

Thèse présentée pour obtenir le titre de  
**DOCTEUR DE L'ÉCOLE NATIONALE  
DES PONTS ET CHAUSSÉES**

Spécialité : Mathématiques appliquées

par

**Julien GUYON**

**Modélisation probabiliste  
en finance et en biologie.  
Théorèmes limites et applications**

Thèse soutenue le 19 juillet 2006 devant le jury composé de :

Jean-François DELMAS	<i>Directeur de thèse</i>
Nicole EL KAROUI	<i>Examineur</i>
Bernard LAPEYRE	<i>Examineur</i>
Sylvie MÉLÉARD	<i>Rapporteur</i>
Philip PROTTER	<i>Rapporteur</i>
Bernard PRUM	<i>Examineur</i>



## Remerciements

Une thèse est une course d'endurance. Un effort solitaire. Au départ, la ligne d'arrivée est on ne sait pas bien où, loin. Rien ne sert d'être un sprinter, mais étrangement il faut sans cesse courir. Tel un marathonien, on arrive au bout en persévérant, en puisant dans ses réserves. Mais on n'y arrive pas sans l'aide et le soutien de son entraîneur-directeur de thèse, de ses coéquipiers-collègues, de sa famille, de ses amis. Qu'il m'est agréable de les remercier ici.

Mes bidons d'eau fraîche ont été les idées de mon directeur de thèse. Disponible, rapide, ouvert à la discussion, Jean-François Delmas a été décisif dans le développement du dernier chapitre de cette thèse et est à l'origine de nombreuses améliorations dans les autres. Il m'a aussi offert de porter, aux Ponts et à l'ENSTA, le costume de professeur que j'ai endossé avec le plus grand des plaisirs. Qu'à ce titre Benjamin Jourdain soit également remercié ici.

Sylvie Méléard et Philip Protter me font l'honneur d'être rapporteurs de cette thèse. Pour ce travail ingrat, ils méritent mes remerciements les plus vifs et une double ration de breuvage.

C'est Bernard Lapeyre qui m'a chaleureusement accueilli au CERMICS où j'ai pu travailler dans les meilleures conditions. A l'écoute de ses ouailles, il n'a pas été avare en conseils qui m'ont été précieux. Sa participation au jury me réjouit, tout comme celle d'un autre Bernard, Prum, spécialiste mondialement reconnu des applications des statistiques en biologie.

Qu'il m'est doux de remercier Nicole El Karoui. Bienveillante comme une mère, depuis l'heure déjà lointaine où j'étais son étudiant à l'Ecole polytechnique, elle n'a jamais compté son temps pour discuter science, pour me conseiller, m'orienter. Elle m'a à plusieurs reprises témoigné une grande confiance dont j'espère être digne, aujourd'hui comme dans le futur. À mon tour de l'assurer ici non seulement de mon profond respect mais aussi de ma franche affection.

Je souhaite aussi remercier pour les échanges scientifiques que nous avons pu avoir et les encouragements qu'ils m'ont prodigués Vlad Bally, Arturo Kohatsu-Higa, Denis Talay, Damien Lambertson, Gilles Pagès, Jean Jacod, Jacques Printems, Christophe Michel, Joe Bonnaud, François Taddéi, Eric Stewart, Ariel Lindner, Grégory Paul, Ariane Bize, Ralph Lavolette, Aurélien Alfonsi, Simone Scotti, Jérôme Lelong et les autres.

Enfin je tiens à remercier mes responsables au Corps des Ponts et Chaussées qui m'ont permis de mener ces travaux de recherche dans d'excellentes conditions morales et matérielles.

Ma famille s'appelle Julien, pas moi, l'autre, Xavier, Simonne, Marie, Jeanne, Jonathan, Yann, Ethan, Noam, Robinson, Fabien, Jean-Baptiste, Gabrielle, Lucile,

Aude, Étienne, François... Mes amis s'appellent Yves, Sophie, Aleksandar, Claude, Emma, Anne-Louise, Marie, Xavier, Maud, Ralph, Soizic, Tina, Pierrette, Danielle, Paolo... Ils me donnent des dessins, des conseils, de la confiture, de l'amour, des randonnées alpines, de l'énergie, des tuyaux musique, des rires, des recettes, du plaisir, des suées, de l'espoir, des cours de dj-ing, de couture ou de jardinage, de la confiance, des framboises, des tomates, de la ciboulette et de la force, et tout ça n'a pas de prix, comme en marché incomplet.

## Résumé

C'est le souci d'une modélisation mathématique à la fois précise et maniable qui constitue le dénominateur commun à ces travaux de thèse. Nous nous sommes en particulier intéressés à deux champs d'application des probabilités : les marchés financiers et la biologie. Le premier chapitre détaille nos motivations. Il résume nos principaux résultats, les compare aux travaux existants et suggère des extensions possibles. Au deuxième chapitre, suite aux articles de Talay et Tubaro (1990) et Bally et Talay (1996), nous mesurons l'erreur que l'on commet lorsque l'on approche la loi de la solution d'une équation différentielle stochastique par celle de son schéma d'Euler. Sous hypothèse d'ellipticité, l'utilisation conjointe de techniques probabilistes et analytiques nous permet d'obtenir un développement limité *fonctionnel*, dans des espaces de fonctions très régulières de type noyau gaussien, du "noyau de transition" du schéma d'Euler, en fonction du pas de temps de discrétisation. Ce résultat trouve une application naturelle en mathématiques financières. Il donne la vitesse de convergence des prix, deltas et gammas d'options européennes pour une classe extrêmement large de payoffs. Il nous permet aussi de construire, au chapitre 3, dans l'analyse d'un modèle à volatilité stochastique proposé par Fouque, Papanicolaou et Sircar (2000), un algorithme d'évaluation et de couverture des options européennes dans lequel l'équilibre entre l'erreur statistique, due à l'échantillonnage "Monte-Carlo", et l'erreur de discrétisation temporelle est assuré de manière adaptative. Enfin, le dernier chapitre a pour thème le vieillissement cellulaire et est le fruit d'une coopération avec des biologistes de la Faculté de Médecine Necker à Paris. Les données expérimentales se présentent sous forme d'un arbre binaire de taux de croissance, à partir duquel nos collègues biologistes souhaitent détecter deux sous-populations. Pour expliquer ces données, nous proposons un modèle autorégressif avec bifurcation, généralisant celui proposé par Cowan et Staudte en 1986, puis construisons et implémentons des procédures permettant d'estimer des paramètres et de tester des hypothèses biologiques. Pour ce faire, nous introduisons le concept de "chaînes de Markov bifurcantes", prouvons que cette famille de processus stochastiques satisfait des théorèmes limites originaux que nous appliquons au modèle et confrontons aux données, confirmant l'intuition et les calculs préliminaires des biologistes.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Enjeux . . . . .	1
1.1.1	Modélisation en finance . . . . .	1
1.1.2	Modélisation en biologie . . . . .	2
1.2	Déroulement de la thèse . . . . .	2
1.3	Erreur faible du schéma d'Euler pour les équations différentielles stochastiques . . . . .	4
1.3.1	Motivation . . . . .	4
1.3.2	Cadre mathématique . . . . .	4
1.3.3	Le schéma d'Euler . . . . .	5
1.3.4	Erreur forte et erreur faible . . . . .	6
1.3.5	Résultats connus . . . . .	7
1.3.6	Résultats de la thèse . . . . .	8
1.3.7	Résultats voisins . . . . .	12
1.3.8	Extensions possibles . . . . .	15
1.4	Analyse d'un modèle à volatilité stochastique . . . . .	15
1.5	Théorèmes limites pour les chaînes de Markov bifurcantes . . . . .	17
1.5.1	Motivation . . . . .	17
1.5.2	La biologie . . . . .	18
1.5.3	Le modèle . . . . .	20
1.5.4	Modèles voisins . . . . .	22
1.5.5	Résultats de la thèse . . . . .	23
1.5.6	Résultats voisins . . . . .	25
1.5.7	Extensions possibles . . . . .	26
<b>2</b>	<b>Schéma d'Euler et distributions tempérées</b>	<b>27</b>
2.1	Introduction and results . . . . .	28
2.1.1	Motivation . . . . .	28
2.1.2	Previous results . . . . .	29
2.1.3	Purpose and method . . . . .	30
2.1.4	A first series of results . . . . .	31

2.1.5	Application to option pricing and hedging . . . . .	33
2.1.6	Some functional spaces . . . . .	34
2.1.7	Main results . . . . .	35
2.1.8	A preliminary result . . . . .	38
2.1.9	Organization of the paper . . . . .	40
2.2	First step: expansion for $\mathbb{E}[f(X_t^{n,x})]$ . . . . .	40
2.2.1	Operators associated with the Euler scheme . . . . .	40
2.2.2	Itô expansions . . . . .	42
2.2.3	Proof of Theorem 13 . . . . .	43
2.3	Second step: expansion for the density of $X_t^{n,x}$ . . . . .	45
2.3.1	Space convolutions . . . . .	45
2.3.2	Proof of Theorems 10 and 12 . . . . .	47
2.4	Appendix . . . . .	52
2.4.1	Kernel of $R^{n,j}$ . . . . .	52
2.4.2	Operators on $\mathcal{G}_t(\mathbb{R}^d)$ . . . . .	55
2.4.3	Moments for the Euler scheme and its derivatives . . . . .	56
<b>3</b>	<b>Un modèle à volatilité stochastique</b> . . . . .	<b>59</b>
3.1	Introduction . . . . .	59
3.1.1	Notations . . . . .	59
3.1.2	Interprétation . . . . .	61
3.1.3	D'autres modèles à volatilité stochastique . . . . .	67
3.1.4	A quoi s'attendre? . . . . .	69
3.2	Prix d'une option européenne . . . . .	70
3.2.1	EDP d'évaluation . . . . .	70
3.2.2	Interprétation probabiliste . . . . .	74
3.3	Analyse asymptotique . . . . .	76
3.3.1	Retour à la moyenne . . . . .	76
3.3.2	Le prix Black-Scholes corrigé . . . . .	79
3.3.3	Stratégies de couverture . . . . .	82
3.4	Calibration . . . . .	88
3.4.1	Le prix corrigé de l'option d'achat . . . . .	88
3.4.2	Surface de volatilité implicite . . . . .	89
3.5	Simulations numériques . . . . .	91
3.5.1	Schéma d'Euler . . . . .	91
3.5.2	Amélioration du schéma . . . . .	95
3.5.3	Mise en œuvre . . . . .	98
3.5.4	Résultats . . . . .	101
3.6	Conclusion . . . . .	113



<b>4</b>	<b>Théorèmes limites pour les chaînes de Markov bifurcantes. Application à la détection du vieillissement cellulaire</b>	<b>115</b>
4.1	Introduction . . . . .	117
4.1.1	Motivation . . . . .	117
4.1.2	The biology . . . . .	117
4.1.3	The mathematical model . . . . .	122
4.2	Bifurcating Markov chains. Limit theorems . . . . .	123
4.2.1	Definitions . . . . .	123
4.2.2	Weak law of large numbers . . . . .	125
4.2.3	Strong law of large numbers . . . . .	133
4.2.4	Central limit theorem . . . . .	135
4.3	Detection of cellular aging . . . . .	137
4.3.1	Limit theorems in the BAR model . . . . .	137
4.3.2	Estimation of the parameters . . . . .	141
4.3.3	Detection of cellular aging . . . . .	144
4.3.4	Data numerical analysis . . . . .	147



# Chapitre 1

## Introduction

C'est le souci d'une modélisation mathématique à la fois précise et maniable qui constitue le dénominateur commun à mes travaux de thèse. Je me suis en particulier intéressé à deux champs d'application des mathématiques : la finance et les sciences du vivant. Ces deux domaines étudient des systèmes aléatoires complexes décrits par des données nombreuses et bruitées. Ils exigent naturellement une expertise en Probabilités et en Statistique. Mes recherches contribuent à la modélisation des interactions dans de tels systèmes ainsi qu'au développement et à l'analyse numérique des méthodes que l'on utilise pour calculer certaines quantités-clés ou pour tirer de l'information pertinente d'importants jeux de données.

### 1.1 Enjeux

#### 1.1.1 Modélisation en finance

Les produits optionnels sont aujourd'hui utilisés de façon intensive par les entreprises, conduisant les établissements financiers à employer des techniques avancées d'analyse stochastique et numérique. Ces établissements sont demandeurs de contact avec le monde de la recherche. Cet intérêt se traduit par des collaborations avec des universitaires, la réalisation de contrats de recherche et le recrutement, au meilleur niveau, de mathématiciens appliqués dans les équipes de recherche et développement des banques. C'est une interface dynamique entre le monde académique et le monde économique, comme en atteste le grand nombre d'emplois et de stages proposés par les banques aux étudiants des filières "ingénieur".

Les mathématiques financières, domaine d'application des probabilités très actif depuis trente ans déjà, fournissent la base rationnelle permettant d'estimer et de couvrir les produits dérivés et, plus généralement, les risques financiers. Or le risque est aujourd'hui omniprésent, et sa gestion est une préoccupation majeure non seulement des banques et des compagnies d'assurance, mais aussi des entre-

prises publiques et privées, qui doivent se couvrir par exemple contre les risques de change, de variation de prix de matières premières ou de défaillance d'un partenaire financier. L'enjeu global des mathématiques financières est donc de développer les techniques quantitatives qui permettent de limiter l'exposition aux changements qui apparaissent dans un monde de plus en plus instable et déréglementé.

### 1.1.2 Modélisation en biologie

La biostatistique, domaine en plein essor, et sans doute l'un des futurs débouchés naturels de l'ingénieur, démontre que le mathématicien appliqué peut apporter beaucoup au biologiste ou au médecin. En effet, la statistique et l'informatique sont des outils indispensables pour traiter la quantité exponentiellement croissante de données que permettent les avancées technologiques modernes, et ceci est particulièrement vrai de l'ingénierie biologique, où les progrès récents de la technique sont source de questions nouvelles et d'une somme impressionnante de données parmi lesquelles il faut savoir tirer l'information pertinente. Les exemples ne manquent pas : séquençage du génôme, expériences de puces à ADN, de dynamique cellulaire, etc.

Avec Jean-François DELMAS, nous avons monté un projet en collaboration avec l'équipe de biologistes que dirige François TADDÉI à la Faculté de médecine Necker à Paris, au sein du Laboratoire de Génétique Moléculaire, Évolutive et Médicale (unité INSERM U571). Cette équipe de tout premier niveau s'est fixé pour but de détecter et de quantifier la variabilité phénotypique à génotype et environnement constants. C'est une question scientifique de toute première importance en ceci qu'elle révisé un des dogmes de la biologie : le fait que les gènes et leur régulation gouvernent directement la forme et le comportement d'un organisme, c'est-à-dire qu'étant donné une condition initiale et un environnement spécifique, la vie se déroule comme programmée, de façon essentiellement déterministe. Or F. TADDÉI et son équipe ont élaboré des expériences originales qui prouvent qu'un groupe d'organismes unicellulaires génétiquement identiques, croissant en environnement constant, peuvent donner à voir des comportements complexes et une grande palette de phénotypes individuels.

## 1.2 Déroulement de la thèse

Dans le domaine des mathématiques financières, mon analyse théorique et numérique de modèles à volatilité stochastique constitue à l'origine la réponse à une commande du Département des Marchés de Capitaux du Crédit Lyonnais - aujourd'hui Calyon - qui m'a accueilli en stage d'août 2001 à août 2002. Ces modèles décrivent l'incertitude sur la volatilité des marchés financiers et constituent donc un raffinement du célèbre modèle de Black-Scholes. Ils expliquent les smiles de volatilité

observés en pratique, smiles qui expriment en particulier le fait que le modèle de Black-Scholes sous-évalue la probabilité d'une variation brutale des cours. Leur calibration permet donc d'estimer finement, et en particulier de ne pas sous-estimer, les options exotiques dont les prix, par manque de liquidité, ne résultent pas d'un équilibre entre l'offre et la demande. Une partie de cette étude constitue aujourd'hui un chapitre des notes du cours de Master "Modèles stochastiques en finance" que dispense Nicole EL KAROUI à l'Université Pierre et Marie Curie à Paris [28].

Ce stage en salle des marchés m'a permis de réaliser l'importance que revêt en pratique l'estimation des erreurs entre des méthodes numériques et le modèle théorique pour des processus aléatoires. C'est la raison pour laquelle je me suis ensuite intéressé à l'analyse de l'erreur de discrétisation temporelle pour les équations différentielles stochastiques. Mes recherches sur des propriétés fines du schéma d'Euler, et leur application à l'estimation des deltas et gammas d'options européennes, ont pris la forme d'un article [40] paru dans *Stochastic Processes and their Applications*.

Quant à mon activité en biostatistique, plus récente, elle s'est développée, comme précisé dans la section précédente, dans le cadre d'une coopération avec des biologistes de la Faculté de Médecine Necker à Paris, et a pour thème le vieillissement cellulaire. Les données expérimentales se présentent sous forme d'un arbre binaire de taux de croissance, à partir duquel nos collègues biologistes souhaitent détecter deux sous-populations, correspondant à deux phénotypes distincts. Mon travail comme mathématicien appliqué a consisté à proposer un modèle pertinent pour ces données, puis à construire et implémenter des procédures rigoureuses permettant d'estimer des paramètres et de tester des hypothèses biologiques. Pour ce faire, j'ai introduit le concept de "chaînes de Markov bifurcantes", prouvé que cette famille de processus stochastiques satisfait des théorèmes limites originaux que j'ai appliqués ensuite à un modèle autoregressif avec bifurcation, et enfin confronté aux données. Non seulement mes résultats confirment l'intuition et les calculs préliminaires des biologistes, mais ils constituent aussi une solide base rationnelle pour mener de manière systématique les études de dynamique cellulaire, de plus en plus nombreuses. Une étude liminaire, menée collectivement au CEMRACS 2004, a été publiée en 2005 dans *ESAIM Proceedings* [42], un article, prépublié [41], a été soumis et un autre, destiné à la communauté des biologistes, est en cours de rédaction.

Nous présentons maintenant plus en détail les trois parties de la thèse, en nous efforçant de les replacer dans leur contexte, en résumant les résultats principaux, et en dégagant quelques pistes de recherche future. Nous ne suivons pas l'ordre chronologique, et commençons cette revue de détail par l'analyse de l'erreur faible du schéma d'Euler pour les équations différentielles stochastiques.

## 1.3 Erreur faible du schéma d'Euler pour les équations différentielles stochastiques

### 1.3.1 Motivation

Les modèles financiers les plus utilisés dans les banques expriment le prix d'une option comme la valeur moyenne inconnue  $\mathbb{E}[f(X_t)]$  d'une quantité  $f(X_t(\omega))$  qui dépend du temps  $t$  et du hasard  $\omega$ .  $X_t$  représente la valeur à l'instant  $t$  de l'actif sous-jacent : action, indice boursier, etc. Le processus  $X = (X_t, t \geq 0)$  est souvent modélisé par une diffusion, c'est-à-dire la solution d'une équation différentielle stochastique (EDS). Il n'est alors pas possible de calculer exactement  $\mathbb{E}[f(X_t)]$  et l'on a naturellement recours à la simulation numérique (voir à ce sujet l'exemple traité au chapitre 3). Tout d'abord une discrétisation temporelle de la dynamique de  $X$  permet de générer une variable aléatoire  $\bar{X}_t$  dont la loi est proche de celle de  $X_t$ . On applique ensuite la méthode de Monte-Carlo : la moyenne arithmétique de  $N$  copies indépendantes de la variable  $f(\bar{X}_t)$  converge vers  $\mathbb{E}[f(\bar{X}_t)]$ . Ce faisant, on commet deux erreurs : une erreur de discrétisation temporelle,  $\mathbb{E}[f(\bar{X}_t)] - \mathbb{E}[f(X_t)]$ , dite *erreur faible*, et une erreur statistique due à l'échantillonnage "Monte-Carlo". Cette dernière, bien connue, de l'ordre de  $1/\sqrt{N}$ , est quantifiée par le Théorème Central Limite. Mes travaux, détaillés au chapitre 2, concernent l'erreur faible que l'on commet lorsque l'on approche  $(X_t, t \geq 0)$  par son schéma d'Euler. Ce schéma est à la fois le plus simple à implémenter et le plus couramment utilisé. Il consiste à "geler" les coefficients de l'équation différentielle stochastique entre deux dates de discrétisation. Précisons cela.

### 1.3.2 Cadre mathématique

Fixons deux entiers  $d, r \geq 1$  et un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  sur lequel vit un mouvement brownien  $r$ -dimensionnel  $B = (B_t, t \geq 0)$ . On note  $\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t)$  la filtration engendrée par ce brownien, complétée des ensembles négligeables. On se donne une fonction  $b = (b_1, \dots, b_d) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  et une fonction  $\sigma = (\sigma_{i,j})_{1 \leq i \leq d, 1 \leq j \leq r} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times r}$ . Pour un multiindice  $\alpha \in \mathbb{N}^d$ ,  $|\alpha| = \alpha_1 + \dots + \alpha_d$  représente sa longueur et  $\partial^\alpha$  est l'opérateur différentiel  $\partial^{|\alpha|} / \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}$ . On munit  $\mathbb{R}^d$  de sa norme euclidienne  $\|\cdot\|$  et on désigne par

- $C_{\text{pol}}^\infty(\mathbb{R}^d)$  l'ensemble des fonctions  $C^\infty f : \mathbb{R}^d \rightarrow \mathbb{R}$  dont les dérivées de tous ordres sont à croissance polynomiale, i.e. telles que pour tout  $\alpha \in \mathbb{N}^d$ , il existe  $c \geq 0$  et  $q \in \mathbb{N}$  tels que pour tout  $x \in \mathbb{R}^d$ ,

$$|\partial^\alpha f(x)| \leq c(1 + \|x\|^q), \quad (1.1)$$

- $C_b^\infty(\mathbb{R}^d)$  l'ensemble des fonctions  $C^\infty f : \mathbb{R}^d \rightarrow \mathbb{R}$  dont les dérivées de tous ordres sont bornées, i.e. telles que  $\partial^\alpha f \in L^\infty(\mathbb{R}^d)$  pour tout  $\alpha \in \mathbb{N}^d$ .

