehat{\mathfrak{l}}_{n, \beta_n(A, B_k), t}(A, B_k, s) - \widetilde{\mathfrak{l}}(A, B_k, s) \right| \xrightarrow{\mathbb{P}_x} 0,$$

quand n tend vers l'infini.

PREUVE. La sous-section 3.4.1 est consacrée à cette démonstration. \square

D'un autre côté, on estime la quantité $\widetilde{H}(A, B_k, t)$ par sa version empirique $\widehat{p}_n(A, B_k, t)$ donnée par

$$\widehat{p}_n(A, B_k, t) = \frac{\sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} > t\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}} \mathbb{1}_{\{Z_i \in A\}}}{\sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}}}. \quad (3.12)$$

On a le résultat suivant de consistance.

PROPOSITION 3.8 *Soit $0 < t < t^*(A)$. Pour tout $x \in E$, on a, quand n tend vers l'infini,*

$$\sup_{0 \leq s \leq t} \left| \widehat{p}_n(A, B_k, s) - \widetilde{H}(A, B_k, t) \right| \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0.$$

PREUVE. La sous-section 3.4.3 est consacrée à cette démonstration. \square

A la lumière de ce qui précède, on peut estimer $f(\xi, s)$, avec $\xi \in A$ et $0 \leq s \leq t < t^*(A)$, par

$$\widehat{f}_n(A, s) = \sum_{k=1}^p \widehat{\mathfrak{l}}_{n, \beta_n(A, B_k), t}(A, B_k, s) \widehat{p}_n(A, B_k, s).$$

La dépendance en t et $(\beta_n(A, B_k))$ est implicite dans l'expression de \widehat{f}_n par souci de lisibilité. Notre principal résultat de convergence est le suivant.

THÉORÈME 3.9 Soient \mathcal{K} un compact inclus dans E et $\xi \in E$. Pour tous $\epsilon, \eta > 0$, il existe un entier n^* et une partition finie $P = (A_l)$ de \mathcal{K} tels que, pour tout $0 < t < \min_l t^*(A_l)$, il existe pour tout couple (l, k) , une suite $(\beta_n(A_l, B_k))$ (dépendant de t), qui converge presque sûrement vers 0, telle que pour tout $n \geq n^*$, pour tous $0 < r_1 < r_2 < t$,

$$\mathbb{P}_\xi \left(\sup_{x \in \mathcal{K}} \sup_{r_1 \leq s \leq r_2} \left| \sum_l \widehat{f}_n(A_l, t) \mathbb{1}_{\{x \in A_l\}} - f(x, s) \right| > \eta \right) < \epsilon.$$

PREUVE. Soit (A_l) une partition de \mathcal{K} . On fixe l . D'après les propositions 3.7 et 3.8, il existe une famille $\{(\beta_n(A_l, B_k))\}_{1 \leq k \leq p}$, telle que

$$\sup_{r_1 \leq s \leq r_2} \left| \sum_{k=1}^p \widehat{l}_{n, \beta_n(A_l, B_k), t}(A_l, B_k, s) \widehat{p}_n(A_l, B_k, s) - \sum_{k=1}^p \widetilde{l}(A_l, B_k, s) \widetilde{H}(A_l, B_k, s) \right| \xrightarrow{\mathbb{P}_\xi} 0,$$

quand n tend vers l'infini. De plus, d'après la proposition 3.6, pour tout $x \in \mathcal{K}$, la distance

$$\sup_{r_1 \leq s \leq r_2} \left| f(x, s) - \sum_{l=1}^{|P|} \mathbb{1}_{\{x \in A_l\}} \sum_{k=1}^p \widetilde{l}(A_l, B_k, s) \widetilde{H}(A_l, B_k, s) \right|$$

est arbitrairement petite, puisqu'on peut choisir une partition (A_l) suffisamment fine. L'inégalité triangulaire entraîne le résultat. \square

3.3 Considérations préliminaires

3.3.1 Loi conditionnelle de S_{n+1} sachant Z_n et Z_{n+1}

Nous allons montrer que la loi de S_{n+1} sachant Z_n et Z_{n+1} admet un taux de saut (voir la proposition 3.15). On a alors un corollaire immédiat qui donne le compensateur du processus de comptage à un seul saut N^{n+1} ,

$$\forall t \geq 0, N^{n+1}(t) = \mathbb{1}_{\{S_{n+1} \leq t\}}, \quad (3.13)$$

dans une certaine filtration (voir le corollaire 3.17). On va d'abord montrer que $G(x, t^*(x))$ et $G\widetilde{Q}(x, t^*(x), y)$ sont strictement positifs pour $x, y \in E$.

REMARQUE 3.10 Soient $x, y \in E$. D'une part, comme M est localement intégrable, on a

$$G(x, t^*(x)) \geq \exp\left(-\int_0^{t^*(x)} M(s)ds\right) > 0, \quad (3.14)$$

D'autre part, par (3.2) et (3.3), on a

$$G\tilde{Q}(x, t^*(x), y) \geq mG(x, t^*(x)) > 0. \quad (3.15)$$

On note R le noyau de Markov de la chaîne (Z_n) des positions lors des sauts du PDMP. On peut donner une formule explicite de R et montrer un résultat de minoration.

REMARQUE 3.11 Le noyau R de la chaîne de Markov (Z_n) peut s'écrire de la façon suivante (voir [31, (34.12), page 116]), pour tous $x \in E$, $B \in \mathcal{B}(E)$,

$$R(x, B) = \int_0^{t^*(x)} f\bar{Q}(x, s, B)ds + G\bar{Q}(x, t^*(x), B), \quad (3.16)$$

où les fonctions $G\bar{Q}$ and $f\bar{Q}$ ont été définies par (3.4) et (3.5). De plus, puisque le noyau de transition \bar{Q} admet \tilde{Q} pour densité par rapport à la mesure μ , on peut aussi écrire,

$$R(x, B) = \int_B \left[\int_0^{t^*(x)} f\tilde{Q}(x, s, y)ds + G\tilde{Q}(x, t^*(x), y) \right] \mu(dy). \quad (3.17)$$

Enfin, par (3.14) et (3.15), on obtient

$$R(x, B) \geq m\mu(B) \exp\left(-\int_0^{t^*(x)} M(s)ds\right). \quad (3.18)$$

En particulier, si $\overset{\circ}{B} \neq \emptyset$, $R(x, B) > 0$ car $\mu(B) > 0$ d'après les hypothèses 3.1.

On note ν_n (respectivement $\tilde{\nu}_n, \eta_n$) la loi de Z_n (respectivement du couple (Z_n, Z_{n+1}) , du triplet (Z_n, Z_{n+1}, S_{n+1})). La remarque suivante explicite le lien entre les mesures $\tilde{\nu}_n$ et ν_n et le noyau R .

REMARQUE 3.12 Soient n un entier et $A \times B \in \mathcal{B}(E)^{\otimes 2}$. On peut écrire

$$\begin{aligned} \tilde{\nu}_n(A \times B) &= \mathbb{P}_{\nu_0}(Z_{n+1} \in B, Z_n \in A) \\ &= \int_A R(x, B)\nu_n(dx). \end{aligned} \quad (3.19)$$

De même, on s'intéresse au lien entre les mesures η_n et ν_n .

REMARQUE 3.13 Soient $t \geq 0$ et $B \in \mathcal{B}(E)$. On a

$$\begin{aligned} \mathbb{P}_{\nu_0}(S_{n+1} > t, Z_{n+1} \in B | Z_n) \\ = \mathbb{1}_{\{0 \leq t < t^*(Z_n)\}} \left[\int_{t \wedge t^*(Z_n)}^{t^*(Z_n)} f\bar{Q}(Z_n, s, B) ds + G\bar{Q}(Z_n, t^*(Z_n), B) \right]. \end{aligned}$$

Cela implique immédiatement que, pour tous $A, B \in \mathcal{B}(E)$ et $t \geq 0$,

$$\begin{aligned} \eta_n(A \times B \times]t, +\infty[) \\ = \int_A \mathbb{1}_{\{0 \leq t < t^*(x)\}} \left[\int_{t \wedge t^*(x)}^{t^*(x)} f\bar{Q}(x, s, B) ds + G\bar{Q}(x, t^*(x), B) \right] \nu_n(dx). \end{aligned} \quad (3.20)$$

Le principal résultat de cette partie est présenté dans la proposition qui suit. Il traite de l'existence d'un taux de saut pour la loi conditionnelle de S_{n+1} sachant Z_n et Z_{n+1} .

REMARQUE 3.14 On peut trouver contre-intuitif que la loi conditionnelle de S_{n+1} dépend de la position Z_{n+1} . Cependant, même si S_{n+1} est générée indépendamment de Z_{n+1} (voir (3.1)), Z_{n+1} dépend de S_{n+1} à travers (3.1). Ces variables ne sont donc pas indépendantes en général sachant Z_n . On peut comprendre cela grâce à l'exemple suivant. Si le flot Φ est injectif et si $Q(x, dy) = \delta_{\{x\}}(dy)$, alors, sachant Z_n et Z_{n+1} , S_{n+1} est l'unique solution de $\Phi(Z_n, S_{n+1}) = Z_{n+1}$.

PROPOSITION 3.15 Soit n un entier. La loi conditionnelle de S_{n+1} sachant (Z_n, Z_{n+1}) vérifie, pour tout $t \geq 0$,

$$\mathbb{P}_{\nu_0}(S_{n+1} > t | Z_n, Z_{n+1}) = \exp \left(- \int_0^{t \wedge t^*(Z_n)} \tilde{\lambda}(Z_n, Z_{n+1}, s) ds \right) \mathbb{1}_{\{0 \leq t < t^*(Z_n)\}},$$

où le taux $\tilde{\lambda}$ est défini, pour tous $x, y \in E$, $0 \leq t \leq t^*(x)$, par

$$\tilde{\lambda}(x, y, t) = \frac{f\tilde{Q}(x, t, y)}{\int_t^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)}. \quad (3.21)$$

PREUVE. Soient $x, y \in E$. $\tilde{\lambda}(x, y, \cdot)$ est une fonction continue sur l'intervalle $[0, t^*(x)]$ car $f(x, \cdot)$ et $\tilde{Q}(x, \cdot, y)$ sont des fonctions continues d'après les hypothèses 3.3. La fonction de survie \tilde{G} associée à $\tilde{\lambda}$ est définie pour tous $x, y \in E$, $0 \leq t \leq t^*(x)$, par

$$\tilde{G}(x, y, t) = \exp\left(-\int_0^t \tilde{\lambda}(x, y, s) ds\right). \quad (3.22)$$

De plus, pour tout $0 \leq t < t^*(x)$, $\tilde{\lambda}(x, y, t) = -u'(t)/u(t)$ avec

$$u(t) = \int_t^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y).$$

Par conséquent, on a

$$\begin{aligned} \int_0^t \tilde{\lambda}(x, y, s) ds &= -\ln\left(\int_t^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)\right) \\ &\quad + \ln\left(\int_0^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)\right) \\ &= -\ln\left(\frac{\int_t^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)}{\int_0^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)}\right). \end{aligned}$$

Finalement, par (3.22),

$$\tilde{G}(x, y, t) = \frac{\int_t^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)}{\int_0^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y)}. \quad (3.23)$$

Afin de prouver le résultat recherché, il est nécessaire de montrer

$$\eta_n(A \times B \times]t, +\infty]) = \int_{A \times B} \tilde{G}(x, y, t \wedge t^*(x)) \mathbb{1}_{\{0 \leq t < t^*(x)\}} \tilde{\nu}_n(dx \times dy), \quad (3.24)$$

pour tous $A, B \in \mathcal{B}(E)$ et $t \geq 0$. Par (3.19), on a

$$\begin{aligned} \int_{A \times B} \tilde{G}(x, y, t \wedge t^*(x)) \mathbb{1}_{\{0 \leq t < t^*(x)\}} \tilde{\nu}_n(dx \times dy) \\ = \int_{A \times B} \tilde{G}(x, y, t \wedge t^*(x)) \mathbb{1}_{\{0 \leq t < t^*(x)\}} R(x, dy) \nu_n(dx). \end{aligned}$$

Donc, avec (3.17) et (3.23), on obtient

$$\begin{aligned}
& \int_{A \times B} \tilde{G}(x, y, t \wedge t^*(x)) \mathbb{1}_{\{0 \leq t < t^*(x)\}} \tilde{\nu}_n(dx \times dy) \\
&= \int_{A \times B} \mathbb{1}_{\{0 \leq t < t^*(x)\}} \left[\int_{t \wedge t^*(x)}^{t^*(x)} f \tilde{Q}(x, s, y) ds + G \tilde{Q}(x, t^*(x), y) \right] \mu(dy) \nu_n(dx) \\
&= \int_A \mathbb{1}_{\{0 \leq t < t^*(x)\}} \left[\int_{t \wedge t^*(x)}^{t^*(x)} f \bar{Q}(x, s, B) ds + G \bar{Q}(x, t^*(x), B) \right] \nu_n(dx).
\end{aligned}$$

Avec l'expression (3.20) de η_n , cela implique directement (3.24). \square

REMARQUE 3.16 Soit n un entier. D'après le théorème de Jirina (voir par exemple le théorème 11.7 de [68]), il existe une famille de noyaux $(\gamma_{x,y}(\cdot))$ telle que, pour tous $A \times B \times \Gamma$ dans $\mathcal{B}(E)^{\otimes 2} \otimes \mathcal{B}(\mathbb{R}_+)$, on a

$$\eta_n(A \times B \times \Gamma) = \int_{A \times B} \gamma_{x,y}(\Gamma) \tilde{\nu}_n(dx \times dy).$$

Soient $x, y \in E^2$. D'après la proposition 3.15, $\gamma_{x,y}$ ne dépend pas de n . De plus, pour $t < t^*(x)$, on a la relation

$$\gamma_{x,y}([t, +\infty[) = \mathbb{P}_{\nu_0}(S_1 > t | Z_0 = x, Z_1 = y) = \tilde{G}(x, y, t).$$

On a aussi la propriété de martingale à temps continu suivante.

COROLLAIRE 3.17 Soit i un entier. Le processus à temps continu M^{i+1} donné par,

$$\forall 0 \leq t < t^*(Z_i), M^{i+1}(t) = N^{i+1}(t) - \int_0^t \tilde{\lambda}(Z_i, Z_{i+1}, u) \mathbb{1}_{\{S_{i+1} \geq u\}} du, \quad (3.25)$$

est une $(\sigma(Z_i, Z_{i+1}) \vee \mathcal{F}_t^{i+1})$ -martingale. On rappelle que N^{i+1} a déjà été défini par (3.13).

PREUVE. La démonstration est analogue à celle du lemme 2.2. En effet, le taux de S_{i+1} sachant Z_i, Z_{i+1} est $\tilde{\lambda}(Z_i, Z_{i+1}, \cdot)$ d'après la proposition 3.15. \square

3.3.2 Ergodicité

Dans cette partie, on s'intéresse au comportement asymptotique des chaînes de Markov (Z_n) , (Z_n, S_{n+1}) and (Z_n, Z_{n+1}, S_{n+1}) ; l'objectif principal est de montrer qu'on peut appliquer le théorème ergodique presque sûr à ces chaînes de Markov.

REMARQUE 3.18 Soient $x, y \in E$ et $B \in \mathcal{B}(E)$. Par (3.14) et (3.15), on a

$$G\tilde{Q}(x, t^*(x), y) \geq m_2, \quad (3.26)$$

où m_2 est donné par

$$m_2 = m \exp \left(- \int_0^{\|t^*\|_\infty} M(s) ds \right) > 0.$$

En particulier, m_2 ne dépend pas de x et y . Avec (3.18), on a

$$R(x, B) \geq m_2 \mu(B). \quad (3.27)$$

On peut alors montrer que la chaîne (Z_n) est ergodique.

PROPOSITION 3.19 On a les résultats suivants.

- a) (Z_n) est μ -irréductible, apériodique et admet une mesure invariante unique, qu'on note ν .
- b) Il existe $\rho > 1$ et $r > 0$ tels que,

$$\forall n \geq 0, \sup_{\xi \in E} \|R^n(\xi, \cdot) - \nu\|_{TV} \leq r\rho^{-n}, \quad (3.28)$$

où $\|\cdot\|_{TV}$ désigne la norme en variation totale (voir [53] pour une définition).

- c) La chaîne de Markov (Z_n) est positive Harris-récurrente.

PREUVE. Par définition et la remarque 3.18, (Z_n) est μ -irréductible et apériodique. De plus, le noyau R satisfait de manière évidente la condition de Doeblin (voir [67, page 396] par exemple),

$$\mu(B) > \epsilon \Rightarrow R(\xi, B) > m_2 \epsilon,$$

par (3.27). D'après le théorème 16.0.2 de [67], (Z_n) admet une loi invariante unique ν . De plus, par le théorème 4.3.3 de [53], (Z_n) est positive Harris-récurrente. \square

Nous allons voir maintenant que les ensembles d'intérieur non vide sont chargés par la mesure ν .

REMARQUE 3.20 *Le noyau R admet une densité par rapport à la mesure μ . Par conséquent, ν et la mesure auxiliaire μ sont équivalentes d'après le théorème 10.4.9 of [67]. Donc, pour tout ensemble A d'intérieur non vide, $\nu(A) > 0$ d'après les hypothèses 3.1.*

On s'intéresse au comportement asymptotique des suites $(\tilde{\nu}_n)$ et (η_n) .

LEMME 3.21 *Pour toute loi initiale $\nu_0 = \delta_{\{x\}}$, $x \in E$,*

$$\lim_{n \rightarrow +\infty} \|\tilde{\nu}_n - \tilde{\nu}\|_{TV} = 0 \quad \text{et} \quad \lim_{n \rightarrow +\infty} \|\eta_n - \eta\|_{TV} = 0,$$

où les lois limites $\tilde{\nu}$ et η sont données par,

$$\forall A \times B \in \mathcal{B}(E)^{\otimes 2}, \tilde{\nu}(A \times B) = \int_A \nu(dx) R(x, B), \quad (3.29)$$

$$\forall A \times B \times \Gamma \in \mathcal{B}(E)^{\otimes 2} \otimes \mathcal{B}(\mathbb{R}_+), \eta(A \times B \times \Gamma) = \int_{A \times B \times \Gamma} \gamma_{x,y}(ds) \tilde{\nu}(dx \times dy). \quad (3.30)$$

PREUVE. Soit g une fonction mesurable bornée par 1. D'après le théorème de Fubini, on a

$$\left| \int_{E \times E} g(x, y) (\tilde{\nu}_n(dx \times dy) - \tilde{\nu}(dx \times dy)) \right| = \left| \int_E (\nu_n(dx) - \nu(dx)) \int_E g(x, y) R(x, dy) \right|,$$

d'après l'expression de $\tilde{\nu}_n$ (3.19) et la définition de $\tilde{\nu}$ (3.29). Donc,

$$\left| \int_{E \times E} g(x, y) (\tilde{\nu}_n(dx \times dy) - \tilde{\nu}(dx \times dy)) \right| = \left| \int_E h(x) (\nu_n(dx) - \nu(dx)) \right|,$$

où $h : x \mapsto \int_E g(x, y) R(x, dy)$ est bornée par 1 car R est un noyau de Markov. Finalement,

$$\|\tilde{\nu}_n - \tilde{\nu}\|_{TV} \leq \|\nu_n - \nu\|_{TV}.$$

On obtient la limite voulue avec (3.28). Maintenant, on établit la seconde limite. Soit g' une fonction mesurable bornée par 1. D'après la remarque 3.16,

$$\left| \int_{E \times E \times \mathbb{R}_+} g'(x, y, s) (\eta_n(dx \times dy \times ds) - \eta(dx \times dy \times ds)) \right| \leq \|h\|_{\infty} \|\tilde{\nu}_n - \tilde{\nu}\|_{TV},$$

d'après le théorème de Fubini et avec la fonction h définie par,

$$\forall x, y \in E, h(x, y) = \int_{\mathbb{R}_+} g'(x, y, s) \gamma_{x,y}(ds).$$

Comme h est bornée par 1, on a

$$\left| \int_{E \times E \times \mathbb{R}_+} g'(x, y, s) (\eta_n(dx \times dy \times ds) - \eta(dx \times dy \times ds)) \right| \leq \|\tilde{\nu}_n - \tilde{\nu}\|_{TV}.$$

On a déjà montré que $\|\tilde{\nu}_n - \tilde{\nu}\|_{TV}$ tend vers 0, donc $\|\eta_n - \eta\|_{TV}$ tend vers 0 également. \square

Ce lemme a pour conséquence le résultat suivant.

PROPOSITION 3.22 *On a les résultats suivants.*

- a) (Z_n, Z_{n+1}) (respectivement (Z_n, Z_{n+1}, S_{n+1})) est $\tilde{\nu}$ -irréductible (respectivement η -irréductible).
- b) (Z_n, Z_{n+1}) et (Z_n, Z_{n+1}, S_{n+1}) sont des chaînes de Markov positives Harris-récurrentes et apériodiques.
- c) $\tilde{\nu}$ (respectivement η) est l'unique mesure invariante de (Z_n, Z_{n+1}) (respectivement de (Z_n, Z_{n+1}, S_{n+1})).

PREUVE. Ce résultat est un corollaire du lemme 3.21. Les arguments utilisés sont similaires à ceux donnés dans la preuve de la proposition 3.19. \square

D'après les résultats obtenus, (Z_n) , (Z_n, Z_{n+1}) and (Z_n, Z_{n+1}, S_{n+1}) sont des chaînes de Markov positives Harris-récurrentes. On peut donc leur appliquer le théorème ergodique presque sûr (voir le théorème 17.1.7 de [67]).

3.4 Démonstrations des résultats principaux

Cette section est dédiée à la présentation des démonstrations des résultats majeurs donnés dans la section 3.2. Dans toute cette partie, on considère un ensemble $A \in \mathcal{C}_2$ et $\epsilon > 0$. En particulier, cela entraîne l'existence d'une famille $\{B_1, \dots, B_p\}$, $p \geq 2$, dépendant de ϵ et satisfaisant l'hypothèse 3.1 b).

3.4.1 Estimation de $\tilde{\lambda}$

On se focalise ici sur l'estimation de la fonction $\tilde{\lambda}$, qui est une approximation du taux $\tilde{\lambda}$. Sa définition est présentée dans le lemme 3.28. Cette sous-section s'appuie sur l'étude réalisée dans le chapitre précédent ; par conséquent, beaucoup de preuves font référence à des démonstrations déjà données. L'estimation de $\tilde{\lambda}$ est une étape clé dans notre procédure d'estimation de f . Rappelons qu'elle est présentée dans la proposition 3.7.

Montrons d'abord des résultats techniques sur \tilde{G} et $\tilde{\lambda}$.

LEMME 3.23 *Soit $1 \leq k \leq p$. Pour tout $0 \leq t < t^*(A)$,*

$$\inf_{x \in A, y \in B_k} \tilde{G}(x, y, t) > 0. \quad (3.31)$$

PREUVE. De (3.23) et (3.26), on a pour tous $x, y \in E$, $0 \leq t < t^*(x)$,

$$\begin{aligned} \tilde{G}(x, y, t) &\geq \frac{G\tilde{Q}(x, t^*(x), y)}{\int_0^{t^*(x)} f\tilde{Q}(x, s, y)ds + G\tilde{Q}(x, t^*(x), y)} \\ &\geq \frac{m_2}{(\|t^*\|_\infty \|f\|_\infty + 1) \|\tilde{Q}\|_\infty}, \end{aligned}$$

car f , \tilde{Q} et t^* sont des fonctions bornées (d'après les hypothèses 3.3). \square

On peut montrer que le taux $\tilde{\lambda}$ est borné.

LEMME 3.24 *Soient $x, y \in E$ et $0 \leq t \leq t^*(x)$. Alors,*

$$|\tilde{\lambda}(x, y, t)| \leq \frac{\|f\|_\infty \|\tilde{Q}\|_\infty}{m_2}.$$

PREUVE. Par (3.21), on a

$$\begin{aligned} \tilde{\lambda}(x, y, t) &= \frac{f\tilde{Q}(x, t, y)}{\int_t^{t^*(x)} f\tilde{Q}(x, s, y)ds + G\tilde{Q}(x, t^*(x), y)} \\ &\leq \frac{f\tilde{Q}(x, t, y)}{G\tilde{Q}(x, t^*(x), y)}. \end{aligned}$$

Donc, avec (3.26) et puisque f et \tilde{Q} sont bornées, on obtient immédiatement le résultat recherché. \square

REMARQUE 3.25 Pour tout $1 \leq k \leq p$, $\tilde{\nu}(A \times B_k) > 0$. En effet,

$$\tilde{\nu}(A \times B_k) = \int_A R(x, B_k) \nu(dx).$$

Pour tout x , $R(x, B_k) \geq m_2 \mu(B_k)$ par (3.27), donc,

$$\tilde{\nu}(A \times B_k) \geq m_2 \mu(B_k) \nu(A).$$

Or $\nu(A) > 0$ d'après la remarque 3.20 et $\mu(B_k) > 0$ car B_k est un ensemble d'intérieur non vide (voir les hypothèses 3.1), d'où le résultat.

Le résultat suivant traite du comportement asymptotique de Y_n . On rappelle que la définition de Y_n est donnée dans (3.9). Son inverse généralisé Y_n^+ a été défini par (3.10).

LEMME 3.26 Soit $1 \leq k \leq p$. Pour tout $x \in E$,

$$\forall 0 \leq t < t^*(A), \quad \frac{Y_n(A, B_k, t)}{n} \xrightarrow{\mathbb{P}_{x\text{-p.s.}}} \int_{A \times B_k} \tilde{G}(x, y, t) \tilde{\nu}(dx \times dy),$$

quand n tend vers l'infini. De plus, cette limite est strictement positive.

PREUVE. Par le théorème ergodique appliqué à (Z_n, Z_{n+1}, S_{n+1}) ,

$$\frac{1}{n} Y_n(A, B_k, t) \xrightarrow{\mathbb{P}_{x\text{-p.s.}}} \eta(A \times B_k \times [t, +\infty[).$$

Avec (3.30) et la remarque 3.16, on a pour tout $0 \leq t < t^*(A)$,

$$\frac{Y_n(A, B_k, t)}{n} \xrightarrow{\mathbb{P}_{x\text{-p.s.}}} \int_{A \times B_k} \tilde{G}(x, y, t) \tilde{\nu}(dx \times dy),$$

quand n tend vers l'infini. Puisque $\inf_{x \in A, y \in B_k} \tilde{G}(x, y, t) > 0$ par (3.31) et $\tilde{\nu}(A \times B_k) > 0$ d'après la remarque 3.25, la limite est strictement positive. \square

LEMME 3.27 Soient $1 \leq k \leq p$, $0 \leq t < t^*(A)$ et $x \in E$. Alors, pour tout entier n ,

$$Y_n(A, B_k, t)^+ \leq 1 \text{ } \mathbb{P}_x\text{-p.s.}$$

et, quand n tend vers l'infini,

$$\begin{aligned} Y_n(A, B_k, t)^+ &\xrightarrow{\mathbb{P}_x\text{-p.s.}} 0, \\ \mathbb{1}_{\{Y_n(A, B_k, t)=0\}} &\xrightarrow{\mathbb{P}_x\text{-p.s.}} 0, \\ \int_0^t \mathbb{1}_{\{Y_n(A, B_k, s)=0\}} ds &\xrightarrow{\mathbb{P}_x\text{-p.s.}} 0. \end{aligned}$$

PREUVE. Ce résultat est un corollaire du lemme 3.26. La démonstration est similaire à celle du lemme 2.14. \square

La fonction \tilde{l} est définie dans la proposition suivante.

PROPOSITION 3.28 Soient $1 \leq k \leq p$, $0 \leq t < t^*(A)$ et $x \in E$. Lorsque n tend vers l'infini,

$$Y_n(A, B_k, t)^+ \sum_{i=0}^{n-1} \tilde{\lambda}(Z_i, Z_{i+1}, t) \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}} \mathbb{1}_{\{S_{i+1} \geq t\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \tilde{l}(A, B_k, t),$$

où

$$\tilde{l}(A, B_k, t) = \frac{\int_{A \times B_k} \tilde{\lambda}(u, v, t) \tilde{G}(u, v, t) \tilde{\nu}(du \times dv)}{\int_{A \times B_k} \tilde{G}(u, v, t) \tilde{\nu}(du \times dv)}. \quad (3.32)$$

De plus, la fonction $\tilde{l}(A, B_k, \cdot)$ est continue sur $[0, t^*(A)[$.

PREUVE. Il s'agit d'une application du théorème ergodique à (Z_n, Z_{n+1}, S_{n+1}) comme dans la preuve de la proposition 2.15. $\tilde{l}(A, B_k, \cdot)$ est continue car $\tilde{\lambda}$ et \tilde{G} sont continues et bornées. On peut donc appliquer le théorème de continuité sous le signe intégrale. \square

On peut maintenant donner des résultats de martingales à temps continu. On donne d'abord une propriété importante d'indépendance conditionnelle. La tribu $\sigma(Z_0, \dots, Z_n)$ est notée \mathcal{G}_n , pour tout entier n .

PROPOSITION 3.29 Soient n un entier et $1 \leq i \leq n$. Pour tout $j \neq i$, soit $t_j \geq 0$ et soit t un réel positif. On a

$$\bigvee_{j \neq i} \mathcal{F}_{t_j}^j \perp_{\mathcal{G}_n} \mathcal{F}_t^i \quad \text{et} \quad \mathcal{F}_t^i \perp_{\sigma(Z_{i-1}, Z_i)} \mathcal{G}_n.$$

On peut alors déduire de la proposition 6.8 de [58] que, pour tous $s < t$, $0 \leq i \leq n-1$,

$$\bigvee_{j \neq i+1} \mathcal{F}_s^j \perp_{\mathcal{G}_n \vee \mathcal{F}_s^{i+1}} \mathcal{F}_t^{i+1} \quad \text{et} \quad \mathcal{F}_t^{i+1} \perp_{\sigma(Z_i, Z_{i+1}) \vee \mathcal{F}_s^{i+1}} \mathcal{G}_n.$$

PREUVE. La preuve technique de ce résultat se trouve dans la section 3.7. □

THÉORÈME 3.30 Soit $1 \leq k \leq p$. Le processus $M_n(A, B_k, \cdot)$, défini pour tout $0 \leq t < t^*(A)$, par

$$M_n(A, B_k, t) = \sum_{i=0}^{n-1} M^{i+1}(t) \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}},$$

est une $(\mathcal{G}_n \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1})$ -martingale à temps continu.

PREUVE. Ce résultat est un corollaire de la proposition 3.29 et du corollaire 3.17. Les arguments sont similaires à ceux du théorème 2.16. □

On peut en déduire un nouveau résultat de martingale à temps continu.

LEMME 3.31 Soient $1 \leq k \leq p$ et n un entier. Le processus $\widetilde{M}_n(A, B_k, \cdot)$, donné pour tout $0 \leq t < t^*(A)$, par

$$\widetilde{M}_n(A, B_k, t) = \int_0^t Y_n(A, B_k, s)^+ dM_n(A, B_k, s),$$

est une martingale à temps continu, dont le crochet $\langle \widetilde{M}_n(A, B_k) \rangle$ vérifie, pour tout $x \in E$,

$$\forall 0 \leq t < t^*(A), \langle \widetilde{M}_n(A, B_k) \rangle(t) \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0,$$

quand n tend vers l'infini.

PREUVE. Il s'agit d'un corollaire des lemmes 3.24 et 3.27 et du théorème 3.30. La démonstration est similaire à celle du lemme 2.17. \square

REMARQUE 3.32 D'après le lemme 3.31, on a, pour tous $0 \leq t < t^*(A)$ et $x \in E$,

$$\sup_{0 \leq s \leq t} |\widetilde{M}_n(A, B_k, s)| \xrightarrow{\mathbb{P}_x} 0,$$

d'après l'inégalité de Lenglart (voir [5], II.5.2.1. Lenglart's inequality).

Dans la suite, on s'intéresse à l'estimation de la quantité $\widetilde{l}(A, B_k, \cdot)$, donnée pour tout $1 \leq k \leq p$ par (3.32). On estime d'abord la fonction $\widetilde{L}(A, B_k, \cdot)$ définie par,

$$\forall 0 \leq t < t^*(A), \widetilde{L}(A, B_k, t) = \int_0^t \widetilde{l}(A, B, s) ds.$$

On considère l'estimateur de type Nelson-Aalen $\widehat{L}_n(A, B_k, \cdot)$ donné par,

$$\forall 0 \leq t < t^*(A), \widehat{L}_n(A, B_k, t) = \int_0^t Y_n(A, B_k, s)^+ dN_n(A, B_k, s),$$

où $N_n(A, B_k, \cdot)$ est le processus de comptage donné par (3.8). D'après tout ce qui précède, le cadre de travail est le même que dans le chapitre précédent.

PROPOSITION 3.33 Soient $1 \leq k \leq p$, $0 < t < t^*(A)$ et $x \in E$. Alors, quand n tend vers l'infini,

$$\sup_{0 \leq s \leq t} \left| \widehat{L}_n(A, B_k, s) - \widetilde{L}(A, B_k, s) \right| \xrightarrow{\mathbb{P}_x} 0.$$

PREUVE. Ce résultat est un corollaire du lemme 3.31 et de la proposition 3.28, et se démontre en utilisant les arguments mis en avant dans les preuves de la proposition 2.19 et du théorème 2.20. \square

On s'intéresse au lissage de cet estimateur par des méthodes à noyau afin d'obtenir une estimation de $\widetilde{L}(A, B_k, \cdot)$. On rappelle qu'on considère un noyau continu K dont le support est $[-1, 1]$, $b > 0$ et $0 < t < t^*(A)$. On peut définir notre estimateur de $\widetilde{L}(A, B_k, \cdot)$ par,

$$\forall 0 \leq s \leq t, \widehat{l}_{n,b,t}(A, B_k, s) = \frac{1}{b} \int_0^t K\left(\frac{s-u}{b}\right) d\widehat{L}_n(A, B_k, u).$$

Un calcul direct montre l'équivalence avec (3.11). On peut maintenant prouver la proposition 3.7.