On fait *dans toute la suite* l'hypothèse suivante :

(A) pour tous  $i \in \{1, \dots, d\}$  et  $j \in \{1, \dots, r\}$ ,  $b_i$  et  $\sigma_{i,j}$  appartiennent à  $C_{\text{pol}}^\infty(\mathbb{R}^d)$  et ont des dérivées premières bornées.

Pour tout  $x \in \mathbb{R}^d$ , il existe alors un unique processus  $X^x = (X_t^x, t \geq 0)$  à valeurs dans  $\mathbb{R}^d$  tel que presque sûrement (p.s.), pour tout  $t \geq 0$ ,

$$X_t^x = x + \int_0^t b(X_s^x) ds + \int_0^t \sigma(X_s^x) dB_s. \quad (1.2)$$

En effet, la condition (A) implique que  $b$  et  $\sigma$  sont lipschitziens, ce qui permet d'appliquer un théorème d'existence et d'unicité pour les EDS, par exemple le théorème 5.2.1 page 66 de [83].

### 1.3.3 Le schéma d'Euler

Soit  $T > 0$  un horizon de temps fini. On peut, sans perte de généralité, prendre  $T$  égal à 1. On cherche à estimer la loi de  $X_T^x = X_1^x$  qui en général n'est pas connue. Pour ce faire, on approche  $X^x$  par son schéma d'Euler d'ordre  $n \geq 1$ , disons  $X^{n,x}$ , défini de la façon suivante. On considère la subdivision régulière  $\mathfrak{S}_n = \{0 = t_0^n < t_1^n < \dots < t_{n-1}^n < t_n^n = 1\}$  de l'intervalle  $[0, 1]$ , i.e.  $t_k^n = k/n$ , et on pose  $X_0^n(x) = x$  et, pour tous  $k \in \{0, \dots, n-1\}$  et  $t \in [t_k^n, t_{k+1}^n]$ ,

$$X_t^{n,x} = X_{t_k^n}^{n,x} + b(X_{t_k^n}^{n,x})(t - t_k^n) + \sigma(X_{t_k^n}^{n,x})(B_t - B_{t_k^n}). \quad (1.3)$$

Le schéma d'Euler est exactement simulable. C'est une petite perturbation de  $X^x$  que l'on peut expliquer de la façon suivante :

- on part à la date 0 de la vraie valeur  $x$ ,
- sur l'intervalle  $[0, t_1^n]$ , on gèle les coefficients de l'EDS (1.2) en leurs valeurs exactes  $b(x)$  et  $\sigma(x)$  à gauche de cet intervalle, et on calcule la valeur  $X_{t_1^n}^{n,x}$  en  $t_1^n$  de la solution de cette nouvelle petite EDS,
- sur l'intervalle  $[t_1^n, t_2^n]$ , on gèle les coefficients de l'EDS (1.2) en leurs valeurs *approchées*  $b(X_{t_1^n}^{n,x})$  et  $\sigma(X_{t_1^n}^{n,x})$  à gauche de cet intervalle, et on calcule la valeur  $X_{t_2^n}^{n,x}$  en  $t_2^n$  de la solution de cette nouvelle petite EDS avec valeur initiale  $X_{t_1^n}^{n,x}(x)$  en  $t_1^n$ ,
- et ainsi de suite...

Les erreurs se propagent : l'erreur commise en  $t_{k+1}^n$  est le fruit de l'erreur en  $t_k^n$  et de l'erreur d'approximation de ce qui se passe entre  $t_k^n$  et  $t_{k+1}^n$ . A l'arrivée, quelle erreur récupère-t-on à la date 1 ? Il faut ici distinguer deux types d'erreur.

### 1.3.4 Erreur forte et erreur faible

L'*erreur forte*, ou erreur trajectorielle, est, pour chaque  $\omega \in \Omega$ , la distance entre  $X_1^x(\omega)$  et  $X_1^{n,x}(\omega)$ . Cette erreur est p.s. de l'ordre de  $1/\sqrt{n}$ , au sens suivant :

$$\forall \alpha \in \left] 0, \frac{1}{2} \right[ , \quad \mathbb{P} \left( \lim_{n \rightarrow \infty} n^\alpha \|X_1^{n,x} - X_1^x\| = 0 \right) = 1.$$

Mieux, ceci est vrai de la distance entre les "trajectoires" :

$$\forall \alpha \in \left] 0, \frac{1}{2} \right[ , \quad \mathbb{P} \left( \lim_{n \rightarrow \infty} n^\alpha \sup_{0 \leq k \leq n} \|X_{t_k^n}^{n,x} - X_{t_k}^x\| = 0 \right) = 1. \quad (1.4)$$

On peut voir sur des exemples simples que cette propriété n'est pas vraie pour  $\alpha = \frac{1}{2}$ . La propriété (1.4) s'obtient, grâce au lemme de Borel-Cantelli, comme une conséquence facile de la majoration suivante (cf [92]) :

$$\forall m \in \mathbb{N}^*, \quad \exists C_m^x \geq 0, \quad \forall n \geq 1, \quad \mathbb{E} \left[ \sup_{0 \leq k \leq n} \|X_{t_k^n}^{n,x} - X_{t_k}^x\|^{2m} \right] \leq C_m^x / n^m. \quad (1.5)$$

L'*erreur faible*, ou erreur en loi, est moralement la distance entre la loi de  $X_1^x$  et celle de  $X_1^{n,x}$ . Comme indiqué au paragraphe 1.3.1, on la mesure en pratique par la distance entre  $\mathbb{E}[f(X_1^x)]$  et  $\mathbb{E}[f(X_1^{n,x})]$ , et c'est elle qui nous intéresse ici. L'opération de moyennage ne peut que diminuer l'erreur. Pour illustrer cette affirmation, jouons un instant aux billes. Si de chaque bille noire on jette une bille rouge à distance au plus  $\delta$ , la moyenne des billes rouges sera à distance au plus  $\delta$  de la moyenne des billes noires. Mieux, si le jet est aléatoire, on peut espérer qu'à la faveur de jets dans des directions différentes, des compensations s'opèrent. Dans ce cas on s'attend à ce que les deux moyennes soient à distance plus petite que  $\delta$ .

En ce qui nous concerne, si  $f$  est lipschitzienne, les billes rouges  $f(X_1^{n,x}(\omega))$  sont jetées sur la droite réelle à distance d'ordre  $1/\sqrt{n}$  des billes noires  $f(X_1^x(\omega))$ . L'erreur en loi

$$\Delta_1^n f(x) = \mathbb{E}[f(X_1^{n,x})] - \mathbb{E}[f(X_1^x)]$$

est donc au plus de l'ordre de  $1/\sqrt{n}$ . Précisément, la majoration (1.5) pour la valeur  $m = 1$ , associée à l'inégalité de Cauchy-Schwarz, montre que

$$|\Delta_1^n f(x)| \leq L_f \mathbb{E}[\|X_1^x - X_1^{n,x}\|] \leq L_f \sqrt{C_1^x/n},$$

où  $L_f$  est la constante de Lipschitz de  $f$ . Cependant,  $f(X_1^{n,x})$  est jeté aléatoirement au-dessus ou au-dessous de  $f(X_1^x)$  et, par le jeu des compensations, on présume une bien meilleure estimation de  $\Delta_1^n f(x)$ . Résumons maintenant les résultats connus à ce sujet.



### 1.3.5 Résultats connus

En itérant des développements d'Itô, Denis TALAY et Luciano TUBARO [93] ont obtenu en 1990, pour chaque fonction  $f \in C_{\text{poi}}^\infty(\mathbb{R}^d)$ , un développement limité de l'erreur faible sous la forme

$$\Delta_1^n f(x) = C/n + O(1/n^2), \quad (1.6)$$

sous l'hypothèse

(B') Les  $b_i$  et les  $\sigma_{i,j}$  sont des fonctions  $C^\infty$  dont toutes les dérivées d'ordre  $\geq 1$  sont bornées.

Ainsi, l'erreur faible associée au schéma d'Euler est proportionnelle au pas de temps  $1/n$ . On dit qu'elle est d'ordre 1. Elle est donc d'un demi-ordre supérieure à l'erreur forte qui est, elle, proportionnelle à la racine carrée du pas de temps  $1/n$ .

En utilisant le calcul de Malliavin, Vlad BALLY et D. TALAY [3] ont, en 1996, étendu ce résultat au cas des fonctions  $f$  mesurables bornées, dans le cas où la diffusion  $X^x$  est uniformément hypoelliptique. Ils obtiennent également [4] un développement limité de l'écart entre les densités<sup>1</sup>  $p(1, x, \cdot)$  de  $X_1^x$  et  $\tilde{p}_n(1, x, \cdot)$  d'une petite perturbation<sup>2</sup> de  $X_1^{n,x}$ . Si  $X^x$  est uniformément elliptique, i.e. sous l'hypothèse (C) (cf paragraphe 1.3.6),  $X_1^{n,x}$  admet une densité  $p_n(1, x, \cdot)$  et les auteurs obtiennent pour tout  $y \in \mathbb{R}^d$  un développement limité sous la forme

$$p_n(1, x, y) - p(1, x, y) = \pi(1, x, y)/n + \pi_n(1, x, y)/n^2. \quad (1.7)$$

Ils montrent aussi que le terme d'erreur principal  $\pi$  et le reste  $\pi_n$  ont des queues de distribution gaussiennes. Précisément, ils trouvent  $c_1 \geq 0$  et  $c_2 > 0$  tels que pour tous  $n \geq 1$  et  $x, y \in \mathbb{R}^d$ ,  $|\pi(1, x, y)| + |\pi_n(1, x, y)| \leq c_1 \exp(-c_2 \|x - y\|^2)$ . Par ailleurs, Valentin KONAKOV et Enno MAMMEN [61] ont proposé une approche analytique de ce problème basée sur la méthode parametrix. Si  $X^x$  est uniformément elliptique, ils obtiennent un développement de l'écart des densités à tout ordre  $p$  mais dont les termes dépendent de  $n$  :

$$p_n(1, x, y) - p(1, x, y) = \sum_{i=1}^{j-1} \pi_{n,i}(1, x, y)/n^i + O(1/n^j).$$

Eux aussi prouvent que les coefficients ont des queues de distribution gaussiennes, uniformément en  $n$  : pour tout  $i$ , ils trouvent des constantes  $c_1 \geq 0$  et  $c_2 > 0$  telles que pour tous  $n \geq 1$  et  $x, y \in \mathbb{R}^d$ ,  $|\pi_{n,i}(1, x, y)| \leq c_1 \exp(-c_2 \|x - y\|^2)$ . Pour ce faire, les auteurs ont besoin de majorations sur les dérivées partielles du noyau de transition de  $X$  - qu'ils trouvent dans les travaux d'Avner FRIEDMAN [35] - et démontrent des majorations analogues pour celui de  $X^n$ .

<sup>1</sup>par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ .

<sup>2</sup>Sous hypothèse d'hypoellipticité, la loi de  $X_1^{n,x}$  n'est pas nécessairement à densité.

### 1.3.6 Résultats de la thèse

#### Énoncés

Le principal résultat de cette partie de ma thèse peut être vu comme une amélioration de (1.7). Il donne le développement limité

$$p_n - p = \pi/n + \pi_n/n^2 \quad (1.8)$$

de  $p_n$ , noyau de transition de  $X^n$ , fonction des trois variables  $t \in (0, 1]$ ,  $x \in \mathbb{R}^d$  et  $y \in \mathbb{R}^d$ , dans des espaces fonctionnels de type “transition gaussienne” que je décris maintenant.

On note  $\mathcal{G}_l(\mathbb{R}^d)$  l’espace vectoriel des fonctions mesurables  $\pi : (0, 1] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  telles que

- pour tout  $t \in (0, 1]$ ,  $\pi(t, \cdot, \cdot)$  est  $C^\infty$ ,
- pour tous  $\alpha, \beta \in \mathbb{N}^d$ , il existe deux constantes  $c_1 \geq 0$  et  $c_2 > 0$  telles que pour tous  $t \in (0, 1]$  et  $x, y \in \mathbb{R}^d$ ,

$$\left| \partial_x^\alpha \partial_y^\beta \pi(t, x, y) \right| \leq c_1 t^{-(|\alpha| + |\beta| + d + l)/2} \exp(-c_2 \|x - y\|^2 / t). \quad (1.9)$$

On introduit aussi l’espace vectoriel  $\mathcal{G}(\mathbb{R}^d)$ , défini comme  $\mathcal{G}_l(\mathbb{R}^d)$  avec (1.9) remplacé par les deux conditions

$$\left| \partial_x^\alpha \partial_y^\beta \pi(t, x, y) \right| \leq c_1 t^{-(|\alpha| + |\beta| + d)/2} \exp(-c_2 \|x - y\|^2 / t), \quad (1.10)$$

$$\left| \partial_x^\alpha \left( \pi \left( t, x, x + y\sqrt{t} \right) \right) \right| \leq c_1 t^{-d/2} \exp(-c_2 \|y\|^2). \quad (1.11)$$

Enfin, on dit qu’une partie  $\mathcal{B} \subset \mathcal{G}_l(\mathbb{R}^d)$  (resp.  $\mathcal{B} \subset \mathcal{G}(\mathbb{R}^d)$ ) est bornée si, dans (1.9) (resp. dans (1.10)-(1.11)),  $c_1$  et  $c_2$  peuvent être choisis indépendamment de  $\pi \in \mathcal{B}$ . On montre alors qu’on a le développement limité (1.8) avec  $p \in \mathcal{G}(\mathbb{R}^d)$ ,  $(p_n, n \geq 1)$  bornée dans  $\mathcal{G}(\mathbb{R}^d)$ ,  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$  et  $(\pi_n, n \geq 1)$  bornée dans  $\mathcal{G}_4(\mathbb{R}^d)$ .

Pour obtenir ce résultat plus précis que (1.7), nous faisons des hypothèses plus fortes :

(B) Pour tous  $i \in \{1, \dots, d\}$  et  $j \in \{1, \dots, r\}$ ,  $b_i$  et  $\sigma_{i,j}$  appartiennent à  $C_b^\infty(\mathbb{R}^d)$ .

(C) Il existe  $\eta > 0$  tel que pour tout  $x, \xi \in \mathbb{R}^d$ ,  $\xi^* a(x) \xi \geq \eta \|\xi\|^2$ .

Dans cette écriture, l’étoile désigne la transposition et  $a = \sigma \sigma^*$ . L’hypothèse (B) ressemble beaucoup à l’hypothèse (B’) utilisée par D. TALAY et L. TUBARO [93] puis par V. BALLY et D. TALAY [3, 4]; elle requiert en plus la bornitude des  $b_i$  et  $\sigma_{i,j}$ . L’hypothèse (C) est la condition d’*uniforme ellipticité*, qui assure que  $X^x$  est “suffisamment stochastique”. Elle est strictement plus forte que la condition d’*uniforme hypoellipticité* employée par V. BALLY et D. TALAY [3, 4], mais a l’avantage de garantir l’existence de  $p_n(t, x, \cdot)$ , densité du schéma d’Euler  $X_t^{n,x}$  à toute date  $t \in (0, 1]$ . En particulier, V. KONAKOV et E. MAMMEN [61] se placent sous cette hypothèse.

On peut donc résumer notre principal résultat sous la forme du

**Theorem 1.** *Sous les hypothèses (B) et (C),*

- (i) *pour tout  $t \in (0, 1]$  et  $x \in \mathbb{R}^d$ ,  $X_t^x$  a une densité  $p(t, x, \cdot)$  et  $p \in \mathcal{G}(\mathbb{R}^d)$ ,*
- (ii) *pour tous  $t \in (0, 1]$ ,  $x \in \mathbb{R}^d$  et  $n \geq 1$ ,  $X_t^{n,x}$  a une densité  $p_n(t, x, \cdot)$  et  $(p_n, n \geq 1)$  est bornée dans  $\mathcal{G}(\mathbb{R}^d)$ ,*
- (iii) *il existe  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$  et une suite bornée  $(\pi_n, n \geq 1)$  dans  $\mathcal{G}_4(\mathbb{R}^d)$  tels que pour tout  $n \geq 1$ ,*

$$p_n - p = \pi/n + \pi_n/n^2. \quad (1.12)$$

Le développement (1.12) implique en particulier des équivalents en  $1/n$ , construits facilement à partir de  $\pi$ , ainsi que des majorations du reste, des erreurs suivantes :

- l'erreur  $\partial_x^\alpha \partial_y^\beta p_n(t, x, y) - \partial_x^\alpha \partial_y^\beta p(t, x, y)$  sur les dérivées spatiales de la densité,
- l'erreur  $\mathbb{E}[f(X_t^{n,x})] - \mathbb{E}[f(X_t^x)]$ , lorsque, pour un  $\mu < 2$ ,  $f$  appartient à l'ensemble  $\mathcal{E}_\mu$  des fonctions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  mesurables à croissance exponentielle d'exposant  $\mu$ , i.e. telles qu'il existe  $c_1, c_2 \geq 0$  tels que pour tout  $y \in \mathbb{R}^d$ ,

$$|f(y)| \leq c_1 \exp(c_2 \|y\|^\mu),$$

- l'erreur  $\partial_x^\alpha \mathbb{E}[f(X_t^{n,x})] - \partial_x^\alpha \mathbb{E}[f(X_t^x)]$  sur les sensibilités de  $\mathbb{E}[f(X_t^{n,x})]$  en la condition initiale  $x$ , pour la même classe de fonctions  $f$ ,
- l'erreur  $\langle S, p_n(t, x, \cdot) \rangle - \langle S, p(t, x, \cdot) \rangle$ , ou  $\langle S, p_n(t, \cdot, y) \rangle - \langle S, p(t, \cdot, y) \rangle$ , lorsque  $S$  est une distribution tempérée. En particulier, à condition de définir  $\mathbb{E}[f(Y)]$  par  $\langle f, p_Y \rangle$  lorsque  $f$  est une distribution tempérée et  $Y$  une variable aléatoire dont la densité  $p_Y$  est dans l'espace de Schwartz des fonctions  $C^\infty$  à décroissance rapide, le développement limité (1.6) est donc valable, sous (B) et (C), dès que  $f$  est une distribution tempérée.

Précisément, on prouve les propositions et théorèmes suivants :

**Proposition 2.** *Sous (B) et (C), pour tous  $\alpha, \beta \in \mathbb{N}^d$ , il existe  $c_1 \geq 0$  et  $c_2 > 0$  tels que pour tous  $n \geq 1$ ,  $t \in (0, 1]$  et  $x, y \in \mathbb{R}^d$ ,*

$$\partial_x^\alpha \partial_y^\beta p_n(t, x, y) - \partial_x^\alpha \partial_y^\beta p(t, x, y) = \frac{1}{n} \partial_x^\alpha \partial_y^\beta \pi(t, x, y) + r_n(t, x, y)$$

et

$$|r_n(t, x, y)| \leq c_1 n^{-2} t^{-(|\alpha| + |\beta| + d + 4)/2} \exp(-c_2 \|x - y\|^2 / t).$$

**Proposition 3.** *Sous (B) et (C), pour tous  $\alpha \in \mathbb{N}^d$ ,  $\mu \in (0, 2)$  et  $f \in \mathcal{E}_\mu$ , il existe  $c_1, c_2 \geq 0$  tels que pour tous  $n \geq 1$ ,  $t \in (0, 1]$  et  $x \in \mathbb{R}^d$ ,*

$$\partial_x^\alpha \mathbb{E}[f(X_t^{n,x})] - \partial_x^\alpha \mathbb{E}[f(X_t^x)] = \frac{1}{n} \int_{\mathbb{R}^d} f(y) \partial_x^\alpha \pi(t, x, y) dy + r_n(t, x)$$

avec

$$|r_n(t, x)| \leq c_1 n^{-2} t^{-(|\alpha| + 4)/2} \exp(c_2 \|x\|^\mu).$$

**Theorem 4.** *Sous (B) et (C), pour tout  $S \in \mathcal{S}'(\mathbb{R}^d)$ , il existe  $c \geq 0$  tel que pour tous  $n \geq 1$ ,  $t \in (0, 1]$  et  $x, y \in \mathbb{R}^d$ ,*

$$\begin{aligned}\langle S, p_n(t, x, \cdot) \rangle - \langle S, p(t, x, \cdot) \rangle &= \frac{1}{n} \langle S, \pi(t, x, \cdot) \rangle + r'_n(t, x), \\ \langle S, p_n(t, \cdot, y) \rangle - \langle S, p(t, \cdot, y) \rangle &= \frac{1}{n} \langle S, \pi(t, \cdot, y) \rangle + r''_n(t, y),\end{aligned}$$

et

$$|r'_n(t, x)| + |r''_n(t, x)| \leq cn^{-2}t^{-(d+4+\#S)/2} \left(1 + \|x\|^{\#S}\right),$$

où  $\#S$  désigne l'ordre de la distribution tempérée  $S$ .

### Schéma de la preuve

Donnons une idée de la preuve du théorème 1. L'assertion (i), connue depuis longtemps, a été démontrée par A. FRIEDMAN (cf [35], théorème 7, page 260). Quant à l'assertion (ii), elle a été essentiellement prouvée par V. KONAKOV et E. MAMMEN [61]. Pour démontrer l'assertion (iii), nous combinons les approches probabiliste et analytique :

- On commence par préciser (cf théorème 13) le résultat obtenu par D. TAYLOR et L. TUBARO [93]. Sous l'hypothèse (A), moins contraignante que l'hypothèse (B') utilisée par ces derniers, l'introduction d'une suite d'opérateurs différentiels, liée à des itérations de la formule d'Itô, permet d'explicitier les endomorphismes de  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $C_t$  et  $R_t^n$ , tels que  $R_t^n = O(1/n^2)$  et

$$P_t^n - P_t = C_t/n + R_t^n, \quad (1.13)$$

où  $P_t f(x) = \mathbb{E}[f(X_t^x)]$  et  $P_t^n f(x) = \mathbb{E}[f(X_t^{n,x})]$ . En particulier, on montre que

$$C_t = \frac{1}{2} \int_0^t P_s L_2^* P_{t-s} ds \quad (1.14)$$

où  $L_2^*$  est un opérateur différentiel d'ordre 3 qui s'exprime en fonction des coefficients  $b$  et  $\sigma$  de l'équation différentielle stochastique (1.2) :

$$\begin{aligned}-L_2^* &= \sum_{i=1}^d \left( b \cdot \nabla b_i + \frac{1}{2} \text{tr}(a \nabla^2 b_i) \right) \partial_i \\ &+ \sum_{i,j=1}^d \left( \frac{1}{2} b \cdot \nabla a_{i,j} + a_j \cdot \nabla b_i + \frac{1}{4} \text{tr}(a \nabla^2 a_{i,j}) \right) \partial_{ij} + \frac{1}{2} \sum_{i,j,k=1}^d a_k \cdot \nabla a_{i,j} \partial_{ijk}.\end{aligned} \quad (1.15)$$

Ici,  $\cdot$ ,  $a_k$ ,  $\text{tr}$ ,  $\nabla$  et  $\nabla^2$  désignent respectivement le produit scalaire de  $\mathbb{R}^d$ , la  $k$ -ème colonne de  $a$ , la trace d'une matrice, le gradient et la matrice hessienne.

- On explicite ensuite, sous les hypothèses (B) et (C),  $C_t$  et  $R_t^n$  sous forme d'opérateurs intégraux, c'est-à-dire qu'on exhibe des fonctions  $\pi$  et  $\pi_n$  telles que pour tout  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$

$$\begin{aligned} C_t f(x) &= \int_{\mathbb{R}^d} f(y) \pi(t, x, y) dy, \\ n^2 R_t^n f(x) &= \int_{\mathbb{R}^d} f(y) \pi_n(t, x, y) dy, \end{aligned}$$

et on montre que  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$  et que  $(\pi_n, n \geq 1)$  est bornée dans  $\mathcal{G}_4(\mathbb{R}^d)$  (cf (2.44) et les lemmes 21, 22, 23 et 24).

Traisons par exemple le cas de  $C_t$ . Sous (B)-(C), pour  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $x \in \mathbb{R}^d$  et  $0 < s < t$ , en utilisant le théorème de dérivation sous le signe somme pour la deuxième égalité,

$$\begin{aligned} P_s L_2^* P_{t-s} f(x) &= \int_{\mathbb{R}^d} p(s, x, z) L_2^* P_{t-s} f(z) dz \\ &= \int_{\mathbb{R}^d} f(y) \left( \int_{\mathbb{R}^d} p(s, x, z) L_2^*(p(t-s, \cdot, y))(z) dz \right) dy, \end{aligned}$$

si bien que  $C_t f(x) = \int_{\mathbb{R}^d} f(y) \pi(t, x, y) dy$  avec

$$\pi(t, x, y) = \int_0^t \int_{\mathbb{R}^d} p(s, x, z) L_2^*(p(t-s, \cdot, y))(z) dz ds. \quad (1.16)$$

Pour montrer que  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$ , on découpe l'intégrale temporelle en deux, suivant que  $s < t/2$  ou  $s \geq t/2$ , et on intègre par parties dans ce dernier cas. Cette technique est similaire à celle qu'utilisent V. BALLY and D. TALAY [3], mais là où ces derniers emploient la formule d'intégration par parties du calcul de Malliavin, nous ne nous servons que de la formule d'intégration par parties usuelle. On utilise ensuite les majorations sur les dérivées partielles des noyaux de  $X$  pour conclure. L'hypothèse d'uniforme ellipticité est cruciale à ce stade, car c'est elle qui garantit les majorations (1.10)-(1.11). Le schéma de démonstration est identique pour le reste  $R_t^n$  - mais le travail un peu plus long.

L'erreur fonctionnelle au premier ordre sur le noyau de transition,  $\pi$ , s'écrit donc en fonction du noyau de transition exact  $p$  et de l'opérateur  $L_2^*$  défini par (1.15) sous la forme (1.16). Dans le cas où  $t = 1$ , (1.16) coïncide avec l'expression de  $\pi$  obtenue par V. BALLY et D. TALAY ([4], définition 2.2, page 100), mais semble préférable puisque, contrairement à l'opérateur  $\mathcal{U}$  dans [4],  $L_2^*$  n'implique ni différentiation temporelle ni différentiation d'ordre 4 en espace.

## Estimation d'erreur sur les deltas et gammas d'une option européenne

Dans le contexte des marchés financiers, on peut considérer  $X^x$  comme le logarithme - coordonnée par coordonnée - des cours d'un panier de  $d$  actifs. Lorsque  $|\alpha| = 1$  (resp.  $|\alpha| = 2$ ),  $\partial_x^\alpha \mathbb{E}[f(X_t^x)]$  s'interprète alors en terme de delta (resp. gamma), quantités qui interviennent dans la couverture dynamique de l'option européenne de maturité  $t$  et de payoff  $f(X_t^x)$ . On obtient ainsi (cf paragraphe 2.1.5) des équivalents en  $1/n$ , ainsi que des majorations des restes, des erreurs commises dans les calculs de ces quantités lorsqu'on utilise un schéma d'Euler pour approcher le log-sous-jacent  $X^x$ .

### 1.3.7 Résultats voisins

L'estimation de l'erreur faible due à la discrétisation temporelle d'équations différentielles stochastiques sert de matière à une littérature très riche. Une référence générale sur la résolution numérique de ces équations est l'ouvrage de Peter KLOEDEN et Eckhard PLATEN [58].

Un grand nombre de travaux récents consacrés à l'erreur faible du schéma d'Euler examine le cas où l'équation différentielle stochastique est gouvernée, non pas par un mouvement brownien comme dans (1.2), mais plus généralement par un processus de Lévy  $L$ . Dans la définition du schéma d'Euler (1.3), l'incrément  $B_t - B_{t_k^n}$  doit alors être remplacé par l'incrément  $L_t - L_{t_k^n}$ . A ma connaissance, l'existence dans cette situation d'un développement limité en puissances du pas de temps  $1/n$  de l'erreur faible du schéma d'Euler  $\Delta_t^n f(x) = \mathbb{E}[f(X_t^{n,x})] - \mathbb{E}[f(X_t^x)]$  n'a été prouvée à ce jour que dans le cas où  $f$  est une fonction régulière. Philip PROTTER et D. TALAY [84] sont les premiers, en 1997, à obtenir un tel résultat sous hypothèse de finitude d'un nombre suffisant de moments de la mesure de Lévy de  $L$ . Jean JACOD, Thomas KURTZ, Sylvie MÉLÉARD et Ph. PROTTER [55] prennent en compte l'impossibilité de simuler exactement, pour un processus de Lévy général  $L$ , les accroissements  $L_t - L_{t_k^n}$ . Sous des hypothèses similaires - régularité de  $f$ , finitude d'un nombre suffisant de moments de la mesure de Lévy de  $L$  -, ils montrent que l'ordre de grandeur de  $\Delta_t^n f(x)$  est alors le maximum de  $1/n$  et d'une "distance" entre les accroissements de  $L$  et leurs approximations effectivement simulées. Ils en précisent également la dépendance en  $f$ ,  $t$  et  $x$ . Notons que dans le cas où les accroissements  $L_t - L_{t_k^n}$  peuvent être exactement simulés, ils obtiennent un développement de l'opérateur d'erreur  $\Delta_t^n$  similaire à (1.13)-(1.14) où l'opérateur différentiel  $L_2^*$  est remplacé par un opérateur intégro-différentiel, détaillé dans leur remarque 9.1, faisant intervenir la mesure de Lévy de  $L$ .

Andreas NEUENKIRCH et Ivan NOURDIN [81], eux, considèrent, en dimension un, le cas où  $B$  est un mouvement brownien fractionnaire d'indice de Hurst  $H > 1/2$ . Dans ce cas,  $\int_0^t \sigma(X_s^x) dB_s$  doit être comprise comme une intégrale anticipante de

Russo-Valois, et les auteurs obtiennent le résultat suivant, assez surprenant puisque n'étant plus vrai à la limite où  $H = 1/2$  correspondant au cas brownien standard :

$$\mathbb{P} \left( \lim_{n \rightarrow +\infty} n^{2H-1} (X_1^{n,x} - X_1^x) = -\frac{1}{2} \int_0^1 \sigma'(X_s^x) D_s X_1^x ds \right) = 1.$$

Ici,  $D_s X_1^x$  désigne la dérivée de Malliavin de  $X_1^x$  à la date  $s$ .

L'erreur faible pour les équations différentielles stochastiques non-linéaires au sens de McKean-Vlasov est étudiée par Mireille BOSSY et D. TALAY [10, 11]. Ces équations ne sont pas linéaires au sens où leurs coefficients dépendent de la loi du processus solution. D. TALAY a aussi étudié avec Ziyu ZHENG [94] l'erreur du schéma d'Euler sur les quantiles de la loi de  $X_1^x$ .

D'autres auteurs se sont penchés sur le cas où les coefficients de l'équation différentielle stochastique (1.2) sont irréguliers. Vigirdas MACKEVICIUS [74] montre que  $|\Delta_t^n f(x)| \leq c(t, f, x)/n$  lorsque  $\sigma$  est constante,  $b$  seulement lipschitzienne et  $f$  suffisamment dérivable. Liqing YAN [97], pour sa part, considérant le cas où le coefficient de dérive  $b$  et le coefficient de diffusion  $\sigma$  dépendent aussi du temps, prouve la convergence en loi du processus  $X^{n,x}$  vers  $X^x$  lorsque  $b$  et  $\sigma$  sont presque partout continus. Il montre également que si  $b$  est lipschitzien en espace et höldérien en temps et si  $\sigma$  est höldérien en espace et en temps, alors  $|\Delta_t^n f(x)| \leq c(t, f, x)/n^\gamma$  dès que  $f$  est lipschitzienne ;  $\gamma \in [0, 1/2]$  dépend de la régularité höldérienne de  $b$  et  $\sigma$ .

Emmanuel GOBET et Stéphane MENOZZI [38] se sont intéressés au cas où le processus  $X^x$  est tué lorsqu'il quitte un domaine régulier  $D$  de  $\mathbb{R}^d$ . Sous l'hypothèse d'uniforme hypoellipticité, et pour des fonctions  $f$  mesurables bornées, ils prouvent que l'erreur faible du schéma d'Euler  $\mathbb{E} [f(X_1^{n,x}) \mathbf{1}_{\{\tau^n > 1\}}] - \mathbb{E} [f(X_1^x) \mathbf{1}_{\{\tau > 1\}}]$  est alors exactement d'ordre  $1/2$ , c'est-à-dire à la fois majorée et minorée par des quantités d'ordre  $1/\sqrt{n}$  ; ici,  $\tau^n = \inf \{t_k^n | X_{t_k^n}^{n,x} \notin D\}$  et  $\tau = \inf \{t | X_t^x \notin D\}$ .

La vitesse de convergence en loi du processus  $X^{n,x}$  vers  $X^x$ ,  $\sqrt{n}$ , a été déterminée par J. JACOD et Ph. PROTTER [57], qui obtiennent plus généralement des vitesses dans le cas où l'équation différentielle stochastique (1.2) est dirigée par une semimartingale. Dans le cas où (1.2) est dirigée par un processus de Lévy sans partie brownienne, J. JACOD [54] montre que cette vitesse dépend de la concentration de la mesure de Lévy autour de 0.

Arturo KOHATSU-HIGA et Roger PETERSON [60] ont développé une méthode originale pour évaluer l'erreur faible  $\Delta_1^n f(x)$ . Ils écrivent  $f(X_1^{n,x}) - f(X_1^x) = (X_1^{n,x} - X_1^x) \int_0^1 f'(\alpha X_1^{n,x} + (1-\alpha)X_1^x) d\alpha$ , résolvent l'équation linéaire satisfaite par le processus d'erreur  $X^{n,x} - X^x$ , et majorent par  $1/n$  la vitesse de convergence en utilisant la propriété de dualité des intégrales stochastiques, ou formule de dualité du calcul de Malliavin. La formule d'intégration par parties leur permet, sous hypothèse d'uniforme ellipticité, d'obtenir cette majoration sur les densités. Cette technique a été utilisée par Emmanuel GOBET et Rémi MUNOS [39] qui présentent différentes méthodes d'estimation de la sensibilité de  $\mathbb{E}[f(X_1^x)]$  à un paramètre  $\theta$  dont dépend

la dynamique de  $X^x$ . Elle a ensuite été étendue par Emmanuelle CLÉMENT, A. KOHATSU-HIGA et Damien LAMBERTON [18] au cas des fonctionnelles régulières de la trajectoire  $(X_t^x, t \in [0, 1])$ , et à une large classe d'équations différentielles stochastiques plus générales que (1.2), incluant notamment les équations différentielles stochastiques avec retard. A ce sujet, citons aussi les travaux de Evelyn BUCKWAR et Tony SHARDLOW [15], E. PLATEN et Uwe KÜCHLER [62, 63] et Yaozhong HU, Salah-Eldin MOHAMMED et Feng YAN [47]. A propos des fonctionnelles régulières de la trajectoire, signalons aussi les travaux de thèse d'Etienne TANRÉ [95].

Paul MALLIAVIN et Anton THALMAIER [75] expriment la partie principale  $C_t$  de l'erreur  $\Delta_t^n = C_t/n + R_t^n$  comme une distribution de Watanabe généralisée.

Notons que la régularité du noyau de transition  $p$  et des majorations de ses dérivées partielles, d'abord obtenues sous hypothèse d'uniforme ellipticité par A. FRIEDMAN [35] en 1964 via la méthode parametrix, peuvent également être démontrées, sous hypothèse d'hypoellipticité, en utilisant le calcul de Malliavin, plus récent. Le livre de David NUALART [82] présente ce calcul et on y trouve une preuve de l'existence et de la régularité de la densité de  $X_t^x$ ,  $t \in (0, 1]$ . Les travaux de Shigeo KUSUOKA et Daniel STROOCK [66, 67, 68] fournissent des estimations fines de  $p$  et de ses dérivées. Mentionnons aussi ici que Chris ROGERS [86] donne une démonstration simple de l'existence et de la régularité de  $p$  dans le cas unidimensionnel et uniformément elliptique. Nicolas FOURNIER et Sébastien GIET [34] donnent des conditions suffisantes pour l'existence de  $p$  lorsque la dynamique de  $X^x$  a aussi une composante de sauts. En utilisant le calcul de Malliavin, Y. HU et Shinzo WATANABE [48] proposent des schémas d'approximation de  $p$ . A ce sujet, A. KOHATSU-HIGA et R. PETTERSON [60] comparent les mérites respectifs de la méthode d'approximation par noyaux et de la méthode utilisant l'intégration par parties de Malliavin, et proposent des choix de poids optimaux pour réduire la variance dans ce dernier cas. Grigori MILSTEIN, John SCHOENMAKERS et Vladimir SPOKOINY [77], eux, estiment  $p$  par retournement du temps.

Signalons aussi que le schéma d'Euler n'est pas la panacée. Le célèbre schéma de Milstein est d'ordre 1, aussi bien pour l'erreur forte que pour l'erreur faible. En dimension supérieure à un, il implique le calcul des intégrales  $\int_0^t B_s dB'_s$ , où  $B$  et  $B'$  sont deux browniens indépendants, intégrales que l'on ne sait malheureusement pas simuler exactement. Ana Bela CRUZEIRO, P. MALLIAVIN et A. THALMAIER [21] en ont proposé une variante dans laquelle ce calcul est évité. Par ailleurs, à la suite des travaux de S. KUSUOKA [65] sur les développements de Taylor stochastiques, Terry LYONS et Nicolas VICTOIR [72], qui obtiennent des formules de cubature sur l'espace de Wiener, puis Syoiti NINOMIYA [80, 79] proposent des schémas d'ordre arbitrairement élevés pour l'erreur faible. Une méthode de pas adaptatifs, visant à réduire la constante  $C$  dans (1.6), a été proposée par Kyoung-Sook MOON, Anders SZEPESSY, Raúl TEMPONE et Georgios ZOURARIS [90, 78]. Pour de tels schémas, l'erreur forte est étudiée en particulier par Norbert HOFMANN, Thomas MÜLLER-



GRONBACH et Klaus RITTER [45, 46].

Signalons enfin que J. JACOD, S. MÉLÉARD et Ph. PROTTER [56] étudient la convergence des portefeuilles de couverture, via la convergence de la représentation martingale de  $X_1^{n,x}$  vers celle de  $X_1^x$ . A ce propos, le calcul de  $\partial_x^\alpha \mathbb{E}[f(X_t^{n,x})]$ , ou d'autres approximations de  $\partial_x^\alpha \mathbb{E}[f(X_t^x)]$ , ou plus généralement de la sensibilité de  $\mathbb{E}[f(X_1^x)]$  à un paramètre  $\theta$  dont dépend la dynamique de  $X^x$ , a fait l'objet d'une très riche littérature, analysant en particulier la méthode des différences finies [37, 36, 69], la méthode de différentiation sous le signe  $\mathbb{E}$  [14] et la méthode d'intégration par parties de Malliavin, en général préférable si  $f$  est discontinue, initiée par Eric FOURNIÉ, Jean-Michel LASRY, Jérôme LEBUCHOUX, Pierre-Louis LIONS et Nizar TOUZI [32, 33], puis développée dans [12, 39, 59].

### 1.3.8 Extensions possibles

Vu cet état de l'art, on peut imaginer des extensions naturelles du théorème 1.

Lorsqu'on remplace le mouvement brownien  $B$  par un processus de Lévy, on a vu que la vitesse de convergence de l'erreur faible n'est toujours pas connue pour des fonctions  $f$  irrégulières. Notre approche pourrait-elle être utile pour traiter ce cas ? Le terme principal de l'opérateur d'erreur,  $C_t$ , est alors connu ([55], remarque 9.1), mais son expression sous forme d'un opérateur intégral, dont on contrôlerait le noyau  $\pi$ , nécessite l'existence et de bonnes majorations de  $p$  dont on ne dispose pas à ce jour. En outre, le même travail devrait être accompli sur le reste  $R_t^n$ , non explicité dans [55], et donc aussi sur  $p_n$ .

L'extension au cas où les coefficient de dérive et de diffusion dépendent aussi du temps me semble possible. En effet, si l'on conserve l'hypothèse (C) d'uniforme intégrabilité, sous la forme  $\xi^* a(t, x) \xi \geq \eta \|\xi\|^2$ , alors  $p \in \mathcal{G}(\mathbb{R}^d)$ , comme l'a prouvé A. FRIEDMAN [35], et il est raisonnable de penser que la suite  $(p_n, n \geq 1)$  est dans ce cas encore bornée dans  $\mathcal{G}(\mathbb{R}^d)$ , même si les travaux de V. KONAKOV et E. MAMMEN [61] ne mentionnent pas ce cas, si bien que la méthode que je propose pourrait être immédiatement adaptée.

Enfin, on peut aussi envisager un affaiblissement de l'hypothèse d'infinie régularité des coefficients  $b$  et  $\sigma$ , résultant en un affaiblissement de la régularité en espace des fonctions  $p$ ,  $p_n$ ,  $\pi$  et  $\pi_n$ .

## 1.4 Analyse d'un modèle à volatilité stochastique

L'analyse de l'erreur en loi du schéma d'Euler est mise en pratique, au chapitre 3, dans la simulation d'un modèle à volatilité stochastique. La famille des modèles à volatilité stochastique, dont on présente quelques exemples au paragraphe 3.1.3 [43, 44, 53, 88], permet de rendre compte des smiles observés sur les marchés, c'est-à-dire du fait que la volatilité implicite, calculée à partir des prix de marché des calls

par inversion de la formule de Black-Scholes, dépend du prix d'exercice  $K$  et de la maturité  $T$ . Elle permet aussi de prendre en considération la corrélation empirique négative entre volatilité et sous-jacent.