PREUVE DE LA PROPOSITION 3.7. L'ensemble des résultats précédents permettent de montrer qu'on est dans le même cadre que dans le chapitre 2 sur l'estimation du taux de saut de processus de renouvellement. La démonstration de la proposition 3.7 est alors identique à celle de la proposition 2.28.

Les différences majeures entre l'étude menée dans le chapitre 2 et le cadre de cette section sont les suivantes : l'indépendance conditionnelle (la proposition 3.29) nécessaire pour avoir le théorème 3.30 se montre différemment, et le taux de saut considéré ici a une double dépendance spatiale.

3.4.2 Approximation de f

Cette sous-section est consacrée à la démonstration de la proposition 3.6. Par souci de lisibilité, on note pour tous $x, y \in E$ et $0 \leq t \leq t^*(x)$,

$$H(x, y, t) = \int_t^{t^*(x)} f\widetilde{Q}(x, s, y) ds + G\widetilde{Q}(x, t^*(x), y). \quad (3.33)$$

Donc, par (3.21), on a

$$H(x, y, t)\tilde{\lambda}(x, y, t) = f\tilde{Q}(x, t, y).$$

Nous allons établir quelques propriétés de H .

LEMME 3.34 *On a les résultats suivants.*

a) $\|H\|_\infty < +\infty$.

b) Pour tous $x, y \in E$, $0 \leq t \leq t^*(x)$, $H(x, y, t) \geq m_2$.

c) Il existe une constante $[H]_{\text{Lip}} > 0$ telle que, pour tous $x, y, u, v \in E$ et $0 \leq t \leq t^*(x) \wedge t^*(u)$,

$$|H(x, y, t) - H(u, v, t)| \leq [H]_{\text{Lip}} d_2((x, y), (u, v)). \quad (3.34)$$

PREUVE. Tout d'abord, par (3.33), on a

$$H(x, y, t) \leq \int_0^{t^*(x)} f\tilde{Q}(x, s, y) ds + G\tilde{Q}(x, t^*(x), y).$$

G est bornée par 1. De plus, f , t^* et \tilde{Q} sont des fonctions bornées d'après les hypothèses 3.1 et 3.3. Donc,

$$H(x, y, t) \leq \|\tilde{Q}\|_\infty (\|t^*\|_\infty \|f\|_\infty + 1).$$

On montre maintenant que m_2 est un minorant de H . En effet, par (3.33) à nouveau, pour tous $x, y \in E$ et $0 \leq t \leq t^*(x)$,

$$H(x, y, t) \geq G\tilde{Q}(x, t^*(x), y) \geq m_2. \quad (3.35)$$

Finalement, on établit que H est Lipschitz. Par (3.33) et l'inégalité triangulaire,

$$\begin{aligned} & |H(x, y, t) - H(u, v, t)| \\ & \leq \int_t^{t^*(x) \vee t^*(u)} |f\tilde{Q}(u, s, v) - f\tilde{Q}(x, s, y)| ds + |G\tilde{Q}(u, t^*(u), v) - G\tilde{Q}(x, t^*(x), y)|. \end{aligned}$$

Donc, puisque le produit $f\tilde{Q}$ est Lipschitz d'après (3.6), on a

$$\begin{aligned} & |H(x, y, t) - H(u, v, t)| \\ & \leq \|t^*\|_\infty [f\tilde{Q}]_{\text{Lip}} d_2((x, y), (u, v)) + |G\tilde{Q}(u, t^*(u), v) - G\tilde{Q}(x, t^*(x), y)|. \end{aligned}$$

Avec (3.7), on obtient (3.34) où

$$[H]_{\text{Lip}} = \|t^*\|_{\infty} [f\tilde{Q}]_{\text{Lip}} + [G\tilde{Q}]_{\text{Lip}},$$

ce qui montre les trois points. \square

On montre aussi que $\tilde{\lambda}$ est Lipschitz.

LEMME 3.35 *Il existe une constante $[\tilde{\lambda}]_{\text{Lip}} > 0$ telle que, pour tous $x, y, u, v \in E$ et $0 \leq t < t^*(x) \wedge t^*(u)$,*

$$|\tilde{\lambda}(x, y, t) - \tilde{\lambda}(u, v, t)| \leq [\tilde{\lambda}]_{\text{Lip}} d_2((x, y), (u, v)).$$

PREUVE. De (3.21) et (3.33), on a

$$|\tilde{\lambda}(x, y, t) - \tilde{\lambda}(u, v, t)| = \frac{|f\tilde{Q}(x, t, y)H(u, v, t) - f\tilde{Q}(u, t, v)H(x, y, t)|}{H(u, v, t)H(x, y, t)}.$$

Donc, avec (3.35), on obtient

$$|\tilde{\lambda}(x, y, t) - \tilde{\lambda}(u, v, t)| \leq \frac{1}{m_2^2} |f\tilde{Q}(x, t, y)H(u, v, t) - f\tilde{Q}(u, t, v)H(x, y, t)|.$$

Par l'inégalité triangulaire, cela entraîne

$$\begin{aligned} & |\tilde{\lambda}(x, y, t) - \tilde{\lambda}(u, v, t)| \\ & \leq \frac{1}{m_2^2} \left(|f\tilde{Q}(x, t, y)H(u, v, t) - f\tilde{Q}(x, t, y)H(x, y, t)| \right. \\ & \quad \left. + |f\tilde{Q}(x, t, y)H(x, y, t) - f\tilde{Q}(u, t, v)H(x, y, t)| \right) \\ & \leq \frac{1}{m_2^2} \left(\|f\|_{\infty} \|\tilde{Q}\|_{\infty} |H(u, v, t) - H(x, y, t)| + \|H\|_{\infty} |f\tilde{Q}(x, t, y) - f\tilde{Q}(u, t, v)| \right). \end{aligned}$$

$f\tilde{Q}$ est Lipschitz d'après les hypothèses 3.3 et H est Lipschitz par le lemme 3.34. Ainsi,

$$|\tilde{\lambda}(x, y, t) - \tilde{\lambda}(u, v, t)| \leq [\tilde{\lambda}]_{\text{Lip}} d_2((x, y), (u, v)),$$

où $[\tilde{\lambda}]_{\text{Lip}}$ est donné par

$$[\tilde{\lambda}]_{\text{Lip}} = \frac{\|f\|_{\infty} \|\tilde{Q}\|_{\infty} [H]_{\text{Lip}} + \|H\|_{\infty} [f\tilde{Q}]_{\text{Lip}}}{m_2^2}.$$

Ceci termine la preuve. \square

REMARQUE 3.36 Sous l'hypothèse 3.1 b), on peut donner une nouvelle propriété de la mesure μ . Par (3.2), on a pour tout $1 \leq k \leq p$,

$$\bar{Q}(x, t, B_k) = \int_{B_k} \tilde{Q}(x, t, y) \mu(dy) \geq m \mu(B_k).$$

En sommant sur k , on a

$$\sum_{k=1}^p \mu(B_k) \leq \frac{1}{m}. \quad (3.36)$$

Dans le lemme suivant, on montre que \tilde{l} est une bonne approximation de $\tilde{\lambda}$.

LEMME 3.37 Soient $1 \leq k \leq p$, $x \in A$ et $y \in B_k$. Soit $0 \leq t < t^*(A)$. Alors,

$$|\tilde{\lambda}(x, y, t) - \tilde{l}(A, B_k, t)| \leq [\tilde{\lambda}]_{\text{Lip}} \text{diam}_2 A \times B_1. \quad (3.37)$$

PREUVE. Par (3.32), on a

$$\begin{aligned} |\tilde{\lambda}(x, y, t) - \tilde{l}(A, B_k, t)| &\leq \left| \tilde{\lambda}(x, y, t) - \frac{\int_{A \times B_k} \tilde{\lambda}(u, v, t) \tilde{G}(u, v, t) \tilde{\nu}(du \times dv)}{\int_{A \times B_k} \tilde{G}(u, v, t) \tilde{\nu}(du \times dv)} \right| \\ &\leq \frac{\int_{A \times B_k} |\tilde{\lambda}(x, y, t) - \tilde{\lambda}(u, v, t)| \tilde{G}(u, v, t) \tilde{\nu}(du \times dv)}{\int_{A \times B_k} \tilde{G}(u, v, t) \tilde{\nu}(du \times dv)} \\ &\leq [\tilde{\lambda}]_{\text{Lip}} \text{diam}_2 A \times B_1, \end{aligned}$$

d'après le lemme 3.35 et l'hypothèse 3.1 b). □

Maintenant, on peut montrer la proposition 3.6. Nous définissons d'abord la fonction $\tilde{H}(A, B_k, \cdot)$ apparaissant dans ce résultat. Elle est donnée, pour tout $0 \leq t < t^*(A)$, par

$$\tilde{H}(A, B_k, t) = \frac{1}{\nu(A)} \int_{A \times B_k} H(x, y, t) \mu(dy) \nu(dx), \quad (3.38)$$

où H a été définie par (3.33).

PREUVE DE LA PROPOSITION 3.6. On montre d'abord que

$$\begin{aligned} \frac{1}{\nu(A)} \left| \int_A f(x, t) \nu(dx) - \sum_{k=1}^p \tilde{l}(A, B_k, t) \int_{A \times B_k} H(x, y, t) \mu(dy) \nu(dx) \right| \\ \leq \frac{1}{m} [\tilde{\lambda}]_{\text{Lip}} \|H\|_{\infty} \text{diam}_2 A \times B_1. \end{aligned} \quad (3.39)$$

Soient $1 \leq k \leq p$, $x \in A$ et $y \in B_k$. En multipliant (3.37) par $H(x, y, t)$, on a

$$\left| f\tilde{Q}(x, t, y) - \tilde{l}(A, B_k, t) H(x, y, t) \right| \leq [\tilde{\lambda}]_{\text{Lip}} \|H\|_{\infty} \text{diam}_2 A \times B_1,$$

car H est bornée d'après le lemme 3.34. On intègre sur B_k contre la mesure $\mu(dy)$. On obtient

$$\left| f\tilde{Q}(x, t, B_k) - \tilde{l}(A, B_k, t) \int_{B_k} H(x, y, t) \mu(dy) \right| \leq \mu(B_k) [\tilde{\lambda}]_{\text{Lip}} \|H\|_{\infty} \text{diam}_2 A \times B_1.$$

En sommant sur k entre 1 et p , on a

$$\left| f(x, t) - \sum_{k=1}^p \tilde{l}(A, B_k, t) \int_{B_k} H(x, y, t) \mu(dy) \right| \leq \frac{1}{m} [\tilde{\lambda}]_{\text{Lip}} \|H\|_{\infty} \text{diam}_2 A \times B_1,$$

avec (3.36). Finalement, on intègre sur A contre $\nu(dx)$.

$$\begin{aligned} \left| \int_A f(x, t) \nu(dx) - \sum_{k=1}^p \tilde{l}(A, B_k, t) \int_{A \times B_k} H(x, y, t) \nu(dx) \mu(dy) \right| \\ \leq \frac{\nu(A)}{m} [\tilde{\lambda}]_{\text{Lip}} \|H\|_{\infty} \text{diam}_2 A \times B_1. \end{aligned} \quad (3.40)$$

En divisant chaque terme de (3.40) par $\nu(A)$, on montre (3.39). De l'autre côté, on a

$$\left| f(\xi, t) - \frac{\int_A f(x, t) \nu(dx)}{\nu(A)} \right| \leq [f]_{\text{Lip}} \text{diam} A, \quad (3.41)$$

puisque f est Lipschitz par les hypothèses 3.3. Finalement, (3.39) et (3.41) donnent le résultat par l'inégalité triangulaire.

3.4.3 Estimation de \tilde{H}

Soient $\xi \in A$ et $0 \leq t < t^*(A)$. Pour estimer $f(\xi, t)$, on estime $\tilde{l}(A, B_k, t)$ et $\tilde{H}(A, B_k, t)$ pour tout $1 \leq k \leq p$, d'après la proposition 3.6. Nous allons voir que la quantité $\tilde{H}(A, B_k, t)$ peut être vue comme une probabilité conditionnelle.

PROPOSITION 3.38 Soit $0 \leq t < t^*(A)$. Alors,

$$\int_{A \times B_k} H(x, y, t) \nu(dx) \mu(dy) = \mathbb{P}_\nu(S_1 > t, Z_1 \in B_k, Z_0 \in A). \quad (3.42)$$

Cela entraine immédiatement que

$$\tilde{H}(A, B_k, t) = \mathbb{P}_\nu(S_1 > t, Z_1 \in B_k | Z_0 \in A).$$

On peut rappeler que $\nu(A)$ est strictement positif d'après la remarque 3.20.

PREUVE. Par (3.30) et la remarque 3.16, on a pour tout $t < t^*(A)$,

$$\begin{aligned} \mathbb{P}_\nu(S_1 > t, Z_1 \in B_k, Z_0 \in A) &= \eta(A \times B_k \times [t, +\infty[) \\ &= \int_{A \times B_k} \tilde{G}(x, y, t) \tilde{\nu}(dx \times dy). \end{aligned} \quad (3.43)$$

Ainsi,

$$\mathbb{P}_\nu(S_1 > t, Z_1 \in B_k, Z_0 \in A) = \int_{A \times B_k} \tilde{G}(x, y, t) \nu(dx) R(x, dy),$$

par (3.29). Donc, avec l'expression de R (3.16), la définition de \tilde{G} (3.23) et la définition de H (3.33), on a

$$\tilde{G}(x, y, t) R(x, dy) = H(x, y, t) \mu(dy),$$

ce qui montre (3.42). \square

Soient $A \in \mathcal{C}_2$, $1 \leq k \leq p$ et $0 \leq t < t^*(A)$. On peut estimer la probabilité conditionnelle

$$\tilde{H}(A, B_k, t) = \mathbb{P}_\nu(S_1 > t, Z_1 \in B_k | Z_0 \in A),$$

par sa version empirique donnée par (3.12). La convergence uniforme presque sûre est donnée dans la proposition 3.8. On montre d'abord la convergence ponctuelle.

LEMME 3.39 Pour tous $0 \leq t < t^*(A)$, $x \in E$, on a, quand n tend vers l'infini,

$$\left| \hat{p}_n(A, B_k, t) - \tilde{H}(A, B_k, t) \right| \xrightarrow{\mathbb{P}_x\text{-p.s.}} 0.$$

PREUVE. On applique le théorème ergodique à (Z_n, Z_{n+1}, S_{n+1}) . Donc, quand n tend vers l'infini,

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}} \mathbb{1}_{\{S_{i+1} > t\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \eta(A \times B_k \times]t, +\infty[)$$

Avec (3.43) et (3.42),

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}} \mathbb{1}_{\{Z_{i+1} \in B_k\}} \mathbb{1}_{\{S_{i+1} > t\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \int_{A \times B_k} H(x, y, t) \nu(dx) \mu(dy). \quad (3.44)$$

De plus, en appliquant le théorème ergodique à (Z_n) , on a

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{Z_i \in A\}} \xrightarrow{\mathbb{P}_x\text{-p.s.}} \nu(A). \quad (3.45)$$

(3.12), (3.44) et (3.45) montrent le résultat. \square

On peut maintenant donner la démonstration de la proposition 3.8, qui établit que l'estimateur $\hat{p}_n(A, B_k, \cdot)$ est uniformément consistant. Ce résultat peut être vu comme un équivalent du théorème de Glivenko-Cantelli mais dans un cadre markovien.

PREUVE DE LA PROPOSITION 3.8. Puisque A et k sont fixés dans la suite, on écrira $\tilde{H}(t)$ à la place de $\tilde{H}(A, B_k, t)$ par souci de lisibilité. D'après le lemme 3.39, pour tout s , $\mathbb{P}_x(\Upsilon_s) = 1$ où l'ensemble Υ_s est donné par

$$\Upsilon_s = \{\hat{p}_n(A, B_k, s) \rightarrow \tilde{H}(s)\}.$$

On montre d'abord que \tilde{H} est une fonction strictement décroissante. Par (3.33) et le théorème de dérivation sous le signe intégrale, \tilde{H}' vérifie,

$$\begin{aligned} \forall 0 \leq s \leq t, \tilde{H}'(s) &= -\frac{1}{\nu(A)} \int_A f(\xi, s) \bar{Q}(\xi, s, B_k) \nu(d\xi) \\ &\leq -\frac{\mu(B_k)m}{\nu(A)} \int_A f(\xi, s) \nu(d\xi), \end{aligned}$$

par (3.2). D'après les hypothèses 3.3, f est strictement positive. Donc, $\int_A f(\xi, s) \nu(d\xi) > 0$ puisque $\nu(A) > 0$. Par conséquent, $\tilde{H}'(s) < 0$ pour tout s , et \tilde{H} est strictement décroissante. En particulier, il s'agit d'une bijection de $[0, t]$ dans $[a, b]$, où

$$[a, b] = [\tilde{H}(t), \mathbb{P}_\nu(Z_1 \in B_k | Z_0 \in A)],$$

puisque, par la proposition 3.38,

$$\tilde{H}(0) = \mathbb{P}_v(Z_1 \in B_k | Z_0 \in A).$$

Pour tout couple d'entiers (l, m) , avec $0 \leq l \leq m$, on considère

$$X(l, m) = \tilde{H}^{-1} \left(a + \frac{(m-l)(b-a)}{m} \right).$$

Par construction, on a pour tout $0 \leq l \leq m-1$,

$$\tilde{H}(X(l+1, m)) - \tilde{H}(X(l, m)) = \frac{b-a}{m}.$$

Ainsi, pour $s \in [X(l, m), X(l+1, m)[$ avec $0 \leq l \leq m-2$, ou pour $s \in [X(m-1, m), t]$, on a

$$\begin{aligned} \hat{p}_n(A, B_k, s) - \tilde{H}(s) &\geq \hat{p}_n(A, B_k, X(l+1, m)) - \tilde{H}(X(l, m)) \\ &\geq \hat{p}_n(A, B_k, X(l+1, m)) - \tilde{H}(X(l+1, m)) + \frac{b-a}{m}, \end{aligned} \quad (3.46)$$

$$\begin{aligned} \hat{p}_n(A, B_k, s) - \tilde{H}(s) &\leq \hat{p}_n(A, B_k, X(l, m)) - \tilde{H}(X(l+1, m)) \\ &\leq \hat{p}_n(A, B_k, X(l, m)) - \tilde{H}(X(l, m)) - \frac{b-a}{m}, \end{aligned} \quad (3.47)$$

car \tilde{H} et $\hat{p}_n(A, B_k, \cdot)$ sont décroissantes. Soit $0 \leq s \leq t$. Il y a deux possibilités : soit $s \in [X(m-1, m), t]$, soit il existe $0 \leq l \leq m-2$ tel que $s \in [X(l, m), X(l+1, m)[$. Avec (3.46) et (3.47),

$$\sup_{0 \leq s \leq t} \left| \hat{p}_n(A, B_k, s) - \tilde{H}(s) \right| \leq \alpha_{n,m} + \frac{b-a}{m},$$

où

$$\alpha_{n,m} = \sup_{0 \leq l \leq m} \left| \hat{p}_n(A, B_k, X(l, m)) - \tilde{H}(X(l, m)) \right|.$$

Soit

$$\Omega_m = \bigcap_{l=0}^m \Upsilon_{X(l,m)} \quad \text{and} \quad \Omega_\infty = \bigcap_{m \geq 2} \Omega_m.$$

Par conséquent, quand n tend vers l'infini

$$\forall \omega \in \Omega_m, \quad \alpha_{n,m}(\omega) \rightarrow 0.$$

De plus, $\mathbb{P}_x(\Omega_m) = 1$ puisque Ω_m est une intersection finie d'ensembles de probabilité 1. Finalement,

$$\limsup_{n \rightarrow +\infty} \sup_{0 \leq s \leq t} \left| \hat{p}_n(A, B_k, s) - \tilde{H}(s) \right| \leq \inf_{m \geq 2} \frac{b-a}{m} \quad \mathbb{P}_x\text{-p.s.},$$

puisque $\mathbb{P}_x(\Omega_\infty) = 1$ comme intersection dénombrable d'ensembles de probabilité 1.

3.5 Simulations

Dans cette section, on présente quelques résultats numériques d'estimation obtenus dans deux applications utilisant les PDMP. La première concerne la modélisation du mouvement d'une bactérie dans un environnement restreint (voir par exemple [47]). Le second exemple traite de l'évolution au cours du temps de la taille d'une cellule (voir notamment [38]), et est un cas particulier de modèle de croissance-fragmentation.

3.5.1 Mouvement d'une bactérie

On considère un PDMP (X_t) défini sur l'espace d'état $\mathcal{D} \times \mathcal{I}$, où

$$\mathcal{D} = \{x \in \mathbb{R}^2 : \|x\|_2 < 1\} \quad \text{et} \quad \mathcal{I} =]0, 2\pi[.$$

On suppose que le point de départ du processus est $X_0 = (0, 0, \pi)$. Soient $x = (x_1, x_2) \in \mathcal{D}$ et $\theta \in \mathcal{I}$. Le flot Φ est donné par,

$$\forall t \geq 0, \quad \Phi((x, \theta), t) = (x_1 + t \cos(\theta), x_2 + t \sin(\theta), \theta).$$

Le processus (X_t) a deux composantes : la première modélise la position dans le disque \mathcal{D} , tandis que la seconde modélise la direction du mouvement qui a lieu dans \mathcal{D} . Le taux de saut λ est donné par $\lambda((x, \theta)) = 5 + \|x\|_2$. Puisque λ ne dépend pas de l'angle θ , on écrira $\lambda(x)$ par souci de lisibilité. Finalement, le noyau de transition Q est défini, pour tous $A \in \mathcal{B}(\mathbb{R}^2)$ et $B \in \mathcal{B}(\mathbb{R})$, par

$$Q((x, \theta), A \times B) = \frac{1}{K_x} \int_A \mathbb{1}_{\mathcal{D}}(y) \exp\left(-\frac{1}{2\sigma^2} \|y - x\|_2^2\right) dy \int_B \mathbb{1}_{\mathcal{I}}(u) du,$$

où K_x est la constante de normalisation adéquate et σ^2 un paramètre de variance. Lorsqu'un saut a lieu, la nouvelle position dans \mathcal{D} est choisie selon une loi normale centrée en la précédente. Le nouvel angle de direction est choisi aléatoirement uniformément dans \mathcal{I} . On a représenté deux trajectoires du processus dans la figure 3.1.

Ce processus peut modéliser le mouvement d'une bactérie dans un environnement clos (voir [47]). La bactérie se déplace sur une ligne droite à vitesse constante. Elle change spontanément et aléatoirement de direction. Pendant la rotation, sa position peut être légèrement perturbée (le paramètre σ^2 est choisi petit, $\sigma^2 = 10^{-4}$ dans les simulations). Ensuite, la bactérie se déplace à nouveau de manière rectiligne. Un changement de direction a lieu également lorsque la bactérie rencontre la frontière de son environnement. Dans notre

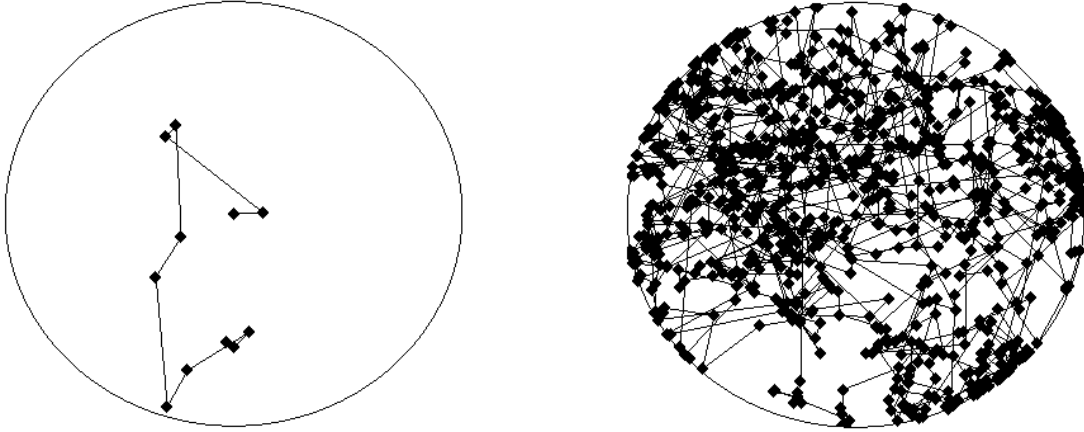


FIGURE 3.1 – Deux trajectoires du processus jusqu’au 10^{ème} saut (à gauche) et jusqu’au 1000^{ème} saut (à droite).

modèle, le taux λ dépend seulement de la distance de la bactérie au centre du disque. Le changement de position de la bactérie lors d’un saut est modélisé par un noyau gaussien afin que les hypothèses 3.1 sur le noyau de transition soient vérifiées.

Dans la suite, on se focalise sur l’estimation de la densité conditionnelle $f(x_0, t)$ pour $x_0 = (0, 0)$. On peut donner la formule explicite de $f(x_0, t)$,

$$\forall t \geq 0, f(x_0, t) = (5 + t) \exp(-t(5 + t/2)).$$

Pour tous $y \in E$ et t , on choisit d’approcher le taux $\tilde{\lambda}(x_0, y, t)$ par $\tilde{l}(A, B_k, t)$, où A est le disque centré en x_0 de rayon ε et $B_1 = A$, $B_2 = \mathcal{D} \setminus A$, c’est-à-dire que (B_k) est la partition de \mathcal{D} la plus simple qu’on puisse considérer. On simule une longue trajectoire du processus : l’observation de 50 000 sauts est disponible pour estimer f . Dans la suite, on choisit $\varepsilon = 0.2$. En particulier, cela implique que $t^*(A) = 0.8$. Par conséquent, on décide d’estimer $\tilde{l}(A, B_k, t)$ sur l’intervalle $[0.05, 0.75]$. La fenêtre de lissage $\beta_n(A, B_k)$ peut s’écrire de la façon suivante,

$$\beta_n(A, B_k) = \frac{1}{h_n(A, B_k)^\alpha},$$

où $h_n(A, B_k)$ est le nombre (aléatoire) de visites dans A suivies par une visite dans B_k , et $\alpha = 1/3$.

On présente nos résultats numériques à partir de différents nombres de sauts observés. Plus précisément, les estimations de $f(x_0, t)$, avec t entre 0 et $t^*(A)$, sont données dans la figure 3.2 à partir de 10 000, 25 000 et 50 000 données. De plus, on donne dans la figure 3.3 les estimations au temps $t = 0.64$ en fonction du nombre de sauts observés, à partir de deux trajectoires observées du PDMP.

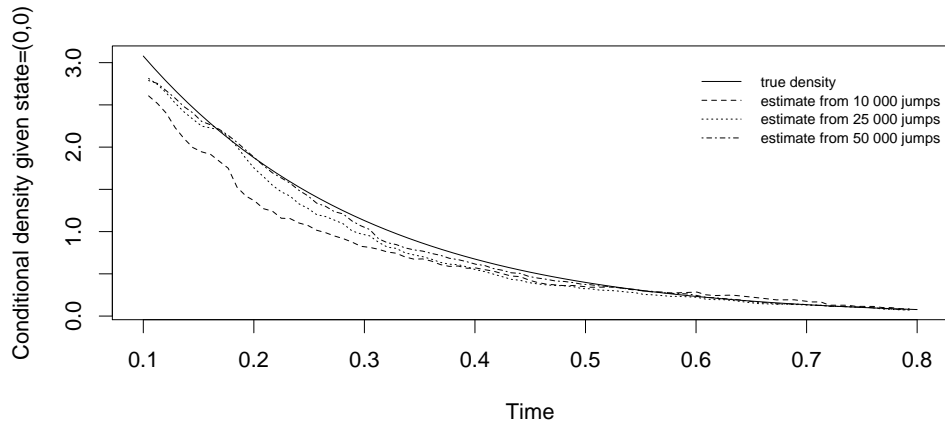


FIGURE 3.2 – Estimation de la densité conditionnelle $f(x_0, t)$, avec $x_0 = (0, 0)$ et $0.05 \leq t \leq 0.75$, à partir de différents nombres de sauts observés.

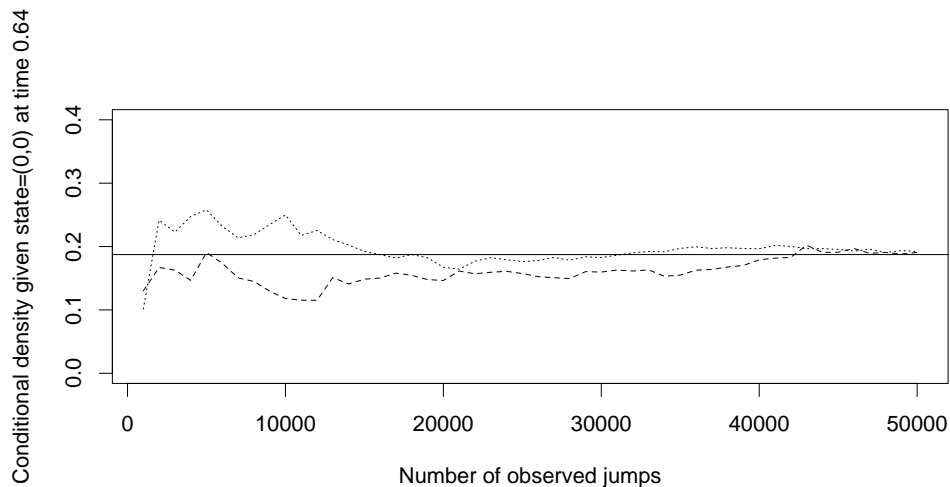


FIGURE 3.3 – Estimation de la densité conditionnelle $f(x_0, t)$ avec $x_0 = (0, 0)$ au temps $t = 0.64$ pour deux trajectoires indépendantes, à partir de différents nombres de sauts observés. La ligne horizontale représente la valeur recherchée.

3.5.2 Un modèle de croissance-fragmentation

Le PDMP (X_t) qu'on considère ici modélise la taille d'une cellule. Celle-ci croît exponentiellement au cours du temps, puis se divise en deux cellules filles à un taux λ qui dépend de sa taille. Dans notre application, on suppose que la taille de la cellule ne peut excéder un certain seuil. Dans la suite, l'espace d'état du processus (X_t) est $E =]0, 3[$. Pour tout $x \in E$, le flot déterministe $\Phi(x, t)$ est donné, pour tout $t \geq 0$, par

$$\Phi(x, t) = x \exp(\tau t),$$

où $\tau = 0.9$ dans nos expériences numériques. On suppose que les temps inter-sauts sont distribués selon la loi de Weibull dont le paramètre de forme est l'inverse de la taille de la cellule. Plus précisément, pour $x \in E$ et $t \geq 0$, la loi conditionnelle $f(x, t)$ est donnée par

$$f(x, t) = \frac{t^{(1-x)/x} \exp(-t^{1/x})}{x}.$$

Finalement, le noyau $Q(x, \cdot)$ est choisi gaussien de moyenne x et de variance σ^2 ($\sigma^2 = 10^{-2}$ dans nos simulations). Une représentation graphique de la densité d'intérêt f et une trajectoire d'un tel PDMP sont représentées dans la figure 3.4.

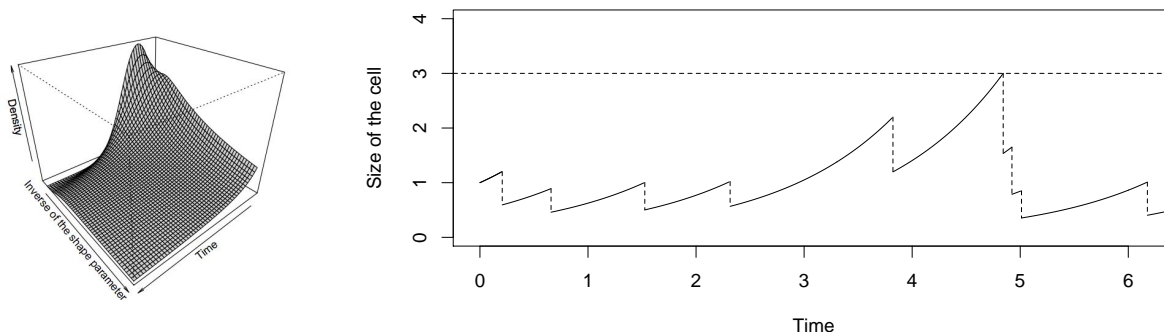


FIGURE 3.4 – Représentation de la densité conditionnelle d'intérêt $f(x, t)$, avec $0.2 \leq x \leq 3$ et $0.2 \leq t \leq 3$ (à gauche) et un exemple d'évolution de la taille de la cellule jusqu'au 10^{ème} saut (à droite).

On s'intéresse à l'estimation de la loi conditionnelle f pour les états $x_1 = 0.315$ et $x_2 = 0.685$. Comme dans la partie précédente, pour tous $y \in E$ et $t \geq 0$, le taux $\tilde{\lambda}(x_i, y, t)$ est approché par $\tilde{l}(A, B_k, t)$, où $A =]x_i - \varepsilon, x_i + \varepsilon[$ et $B_1 = A, B_2 = \mathcal{D} \setminus A$. Dans les deux

cas, $t^*(A) > 1.5$ pour $\varepsilon \leq 0.2$. Par conséquent, on présente les résultats d'estimation sur l'intervalle $[0, 1.5]$. On donne dans la figure 3.5 les estimations à partir de différents nombres de sauts observés, avec $\varepsilon = 0.1$. La fenêtre de lissage $\beta_n(A, B_k)$ est choisie de la même manière que dans l'exemple précédent. Pour le temps $t = 0.25$ et les états x_1 et x_2 , on donne les diagrammes en boîte de notre estimateur à partir de 50 réplifications dans la figure 3.6.