Dans cette partie de la thèse, après avoir rappelé les propriétés du modèle de Black-Scholes, on s'intéresse à celles du modèle suivant

$$\begin{cases} dX_t = \mu X_t dt + \sigma_t X_t dW_t, \\ \sigma_t = f(Y_t), \\ dY_t = \alpha(m - Y_t) dt + \beta d\hat{Z}_t, \end{cases} \quad (1.17)$$

proposé en 2000 par Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR dans leur ouvrage [29]. Dans cette écriture :

- $X$  représente le sous-jacent,  $X_t$  son cours à la date  $t$ ,
- $\mu$  est le rendement instantané, supposé constant,
- $\sigma_t$  est la valeur à la date  $t$  de la volatilité du cours du sous-jacent ; elle mesure l'intensité du bruit  $\sigma_t X_t dW_t$  auquel est soumis le cours du sous-jacent,
- $W$  est un mouvement brownien standard, la volatilité  $\sigma$  est elle-même un processus stochastique, fonction déterministe du processus  $Y$  ; la fonction  $f$  est définie sur  $\mathbb{R}$  et à valeurs strictement positives,
- $Y$  est un processus d'Ornstein-Uhlenbeck, de moyenne à long terme  $m$  et de variance à long terme  $\frac{\beta^2}{2\alpha}$ ,
- $\hat{Z}$  est un mouvement brownien standard éventuellement corrélé à  $W$ .

Nous résumons d'abord le travail théorique des auteurs. En particulier, on montre que le marché est incomplet et on exhibe une famille de probabilités risque-neutre équivalentes à la probabilité historique, indexée par une famille de primes de risque de volatilité. Ces primes de risque traduisent la surfacturation due à l'incertitude sur la volatilité du marché. Les auteurs obtiennent ensuite des développements asymptotiques du prix et de différentes stratégies de couverture des options vanille lorsque la volatilité retourne vite à sa moyenne. Dans ce cas, on facture au client un risque de volatilité qui dépend peu du strike. Aussi, afin de générer des smiles de grande amplitude ressemblant à ceux souvent observés sur les marchés, nous proposons ensuite des exemples de simulation informatique originaux dans le cas non asymptotique où la constante de temps de retour à la moyenne de la volatilité,  $1/\alpha$ , est comparable à la maturité de l'option étudiée. Nous décrivons l'influence des différents paramètres et devinons quelques propriétés du modèle au vu des expériences numériques. En particulier, faire varier la force de rappel de l'Ornstein-Uhlenbeck et observer l'effet sur le smile illustre la propriété d'ergodicité démontrée par J.-P. FOUQUE, G. PAPANICOLAOU et K. R. SIRCAR.

Dans l'algorithme que nous proposons, l'équilibre entre l'erreur statistique, due à l'échantillonnage de la méthode de Monte-Carlo, et l'erreur de discrétisation tempo-

relle, étudiée au chapitre 2, est assurée de manière adaptative. En fait, nous n'utilisons pas le schéma d'Euler mais préférons exploiter certaines particularités du modèle (1.17), en particulier le fait que  $Y$  puisse être simulé exactement et de manière autonome. Ensuite, le logarithme de  $X$  est simulé sachant  $Y$  en utilisant le schéma d'Euler. En outre, on utilise la méthode d'extrapolation de Romberg et plusieurs méthodes de réduction de variance qui s'avèrent numériquement efficaces.

Une direction possible de recherche concerne la calibration de ce modèle. C'est certainement un problème mathématique délicat. J.-P. FOUQUE, G. PAPANICOLAOU et K. R. SIRCAR montrent comment calibrer le modèle dans le cas où  $\alpha$  est grand devant  $1/T$ , mais pas dans le cas général - et pertinent - où il ne l'est pas. Par ailleurs, les simulations montrent que la volatilité implicite en dehors de la monnaie tend vers plus l'infini quand la maturité tend vers 0 et que la volatilité implicite à la monnaie ne dépend de  $\alpha$  et  $T$  que *via* le produit  $\alpha T$ . Les explications mathématiques de ces propriétés restent à trouver.

## 1.5 Théorèmes limites pour les chaînes de Markov bifurcantes

### 1.5.1 Motivation

Cette partie de ma thèse, la dernière, est née d'une coopération avec le Laboratoire de Génétique Moléculaire, Évolutive et Médicale (unité INSERM U571, Faculté de Médecine Necker, Paris). F. TADDÉI, Eric STEWART, Ariel LINDNER and Grégory PAUL, avec Richard MADDEN de l'Institut des Hautes Études Scientifiques, ont étudié le phénomène de vieillissement chez *Escherichia Coli*. *Escherichia Coli* est l'organisme unicellulaire "de référence". Il a été très largement étudié par la communauté des biologistes qui a produit à son sujet une très vaste littérature. C'est vraisemblablement à cette échelle microscopique que l'on a le plus de chances de décrire et de quantifier les processus moléculaires responsables du vieillissement. Cependant, il n'est pas évident de mesurer le vieillissement chez les êtres unicellulaires, et en particulier chez *Escherichia Coli* qui se reproduit en se divisant en deux de manière apparemment symétrique et sans montrer de phase juvénile. E. J. STEWART *et al.* [89] ont construit une expérience qui permet d'identifier et de mesurer le vieillissement chez *Escherichia Coli*. Nous proposons une étude statistique des données expérimentales.

Ces données se présentent sous forme d'un arbre binaire de taux de croissance, à partir duquel nos collègues biologistes souhaitent détecter deux sous-populations, correspondant à deux phénotypes distincts : un phénotype "taux de croissance faible", et un phénotype "taux de croissance élevé". L'arbre binaire est l'arbre

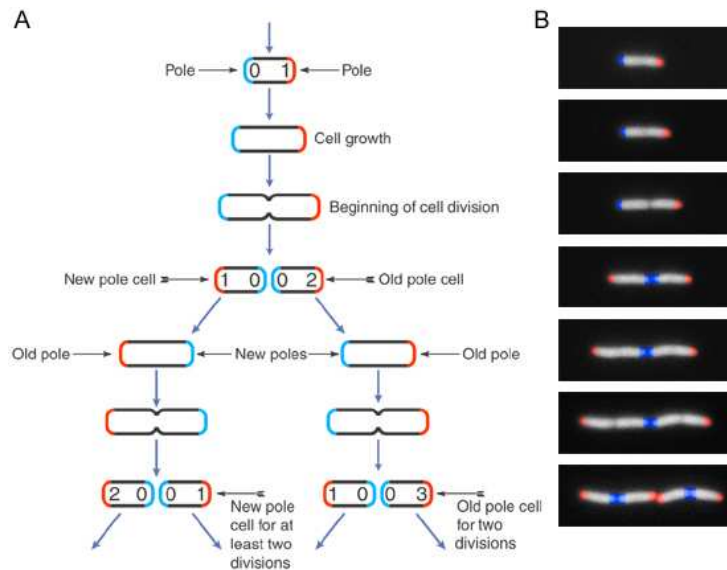


FIG. 1.1 – Le cycle de vie d'*E. coli*, d'après E. J. STEWART *et al.* [89]

généalogique issu d'un ancêtre tiré au hasard dans une méta-colonie. Les cellules se divisent en deux toutes les demi-heures environ, et en général, pour des raisons techniques, on ne mesure le développement de la colonie que jusqu'à la septième ou la huitième génération. Entre deux divisions, une dizaine de photographies montre que la longueur de la cellule, qui a une forme de bâtonnet, croît exponentiellement vite; on calcule ainsi un taux de croissance pour chaque individu de l'arbre.

## 1.5.2 La biologie

La division, qui s'opère par le milieu de la cellule, semble symétrique. En réalité, les deux cellules filles sont physiologiquement distinguables (cf. figure 1.1, où les âges des pôles de la cellule du haut - 0 et 1 - ont été arbitrairement choisis). En effet, après division, pour chaque cellule fille, une des deux extrémités est neuve, l'autre préexistant à la division. A la division suivante, une des deux cellules filles récupère l'extrémité neuve (on parle de nouveau pôle), l'autre héritant de l'extrémité "vieille" (on parle de vieux pôle). Ainsi, à chaque division, exceptée la toute première, on distingue la cellule fille "nouveau pôle" de la cellule fille "vieux pôle". On s'attend à ce que tout constituant cellulaire se formant dans les extrémités et diffusant peu s'accumule dans les vieux pôles. Les cellules "vieux pôles" et les cellules "nouveaux pôles" pourraient ainsi être physiologiquement asymétriques. Notre objectif est de montrer que les cellules "nouveaux pôles" croissent plus vite que les cellules "vieux pôles" et peuvent donc être considérées comme des "enfants" des "vieux pôles".

Malheureusement pour le biologiste, et heureusement pour le statisticien, cela

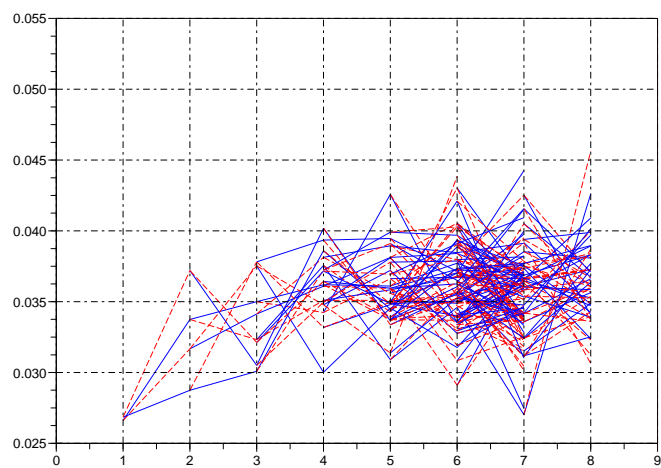
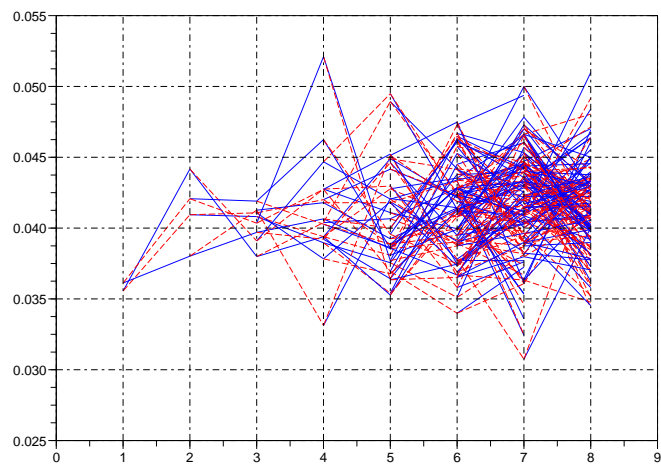


FIG. 1.2 – Deux arbres de données

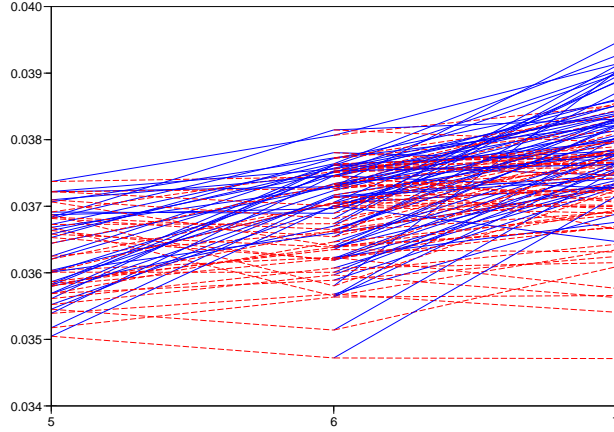


FIG. 1.3 – L'arbre moyen

ne se distingue pas à l'œil nu. Les données, dont deux exemples sont reproduits en figure 1.2, sont en effet extrêmement bruitées. Sur ces graphes, les nouveaux pôles sont en trait plein et les vieux pôles en pointillés. On a représenté le numéro de la division en abscisse et le taux de croissance en ordonnée.

Afin de débruiter ces données, E. J. STEWART *et al.* ont moyenné les 95 arbres constituant leur jeu de données. La figure 1.3 représente l'arbre moyen, au niveau des générations 5 à 7. Elle montre clairement une ségrégation entre nouveaux et vieux pôles. Cependant, afin d'éviter de modéliser la répétition des expériences (ont-elles même loi? sont-elles indépendantes?), il est souhaitable de pouvoir opérer cette ségrégation expérience par expérience. C'est ce que nous nous proposons de faire dans cette partie de la thèse.

### 1.5.3 Le modèle

Notons  $n \geq 1$  la mère de  $2n$  - la cellule fille nouveau pôle - et de  $2n + 1$  - la cellule fille vieux pôle - (cf. figure 1.4) et  $X_i$  le taux de croissance de l'individu  $i$ . On propose le modèle markovien de mémoire un suivant pour expliquer les données. Le taux de croissance de l'ancêtre,  $X_1$ , suit la loi  $\nu$  et pour tout  $n \geq 1$ ,

$$\begin{cases} X_{2n} &= \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} &= \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \quad (1.18)$$

Les paramètres  $\alpha_0, \alpha_1$  appartiennent à  $(-1, 1)$ ;  $\beta_0, \beta_1 \in \mathbb{R}$ ; le bruit est modélisé par la suite  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$  de vecteurs gaussiens bidimensionnels indépendants,

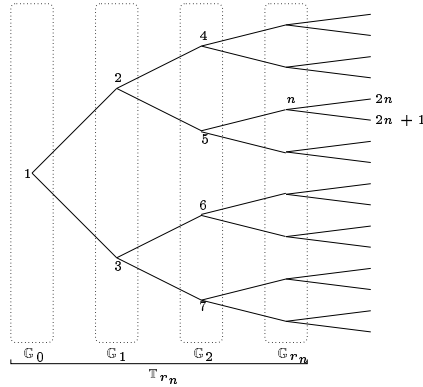


FIG. 1.4 – L'arbre binaire  $T$

centrés et de matrice de covariance

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1).$$

Ce modèle décrit une structure de dépendance markovienne de mémoire un, puisque le taux de croissance d'une cellule est expliqué par celui de sa mère. Une telle modélisation est naturelle, puisque deux cellules filles se partagent la matière de leur mère. Un modèle markovien de mémoire deux prendrait aussi en compte le taux de croissance de la grand-mère. En un certain sens, (1.18) est le modèle raisonnable "le plus simple" décrivant une dépendance au sein de la colonie : l'explication des taux de croissance de la progéniture est linéaire en celui de la mère, et le bruit est gaussien. On peut le baptiser modèle autoregressif bifurcant, de mémoire un, linéaire et gaussien. Le paramètre  $\rho$  prend en compte le fait que deux cellules filles se partagent une même matière (en particulier une protéine présente en exemplaire unique chez la mère ne se transmet qu'à une des deux filles), ce qui induit une corrélation négative, mais aussi le fait qu'elles croissent en même temps en étant géographiquement voisines l'une de l'autre, dans le même environnement, ce qui induit une corrélation positive. L'indépendance des vecteurs  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$  revient à négliger le facteur environnement pour des cellules cousines, mère-fille, ou plus éloignées encore. Notons que puisqu'une variable aléatoire réelle gaussienne prend des valeurs négatives arbitrairement grandes, le modèle autorise les taux de croissance à être négatifs. Cependant, vu les valeurs observées des paramètres, ceci ne se produit qu'avec une probabilité extrêmement faible.

Avec ce formalisme, on peut reformuler nos objectifs sous la forme suivante :

- (1) estimer le paramètre  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ ,  $\rho$  et  $\sigma^2$ ,
- (2) tester l'hypothèse nulle  $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$  contre son alternative  $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$ .

Le point (2) est crucial pour les biologistes : rejeter  $H_0$  revient à accepter le fait que les dynamiques des taux de croissance des nouveaux pôles et des vieux pôles

sont différentes. On verra plus précisément que, dans les conditions expérimentales étudiées, les nouveaux pôles croissent en moyenne environ 3% plus vite que les vieux pôles.

#### 1.5.4 Modèles voisins

Plusieurs auteurs ont déjà étudié les modèles autorégressifs bifurcants, ou modèles BAR, pour “Bifurcating AutoRegressive models”. Cependant, à ma connaissance, aucun de ces auteurs ne distingue deux types de filles, tels que les types “nouveau pôle” et “vieux pôle”. Mathématiquement parlant, à ce jour, tous les travaux sur les modèles BAR supposent que la loi de  $X_{2n}$  sachant  $X_n$  est identique à la loi de  $X_{2n+1}$  sachant  $X_n$ . Or, l’enjeu essentiel est ici de détecter une différence entre ces lois. De ce point de vue, le modèle (1.18) doit être perçu comme une généralisation des modèles BAR existants qui permet de rendre compte ou de révéler une dissymétrie systémique entre les sœurs. Cette généralisation est source de difficultés mathématiques. En particulier, sous  $H_1$ , il n’existe pas de loi stationnaire, c’est-à-dire de loi commune à tous les taux de croissance des cellules de la colonie.

Richard COWAN et Robert STAUDTE [20] sont les premiers à proposer un modèle BAR en 1986, déjà pour étudier des données de lignages cellulaires. C’est exactement le modèle (1.18), mais étudié dans le cas où  $(\alpha_0, \beta_0) = (\alpha_1, \beta_1)$ , i.e. sous  $H_0$ . Les auteurs donnent les équations satisfaites par l’estimateur du maximum de vraisemblance du paramètre  $(\alpha, \beta, \rho, \sigma^2)$  et simulent le modèle. Richard HUGGINS et R. STAUDTE [52] proposent des estimateurs dits robustes pour des variantes du modèle (1.18), basés sur l’observation d’un grand nombre d’arbres indépendants. R. HUGGINS [49] traite ensuite le cas où un seul arbre, dont la taille tend vers l’infini, est observé. R. STAUDTE, R. HUGGINS, J. ZHANG, David AXELROD et Marek KIMMEL [87] autorisent certains paramètres à dépendre de la génération et de l’expérience. Quang BUI et R. HUGGINS [16] étendent le modèle au cas où  $X$  est bidimensionnel et l’appliquent à des données pour estimer la corrélation entre la durée de vie d’une cellule et son diamètre final. Ils proposent aussi [17] de rendre le coefficient  $\alpha$  aléatoire afin de modéliser le cas où ce paramètre dépend de l’environnement. R. HUGGINS et Ishwar BASAWA [50] proposent le modèle BARMA( $p, q$ ), qui est au modèle ARMA( $p, q$ ) (AutoRegressive Moving Average de paramètre ( $p, q$ )) ce que le modèle BAR est au modèle AR(1). En particulier, le modèle BARMA( $p, q$ ) explique le taux de croissance d’un individu par ceux de ses  $p$  plus proches ancêtres;  $q$  est une profondeur de moyenne mobile (Moving Average en anglais) sur le bruit. Les auteurs donnent ensuite [51] les propriétés de l’estimateur du maximum de vraisemblance dans le modèle BARMA( $p, 0$ ). Enfin, I. V. BASAWA et Jin ZHOU [5] traitent des exemples où le bruit  $\varepsilon$  n’est pas gaussien.



### 1.5.5 Résultats de la thèse

Nous proposons une approche nouvelle de ces problèmes basée uniquement sur le caractère markovien du processus  $X$  et indépendante du choix des fonctions de régression et du bruit. Précisément, nous proposons une méthode générale pour étudier des données dépendantes sur un arbre binaire : chaque individu génère deux filles, une du type 0 - disons “nouveau pôle” - et une du type 1 - “vieux pôle”.

#### Quelques notations

Comme indiqué sur la figure 1.4, on note 1 l’ancêtre de la colonie, et  $n \geq 1$  la mère de  $2n$  - la cellule fille de type 0 - et de  $2n + 1$  - la cellule fille de type 1. On note également  $\mathbb{G}_q = \{2^q, 2^q + 1, \dots, 2^{q+1} - 1\}$  la génération  $q$  et  $\mathbb{T}_r = \cup_{q=0}^r \mathbb{G}_q$  le sous-arbre constitué des  $r + 1$  premières générations. Leurs cardinaux sont  $|\mathbb{G}_q| = 2^q$  et  $|\mathbb{T}_r| = 2^{r+1} - 1$ .

On s’intéresse à une caractéristique ( $x \in S$ , un espace métrique muni de sa tribu borélienne  $\mathcal{S}$ ) de ces individus - par exemple leur taux de croissance -, que l’on suppose stochastique et dépendant des caractéristiques des ancêtres seulement à travers la caractéristique de la mère. On décrit la structure de dépendance par une transition de probabilité  $P(x, dydz)$  qui donne la probabilité que le couple des caractéristiques des filles soit autour de  $(y, z)$  lorsque la caractéristique de la mère vaut  $x$ . Notez que  $y$ , la caractéristique de la fille de type 0, et  $z$ , celle de la fille de type 1, peuvent être conditionnellement dépendants sachant  $x$ , et que leurs lois conditionnelles peuvent être différentes. La structure de dépendance au sein de l’arbre est complètement décrite lorsqu’on suppose en outre que, étant données les caractéristiques  $(X_i, i \in \mathbb{T}_q)$  des individus des générations 0 à  $q$ , on construit celles de la génération  $q + 1$  en tirant  $2^q$  couples *indépendants*  $(X_{2i}, X_{2i+1})$  suivant  $P(X_i, \cdot)$ ,  $i \in \mathbb{G}_q$ . On parle de chaînes de Markov bifurcantes.

Pour de tels processus stochastiques, on démontre d’abord des lois des grands nombres et des théorèmes de la limite centrale (paragraphe 4.2) que l’on applique ensuite à la détection du vieillissement cellulaire chez *Escherichia Coli*, en utilisant les données de E. J. STEWART *et al.* et le modèle (1.18) (paragraphe 4.3). Résumons brièvement nos résultats.