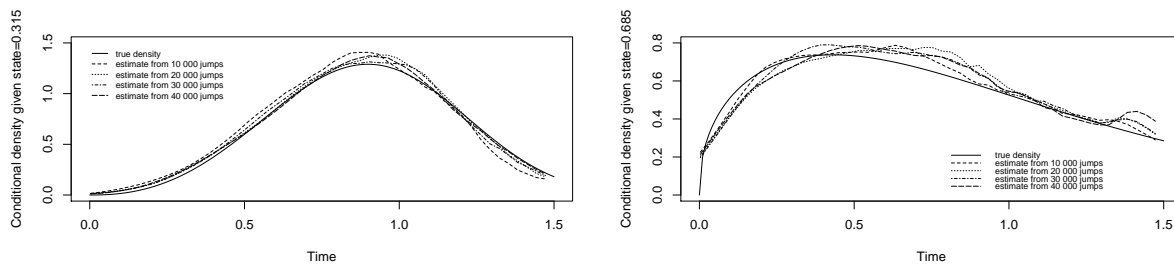


FIGURE 3.5 – Estimation de la densité conditionnelle $f(x, t)$, pour $x = x_1$ (à gauche) ou $x = x_2$ (à droite), à partir de différents nombres de sauts observés. La vraie densité est en ligne pleine.

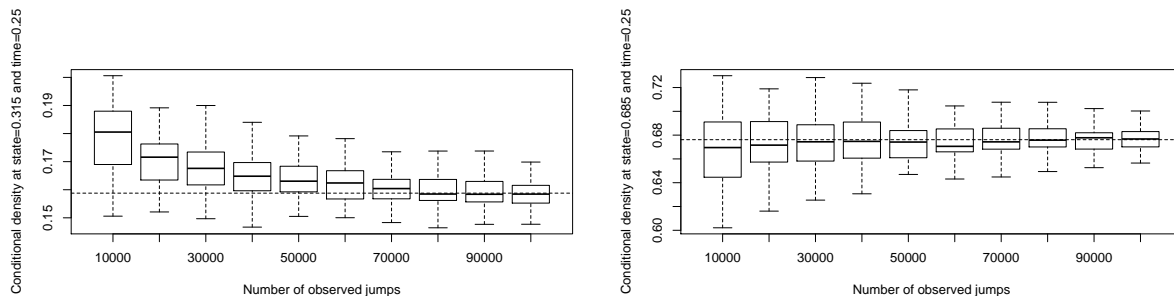


FIGURE 3.6 – Diagrammes en boîte à partir de 50 réplifications de notre estimateur de $f(x, t)$, pour $x = x_1$ (à gauche) et $x = x_2$ (à droite) au temps $t = 0.64$, à partir de différents nombres de sauts observés. La ligne horizontale en pointillés représente la vraie valeur.

Un problème naturel est le choix du paramètre ε . Lorsque ε est grand, le nombre de données disponibles est grand, mais $\tilde{l}(A, B_k, t)$ peut être une mauvaise approximation de $\tilde{\lambda}(x, y, t)$. On se propose d'illustrer ces considérations en estimant $f(x, t)$ pour les états x_1 et x_2 au temps $t = 0.64$, pour différentes valeurs de ε . On peut observer sur la figure 3.7 que $\varepsilon = 0.2$ n'est pas un choix judicieux, bien que le nombre de visites dans A soit approximativement de 27%. Pour $\varepsilon = 0.1$ (respectivement $\varepsilon = 0.01$), le nombre de visites

dans A est proche de 13% (respectivement de 7%). Un choix approprié du paramètre ε est le résultat d'un bon compromis entre un grand nombre de visites dans A et la qualité de l'approximation du taux de saut $\tilde{\lambda}$.

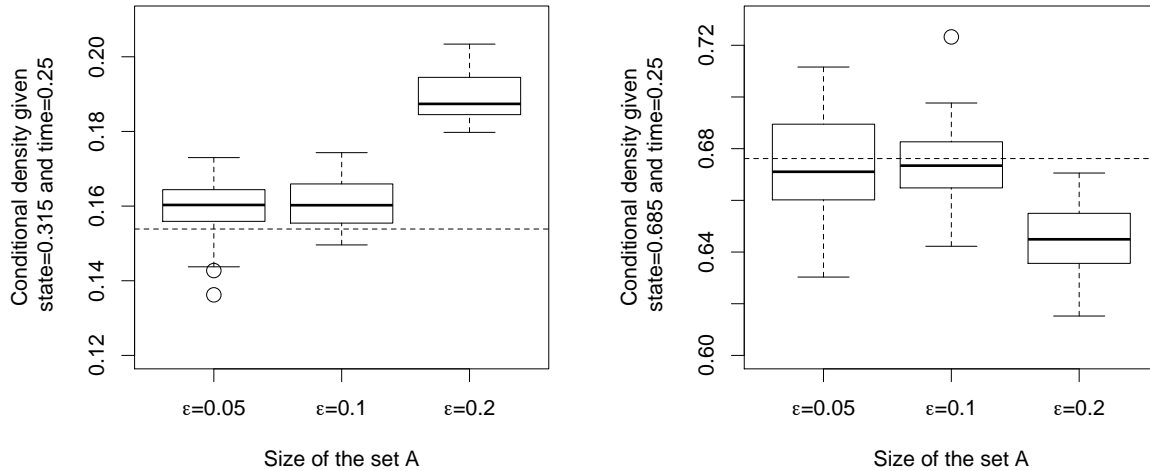


FIGURE 3.7 – Diagrammes en boîte à partir de 50 réplifications de notre estimateur de $f(x, t)$, pour $x = x_1$ (à gauche) et $x = x_2$ (à droite) au temps $t = 0.64$, à partir de 50 000 sauts observés et pour différentes valeurs de ε . La ligne horizontale représente la valeur recherchée.

3.6 Conclusion du chapitre

Dans ce chapitre, on a proposé et analysé une méthode d'estimation non paramétrique de la densité conditionnelle des temps inter-sauts d'un PDMP défini sur un espace métrique. Une partie de cette étude concerne l'estimation de la loi des durées inter-sauts sachant les positions passée et future du processus lors des sauts. Pour cela, on s'est appuyé sur la méthodologie mise en œuvre dans le chapitre 2. En analysant les liens entre la densité conditionnelle d'intérêt f et ce taux de saut, on a introduit et estimé une approximation spatiale de f . Sous des hypothèses reliées aux caractéristiques du processus, on a montré la convergence uniforme en probabilité sur tout compact de notre estimateur de f . Les résultats numériques obtenus apparaissent satisfaisants dès lors que le nombre de données est suffisamment grand.

Comme dans le chapitre 2, le choix de la fenêtre de lissage (β_n) et la discrétisation de l'espace doivent être analysés et discutés d'un point de vue théorique et numérique. En effet, pour des applications statistiques, il est important de choisir de manière adéquate ces paramètres. Enfin, à partir de la relation $\bar{\lambda} = f/G$, on pourrait estimer la quantité $\bar{\lambda}$ *via* une estimation de la fonction de survie conditionnelle G . Cela permettrait de proposer une procédure d'estimation non paramétrique pour le « vrai » taux de saut λ du PDMP.

3.7 Preuve de la proposition 3.29

Pour la première indépendance conditionnelle, soient h_1, \dots, h_n des fonctions mesurables bornées de \mathbb{R}_+ dans \mathbb{R} . On a $\mathcal{G}_n \subset \mathcal{G}_n \vee \sigma(\delta_{n-1})$ et, d'après (3.1), $h_n(S_n)$ est $\mathcal{G}_n \vee \sigma(\delta_{n-1})$ -mesurable. Donc,

$$\begin{aligned} & \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] \\ &= \mathbb{E}_{\nu_0} \left[\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \sigma(\delta_{n-1}) \vee \mathcal{G}_n] | \tilde{\mathcal{G}}_n \right] \\ &= \mathbb{E}_{\nu_0} \left[h_n(S_n) \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \sigma(\delta_{n-1}) \vee \mathcal{G}_n] | \mathcal{G}_n \right]. \end{aligned} \quad (3.48)$$

De plus, toujours par (3.1), $\sigma(\delta_{n-1}) \vee \mathcal{G}_n \subset \sigma(\delta_{n-1}) \vee \mathcal{G}_{n-1} \vee \sigma(\varepsilon_{n-1})$. Par conséquent,

$$\begin{aligned} & \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \sigma(\delta_{n-1}) \vee \mathcal{G}_n] \\ &= \mathbb{E}_{\nu_0} \left[\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \sigma(\delta_{n-1}) \vee \mathcal{G}_{n-1} \vee \sigma(\varepsilon_{n-1})] | \sigma(\delta_{n-1}) \vee \mathcal{G}_n \right]. \end{aligned} \quad (3.49)$$

Néanmoins, $h_1(S_1) \dots h_{n-1}(S_{n-1})$ est $\mathcal{G}_{n-1} \vee \sigma(\delta_0, \dots, \delta_{n-2})$ -mesurable, et de plus,

$$\sigma(\varepsilon_{n-1}) \vee \sigma(\delta_{n-1}) \perp \mathcal{G}_{n-1} \vee \sigma(\delta_0, \dots, \delta_{n-2}).$$

Avec (3) [19, page 308],

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \sigma(\delta_{n-1}) \vee \mathcal{G}_{n-1} \vee \sigma(\varepsilon_{n-1})] = \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_{n-1}].$$

Finalement, avec (3.49),

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \sigma(\delta_{n-1}) \vee \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \mathcal{G}_{n-1}].$$

Alors, par (3.48),

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_1(S_1) \dots h_{n-1}(S_{n-1}) | \tilde{\mathcal{G}}_{n-1}] \mathbb{E}_{\nu_0} [h_n(S_n) | \mathcal{G}_n].$$

Par une récurrence immédiate, on a

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] = \prod_{i=1}^n \mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_i].$$

Prenons, pour $j \neq i$, $h_j = \mathbb{1}$. Alors,

$$\mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_i]. \quad (3.50)$$

Ainsi,

$$\mathbb{E}_{\nu_0} [h_1(S_1) \dots h_n(S_n) | \mathcal{G}_n] = \prod_{i=1}^n \mathbb{E}_{\nu_0} [h_i(S_i) | \mathcal{G}_i].$$

Cela montre que

$$\bigvee_{j \neq i} \sigma(S_j) \perp_{\mathcal{G}_n} \sigma(S_i),$$

et cela entraîne directement le résultat recherché. Pour la seconde indépendance conditionnelle, soient maintenant $h_1 : \mathbb{R}_+ \rightarrow \mathbb{R}$ et $h_2 : \mathbb{E}^{n+1} \rightarrow \mathbb{R}$ des fonctions mesurables bornées. En remarquant que $\sigma(Z_{i-1}, Z_i) \subset \mathcal{G}_n$, on a

$$\begin{aligned} & \mathbb{E}_{\nu_0} [h_1(S_i) h_2(Z_0, \dots, Z_n) | \sigma(Z_{i-1}, Z_i)] \\ &= \mathbb{E}_{\nu_0} \left[\mathbb{E}_{\nu_0} [h_1(S_i) h_2(Z_0, \dots, Z_n) | \mathcal{G}_n] | \sigma(Z_{i-1}, Z_i) \right] \\ &= \mathbb{E}_{\nu_0} \left[h_2(Z_0, \dots, Z_n) \mathbb{E}_{\nu_0} [h_1(S_i) | \mathcal{G}_n] | \sigma(Z_{i-1}, Z_i) \right]. \end{aligned} \quad (3.51)$$

On va prouver que

$$\mathbb{E}_{\nu_0} [h_1(S_i) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_1(S_i) | \sigma(Z_{i-1}, Z_i)]. \quad (3.52)$$

Par (3.50), on a

$$\mathbb{E}_{\nu_0} [h_1(S_i) | \mathcal{G}_n] = \mathbb{E}_{\nu_0} [h_1(S_i) | \sigma(Z_0, \dots, Z_i)].$$

Donc, pour établir (3.52), on peut se contenter de montrer

$$\sigma(S_i) \perp_{\sigma(Z_{i-1}, Z_i)} \sigma(Z_0, \dots, Z_{i-2}). \quad (3.53)$$

De (3.1), on a

$$\sigma(\delta_{i-1}, \varepsilon_{i-1}) \perp \sigma(Z_0, \dots, Z_{i-2}) \vee \sigma(Z_{i-1}).$$

Donc, d'après la proposition 6.8 de [58] dans le sens \Rightarrow , on a

$$\sigma(\delta_{i-1}, \varepsilon_{i-1}) \perp_{\sigma(Z_{i-1})} \sigma(Z_0, \dots, Z_{i-2}). \quad (3.54)$$

De plus, il est facile de voir que

$$\sigma(Z_0, \dots, Z_{i-2}) \perp_{\sigma(Z_{i-1})} \sigma(Z_{i-1}). \quad (3.55)$$

Donc, avec (3.54), (3.55) et la proposition 6.8 de [58] à nouveau, mais dans le sens \Leftarrow , on a

$$\sigma(Z_{i-1}, \delta_{i-1}, \varepsilon_{i-1}) \perp_{\sigma(Z_{i-1})} \sigma(Z_0, \dots, Z_{i-2}).$$

On a de plus l'égalité des tribus,

$$\sigma(Z_{i-1}, Z_i, \delta_{i-1}, \varepsilon_{i-1}) = \sigma(Z_{i-1}, \delta_{i-1}, \varepsilon_{i-1}),$$

puisque la variable Z_i est générée par Z_{i-1} et les erreurs δ_{i-1} et ε_{i-1} . Ainsi,

$$\sigma(Z_i) \vee \sigma(Z_{i-1}, \delta_{i-1}, \varepsilon_{i-1}) \perp_{\sigma(Z_{i-1})} \sigma(Z_0, \dots, Z_{i-2}).$$

Donc, à nouveau par la proposition 6.8 de [58], dans le sens \Rightarrow , on a

$$\sigma(Z_{i-1}, \delta_{i-1}, \varepsilon_{i-1}) \perp_{\sigma(Z_{i-1}, Z_i)} \sigma(Z_0, \dots, Z_{i-2}).$$

Cela montre (3.53) puisque le temps S_i est généré à partir de Z_{i-1} et de l'erreur δ_{i-1} . On a donc montré (3.52). Par conséquent, (3.51) et (3.52) mènent à

$$\begin{aligned} & \mathbb{E}_{\nu_0} [\mathfrak{h}_1(S_i) \mathfrak{h}_2(Z_0, \dots, Z_n) | \sigma(Z_{i-1}, Z_i)] \\ &= \mathbb{E}_{\nu_0} [\mathfrak{h}_1(S_i) | \sigma(Z_{i-1}, Z_i)] \mathbb{E}_{\nu_0} [\mathfrak{h}_2(Z_0, \dots, Z_n) | \sigma(Z_{i-1}, Z_i)], \end{aligned}$$

ce qui montre le résultat voulu.

Deuxième partie

ESTIMATION D'UNE TRANSITION

APPLICATION DE LA QUANTIFICATION OPTIMALE À LA MÉTHODE SIR

Ce chapitre est consacré à une application de la quantification optimale en norme \mathbb{L}^p à la méthode SIR (pour l'anglais Sliced Inverse Regression) dans le but d'étudier le modèle de régression semi-paramétrique $Y = f(\beta'X, \varepsilon)$. Ce type de modèle est utilisé lorsque la dimension de la variable explicative X est grande. On cherche alors à déterminer quelles composantes de X sont importantes pour décrire le comportement de la variable d'intérêt Y . Dans ce chapitre, on approche la variable X par sa projection optimale en norme \mathbb{L}^p sur une grille finie à N points, et on détermine une approximation β_N de β . Dans un second temps, on approche la loi conditionnelle de Y sachant $\beta'X$ par la loi conditionnelle de la projection de Y sachant la projection de $\beta'_N X$. Les résultats d'approximation sont à chaque fois donnés avec les vitesses de convergence correspondantes et sont illustrés par une étude de simulations. Ce travail a donné lieu à l'article *Optimal quantization applied to Sliced Inverse Regression* [11], avec Anne Gégout-Petit et Jérôme Saracco, publié dans *Journal of Statistical Planning and Inference*.

Le chapitre est organisé de la façon suivante. La section 4.1 est introductive sur l'application de la quantification à la méthode SIR. La section 4.2 est entièrement consacrée à un panorama de la quantification optimale de vecteurs aléatoires et de chaînes de Markov. Cette partie est indépendante et n'est pas extraite de l'article [11]. La suite du chapitre reprend l'ensemble de l'étude publiée dans [11]. Dans la section 4.3, on propose une méthode d'approximation du modèle de régression non paramétrique $Y = \tilde{f}(U, \varepsilon)$. Le modèle semi-paramétrique est étudié dans la section 4.4 : on montre comment approcher le paramètre inconnu β dans la sous-section 4.4.1 alors que la prédiction pour le modèle de régression semi-paramétrique est étudiée dans la sous-section 4.4.2. On trouve l'ensemble des preuves des résultats des sections 4.3 et 4.4 dans la section 4.7. Des simulations illustrent les méthodes mises en œuvre dans la section 4.5. Une conclusion et des perspectives à ce travail sont données dans la section 4.6

SOMMAIRE DU CHAPITRE

4.1	Introduction	97
4.2	Quantification optimale en norme \mathbb{L}^p	98
4.2.1	Quantification d'un vecteur aléatoire	99
4.2.2	Quantification d'une chaîne de Markov	104
4.3	Étude d'un modèle non paramétrique	107
4.4	Étude du modèle semi-paramétrique	109
4.4.1	Approximation de la direction de β	109
4.4.2	Prédiction dans le cadre semi-paramétrique	111
4.5	Simulations	112
4.5.1	Approximation du sous-espace engendré par β	113
4.5.2	Approximation de la loi de Y sachant $\beta'X$	114
4.6	Conclusion du chapitre	115
4.7	Preuves des résultats des sections 4.3 et 4.4	120
4.7.1	Preuve de la proposition 4.9	120
4.7.2	Preuve de la proposition 4.12	121
4.7.3	Preuve de la proposition 4.11	123
4.7.4	Preuve du théorème 4.13	123

4.1 Introduction

Lorsqu'on étudie la régression d'une variable réponse Y sur une variable explicative X d -dimensionnelle, on cherche une représentation parcimonieuse de la loi conditionnelle qui nous intéresse. Dans beaucoup d'applications pratiques, la dimension d de la variable X est très grande, ce qui induit une étude statistique difficile. Une approche classique lorsqu'on rencontre ce type de problème est la réduction de la dimension de la partie explicative du modèle sans perte d'information, et sans considérer un modèle trop spécifique. Pour cela, il existe différentes méthodes de réduction de dimension dont l'objet est de remplacer la variable explicative X par une combinaison linéaire de ses composantes.

Dans ce chapitre, nous considérons le modèle semi-paramétrique suivant,

$$Y = f(\beta'X, \varepsilon), \quad (4.1)$$

où la variable réponse Y est réelle et est liée, *via* une fonction de lien inconnue f , au vecteur aléatoire d -dimensionnel X à travers un paramètre inconnu β , lui aussi de dimension d . La variable aléatoire ε est un terme d'erreur indépendant de X . Le modèle (4.1) peut aussi être défini par $Y \perp X | \beta'X$ où le symbole \perp représente l'indépendance. En particulier, cela signifie que pour la régression de Y sur X , une statistique suffisante est la combinaison linéaire $\beta'X$.

Il existe de nombreuses méthodes statistiques pour identifier et estimer le sous-espace vectoriel engendré par le paramètre β . On peut citer notamment la méthode Sliced Inverse Regression (SIR) introduite par Duan et Li en 1991 dans [40], ou les méthodes Principal Hessian Directions (voir par exemple [22]) et Sliced Average Variance Estimation (SAVE) (voir [23]) développées par Cook. Sans hypothèse supplémentaire, seul le sous-espace engendré par β est identifiable dans ce modèle. Il est souvent appelé sous-espace central de réduction de la dimension ou sous-espace EDR pour l'anglais Effective Dimension Reduction. Pour estimer ce sous-espace, la méthode SIR utilise des propriétés de l'espérance conditionnelle de la variable explicative X sachant la variable d'intérêt Y sous certaines conditions sur la loi de X , tandis que la méthode SAVE est basée sur des propriétés de la variance conditionnelle de X sachant Y . Dans ce papier, on se focalise uniquement sur l'approche SIR qui est devenue au fil du temps la méthode standard, notamment grâce à sa simplicité de mise en œuvre. L'idée principale est de trancher l'espace d'arrivée de Y en H sous-ensembles S_h . On pose alors $\hat{Y} = \{h : Y \in S_h\}$ pour $h \in \{1, \dots, H\}$. Il s'agit d'une version discrète de la variable réponse Y . En particulier \hat{Y} peut être vue comme une projection de Y sur une grille finie. Sous une hypothèse portant sur la loi de X et

une autre qui permet d'assurer que le modèle de régression n'est pas un cas pathologique, on peut montrer que le sous-espace propre associé à la plus grande valeur propre de la matrice $\Sigma^{-1}\widehat{\Gamma}$ est colinéaire au vecteur inconnu β , où la matrice $\widehat{\Gamma}$ est donnée par $\widehat{\Gamma} = \sum_{h=1}^H \mathbb{P}(Y \in S_h)(\mathbb{E}[X|Y \in S_h] - \mathbb{E}[X])(\mathbb{E}[X|Y \in S_h] - \mathbb{E}[X])'$ et peut facilement être estimée par sa version empirique.

Dans ce travail, on propose d'utiliser la quantification optimale afin de déterminer une approximation numérique du vecteur β et de la loi conditionnelle de Y sachant $\beta'X$. Un panorama de la quantification optimale d'un vecteur aléatoire et d'une chaîne de Markov est donné dans la section 4.2 afin de bien présenter cet outil qui est un point clé de notre approche. Dans la section 4.3, on utilise la quantification optimale dans le but d'approcher un premier modèle de régression non paramétrique. Puis, dans la section 4.4, on étudie le modèle semi-paramétrique (4.1). On propose d'abord, dans la sous-section 4.4.1, une méthode d'approximation de la direction de β . L'idée est de remplacer X par sa projection optimale \widehat{X}^N . On note $\widehat{\Gamma}_N = \text{Var}(\mathbb{E}[\widehat{X}^N|Y])$. On montre que $\widehat{\Gamma}_N$ converge vers la matrice $\widehat{\Gamma}$ quand N tend vers l'infini, et on contrôle la vitesse de convergence. De ce résultat, on montre la convergence de la suite des sous-espaces propres principaux de $(\widehat{\Gamma}_N)$ vers la direction de β . En pratique, la quantification optimale est souvent utilisée pour approcher des espérances conditionnelles (voir par exemple [34] et [35]). Dans la sous-section 4.4.2, on propose d'utiliser la quantification optimale dans le but d'approcher la loi de Y sachant X , c'est-à-dire sachant $\beta'X$ pour le modèle considéré, et nous montrons – en nous appuyant sur l'étude de la section 4.3, – un résultat de convergence quand le nombre de points des grilles tend vers l'infini. Cette approche est intéressante car la plupart des papiers portant sur la méthode SIR ne s'attachent pas à proposer une méthode d'estimation de la fonction de lien. Seuls quelques papiers sur ce sujet sont disponibles dans la littérature, par exemple [48]. Dans la section 4.5, nous illustrons notre approche par des simulations et nous illustrons le bon comportement de nos approximations pour différentes tailles d'échantillons. Les preuves de l'ensemble des résultats des sections 4.3 et 4.4 se trouvent dans la section 4.7.

4.2 Quantification optimale en norme \mathbb{L}^p

La quantification optimale en norme \mathbb{L}^p consiste à déterminer une ou la meilleure approximation prenant un nombre fini et fixé de valeurs d'une variable aléatoire. Les applications de la quantification optimale sont nombreuses, notamment en intégration numérique et en finance (voir par exemple [69, 71]). Récemment, la quantification optimale

a été un outil clé dans le traitement de problèmes numériques (arrêt optimal, contrôle impulsif, moments des temps de sortie) pour les processus markoviens déterministes par morceaux (voir notamment [16, 33, 35]). Dans cette section, on s'intéresse d'abord à la quantification d'un vecteur aléatoire, puis à deux méthodes d'approximation numérique pour les chaînes de Markov. Nous nous sommes inspirés de celles-ci dans le but de proposer une version quantifiée du modèle semi-paramétrique d'intérêt (4.1).

4.2.1 Quantification d'un vecteur aléatoire

On présente ici quelques résultats classiques en quantification d'un vecteur aléatoire. Ils sont extraits de [50, 69, 71]. On montre notamment que considérer une meilleure approximation en norme \mathbb{L}^p d'une variable aléatoire a du sens. On donne aussi des résultats asymptotiques sur l'erreur de quantification et on s'intéresse à un algorithme de gradient permettant d'obtenir la version quantifiée d'une variable aléatoire.

On considère un vecteur aléatoire X sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans \mathbb{R}^d , muni de la tribu borélienne $\mathcal{B}(\mathbb{R}^d)$ pour la norme euclidienne $|\cdot|$. Dans toute la suite, λ_d désigne la mesure de Lebesgue sur \mathbb{R}^d . Pour toute grille $\gamma = \{x^1, \dots, x^N\}$ de N points de \mathbb{R}^d , la projection sur γ au plus proche voisin est donnée par

$$\text{Proj}_\gamma : x \mapsto \text{Proj}_\gamma(x) = \sum_{i=1}^N x^i \mathbb{1}_{C_i(\gamma)}(x),$$

où $\{C_1(\gamma), \dots, C_N(\gamma)\}$ est une partition d'ensembles mesurables de \mathbb{R}^d vérifiant,

$$\forall 1 \leq i \leq N, \quad C_i(\gamma) \subset \left\{ x \in \mathbb{R}^d : |x - x^i| = \min_{1 \leq j \leq N} |x - x^j| \right\}.$$

Soit $p \geq 1$; nous supposons que le moment d'ordre p de la loi \mathbb{P}_X de X est fini. On considère l'application $Q_N^p(\mathbb{P}_X)(\cdot)$ définie sur \mathbb{R}^{dN} par

$$Q_N^p(\mathbb{P}_X) : (x^1, \dots, x^N) \mapsto \|X - \text{Proj}_\gamma(X)\|_p^p,$$

avec $\gamma = \{x^1, \dots, x^N\}$. On désigne par \mathcal{C} l'enveloppe convexe fermée de $X(\Omega)$ et par \mathcal{D}_N l'ensemble des N -grilles dont les éléments sont deux à deux distincts. Dans la suite, nous noterons simplement Q_N^p pour $Q_N^p(\mathbb{P}_X)(\cdot)$.

PROPOSITION 4.1 *On suppose que le cardinal de $X(\Omega)$ est infini. Alors l'application Q_N^p est continue et différentiable λ_{dN} -presque partout, et atteint son minimum sur $\mathcal{C}^N \cap \mathcal{D}_N$.*

En vertu de ce résultat, parler de meilleure approximation en norme \mathbb{L}^p d'une variable aléatoire a du sens. Pour une variable aléatoire X , un problème naturel est alors de déterminer une N -grille γ^* telle que l'erreur \mathbb{L}^p induite par la projection sur γ_N^* est minimale, *id est*

$$\gamma_N^* \in \arg \min_{|\gamma| \leq N} \|X - \text{Proj}_\gamma(X)\|_p, \quad (4.2)$$

où, lorsque A est un ensemble, $|A|$ désigne son cardinal. Une approximation de X est alors la variable aléatoire $\text{Proj}_{\gamma_N^*}(X)$. Nous verrons plus loin que la grille de quantification optimale pour la loi uniforme est unique et peut s'écrire explicitement. Pour la grande majorité des lois, ce n'est malheureusement pas le cas, mais nous donnerons un algorithme de gradient qui permet de résoudre numériquement ce problème.

La proposition et le théorème qui suivent précisent le comportement asymptotique de l'erreur de quantification. Le premier est une généralisation d'un résultat de Pierce de 1970 (voir [73]). Nous utiliserons ce résultat dans la suite de ce chapitre. Dans la suite, γ_N^* désigne une grille satisfaisant (4.2).

PROPOSITION 4.2 *Si X est dans $\mathbb{L}^{p+\delta}$ pour un certain $\delta > 0$, alors il existe des constantes strictement positives D_1, D_2, D_3 telles que, pour tout $N \geq D_3$, on a*

$$\|X - \text{Proj}_{\gamma_N^*}(X)\|_p^p \leq N^{-p/d} \left[D_1 \|X\|_{p+\delta}^{p+\delta} + D_2 \right]. \quad (4.3)$$

L'erreur de quantification est donc de l'ordre de $N^{-p/d}$. On peut même être plus précis puisqu'il s'agit de l'ordre exact de convergence. Le théorème suivant (voir le chapitre II de [50]) est une généralisation du théorème de Zador de 1962 (voir [87]) qui avait démontré ce résultat pour $p = 2$.

THÉORÈME 4.3 *On suppose qu'il existe $\delta > 0$ tel que X est dans $\mathbb{L}^{p+\delta}$ et qu'on a la décomposition $\mathbb{P}_X(dx) = g(x)\lambda_d(dx) + \nu(dx)$ où ν est étrangère à λ_d . Alors, il existe une constante $J_{p,d}$ indépendante de la loi de X telle que, quand N tend vers l'infini,*

$$N^{p/d} \|X - \text{Proj}_{\gamma_N^*}(X)\|_p^p \rightarrow J_{p,d} \|g\|_{d/(d+p)}.$$

La constante $J_{p,d}$ est la limite quand N tend vers l'infini de la partie de gauche lorsque X suit la loi uniforme sur l'hypercube unité de dimension d . On ne sait pas calculer cette constante en général sauf dans le cas $d = 1$ comme on le verra plus loin. L'expression de la constante $J_{p,1}$ est donnée dans la suite (voir (4.5)).

REMARQUE 4.4 *L'hypothèse d'existence d'un $\delta > 0$ tel que le moment d'ordre $p + \delta$ de X est fini est importante. Pour s'en convaincre, on peut considérer l'exemple suivant. Soit pour tout $k \geq 2$, $x_k = 3 \times 2^{k-1}$. On considère la variable aléatoire discrète X à valeurs dans $\{x_k : k \geq 2\}$ dont la loi est donnée, pour tout $k \geq 2$, par $\mathbb{P}(X = x_k) = C/2^{kp} k \ln^2(k)$, où C est la constante de normalisation. Alors X est dans \mathbb{L}^p mais jamais dans $\mathbb{L}^{p+\delta}$ pour $\delta > 0$, et on a*

$$\lim_{N \rightarrow +\infty} N^p Q_N^p(\gamma_N^*) = +\infty.$$

Nous venons d'étudier l'existence d'une grille de quantification optimale ainsi que le comportement asymptotique de l'erreur de quantification. Mais, en pratique, comment obtenir une grille optimale ? Pour la loi uniforme, on peut mener des calculs exacts et obtenir une telle grille. Pour la plupart des lois, cette démarche n'a aucune chance d'aboutir ; ce problème est alors résolu numériquement par un algorithme de gradient stochastique. Celui-ci est justifié par la proposition 4.5 et le théorème 4.1. Par convention, $x/|x| = 0$ lorsque $x = 0$.

PROPOSITION 4.5 *Si $p > 1$ et si le moment d'ordre p de X est fini, alors Q_N^p est continûment différentiable en toute grille $\gamma \in \mathcal{D}_N$ telle que*

$$\mathbb{P}_X \left(\bigcup_{i=1}^N \partial C_i(\gamma) \right) = 0.$$

En une telle grille, le gradient est donné par

$$\nabla Q_N^p(\gamma) = p \left(\int_{C_i(\gamma)} |x^i - \xi|^{p-1} \frac{x^i - \xi}{|x^i - \xi|} \mathbb{P}_X(d\xi) \right)_{1 \leq i \leq N}.$$

Lorsque $p = 1$, le résultat est encore vrai en toute grille γ telle que $\mathbb{P}_X(\{x^i\}) = 0$ pour tout $1 \leq i \leq N$.

Les grilles γ qui minimisent l'erreur \mathbb{L}^p de quantification sont parmi les grilles qui annulent le gradient ∇Q_N^p . Si \widehat{X} est la projection sur une telle grille de X , on dit que \widehat{X} est un quantifieur stationnaire, et on a la propriété qui suit.

PROPOSITION 4.6 Pour tout N et pour tout $p \geq 1$, si γ^* est une N -grille telle que $\nabla Q_N^p(\gamma^*) = 0$, alors on a

$$\mathbb{E}[X|\widehat{X}] = \widehat{X}, \quad (4.4)$$

où $\widehat{X} = \text{Proj}_{\gamma^*}(X)$.

On en déduit deux corollaires immédiats : on a l'égalité $\mathbb{E}[\widehat{X}] = \mathbb{E}[X]$, et pour tout $r \geq 1$, $\|\widehat{X}\|_r \leq \|X\|_r$, par contraction dans \mathbb{L}^r de l'espérance conditionnelle.

On cherche une grille stationnaire *via* un algorithme de gradient stochastique. Dans la suite, pour toute grille $\gamma = \{x^1, \dots, x^N\}$ et tout point ξ de \mathbb{R}^d , on pose

$$\nabla_x q_N^p(\gamma, \xi) = p \left(|x^i - \xi|^{p-1} \frac{x^i - \xi}{|x^i - \xi|} \mathbb{1}_{C_i(\gamma)}(\xi) \right)_{1 \leq i \leq N}.$$

Il s'agit de l'erreur locale de quantification. L'algorithme de gradient peut être défini ainsi.