#### Théorèmes limites

Dans les théorèmes limites que nous démontrons, la loi stationnaire, disons  $\mu$ , de la chaîne de Markov  $(Y_r, r \in \mathbb{N})$ , obtenue en choisissant une lignée “au hasard” dans l’arbre, joue un rôle capital. “Au hasard” signifie que l’on part de l’ancêtre, que l’on choisit une de ses deux filles à pile ou face, et qu’on recommence indéfiniment, tous les tirages à pile ou face étant indépendants. Précisément, on se place dans la cas où la chaîne  $(Y_r, r \in \mathbb{N})$  converge en loi vers une probabilité  $\mu$  indépendante de sa

condition initiale déterministe  $x$ . La loi  $\mu$  est alors l'unique probabilité invariante de la chaîne  $(Y_r, r \in \mathbb{N})$ , et on dit que la chaîne  $(Y_r, r \in \mathbb{N})$  est ergodique. Sous cette hypothèse, on montre que la moyenne empirique de  $f$  sur la génération  $\mathbb{G}_q$ ,

$$\overline{M}_{\mathbb{G}_q}(f) = |\mathbb{G}_q|^{-1} \sum_{i \in \mathbb{G}_q} f(X_i, X_{2i}, X_{2i+1}),$$

ainsi que celle sur les  $r + 1$  premières générations,

$$\overline{M}_{\mathbb{T}_r}(f) = |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} f(X_i, X_{2i}, X_{2i+1}),$$

convergent en moyenne quadratique vers  $\int Pf d\mu$ , où  $Pf(x) = \int_{S^2} f(x, y, z)P(x, dydz)$ , et ce pour toute fonction  $f$  continue bornée. En vue de l'appliquer au modèle (1.18), on montre que ce résultat est également valable pour des ensembles de fonctions  $f$  plus généraux, pouvant en particulier contenir des fonctions non bornées (cf théorème 39). On obtient aussi des conditions suffisantes de convergence presque sûre, incluant notamment le cas où la chaîne  $(Y_r, r \in \mathbb{N})$  converge vite en loi - par exemple si elle est géométriquement ergodique (cf théorème 41 et corollaire 42). En outre, on montre que la convergence de  $\overline{M}_{\mathbb{T}_r}(f)$  vers  $\int Pf d\mu$  a lieu à vitesse  $|\mathbb{T}_r|^{1/2}$ . En particulier, pour toute fonction  $f$  continue bornée à valeurs réelles telle que  $Pf = 0$ ,  $|\mathbb{T}_r|^{1/2} \overline{M}_{\mathbb{T}_r}(f)$  converge en loi vers la gaussienne  $\mathcal{N}(0, \Sigma^2)$ , où  $\Sigma^2 = \int P(f^2) d\mu$  (voir le théorème 46 et les corollaires 47 et 51 pour des énoncés plus généraux).

## Détection du vieillissement cellulaire

On applique ensuite ces résultats au modèle (1.18) : étant donné l'observation des  $r + 1$  premières générations, l'estimateur du maximum de vraisemblance de  $(\alpha_0, \beta_0, \alpha_1, \beta_1, \rho, \sigma^2)$  est fortement convergent quand  $r \rightarrow +\infty$  (cf propositions 58 et 61) ; la version multidimensionnelle du théorème central limite indiqué ci-dessus (corollaire 51) permet de construire des régions de confiance sur les paramètres et des statistiques de test (cf propositions 60, 62, 63, 64 et 65). Les données nous conduisent alors à rejeter fortement  $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ , ainsi que l'hypothèse  $H'_0 = \{\beta_0/(1 - \alpha_0) = \beta_1/(1 - \alpha_1)\}$  d'égalité des points fixes des deux droites de régression du modèle. Le nuage des estimations de ces points fixes (figure 4.16) indique alors que, dans les conditions expérimentales étudiées, les nouveaux pôles croissent en moyenne environ 3% plus vite que les vieux pôles. En pratique, ces travaux ont donc permis de valider l'hypothèse de vieillissement chez *E. Coli*, de quantifier ce dernier et d'étudier sa dépendance en les conditions expérimentales (température, stress, utilisation d'une cellule mutant, quantité de nutriments, etc.).

## 1.5.6 Résultats voisins

Nous avons déjà évoqué au paragraphe 1.5.4 des modèles de lignages cellulaires, proches de (1.18), mais ne tenant pas compte d'une dissymétrie entre les sœurs. Je mentionne maintenant un autre ensemble de travaux [2, 6, 70, 98, 71, 73, 91], plus théoriques, qui étudient les chaînes de Markov indexées par un arbre. Les différents auteurs considèrent des arbres plus généraux que l'arbre binaire, et éventuellement aléatoires, mais tous supposent qu'étant donné la caractéristique de leur mère, les filles ont des caractéristiques indépendantes et de même loi. L'indépendance conditionnelle correspond, dans le cas de l'arbre binaire, à la situation où  $P$  se factorise en un produit  $P_0 \otimes P_1$  de deux probabilités de transition, i.e.  $P(x, dydz) = P_0(x, dy) \otimes P_1(x, dz)$  pour tout  $x \in S$ . L'égalité en loi conditionnelle signifie qu'en plus  $P_0 = P_1$ . Ces travaux de thèse sont donc originaux en ce sens qu'ils autorisent à la fois une dissymétrie entre les sœurs et leur dépendance conditionnelle. En outre, hormis la deuxième partie de [2], tous les articles mentionnés ci-dessus ne traitent que le cas où l'espace d'états  $S$  est discret, alors que nous considérons des espaces d'états généraux.

Amir DEMBO, Peter MÖRTERS et Scott SHEFFIELD [22] démontrent des principes de grandes déviations pour des processus de Galton-Watson multitype. Faïza MAËOUÏA et Abderrahmen TOUATI [73] construisent une région de confiance asymptotique et globale pour les moyennes et les matrices de covariance des lois de reproduction d'un processus de branchement multitype et supercritique. Wen LIU, Weiguo YANG et Liying WANG [70, 98, 71], puis Christiane TAKACS [91] démontrent des lois fortes des grands nombres pour les chaînes de Markov à espace d'états fini indexées par des arbres déterministes particuliers. Krishna ATHREYA and Hye-Jeong KANG utilisent une hypothèse d'ergodicité similaire à celle que nous faisons sur la chaîne  $Y$  pour démontrer des lois des grands nombres. Précisément, ils considèrent des arbres de Galton-Watson sur lesquels se meuvent des particules suivant une chaîne de Markov de transition  $R$  sur l'espace d'états  $S$  et ils supposent qu'il existe une probabilité  $\mu$  telle que  $\lim_{m \rightarrow \infty} R^m(x, A) = \mu(A)$  pour tout événement  $A \in \mathcal{S}$  et pour tout  $x \in S$ . Si cette convergence a lieu uniformément en  $x$  sur les compacts de  $S$ , ils obtiennent une loi faible. Si elle a lieu uniformément sur  $S$ , ils obtiennent une loi forte. Notons que nos résultats ne nécessitent pas d'uniformité en  $x$  - mais notre arbre est déterministe. Signalons aussi les travaux d'Itai BENJAMINI et Yuval PERES [6]. Comme nous, ils considèrent une chaîne de Markov indexée par un arbre déterministe (non nécessairement binaire). Cependant, ils ne s'intéressent pas à des lois des grands nombres, mais étudient des propriétés trajectorielles de ces chaînes. Ils donnent en particulier des conditions nécessaires et suffisantes pour que, presque sûrement, il existe un chemin de longueur infinie dans l'arbre sur lequel la chaîne de Markov prend des valeurs bornées.

Signalons enfin des utilisations des chaînes de Markov indexées par des arbres en

agronomie (par exemple [27]), en classification des sites internet (par exemple [96]) et en classification d'images (par exemple [24]).

### 1.5.7 Extensions possibles

À la suite de cette étude, on peut envisager un grand nombre de directions de recherche possibles. Concluons cette présentation en examinant quelques unes d'entre elles.

Tout d'abord, le modèle (1.18) est source de questions mathématiques encore ouvertes. En particulier, dispose-t-on d'un théorème central limite sur la moyenne  $\overline{M}_{\mathbb{G}_q}(f)$  de  $f$  sur la génération  $q$ ? Peut-on préciser la loi asymptotique de la mesure empirique  $|\mathbb{G}_q|^{-1} \sum_{i \in \mathbb{G}_q} \delta_{X_i}$  des taux de croissance au sein de la génération  $q$ ? Que dire des moyennes empiriques de  $f$  sur des sous-arbres incomplets - différents d'un  $\mathbb{T}_r$  - dont la taille tend vers  $+\infty$ ?

Ensuite, on peut imaginer plusieurs raffinements du modèle (1.18). On peut étudier des modèles markoviens de mémoire supérieure à un, qui expliquent *a priori* le taux de croissance d'un individu non seulement par celui de sa mère, mais aussi par celui de sa grand-mère, voire de son arrière grand-mère, etc. Un test d'utilité des régresseurs permettrait la sélection du modèle adéquat. Par ailleurs, la transformation de l'arbre déterministe  $\mathbb{T}$  en un arbre de Galton-Watson de loi de reproduction  $p_0 = \eta, p_2 = 1 - \eta$  permettrait de rendre compte de la possibilité, observée en pratique, qu'une cellule meure, et lui affecterait la probabilité  $\eta$ . De plus, les données ne fournissent pas uniquement les taux de croissance, mais aussi des données géographiques - coordonnées du centre de la cellule, angle par rapport à une direction fixée - dont on pourrait mesurer le lien éventuel avec les taux de croissance. Les données les plus récentes, obtenues par A. LINDNER, détectent en outre la présence éventuelle d'agrégats de protéines mal formées, vraisemblablement corrélée négativement au taux de croissance. On peut imaginer que le vieillissement est dû à un problème qui est une variable cachée  $C$ , liée à nos observables "taux de croissance" et "présence d'agrégats", et qu'il faudrait inférer au vu de ces derniers. En bref, les questions ne manquent pas, qu'elles concernent les mathématiques ou la modélisation.

# Chapitre 2

## Schéma d'Euler et distributions tempérées

Dans ce chapitre, on reproduit un article paru dans *Stochastic Processes and Their Applications* [40].

Given a smooth  $\mathbb{R}^d$ -valued diffusion  $(X_t^x, t \in [0, 1])$  starting at point  $x$ , we study how fast the Euler scheme  $X_1^{n,x}$  with time step  $1/n$  converges in law to the random variable  $X_1^x$ . Precisely, we look for which class of test functions  $f$  the approximate expectation  $\mathbb{E}[f(X_1^{n,x})]$  converges with speed  $1/n$  to  $\mathbb{E}[f(X_1^x)]$ .

When  $f$  is smooth with polynomially growing derivatives or, under a uniform hypoellipticity condition for  $X$ , when  $f$  is only measurable and bounded, it is known that there exists a constant  $C_1 f(x)$  such that

$$\mathbb{E}[f(X_1^{n,x})] - \mathbb{E}[f(X_1^x)] = C_1 f(x)/n + O(1/n^2). \quad (2.1)$$

If  $X$  is uniformly elliptic, we expand this result to the case when  $f$  is a tempered distribution. In such a case,  $\mathbb{E}[f(X_1^x)]$  (resp.  $\mathbb{E}[f(X_1^{n,x})]$ ) has to be understood as  $\langle f, p(1, x, \cdot) \rangle$  (resp.  $\langle f, p_n(1, x, \cdot) \rangle$ ) where  $p(t, x, \cdot)$  (resp.  $p_n(t, x, \cdot)$ ) is the density of  $X_t^x$  (resp.  $X_t^{n,x}$ ). In particular, (2.1) is valid when  $f$  is a measurable function with polynomial growth, a Dirac mass or any derivative of a Dirac mass. We even show that (2.1) remains valid when  $f$  is a measurable function with exponential growth. Actually our results are symmetric in the two space variables  $x$  and  $y$  of the transition density and we prove that

$$\partial_x^\alpha \partial_y^\beta p_n(t, x, y) - \partial_x^\alpha \partial_y^\beta p(t, x, y) = \partial_x^\alpha \partial_y^\beta \pi(t, x, y)/n + r_n(t, x, y)$$

for a function  $\partial_x^\alpha \partial_y^\beta \pi$  and a  $O(1/n^2)$  remainder  $r_n$  which are shown to have gaussian tails and whose dependence on  $t$  is precised. We give applications to option pricing and hedging, proving numerical convergence rates for prices, deltas and gammas.

## 2.1 Introduction and results

Let  $d, r \geq 1$  be two integers. Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space on which lives a  $r$ -dimensional Brownian motion  $B$ . We denote by  $\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t)$  the filtration generated by  $B$ . Let us give two functions  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times r}$ . We systematically use (column) vector and matrix notations, so that  $b(x)$  should be thought of as a vector of size  $d$  and  $\sigma(x)$  as a matrix of size  $d \times r$ . We denote transposition by a star and define a  $d \times d$  matrix-valued function by putting  $a = \sigma \sigma^*$ . For a multiindex  $\alpha \in \mathbb{N}^d$ ,  $|\alpha| = \alpha_1 + \dots + \alpha_d$  is its length and  $\partial^\alpha$  is the differential operator  $\partial^{|\alpha|} / \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}$ . Equipping  $\mathbb{R}^d$  with the euclidian norm  $\|\cdot\|$ , we denote by

- $C_{\text{pol}}^\infty(\mathbb{R}^d)$  the set of infinitely differentiable functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  with polynomially growing derivatives of any order, i.e. such that for all  $\alpha \in \mathbb{N}^d$ , there exists  $c \geq 0$  and  $q \in \mathbb{N}$  such that for all  $x \in \mathbb{R}^d$ ,

$$|\partial^\alpha f(x)| \leq c(1 + \|x\|^q), \quad (2.2)$$

- $C_b^\infty(\mathbb{R}^d)$  the set of infinitely differentiable functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  with bounded derivatives of any order, i.e. such that  $\partial^\alpha f \in L^\infty(\mathbb{R}^d)$  for all  $\alpha \in \mathbb{N}^d$ .

We shall make use of the following assumptions :

- (A) For all  $i \in \{1, \dots, d\}$  and  $j \in \{1, \dots, r\}$ ,  $b_i$  and  $\sigma_{i,j}$  belong to  $C_{\text{pol}}^\infty(\mathbb{R}^d)$  and have bounded first derivatives.
- (B) For all  $i \in \{1, \dots, d\}$  and  $j \in \{1, \dots, r\}$ ,  $b_i$  and  $\sigma_{i,j}$  belong to  $C_b^\infty(\mathbb{R}^d)$ .
- (C) There exists  $\eta > 0$  such that for all  $x, \xi \in \mathbb{R}^d$ ,  $\xi^* a(x) \xi \geq \eta \|\xi\|^2$ .

(C) is known as the uniform ellipticity condition.

It is well known that, given  $x \in \mathbb{R}$ , the hypothesis (A) guarantees the existence and the  $\mathbb{P}$ -almost sure uniqueness of a solution  $X^x = (X_t^x, t \geq 0)$  of the stochastic differential equation (SDE)

$$X_t^x = x + \int_0^t b(X_s^x) ds + \int_0^t \sigma(X_s^x) dB_s. \quad (2.3)$$

### 2.1.1 Motivation

Let us fix a time horizon  $T > 0$ . Without loss of generality, we can and do assume that  $T = 1$ . We try to estimate the law of  $X_1^x$ . To do so, the most natural idea is to approach  $X^x$  by its Euler scheme of order  $n \geq 1$ , say  $X^{n,x} = (X_t^{n,x}, t \geq 0)$ , defined as follows. We consider the regular subdivision  $\mathfrak{S}_n = \{0 = t_0^n < t_1^n < \dots < t_{n-1}^n < t_n^n = 1\}$  of the interval  $[0, 1]$ , i.e.  $t_k^n = k/n$ , and we put  $X_0^{n,x} = x$  and, for all  $k \in \{0, \dots, n-1\}$  and  $t \in [t_k^n, t_{k+1}^n]$ ,

$$X_t^{n,x} = X_{t_k^n}^{n,x} + b\left(X_{t_k^n}^{n,x}\right) (t - t_k^n) + \sigma\left(X_{t_k^n}^{n,x}\right) (B_t - B_{t_k^n}). \quad (2.4)$$

Then the random variable  $X_1^{n,x}$  is exactly simulatable and should be close in law of  $X_1^x$ . Precisely, we measure the weak error between  $X_1^{n,x}$  and  $X_1^x$  by the quantities

$$\Delta_1^n f(x) = \mathbb{E}[f(X_1^{n,x})] - \mathbb{E}[f(X_1^x)]$$

and we try to find the largest space of test functions  $f$  for which, for each  $x$ , there exists a constant  $C_1 f(x)$  such that

$$\Delta_1^n f(x) = C_1 f(x)/n + O(1/n^2). \quad (2.5)$$

Practical interest of such an expansion has to be underlined (see, for instance, [55, 93]). When (2.5) holds, one can use the Euler scheme plus a Monte-Carlo method to estimate  $\mathbb{E}[f(X_1^x)]$  and then, in a time of order  $nN$ , gets an error of order  $1/\sqrt{N} + 1/n$ , where  $N$  stands for the number of independant copies of  $X_1^{n,x}$  generated by the Monte-Carlo procedure. Given a tolerance  $\varepsilon \ll 1$ , in order to minimize the time of calculus, one should then choose  $N = O(n^2)$  and gets a result in a time of order  $1/\varepsilon^3$ .

One can even do better using Romberg's extrapolation technique: if one runs  $N$  independant copies  $(X_{i,1}^{2n,x}, X_{i,1}^{n,x})$  of the couple  $(X_1^{2n,x}, X_1^{n,x})$ , which still requires a time of order  $nN$ , then computing  $\frac{1}{N} \sum_{i=1}^N (2f(X_{i,1}^{2n,x}) - f(X_{i,1}^{n,x}))$  one gets an estimate of  $\mathbb{E}[f(X_1^x)]$  whose accuracy is of order  $1/\sqrt{N} + 1/n^2$ , since (2.5) implies that  $\mathbb{E}[2f(X_1^{2n,x}) - f(X_1^{n,x})] = \mathbb{E}[f(X_1^x)] + O(1/n^2)$ . Given a tolerance  $\varepsilon \ll 1$ , one should now choose  $N = O(n^4)$  and gets a result in a time of order  $1/\varepsilon^{5/2}$ .

## 2.1.2 Previous results

Using Itô expansions, D. TALAY and L. TUBARO [93] have shown that (2.5) holds when  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  under condition

**(B')** The  $b_i$ 's and the  $\sigma_{i,j}$ 's are infinitely differentiable functions with bounded derivatives of any order  $\geq 1$ .

Hypothesis (B') is almost (B) but in (B') the functions  $b_i$  and  $\sigma_{i,j}$  are not supposed bounded themselves. Using Malliavin calculus, V. BALLY and D. TALAY [3] have extended this result to the case of measurable and bounded  $f$ 's, with the extra hypothesis that  $X$  is uniformly hypoelliptic. If (C) holds,  $X_1^{n,x}$  and  $X_1^x$  have densities, say  $p_n(1, x, \cdot)$  and  $p(1, x, \cdot)$  respectively (in this paper, densities are always taken with respect to the Lebesgue measure). Then, for each pair  $(x, y)$ , the authors [4] get an expansion of the error on the density itself of the form

$$p_n(1, x, y) - p(1, x, y) = \pi(1, x, y)/n + \pi_n(1, x, y)/n^2. \quad (2.6)$$

They also show that the principal error term  $\pi$  and the remainder  $\pi_n$  have gaussian tails. Namely, they find constants  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $n \geq 1$  and  $x, y \in \mathbb{R}^d$ ,  $|\pi(1, x, y)| + |\pi_n(1, x, y)| \leq c_1 \exp(-c_2 \|x - y\|^2)$ .

Besides, V. KONAKOV and E. MAMMEN [61] have proposed an analytical approach for this problem based on the so-called parametrix method. If (B) and (C) hold, for each pair  $(x, y)$ , they get an expansion of arbitrary order  $j$  of  $p_n(1, x, y)$  but whose terms depend on  $n$ :

$$p_n(1, x, y) - p(1, x, y) = \sum_{i=1}^{j-1} \pi_{n,i}(1, x, y)/n^i + O(1/n^j). \quad (2.7)$$

They also prove that the coefficients have gaussian tails, uniformly in  $n$ : for each  $i$ , they find constants  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $n \geq 1$  and  $x, y \in \mathbb{R}^d$ ,  $|\pi_{n,i}(1, x, y)| \leq c_1 \exp(-c_2 \|x - y\|^2)$ . To do so, the authors use upper bounds on the partial derivatives of  $p$  - which they find in [35] - and prove analogous bounds on  $p_n$ 's ones.

A link with generalized Watanabe distributions on Wiener's space is exhibited in [75]. For the general case of Lévy driven stochastic differential equations, (2.5) holds under regularity assumptions on  $f$  and integrability conditions on the Lévy process, see [55, 84]. The rate of convergence of the process  $(X_t^{n,x} - X_t^x, t \in [0, 1])$  is given in [57, 54]. As for the simulation of densities, see for instance [60].