1. γ_0 est une grille initiale de \mathcal{D}_N .
2. Lorsqu'on dispose de la grille γ_s , on propose

$$\gamma_{s+1} = \gamma_s - \frac{\delta_{s+1}}{p} \nabla_x q_N^p(\gamma_s, \xi^{s+1}),$$

où la suite (ξ_s) est une suite de variables aléatoires indépendantes de loi \mathbb{P}_X , et la suite (δ_s) est déterministe, positive, et vérifie

$$\sum_{s \geq 1} \delta_s = +\infty \quad \text{et} \quad \sum_{s \geq 1} \delta_s^2 < +\infty.$$

Sous certaines conditions, on a convergence de l'algorithme vers une grille optimale (voir dans [72] le théorème 2.4). L'algorithme de gradient stochastique permet d'obtenir en général de bons résultats numériques, pour certaines lois et lorsque la suite (δ_s) est bien choisie. Dans le cas de la quantification optimale de la gaussienne centrée réduite en norme \mathbb{L}^2 pour N quantifieurs, on prendra comme conseillé dans [72],

$$\forall s \geq 1, \delta_s = \frac{4N^{1/d}}{4N^{1/d} + \pi^2 N^{-2/d} s}.$$

Des résultats graphiques – les figures 4.1 et 4.2 – sont présentés plus bas. Il s'agit de grilles obtenues après dix mille et cent mille itérations de l'algorithme de gradient stochastique.

Dans le cas de la loi uniforme sur le segment unité, les calculs peuvent être menés de manière exacte. En effet, pour cette loi, pour tout entier $N \geq 1$ et pour tout $p \geq 1$, on a

$$\nabla Q_N^p(\gamma) = \left(\int_0^{\frac{x^1+x^2}{2}} |x^1 - \xi|^{p-1} \frac{x^1 - \xi}{|x^1 - \xi|} d\xi, \dots, \int_{\frac{x^{N-1}+x^N}{2}}^1 |x^N - \xi|^{p-1} \frac{x^N - \xi}{|x^N - \xi|} d\xi \right),$$

en toute grille $\gamma = \{x^1, \dots, x^N\} \subset [0, 1]$ dont les éléments sont rangés dans l'ordre croissant. Ecrire $\nabla Q_N^p(\gamma) = 0$ mène au système suivant,

$$\begin{cases} \frac{1}{p} \left((x^1)^p - \left(\frac{x^2-x^1}{2} \right)^p \right) & = 0, \\ \vdots \\ \frac{1}{p} \left(\left(\frac{x^{i+1}-x^i}{2} \right)^p - \left(\frac{x^{i+2}-x^{i+1}}{2} \right)^p \right) & = 0, \\ \vdots \\ \frac{1}{p} \left(\left(\frac{x^N-x^{N-1}}{2} \right)^p - (1-x^N)^p \right) & = 0. \end{cases}$$

L'unique solution de ce système est donnée, pour tout $1 \leq i \leq N$, par

$$x^i = \frac{1}{2N} + \frac{i-1}{N}.$$

Ainsi, l'unique N -grille \mathbb{L}^p -optimale est

$$\gamma^* = \left\{ \frac{1}{2N}, \dots, \frac{1}{2N} + \frac{i}{N}, \dots, \frac{1}{2N} + \frac{N-1}{N} \right\}.$$

On peut calculer l'erreur de quantification associée à cette grille,

$$Q_N^p(\gamma^*) = 2N \int_0^{\frac{1}{2N}} \left(\frac{1}{2N} - \xi \right)^p d\xi = \frac{1}{2^p(p+1)N^p}.$$

En particulier, on déduit de ce calcul la valeur de la constante $J_{p,1}$,

$$\lim_{N \rightarrow +\infty} N^p Q_N^p(\gamma^*) = J_{p,1} = \frac{1}{2^p(p+1)}. \quad (4.5)$$

4.2.2 Quantification d'une chaîne de Markov

Les résultats de cette partie sont présentés et démontrés dans [71]. Ils concernent deux méthodes d'approximation numérique pour les chaînes de Markov basées sur la quantification optimale en norme \mathbb{L}^p . Dans cette partie, par convention, $\frac{1-u^m}{1-u} = m$ pour m entier et $u = 1$. La constante de Lipschitz d'une application f lipschitzienne sera notée $[f]_{\text{Lip}}$.

4.2.2.1 Quantification des lois marginales

On considère une chaîne de Markov (X_k) sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans \mathbb{R}^d . La loi initiale de la chaîne est notée μ_0 et la suite de ses noyaux de transition (P_k) ,

$$\forall k \geq 1, \forall x \in \mathbb{R}^d, \forall A \in \mathcal{B}(\mathbb{R}^d), \quad \mathbb{P}(X_k \in A | X_{k-1} = x) = \int_A P_k(x, dy).$$

Soit n un entier strictement positif fixé, on s'intéresse à la quantification optimale du processus pour k entre 0 et n . On suppose que pour tout k , on dispose d'une grille \mathbb{L}^p -optimale $\gamma_k = \{x^1, \dots, x^{N_k}\}$ de N_k points de \mathbb{R}^d , *id est* $\widehat{X}_k = \text{Proj}_{\gamma_k}(X_k)$ est la meilleure approximation de X_k prenant N_k valeurs.

Le processus (\widehat{X}_k) n'est évidemment pas markovien. On peut néanmoins considérer sa loi initiale $\widehat{\mu}_0$ ainsi que les matrices de transition \widehat{P}_k ,

$$\forall k \geq 1, \forall x^i \in \gamma_{k-1}, \forall x^j \in \gamma_k, \quad \widehat{P}_k^{ij} = \mathbb{P}(\widehat{X}_k = x^j | \widehat{X}_{k-1} = x^i).$$

On considère la chaîne de Markov (\widehat{X}_k^c) de loi initiale $\widehat{\mu}_0$ et de suite des matrices de transition (\widehat{P}_k) . On cherche à évaluer l'erreur commise lorsqu'on approche la loi de (X_k) par celle de (\widehat{X}_k^c) . Dans ce but, on évalue la différence

$$\left| \mathbb{E}[\varphi_0(X_0) \dots \varphi_n(X_n)] - \mathbb{E}[\varphi_0(\widehat{X}_0^c) \dots \varphi_n(\widehat{X}_n^c)] \right|,$$

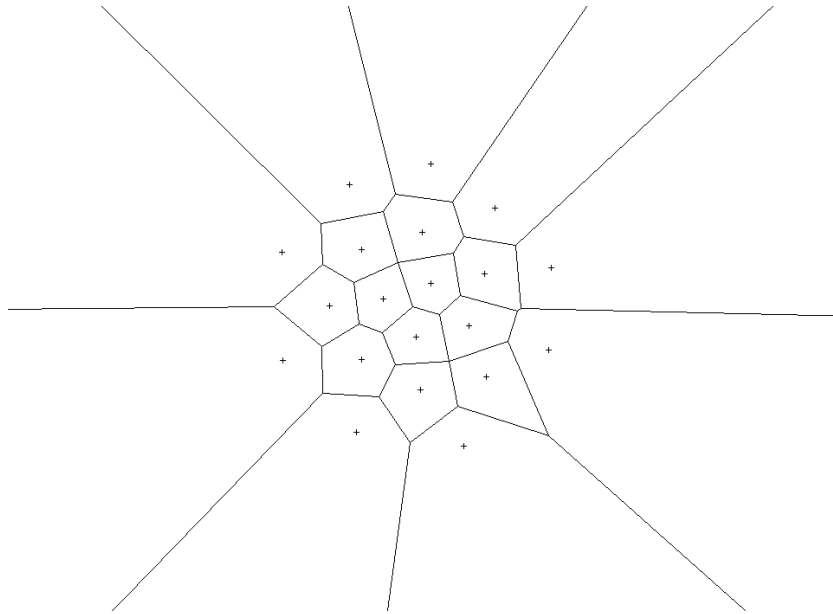


FIGURE 4.1 – Centres et cellules d’une grille optimale pour la loi normale avec $N = 20$.

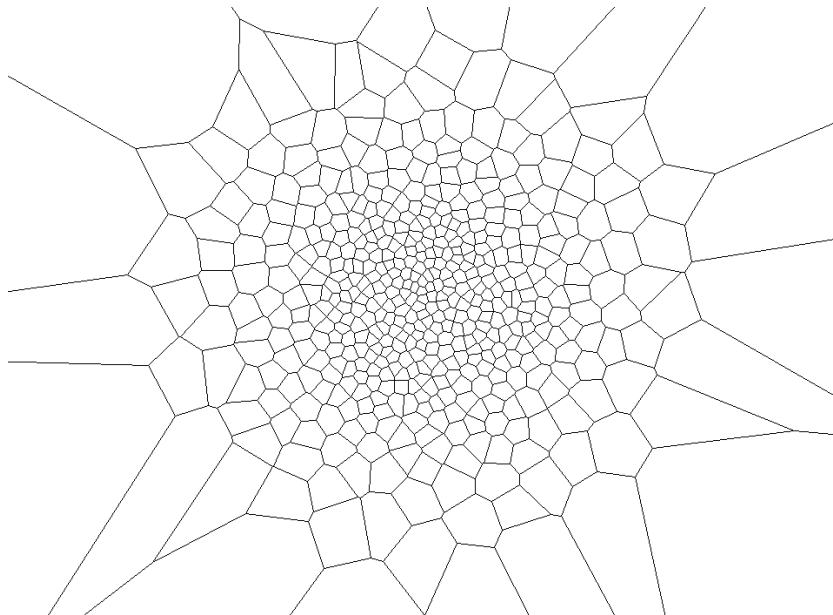


FIGURE 4.2 – Cellules d’une grille optimale pour la loi normale avec $N = 500$.

où les applications $\varphi_0, \dots, \varphi_n$ sont bornées et lipschitziennes. On pose

$$\begin{aligned}\|\varphi\|_{\text{sup}} &= \max_{0 \leq k \leq n} \|\varphi_k\|_{\text{sup}}, \\ [\varphi]_{\text{Lip}} &= \max_{0 \leq k \leq n} [\varphi_k]_{\text{Lip}}.\end{aligned}$$

Nous faisons l'hypothèse de Lipschitz suivante sur les noyaux de transition P_k de la chaîne (X_k) . Pour tout $k \geq 1$, il existe une constante $[P_k]_{\text{Lip}}$ telle que, pour toute application φ lipschitzienne sur \mathbb{R}^d , pour tous $x, \hat{x} \in \mathbb{R}^d$,

$$\left| \int_{\mathbb{R}^d} \varphi(y) P_k(x, dy) - \int_{\mathbb{R}^d} \varphi(y) P_k(\hat{x}, dy) \right| \leq [P_k]_{\text{Lip}} [\varphi]_{\text{Lip}} |x - \hat{x}|. \quad (4.6)$$

Comme précédemment, on note $[P]_{\text{Lip}} = \max_{1 \leq k \leq n} [P_k]_{\text{Lip}}$. On a alors le résultat suivant.

PROPOSITION 4.7 *Soient $p \geq 1$, $n \in \mathbb{N}^*$, $\varphi_0, \dots, \varphi_n$ des applications lipschitziennes bornées. Sous l'hypothèse (4.6) et lorsque $[\varphi]_{\text{Lip}}$ et $\|\varphi\|_{\text{sup}}$ sont inférieurs à 1, on a la majoration de l'erreur de quantification des lois marginales,*

$$\begin{aligned}|\mathbb{E}[\varphi_0(X_0) \dots \varphi_n(X_n)] - \mathbb{E}[\varphi_0(\hat{X}_0^c) \dots \varphi_n(\hat{X}_n^c)]| \\ \leq \sum_{k=0}^n \left(1 + 2 \frac{1 - [P]_{\text{Lip}}^{n-k+1}}{1 - [P]_{\text{Lip}}} \right) \|X_k - \hat{X}_k\|_p.\end{aligned}$$

4.2.2.2 Quantification markovienne

Cette fois, on considère une chaîne de Markov (X_k) sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^d dont les transitions sont données, pour tout $k \geq 1$, par

$$X_k = F_k(X_{k-1}, \varepsilon_k),$$

où ε_k est indépendant de X_0, \dots, X_{k-1} . On fixe un entier n et pour tout $1 \leq k \leq n$, on dispose d'une N_k -grille γ_k de points de \mathbb{R}^d . On définit le processus quantifié (\hat{X}_k) par

$$\begin{aligned}\hat{X}_0 &= \text{Proj}_{\gamma_0}(X_0), \\ \hat{X}_k &= \text{Proj}_{\gamma_k}(F_k(\hat{X}_{k-1}, \varepsilon_k)),\end{aligned}$$

pour tout $k \geq 1$. On souhaite comme dans la partie précédente évaluer la quantité

$$\left| \mathbb{E}[\varphi_0(X_0) \dots \varphi_n(X_n)] - \mathbb{E}[\varphi_0(\widehat{X}_0) \dots \varphi_n(\widehat{X}_n)] \right|,$$

où les φ_i sont lipschitziennes et bornées. Nous faisons également l'hypothèse de Lipschitz supplémentaire sur les transitions F_k . Pour tout $k \geq 1$, il existe une constante $[F_k]_{\text{Lip}}$ telle que, pour tous $x, \widehat{x} \in \mathbb{R}^d$,

$$\mathbb{E} [|F_k(x, \varepsilon_k) - F_k(\widehat{x}, \varepsilon_k)|] \leq [F_k]_{\text{Lip}} |x - \widehat{x}|. \quad (4.7)$$

On note $[F]_{\text{Lip}} = \max_{1 \leq k \leq n} [F_k]_{\text{Lip}}$. On a la proposition qui suit.

PROPOSITION 4.8 *Soient $p \geq 1$, $n \in \mathbb{N}^*$, $\varphi_0, \dots, \varphi_n$ des applications lipschitziennes bornées. Sous l'hypothèse (4.7) et lorsque $[\varphi]_{\text{Lip}}$ et $\|\varphi\|_{\text{sup}}$ sont inférieurs à 1, on a la majoration de l'erreur de quantification markovienne,*

$$\left| \mathbb{E}[\varphi_0(X_0) \dots \varphi_n(X_n)] - \mathbb{E}[\varphi_0(\widehat{X}_0^c) \dots \varphi_n(\widehat{X}_n^c)] \right| \leq \sum_{k=0}^n \frac{1 - [F]_{\text{Lip}}^{n-k+1}}{1 - [F]_{\text{Lip}}} \|\Delta_k\|_1.$$

où $\Delta_k = F_k(\widehat{X}_{k-1}, \varepsilon_k) - \widehat{X}_k$ pour $k \geq 1$, et $\Delta_0 = \widehat{X}_0 - X_0$.

4.3 Étude d'un modèle non paramétrique

On s'intéresse ici au modèle de régression non paramétrique

$$Y = \tilde{f}(U, \varepsilon), \quad (4.8)$$

où U est une covariable aléatoire dans \mathbb{R}^d , ε un terme d'erreur indépendant de U et \tilde{f} une fonction de lien inconnue à valeurs dans \mathbb{R} . Nous proposons d'étudier une approximation de la loi conditionnelle de Y sachant U par une approche basée sur la quantification optimale en norme \mathbb{L}^p des deux variables observables : U et la quantité d'intérêt Y . \widehat{U}^N et \widehat{Y}^N désignent les approximations optimales prenant N valeurs de U et Y . \widehat{P} est la matrice de transition de \widehat{U}^N vers \widehat{Y}^N , *id est*, si γ_N et δ_N sont les N -grilles optimales pour la quantification de U et Y , alors,

$$\forall \widehat{u} \in \gamma_N, \forall \widehat{y} \in \delta_N, \widehat{P}(\widehat{u}, \widehat{y}) = \mathbb{P}(\widehat{Y}^N = \widehat{y} | \widehat{U}^N = \widehat{u}).$$

On considère aussi la variable aléatoire \widehat{Y}^c telle que $(\widehat{U}^N, \widehat{Y}^c)$ est une chaîne de Markov arrêtée dont la matrice de transition est \widehat{P} . Un prédicteur naturel de Y est donc la variable aléatoire discrète \widehat{Y}^c . En effet, nous allons montrer qu'une bonne approximation de la loi de Y sachant U est la loi de \widehat{Y}^c sachant \widehat{U}^N , c'est-à-dire la loi de \widehat{Y}^N sachant \widehat{U}^N . Précisément, pour un u fixé, on approche $Y = \tilde{f}(u, \varepsilon)$ par la loi conditionnelle de \widehat{Y}^N sachant $\widehat{U}^N = \text{Proj}_{\mathcal{Y}^N}(u)$. Nous établissons dans la proposition 4.9 que, pour tout u fixé, la loi discrète donnée par $\widehat{P}(\text{Proj}_{\mathcal{Y}^N}(u), \widehat{y})$ sur les N points \widehat{y} de la grille δ_N , converge vers la loi de Y sachant $U = u$ et nous donnons la vitesse de convergence. De cela on peut déduire une approximation de la moyenne et de la variance conditionnelle de Y sachant $U = u$. On peut aussi obtenir des quantiles conditionnels ainsi qu'un intervalle de prédiction pour Y avec un niveau de confiance donné.

Avant de présenter le résultat qui nous intéresse, on introduit les hypothèses suivantes :

- (\mathcal{H}_1) $\exists p \geq 1$ tel que $U, Y \in \mathbb{L}^p$,
- (\mathcal{H}'_1) $\exists \delta > 0$ tel que $U, Y \in \mathbb{L}^{p+\delta}$,
- (\mathcal{H}_2) $\exists [\tilde{f}]_{\text{Lip}} > 0$ tel que, pour tous $u, v \in \mathbb{R}^d$, $\|\tilde{f}(u, \varepsilon) - \tilde{f}(v, \varepsilon)\|_p \leq [\tilde{f}]_{\text{Lip}}|u - v|$,
- (\mathcal{H}_3) les lois de U et Y ne chargent pas les hyperplans.

Le résultat suivant donne la convergence de la loi conditionnelle de \widehat{Y}^N sachant \widehat{U}^N vers la loi de Y sachant U dans le modèle de régression (4.8). Précisément, pour tout fonction Lipschitz φ , on montre que la norme \mathbb{L}^1 de $\mathbb{E}[\varphi(Y)|U] - \mathbb{E}[\varphi(\widehat{Y}^N)|\widehat{U}^N]$ est bornée par une quantité dans laquelle interviennent les erreurs de quantification des variables U et Y . Grâce à la proposition 4.2, l'erreur de prédiction tend vers 0 quand N tend vers l'infini à la vitesse $N^{-1/d}$.

PROPOSITION 4.9 *Pour toute fonction Lipschitz φ dont la constante de Lipschitz est $[\varphi]_{\text{Lip}}$, sous les Hypothèses $(\mathcal{H}_{1 \rightarrow 3})$, on a*

$$\left\| \mathbb{E}[\varphi(Y)|U] - \mathbb{E}[\varphi(\widehat{Y}^N)|\widehat{U}^N] \right\|_1 \leq 2[\varphi]_{\text{Lip}}[\tilde{f}]_{\text{Lip}}\|U - \widehat{U}^N\|_p + [\varphi]_{\text{Lip}}\|Y - \widehat{Y}^N\|_p.$$

De plus, si l'hypothèse (\mathcal{H}_1) est remplacée par l'hypothèse (\mathcal{H}'_1) , la vitesse de convergence est donnée par

$$\left\| \mathbb{E}[\varphi(Y)|U] - \mathbb{E}[\varphi(\widehat{Y}^N)|\widehat{U}^N] \right\|_1 = \mathcal{O}\left(\frac{1}{N^{1/d}}\right).$$

PREUVE. La démonstration de cette proposition se trouve dans la sous-section 4.7.1. \square

Dans la section qui suit, on propose d’approcher le sous-espace vectoriel engendré par β dans le modèle (4.1) en utilisant la quantification optimale. Puis, en combinant l’approximation obtenue du sous-espace EDR et le résultat précédent dans le cadre non paramétrique, on s’intéresse à la loi conditionnelle de Y sachant $\beta'X$ dans le modèle (4.1).

4.4 Étude du modèle semi-paramétrique

Dans cette partie, on étudie le modèle de régression semi-paramétrique donné par (4.1). Cela revient à considérer le modèle (4.8) avec $U = \beta'X$.

4.4.1 Approximation de la direction de β

Puisque β et f sont simultanément inconnus, seul le sous-espace vectoriel engendré par le paramètre β est identifiable. Nous rappelons que $\widehat{\Gamma}_N$ est la matrice de covariance de l’espérance conditionnelle $\mathbb{E}[\widehat{X}^N | \widehat{Y}]$ où $\widehat{Y} = \text{Proj}_\gamma(Y)$ est la projection sur une grille (pas nécessairement optimale) γ de \mathbb{R} . $\tilde{\beta}_N$ désigne un vecteur principal de la matrice $\Sigma^{-1}\widehat{\Gamma}_N$. Nous montrons dans la proposition 4.11 que, pour N grand, la direction de $\tilde{\beta}_N$ est une bonne approximation de celle de β . En fait, nous donnons même dans le théorème 4.12 un résultat plus fort : il existe une suite (β_N) de vecteurs propres principaux de $\Sigma^{-1}\widehat{\Gamma}_N$ qui converge vers β quand le nombre de quantifieurs de X tend vers l’infini. Ce résultat provient de la convergence de la matrice $\widehat{\Gamma}_N$ vers $\widehat{\Gamma}$ (voir le lemme 4.14 donné dans la sous-section 4.7.2).

Nous avons besoin des hypothèses supplémentaires suivantes, qui sont classiques lorsqu’on travaille sur la méthode SIR :

$$\begin{aligned} (\mathcal{H}_5) \quad & \exists \widehat{y} \in \gamma, \mathbb{E}[(X - \mu)' \beta | \widehat{Y} = \widehat{y}] \neq 0, \\ (\mathcal{H}_6) \quad & X \text{ a une loi à symétrie elliptique.} \end{aligned}$$

On rappelle que la loi d’une variable aléatoire Z est dite à symétrie sphérique lorsque pour toute matrice orthogonale \mathcal{O} , $\mathcal{O}Z$ et Z ont même loi. L’hypothèse (\mathcal{H}_6) signifie que X s’écrit $X = \mu + \mathcal{A}Z$, où Z est à symétrie sphérique et \mathcal{A} est inversible.

REMARQUE 4.10 *Projeter Y sur une seule classe mène à $\mathbb{E}[(X - \mu)' \beta | \widehat{Y} = \widehat{y}] = 0$. L’hypothèse (\mathcal{H}_5) n’est alors jamais vérifiée.*

Les hypothèses suivantes assurent la validité des résultats classiques de quantification :

- (\mathcal{H}_7) $\exists p \geq 1$ tel que $X \in \mathbb{L}^p \cap \mathbb{L}^q$ avec $\frac{1}{p} + \frac{1}{q} = 1$,
- (\mathcal{H}_8) la loi de X ne charge pas les hyperplans,
- (\mathcal{H}_9) $\exists \delta > 0$ tel que $X \in \mathbb{L}^{p+\delta}$.

Pour toute suite $(\tilde{\beta}_N)$ de vecteurs propres principaux de la suite de matrices $(\Sigma^{-1}\hat{\Gamma}_N)$, le résultat suivant donne la convergence de la direction de la suite $(\tilde{\beta}_N)$ vers la direction de β quand le nombre N de quantifieurs tend vers l'infini. Pour cela, nous introduisons une mesure de la colinéarité des deux vecteurs $\tilde{\beta}_N$ et β ,

$$\cos^2(\tilde{\beta}_N, \beta) = \frac{(\tilde{\beta}'_N \beta)^2}{(\tilde{\beta}'_N \tilde{\beta}_N) \times (\beta' \beta)}.$$

PROPOSITION 4.11 *Sous les hypothèses $(\mathcal{H}_{5 \rightarrow 8})$, pour toute suite $(\tilde{\beta}_N)$ de vecteurs propres principaux de la suite $(\Sigma^{-1}\hat{\Gamma}_N)$, on a, quand N tend vers l'infini,*

$$\cos^2(\tilde{\beta}_N, \beta) \rightarrow 1.$$

PREUVE. La preuve de ce résultat se trouve dans la sous-section 4.7.3 □

Avant de présenter le résultat suivant, on rappelle que pour tout $x \in \mathbb{R}^d$, la norme euclidienne de x est notée $|x|$.

PROPOSITION 4.12 *Sous les hypothèses $(\mathcal{H}_{5 \rightarrow 8})$, il existe une suite de vecteurs propres principaux de $\Sigma^{-1}\hat{\Gamma}_N$, notée (β_N) qui converge vers β . En effet, il existe des constantes $C_1, C_2 > 0$ telles que,*

$$\forall N \geq C_2, \quad |\beta_N - \beta| \leq \frac{2d}{C_1} \|\Sigma^{-1}\|_\infty \|X - \hat{X}^N\|_p \|X\|_q.$$

De plus, sous l'hypothèse supplémentaire (\mathcal{H}_9) , on contrôle la vitesse de convergence,

$$|\beta_N - \beta| = \mathcal{O}\left(\frac{1}{N^{1/d}}\right).$$

PREUVE. La preuve de ce résultat se trouve dans la sous-section 4.7.2 □

4.4.2 Prédiction dans le cadre semi-paramétrique

Maintenant, nous utilisons les résultats d'approximation de β et ceux établis dans le cadre non paramétrique afin de prédire Y dans le modèle (4.1). Pour un β_N fixé, défini dans la section 4.4.1, le modèle

$$Y = f(\beta'_N X, \varepsilon)$$

est une bonne approximation du modèle initial. Par conséquent, la méthode utilisée pour étudier la loi de Y dans le modèle non paramétrique (4.8) quand $U = \beta'_N X$ peut permettre d'obtenir une bonne prédiction de Y dans le modèle (4.1). Pour présenter les résultats asymptotiques, nous considérons deux paramètres : le nombre N de quantifieurs de X dont dépend l'approximation β_N de β (voir la sous-section 4.4.1) et le nombre m de quantifieurs des variables $U = \beta'_N X$ et Y (voir la section 4.3). Nous allons montrer que l'erreur d'approximation de la loi de Y sachant $\beta'X$ tend vers 0 quand m et N tendent simultanément vers l'infini.

D'après la proposition 4.12, il existe des constantes $\Delta_0, C_0 > 0$ telles que, pour tout $N \geq C_0$,

$$|\beta_N - \beta| \leq \Delta_0 \|X - \widehat{X}^N\|_p. \quad (4.9)$$

Nous avons besoin d'hypothèses supplémentaires pour établir notre résultat asymptotique. Nous supposons que f est Lipschitz et que la quantification optimale en norme \mathbb{L}^p de la variable Y est possible :

- (\mathcal{H}_{10}) $\exists [f]_{\text{Lip}} > 0$ tel que, pour tous $u, v \in \mathbb{R}$, $\|f(u, \varepsilon) - f(v, \varepsilon)\|_p \leq [f]_{\text{Lip}} |u - v|$,
- (\mathcal{H}_{11}) $Y \in \mathbb{L}^{p+\delta}$,
- (\mathcal{H}_{12}) la loi de Y ne charge pas les hyperplans.

On prédit Y sachant X par la variable aléatoire \widehat{Y}^c définie telle que $(\widehat{\beta}'_N X^m, \widehat{Y}^c)$ est une chaîne de Markov arrêtée avec le même noyau de transition que la chaîne arrêtée $(\widehat{\beta}'_N X^m, \widehat{Y}^m)$ où $\widehat{\beta}'_N X^m$ et \widehat{Y}^m sont les approximations discrètes optimales (en norme \mathbb{L}^p) de $\beta'_N X$ et Y avec m quantifieurs.

THÉORÈME 4.13 *Sous les hypothèses ($\mathcal{H}_{5 \rightarrow 12}$), pour toute fonction Lipschitz φ , il existe des nombres réels A_1, A_2, A_3 , une suite (g_N) qui admet une limite strictement positive et deux entiers \bar{m} et \bar{N} tels que pour tous $m \geq \bar{m}$ et $N \geq \bar{N}$, on a*

$$\left\| \mathbb{E}[\varphi(Y) | \beta'X] - \mathbb{E}[\varphi(\widehat{Y}^m) | \widehat{\beta}'_N X^m] \right\|_1 \leq \frac{A_1}{N^{1/d}} + \frac{A_2}{m} g_N + \frac{A_3}{m}.$$

PREUVE. La preuve de ce résultat se trouve dans la sous-section 4.7.4 □

Un corollaire immédiat de ce résultat est que l'erreur de prédiction tend vers 0 quand N et m tendent vers l'infini.

4.5 Simulations

Dans l'étude de simulation qui suit, on s'intéresse à la fois à l'approximation du paramètre vectoriel β et à l'approximation de la loi de Y sachant $X = x$ dans le modèle de régression semi-paramétrique (4.1). Dans la première partie, seule la direction de β est identifiable et on évalue la qualité de notre procédure en étudiant le carré du cosinus de l'angle entre la vraie direction de β et son approximation. L'approximation est de bonne qualité lorsque la mesure de la colinéarité est proche de 1. Dans la seconde partie, on se focalise sur l'approximation de la loi de Y sachant X . On compare l'espérance et la variance conditionnelles à leur approximation en utilisant notre procédure. Pour cela, on suppose que le signe et la norme de β sont connus ; le paramètre β est alors identifiable.

Dans cette étude, on considère les trois modèles de régression suivants :

$$\begin{aligned}(\mathcal{M}_1) \quad Y &= (\beta'X)^3 + \varepsilon, \\(\mathcal{M}_2) \quad Y &= (\beta'X)^3 + \beta'X\varepsilon, \\(\mathcal{M}_3) \quad Y &= (\beta'X)^2 \exp\left(\frac{\beta'X}{\theta}\right) + \varepsilon,\end{aligned}$$

où X suit la loi normale de dimension d $\mathcal{N}(0, I_d)$ et ε est un bruit gaussien centré réduit, indépendant de la covariable X .

Ces modèles sont classiques lorsqu'on étudie la qualité d'une méthode d'estimation semi-paramétrique de type SIR. Dans la première (respectivement la seconde) partie de l'étude, la dimension de X est $d = 10$ (respectivement $d = 4$) avec $\beta = (1, -1, 0, \dots, 0)'$. Le premier et le troisième modèle sont homoscédastiques (la variance de l'erreur ne dépend pas de la covariable X) tandis que le second est hétéroscédastique (la variance de l'erreur dépend de X). Le paramètre θ qui intervient dans le modèle (\mathcal{M}_3) nous permet de tester la performance de notre méthode d'approximation lorsqu'une dépendance symétrique entre Y et $\beta'X$ intervient. Lorsque $\theta = 1$, il n'y a pas de dépendance symétrique entre la variable d'intérêt et la covariable et la méthode SIR classique fonctionne bien. La dépendance symétrique apparaît lorsque θ augmente. On la qualifie de modérée lorsque $\theta = 5$ et de forte lorsque $\theta = 10$. De tels cas sont considérés comme pathologiques pour la méthode

SIR ; elle ne parvient en général pas à bien estimer la direction de β . Nous verrons que notre approche apparaît comme robuste en présence de tels modèles à dépendance symétrique.

4.5.1 Approximation du sous-espace engendré par β

On suppose qu'on a un échantillon $(X_i, Y_i)_{1 \leq i \leq n}$ de variables aléatoires générées selon l'un des trois modèles de régression précédents. La matrice de covariance Σ de X est estimée par sa version empirique notée S . Les variables X et Y sont quantifiées en norme \mathbb{L}^2 par l'algorithme usuel donné dans la partie 4.2. Ainsi, on obtient les versions discrètes \widehat{X}^N et \widehat{Y}^m de X et Y . Enfin, on calcule la matrice de covariance $\widehat{\Gamma}_N = \text{Var}(\mathbb{E}[\widehat{X}^N | \widehat{Y}^m])$.

Dans la suite, on travaille avec deux tailles d'échantillon : $n = 300$ et $n = 1000$. La taille d'échantillon n est petite pour obtenir des grilles de quantification de bonne qualité à partir de l'algorithme de gradient utilisé. Afin de pallier cet éventuel problème, on propose d'utiliser une idée introduite par Saracco dans [78]. On quantifie B fois les variables X et Y à partir du même échantillon. On obtient B estimations T_1, \dots, T_B de $\widehat{\Gamma}_N$, et on travaille avec leur moyenne

$$\bar{T} = \frac{1}{B} \sum_{b=1}^B T_b.$$

Notons que, puisque le sous-espace propre principal de chaque matrice $\Sigma^{-1}T_b$ est colinéaire à β , alors le sous-espace propre principal de $\Sigma^{-1}\bar{T}$ est aussi colinéaire à β . On note $\widehat{\beta}_N$ un vecteur propre principal de $S^{-1}\bar{T}$ et on utilise ce vecteur pour approcher la direction de β . Dans les résultats graphiques présentés plus bas, le nombre de grilles de quantification est $B = 5$, le nombre de quantifieurs pour Y est $m = 5$ et le nombre de quantifieurs pour X est $N \in \{20, 30, 50, 100, 200\}$.