### 2.1.3 Purpose and method

Equations (2.6) and (2.7) can be seen as expansions of  $\Delta_1^n f(x) = \mathbb{E}[f(X_1^{n,x})] - \mathbb{E}[f(X_1^x)]$  in the special case when  $f = \delta_y$ , the Dirac mass at point  $y \in \mathbb{R}^d$ . We aim at giving a precise sense to such quantities when  $f$  is any tempered distribution, and at proving that expansions in powers of  $1/n$  remain valid in this extremely general setting. Moreover, we will derive expansions that are valid not only for  $t = 1$ , but also for any time  $t \in (0, 1]$ , the stepsize  $1/n$  being fixed, and we shall make explicit, in these expansions, the way the coefficients and the remainders depend on  $t$ ,  $f$  and  $x$ .

To get these precise results, we shall place ourselves in a strong situation. Namely, we will assume infinite regularity and boundedness of the coefficients of the SDE (2.3), that is condition (B), and uniform ellipticity, that is condition (C). The reason for this is the following. Let us write  $P_t f(x) = \mathbb{E}[f(X_t^x)]$  and  $P_t^n f(x) = \mathbb{E}[f(X_t^{n,x})]$ . We first expand  $\Delta_t^n = P_t^n - P_t$  as an endomorphism of  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ , in powers of  $1/n$ . This can be done under nothing more than hypothesis (A), see Theorems 13 and 14 in Section 2.1.8. The coefficients in these expansions are operators of the form  $\int_0^t P_s D P_{t-s} ds$  or  $\sum_{t_k^n < t} P_{t_k^n}^n D P_{t-t_k^n}$ , where  $D$  is a differential operator. Now, under (B) and (C), both  $X_t^x$  and  $X_t^{n,x}$  have regular densities, say  $p(t, x, \cdot)$  and  $p_n(t, x, \cdot)$ , with gaussian tails, as soon as  $t > 0$ , so that we may express these operators as



integral operators on  $\mathbb{R}^d$ . For instance, for  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $x \in \mathbb{R}^d$  and  $0 < s < t$ ,

$$\begin{aligned} P_s DP_{t-s} f(x) &= \int_{\mathbb{R}^d} p(s, x, z) DP_{t-s} f(z) dz \\ &= \int_{\mathbb{R}^d} f(y) \left( \int_{\mathbb{R}^d} p(s, x, z) D(p(t-s, \cdot, y))(z) dz \right) dy. \end{aligned}$$

Now the expansions read on the density itself, with coefficients of the form

$$\pi(t, x, y) = \int_0^t \int_{\mathbb{R}^d} p(s, x, z) D(p(t-s, \cdot, y))(z) dz ds, \quad (2.8)$$

$$\text{or } \pi_n(t, x, y) = \sum_{t_k^n < t} \int_{\mathbb{R}^d} p_n(t_k^n, x, z) D(p(t-t_k^n, \cdot, y))(z) dz. \quad (2.9)$$

At this step, the key point is to prove that these coefficients, as well as any of their spatial derivatives, have gaussian tails (see Proposition 9). To do so, we split the above time integral (resp. sum) depending on whether  $s$  (resp.  $t_k^n$ ) is small or large, and integrate by parts in the latter case. This is very similar to V. BALLY and D. TALAY's technique [3], but they use the Malliavin calculus integration by parts formula whereas we only use the genuine one. Then we use upper bounds on the partial derivatives of  $p$  and  $p_n$ , as is done in V. KONAKOV and E. MAMMEN's work [61]. Here the uniform ellipticity hypothesis is crucial: it provides upper bounds that have enough quality in  $t$  to allow us to conclude.

The same analysis, with a bit more work, can be done for the remainders. We then get *functional* expansions of the form

$$p_n - p = \pi/n + \pi_n/n^2 \quad \text{or} \quad p_n - p = \sum_{i=1}^j \pi_{n,i}/n^i \quad (2.10)$$

where  $\pi$ ,  $\pi_n$  and the  $\pi_{n,i}$ 's and all their spatial derivatives have gaussian tails, uniformly in  $n$ . We then achieve to give a distributional sense to expansion (2.5) by a duality approach: any tempered distribution can be integrated or bracketed in the variable  $y$  with the expansions. Theorems 10, 11 and 12 provide precise statements, see Section 2.1.7.

### 2.1.4 A first series of results

Stating Theorems 10 and 12 requires a bit of preparation, namely defining appropriate functional spaces in which will live the coefficients  $\pi$ ,  $\pi_n$  and  $\pi_{n,i}$  in expansions (2.10). Before doing this, to encourage the reader, we would like to state a series of easy consequences of Theorem 10, including an application to financial markets. They will be proved in Section 2.1.7. The function  $\pi$  which appears in them is the

principal *functional* error term. It is defined by (2.21)-(2.22). Note that analogous corollaries can be derived from Theorem 12 as well. The first result gives the rate of convergence of the spatial derivatives of the density:

**Proposition 5.** *Under (B) and (C), for all  $\alpha, \beta \in \mathbb{N}^d$ , there exists  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $n \geq 1$ ,  $t \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,*

$$\partial_x^\alpha \partial_y^\beta p_n(t, x, y) - \partial_x^\alpha \partial_y^\beta p(t, x, y) = \frac{1}{n} \partial_x^\alpha \partial_y^\beta \pi(t, x, y) + r_n(t, x, y)$$

and

$$|r_n(t, x, y)| \leq c_1 n^{-2} t^{-(|\alpha| + |\beta| + d + 4)/2} \exp(-c_2 \|x - y\|^2 / t).$$

The next proposition states that (2.5) is valid for measurable and polynomially growing  $f$ 's:

**Proposition 6.** *Assume (B) and (C). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a measurable function such that there exists  $c' \geq 0$  and  $q \in \mathbb{N}$  such that for all  $x \in \mathbb{R}^d$ ,  $|f(x)| \leq c'(1 + \|x\|^q)$ . Then there exists  $c \geq 0$  such that for all  $n \geq 1$ ,  $t \in (0, 1]$  and  $x \in \mathbb{R}^d$ ,*

$$\mathbb{E}[f(X_t^{n,x})] - \mathbb{E}[f(X_t^x)] = \frac{1}{n} \int_{\mathbb{R}^d} f(y) \pi(t, x, y) dy + r_n(t, x) \quad (2.11)$$

and

$$|r_n(t, x)| \leq cn^{-2} t^{-2} (1 + \|x\|^q).$$

As far as extending the class of  $f$ 's for which (2.5) holds is concerned, we can even do better. Indeed, if for  $\mu \in (0, 2)$  we denote by  $\mathcal{E}_\mu$  the set of all measurable functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  such that there exists  $c_1, c_2 \geq 0$  such that for all  $y \in \mathbb{R}^d$ ,

$$|f(y)| \leq c_1 \exp(c_2 \|y\|^\mu),$$

we have

**Proposition 7.** *Under (B) and (C), for all  $\mu \in (0, 2)$  and  $f \in \mathcal{E}_\mu$ , there exists  $c_1, c_2 \geq 0$  such that for all  $n \geq 1$ ,  $t \in (0, 1]$  and  $x \in \mathbb{R}^d$ ,  $f(X_t^x)$  and  $f(X_t^{n,x})$  are integrable and*

$$\mathbb{E}[f(X_t^{n,x})] - \mathbb{E}[f(X_t^x)] = \frac{1}{n} \int_{\mathbb{R}^d} f(y) \pi(t, x, y) dy + r_n(t, x) \quad (2.12)$$

with

$$|r_n(t, x)| \leq c_1 n^{-2} t^{-2} \exp(c_2 \|x\|^\mu).$$

In particular, (2.5) remains true under (B) and (C) when  $f \in \mathcal{E} = \cup_{\mu \in (0, 2)} \mathcal{E}_\mu$ . More generally, Theorem 10 leads to

**Proposition 8.** Under (B) and (C), for all  $\alpha \in \mathbb{N}^d$ ,  $\mu \in (0, 2)$  and  $f \in \mathcal{E}_\mu$ , there exists  $c_1, c_2 \geq 0$  such that for all  $n \geq 1$ ,  $t \in (0, 1]$  and  $x \in \mathbb{R}^d$ ,

$$\partial_x^\alpha \mathbb{E}[f(X_t^{n,x})] - \partial_x^\alpha \mathbb{E}[f(X_t^x)] = \frac{1}{n} \int_{\mathbb{R}^d} f(y) \partial_x^\alpha \pi(t, x, y) dy + r_n(t, x) \quad (2.13)$$

with

$$|r_n(t, x)| \leq c_1 n^{-2} t^{-(|\alpha|+4)/2} \exp(c_2 \|x\|^\mu).$$

This result can now be used in the context of financial markets.

## 2.1.5 Application to option pricing and hedging

Let  $S^v = (S^{v,1}, \dots, S^{v,d})$  be a basket of assets satisfying

$$\frac{dS_t^{v,i}}{S_t^{v,i}} = \mu_i(S_t^v) dt + \sum_{j=1}^r \sigma_{i,j}(S_t^v) dB_t^j, \quad S_0^{v,i} = v^i > 0,$$

with  $\mu, \sigma \in C_b^\infty(\mathbb{R}^d)$  and  $\sigma$  satisfying (C). Given a measurable and polynomially growing function  $\phi$ , we try to estimate the price  $\text{Price} = \mathbb{E}[\phi(S_t^v)]$ , the deltas  $\text{Delta}_i = \partial_v^{e_i} \mathbb{E}[\phi(S_t^v)]$  and the gammas  $\text{Gamma}_{i,j} = \partial_v^{e_i + e_j} \mathbb{E}[\phi(S_t^v)]$  of the european option of maturity  $t$  and payoff  $\phi$  ( $(e_1, \dots, e_d)$  is the canonical base of  $\mathbb{R}^d$ ). To do so, let us set  $x = \ln v$  (i.e.  $x^i = \ln v^i$ ) and  $X_t^{x,i} = \ln(S_t^{v,i})$ . Then  $X$  is the solution of (2.3) with  $b = \mu - \|\sigma\|^2/2 \in C_b^\infty(\mathbb{R}^d)$ , where  $\|\sigma\|_i^2(x) = \sum_{j=1}^r \sigma_{i,j}^2(x)$ . If we set  $\exp(x) = (\exp(x^1), \dots, \exp(x^d))$  and  $f(x) = \phi(\exp(x))$ , we define a function  $f \in \mathcal{E}_1$  and, since  $\text{Price} = \mathbb{E}[f(X_t^x)]$ , (2.12) leads to

$$\text{Price}^n - \text{Price} = C_t^{\text{Price}} \phi(v)/n + O(n^{-2} t^{-2} \exp(c_2 \|\ln v\|)),$$

where  $\text{Price}^n$  stands for the approximated price  $\mathbb{E}[f(X_t^{n,x})]$  and

$$C_t^{\text{Price}} \phi(v) = \int_{(\mathbb{R}_+^*)^d} \phi(u) \frac{\pi(t, \ln v, \ln u)}{u_1 \cdots u_d} du.$$

Besides, if we set  $\text{Delta}_i^n = \partial_v^{e_i} \mathbb{E}[f(X_t^{n, \ln v})]$  and  $\text{Gamma}_{i,j}^n = \partial_v^{e_i + e_j} \mathbb{E}[f(X_t^{n, \ln v})]$ , (2.13) shows that

$$\begin{aligned} \text{Delta}^n - \text{Delta} &= C_t^{\text{Delta}} \phi(v)/n + O(n^{-2} t^{-5/2} \exp(c_2 \|\ln v\|)), \\ \text{Gamma}^n - \text{Gamma} &= C_t^{\text{Gamma}} \phi(v)/n + O(n^{-2} t^{-3} \exp(c_2 \|\ln v\|)), \end{aligned}$$

where

$$\begin{aligned} C_t^{\text{Delta}} \phi(v)_i &= \frac{1}{v_i} \int_{(\mathbb{R}_+^*)^d} \phi(u) \frac{\partial_2^{e_i} \pi(t, \ln v, \ln u)}{u_1 \cdots u_d} du, \\ C_t^{\text{Gamma}} \phi(v)_{i,j} &= \frac{1}{v_i v_j} \int_{(\mathbb{R}_+^*)^d} \phi(u) \frac{\partial_2^{e_i + e_j} \pi(t, \ln v, \ln u) - \mathbf{1}_{\{i=j\}} \partial_2^{e_i} \pi(t, \ln v, \ln u)}{u_1 \cdots u_d} du. \end{aligned}$$

Eventually we have proved that applying the Euler scheme of order  $n$  to the logarithm of the underlying leads to approximations of the price, the deltas and the gammas which converge to the true price, deltas and gammas with speed  $1/n$ , at least when the drift and volatility of the underlying satisfy (B) and (C), which in the context of financial markets seems not to be a restricting hypothesis. Note that the principal part of the error explodes as  $t$  tends to 0 as  $t^{-1/2}$  for the prices,  $t^{-1}$  for the deltas and  $t^{-3/2}$  for the gammas.

### 2.1.6 Some functional spaces

In order to state our main results (Proposition 9 and Theorems 10 and 12) precisely and shortly, let us introduce some families of functional spaces. Functional expansions like (2.10) will take place in such spaces. For  $l \in \mathbb{Z}$ , we first define  $\mathcal{G}_l(\mathbb{R}^d)$  as the set of all measurable functions  $\pi : (0, 1] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that

- for all  $t \in (0, 1]$ ,  $\pi(t, \cdot, \cdot)$  is infinitely differentiable,
- for all  $\alpha, \beta \in \mathbb{N}^d$ , there exists two constants  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $t \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,

$$|\partial_x^\alpha \partial_y^\beta \pi(t, x, y)| \leq c_1 t^{-(|\alpha| + |\beta| + d + l)/2} \exp(-c_2 \|x - y\|^2 / t). \quad (2.14)$$

We say that a subset  $\mathcal{B} \subset \mathcal{G}_l(\mathbb{R}^d)$  is bounded if, in (2.14),  $c_1$  and  $c_2$  can be chosen independently on  $\pi \in \mathcal{B}$ . We also introduce the space  $\mathcal{G}(\mathbb{R}^d)$  defined in the same way as  $\mathcal{G}_l(\mathbb{R}^d)$  with (2.14) replaced by the following two conditions :

$$|\partial_x^\alpha \partial_y^\beta \pi(t, x, y)| \leq c_1 t^{-(|\alpha| + |\beta| + d)/2} \exp(-c_2 \|x - y\|^2 / t), \quad (2.15)$$

$$\left| \partial_x^\alpha \left( \pi \left( t, x, x + y\sqrt{t} \right) \right) \right| \leq c_1 t^{-d/2} \exp(-c_2 \|y\|^2). \quad (2.16)$$

Note that we may always take the couple of constants  $(c_1, c_2)$  to be the same in both equations (2.15) and (2.16). Indeed, if they hold with two couples  $(c'_1, c'_2)$  and  $(c''_1, c''_2)$ , they both hold with  $(c_1, c_2)$  if we take  $c_1 = c'_1 \vee c''_1$  and  $c_2 = c'_2 \wedge c''_2$ . We say that a subset  $\mathcal{B} \subset \mathcal{G}(\mathbb{R}^d)$  is bounded if, in (2.15) and (2.16),  $c_1$  and  $c_2$  can be chosen independently on  $\pi \in \mathcal{B}$ . Note that in equation (2.16), the upper bound keeps the same quality in  $t$ , namely  $t^{-d/2}$ , whatever the “number”  $\alpha$  of times one differentiates the mapping  $x \mapsto \pi(t, x, x + y\sqrt{t})$ . This will be crucial when proving Proposition 9.

It is convenient to extend these definitions to mappings that also depend on an intermediate time  $s \in (0, t)$ . To do so, let us denote by  $\mathcal{T}_1$  the unit triangle  $\{(s, t) \in \mathbb{R}^2 | 0 < s < t \leq 1\}$  and, for  $l \in \mathbb{Z}$ , let us define  $\mathcal{H}_l(\mathbb{R}^d)$  as the space of measurable functions  $\rho : \mathcal{T}_1 \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that

- for all  $(s, t) \in \mathcal{T}_1$ ,  $\rho(s, t, \cdot, \cdot)$  is infinitely differentiable,

- for all  $\alpha, \beta \in \mathbb{N}^d$ , there exists two constants  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $(s, t) \in \mathcal{T}_1$  and  $x, y \in \mathbb{R}^d$ ,

$$|\partial_x^\alpha \partial_y^\beta \rho(s, t, x, y)| \leq c_1 t^{-(|\alpha| + |\beta| + d + l)/2} \exp(-c_2 \|x - y\|^2 / t). \quad (2.17)$$

Again we say that a subset  $\mathcal{B} \subset \mathcal{H}_l(\mathbb{R}^d)$  is bounded if, in (2.17),  $c_1$  and  $c_2$  can be chosen independently on  $\rho \in \mathcal{B}$ . We also introduce the space  $\mathcal{H}(\mathbb{R}^d)$  which is defined in the same way as  $\mathcal{H}_l(\mathbb{R}^d)$  with (2.17) replaced by

$$|\partial_x^\alpha \partial_y^\beta \rho(s, t, x, y)| \leq c_1 t^{-(|\alpha| + |\beta| + d)/2} \exp(-c_2 \|x - y\|^2 / t), \quad (2.18)$$

$$\left| \partial_x^\alpha \left( \rho \left( s, t, x, x + y\sqrt{t} \right) \right) \right| \leq c_1 t^{-d/2} \exp(-c_2 \|y\|^2), \quad (2.19)$$

and we say that a subset  $\mathcal{B} \subset \mathcal{H}(\mathbb{R}^d)$  is bounded if, in (2.18) and (2.19),  $c_1$  and  $c_2$  can be chosen independently on  $\rho \in \mathcal{B}$ . Again we may always choose the couple  $(c_1, c_2)$  to be the same in both equations (2.18) and (2.19). Note that the upper bounds in (2.17), (2.18) and (2.19) are exactly the same as the ones in (2.14), (2.15) and (2.16). In particular, they do not depend on  $s$ .

Eventually, for  $\pi_1, \pi_2 \in \mathcal{G}(\mathbb{R}^d)$ ,  $g \in C_b^\infty(\mathbb{R}^d)$  and  $\gamma \in \mathbb{N}^d$ , we define a function  $\pi_1 *_{g, \gamma} \pi_2$  on  $\mathcal{T}_1 \times \mathbb{R}^d \times \mathbb{R}^d$  by putting

$$(\pi_1 *_{g, \gamma} \pi_2)(s, t, x, y) = \int_{\mathbb{R}^d} g(z) \pi_1(s, x, z) \partial_2^\gamma \pi_2(t - s, z, y) dz.$$

Notation  $\partial_2$  means differentiation with respect to the second argument, here  $z$ . Operation  $*_{g, \gamma}$  is a space convolution which naturally appears when developping the differential operator  $D$  in equations (2.8) and (2.9).

### 2.1.7 Main results

We are now able to state our main results as follows.

**Proposition 9.** *Let  $\mathcal{B}_1$  and  $\mathcal{B}_2$  be two bounded subsets of  $\mathcal{G}(\mathbb{R}^d)$ ,  $g \in C_b^\infty(\mathbb{R}^d)$  and  $\gamma \in \mathbb{N}^d$ . Then*

- (i)  $\{\pi_1 *_{g, \gamma} \pi_2 \mid \pi_1 \in \mathcal{B}_1, \pi_2 \in \mathcal{B}_2\}$  is a bounded subset of  $\mathcal{H}_{|\gamma|}(\mathbb{R}^d)$ ,
- (ii)  $\{\pi_1 *_{g, 0} \pi_2 \mid \pi_1 \in \mathcal{B}_1, \pi_2 \in \mathcal{B}_2\}$  is a bounded subset of  $\mathcal{H}(\mathbb{R}^d)$ .

**Theorem 10.** *Under (B) and (C),*

- (i) for all  $t \in (0, 1]$  and  $x \in \mathbb{R}^d$ ,  $X_t^x$  has a density  $p(t, x, \cdot)$  and  $p \in \mathcal{G}(\mathbb{R}^d)$ ,
- (ii) for all  $t \in (0, 1]$ ,  $x \in \mathbb{R}^d$  and  $n \geq 1$ ,  $X_t^{n, x}$  has a density  $p_n(t, x, \cdot)$  and  $(p_n, n \geq 1)$  is a bounded sequence in  $\mathcal{G}(\mathbb{R}^d)$ ,
- (iii) there exists  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$  and a bounded sequence  $(\pi_n, n \geq 1)$  in  $\mathcal{G}_4(\mathbb{R}^d)$  such that for all  $n \geq 1$ ,

$$p_n - p = \pi/n + \pi_n/n^2. \quad (2.20)$$

These results are proved in Section 2.3.2. In Theorem 10, statement (i) is already known, see [35], Theorem 7, page 260, and statement (ii) has essentially been proved in [61]. As explained in Section 2.1.3, Proposition 9, together with these two statements, is the key to derive statement (iii).

The function  $\pi$  can be expressed in terms of  $p$  by

$$\pi(t, x, y) = \frac{1}{2} \int_0^t \int_{\mathbb{R}^d} p(s, x, z) L_2^*(p(t-s, \cdot, y))(z) dz ds, \quad (2.21)$$

where the differential operator  $L_2^*$  is explicitly given in terms of the functions  $a$  and  $b$  by

$$\begin{aligned} -L_2^* &= \sum_{i=1}^d \left( b \cdot \nabla b_i + \frac{1}{2} \operatorname{tr} (a \nabla^2 b_i) \right) \partial_i \\ &+ \sum_{i,j=1}^d \left( \frac{1}{2} b \cdot \nabla a_{i,j} + a_j \cdot \nabla b_i + \frac{1}{4} \operatorname{tr} (a \nabla^2 a_{i,j}) \right) \partial_{ij} + \frac{1}{2} \sum_{i,j,k=1}^d a_k \cdot \nabla a_{i,j} \partial_{ijk}. \end{aligned} \quad (2.22)$$

Here,  $\cdot$ ,  $a_k$ ,  $\operatorname{tr}$ ,  $\nabla$  and  $\nabla^2$  respectively stand for the inner product in  $\mathbb{R}^d$ , the  $k$ -th column of  $a$ , the trace of a matrix, the gradient vector and the hessian matrix. In the case when  $t = 1$ , (2.21) agrees with V. BALLY and D. TALAY's expression for  $\pi$  ([4], definition 2.2, page 100), but seems preferable because it does not involve differentiation with respect to  $t$  and makes explicit that the space differential operator  $L_2^*$  is of order less than 3, when V. BALLY and D. TALAY's operator  $\mathcal{U}$  involves a fourth order differentiation in space.