On génère 100 échantillons pour chaque modèle et chaque taille d'échantillon. Pour chaque échantillon simulé, on calcule $\widehat{\beta}_N$ ainsi que la mesure de qualité de l'approximation $\cos^2(\beta, \widehat{\beta}_N)$. On représente dans la figure 4.3 (respectivement la figure 4.4) les diagrammes en boîte des cosinus carrés en fonction du nombre N de quantifieurs pour X pour les différents modèles quand $n = 300$ (respectivement quand $n = 1000$). On compare aussi notre méthode d'estimation avec la méthode SIR classique quand le nombre H de tranches est égal à 5 (c'est-à-dire le nombre m de quantifieurs pour Y dans notre procédure). On observe clairement que, quand le nombre N de quantifieurs, augmente, la qualité de notre approximation augmente également pour chacun des modèles. Bien sûr, on observe la même dynamique quand la taille d'échantillon n grandit. On peut voir que la méthode SIR classique est plus performante que notre procédure d'approximation pour les modèles

(\mathcal{M}_1) , (\mathcal{M}_2) et (\mathcal{M}_3) avec $\theta = 1$, qui sont les modèles favorables à une approche SIR classique. Cependant, les performances numériques obtenues sont proches de celles obtenues par l'approche SIR usuelle lorsque $N = 200$ et $n = 300$ ou $n = 1000$. Quand une dépendance symétrique apparaît – lorsque $\theta = 5$ ou $\theta = 10$ pour le modèle (\mathcal{M}_3) , – notre approche permet d'obtenir des estimations de bien meilleure qualité que la méthode SIR classique, surtout lorsque $n = 1000$. En pratique, à partir d'un jeu de données réel, il est impossible de savoir si la variable d'intérêt Y dépend de manière symétrique ou non de la covariable X . Dans ce cas, il est raisonnable de penser que notre procédure d'estimation de la direction de β est plus robuste que la méthode SIR.

4.5.2 Approximation de la loi de Y sachant $\beta'X$

Pour une bonne prédiction, nous avons besoin d'échantillons plus grands que ceux utilisés dans la partie précédente. En effet, l'estimation du sous-espace EDR ne nécessite pas le calcul des probabilités associées aux points des grilles, mais seulement la position de ces points. Pour la prédiction, nous avons à estimer la loi jointe de $(\widehat{\beta'_N X^m}, \widehat{Y^m})$. Nous considérons des échantillons de taille $n = 10\,000$ générés à partir des modèles de régression semi-paramétrique présentés précédemment où la dimension de la covariable X est $d = 4$.

Pour chaque échantillon simulé, on propose d'abord d'approcher β par la procédure décrite dans la sous-section précédente. Néanmoins, on utilise une seule quantification ($B = 1$) des variables X et Y car nous travaillons ici avec suffisamment de données pour obtenir une grille de quantification optimale satisfaisante. Pour cela, nous utilisons $N = 200$ quantifieurs pour X et $m = 5$ quantifieurs pour Y . Ici, nous avons besoin d'une approximation β_N de β et pas seulement de sa direction. Dans la suite, nous supposons donc que le signe et la norme du vecteur β sont connus.

Finalement, afin d'approcher la loi conditionnelle de Y sachant X , on discrétise la variable $\beta'_N X$ avec $m = 100$ quantifieurs, et on utilise aussi $m = 100$ quantifieurs pour la discrétisation de la variable Y à cette étape. On obtient ainsi la loi de $\widehat{Y^m}$ sachant $\widehat{\beta'_N X^m}$. Nous pouvons donc approcher la moyenne et la variance conditionnelles de Y sachant X par celles de $\widehat{Y^m}$ sachant $\widehat{\beta'_N X^m}$.

Dans le tableau 4.1 (respectivement 4.2), on présente quelques résultats d'approximation de la moyenne et de la variance conditionnelles de Y sachant $X = (0.5, -0.5, 1, 0)'$ (respectivement sachant $X = (-1/3, 0.5, 1, 1)'$). Pour chaque modèle, on compare les approximations obtenues aux vraies valeurs et on évalue l'erreur relative correspondante. On peut voir que les valeurs approchées obtenues avec notre méthode sont très proches des

valeurs théoriques aussi bien pour la moyenne et que pour la variance conditionnelle, et ce pour tous les modèles. Les erreurs relatives (en valeur absolue) sont inférieures à 16% dans la plupart des cas considérés.

Finalement, nous simulons 10 valeurs x_j indépendantes et de loi uniforme sur l'hypercube $[-2, 2]^4$. Pour chaque valeur simulée et pour chaque modèle, on approche les moyenne et variance conditionnelles de Y sachant $X = x_j$. La figure 4.6 donne les diagrammes en boîtes de l'erreur relative de ces moments conditionnels pour le modèle (\mathcal{M}_3) avec $\theta = 5$. On peut voir que notre procédure d'approximation permet d'obtenir des erreurs relatives raisonnables. On obtient des résultats similaires pour les autres modèles.

Dans cette partie, on ne s'est intéressé qu'à l'approximation des deux premiers moments conditionnels de Y sachant X . Cependant, il est possible – avec l'approche proposée, – d'obtenir également la médiane conditionnelle ou des quantiles conditionnels. Ces derniers permettent notamment de construire un intervalle de prédiction pour Y avec un niveau de confiance choisi.

4.6 Conclusion du chapitre

Dans ce chapitre, on a présenté une méthode d'approximation, basée sur la quantification optimale, du sous-espace EDR et de la loi conditionnelle de la variable d'intérêt Y sachant le prédicteur X dans le modèle (4.1). On a démontré la convergence de nos approximations et donné les vitesses de convergence. A notre connaissance, l'utilisation de la quantification optimale pour traiter des problèmes issus de la statistique non paramétrique ou semi-paramétrique est nouvelle. Les simulations présentées montrent que les résultats numériques obtenus par cette méthode sont comparables à ceux obtenus par l'approche SIR classique, et même meilleurs lorsqu'une dépendance symétrique entre Y et X apparaît. En pratique, sur des jeux de données réelles, notre procédure est donc vraisemblablement plus robuste que la méthode SIR usuelle.

Comme on l'a vu dans la sous-section 4.5.2, la taille requise de l'échantillon pour l'étape de prédiction est certainement l'inconvénient majeur de notre méthode. Notons que c'est aussi le cas pour toutes les méthodes numériques utilisant l'outil quantification optimale. Dans une optique applicative, il serait judicieux de développer un outil pour choisir les nombre optimaux de quantifieurs N et m en fonction du nombre n de données disponibles. Les résultats partiels présentés dans [70] montrent que la vitesse de convergence de l'algorithme de gradient décroît quand le nombre de quantifieurs et la dimension de la variable

augmentent. Le choix est donc un compromis entre la finesse de la grille obtenue et la qualité des estimations des poids des cellules.

D'un point de vue numérique, on peut déterminer le nombre optimal de quantifieurs N pour approcher β_N de la façon suivante : à partir d'un échantillon de taille n , on génère \mathcal{B} répliques bootstrap. D'après [43], un estimateur bootstrap naïf de la qualité de l'approximation $\cos^2(\beta_N, \beta)$ peut être défini par

$$Q(N) = \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \cos^2(\beta_N^{(b)}, \beta_N),$$

où $\beta_N^{(b)}$ est l'approximation du sous-espace EDR basée sur le $b^{\text{ème}}$ échantillon bootstrap et utilisant N quantifieurs. En pratique, on calcule ce critère entre deux valeurs N_{\min} et N_{\max} afin de choisir N . Finalement, le nombre optimal N de quantifieurs pourrait être $\hat{N} = \arg \max_N Q(N)$.

Les résultats sur l'approximation de la loi conditionnelle de Y sachant $\beta'X$ par quantification seront sans aucun doute affinés et complétés par Isabelle Charlier au cours de sa thèse sur la régression quantile.

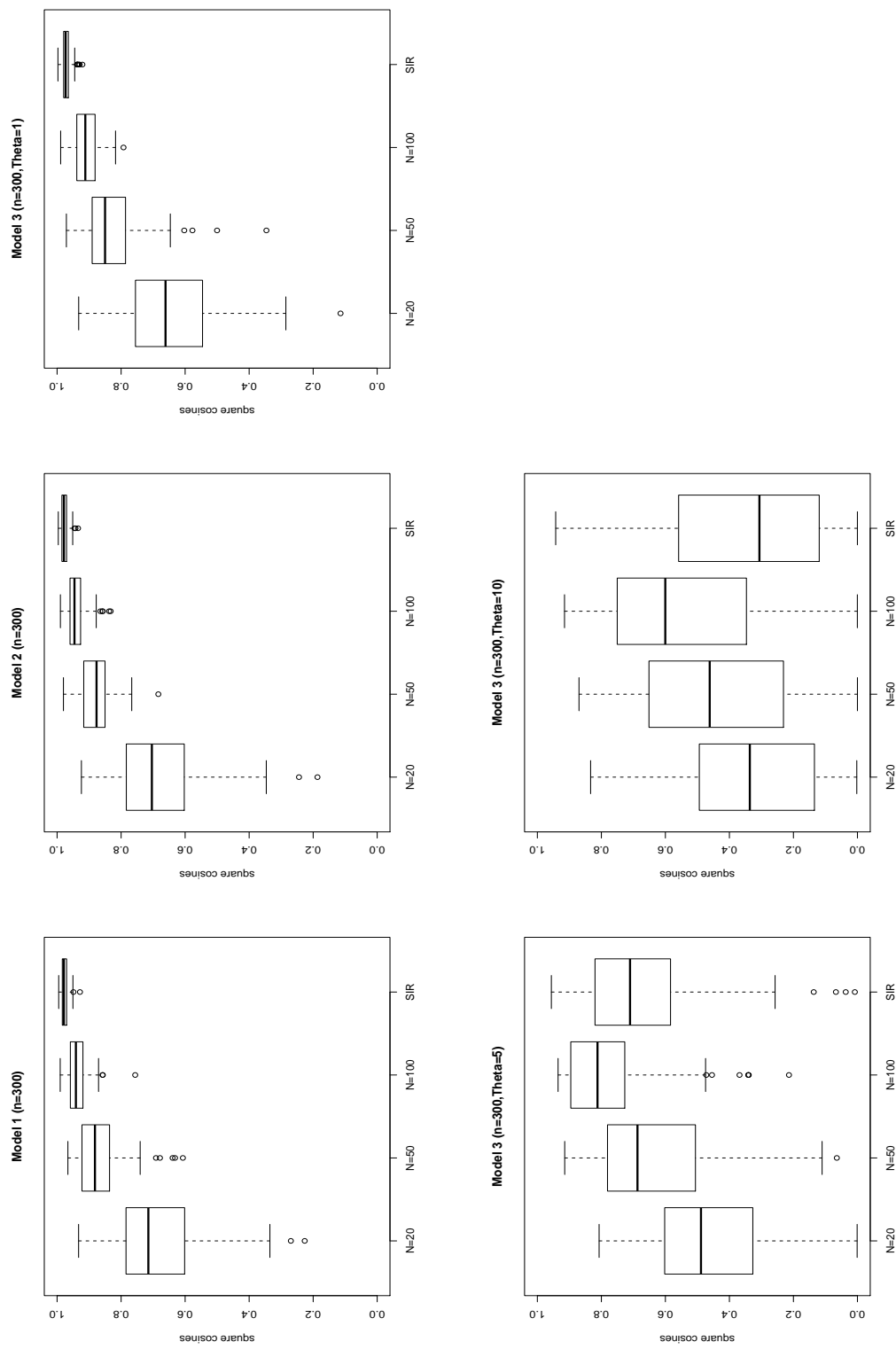


FIGURE 4.3 – Diagrammes en boîte des cosinus carrés pour l’approximation du paramètre β avec $n = 300$.

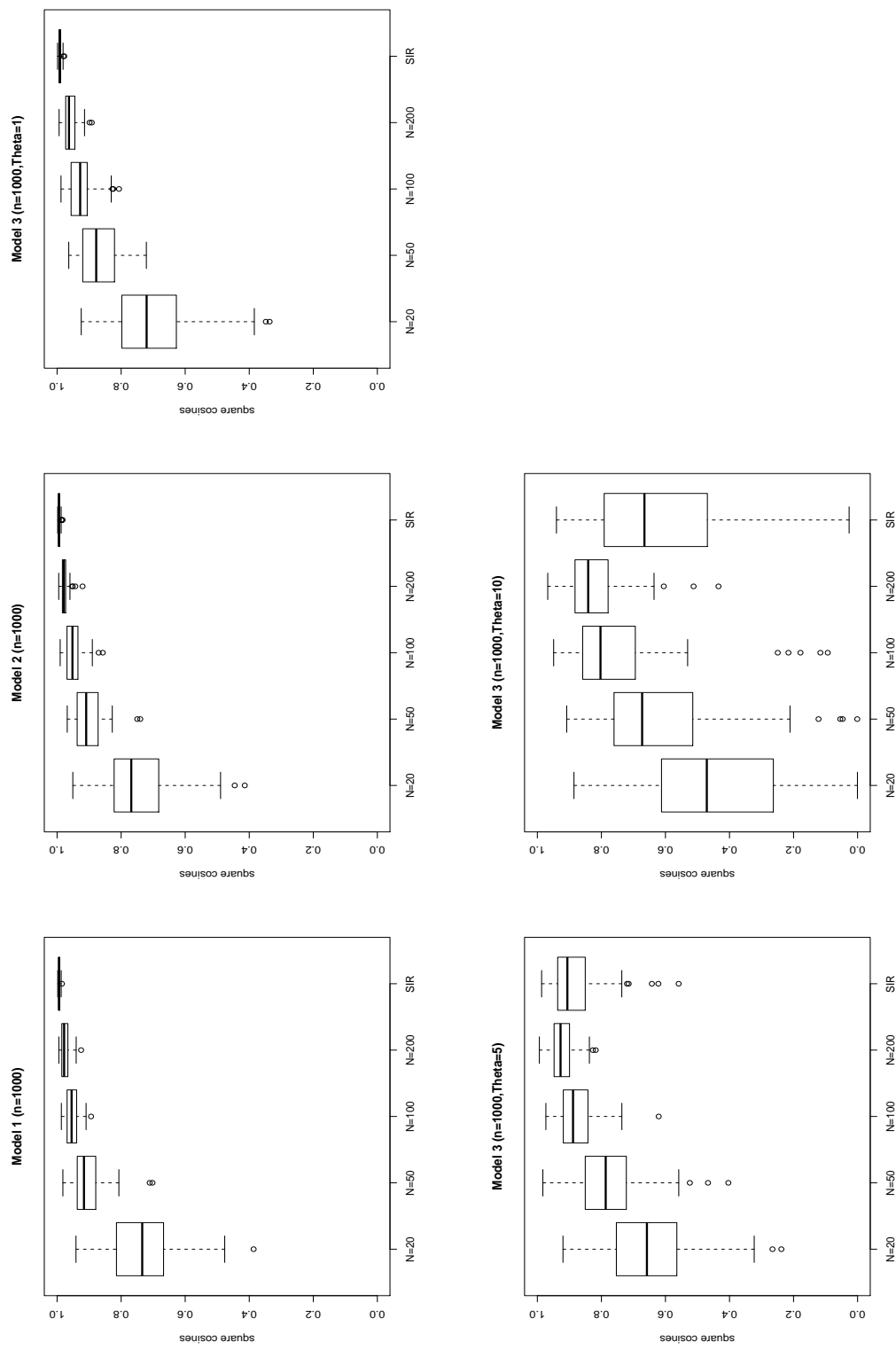


FIGURE 4.4 – Diagrammes en boîte des cosinus carrés pour l'approximation du paramètre β avec $n = 1000$.

ESPÉRANCE CONDITIONNELLE					
modèle	(\mathcal{M}_1)	(\mathcal{M}_2)	(\mathcal{M}_3)	(\mathcal{M}_3)	(\mathcal{M}_3)
			avec $\theta = 1$	avec $\theta = 5$	avec $\theta = 10$
vraie valeur	1	1	2.72	1.22	1.11
approximation	1.07	0.84	2.56	1.02	1.11
erreur relative	0.07	-0.16	-0.06	-0.16	0.00

VARIANCE CONDITIONNELLE					
modèle	(\mathcal{M}_1)	(\mathcal{M}_2)	(\mathcal{M}_3)	(\mathcal{M}_3)	(\mathcal{M}_3)
			avec $\theta = 1$	avec $\theta = 5$	avec $\theta = 10$
vraie valeur	1	1	1	1	1
approximation	1.01	0.92	1.14	1.05	1.07
erreur relative	0.01	-0.08	0.14	0.05	0.07

TABLE 4.1 – Approximation de la loi de Y sachant $X = (0.5, -0.5, 1, 0)'$.

ESPÉRANCE CONDITIONNELLE					
modèle	(\mathcal{M}_1)	(\mathcal{M}_2)	(\mathcal{M}_3)	(\mathcal{M}_3)	(\mathcal{M}_3)
			avec $\theta = 1$	avec $\theta = 5$	avec $\theta = 10$
vraie valeur	-0.56	-0.58	0.30	0.59	0.64
approximation	-0.64	-0.51	0.52	0.80	0.61
erreur relative	0.11	-0.12	0.71	0.36	-0.05

VARIANCE CONDITIONNELLE					
modèle	(\mathcal{M}_1)	(\mathcal{M}_2)	(\mathcal{M}_3)	(\mathcal{M}_3)	(\mathcal{M}_3)
			avec $\theta = 1$	avec $\theta = 5$	avec $\theta = 10$
vraie valeur	1	0.69	1	1	1
approximation	1.01	0.80	0.77	1.01	1.09
erreur relative	0.01	0.15	-0.23	0.01	0.09

TABLE 4.2 – Approximation de la loi de Y sachant $X = (-1/3, 0.5, 1, 1)'$.

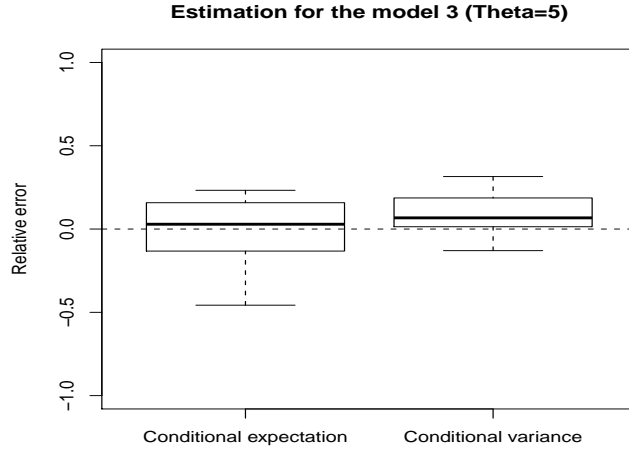


FIGURE 4.5 – Erreurs relatives pour l’approximation de la moyenne et de la variance conditionnelles.

4.7 Preuves des résultats des sections 4.3 et 4.4

4.7.1 Preuve de la proposition 4.9

Définissons $F(\mathbf{u}) = \mathbb{E}[\varphi(Y)|\mathbf{U} = \mathbf{u}]$ et $\hat{F}(\hat{\mathbf{u}}) = \mathbb{E}[\varphi(\hat{Y})|\hat{\mathbf{U}} = \hat{\mathbf{u}}]$. Par l’inégalité triangulaire, on a

$$\begin{aligned} \|F(\mathbf{U}) - \hat{F}(\hat{\mathbf{U}})\|_p &\leq \|F(\mathbf{U}) - F(\hat{\mathbf{U}})\|_p + \|F(\hat{\mathbf{U}}) - \mathbb{E}[\varphi(Y)|\hat{\mathbf{U}}]\|_p \\ &\quad + \|\mathbb{E}[\varphi(Y)|\hat{\mathbf{U}}] - \hat{F}(\hat{\mathbf{U}})\|_p. \end{aligned}$$

De plus, en utilisant la propriété de Lipschitz de φ et de \tilde{f} , et par indépendance de \mathbf{U} et ε , on a,

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d, \quad |F(\mathbf{u}) - F(\mathbf{v})| \leq [\varphi]_{\text{Lip}}[\tilde{f}]_{\text{Lip}}|\mathbf{u} - \mathbf{v}|.$$

Alors, par la propriété de contraction dans \mathbb{L}^p de l’espérance conditionnelle, on obtient

$$\|F(\hat{\mathbf{U}}) - \mathbb{E}[\varphi(Y)|\hat{\mathbf{U}}]\|_p \leq \|F(\mathbf{U}) - F(\hat{\mathbf{U}})\|_p \leq [\varphi]_{\text{Lip}}[\tilde{f}]_{\text{Lip}}\|\mathbf{U} - \hat{\mathbf{U}}\|_p.$$

De plus, puisque $\hat{F}(\hat{\mathbf{U}}) - \mathbb{E}[\varphi(Y)|\hat{\mathbf{U}}] = \mathbb{E}[\varphi(\hat{Y}) - \varphi(Y)|\hat{\mathbf{U}}]$, et en utilisant à nouveau la contraction dans \mathbb{L}^p de l’espérance conditionnelle et la propriété de Lipschitz de φ , on a

$$\|\hat{F}(\hat{\mathbf{U}}) - \mathbb{E}[\varphi(Y)|\hat{\mathbf{U}}]\|_p \leq \|\varphi(Y) - \varphi(\hat{Y})\|_p \leq [\varphi]_{\text{Lip}}\|Y - \hat{Y}\|_p.$$

D'où l'inégalité qu'on voulait montrer.

4.7.2 Preuve de la proposition 4.12

Avant de montrer la proposition 4.12, on donne d'abord, dans le lemme 4.14, une borne sur la distance entre $\widehat{\Gamma}$ et $\widehat{\Gamma}_N$.

LEMME 4.14 *Sous les hypothèses $(\mathcal{H}_{5 \rightarrow 10})$, on a $\|\widehat{\Gamma} - \widehat{\Gamma}_N\|_\infty \leq 2d\|X - \widehat{X}^N\|_p\|X\|_q$.*

PREUVE. Soit $A_N = \widehat{\Gamma} - \widehat{\Gamma}_N$. On contrôle tous les termes de cette matrice symétrique comme suit.

– ETUDE DES TERMES DIAGONAUX. Pour tout $1 \leq i \leq d$, on a

$$\begin{aligned} A_N(i, i) &= \text{Var}[\mathbb{E}[X|\widehat{Y}]_i] - \text{Var}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_i] \\ &= \mathbb{E}[\mathbb{E}[X|\widehat{Y}]_i^2] - \mathbb{E}[X]_i^2 - \mathbb{E}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_i^2] + \mathbb{E}[\widehat{X}^N]_i^2. \end{aligned}$$

En utilisant la propriété de stationnarité (4.4), on obtient

$$\begin{aligned} A_N(i, i) &= \mathbb{E}[\mathbb{E}[X|\widehat{Y}]_i^2] - \mathbb{E}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_i^2] \\ &= \mathbb{E}\left[\left(\mathbb{E}[X|\widehat{Y}]_i - \mathbb{E}[\widehat{X}^N|\widehat{Y}]_i\right)\left(\mathbb{E}[X|\widehat{Y}]_i + \mathbb{E}[\widehat{X}^N|\widehat{Y}]_i\right)\right]. \end{aligned}$$

Par l'inégalité de Hölder, la contraction dans \mathbb{L}^p et dans \mathbb{L}^q de l'espérance conditionnelle et à nouveau (4.4), on a

$$\begin{aligned} |A_N(i, i)| &\leq \|\mathbb{E}[X - \widehat{X}^N|\widehat{Y}]_i\|_p \|\mathbb{E}[X + \widehat{X}^N|\widehat{Y}]_i\|_q \\ &\leq \|X - \widehat{X}^N\|_p \|X + \widehat{X}^N\|_q \\ &\leq \|X - \widehat{X}^N\|_p (\|X\|_q + \|\widehat{X}^N\|_q) \\ &\leq 2\|X - \widehat{X}^N\|_p \|X\|_q. \end{aligned}$$

– ETUDE DES TERMES NON DIAGONAUX. Pour tout couple $(i, j) \in \{1, \dots, d\}^2$ tel que $i \neq j$, on a

$$\begin{aligned} A_N(i, j) &= \text{Cov}[\mathbb{E}[X|\widehat{Y}]_i, \mathbb{E}[X|\widehat{Y}]_j] - \text{Cov}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_i, \mathbb{E}[\widehat{X}^N|\widehat{Y}]_j] \\ &= \mathbb{E}[\mathbb{E}[X|\widehat{Y}]_i \mathbb{E}[X|\widehat{Y}]_j] - \mathbb{E}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_i \mathbb{E}[\widehat{X}^N|\widehat{Y}]_j] \\ &= \mathbb{E}\left[\left(\mathbb{E}[X|\widehat{Y}]_i - \mathbb{E}[\widehat{X}^N|\widehat{Y}]_i\right)\left(\mathbb{E}[X|\widehat{Y}]_j + \mathbb{E}[\widehat{X}^N|\widehat{Y}]_j\right)\right] \\ &\quad + \mathbb{E}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_i \mathbb{E}[X|\widehat{Y}]_j] - \mathbb{E}[\mathbb{E}[\widehat{X}^N|\widehat{Y}]_j \mathbb{E}[X|\widehat{Y}]_i]. \end{aligned}$$

Par symétrie, on a aussi

$$A_N(i, j) = \mathbb{E} \left[\left(\mathbb{E}[X|\hat{Y}]_j - \mathbb{E}[\hat{X}^N|\hat{Y}]_j \right) \left(\mathbb{E}[X|\hat{Y}]_i + \mathbb{E}[\hat{X}^N|\hat{Y}]_i \right) \right] \\ + \mathbb{E} \left[\mathbb{E}[\hat{X}^N|\hat{Y}]_j \mathbb{E}[X|\hat{Y}]_i - \mathbb{E}[\hat{X}^N|\hat{Y}]_i \mathbb{E}[X|\hat{Y}]_j \right].$$

Alors, en sommant les deux égalités précédentes, on obtient

$$2A_N(i, j) = \mathbb{E} \left[\left(\mathbb{E}[X|\hat{Y}]_i - \mathbb{E}[\hat{X}^N|\hat{Y}]_i \right) \left(\mathbb{E}[X|\hat{Y}]_j + \mathbb{E}[\hat{X}^N|\hat{Y}]_j \right) \right] \\ + \mathbb{E} \left[\left(\mathbb{E}[X|\hat{Y}]_j - \mathbb{E}[\hat{X}^N|\hat{Y}]_j \right) \left(\mathbb{E}[X|\hat{Y}]_i + \mathbb{E}[\hat{X}^N|\hat{Y}]_i \right) \right]$$

Comme pour les termes diagonaux, en utilisant l'inégalité de Hölder et la contraction dans \mathbb{L}^p et \mathbb{L}^q de l'espérance conditionnelle, on obtient

$$|A_N(i, j)| \leq 2\|X - \hat{X}^N\|_p \|X\|_q.$$

Finalement, on a

$$\|A_N\|_\infty = \max_{1 \leq i \leq d} \sum_{j=1}^d |A_N(i, j)| \leq 2d\|X - \hat{X}^N\|_p \|X\|_q,$$

ce qui termine la preuve du lemme. □

Une conséquence immédiate du lemme 4.14 est que les valeurs propres de $\Sigma^{-1}\hat{\Gamma}_N$ convergent vers les valeurs propres de $\Sigma^{-1}\hat{\Gamma}$ quand N tend vers l'infini. En particulier, si $\hat{\rho}$ et $\hat{\rho}_N$ désignent les plus grandes valeurs propres de $\Sigma^{-1}\hat{\Gamma}$ et $\Sigma^{-1}\hat{\Gamma}_N$, on a, quand N tend vers l'infini,

$$\hat{\rho}_N \rightarrow \hat{\rho} > 0.$$

On considère z_N le vecteur propre principal de norme 1 de $\Sigma^{-1}\hat{\Gamma}_N$ qui s'écrit sous la forme $z_N = \alpha_N \beta + z_N^\perp$ où $z_N^\perp \perp \beta$ et $\alpha_N \geq 0$. Alors, on a

$$\hat{\rho}_N z_N = \Sigma^{-1}\hat{\Gamma}_N z_N \\ = \Sigma^{-1} \left[(\hat{\Gamma}_N - \hat{\Gamma})z_N + \hat{\Gamma}(\alpha_N \beta + z_N^\perp) \right].$$

De la théorie sur la méthode SIR, le rang de $\hat{\Gamma}$ est 1 et β est un vecteur propre principal de $\Sigma^{-1}\hat{\Gamma}$. Donc le noyau de la matrice $\Sigma^{-1}\hat{\Gamma}$ est de dimension $d - 1$ et contient le vecteur z_N^\perp . Par conséquent, on a $\hat{\rho}_N z_N - \alpha_N \hat{\rho} \beta = \Sigma^{-1}(\hat{\Gamma}_N - \hat{\Gamma})z_N$. Maintenant, on montre que la

suite $(\alpha_N)_{N \geq 1}$ a une limite strictement positive. Comme z_N est de norme 1 pour tout N , on a

$$|\widehat{\rho}_N z_N - \alpha_N \widehat{\rho} \beta| \leq \|\Sigma^{-1}\|_\infty \|\widehat{\Gamma}_N - \widehat{\Gamma}\|_\infty \rightarrow 0,$$

quand N tend vers l'infini. Puisque $(\widehat{\rho}_N)$ a une limite non nulle et puisque z_N est de norme 1 pour tout N , on a $\lim_{N \rightarrow +\infty} \alpha_N = \frac{1}{|\beta|}$. Soit N_1 un entier tel que pour tout $N \geq N_1$, $\alpha_N > 0$.

Soit $C = \min_{N \geq N_1} \alpha_N$ et, pour tout $N \geq N_1$, soit $\beta_N = \frac{z_N \widehat{\rho}_N}{\alpha_N \widehat{\rho}}$. Alors, pour tout $N \geq N_1$, on a

$$|\beta_N - \beta| \leq \frac{1}{C} \|\Sigma^{-1}\|_\infty \|\widehat{\Gamma}_N - \widehat{\Gamma}\|_\infty.$$

En vertu du lemme 4.14 et de la proposition 4.2, on obtient le résultat voulu.

4.7.3 Preuve de la proposition 4.11

On considère la suite (β_N) définie dans la preuve précédente. Alors, pour tout $N \geq N_1$, il existe λ_N tel que $\tilde{\beta}_N = \lambda_N \beta_N$. Alors,

$$\begin{aligned} \lim_{N \rightarrow +\infty} \cos^2(\tilde{\beta}_N, \beta) &= \lim_{N \rightarrow +\infty} \cos^2(\beta_N, \beta) \\ &= \lim_{N \rightarrow +\infty} \left(\frac{\widehat{\rho}_N}{\widehat{\rho}} \right)^2 \frac{|\beta|^2}{|\beta_N|^2} \\ &= 1. \end{aligned}$$

4.7.4 Preuve du théorème 4.13

Soit $Y_N = f(\beta'_N X, \varepsilon)$ et soit \widehat{Y}_N^m sa projection sur une grille optimale en norme \mathbb{L}^p de cardinal m . On commence par montrer que l'erreur de prédiction est inférieure à une somme de quatre termes donnée dans le lemme 4.15. Puis, on contrôle chacun des termes afin de terminer la démonstration.