We shall now prove that if  $X$  is elliptic the expansion (2.5) is valid in the very general case when  $f$  is a tempered distribution. Let us denote by  $\mathcal{S}(\mathbb{R}^d)$  Schwartz's space, i.e. the space of infinitely differentiable functions  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $x \mapsto x^\alpha \partial^\beta \varphi(x) \in L^\infty(\mathbb{R}^d)$  for all  $\alpha, \beta \in \mathbb{N}^d$  ( $x^\alpha$  stands for  $x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ ), and let us denote by  $\mathcal{S}'(\mathbb{R}^d)$  the space of tempered distributions. The seminorms  $(N_q, q \in \mathbb{N})$  are defined on  $\mathcal{S}(\mathbb{R}^d)$  by

$$N_q(\varphi) = \sum_{|\alpha| \leq q, |\beta| \leq q} \sup_{x \in \mathbb{R}^d} |x^\alpha \partial^\beta \varphi(x)|,$$

and the order  $\#S$  of  $S \in \mathcal{S}'(\mathbb{R}^d)$  is the smallest integer  $q$  such that there is a  $c \geq 0$  such that  $|\langle S, \varphi \rangle| \leq c N_q(\varphi)$  for all  $\varphi \in \mathcal{S}(\mathbb{R}^d)$ . Note that whenever  $\pi \in \mathcal{G}_l(\mathbb{R}^d)$ ,  $\pi(t, x, \cdot)$  and  $\pi(t, \cdot, y)$  belong to  $\mathcal{S}(\mathbb{R}^d)$ . More precisely, for  $\mathcal{B} \subset \mathcal{G}_l(\mathbb{R}^d)$  bounded, there exists  $c \geq 0$  such that for all  $\pi \in \mathcal{B}$ ,  $t \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,

$$N_q(\pi(t, x, \cdot)) \leq ct^{-(d+l+q)/2} (1 + \|x\|^q) \quad \text{and} \quad N_q(\pi(t, \cdot, y)) \leq ct^{-(d+l+q)/2} (1 + \|y\|^q).$$

Applying a tempered distribution  $S$  to (2.20),  $t$  and  $x$  or  $t$  and  $y$  being fixed, we immediately deduce from Theorem 10

**Theorem 11.** *Under (B) and (C), for all  $S \in \mathcal{S}'(\mathbb{R}^d)$ , there exists  $c \geq 0$  such that for all  $n \geq 1$ ,  $t \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,*

$$\begin{aligned}\langle S, p_n(t, x, \cdot) \rangle - \langle S, p(t, x, \cdot) \rangle &= \frac{1}{n} \langle S, \pi(t, x, \cdot) \rangle + r'_n(t, x), \\ \langle S, p_n(t, \cdot, y) \rangle - \langle S, p(t, \cdot, y) \rangle &= \frac{1}{n} \langle S, \pi(t, \cdot, y) \rangle + r''_n(t, y),\end{aligned}$$

and

$$|r'_n(t, x)| + |r''_n(t, x)| \leq cn^{-2}t^{-(d+4+\#S)/2} \left(1 + \|x\|^{\#S}\right).$$

Let us define  $\mathbb{E}[S(Y)]$  by  $\langle S, p_Y \rangle$  when  $S \in \mathcal{S}'(\mathbb{R}^d)$  and  $Y$  is a random variable with density  $p_Y \in \mathcal{S}(\mathbb{R}^d)$ . Note that, when  $S$  is a measurable and polynomially growing function, this definition coincides with the usual expectation. We then have proved that, under (B) and (C), (2.5) is valid for  $f$ 's being only tempered distributions, and not only for  $t = 1$ , but also for any time  $t \in (0, 1]$ , and we have even precised the way the  $O(1/n^2)$  remainder depends on  $t$ ,  $f$  and  $x$ . Precisely, this remainder grows slower than  $\|x\|^{\#f}$  as  $x$  tends to infinity, and explodes slower than  $t^{-(\#f+d+4)/2}$  as  $t$  tends to 0.

We can now prove the propositions stated in Section 2.1.4. Proposition 5 is immediate from Theorem 10. In the special case when  $S$  is a measurable and polynomially growing function, we get Proposition 6:

*Proof of Proposition 6.* Multiplying (2.20) by  $f(y)$  and integrating in  $y$  leads to (2.11) with the remainder  $r_n(t, x) = n^{-2} \int_{\mathbb{R}^d} f(y) \pi_n(t, x, y) dy$ . Since  $|f(y)| \leq c'(1 + \|y\|^q)$  and  $(\pi_n, n \geq 1)$  is bounded in  $\mathcal{G}_4(\mathbb{R}^d)$ , we can find  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $n \geq 1$ ,  $t \in (0, 1]$  and  $x \in \mathbb{R}^d$ ,  $|r_n(t, x)| \leq c_1 n^{-2} t^{-(d+4)/2} \int_{\mathbb{R}^d} (1 + \|y\|^q) \exp(-c_2 \|x - y\|^2/t) dy$ . Setting  $\zeta = (y - x)/\sqrt{t}$  leads to  $|r_n(t, x)| \leq c_1 n^{-2} t^{-2} \int_{\mathbb{R}^d} (1 + \|x + \zeta\sqrt{t}\|^q) \exp(-c_2 \|\zeta\|^2) d\zeta$ . To complete the proof, it remains to observe that there exists  $c \geq 0$  such that for all  $t \in (0, 1]$  and  $x, \zeta \in \mathbb{R}^d$ ,  $\|x + \zeta\sqrt{t}\|^q \leq c(\|x\|^q + \|\zeta\|^q)$ .  $\square$

It is easy to adapt the preceding proof to get Proposition 7. In the same way, differentiating (2.5)  $\alpha$  times in  $x$ , multiplying by  $f(y)$  and integrating in  $y$  leads to Proposition 8.

Expansion (2.20) should be seen as an improvement of (2.6): it allows for infinite differentiation in  $x$  and  $y$  and also precises the way the coefficients explode when  $t$  tends to 0. We have an analogous improvement for expansion (2.7):

**Theorem 12.** *Under (B) and (C), for each  $i \geq 1$ , there exists a bounded family  $(\pi_{n,i}, n \geq 1)$  in  $\mathcal{G}_{2i-2}(\mathbb{R}^d)$  and two bounded families  $(\pi'_{n,i}, n \geq 1)$  and  $(\pi''_{n,i}, n \geq 1)$  in  $\mathcal{G}_{2i}(\mathbb{R}^d)$  such that for all  $j, n \geq 1$ ,*

$$p_n - p = \sum_{i=1}^{j-1} \frac{\pi_{n,i}}{n^i} + \sum_{i=2}^j (t - \lfloor nt \rfloor / n)^i \pi'_{n,i} + \frac{\pi''_{n,j}}{n^j}. \quad (2.23)$$

Here and in all the sequel we use the convention that a sum over an empty set is zero, and  $\lfloor nt \rfloor$  denotes the greatest integer less than or equal to  $nt$ . Expressions involving  $\lfloor nt \rfloor$  do not appear in (2.20) since they are hidden in the remainder. When  $t = 1$  and no differentiation is applied neither in  $x$  nor in  $y$ , (2.23) boils down to the result of V. KONAKOV and E. MAMMEN [61]. Again note that (2.23) is much richer in the sense that it allows for infinite differentiation in space and also precises the dependence on  $t$ . Theorem 12 will also be proved in Section 2.3.2.

### 2.1.8 A preliminary result

As explained in section 2.1.3, in order to prove point (iii) in Theorem 10, we first seek an expansion for the error operator

$$\Delta_t^n = P_t^n - P_t$$

where, for  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ , we have set  $P_t f(x) = \mathbb{E}[f(X_t^x)]$  and  $P_t^n f(x) = \mathbb{E}[f(X_t^{n,x})]$ . Precisely, we look for operators  $C_t$  and  $R_t^n$  such that  $R_t^n = O(1/n^2)$  and  $\Delta_t^n = C_t/n + R_t^n$ . The following theorem, interesting in itself, is proved in Section 2.2. It can be seen as an improvement of [93]. It not only gives explicit formulas for  $C_t f(x)$  and  $R_t^n f(x)$  but also provides useful information about their dependencies on  $n, t, f$  and  $x$ . Note that it does not require neither (B) nor (B') nor (C). In order to state it shortly, let us

- denote by  $\mathcal{L}(C_{\text{pol}}^\infty(\mathbb{R}^d))$  the space of endomorphisms of  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,
- say that a subset  $\mathcal{B} \subset C_{\text{pol}}^\infty(\mathbb{R}^d)$  is bounded if, in (2.2),  $c$  and  $q$  can be chosen independently on  $f \in \mathcal{B}$ ,
- say that  $T \in \mathcal{L}(C_{\text{pol}}^\infty(\mathbb{R}^d))$  is bounded if for all bounded  $\mathcal{B} \subset C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $\{Tf | f \in \mathcal{B}\}$  is a bounded subset of  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,
- denote by  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$  the space of bounded endomorphisms of  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,
- say that a  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ -valued family  $(T_i, i \in I)$  is bounded if for all bounded  $\mathcal{B} \subset C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $\{T_i f | f \in \mathcal{B}, i \in I\}$  is a bounded subset of  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,
- say that  $(T_i, i \in I)$  is a  $O(h(i))$  family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$  if the family  $(h(i)^{-1}T_i, i \in I)$  is bounded.

It is already known that, under (A),  $(P_t, t \in [0, 1])$  is a bounded family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . A proof can be found in [67], Lemma 3.9, page 15. Using Lemma 29, this proof



straightforwardly adapts uniformly in  $n$  so that  $(P_t^n, t \in [0, 1], n \geq 1)$  is also bounded in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . We are now in the position to state the main result of the first step:

**Theorem 13.** *Under (A),  $(\Delta_t^n, t \in [0, 1], n \geq 1)$  is a  $O(t/n)$  family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , and there exists a  $O(t)$  process  $(C_t, t \in [0, 1])$  and a  $O(1/n^2)$  family  $(R_t^n, t \in [0, 1], n \geq 1)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$  such that*

$$\Delta_t^n = C_t/n + R_t^n.$$

Moreover,  $C_t$  is explicitly given in terms of  $(P_t, t \in [0, 1])$  and of  $L_2^*$  (see (2.22)) by

$$C_t = \frac{1}{2} \int_0^t P_s L_2^* P_{t-s} ds. \quad (2.24)$$

Note that this theorem covers the result of D. TALAY and L. TUBARO [93] since it implies that for any  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  we can find a  $q \in \mathbb{N}$  such that for all  $x \in \mathbb{R}^d$ ,  $t \in [0, 1]$  and  $n \geq 1$ ,

$$\Delta_t^n f(x) = C_t f(x)/n + O\left(\frac{1 + \|x\|^q}{n^2}\right).$$

It even improves it a bit since we see that this holds under nothing more than condition (A), whereas D. TALAY and L. TUBARO state their result under the stronger condition (B'). Note also that if we restrict ourselves to times  $t$  belonging to the discretization grid  $\mathfrak{S}_n$ , we get a better control, of order  $O(t/n^2)$ , of the remainder, see Remark 18.

Instead of Theorem 13, in order to derive Theorem 12, we shall need

**Theorem 14.** *Under (A), there exists a sequence of differential operators  $(L_j^*, j \geq 2)$ , recursively defined by (2.27)-(2.28), and for each  $i \geq 1$  a  $O(t/n^i)$  family  $R^i = (R_t^{n,i}, t \in [0, 1], n \geq 1)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$  such that for all  $j \geq 1$ ,*

$$\Delta_t^n = \sum_{i=2}^j \frac{1}{i!n^i} \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k^n}^n L_i^* P_{t-t_k^n} + R_t^{n,j} + \sum_{i=2}^j \frac{(t - \lfloor nt \rfloor/n)^i}{i!} P_{\lfloor nt \rfloor/n}^n L_i^* P_{t-\lfloor nt \rfloor/n}. \quad (2.25)$$

Observe that the main term in (2.25) is

$$\frac{1}{n} \left( \frac{1}{2n} \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k^n}^n L_2^* P_{t-t_k^n} \right) \approx \frac{C_t}{n},$$

and the remainder is of order  $1/n^2$ . Note also that if we restrict ourselves to times belonging to the discretization grid  $\mathfrak{S}_n$ , we get the following expansion in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ :

$$\Delta_{\lfloor nt \rfloor/n}^n = \sum_{i=2}^j \frac{1}{i!n^i} \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k^n}^n L_i^* P_{\lfloor nt \rfloor/n - t_k^n} + O\left(\frac{t}{n^j}\right).$$

Theorem 14 is also proved in Section 2.2.

## 2.1.9 Organization of the paper

Section 2.2 deals with the expansion for the expectation: it is dedicated to the proofs of Theorems 13 and 14.

Section 2.3 is our second and final step. It is devoted to the proofs of Theorems 10 and 12. It begins with the proof of Proposition 9 concerning the space convolution  $*_{g,\gamma}$  in  $\mathcal{G}(\mathbb{R}^d)$ .

Eventually, Section 2.4 is an appendix where we have gathered useful results on the Euler scheme and technical lemmas that are used in Sections 2.2 and 2.3.

## 2.2 First step: expansion for $\mathbb{E}[f(X_t^{n,x})]$

In this section we seek to expand  $\Delta_t^n f(x) = \mathbb{E}[f(X_t^{n,x})] - \mathbb{E}[f(X_t^x)]$  in powers of the time step  $1/n$  when  $f$  is a regular function, say  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ . The idea is the following. Recall the discussion preceding Theorem 13: under (A), both  $P_t$  and  $P_t^n$  are endomorphisms of  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ . In  $\mathcal{L}(C_{\text{pol}}^\infty(\mathbb{R}^d))$  we then write

$$\Delta_t^n = P_t^n - P_t = \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k^n}^n \Delta_{1/n}^n P_{t-t_{k+1}^n} + P_{\lfloor nt \rfloor / n}^n \Delta_{t - \lfloor nt \rfloor / n}^n. \quad (2.26)$$

There is a subtle point here:  $(X_t^{n,x}, t \in [0, 1])$  is not a Markov process, since the future of  $X_t^{n,x}$  depends on the past value  $X_{\lfloor nt \rfloor / n}^{n,x}$ , see (2.4). Nevertheless, it is easy to check by conditioning on  $\mathcal{F}_{t_k^n}$  that we have  $P_{t_k^n}^n P_s^n = P_{t_k^n + s}^n$  for all  $s \geq 0$  - but beware: this is different from  $P_s^n P_{t_k^n}^n$  as soon as  $ns$  is not an integer.

Equation (2.26) leads us to expand  $\Delta_t^n$  for small  $t$ , namely for  $t \leq 1/n$ . This naturally involves a series of differential operators as we shall now see.

### 2.2.1 Operators associated with the Euler scheme

Let us denote by  $L$  the infinitesimal generator of the diffusion  $X$  and by  $(L^x, x \in \mathbb{R}^d)$  its tangent infinitesimal generator, i.e.

$$L = \sum_{i=1}^d b_i \partial^{e_i} + \frac{1}{2} \sum_{i,j=1}^d a_{i,j} \partial^{e_i + e_j} \quad \text{and} \quad L^x = \sum_{i=1}^d b_i(x) \partial^{e_i} + \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x) \partial^{e_i + e_j}.$$

We use the convention that  $L$  and  $L^x$  act on the variable  $y$ , so that, for instance,  $L\psi(t, x, y)$  and  $L^x\psi(t, x, y)$  respectively stand for  $L(\psi(t, x, \cdot))(y)$  and  $L^x(\psi(t, x, \cdot))(y)$ .  $L^x$  is the infinitesimal generator of the Euler scheme  $(X_t^{n,x}, t \in [0, 1/n])$  starting from  $x$ , over the first discretization time interval:  $L^x$  is built from  $L$  in the same way as  $X^n$  is built from  $X$ , by freezing the drift  $b$  and the volatility  $\sigma$  to their initial value on

discretization intervals. Besides, for each  $x \in \mathbb{R}^d$  we define a sequence of differential operators  $(L_j^x, j \in \mathbb{N})$  by putting  $L_0^x = I$  (the identity operator) and

$$L_{j+1}^x = L^x L_j^x - L_j^x L, \quad (2.27)$$

and we set

$$L_j^* f(x) = L_j^x f(x). \quad (2.28)$$

Observe that  $L_1^* = 0$ . Besides,  $L_2^*$  is given by (2.22) so that, under (A),  $L_2^* \in \mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$  and there exists a family  $(g_{2,\alpha}^*, 1 \leq |\alpha| \leq 3)$  in  $C_{\text{pol}}^\infty(\mathbb{R}^d)$  such that

$$L_2^* = \sum_{1 \leq |\alpha| \leq 3} g_{2,\alpha}^* \partial^\alpha. \quad (2.29)$$

$L_2^*$  gives the exact principal error term in the expansion of  $\Delta_t^n$ , see (2.24) and (2.21).  $L_j^*$  is the differential operator appearing in (2.25). It does not give the exact expansion in powers of  $1/n$  but an approximated version, in the spirit of [61], since in (2.25) the coefficients depend on  $n$  - but should themselves be expanded in powers of  $1/n$ . See [55], equations (6.35) and (6.36), for an expression of the operators involved in the exact expansion.

Under (A),  $L$  and  $L^x$  belong to  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$  for each  $x \in \mathbb{R}^d$ , and, by induction, so does  $L_j^x$ . We can describe  $L_j^x$  more precisely. Indeed, defining the powers of an operator  $A$  by  $A^0 = I$  and  $A^{j+1} = AA^j$ , inductions on  $j$  lead to  $L_j^x = \sum_{i=0}^j (-1)^i \binom{j}{i} (L^x)^{j-i} L^i$  and to the existence of a family  $(g_{j,\alpha}, h_{j,\alpha}, j \in \mathbb{N}^*, 1 \leq |\alpha| \leq 2j)$  in  $C_{\text{pol}}^\infty(\mathbb{R}^d)$  such that

$$\forall x \in \mathbb{R}^d, \quad (L^x)^j = \sum_{1 \leq |\alpha| \leq 2j} g_{j,\alpha}(x) \partial^\alpha \quad \text{and} \quad L^j = \sum_{1 \leq |\alpha| \leq 2j} h_{j,\alpha} \partial^\alpha.$$

Hence, for each  $j \in \mathbb{N}^*$  one can find a family  $(m_{j,\alpha}, 1 \leq |\alpha| \leq 2j)$  of integers and a family  $(g_{j,\alpha,l}, h_{j,\alpha,l}, 1 \leq |\alpha| \leq 2j, 1 \leq l \leq m_{j,\alpha})$  in  $C_{\text{pol}}^\infty(\mathbb{R}^d)$  such that for all  $x \in \mathbb{R}^d$ ,

$$L_j^x = \sum_{1 \leq |\alpha| \leq 2j} \left( \sum_{l=1}^{m_{j,\alpha}} g_{j,\alpha,l}(x) h_{j,\alpha,l} \right) \partial^\alpha. \quad (2.30)$$

**Remark 15.** Note that when (B) holds, the functions  $g_{j,\alpha,l}$ ,  $h_{j,\alpha,l}$  and  $g_{2,\alpha}^*$  all belong to  $C_b^\infty(\mathbb{R}^d)$  (in fact they are polynomial in  $b$ ,  $\sigma$  and their derivatives).

We are now in the position to define a family of operators  $\Phi^j = (\Phi_{s,t}^{n,j}, n \geq 1, 0 \leq s \leq t \leq 1/n)$  as follows :

$$\forall f \in C_{\text{pol}}^\infty(\mathbb{R}^d), \quad \Phi_{s,t}^{n,j} f(x) = \mathbb{E} [L_j^x P_{t-s} f(X_s^{n,x})]. \quad (2.31)$$

Observe that  $\Phi_{0,t}^{n,j} = L_j^* P_t$  and that, from (2.30),

$$\Phi_{s,t}^{n,j} = \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} g_{j,\alpha,l} P_s^n (h_{j,\alpha,l} \partial^\alpha P_{t-s}). \quad (2.32)$$

Boundedness is a key property of this family:

**Proposition 16.** *Under (A),  $\Phi^j$  is a bounded family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ .*

*Proof.*  $(P_t, t \in [0, 1])$  and  $(P_t^n, t \in [0, 1], n \geq 1)$  are bounded families in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , see the discussion preceding Theorem 13. Besides, multiplication by a function in  $C_{\text{pol}}^\infty(\mathbb{R}^d)$  and differentiation are bounded operators on  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ . As a sum of compositions of bounded families in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ ,  $\Phi^j$  is a bounded family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ .  $\square$

The family  $\Phi^j$  naturally appears when we recursively use Itô's formula to expand  $\Delta_t^n$  for small  $t$ , as we now explain.