LEMME 4.15 Pour toute fonction Lipschitz φ , on a, pour tous N et m ,

$$\begin{aligned} \left\| \mathbb{E}[\varphi(Y)|\beta'X] - \mathbb{E}[\varphi(\widehat{Y}^m)|\widehat{\beta}'_N X^m] \right\|_1 &\leq 4[f]_{\text{Lip}}[\varphi]_{\text{Lip}}\|X\|_p|\beta - \beta_N| \\ &+ 2[\varphi]_{\text{Lip}}\|Y_N - \widehat{Y}_N^m\|_p \\ &+ 2[f]_{\text{Lip}}[\varphi]_{\text{Lip}}\|\beta'_N X - \widehat{\beta}'_N X^m\|_p \\ &+ [\varphi]_{\text{Lip}}\|Y - \widehat{Y}^m\|_p. \end{aligned}$$

PREUVE. Soient $V = \beta'X$, $V_N = \beta'_N X$ et $F(v) = \mathbb{E}[\varphi(Y)|V = v]$. Alors, l'erreur de prédiction peut s'écrire de la façon suivante,

$$\begin{aligned} \mathbb{E}[\varphi(Y)|\beta'X] - \mathbb{E}[\varphi(\widehat{Y}^m)|\widehat{\beta}'_N X^m] &= F(V) - F(V_N) \\ &+ F(V_N) - \mathbb{E}[\varphi(Y)|\beta'_N X] \\ &+ \mathbb{E}[\varphi(Y)|\beta'_N X] - \mathbb{E}[\varphi(Y_N)|\beta'_N X] \\ &+ \mathbb{E}[\varphi(Y_N)|\beta'_N X] - \mathbb{E}[\varphi(\widehat{Y}_N^m)|\widehat{\beta}'_N X^m] \\ &+ \mathbb{E}[\varphi(\widehat{Y}_N^m)|\widehat{\beta}'_N X^m] - \mathbb{E}[\varphi(Y_N)|\widehat{\beta}'_N X^m] \\ &+ \mathbb{E}[\varphi(Y_N)|\widehat{\beta}'_N X^m] - \mathbb{E}[\varphi(\widehat{Y}^m)|\widehat{\beta}'_N X^m]. \end{aligned}$$

On donne un majorant de chacun des termes en norme \mathbb{L}^p . Pour le premier, on a

$$\begin{aligned} \|F(V) - F(V_N)\|_p &\leq [\varphi]_{\text{Lip}}[f]_{\text{Lip}}\|V - V_N\|_p \\ &\leq [\varphi]_{\text{Lip}}[f]_{\text{Lip}}\|X\|_p|\beta - \beta_N|. \end{aligned}$$

Pour le second terme, puisque $\sigma(\beta'_N X) \subset \sigma(X)$, on a

$$F(V_N) - \mathbb{E}[\varphi(Y)|\beta'_N X] = \mathbb{E}[F(V_N) - \mathbb{E}[\varphi(Y)|X]|\beta'_N X].$$

Puisque $\mathbb{E}[\varphi(Y)|X] = \mathbb{E}[\varphi(Y)|\beta'X] = F(V)$ par hypothèse sur le modèle, on obtient, par contraction dans \mathbb{L}^p de l'espérance conditionnelle,

$$\begin{aligned} \|F(V_N) - \mathbb{E}[\varphi(Y)|\beta'_N X]\|_p &\leq \|F(V_N) - F(V)\|_p \\ &\leq [\varphi]_{\text{Lip}}[f]_{\text{Lip}}\|X\|_p|\beta - \beta_N|. \end{aligned}$$

Pour le troisième terme, on a

$$\begin{aligned} \left\| \mathbb{E}[\varphi(Y)|\beta'_N X] - \mathbb{E}[\varphi(Y_N)|\beta'_N X] \right\|_p &\leq \|\varphi(Y) - \varphi(Y_N)\|_p \\ &\leq [\varphi]_{\text{Lip}} [f]_{\text{Lip}} \|X\|_p |\beta - \beta_N|. \end{aligned}$$

Pour le quatrième terme, on applique la proposition 4.9 à $Y_N = f(\beta'_N X, \varepsilon)$ et on obtient

$$\begin{aligned} \left\| \mathbb{E}[\varphi(Y_N)|\beta'_N X] - \mathbb{E}[\varphi(\widehat{Y}_N^m)|\widehat{\beta}'_N X^m] \right\|_p &\leq 2[\varphi]_{\text{Lip}} [f]_{\text{Lip}} \|\beta'_N X - \widehat{\beta}'_N X^m\|_p \\ &\quad + [\varphi]_{\text{Lip}} \|Y_N - \widehat{Y}_N^m\|_p. \end{aligned}$$

Pour le cinquième terme, on a

$$\begin{aligned} \left\| \mathbb{E}[\varphi(\widehat{Y}_N^m)|\widehat{\beta}'_N X^m] - \mathbb{E}[\varphi(Y_N)|\widehat{\beta}'_N X^m] \right\|_p &\leq \|\varphi(Y_N) - \varphi(\widehat{Y}_N^m)\|_p \\ &\leq [\varphi]_{\text{Lip}} \|Y_N - \widehat{Y}_N^m\|_p. \end{aligned}$$

Enfin, pour le sixième et dernier terme,

$$\begin{aligned} \left\| \mathbb{E}[\varphi(Y_N)|\widehat{\beta}'_N X^m] - \mathbb{E}[\varphi(\widehat{Y}^m)|\widehat{\beta}'_N X^m] \right\|_p &\leq \|\varphi(Y_N) - \varphi(Y)\|_p + \|\varphi(Y) - \varphi(\widehat{Y}^m)\|_p \\ &\leq [\varphi]_{\text{Lip}} \|Y_N - Y\|_p + [\varphi]_{\text{Lip}} \|Y - \widehat{Y}^m\|_p \\ &\leq [\varphi]_{\text{Lip}} [f]_{\text{Lip}} \|X\|_p |\beta - \beta_N| + [\varphi]_{\text{Lip}} \|Y - \widehat{Y}^m\|_p. \end{aligned}$$

En sommant ces six inégalités, on obtient le résultat recherché. \square

La suite de la preuve du théorème 4.13 se divise en quatre parties.

- (i) En utilisant (4.9) et la proposition 4.2, puisque $\|X\|_{p+\delta} < +\infty$, il existe des constantes D_1, D_2, D_3 telles que, pour tout $N \geq \max(C_0, D_3)$, on a

$$|\beta - \beta_N| \leq \frac{\Delta_0}{N^{1/d}} \left[D_1 \|X\|_{p+\delta}^{p+\delta} + D_2 \right]^{1/p}. \quad (4.10)$$

- (ii) En utilisant la proposition 4.2 à nouveau, il existe des constantes C_1, C_2, C_3 telles que, pour tout $m \geq C_3$, on a

$$\|Y_N - \widehat{Y}_N^m\|_p \leq \frac{1}{m} \left[C_1 \|Y_N\|_{p+\delta}^{p+\delta} + C_2 \right]^{1/p}.$$

Par l'inégalité triangulaire,

$$\begin{aligned}
\|Y_N\|_{p+\delta} &\leq \|Y\|_{p+\delta} + \|Y - Y_N\|_{p+\delta} \\
&\leq \|Y\|_{p+\delta} + [f]_{\text{Lip}} \|X\|_{p+\delta} |\beta - \beta_N| \\
&\leq \|Y\|_{p+\delta} + [f]_{\text{Lip}} \|X\|_{p+\delta} \Delta_0 \|X - \widehat{X}^N\|_p \quad \text{pour } N \geq C_0.
\end{aligned}$$

Avec (4.3), pour tout $N \geq \max(C_0, D_3)$, on a

$$\|Y_N\|_{p+\delta}^{p+\delta} \leq \left[\|Y\|_{p+\delta} + \frac{\Delta_1 \Delta_2}{N^{1/d}} \right]^{p+\delta},$$

où $\Delta_1 = \Delta_0 [f]_{\text{Lip}} \|X\|_{p+\delta}$ and $\Delta_2 = \left(D_1 \|X\|_{p+\delta}^{p+\delta} + D_2 \right)^{1/p}$. Finalement, pour tout $m \geq C_3$ et pour tout $N \geq \max(C_0, D_3)$,

$$\left\| Y_N - \widehat{Y}_N^m \right\|_p \leq \frac{1}{m} \left[C_1 \left(\frac{\Delta_1 \Delta_2}{N^{1/d}} + \|Y\|_{p+\delta} \right)^{p+\delta} + C_2 \right]^{1/p}. \quad (4.11)$$

(iii) A nouveau grâce à la proposition 4.2, il existe des constantes $C'_1, C'_2, C'_3 > 0$ telles que, pour tout $m \geq C'_3$,

$$\begin{aligned}
\left\| \beta'_N X - \widehat{\beta}'_N X^m \right\|_p &\leq \frac{1}{m} \left[C'_1 \|\beta'_N X\|_{p+\delta}^{p+\delta} + C'_2 \right]^{1/p} \\
&\leq \frac{1}{m} \left[C'_1 |\beta_N|^{p+\delta} \|X\|_{p+\delta}^{p+\delta} + C'_2 \right]^{1/p}.
\end{aligned}$$

Par (4.9), on a alors

$$\forall N \geq C_0, |\beta_N| \leq |\beta| + \Delta_0 \|X - \widehat{X}^N\|_p.$$

Avec (4.3), on a, pour tout $N \geq \max(C_0, D_3)$,

$$\left\| \beta'_N X - \widehat{\beta}'_N X^m \right\|_p \leq \frac{1}{m} \left[C'_1 \|X\|_{p+\delta}^{p+\delta} \left(\frac{\Delta_0 \Delta_2}{N^{1/d}} + |\beta| \right)^{p+\delta} + C'_2 \right]^{1/p}. \quad (4.12)$$

(iv) En utilisant la proposition 4.2 à nouveau, il existe des constantes C''_1, C''_2, C''_3 telles que, pour tout $m \geq C''_3$, on a

$$\|Y - \widehat{Y}^m\|_p \leq \frac{1}{m} \left[C''_1 \|Y\|_{p+\delta}^{p+\delta} + C''_2 \right]^{1/p}, \quad (4.13)$$

car $Y \in \mathbb{L}^{p+\delta}$.

Avec (4.10), (4.11), (4.12), (4.13) et l'inégalité du lemme 4.15, on a, pour tout $m \geq \max(C_3, C'_3, C''_3)$ et pour tout $N \geq \max(C_0, D_3)$,

$$\begin{aligned}
& \left\| \mathbb{E}[\varphi(Y)|\beta'X] - \mathbb{E}[\varphi(\widehat{Y}^m)|\widehat{\beta}'_N X^m] \right\|_1 \\
& \leq \frac{1}{N^{1/d}} 4[f]_{\text{Lip}}[\varphi]_{\text{Lip}} \|X\|_p \Delta_0 \left[D_1 \|X\|_{p+\delta}^{p+\delta} + D_2 \right]^{1/p} \\
& \quad + \frac{1}{m} 2[\varphi]_{\text{Lip}} \left[C_1 \left(\frac{\Delta_1 \Delta_2}{N^{1/d}} + \|Y\|_{p+\delta} \right)^{p+\delta} + C_2 \right]^{1/p} \\
& \quad + \frac{1}{m} 2[f]_{\text{Lip}}[\varphi]_{\text{Lip}} \left[C'_1 \|X\|_{p+\delta}^{p+\delta} \left(\frac{\Delta_0 \Delta_2}{N^{1/d}} + |\beta| \right)^{p+\delta} + C'_2 \right]^{1/p} \\
& \quad + \frac{1}{m} \left[C''_1 \|Y\|_{p+\delta}^{p+\delta} + C''_2 \right]^{1/p}.
\end{aligned}$$

Ceci termine la preuve.

ESTIMATION RÉCURSIVE ET NON PARAMÉTRIQUE DU NOYAU D'UN PDMP

Ce chapitre est dédié à une méthode d'estimation non paramétrique du noyau d'un PDMP dont l'espace d'état est \mathbb{R}^d . En s'inspirant des techniques classiques d'estimation à noyau pour les chaînes de Markov, on propose un estimateur récursif pour Q et on démontre sa consistance forte (c'est-à-dire la convergence presque sûre de l'estimateur vers la quantité d'intérêt). Ce travail se démarque des résultats de la littérature car on n'a pas accès ici à une chaîne de Markov gouvernée par le noyau de transition Q . En effet, des difficultés techniques apparaissent à la fois à cause de la structure particulière des chaînes immergées du processus et de l'existence de la frontière de l'espace d'état. Les résultats présentés ici ont été soumis pour publication sous le titre *A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process* [7].

Le chapitre est organisé de la façon suivante. Dans la première partie, on propose une revue de la littérature sur l'estimation non paramétrique des noyaux de Markov à partir de l'observation d'une trajectoire de la chaîne en temps long. Ensuite, on définit notre cadre de travail et on montre que notre problème d'estimation peut se résumer à l'estimation de deux lois invariantes : les sections 5.3 et 5.4 sont consacrées à ces deux sous-problèmes. Enfin, on propose d'illustrer nos résultats asymptotiques par des simulations dans la section 5.5. Des perspectives possibles à ce travail sont présentées dans la dernière partie de ce chapitre.

SOMMAIRE DU CHAPITRE

5.1	Introduction	131
5.2	Considérations préliminaires	133
5.3	Estimation de la loi invariante de (Z_n^-)	135
5.3.1	Quelques propriétés de (Z_n^-)	136
5.3.2	Estimation de p	143
5.4	Estimation de la loi invariante de (Z_n^-, Z_n)	146
5.4.1	Quelques propriétés de (Z_n^-, Z_n)	146
5.4.2	Estimation de h	148
5.5	Illustration numérique	155
5.5.1	Premières simulations	155
5.5.2	Le processus TCP window size	157
5.6	Conclusion du chapitre	158

5.1 Introduction

Dans ce chapitre, on propose un estimateur non paramétrique récursif du noyau de transition d'un processus markovien déterministe par morceaux (PDMP) et on étudie son comportement asymptotique. La procédure d'estimation peut être réalisée à partir de l'observation en temps long d'une trajectoire du processus.

Rappelons ici qu'un des objectifs de cette thèse est de proposer une méthode d'estimation non paramétrique des caractéristiques λ et Q qui contrôlent l'aléa de la dynamique d'un PDMP. Le flot Φ est souvent donné par des lois déterministes de la physique ou de la biologie ; nous ne nous sommes donc pas intéressés à son estimation. Le cadre de travail choisi est l'estimation à partir de l'observation d'une trajectoire en temps long du processus, et non pas à partir de l'observation d'un grand nombre de processus indépendants et distribués selon les mêmes caractéristiques. On s'est intéressé à l'estimation de la densité associée à λ dans le chapitre 3 de ce manuscrit ; dans ce chapitre, on se propose d'estimer le noyau markovien Q d'un PDMP défini sur \mathbb{R}^d sous l'hypothèse d'existence d'une densité q par rapport à la mesure de Lebesgue. Comme lorsqu'on estime la densité associée à λ , on travaille à partir de l'ergodicité des chaînes immergées du PDMP. Elle sera assurée ici par la condition de Doeblin.

Des méthodes d'estimation non paramétrique pour les chaînes de Markov stationnaires ont beaucoup été étudiées, tout d'abord par Roussas dans [77]. Il a proposé des méthodes à noyau pour estimer la loi stationnaire et le noyau de transition d'une chaîne de Markov vérifiant la condition de Doeblin. Plus tard, Rosenblatt a proposé dans [76] des résultats sur le biais et la variance de cet estimateur sous des hypothèses plus faibles. Puis, Yakowitz a amélioré dans [86] les résultats existant sur la normalité asymptotique des estimateurs en utilisant une condition de Harris. Masry et Györfi dans [65], et Basu et Sahoo dans [12], ont complété ces études en affinant certains résultats. Il existe également dans la littérature de nombreux papiers sur l'estimation pour les chaînes de Markov non stationnaires. Sans prétendre présenter une étude exhaustive sur le sujet, on peut suggérer au lecteur les références [20, 37, 41, 51, 60, 61, 63]. Dans ce nouveau contexte, Doukhan et Ghindès ont étudié dans [37] une borne de l'erreur \mathbb{L}^1 de leur estimateur. Hernández-Lerma *et al.* dans [51] et Duflo dans [41] se sont intéressés à des méthodes récursives pour estimer le noyau de transition et la loi invariante d'une chaîne de Markov. Liebscher a donné dans [63] des résultats sous une condition plus faible que l'hypothèse de Doeblin. Plus récemment, Cléménçon dans [20] a proposé un estimateur quotient en utilisant des ondelettes et a donné la borne inférieure du risque minimax \mathbb{L}^p . Lacour a proposé dans [61] un estimateur par

projection avec sélection de modèle, puis elle a introduit dans [60] un estimateur original en considérant une nouvelle expression de contraste dérivée d'un problème de régression.

Les résultats présentés dans ce chapitre et ceux de la littérature sont à la fois différents et complémentaires. Dans ce travail, on propose d'estimer la densité de transition q d'un PDMP : en particulier, on ne dispose pas de l'observation d'une chaîne de Markov dont le noyau de transition est donné par q . On peut néanmoins écrire q comme le rapport de deux lois invariantes : la loi invariante du couple formé par la position juste avant le saut et celle lors du saut, divisée par la loi invariante de la position juste avant le saut. En effet, $Q(x, A)$ est défini comme la probabilité conditionnelle que la position lors du saut soit dans A , sachant que le processus se trouve en x juste avant. Il est donc naturel d'estimer ces deux lois invariantes afin d'obtenir un estimateur de leur rapport. Une première difficulté apparaît lorsqu'on cherche à estimer la loi invariante de la position juste avant le saut : le noyau de transition de cette chaîne de Markov peut charger la frontière de l'espace d'état. Par conséquent, le noyau de transition et la loi invariante qui nous intéresse admettent tous deux une densité seulement sur l'intérieur de leur espace d'état. Mais la difficulté principale rencontrée dans cette étude intervient dans l'étude de la loi invariante bidimensionnelle du processus « position juste avant le saut, position lors du saut ». Cette chaîne de Markov admet une structure particulière : sa loi invariante est à densité sur l'intérieur de l'espace d'état, mais pas son noyau de transition. En effet, la position juste avant le saut est distribuée sur la courbe définie par le flot déterministe initialisée en la position lors du saut précédent. Cela nous oblige à proposer une nouvelle approche pour étudier le comportement asymptotique de l'estimateur de cette loi invariante. La méthode proposée est plus universelle (elle permet d'obtenir la convergence presque sûre sans que le noyau soit à densité sur l'intérieur de l'espace), mais plus restrictive sur l'allure de la fenêtre de lissage. Des complications interviennent à cause de la présence des sauts déterministes lorsque le flot tente de franchir la frontière de l'espace d'état. Cela entraîne notamment que les lois invariantes mentionnées précédemment peuvent charger un ensemble de mesure de Lebesgue nulle.

La suite du chapitre est scindée en quatre parties. La section 5.2 est dédiée à la formulation du problème qui nous intéresse et à l'introduction de quelques notations. La loi invariante de la suite des positions juste avant les sauts est estimée dans la section 5.3 alors qu'on traite de la loi invariante du processus bidimensionnel « position juste avant le saut, position lors du saut » dans la section 5.4. Enfin, la section 5.5 est consacrée à des simulations numériques afin d'illustrer le comportement asymptotique de notre estimateur.

5.2 Considérations préliminaires

Dans cette partie, on rappelle en premier lieu la définition d'un PDMP sur \mathbb{R}^d , où d est un entier supérieur ou égal à 1. Le processus évolue dans un sous-ensemble ouvert E de \mathbb{R}^d , muni de la norme euclidienne notée $|\cdot|$. Le comportement au cours du temps du processus est décrit par les trois caractéristiques (λ, Q, Φ) .

– $\Phi : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ est le flot déterministe. Il satisfait,

$$\forall \xi \in \mathbb{R}^d, \forall s, t \in \mathbb{R}, \Phi_\xi(t+s) = \Phi_{\Phi_\xi(t)}(s).$$

Pour tout $\xi \in E$, $t^+(\xi)$ désigne le temps déterministe de sortie de E ,

$$t^+(\xi) = \inf \{t > 0 : \Phi_\xi(t) \in \partial E\},$$

avec la convention habituelle $\inf \emptyset = +\infty$.

– $\lambda : \mathbb{R}^d \rightarrow \mathbb{R}_+$ est le taux de saut du processus. Il s'agit d'une fonction mesurable qui vérifie,

$$\forall \xi \in \mathbb{R}^d, \exists \varepsilon > 0, \int_0^\varepsilon \lambda(\Phi_\xi(s)) ds < +\infty.$$

– Q est un noyau de Markov sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui vérifie,

$$\forall \xi \in \bar{E}, Q(\xi, E \setminus \{\xi\}) = 1.$$

REMARQUE 5.1 Dans ce chapitre, on utilise la notation t^+ pour désigner le temps (déterministe) d'atteinte de la frontière, au lieu de l'habituel t^ . On propose cette nouvelle notation car, dans la suite, on fera intervenir le temps déterministe d'atteinte de la frontière en parcourant le flot à rebours : cette quantité, négative, sera noté t^- .*

Partant d'un point $x \in E$, l'évolution du processus peut être décrite comme suit. T_1 est une variable aléatoire positive dont la fonction de survie est donnée, pour tout $t \geq 0$, par

$$\mathbb{P}(T_1 > t | X_0 = x) = \exp\left(-\int_0^t \lambda(\Phi_x(s)) ds\right) \mathbb{1}_{\{0 \leq t < t^+(x)\}}.$$

On choisit ensuite une variable aléatoire Z_1 à valeurs dans E selon la loi $Q(\Phi_x(T_1), \cdot)$. Remarquons ici que la nouvelle position du processus juste après le saut dépend seulement de la limite à gauche de la trajectoire avant le saut $\Phi_x(T_1)$. Entre les instants 0 et T_1 , le processus évolue selon

$$X_t = \begin{cases} \Phi_x(t) & \text{pour } 0 \leq t < T_1, \\ Z_1 & \text{pour } t = T_1. \end{cases}$$

Maintenant, partant de X_{T_1} , le temps $S_2 = T_2 - T_1$ et la position Z_2 sont générés selon la même méthode que précédemment, et ainsi de suite. On obtient ainsi un processus markovien fort dont la suite des instants de saut est notée (T_k) , avec $T_0 = 0$ par convention. On considère souvent la chaîne immergée (Z_n, S_n) associée au processus (X_t) , où Z_n est la position lors du $n^{\text{ème}}$ saut ($Z_n = X_{T_n}$) et $S_n = T_n - T_{n-1}$ est la durée entre le $n^{\text{ème}}$ et le $(n-1)^{\text{ème}}$ saut, avec $S_0 = 0$. Notons que la chaîne immergée (Z_n, S_n) contient tout l'aléa du processus à temps continu (X_t) . Dans tout ce chapitre, on suppose que le noyau Q admet une densité q par rapport à la mesure de Lebesgue. Notre objectif principal est l'estimation de cette fonction à partir de l'observation d'une trajectoire en temps long de la chaîne immergée du processus.

HYPOTHÈSE 5.2 *On suppose que le noyau de transition Q admet une densité conditionnelle q , c'est-à-dire,*

$$\forall \xi \in \mathbb{R}^d, \forall B \in \mathcal{B}(\mathbb{R}^d), Q(\xi, B) = \int_B q(\xi, z) dz.$$

De plus, q est supposée continue par morceaux.

Dans la suite, on considère le processus à temps discret (Z_n^-) défini par,

$$\forall n \geq 1, Z_n^- = \Phi_{Z_{n-1}}(S_n).$$

Il est naturel de considérer cette suite puisque le noyau Q décrit les transitions de Z_n^- vers Z_n . Z_n^- représente la limite à gauche de la trajectoire du processus (X_t) lors du $n^{\text{ème}}$ saut. Nous allons prouver dans le lemme 5.3 que (Z_n^-) est une chaîne de Markov homogène et nous expliciterons son noyau de transition.

Notre objectif ici est de donner un estimateur récursif de la quantité $q(x, y)$. L'estimateur récursif de $q(x, y)$ que nous proposons est défini par

$$\hat{q}_n(x, y) = \frac{\sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j - y}{w_j}\right)}{\sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right)},$$

où $w_j = w_1 j^{-\beta}$, $v_j = v_1 j^{-\alpha}$, avec α , β , w_1 et v_1 des réels strictement positifs, et K est une fonction noyau satisfaisant les hypothèses 5.11 présentées plus loin. Sous certaines hypothèses annoncées dans la suite, nous allons montrer dans le théorème 5.21 que cet estimateur est fortement consistant, c'est-à-dire,

$$\widehat{q}_n(x, y) \xrightarrow{\text{p.s.}} q(x, y),$$

quand n tend vers l'infini.

Avant d'aller plus loin, on introduit un certain nombre de notations. Dans toute la suite, f et G désignent la densité et la fonction de survie conditionnelles associées à $\lambda(\Phi(\cdot))$. Plus précisément, pour tout $z \in \mathbb{R}^d$ et pour tout $t \geq 0$,

$$\begin{aligned} G(z, t) &= \exp\left(-\int_0^t \lambda(\Phi_z(s)) ds\right), \\ f(z, t) &= \lambda(\Phi_z(t))G(z, t). \end{aligned}$$

De plus, on note \mathcal{S} la loi conditionnelle du temps inter-saut S_{n+1} sachant Z_n , pour tout entier n . Alors, pour tout $z \in E$ et pour tout $\Gamma \in \mathcal{B}(\mathbb{R}_+)$, on a

$$\begin{aligned} \mathcal{S}(z, \Gamma) &= \mathbb{P}(S_{n+1} \in \Gamma | Z_n = z, \sigma(Z_i, S_i : 0 \leq i \leq n)) \\ &= \mathbb{P}(S_{n+1} \in \Gamma | Z_n = z) \\ &= \int_{\Gamma \cap [0, t^+(z)[} f(z, s) ds + \mathbb{1}_{\Gamma}(t^+(z))G(z, t^+(z)). \end{aligned} \tag{5.1}$$

Par définition, la relation entre \mathcal{S} et la fonction de survie conditionnelle G est donnée, pour tous $z \in \mathbb{R}^d$ et $t \geq 0$, par $G(z, t) = \mathcal{S}(z,]t, +\infty[)$.

Dans la partie suivante, on s'intéresse à la loi invariante du processus (Z_n^-) et on étudie un estimateur non-paramétrique de cette loi.

5.3 Estimation de la loi invariante de (Z_n^-)

L'objectif principal de cette partie est l'estimation de la loi invariante de la chaîne de Markov (Z_n^-) . Cette section est divisée en deux parties. Dans la première, on s'intéresse à l'existence et à l'unicité de la mesure invariante de (Z_n^-) , et aux propriétés de son noyau de transition \mathcal{R} . Dans la seconde partie, on propose un estimateur récursif de la partie à densité de la loi invariante de (Z_n^-) et on étudie son comportement asymptotique.

5.3.1 Quelques propriétés de (Z_n^-)

Dans cette partie, on se focalise sur le processus (Z_n^-) , qui est une chaîne de Markov sur \bar{E} . En particulier, on s'intéresse à son noyau de transition \mathcal{R} et à l'existence d'une mesure invariante. Pour tout z , $\Phi_z^{-1}(B)$ est la préimage du sous-ensemble B de \mathbb{R}^d par l'application Φ_z ,

$$\Phi_z^{-1}(B) = \{t \in \mathbb{R} : \Phi_z(t) \in B\}.$$

LEMME 5.3 (Z_n^-) est une chaîne de Markov dont le noyau de transition \mathcal{R} est donné, pour tous $y \in \bar{E}$ et $B \in \mathcal{B}(\bar{E})$, par

$$\mathcal{R}(y, B) = \int_{\bar{E}} Q(y, dz) \mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbb{R}_+), \quad (5.2)$$

où la loi conditionnelle \mathcal{S} a déjà été définie par (5.1).

PREUVE. Pour tout entier n , par (5.1), on a

$$\begin{aligned} & \mathbb{P}(Z_{n+1}^- \in B | Z_n = z, Z_n^-, \dots, Z_1^-) \\ &= \mathbb{P}(S_{n+1} \in \Phi_z^{-1}(B) \cap \mathbb{R}_+ | Z_n = z, Z_n^-, \dots, Z_1^-) \\ &= \mathbb{E} \left[\mathbb{E} [\mathbb{1}_{\{S_{n+1} \in \Phi_z^{-1}(B) \cap \mathbb{R}_+\}} | Z_n = z, \sigma(Z_i, S_i : 0 \leq i \leq n)] | Z_n = z, Z_n^-, \dots, Z_1^- \right] \\ &= \mathbb{E} [\mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbb{R}_+) | Z_n = z, Z_n^-, \dots, Z_1^-] \\ &= \mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbb{R}_+). \end{aligned}$$

Par conséquent,

$$\begin{aligned} \mathbb{P}(Z_{n+1}^- \in B | Z_n^-, \dots, Z_1^-) &= \int_{\bar{E}} \mathbb{P}(Z_{n+1}^- \in B | Z_n = z, Z_n^-, \dots, Z_1^-) Q(Z_n^-, dz) \\ &= \int_{\bar{E}} \mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbb{R}_+) Q(Z_n^-, dz), \end{aligned}$$

ce qui donne le résultat. □

On s'intéresse à l'ergodicité de la chaîne de Markov (Z_n^-) à partir de la condition de Doeblin.

HYPOTHÈSE 5.4 *On suppose que le noyau de transition \mathcal{R} vérifie la condition de Doeblin (voir [67, page 396] par exemple), c'est-à-dire qu'il existe une mesure de probabilité μ sur $(\bar{E}, \mathcal{B}(\bar{E}))$, un réel ε et un entier k tels que,*

$$\forall y \in \bar{E}, \forall B \in \mathcal{B}(\bar{E}), \mathcal{R}^k(y, B) \geq \varepsilon \mu(B). \quad (5.3)$$

La condition de Doeblin pour le noyau \mathcal{R} peut sembler délicate à vérifier en pratique. Elle est toutefois satisfaite dans les applications présentées dans le chapitre 3. Sous cette hypothèse, on peut établir des résultats sur le processus (Z_n^-) .

PROPOSITION 5.5 *On a les résultats suivants.*

- *La chaîne de Markov (Z_n^-) est μ -irréductible, apériodique et admet une mesure invariante unique, qu'on notera π .*
- *Il existe $\rho > 1$ et $\kappa > 0$ tels que,*

$$\forall n \geq 1, \sup_{\xi \in \bar{E}} \|\mathcal{R}^n(\xi, \cdot) - \pi\|_{TV} \leq \kappa \rho^{-n}, \quad (5.4)$$

où $\|\cdot\|_{TV}$ désigne la norme en variation totale.

- *De plus, (Z_n^-) est positive Harris-récurrente.*

PREUVE. Par définition et l'inégalité (5.3), (Z_n^-) est μ -irréductible et apériodique (voir [67, page 114]). De plus, d'après le théorème 16.0.2 de [67], (Z_n^-) admet une loi invariante unique π , puisqu'elle est apériodique, et on a (5.4). Finalement, en vertu du théorème 4.3.3 de [53], (Z_n^-) est positive Harris-récurrente. \square

Maintenant, on impose des conditions sur le flot déterministe du processus afin d'obtenir une expression plus agréable du noyau \mathcal{R} . Dans la suite, pour $\xi \in E$, $t^-(\xi)$ est le temps déterministe d'atteinte de la frontière de E mais en suivant le flot à rebours,

$$t^-(\xi) = \sup\{t < 0 : \Phi_\xi(t) \in \partial E\},$$

avec la convention habituelle $\sup \emptyset = -\infty$. Notons que $t^-(\xi)$ est négatif.

HYPOTHÈSES 5.6

(i) Le flot Φ est \mathcal{C}^1 . Pour tout couple $(z, t) \in \mathbb{R}^d \times \mathbb{R}$, $D\Phi_z(t)$ est donné par

$$D\Phi_z(t) = \left| \det \left(\frac{\partial \Phi_x^{(i)}(t)}{\partial x_j} \right)_{1 \leq i, j \leq d} \right|. \quad (5.5)$$

(ii) Quelque soit $t \in \mathbb{R}$, l'application $\varphi_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$, donnée par $\varphi_t(x) = \Phi_x(t)$, est injective.

On donne maintenant une expression du noyau \mathcal{R} . Le résultat technique qui suit est un point clé dans l'estimation de la loi invariante de la chaîne de Markov (Z_n^-) .

PROPOSITION 5.7 Soit $y \in E$ et $B \in \mathcal{B}(\bar{E})$. On a

$$\mathcal{R}(y, B) = \int_{B \cap E} r(y, z) dz + \mathcal{R}(y, B \cap \partial E), \quad (5.6)$$

où la densité conditionnelle r est donnée par,

$$\forall z \in E, r(y, z) = \int_0^{-t^-(z)} q(y, \Phi_z(-s)) f(\Phi_z(-s), s) D\Phi_z(-s) ds. \quad (5.7)$$

PREUVE. On suppose $B \subset E$. Fixons $t \geq 0$. On définit l'ensemble A_t par

$$A_t = \{\Phi_\xi(-t) : \xi \in B\} \subset \mathbb{R}^d,$$

et on considère la fonction $\varphi_t : A_t \rightarrow B$ donnée par $\varphi_t(x) = \Phi_x(t)$, pour $x \in A_t$. φ_t est une application injective et \mathcal{C}^1 d'après les hypothèses 5.6. De plus, quelque soit $z \in B$,

$$\varphi_t(\Phi_z(-t)) = \Phi_{\Phi_z(-t)}(t) = z,$$

avec $\Phi_z(-t) \in A_t$ par définition de A_t . Par conséquent, φ_t est une bijection \mathcal{C}^1 . Sa réciproque φ_t^{-1} est donnée par $\varphi_t^{-1}(x) = \Phi_x(-t)$, donc elle est également \mathcal{C}^1 . Donc, φ_t est

un C^1 -difféomorphisme de A_t dans B , ce qui nous permet de le considérer en tant que changement de variable. En particulier, on a

$$(z \in E, t \in \mathbb{R}_+, \Phi_z(t) \in B) \Leftrightarrow (t \in \mathbb{R}_+, z \in E, z \in A_t). \quad (5.8)$$

La correspondance entre les ensembles A_t et B est illustrée sur la figure 5.1. De plus, la matrice jacobienne $\mathbb{J}_{\varphi_t^{-1}}$ de la réciproque φ_t^{-1} satisfait,

$$\forall x \in \mathbb{R}^d, \mathbb{J}_{\varphi_t^{-1}}(x) = \left(\frac{\partial \Phi_x^{(i)}(-t)}{\partial x_j} \right)_{1 \leq i, j \leq d}.$$

Par (5.1) et (5.2), on a

$$\mathcal{R}(y, B) = \int_E \left(\int_{\Phi_z^{-1}(B) \cap \mathbb{R}_+} q(y, z) f(z, t) dt \right) dz.$$

Avec (5.8), on obtient

$$\mathcal{R}(y, B) = \int_{\mathbb{R}_+} \left(\int_{A_t} \mathbb{1}_E(z) f(z, t) q(y, z) dz \right) dt.$$

Par le changement de variable φ_t , on a

$$\begin{aligned} \mathcal{R}(y, B) &= \int_{\mathbb{R}_+} \left(\int_B \mathbb{1}_E(\varphi_t^{-1}(\xi)) f(\varphi_t^{-1}(\xi), t) q(y, \varphi_t^{-1}(\xi)) \left| \det \mathbb{J}_{\varphi_t^{-1}}(\xi) \right| d\xi \right) dt \\ &= \int_{\mathbb{R}_+} \left(\int_B \mathbb{1}_E(\Phi_\xi(-t)) f(\Phi_\xi(-t), t) q(y, \Phi_\xi(-t)) D\Phi_\xi(-t) d\xi \right) dt, \end{aligned}$$

où $D\Phi_\xi$ est donné par (5.5). On remarque que

$$(\Phi_\xi(-t) \in E) \Leftrightarrow (0 \leq t < -t^-(\xi)),$$

donc $\mathbb{1}_E(\Phi_\xi(-t)) = \mathbb{1}_{\{0 \leq t < -t^-(\xi)\}}$. On obtient le résultat voulu par le théorème de Fubini. \square

A la lumière de ce résultat, on peut obtenir la propriété suivante sur la loi invariante π de la chaîne de Markov (Z_n^-) : π admet une densité sur l'intérieur de l'espace d'état. De plus, on peut écrire le lien entre cette densité et r .

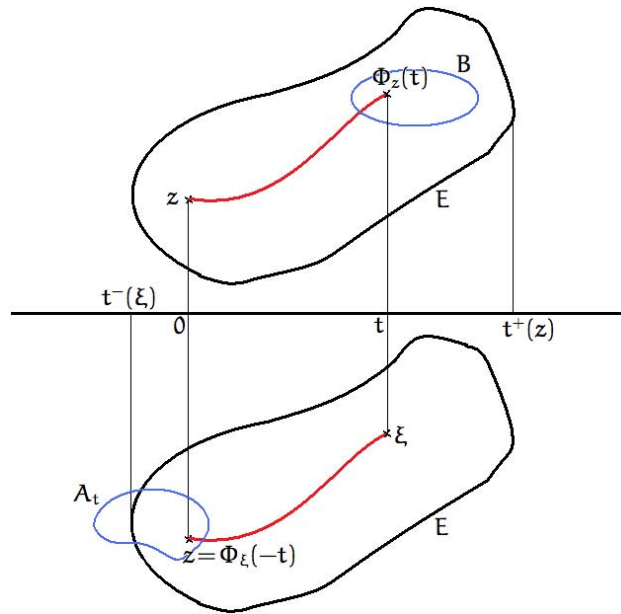


FIGURE 5.1 – Correspondance entre les ensembles A_t et B .

COROLLAIRE 5.8 Il existe une fonction positive p telle que,

$$\forall B \in \mathcal{B}(\bar{E}), \pi(B) = \int_{B \cap E} p(x) dx + \pi(B \cap \partial E). \quad (5.9)$$

De plus, p vérifie,

$$\forall x \in E, p(x) = \int_{\bar{E}} \pi(dy) r(y, x). \quad (5.10)$$

PREUVE. D'après la proposition 5.5, μ est une mesure d'irréductibilité pour le noyau markovien \mathcal{R} . D'après le théorème 4.2.2 de [67], la mesure d'irréductibilité maximale $\tilde{\mu}$ est équivalente à la mesure $\tilde{\mu}'$, donnée pour tout $B \in \mathcal{B}(\bar{E})$, par

$$\tilde{\mu}'(B) = \int_{\bar{E}} \mu(dy) \sum_{n \geq 0} \mathcal{R}^n(y, B) \frac{1}{2^{n+1}}.$$

\mathcal{R} admet une densité sur l'intérieur E de l'espace d'état \bar{E} de (Z_n^-) (voir la proposition 5.7). Par conséquent, c'est aussi le cas pour \mathcal{R}^n . En effet, quelque soit l'ensemble B tel que

$\lambda_d(B \cap E) = 0$, on a

$$\mathcal{R}^n(\mathbf{y}, B \cap E) = \int_{\bar{E}} \mathcal{R}^{n-1}(\mathbf{y}, d\mathbf{z}) \mathcal{R}(\mathbf{z}, B \cap E) = 0.$$

Donc, $\tilde{\mu}'(B \cap E) = 0$. Finalement,

$$\lambda_d(B \cap E) = 0 \Rightarrow \tilde{\mu}(B \cap E) = 0.$$

Puisque π et la mesure d'irréductibilité maximale $\tilde{\mu}$ sont équivalentes, π admet une densité sur E . On s'intéresse maintenant à l'expression de cette densité. Par le théorème de Fubini, on a pour $B \subset E$,

$$\begin{aligned} \pi(B) &= \int_{\bar{E}} \pi(d\mathbf{y}) \mathcal{R}(\mathbf{y}, B) \\ &= \int_{\bar{E}} \pi(d\mathbf{y}) \int_B r(\mathbf{y}, \mathbf{x}) d\mathbf{x} \\ &= \int_B \left(\int_{\bar{E}} \pi(d\mathbf{y}) r(\mathbf{y}, \mathbf{x}) \right) d\mathbf{x}. \end{aligned}$$

Par conséquent, on peut identifier p et la fonction $\int_{\bar{E}} \pi(d\mathbf{y}) r(\mathbf{y}, \cdot)$. \square

La régularité de la densité conditionnelle r est importante dans la suite. Nous introduisons de nouvelles hypothèses sous lesquelles r est Lipschitz.

HYPOTHÈSES 5.9 *On fait les hypothèses suivantes.*

(i) t^- est bornée et Lipschitz,

$$\exists [t^-]_{\text{Lip}} > 0, \forall \mathbf{x}, \mathbf{y} \in E, |t^-(\mathbf{x}) - t^-(\mathbf{y})| \leq [t^-]_{\text{Lip}} |\mathbf{x} - \mathbf{y}|.$$

(ii) Le flot Φ est Lipschitz,

$$\exists [\Phi]_{\text{Lip}} > 0, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \forall t \in \mathbb{R}, |\Phi_{\mathbf{x}}(t) - \Phi_{\mathbf{y}}(t)| \leq [\Phi]_{\text{Lip}} |\mathbf{x} - \mathbf{y}|.$$

(iii) f est bornée et Lipschitz,

$$\exists [f]_{\text{Lip}} > 0, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \forall t \in \mathbb{R}, |f(\mathbf{x}, t) - f(\mathbf{y}, t)| \leq [f]_{\text{Lip}} |\mathbf{x} - \mathbf{y}|.$$

(iv) q est bornée et il existe une constante $[q]_{\text{Lip}} > 0$ telle que, pour tous $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d$,

$$|q(\mathbf{x}, \mathbf{y}) - q(\mathbf{x}, \mathbf{z})| \leq [q]_{\text{Lip}} |\mathbf{y} - \mathbf{z}| \quad \text{et} \quad |q(\mathbf{x}, \mathbf{z}) - q(\mathbf{y}, \mathbf{z})| \leq [q]_{\text{Lip}} |\mathbf{x} - \mathbf{y}|.$$

(v) $D\Phi$ est bornée et Lipschitz,

$$\exists [D\Phi]_{\text{Lip}} > 0, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \forall t \in \mathbb{R}, |D\Phi_{\mathbf{x}}(t) - D\Phi_{\mathbf{y}}(t)| \leq [D\Phi]_{\text{Lip}} |\mathbf{x} - \mathbf{y}|.$$

On a le résultat suivant.

PROPOSITION 5.10 *r est une fonction bornée. De plus, il existe une constante $[r]_{\text{Lip}} > 0$ telle que, pour tous $x \in \bar{E}$, $y \in E$ et $u \in \mathbb{R}^d$ tels que $y + u \in E$, on a*

$$|r(x, y + u) - r(x, y)| \leq [r]_{\text{Lip}} |u|.$$

PREUVE. D'après (5.7), on a

$$\|r\|_{\infty} \leq \|t^{-}\|_{\infty} \|q\|_{\infty} \|f\|_{\infty} \|D\Phi\|_{\infty}.$$

Pour le second point, on considère la fonction γ définie par,

$$\forall (y, t) \in \mathbb{R}^d \times \mathbb{R}, \gamma(y, t) = q(x, \Phi_y(-t)) f(\Phi_y(-t), t) D\Phi_y(-t).$$

Cette fonction est Lipschitz comme produit et composition de fonctions Lipschitz et bornées (voir les hypothèses 5.9). $[\gamma]_{\text{Lip}}$ désigne sa constante de Lipschitz. On a donc

$$|\gamma(y, t) - \gamma(y + u, t)| \leq [\gamma]_{\text{Lip}} |u|.$$

De plus, par (5.7), la fonction $r(x, y)$ est donnée par $r(x, y) = \int_0^{-t^-(y)} \gamma(y, s) ds$. On suppose que $-t^-(y) \leq -t^-(y + u)$ (on rappelle que t^- est une fonction négative). On a

$$r(x, y + u) - r(x, y) = \int_0^{-t^-(y)} (\gamma(y + u, s) - \gamma(y, s)) ds + \int_{-t^-(y)}^{-t^-(y+u)} \gamma(y + u, s) ds.$$

Par conséquent,

$$\begin{aligned} |r(x, y + u) - r(x, y)| &\leq \int_0^{\|t^-\|_{\infty}} |\gamma(y + u, s) - \gamma(y, s)| ds \\ &\quad + \|q\|_{\infty} \|f\|_{\infty} \|D\Phi\|_{\infty} |t^-(y) - t^-(y + u)| \\ &\leq \|t^-\|_{\infty} [\gamma]_{\text{Lip}} |u| + [t^-]_{\text{Lip}} \|q\|_{\infty} \|f\|_{\infty} \|D\Phi\|_{\infty} |u|. \end{aligned}$$

L'inégalité obtenue pour $-t^-(y) > -t^-(y + u)$ est exactement la même. Ceci termine la preuve. \square

5.3.2 Estimation de p

On propose un estimateur non paramétrique et récursif de la fonction p donnée dans le corollaire 5.8. Dans toute la suite, on considère une fonction noyau K qui satisfait les hypothèses suivantes.

HYPOTHÈSES 5.11 *Le noyau $K : \mathbb{R}^d \rightarrow \mathbb{R}_+$ vérifie :*

- (i) $\text{supp } K \subset B(0_{\mathbb{R}^d}, \delta)$, où $\delta > 0$ et $B(x, r)$ désigne la boule ouverte centrée en x et dont le rayon est r .
- (ii) K est borné.

Sous les hypothèses 5.11, la quantité $\int_{\mathbb{R}^d} K^2(y) dy$ est finie. On note τ^2 cette intégrale dans la suite. Pour tout entier n , l'estimateur récursif \hat{p}_n de p que nous proposons est donné, pour tout $x \in E$, par

$$\hat{p}_n(x) = \frac{1}{n} \sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right), \quad (5.11)$$

où la fenêtre de lissage v_j satisfait

$$v_j = v_1 j^{-\alpha}, \text{ avec } \alpha, v_1 > 0.$$

REMARQUE 5.12 *Soient $x \in E$ et $j \geq 1$. Puisque la suite (v_n) est décroissante, on a*

$$\text{supp } K\left(\frac{\cdot - x}{v_j}\right) \subset \text{supp } K\left(\frac{\cdot - x}{v_1}\right) \subset B(x, v_1 \delta).$$

Donc, si $v_1 \delta < \text{dist}(x, \partial E)$, on a

$$\text{supp } K\left(\frac{\cdot - x}{v_j}\right) \subset E.$$

Dans la proposition suivante, on établit la convergence ponctuelle presque sûre de \hat{p}_n .

PROPOSITION 5.13 *Soit $x \in E$. On choisit v_1 tel que $v_1 \delta < \text{dist}(x, \partial E)$ et α tel que $\alpha d < 1$. Alors,*

$$\hat{p}_n(x) \xrightarrow{\text{p.s.}} p(x),$$

quand n tend vers l'infini.

PREUVE. D'après l'expression de $p(x)$ donnée par (5.10), la différence $\hat{p}_n(x) - p(x)$ peut s'écrire de la façon suivante,

$$\begin{aligned}\hat{p}_n(x) - p(x) &= \frac{1}{n} \sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right) - \int_{\bar{E}} r(u, x) \pi(du) \\ &= \frac{1}{nv_1^d} K\left(\frac{Z_1^- - x}{v_1}\right) + \frac{1}{n} M_n + R_n^{(1)} + R_n^{(2)},\end{aligned}\quad (5.12)$$

où M_n est donné par

$$M_n = \sum_{j=1}^n \left[\frac{1}{v_{j+1}^d} K\left(\frac{Z_{j+1}^- - x}{v_{j+1}}\right) - \int_{\mathbb{R}^d} r(Z_j^-, x + yv_{j+1}) K(y) dy \right],$$

et les termes de reste $R_n^{(1)}$ et $R_n^{(2)}$ sont donnés par

$$\begin{aligned}R_n^{(1)} &= \frac{1}{n} \sum_{j=1}^n \int_{\mathbb{R}^d} [r(Z_j^-, x + yv_{j+1}) - r(Z_j^-, x)] K(y) dy, \\ R_n^{(2)} &= \frac{1}{n} \sum_{j=1}^n r(Z_j^-, x) - \int_{\bar{E}} r(u, x) \pi(du).\end{aligned}$$

La dépendance en x est implicite et n'est pas rappelée dans les notations M_n , $R_n^{(1)}$ et $R_n^{(2)}$. Dans la décomposition (5.12), le premier terme va clairement vers 0 quand n tend vers l'infini. La suite de la preuve est découpée en trois parties : dans la première, on montre que $R_n^{(2)}$ converge vers 0 par le théorème ergodique. Dans la seconde, on se focalise sur $R_n^{(1)}$ et on prouve que ce terme converge aussi vers 0. Finalement, on montre que M_n/n tend vers 0 en utilisant la seconde loi des grands nombres pour les martingales. On rappelle que la chaîne de Markov (Z_n^-) est positive Harris-récurrente et que sa mesure invariante est π , d'après la proposition 5.5. Comme r est bornée (voir la proposition 5.10), on peut donc appliquer le théorème ergodique presque sûr (voir par exemple le théorème 17.1.7 de [67]) et on obtient que $R_n^{(2)}$ converge presque sûrement vers 0. Pour $R_n^{(1)}$, on a

$$\begin{aligned}|R_n^{(1)}| &\leq \frac{1}{n} \sum_{j=1}^n \int_{\mathbb{R}^d} |r(Z_j^-, x + yv_{j+1}) - r(Z_j^-, x)| K(y) dy \\ &\leq \frac{1}{n} \sum_{j=1}^n \int_{\mathbb{R}^d} [r]_{\text{Lip}} |y| v_{j+1} K(y) dy \\ &\leq \frac{1}{n} \sum_{j=1}^n v_{j+1} \left(\int_{\mathbb{R}^d} |y| K(y) dy \right) [r]_{\text{Lip}}.\end{aligned}$$

Ce majorant tend vers 0 d'après le lemme de Cesaro car la limite de la suite (v_n) est 0. Donc, $R_n^{(1)}$ converge vers 0 quand n tend vers l'infini. Finalement on s'intéresse au terme M_n/n . On commence par montrer que la suite (M_n) est une martingale à temps discret par rapport à la filtration (\mathcal{F}_n) donnée par,

$$\forall n \geq 1, \mathcal{F}_n = \sigma(Z_1^-, \dots, Z_{n+1}^-).$$

On a

$$\mathbb{E}[M_n | \mathcal{F}_{n-1}] = M_{n-1} + \mathbb{E} \left[\frac{1}{v_{n+1}^d} K \left(\frac{Z_{n+1}^- - x}{v_{n+1}} \right) \middle| Z_n^- \right] - \int_{\mathbb{R}^d} r(Z_n^-, x + yv_{n+1}) K(y) dy.$$

Par conséquent, on a juste à prouver que

$$\mathbb{E} \left[\frac{1}{v_{n+1}^d} K \left(\frac{Z_{n+1}^- - x}{v_{n+1}} \right) \middle| Z_n^- \right] = \int_{\mathbb{R}^d} r(Z_n^-, x + yv_{n+1}) K(y) dy. \quad (5.13)$$

On a

$$\mathbb{E} \left[\frac{1}{v_{n+1}^d} K \left(\frac{Z_{n+1}^- - x}{v_{n+1}} \right) \middle| Z_n^- \right] = \frac{1}{v_{n+1}^d} \int_{\bar{E}} K \left(\frac{u - x}{v_{n+1}} \right) \mathcal{R}(Z_n^-, du).$$

Par hypothèse sur v_1 et la remarque 5.12,

$$\begin{aligned} \int_{\bar{E}} K \left(\frac{u - x}{v_{n+1}} \right) \mathcal{R}(Z_n^-, du) &= \int_E K \left(\frac{u - x}{v_{n+1}} \right) \mathcal{R}(Z_n^-, du) \\ &= \int_E K \left(\frac{u - x}{v_{n+1}} \right) r(Z_n^-, u) du, \end{aligned}$$

d'après (5.6). Finalement, le changement de variable $u = yv_{n+1} + x$ établit (5.13). Donc, (M_n) est bien une martingale. On va étudier le comportement asymptotique de son crochet $\langle M \rangle$. Un calcul simple montre que

$$\begin{aligned} (M_n - M_{n-1})^2 &= \frac{1}{v_{n+1}^{2d}} K^2 \left(\frac{Z_{n+1}^- - x}{v_{n+1}} \right) + \left[\int_E r(Z_n^-, x + yv_{n+1}) K(y) dy \right]^2 \\ &\quad - \frac{2}{v_{n+1}^d} K \left(\frac{Z_{n+1}^- - x}{v_{n+1}} \right) \int_E r(Z_n^-, x + yv_{n+1}) K(y) dy. \end{aligned}$$

En utilisant la méthode mise en œuvre pour montrer (5.13), on en déduit que

$$\begin{aligned} \mathbb{E} [(M_n - M_{n-1})^2 | \mathcal{F}_{n-1}] &= \frac{1}{v_{n+1}^d} \int_E K^2(y) r(Z_n^-, x + yv_{n+1}) dy \\ &\quad - \left[\int_E r(Z_n^-, x + yv_{n+1}) K(y) dy \right]^2. \end{aligned}$$

Par conséquent,

$$\langle M \rangle_n \leq \sum_{j=1}^n \left(\frac{1}{v_{j+1}^d} \|r\|_\infty \tau^2 + \|r\|_\infty^2 \right) \sim \text{cst } n^{\alpha d + 1},$$

quand n tend vers l'infini. Par la seconde loi des grands nombres pour les martingales (voir le théorème 1.3.15 de [41]), on a

$$M_n^2 = \mathcal{O} \left(\langle M \rangle_n (\ln(\langle M \rangle_n))^{1+\gamma} \right), \quad \text{avec } \gamma > 0.$$

Par conséquent,

$$\frac{M_n}{n} = \mathcal{O} \left(\sqrt{n^{\alpha d - 1} (\ln(n^{\alpha d + 1}))^{1+\gamma}} \right).$$

Donc, M_n/n tend presque sûrement vers 0 lorsque $\alpha d < 1$. Ceci termine la preuve. \square

5.4 Estimation de la loi invariante de (Z_n^-, Z_n)

Dans cette partie, on montre que la chaîne de Markov (Z_n^-, Z_n) admet une loi invariante unique et on s'intéresse à l'estimation récursive de cette mesure.

5.4.1 Quelques propriétés de (Z_n^-, Z_n)

On se focalise sur le comportement asymptotique de la suite (Z_n^-, Z_n) . Dans toute la suite, η_n (respectivement π_n) désigne la loi du couple (Z_n^-, Z_n) (respectivement de Z_n^-) pour tout entier n . On a les relations (évidentes) suivantes entre η_n , Q ou q et π_n ,

$$\begin{aligned} \eta_n(A \times B) &= \int_A Q(z, B) \pi_n(dz) \\ &= \int_{A \times B} q(z, y) \pi_n(dz) dy. \end{aligned} \quad (5.14)$$

LEMME 5.14 *On a*

$$\lim_{n \rightarrow +\infty} \|\eta_n - \eta\|_{TV} = 0,$$

où la loi limite η est donnée, pour tout $A \times B \in \mathcal{B}(\bar{E} \times E)$, par

$$\eta(A \times B) = \int_{A \times B} q(z, y) \pi(dz) dy. \quad (5.15)$$

PREUVE. Soit g une fonction mesurable bornée par 1. Par le théorème de Fubini, on a

$$\left| \int_{\bar{E} \times E} g(x, y) (\eta_n(dx \times dy) - \eta(dx \times dy)) \right| \leq \left| \int_{\bar{E}} (\pi_n(dx) - \pi(dx)) \int_E g(x, y) q(x, y) dy \right|,$$

d'après (5.14) et (5.15). Donc,

$$\left| \int_{\bar{E} \times E} g(x, y) (\eta_n(dx \times dy) - \eta(dx \times dy)) \right| \leq \left| \int_{\bar{E}} \tilde{g}(x) (\pi_n(dx) - \pi(dx)) \right|,$$

où la fonction $\tilde{g} : x \mapsto \int_E g(x, y) q(x, y) dy$ est bornée par 1 car g est bornée par 1 et q est la densité conditionnelle associée au noyau de Markov Q . Par conséquent,

$$\|\eta_n - \eta\|_{TV} \leq \|\pi_n - \pi\|_{TV}. \quad (5.16)$$

On obtient la limite voulue avec (5.4). \square

Montrons maintenant que η admet une densité sur $E \times E$.

LEMME 5.15 *Il existe une fonction positive h telle que*

$$\eta(A \times B) = \int_{A \times B} h(x, y) dx dy,$$

pour tout $A \times B \in \mathcal{B}(\bar{E} \times E)$ tel que $A \subset E$. De plus, h est donnée, pour tous $x, y \in E$, par

$$h(x, y) = p(x)q(x, y). \quad (5.17)$$

PREUVE. De (5.15), on a

$$\begin{aligned} \eta(A \times B) &= \int_{A \times B} q(z, y) \pi(dz) dy \\ &= \int_{A \times B} q(z, y) p(z) dz dy, \end{aligned}$$

par (5.9) puisque $A \subset E$. On conclut en identifiant. \square

5.4.2 Estimation de h

On se propose d'estimer la fonction h par la quantité \widehat{h}_n donnée, pour tout couple $(x, y) \in E^2$, par

$$\widehat{h}_n(x, y) = \frac{1}{n} \sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j^- - y}{w_j}\right), \quad (5.18)$$

où la fenêtre de lissage w_j est donnée par

$$w_j = w_1 j^{-\beta}, \text{ avec } w_1, \beta > 0.$$

La fonction noyau K vérifie les hypothèses 5.11. Dans la suite, on s'intéresse à la convergence ponctuelle presque sûre de l'estimateur \widehat{h}_n en un point $(x, y) \in E^2$. On suppose que w_1 est tel que $w_1 \delta < \text{dist}(x, \partial E)$, où δ est le rayon de la boule ouverte qui contient le support de K . Dans ce cas, la remarque 5.12 est encore valide et on a les inclusions suivantes, pour tout entier j ,

$$\text{supp } K\left(\frac{\cdot - x}{w_j}\right) \subset B(x, w_1 \delta) \subset E. \quad (5.19)$$

Notre objectif principal est d'établir dans la proposition 5.20 que $\widehat{h}_n(x, y)$ converge presque sûrement vers $h(x, y)$. On commence par montrer que cet estimateur est asymptotiquement non biaisé (voir la proposition 5.17).

Pour cela, nous devons établir de nouvelles propriétés des mesures π_n et π . On rappelle que π_n est la loi de Z_n^- , tandis que π est la loi invariante de la chaîne de Markov (Z_n^-) .

LEMME 5.16 *On a les résultats suivants.*

- Pour tout entier n , π_n admet une densité p_n sur E .
- p_n est bornée par $\|r\|_\infty$ et est $[r]_{\text{Lip}}$ -Lipschitz.
- p est Lipschitz.
- Pour tout entier n , on a

$$\sup_{x \in E} |p_n(x) - p(x)| \leq \|r\|_\infty \kappa \rho^{-(n-1)}. \quad (5.20)$$

PREUVE. Pour le premier point, soit $B \in \mathcal{B}(\bar{E})$ tel que $B \subset E$. On a

$$\begin{aligned}\pi_n(B) &= \int_{\bar{E}} \int_B \mathcal{R}(\xi, dy) \pi_{n-1}(d\xi) \\ &= \int_B \int_{\bar{E}} r(\xi, y) \pi_{n-1}(d\xi) dy,\end{aligned}$$

où r est la densité conditionnelle associée au noyau \mathcal{R} donnée par (5.6). Par conséquent, on peut identifier

$$p_n(y) = \int_{\bar{E}} r(\xi, y) \pi_{n-1}(d\xi). \quad (5.21)$$

Pour le second point, on a déjà montré dans la proposition 5.10 que r est bornée. Ainsi,

$$|p_n(y)| \leq \|r\|_{\infty} \pi_{n-1}(\bar{E}) = \|r\|_{\infty}.$$

De plus, puisque r est Lipschitz,

$$\begin{aligned}|p_n(y) - p_n(z)| &\leq \int_{\bar{E}} |r(\xi, y) - r(\xi, z)| \pi_{n-1}(d\xi) \\ &\leq [r]_{\text{Lip}} |y - z| \pi_{n-1}(\bar{E}) \\ &= [r]_{\text{Lip}} |y - z|.\end{aligned}$$

Pour le troisième point, p est Lipschitz pour les mêmes raisons que p_n puisque p vérifie (5.10). Finalement, pour le quatrième et dernier point, on a, par (5.10) et (5.21),

$$\begin{aligned}|p_n(x) - p(x)| &\leq \int_{\bar{E}} r(y, x) |\pi_{n-1}(dy) - \pi(dy)| \\ &\leq \|r\|_{\infty} \|\pi_{n-1} - \pi\|_{\text{TV}} \\ &\leq \|r\|_{\infty} \kappa \rho^{-(n-1)},\end{aligned}$$

par (5.4). Cela termine la preuve. □

Maintenant, on peut montrer que $\hat{h}_n(x, y)$ est un estimateur asymptotiquement sans biais de $h(x, y)$. On rappelle que $\hat{h}_n(x, y)$ a déjà été défini par (5.18).

PROPOSITION 5.17 *Quand n tend vers l'infini,*

$$\mathbb{E} \left[\hat{h}_n(x, y) \right] \rightarrow h(x, y).$$

PREUVE. Il suffit de montrer que

$$\mathbb{E} \left[\widehat{h}_n(x, y) \right] - \frac{n+1}{n} h(x, y)$$

tend vers 0. On a, par (5.14) et (5.18),

$$\begin{aligned} \mathbb{E} \left[\widehat{h}_n(x, y) \right] - \frac{n+1}{n} h(x, y) &= \frac{1}{n} \sum_{j=1}^{n+1} \left[\int_{\bar{E} \times E} \frac{1}{w_j^{2d}} K \left(\frac{u-x}{w_j} \right) K \left(\frac{v-y}{w_j} \right) \pi_j(du) q(u, v) dv \right. \\ &\quad \left. - \int_{E \times E} h(x, y) K(u) K(v) dudv \right]. \end{aligned}$$

D'après (5.19), on peut remplacer \bar{E} par E dans la première intégrale. On peut donc aussi remplacer $\pi_j(du)$ par $p_j(u)du$ (voir le lemme 5.16). Avec (5.17) et un changement de variable,

$$\begin{aligned} \mathbb{E} \left[\widehat{h}_n(x, y) \right] - \frac{n+1}{n} h(x, y) &= \frac{1}{n} \sum_{j=1}^{n+1} \int_{E \times E} K(u) K(v) \left(p_j(x + uw_j) q(x + uw_j, y + vw_j) - p(x) q(x, y) \right) du dv. \end{aligned}$$

De plus, puisque p_j est $[r]_{\text{Lip}}$ -Lipschitz et bornée par $\|r\|_\infty$ d'après le lemme 5.16, un calcul élémentaire mène à

$$\begin{aligned} &|p_j(x + uw_j) q(x + uw_j, y + vw_j) - p(x) q(x, y)| \\ &\leq \|r\|_\infty [q]_{\text{Lip}} w_j (|u| + |v|) + \|q\|_\infty [r]_{\text{Lip}} |u| w_j + \|q\|_\infty |p_j(x) - p(x)| \\ &\leq w_j (|u| (\|q\|_\infty [r]_{\text{Lip}} + \|q\|_{\text{Lip}} [r]_\infty) + |v| \|r\|_\infty [q]_{\text{Lip}}) + \|q\|_\infty \|r\|_\infty \kappa \rho^{-(j-1)}, \end{aligned}$$

avec (5.20). Finalement, on obtient

$$\begin{aligned} &\left| \mathbb{E} \left[\widehat{h}_n(x, y) \right] - \frac{n+1}{n} h(x, y) \right| \\ &\leq \frac{(2\|r\|_\infty [q]_{\text{Lip}} + \|q\|_\infty [r]_{\text{Lip}}) \int_E K(u) |u| du}{n} \sum_{j=1}^{n+1} w_j + \frac{\|r\|_\infty \|q\|_\infty \kappa}{n} \sum_{j=1}^{n+1} \rho^{-(j-1)} \\ &\leq \frac{(2\|r\|_\infty [q]_{\text{Lip}} + \|q\|_\infty [r]_{\text{Lip}}) \int_E K(u) |u| du}{n} \sum_{j=1}^{n+1} w_j + \frac{\|r\|_\infty \|q\|_\infty \kappa}{n(1-\rho^{-1})}, \end{aligned}$$

ce qui tend vers 0 par le lemme de Cesaro. □

Dans la suite, on va s'intéresser au processus à temps discret

$$(A_n) = \left(\frac{1}{w_n^{2d}} \mathbb{K} \left(\frac{Z_n^- - x}{w_n} \right) \mathbb{K} \left(\frac{Z_n - y}{w_n} \right) \right), \quad (5.22)$$

qui apparaît naturellement lorsqu'on étudie l'estimateur $\hat{h}_n(x, y)$. Précisément, on propose d'étudier sa fonction d'autocovariance afin d'obtenir le comportement asymptotique de la variance de $\hat{h}_n(x, y)$.

PROPOSITION 5.18 *Il existe deux constantes B et $b > 1$ telles que, pour tous entiers $n \geq k$,*

$$|\text{Cov}(A_k, A_n)| \leq \frac{\|\mathbb{K}\|_\infty^4 B}{w_n^{4d}} b^{k-n} (1 + b^{-k}). \quad (5.23)$$

En particulier, pour $k = n$ et en utilisant que $b^{-n} \leq 1$, on a

$$\text{Var}(A_n) \leq \frac{2\|\mathbb{K}\|_\infty^4 B}{w_n^{4d}}. \quad (5.24)$$

PREUVE. On a

$$\begin{aligned} & \text{Cov}(A_k, A_n) \\ &= \frac{\|\mathbb{K}\|_\infty^4}{w_n^{2d} w_k^{2d}} \text{Cov} \left(\frac{1}{\|\mathbb{K}\|_\infty^2} \mathbb{K} \left(\frac{Z_k^- - x}{w_k} \right) \mathbb{K} \left(\frac{Z_k - y}{w_k} \right), \frac{1}{\|\mathbb{K}\|_\infty^2} \mathbb{K} \left(\frac{Z_n^- - x}{w_n} \right) \mathbb{K} \left(\frac{Z_n - y}{w_n} \right) \right), \end{aligned}$$

où les deux termes intervenant dans la covariance sont bornés par 1. On applique le théorème 16.1.5 de [67] avec $V = 1$, $\Phi = (Z^-, Z)$,

$$g = \frac{1}{\|\mathbb{K}\|_\infty^2} \mathbb{K} \left(\frac{\cdot - x}{w_n} \right) \mathbb{K} \left(\frac{\cdot - y}{w_n} \right) \quad \text{et} \quad h = \frac{1}{\|\mathbb{K}\|_\infty^2} \mathbb{K} \left(\frac{\cdot - x}{w_k} \right) \mathbb{K} \left(\frac{\cdot - y}{w_k} \right).$$

Les conditions du théorème sont vérifiées d'après (5.4) et (5.16). On obtient

$$|\text{Cov}(A_k, A_n)| \leq \frac{\|\mathbb{K}\|_\infty^4}{w_n^{2d} w_k^{2d}} B b^{k-n} (1 + b^{-k}).$$

Avec $1/w_k^{2d} \leq 1/w_n^{2d}$, cela montre (5.23). □

Dans le résultat suivant, on donne une borne de la variance de $\hat{h}_n(x, y)$. C'est un corollaire de la proposition 5.18.

COROLLAIRE 5.19 *Soit n un entier. On a*

$$\text{Var} \left(\hat{h}_n(x, y) \right) \leq \frac{8}{nw_{n+1}^{4d}} \frac{\|K\|_{\infty}^4 B}{1 - b^{-1}}. \quad (5.25)$$

Par conséquent, la variance tend vers 0 lorsque $4\beta d < 1$ (on rappelle que $w_n = w_1 n^{-\beta}$).

PREUVE. Cette inégalité est une conséquence de (5.23) établie dans la proposition 5.18. En effet, d'après l'expression de A_n (5.22) et celle de $\hat{h}_n(x, y)$ (5.18), on a

$$\begin{aligned} \text{Var} \left(\hat{h}_n(x, y) \right) &= \frac{2}{n^2} \sum_{k=1}^{n+1} \sum_{l=k}^{n+1} \text{Cov}(A_l, A_k) \\ &\leq \frac{2}{n^2} \|K\|_{\infty}^4 B \sum_{k=1}^{n+1} \sum_{l=k}^{n+1} b^{k-l} (1 + b^{-k}) w_l^{-4d}. \end{aligned}$$

En utilisant que $b^{-k} \leq 1$ et $w_l^{-4d} \leq w_{n+1}^{-4d}$,

$$\begin{aligned} \text{Var} \left(\hat{h}_n(x, y) \right) &\leq \frac{4}{n^2 w_{n+1}^{4d}} \|K\|_{\infty}^4 B \sum_{k=1}^{n+1} b^k \sum_{l=k}^{n+1} b^{-l} \\ &\leq \frac{4}{n^2 w_{n+1}^{4d}} \|K\|_{\infty}^4 B \sum_{k=1}^{n+1} b^k \frac{b^{-k}}{1 - b^{-1}} \\ &\leq \frac{8}{nw_{n+1}^{4d}} \frac{\|K\|_{\infty}^4 B}{1 - b^{-1}}, \end{aligned}$$

avec $(n+1)/n \leq 2$. □

On peut maintenant montrer la consistance de notre estimateur de $h(x, y)$.

PROPOSITION 5.20 *Soit $(x, y) \in E^2$. On choisit $w_1 \delta < \text{dist}(x, \partial E)$ et $8\beta d < 1$. Alors,*

$$\hat{h}_n(x, y) \xrightarrow{\text{p.s.}} h(x, y),$$

quand n tend vers l'infini.

PREUVE. D'après la proposition 5.17, il suffit de prouver que

$$Y_n = \left(\widehat{h}_n(x, y) - \mathbb{E} \left[\widehat{h}_n(x, y) \right] \right)^2 \quad (5.26)$$

converge presque sûrement vers 0. Dans la suite de la preuve, on montre qu'il existe une variable aléatoire Y telle que $Y_n \xrightarrow{p.s.} Y$. Puisque la suite (Y_n) tend vers 0 dans \mathbb{L}^1 (on peut remarquer que $\mathbb{E}[Y_n] = \text{Var}(\widehat{h}_n(x, y))$, et utiliser le corollaire 5.19), alors Y vaut 0 presque sûrement et cela entraîne le résultat recherché. Afin de montrer la convergence presque sûre de (Y_n) , on utilise le lemme de Van Ryzin (voir [85]). D'après ce résultat, si la suite (Y_n) vérifie les conditions suivantes,

(i) $Y_n \geq 0$ p.s.,

(ii) $\mathbb{E}[Y_1] < +\infty$,

(iii) $\mathbb{E}[Y_{n+1} | \mathcal{S}_n] \leq Y_n + Y'_n$ p.s., où $\mathcal{S}_n = \sigma(Y_1, \dots, Y_n)$ et Y'_n est \mathcal{S}_n -mesurable,

(iv) $\sum_{n \geq 1} \mathbb{E}[|Y'_n|] < +\infty$,

alors Y_n converge presque sûrement vers une variable aléatoire Y . Dans notre contexte, les points (i) et (ii) sont évidemment vérifiés. On pose

$$V_n = \frac{1}{n} \sum_{k=1}^{n+1} (A_k - \mathbb{E}[A_k]).$$

Par (5.18), (5.22) et (5.26), on a $V_n^2 = Y_n$. On a aussi la relation de récurrence

$$V_n = \frac{A_{n+1} - \mathbb{E}[A_{n+1}] + (n-1)V_{n-1}}{n}.$$

En élevant au carré, on a

$$\begin{aligned} Y_n &= \left(\frac{n-1}{n} \right)^2 Y_{n-1} + \frac{(A_{n+1} - \mathbb{E}[A_{n+1}])^2 + 2(n-1)V_{n-1}(A_{n+1} - \mathbb{E}[A_{n+1}])}{n^2} \\ &\leq Y_{n-1} + \frac{(A_{n+1} - \mathbb{E}[A_{n+1}])^2 + 2(n-1)V_{n-1}(A_{n+1} - \mathbb{E}[A_{n+1}])}{n^2}. \end{aligned}$$

Finalement, $\mathbb{E}[Y_n | \mathcal{S}_{n-1}] \leq Y_{n-1} + Y'_{n-1}$, où

$$Y'_{n-1} = \frac{1}{n^2} \mathbb{E} \left[(A_{n+1} - \mathbb{E}[A_{n+1}])^2 + 2(n-1)V_{n-1}(A_{n+1} - \mathbb{E}[A_{n+1}]) \middle| \mathcal{S}_{n-1} \right]. \quad (5.27)$$

Donc, (iii) est vérifié. Il reste à vérifier le point (iv). Par (5.27) et l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}\mathbb{E}[|Y'_{n-1}|] &\leq \frac{1}{n^2} \text{Var}(A_{n+1}) + \frac{2(n-1)}{n^2} \mathbb{E}[|V_{n-1}| |A_{n+1} - \mathbb{E}[A_{n+1}]|] \\ &\leq \frac{1}{n^2} \text{Var}(A_{n+1}) + \frac{2}{n} \sqrt{\mathbb{E}[Y_{n-1}] \text{Var}(A_{n+1})} \\ &\leq \frac{1}{n^2} \text{Var}(A_{n+1}) + \frac{2}{n} \sqrt{\text{Var}(\widehat{h}_{n-1}(x, y)) \text{Var}(A_{n+1})}.\end{aligned}$$

Donc, par (5.24) et (5.25), il existe c_1 et c_2 tels que

$$\begin{aligned}\mathbb{E}[|Y'_{n-1}|] &\leq \frac{c_1}{n^2 w_n^{4d}} + \frac{c_2}{n^{3/2} w_n^{4d}} \\ &\leq \frac{c_1}{w_1 n^{2-4\beta d}} + \frac{c_2}{w_1 n^{3/2-4\beta d}}.\end{aligned}$$

Ainsi, $\sum \mathbb{E}[|Y'_n|]$ est une série convergente dès que $8\beta d < 1$. \square

Notre résultat principal de convergence est établi dans le théorème suivant. L'estimateur récursif de $q(x, y)$ que nous considérons peut s'écrire,

$$\widehat{q}_n(x, y) = \frac{\sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j - y}{w_j}\right)}{\sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right)},$$

où le couple (x, y) est dans E^2 , $v_j = v_1 j^{-\alpha}$, $w_j = w_1 j^{-\beta}$, avec $\alpha, \beta > 0$, et K est un noyau satisfaisant les hypothèses 5.11. Il s'agit du rapport des estimateurs $\widehat{h}_n(x, y)$ donné par (5.18) et $\widehat{p}_n(x)$ donné par (5.11) : $\widehat{q}_n(x, y) = \widehat{h}_n(x, y)/\widehat{p}_n(x)$.

THÉORÈME 5.21 *On choisit v_1 et w_1 tels que $\max(v_1, w_1)\delta < \text{dist}(x, \partial E)$. Si $p(x) > 0$, $\alpha d < 1$ et $8\beta d < 1$, alors, quand n tend vers l'infini,*

$$\widehat{q}_n(x, y) \xrightarrow{\text{p.s.}} q(x, y).$$

PREUVE. D'après (5.17), si $p(x) > 0$, on peut écrire $q(x, y) = h(x, y)/p(x)$. Le résultat est alors un corollaire des propositions 5.13 et 5.20. \square

5.5 Illustration numérique

Le but de cette partie est d'illustrer le comportement de notre estimateur récursif à travers des simulations dans le cas unidimensionnel. Plus précisément, on s'intéresse d'abord à un exemple simple pour lequel les transitions sont exponentiellement distribuées, puis au célèbre processus TCP window size.

5.5.1 Premières simulations

On considère un PDMP (X_t) défini sur $E =]0, 1[$, dont le point de départ se situe au milieu de E , $X_0 = 0.5$. Les caractéristiques locales de (X_t) sont les suivantes. Le flot Φ vérifie $\Phi_x(t) = x + t$ pour tous $x \in E$ et $t \in \mathbb{R}$. De plus, le taux de saut λ est choisi constant égal à C_1 . Finalement, pour $x \in \bar{E}$, $Q(x, \cdot)$ est la loi exponentielle de paramètre $C_2 + x$, restreinte à E . Toutes les hypothèses posées dans ce chapitre sont bien sûr vérifiées dans cet exemple. En particulier, le noyau de transition Q admet une densité q par rapport à la mesure de Lebesgue.

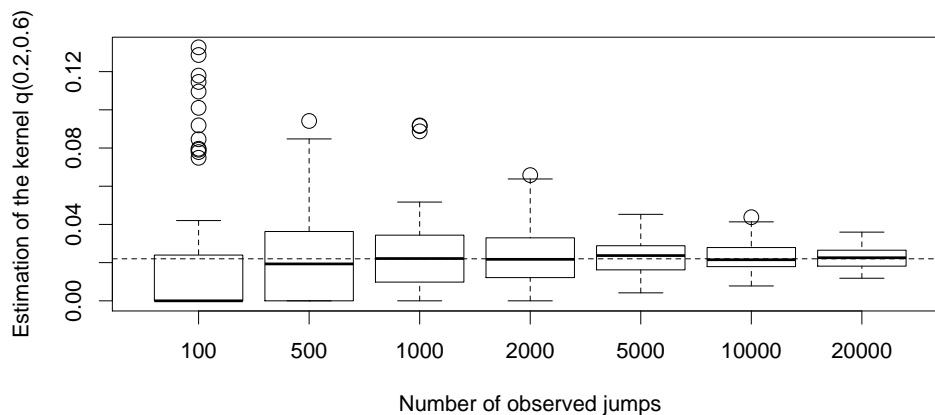


FIGURE 5.2 – Diagrammes en boîte, sur 100 répliques, de notre estimation de $q(x, y)$, avec $x = 0.2$, $y = 0.6$ et $C_1 = C_2 = 10$, à partir de différents nombres de sauts observés. Lorsque $\hat{p}_n(x) = 0$, on remplace $\hat{q}_n(x, y)$ par 0. Par conséquent, il y a beaucoup de zéros parmi les 100 répliques pour n petit.

Dans la suite, on se focalise sur l'estimation non paramétrique de la densité conditionnelle

q associée à Q . Pour tout couple $(x, y) \in \bar{E} \times E$, la fonction d'intérêt $q(x, y)$ vérifie,

$$q(x, y) = \frac{1}{K_x}(C_2 + x) \exp(-(C_2 + x)y),$$

où K_x est la constante de normalisation. On choisit le noyau d'Epanechnikov K donné, pour tout $x \in \mathbb{R}$, par

$$K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{(|x| \leq 1)}.$$

Ce noyau vérifie les hypothèses imposées dans la partie théorique de ce chapitre. Finalement, on choisit les paramètres α et β ,

$$\alpha = \beta = \frac{1}{8} - 10^{-2}.$$

La figure 5.2 montre la distribution empirique à partir de 100 répliques de l'estimation $\hat{q}_n(x, y)$, avec $x = 0.2$, $y = 0.6$ et $C_1 = C_2 = 10$, à partir de différents nombres de sauts observés. Sur les échantillons petits, notre procédure d'estimation est peu satisfaisante. Néanmoins, notre estimation apparaît comme très performante lorsque n est suffisamment grand, surtout au-delà de 10 000 sauts observés. De plus, la courbe $q(x, y)$, avec $x = 0.5$ et $0.1 \leq y \leq 0.9$, et son estimation à partir de 20 000 sauts sont très proches (voir la figure 5.3). L'estimation est donnée seulement sur l'intervalle $[0.1, 0.9] \subset E$, puisque la convergence presque sûre n'a été établie que sur l'intérieur de l'espace d'état.

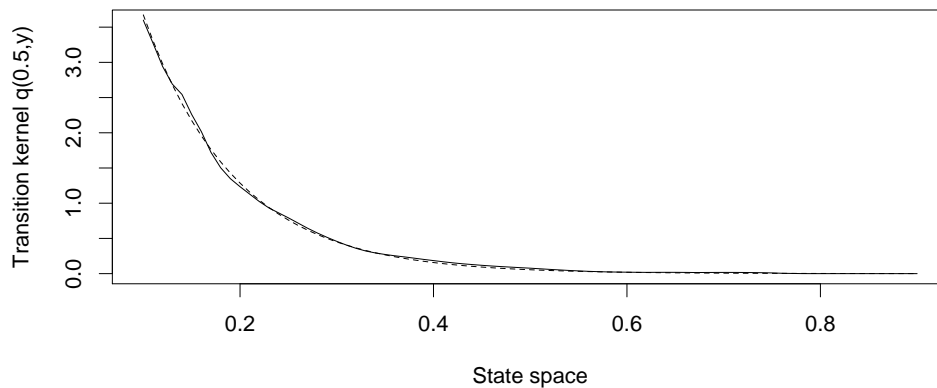


FIGURE 5.3 – Estimation de $q(x, y)$, $x = 0.5$, pour $0.1 \leq y \leq 0.9$ à partir de 20 000 sauts observés, avec $C_1 = C_2 = 10$. La densité théorique est en pointillés, tandis que son estimation est en ligne pleine.

5.5.2 Le processus TCP window size

Le processus TCP window size apparaît naturellement dans la modélisation du célèbre Transmission Control Protocol utilisé pour la transmission de données sur le réseau Internet. Pour ce PDMP, les trajectoires déterministes sont linéaires et tout l'aléa provient des temps de saut. L'évolution du processus TCP window size (X_t) sur $E =]0, +\infty[$ est gouvernée par ses caractéristiques Φ , λ et Q . Le flot Φ est linéaire, $\Phi_x(t) = x + t$ pour tous $x \in E$ et $t \in \mathbb{R}$. Le taux de saut λ vérifie $\lambda(x) = x$. Finalement, le noyau Q est donné, pour tout couple $(x, y) \in E^2$, par $Q(x, dy) = \delta_{\lfloor x/2 \rfloor}(dy)$. On peut diriger le lecteur intéressé vers [18] pour une étude sur le comportement en temps long de ce processus.

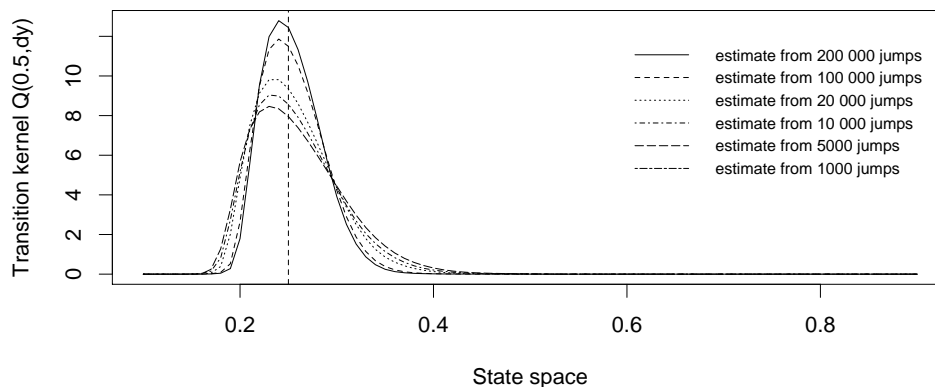


FIGURE 5.4 – Estimation de $Q(x, dy) = \delta_{\lfloor x/2 \rfloor}(dy)$, avec $x = 0.5$, à partir de différents nombres de sauts observés. La ligne verticale correspond à la valeur $x/2 = 0.25$.

Comme on l'a dit, le noyau de transition Q n'admet pas de densité par rapport à la mesure de Lebesgue. Par conséquent, notre approche n'est pas adaptée pour traiter l'estimation de Q dans ce problème. On applique néanmoins notre procédure d'estimation dans ce cadre inapproprié afin de se rendre compte des limites de notre travail. Le noyau K et les constantes α et β ont été choisis comme dans l'exemple précédent. On peut observer sur la figure 5.4 que les résultats numériques paraissent acceptables sans être pleinement satisfaisants : l'estimation se concentre lentement autour de la valeur cherchée.

5.6 Conclusion du chapitre

Dans ce chapitre, on a montré que l'estimateur à noyau récursif qu'il est naturel de considérer converge presque sûrement vers le noyau de Markov du PDMP, sous des conditions de régularité du processus et d'ergodicité. La difficulté principale apparaît dans l'estimation de la loi invariante du couple (Z_n^-, Z_n) , dont le noyau n'est pas à densité, même sur l'intérieur de l'espace d'état. La méthode mise en œuvre pour estimer la loi invariante de (Z_n^-) (le dénominateur) est fondée sur des résultats classiques de martingales à temps discret, et permettrait d'obtenir soit la convergence uniforme, soit des vitesses de convergence sous des conditions de Lipschitz mixing grâce au théorème central limite pour les martingales. Néanmoins, la technique plus universelle utilisée pour étudier le numérateur ne semble pas suffisamment souple pour obtenir ce même type de résultat pour l'estimateur du numérateur. Les perspectives de travail sur ce sujet (convergence uniforme, théorème central limite) semblent donc très ouvertes.

CONCLUSION ET PERSPECTIVES

L'objectif principal de cette thèse était de proposer des estimateurs pour les lois conditionnelles des deux aléas intervenant dans la dynamique des processus markoviens déterministes par morceaux. Dès le début de mes réflexions avec François Dufour et Anne Gégout-Petit, nous avons décidé de travailler dans un cadre non paramétrique et pour des processus observés en temps long.

Dans ce contexte, on a donné des estimateurs non paramétriques des deux caractéristiques gouvernant l'aléa d'un PDMP : la densité conditionnelle des temps inter-sauts f dans le chapitre 3, et le noyau de transition Q dans le chapitre 5. On a également montré des résultats de convergence et illustré ces résultats par des simulations sur différents exemples. Notre étude de l'estimateur de f s'appuie en partie sur le travail réalisé dans le chapitre 2 sur l'estimation du taux de saut pour une certaine classe de processus de renouvellement. Enfin, on a présenté dans le chapitre 4 un travail indépendant sur une méthode d'approximation numérique pour la méthode SIR par quantification optimale en norme \mathbb{L}^p .

Même si nous avons donné des méthodes d'estimation pour les deux caractéristiques qui nous intéressaient, il reste bien sûr de nombreuses pistes de recherche à explorer. J'en détaille quelques unes dans la suite.

ESTIMATION DES LOIS DES TEMPS INTER-SAUTS

Sur ce sujet, un point méthodologique délicat sur lequel il faut encore travailler est la suite des fenêtres (β_n) . D'un point de vue théorique, mais aussi numérique, il serait tout à fait pertinent de déterminer explicitement une suite adéquate. On pourrait même réfléchir au choix optimal d'une telle fenêtre. Dans le même ordre d'idée, le choix de la partition de l'espace d'état et le compromis entre les erreurs déterministe et stochastique réalisées doivent être discutés. De plus, nous avons choisi d'utiliser des méthodes de lissage à noyau dans la direction temporelle, mais d'autres techniques existent : par exemple, on peut décomposer la fonction d'intérêt dans une base et estimer ses coordonnées. Enfin, il serait fort intéressant d'étudier l'estimation en un point x au travers d'une suite décroissante

(A_n) d'ensembles contenant x . Il faudrait alors s'intéresser précisément à la vitesse de décroissance de cette suite, ainsi qu'aux liens avec les fenêtres de lissage.

ESTIMATION DES TRANSITIONS

On a montré la convergence ponctuelle presque sûre de l'estimateur récursif naturel du noyau de transition. Des pistes à envisager sont l'étude de la convergence uniforme et la mise en œuvre d'un théorème central limite. Des conditions de Lipschitz mixing permettent d'obtenir des convergences en loi vers des gaussiennes pour l'estimation des lois invariantes. Mais c'est l'étude du rapport des estimateurs qui rend difficile, mais aussi attrayante, la suite des investigations sur ce sujet.

Enfin, même si nous avons pris le parti d'étudier des processus très généraux, il serait intéressant d'étudier des problèmes d'estimation pour certains PDMP, pour lesquels le flot ou les transitions sont donnés explicitement et dépendent par exemple d'un paramètre. Dans ce type de problème, on peut sans aucun doute donner des résultats fins de convergence des estimateurs. Précisément, on peut s'intéresser par exemple à l'estimation semi-paramétrique pour le processus TCP window size. Les sauts sont des masses de Dirac en une certaine proportion θ de la valeur du processus juste avant le saut. On peut alors estimer à la fois le paramètre réel θ et la loi conditionnelle des temps inter-sauts. Alors qu'on n'a jamais utilisé la forme du flot dans l'estimation des temps de saut dans les travaux présentés dans ce manuscrit, la structure très particulière des trajectoires déterministes pour le processus TCP window size permet de relier de manière évidente le temps et la variable spatiale.

Un autre sujet attrayant, mais à plus long terme, est de proposer un test pour vérifier si le taux de saut λ d'un PDMP dépend réellement de l'espace d'état. On imagine facilement les nombreuses applications pratiques d'un tel résultat en statistique. Dans la suite, on pourrait également chercher une procédure permettant de tester l'égalité des taux de saut de deux PDMP indépendants. Mais tout cela fera sans doute partie d'une autre aventure!

BIBLIOGRAPHIE

- [1] O. O. AALEN : *Statistical inference for a family of counting processes*. ProQuest LLC, Ann Arbor, MI, 1975. Thesis (Ph.D.)—University of California, Berkeley.
- [2] O. O. AALEN : Weak convergence of stochastic integrals related to counting processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 38(4):261–277, 1977.
- [3] O. O. AALEN : Nonparametric inference for a family of counting processes. *Ann. Statist.*, 6(4):701–726, 1978.
- [4] O. O. AALEN, P. K. ANDERSEN, Ø. BORGAN, R. D. GILL et N. KEIDING : History of applications of martingales in survival analysis. *J. Électron. Hist. Probab. Stat.*, 5(1):28, 2009.
- [5] P. K. ANDERSEN, Ø. BORGAN, R. D. GILL et Niels KEIDING : *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.
- [6] T. AVEN et U. JENSEN : *Stochastic models in reliability*, volume 41 de *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999.
- [7] R. AZAÏS : A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process. *Preprint arXiv :1211.5579*, 2012.
- [8] R. AZAÏS, R. COUDRET et G. DURRIEU : A hidden renewal model for monitoring aquatic systems biosensors. *Preprint*, 2013.
- [9] R. AZAÏS, F. DUFOUR et A. GÉGOUT-PETIT : Nonparametric estimation of the conditional density of the inter-jumping times for piecewise-deterministic Markov processes. *Preprint arXiv :1202.2212v2*, 2012.
- [10] R. AZAÏS, F. DUFOUR et A. GÉGOUT-PETIT : Nonparametric estimation of the jump rate for nonhomogeneous marked renewal processes. *To appear in Annales de l'Institut Henri Poincaré*, 2013.
- [11] R. AZAÏS, A. GÉGOUT-PETIT et J. SARACCO : Optimal quantization applied to sliced inverse regression. *J. Statist. Plann. Inference*, 142(2):481–492, 2012.

- [12] A. K. BASU et D. K. SAHOO : On Berry-Esseen theorem for nonparametric density estimation in Markov sequences. *Bull. Inform. Cybernet.*, 30(1):25–39, 1998.
- [13] M. BENAÏM et N. EL KAROUI : *Promenade aléatoire, Chaînes de Markov et simulations ; martingales et stratégie*. Les éd. de l'École polytechnique, Palaiseau, Paris, 2004.
- [14] J. BERAN : Nonparametric regression with randomly censored survival data. *Technical report, Dept. Statist. Univ. California, Berkeley*, 1981.
- [15] A. BRANDEJSKY, B. de SAPORTA et F. DUFOUR : Numerical method for expectations of piecewise-deterministic markov processes. *CAMCoS*, 7(1):63–104, 2012.
- [16] A. BRANDEJSKY, B. DE SAPORTA et F. DUFOUR : Numerical methods for the exit time of a piecewise-deterministic Markov process. *Adv. in Appl. Probab.*, 44(1):196–225, 2012.
- [17] E. BUCKWAR et M.G. RIEDLER : An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution. *Journal of mathematical biology*, 63(6):1051–1093, 2011.
- [18] D. CHAFAÏ, F. MALRIEU et K. PAROUX : On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8):1518–1534, 2010.
- [19] K. L. CHUNG : *A course in probability theory*. Academic Press, New York-London, second édition, 1974. Probability and Mathematical Statistics, Vol. 21.
- [20] S. J. M. CLÉMENÇON : Adaptive estimation of the transition density of a regular Markov chain. *Math. Methods Statist.*, 9(4):323–357, 2000.
- [21] F. COMTE, S. GAÏFFAS et A. GUILLOUX : Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 47(4):1171–1196, 2011.
- [22] R. D. COOK : Principal hessian directions revisited. *JASA*, 93(441):84–94, 1998.
- [23] R. D. COOK : Save : A method for dimension reduction and graphics in regression. *Com in Stat - Theory Methods*, 29:2109–2121, 2000.
- [24] O. L V COSTA : Impulse control of piecewise-deterministic processes via linear programming. *Automatic Control, IEEE Transactions on*, 36(3):371–375, 1991.
- [25] O. L. V. COSTA et F. DUFOUR : Stability and ergodicity of piecewise deterministic markov processes. *SIAM J. Control and Optimization*, 47(2):1053–1077, 2008.
- [26] O. L. V. COSTA et C. A. B. RAYMUNDO : Impulse and continuous control of piecewise deterministic markov processes. *Stochastics and Stochastics Reports*, 70(1-2):75–107, 2000.

- [27] R. COUDRET, G. DURRIEU et J. SARACCO : Comparison of kernel density estimators with assumption on number of modes. *Communication in Statistics - Simulation and Computation*, In press, 2013.
- [28] D. R. COX : Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972.
- [29] A. CRUDU, A. DEBUSSCHE et O. RADULESCU : Hybrid stochastic simplifications for multiscale gene networks. *BMC Systems Biology*, 3(1):89, 2009.
- [30] D. M. DABROWSKA : Nonparametric regression with censored survival time data. *Scand. J. Statist.*, 14(3):181–197, 1987.
- [31] M. H. A. DAVIS : *Markov models and optimization*, volume 49 de *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.
- [32] M.H.A. DAVIS, M.A.H. DEMPSTER, S.P. SETHI et D. VERMES : Optimal capacity expansion under uncertainty. *Advances in Applied Probability*, 19(1):156–176, 1987.
- [33] B. de SAPORTA et F. DUFOUR : Numerical method for impulse control of piecewise deterministic Markov processes. *Automatica J. IFAC*, 48(5):779–793, 2012.
- [34] B. de SAPORTA, F. DUFOUR, C. ELEGBEDE et H. ZHANG : Arrêt optimal pour la maintenance prédictive. *In Proceedings of $\lambda\mu$* , La Rochelle, France, 2010.
- [35] B. de SAPORTA, F. DUFOUR et K. GONZALES : Numerical method for optimal stopping of piecewise deterministic processes. *Ann. Appl. Probab.*, 20(5):1607–1637, 2010.
- [36] B. de SAPORTA, F. DUFOUR, H. ZHANG et C. ELEGBEDE : Optimal stopping for the predictive maintenance of a structure subject to corrosion. *Journal of Risk and Reliability*, 226 (2):169–181, 2012.
- [37] P. DOUKHAN et M. GHINDÈS : Estimation de la transition de probabilité d’une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d’ordre 1. *Stochastic Process. Appl.*, 15(3):271–293, 1983.
- [38] M. DOUMIC, M. HOFFMANN, N. KRELL et L. ROBERT : Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Preprint*, 2012.
- [39] M. DOUMIC, M. HOFFMANN, P. REYNAUD-BOURET et V. RIVOIRARD : Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis*, 50(2):925–950, décembre 2011.
- [40] N. DUAN et K.-C. LI : Slicing regression : a link-free regression method. *Ann. Appl. Stat.*, 19(2):505–530, 1991.
- [41] M. DUFLO : *Random iterative models*. Applications of Mathematics. Springer-Verlag, Berlin, 1997.

- [42] F. DUFOUR et O. L. V. COSTA : Stability of piecewise-deterministic markov processes. *SIAM J. Control Optim.*, 37(5):1483–1502, août 1999.
- [43] B. EFRON : *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38. SIAM, Philadelphia, 1982.
- [44] A. FAGGIONATO, D. GABRIELLI et M.R. CRIVELLARI : Averaging and large deviation principles for fully-coupled piecewise deterministic markov processes and applications to molecular motors. *Markov Processes and Related Fields*, 16(3):497–548, 2010.
- [45] A. FAGGIONATO, D. GABRIELLI et M. RIBEZZI CRIVELLARI : Non-equilibrium thermodynamics of piecewise deterministic Markov processes. *J. Stat. Phys.*, 137(2):259–304, 2009.
- [46] J. A. FILAR et A. HAURIE : A two-factor stochastic production model with two time scales. *Automatica*, 37(10):1505–1513, octobre 2001.
- [47] J. FONTBONA, H. GUÉRIN et F. MALRIEU : Quantitative estimates for the long time behavior of an ergodic variant of the telegraph process. *Advances in Applied Probability*, 44(4):977–994, 2012.
- [48] A. GANNOUN, S. GIRARD, C. GUINOT et J. SARACCO : Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46(1):103–122, 2004.
- [49] A. GENADOT et M. THIEULLEN : Averaging for a fully coupled piecewise-deterministic markov process in infinite dimensions. *Advances in Applied Probability*, 44(3):749–773, 2012.
- [50] S. GRAPH et H. LUSCHGY : *Foundations of quantization for random vectors*, volume 1730 de *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000.
- [51] O. HERNÁNDEZ-LERMA, S. O. ESPARZA et B. S. DURAN : Recursive nonparametric estimation of nonstationary Markov processes. *Bol. Soc. Mat. Mexicana (2)*, 33(2):57–69, 1988.
- [52] O. HERNÁNDEZ-LERMA et J.-B. LASSERRE : Further criteria for positive Harris recurrence of Markov chains. *Proc. Amer. Math. Soc.*, 129(5):1521–1524, 2001.
- [53] O. HERNÁNDEZ-LERMA et J.-B. LASSERRE : *Markov chains and invariant probabilities*, volume 211 de *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2003.
- [54] J. HU, W.C. WU et S.S. SASTRY : *Modeling subtilin production in bacillus subtilis using stochastic hybrid systems*. In R. Alur and G.J. Pappas, editors, *Hybrid Systems : Computation and Control*, number 2993 in LNCS, Springer-Verlag, Berlin, 2004.

- [55] M. JACOBSEN : *Statistical analysis of counting processes*, volume 12 de *Lecture Notes in Statistics*. Springer-Verlag, New York, 1982.
- [56] M. JACOBSEN : *Point Process Theory and Applications : Marked Point and Piecewise Deterministic Processes*. Probability and its Applications. Birkhäuser, Boston-Basel-Berlin, 2006.
- [57] L. J. JOU et C. M. LIAO : A dynamic artificial clam (*corbicula fluminea*) allows parcimony on-line measurement of waterborne metals. *Environmental Pollution*, 144:172–183, 2006.
- [58] O. KALLENBERG : *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second édition, 2002.
- [59] A. KLENKE : *Probability theory*. Universitext. Springer-Verlag London Ltd., London, 2008. A comprehensive course.
- [60] C. LACOUR : Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(5):571–597, 2007.
- [61] C. LACOUR : Nonparametric estimation of the stationary density and the transition density of a Markov chain. *Stochastic Process. Appl.*, 118(2):232–260, 2008.
- [62] G. LI et H. DOSS : An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.*, 23(3):787–823, 1995.
- [63] E. LIEBSCHER : Density estimation for Markov chains. *Statistics*, 23(1):27–48, 1992.
- [64] T. MARTINUSSEN et T. H. SCHEIKE : *Dynamic regression models for survival data*. Statistics for Biology and Health. Springer, New York, 2006.
- [65] E. MASRY et L. GYÖRFI : Strong consistency and rates for recursive probability density estimators of stationary processes. *J. Multivariate Anal.*, 22(1):79–93, 1987.
- [66] I. W. MCKEAGUE et K. J. UTIKAL : Inference for a nonlinear counting process regression model. *Ann. Statist.*, 18(3):1172–1187, 1990.
- [67] S. MEYN et R. L. TWEEDIE : *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second édition, 2009.
- [68] J.-Y. OUVRARD : *Probabilités 2*. Deuxième édition, Master, Agrégation. Cassini, 2004.
- [69] G. PAGÈS : A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 89(1):1–38, 1998.
- [70] G. PAGÈS et J.C. FORT : Asymptotics of optimal quantizers for some scalar distributions. *J. Computational & Applied Mathematics*, 146(2):253–275, 2002.

- [71] G. PAGÈS, H. PHAM et J. PRINTEMS : Optimal quantization methods and applications to numerical problems in finance. *In Handbook of computational and numerical methods in finance*, pages 253–297. Birkhäuser Boston, Boston, MA, 2004.
- [72] G. PAGÈS et J. PRINTEMS : Optimal quadratic quantization for numerics : the Gaussian case. *Monte Carlo Methods Appl.*, 9(2):135–165, 2003.
- [73] J.N. PIERCE : Asymptotic quantizing error for unbounded random variables. *IEEE Trans. Inform. Theory*, 16:101–112, 1970.
- [74] H. RAMLAU-HANSEN : Smoothing counting process intensities by means of kernel functions. *Ann. Statist.*, 11(2):453–466, 1983.
- [75] A. ROBSON, R. WILSON et C. Garcia de LEANIZ : Mussels flexing their muscles : a new method for quantifying bivalve behavior. *Marine Biology*, 151:1195–1204, 2007.
- [76] M. ROSENBLATT : Density estimates and Markov sequences. *In Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, pages 199–213. Cambridge Univ. Press, London, 1970.
- [77] G. G. ROUSSAS : Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.*, 21:73–87, 1969.
- [78] J. SARACCO : Pooled slicing methods versus slicing methods. *Com in Stat - Simulation Computation*, 30(3):489–511, 2001.
- [79] F. G. SCHMITT, M. DE ROSA, G. DURRIEU, M. SOW, P. CIRET, D. TRAN et J. C. MASSABUAU : Statistical analysis of bivalve high frequency microclosing behavior : scaling properties and shot noise modeling. *International Journal of Bifurcation and Chaos*, 21(12):3565–3576, 2011.
- [80] M. SOW, G. DURRIEU et L. BRIOLLAIS : Water quality assessment by means of HFNI valvometry and high-frequency data modeling. *Environmental Monitoring and Assessment*, 182(1-4):155–170, 2011.
- [81] W. STUTE : Conditional empirical processes. *Ann. Statist.*, 14(2):638–647, 1986.
- [82] D. TRAN, P. CIRET, A. CIUTAT, G DURRIEU et J.-C. MASSABUAU : Estimation of potential and limits of bivalve closure response to detect contaminants : application to cadmium. *Environmental Toxicology and Chemistry*, 22(4):914–920, 2003.
- [83] K. J. UTIKAL : Nonparametric inference for a doubly stochastic Poisson process. *Stochastic Process. Appl.*, 45(2):331–349, 1993.
- [84] K. J. UTIKAL : Nonparametric inference for Markovian interval processes. *Stochastic Process. Appl.*, 67(1):1–23, 1997.

- [85] J. VAN RYZIN : On strong consistency of density estimates. *Ann. Math. Statist.*, 40:1765–1772, 1969.
- [86] S. YAKOWITZ : Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal.*, 30(1):124–136, 1989.
- [87] P.L ZADOR : *Development and evaluation of procedures for quantizing multivariate distributions*. Thèse de doctorat, Stanford University, Stanford, CA 94305, 1963.