## 2.2.2 Itô expansions

We recall (see [67], theorem 3.11, page 16) that for  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $(s, y) \mapsto P_{t-s}f(y)$  is infinitely differentiable on  $[0, t] \times \mathbb{R}^d$  and

$$\forall (s, y) \in [0, t] \times \mathbb{R}^d, \quad (\partial_s + L)P_{t-s}f(y) = 0. \quad (2.33)$$

Since  $\partial_s$  and  $L_j^x$  commute, (2.33) and the definition of  $L_j^x$  imply

$$(\partial_s + L^x)L_j^x P_{t-s} = (L^x L_j^x - L_j^x L)P_{t-s} = L_{j+1}^x P_{t-s}. \quad (2.34)$$

For a measurable family  $(A_s)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , we denote by  $\int_{t_1}^{t_2} A_s ds$  the element of  $\mathcal{L}(C_{\text{pol}}^\infty(\mathbb{R}^d))$  which maps  $f$  to  $x \mapsto \int_{t_1}^{t_2} A_s f(x) ds$ . The following lemma states that  $\Phi_{\cdot,t}^{n,j+1}$  is the derivative of  $\Phi_{\cdot,t}^{n,j}$  on the interval  $[0, t]$ .

**Lemma 17.** *Under (A), for all  $j \in \mathbb{N}$ ,  $n \geq 1$  and  $0 \leq s \leq t \leq 1/n$ ,*

$$\Phi_{s,t}^{n,j} = L_j^* P_t + \int_0^s \Phi_{s',t}^{n,j+1} ds'. \quad (2.35)$$

*Proof.* For  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $(s, y) \mapsto L_j^x P_{t-s}f(y)$  is infinitely differentiable on  $[0, t] \times \mathbb{R}^d$  so that we can apply Itô's formula to it and to the semimartingale  $X^{n,x}$  between 0 and  $s$ . Using (2.34) for the second equality, we get

$$\begin{aligned} L_j^x P_{t-s}f(X_s^{n,x}) - L_j^x P_t f(x) - M_s \\ = \int_0^s \left( \frac{\partial}{\partial s} + L^x \right) L_j^x P_{t-s'}f(X_{s'}^{n,x}) ds' = \int_0^s L_{j+1}^x P_{t-s'}f(X_{s'}^{n,x}) ds' \end{aligned}$$

where  $M_s = \sum_{i=1}^d \sum_{j=1}^r \sigma_{i,j}(x) \int_0^s \partial^{e_i} (L_j^x P_{t-s'} f(X_{s'}^{n,x})) dB_{s'}^j$ . Since  $\{L_j^x P_{t-s'} f|_{s'} \in [0, t]\}$  is bounded in  $C_{\text{pol}}^\infty(\mathbb{R}^d)$ , (2.59) implies that  $(M_s, s \in [0, t])$  is a square-integrable martingale and thus has zero mean. Hence, taking expectations and using (2.31) and Fubini's theorem, we have

$$\Phi_{s,t}^{n,j} f(x) - L_j^* P_t f(x) = \int_0^s \mathbb{E} [L_{j+1}^x P_{t-s'} f(X_{s'}^{n,x})] ds' = \int_0^s \Phi_{s',t}^{n,j+1} f(x) ds',$$

which concludes the proof.  $\square$

For  $t \in [0, 1/n]$ , since  $\Delta_t^n = \Phi_{t,t}^{n,0} - \Phi_{0,t}^{n,0}$ , by iterating (2.35) we get

$$\Delta_t^n = \sum_{i=2}^j \frac{t^i}{i!} L_i^* P_t + I_t^{n,j+1}, \quad (2.36)$$

where

$$I_t^{n,j+1} = \int_0^t \int_0^{s_1} \cdots \int_0^{s_j} \Phi_{s_{j+1},t}^{n,j+1} ds_{j+1} \cdots ds_2 ds_1. \quad (2.37)$$

The crucial point here is that, by construction,  $L_1^* = 0$  so that the sum in (2.36) begins with  $i = 2$ .

Injecting this in (2.26), we eventually get for all  $t \in [0, 1]$  and  $n \geq 1$

$$\Delta_t^n = \sum_{i=2}^j \frac{1}{i! n^i} \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k}^n L_i^* P_{t-t_k} + R_t^{n,j} + \sum_{i=2}^j \frac{(t - \lfloor nt \rfloor / n)^i}{i!} P_{\lfloor nt \rfloor / n}^n L_i^* P_{t - \lfloor nt \rfloor / n}, \quad (2.38)$$

where

$$R_t^{n,j} = \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k}^n I_{1/n}^{n,j+1} P_{t-t_{k+1}} + P_{\lfloor nt \rfloor / n}^n I_{t - \lfloor nt \rfloor / n}^{n,j+1}. \quad (2.39)$$

From Proposition 16,  $(I_t^{n,j+1}, n \geq 1, t \in [0, 1/n])$  is a  $O(t^{j+1})$  family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . Recalling the boundedness of  $(P_t, t \in [0, 1])$  and  $(P_t^n, t \in [0, 1], n \geq 1)$ , we get that the family  $R^i = (R_t^{n,i}, t \in [0, 1], n \geq 1)$  is  $O(t/n^i)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . Theorem 14 is thus proved. We are now in good position to prove Theorem 13.

### 2.2.3 Proof of Theorem 13

In the particular case when  $j = 1$ , (2.38) reads  $\Delta_t^n = R_t^{n,1}$  so that we have proved that  $(\Delta_t^n, t \in [0, 1], n \geq 1)$  is  $O(t/n)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , which was the first statement of Theorem 13.

In the particular case when  $j = 2$ , if we set

$$C_t = \frac{1}{2} \int_0^t P_s L_2^* P_{t-s} ds, \quad (2.40)$$

$$A_{1,t}^n = \frac{1}{2n} \left( \frac{1}{n} \sum_{k=0}^{\lfloor nt \rfloor - 1} P_{t_k^n} L_2^* P_{t-t_k^n} - \int_0^t P_s L_2^* P_{t-s} ds \right), \quad (2.41)$$

$$A_{2,t}^n = \frac{1}{2n^2} \sum_{k=0}^{\lfloor nt \rfloor - 1} \left( P_{t_k^n}^n L_2^* P_{t-t_k^n} - P_{t_k^n} L_2^* P_{t-t_k^n} \right), \quad (2.42)$$

$$A_{3,t}^n = R_t^{n,2} + \frac{(t - \lfloor nt \rfloor / n)^2}{2} P_{\lfloor nt \rfloor / n}^n L_2^* P_{t - \lfloor nt \rfloor / n} \quad (2.43)$$

$$R_t^n = A_{1,t}^n + A_{2,t}^n + A_{3,t}^n, \quad (2.44)$$

equation (2.38) reads

$$\Delta_t^n = C_t / n + R_t^n. \quad (2.45)$$

As a composition of bounded families,  $(P_s L_2^* P_{t-s}, 0 \leq s \leq t \leq 1)$  is a bounded family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , so that  $(C_t, t \in [0, 1])$  is  $O(t)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . It remains to prove that  $(R_t^n, t \in [0, 1], n \geq 1)$  is  $O(1/n^2)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . We have already proved that it is true of  $(R_t^{n,2}, t \in [0, 1], n \geq 1)$ . It is obviously also true of  $((t - \lfloor nt \rfloor / n)^2 P_{\lfloor nt \rfloor / n}^n L_2^* P_{t - \lfloor nt \rfloor / n}, t \in [0, 1], n \geq 1)$ , so that  $(A_{3,t}^n, t \in [0, 1], n \geq 1)$  is  $O(1/n^2)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ .

For  $(A_{1,t}^n, t \in [0, 1], n \geq 1)$ , observe that, if we set  $L_3^\# = LL_2^* - L_2^*L \in \mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , as  $\partial_s P_s = LP_s = P_s L$ , we have  $\partial_s P_s L_2^* P_{t-s} = P_s LL_2^* P_{t-s} - P_s L_2^* LP_{t-s} = P_s L_3^\# P_{t-s}$ . Hence the family  $(P_{t_k^n} L_2^* P_{t-t_k^n} - P_s L_2^* P_{t-s}, t \in [0, 1], n \geq 1, k \in \{0, \dots, \lfloor nt \rfloor - 1\}, s \in [t_k^n, t_{k+1}^n])$  satisfies

$$P_{t_k^n} L_2^* P_{t-t_k^n} - P_s L_2^* P_{t-s} = - \int_{t_k^n}^s P_u L_3^\# P_{t-u} du \quad (2.46)$$

and thus is  $O(1/n)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . As a consequence,

$$A_{1,t}^n = \frac{1}{2n} \sum_{k=0}^{\lfloor nt \rfloor - 1} \int_{t_k^n}^{t_{k+1}^n} (P_{t_k^n} L_2^* P_{t-t_k^n} - P_s L_2^* P_{t-s}) ds - \frac{1}{2n} \int_{\lfloor nt \rfloor / n}^t P_s L_2^* P_{t-s} ds \quad (2.47)$$

is  $O(1/n^2)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ .

As for  $(A_{2,t}^n, t \in [0, 1], n \geq 1)$ , note that  $P_{t_k^n}^n L_2^* P_{t-t_k^n} - P_{t_k^n} L_2^* P_{t-t_k^n} = \Delta_{t_k^n}^n L_2^* P_{t-t_k^n}$ . Since  $(\Delta_t^n, t \in [0, 1], n \geq 1)$  is  $O(1/n)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ , so is the family  $(P_{t_k^n}^n L_2^* P_{t-t_k^n} - P_{t_k^n} L_2^* P_{t-t_k^n}, t \in [0, 1], n \geq 1, k \in \{0, \dots, \lfloor nt \rfloor - 1\})$ , as the composition of a bounded family by a  $O(1/n)$  family in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . This completes the proof of Theorem 13.

**Remark 18.** It is noteworthy that the family  $(R_t^n, t \in [0, 1], n \geq 1)$  defined by

$$R_t^n = R_t^n + \frac{1}{2n} \int_{\lfloor nt \rfloor/n}^t P_s L_2^* P_{t-s} ds - \frac{(t - \lfloor nt \rfloor/n)^2}{2} P_{\lfloor nt \rfloor/n}^n L_2^* P_{t-\lfloor nt \rfloor/n}$$

is  $O(t/n^2)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ . In particular,  $(R_{\lfloor nt \rfloor/n}^n, t \in [0, 1], n \geq 1)$  is  $O(t/n^2)$  in  $\mathcal{L}_b(C_{\text{pol}}^\infty(\mathbb{R}^d))$ .

## 2.3 Second step: expansion for the density of $X_t^{n,x}$

This section is devoted to the proofs of Theorems 10 and 12.

### 2.3.1 Space convolutions

We begin by proving Proposition 9 which is the key argument. Recall the definitions of Section 2.1.6. Let  $\mathcal{B}_1$  and  $\mathcal{B}_2$  be two bounded subsets of  $\mathcal{G}(\mathbb{R}^d)$ ,  $g \in C_b^\infty(\mathbb{R}^d)$  and  $\gamma \in \mathbb{N}^d$ . We want to prove that

- (i)  $\{\pi_1 *_{g,\gamma} \pi_2 \mid \pi_1 \in \mathcal{B}_1, \pi_2 \in \mathcal{B}_2\}$  is a bounded subset of  $\mathcal{H}_{|\gamma|}(\mathbb{R}^d)$ ,
- (ii)  $\{\pi_1 *_{g,0} \pi_2 \mid \pi_1 \in \mathcal{B}_1, \pi_2 \in \mathcal{B}_2\}$  is a bounded subset of  $\mathcal{H}(\mathbb{R}^d)$ .

The functions  $\pi_1 *_{g,\gamma} \pi_2$  depend on  $(s, t, x, y)$ . We shall proceed differently depending on  $s$  is small or large with respect to  $t$ . The main trick is to integrate by parts in the latter case, so that the derivatives should always rest on the regularizing part of the integral. This is analogous to V. BALLY and D. TALAY's use of Malliavin calculus integration by parts formula [3]. This is the reason why we partition the unit triangle  $\mathcal{T}_1$  into  $\mathcal{T}_1^- = \{(s, t) \in \mathcal{T}_1 \mid 0 < s \leq t/2\}$  and  $\mathcal{T}_1^+ = \{(s, t) \in \mathcal{T}_1 \mid t/2 < s < t\}$ , and, for  $\epsilon = \pm$ , we define  $(\pi_1 *_{g,\gamma} \pi_2)_\epsilon(s, t, x, y) = \mathbf{1}_{\mathcal{T}_1^\epsilon}(s, t) (\pi_1 *_{g,\gamma} \pi_2)(s, t, x, y)$ . We then have  $\pi_1 *_{g,\gamma} \pi_2 = (\pi_1 *_{g,\gamma} \pi_2)_- + (\pi_1 *_{g,\gamma} \pi_2)_+$ .

Before proving Proposition 9 and for the sake of clarity, let us state apart the following technical lemma, whose proof is a straightforward application of Lebesgue's dominated convergence theorem:

**Lemma 19.** *Let  $l \in \mathbb{Z}$ ,  $(\chi_i, i \in I)$  be a family of measurable functions mapping  $\mathcal{T}_1 \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$  into  $\mathbb{R}$  such that*

- for all  $i \in I$ ,  $(s, t) \in \mathcal{T}_1$  and  $\zeta \in \mathbb{R}^d$ ,  $\chi_i(s, t, \cdot, \cdot, \zeta)$  is infinitely differentiable,
- for all  $\alpha, \beta \in \mathbb{N}^d$ , there exists two constants  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $i \in I$ ,  $(s, t) \in \mathcal{T}_1$  and  $x, y, \zeta \in \mathbb{R}^d$ ,

$$\left| \partial_x^\alpha \partial_y^\beta \chi_i(s, t, x, y, \zeta) \right| \leq c_1 t^{-(|\alpha| + |\beta| + d + l)/2} \exp(-c_2 \|x - y\|^2 / t - c_2 \|\zeta\|^2), \quad (2.48)$$

and let us define  $\mathcal{I}(\chi_i)(s, t, x, y) = \int_{\mathbb{R}^d} \chi_i(s, t, x, y, \zeta) d\zeta$ . Then  $\{\mathcal{I}(\chi_i) \mid i \in I\}$  is a bounded subset of  $\mathcal{H}_l(\mathbb{R}^d)$ .

*Proof of Proposition 9-(i).* It is enough to show that both  $\mathcal{B}_\epsilon \equiv \{(\pi_1 *_{g,\gamma} \pi_2)_\epsilon | \pi_1 \in \mathcal{B}_1, \pi_2 \in \mathcal{B}_2\}$  are bounded.

**Step 1.** Let us first treat  $\mathcal{B}_-$ , i.e. the case when  $s$  is small. After the change of variables  $z = x + \zeta \sqrt{s}$ , we get  $(\pi_1 *_{g,\gamma} \pi_2)_- = \mathcal{I}(\chi_{\pi_1, \pi_2}^-)$  with

$$\chi_{\pi_1, \pi_2}^-(s, t, x, y, \zeta) = \mathbf{1}_{\mathcal{T}_1^-}(s, t) s^{d/2} g(x + \zeta \sqrt{s}) \pi_1(s, x, x + \zeta \sqrt{s}) \partial_2^\gamma \pi_2(t - s, x + \zeta \sqrt{s}, y).$$

It is enough to check that the family  $(\chi_{\pi_1, \pi_2}^-, (\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2)$  satisfies the assumptions of Lemma 19 with  $l = |\gamma|$ . The first point is obvious. In order to check the second one, let us fix  $\alpha, \beta \in \mathbb{N}^d$ . According to Leibniz's formula,  $\partial_x^\alpha \partial_y^\beta \chi_{\pi_1, \pi_2}^-(s, t, x, y, \zeta)$  can be written as a weighted sum of terms of the form

$$\begin{aligned} \chi_{\pi_1, \pi_2}^-, \alpha_1, \alpha_2, \alpha_3(s, t, x, y, \zeta) &= \mathbf{1}_{\mathcal{T}_1^-}(s, t) s^{d/2} \partial^{\alpha_1} g(x + \zeta \sqrt{s}) \\ &\quad \partial_x^{\alpha_2} (\pi_1(s, x, x + \zeta \sqrt{s})) \partial_2^{\gamma + \alpha_3} \partial_3^\beta \pi_2(t - s, x + \zeta \sqrt{s}, y), \end{aligned}$$

with  $|\alpha_1| + |\alpha_2| + |\alpha_3| = |\alpha|$ , so that in order to check (2.48) it is enough to show that for each such  $(\alpha_1, \alpha_2, \alpha_3)$  one can find  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $(\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ ,  $(s, t) \in \mathcal{T}_1$  and  $x, y, \zeta \in \mathbb{R}^d$ ,  $|\chi_{\pi_1, \pi_2}^-, \alpha_1, \alpha_2, \alpha_3(s, t, x, y, \zeta)|$  is less than the r.h.s. of (2.48), with  $l = |\gamma|$ . Now,  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are bounded subsets of  $\mathcal{G}(\mathbb{R}^d)$  so that from (2.15)-(2.16) one can find  $c_3, c_5 \geq 0$  and  $c_4 > 0$  such that for all  $(\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ ,  $(s, t) \in \mathcal{T}_1$  and  $x, y, \zeta \in \mathbb{R}^d$ ,

$$\left| \partial_x^{\alpha_2} (\pi_1(s, x, x + \zeta \sqrt{s})) \right| \leq c_3 s^{-d/2} \exp(-c_4 \|\zeta\|^2)$$

and

$$\begin{aligned} \mathbf{1}_{\mathcal{T}_1^-}(s, t) \left| \partial_2^{\gamma + \alpha_3} \partial_3^\beta \pi_2(t - s, x + \zeta \sqrt{s}, y) \right| \\ \leq \mathbf{1}_{\mathcal{T}_1^-}(s, t) c_3 (t - s)^{-(|\alpha_3| + |\beta| + |\gamma| + d)/2} \exp\left(-c_4 \|x - y + \zeta \sqrt{s}\|^2 / (t - s)\right) \\ \leq \mathbf{1}_{\mathcal{T}_1^-}(s, t) c_5 t^{-(|\alpha| + |\beta| + |\gamma| + d)/2} \exp\left(-c_4 \|x - y + \zeta \sqrt{s}\|^2 / t\right) \end{aligned}$$

where, for the last inequality, we have used the fact that when  $(s, t) \in \mathcal{T}_1^-$ ,  $t/2 \leq t - s \leq t \leq 1$ . Now, using the fact that  $\|x - z\|^2 \geq \|x\|^2 / 2 - \|z\|^2$  for all  $x, z \in \mathbb{R}^d$ , we see that for all  $(s, t) \in \mathcal{T}_1^-$ ,  $\|\zeta\|^2 + \|x - y + \zeta \sqrt{s}\|^2 / t \geq (\|x - y\|^2 / t + \|\zeta\|^2) / 2$ . Since  $g \in C_b^\infty(\mathbb{R}^d)$ , we can eventually find  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $(\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ ,  $(s, t) \in \mathcal{T}_1$  and  $x, y, \zeta \in \mathbb{R}^d$ ,

$$\left| \chi_{\pi_1, \pi_2}^-, \alpha_1, \alpha_2, \alpha_3(s, t, x, y, \zeta) \right| \leq c_1 t^{-(|\alpha| + |\beta| + d + |\gamma|)/2} \exp\left(-c_2 \|x - y\|^2 / t - c_2 \|\zeta\|^2\right),$$

which completes Step 1.



**Step 2.** Let us now treat  $\mathcal{B}_+$ , i.e. the case when  $s$  is large. After  $|\gamma|$  integrations by parts, we have

$$(\pi_1 *_{g,\gamma} \pi_2)_+(s, t, x, y) = \mathbf{1}_{\mathcal{T}_1^+}(s, t) \int_{\mathbb{R}^d} \partial_z^\gamma (g(z) \pi_1(s, x, z)) \pi_2(t - s, z, y) dz.$$

Using Leibniz's formula and making the change of variables  $z = y - \zeta \sqrt{t - s}$ , we get that  $(\pi_1 *_{g,\gamma} \pi_2)_+$  is a weighted sum of terms of the form  $\mathcal{I}(\chi_{\pi_1, \pi_2}^{+, \gamma_1, \gamma_2})$  with

$$\begin{aligned} \chi_{\pi_1, \pi_2}^{+, \gamma_1, \gamma_2}(s, t, x, y, \zeta) &= \mathbf{1}_{\mathcal{T}_1^+}(s, t) (t - s)^{d/2} \partial^{\gamma_1} g(y - \zeta \sqrt{t - s}) \\ &\quad \partial_3^{\gamma_2} \pi_1(s, x, y - \zeta \sqrt{t - s}) \pi_2(t - s, y - \zeta \sqrt{t - s}, y) \end{aligned}$$

and  $|\gamma_1| + |\gamma_2| = |\gamma|$ , so that we are now in the position to apply the same arguments as in Step 1 and get that the family  $(\chi_{\pi_1, \pi_2}^{+, \gamma_1, \gamma_2}, (\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2)$  satisfies the assumptions of Lemma 19 with  $l = |\gamma|$ , which completes the proof.  $\square$

*Proof of Proposition 9-(ii).* From (i), we know that  $\{\pi_1 *_{g,0} \pi_2 \mid \pi_1 \in \mathcal{B}_1, \pi_2 \in \mathcal{B}_2\}$  is a bounded subset of  $\mathcal{H}_0(\mathbb{R}^d)$ . It remains to prove that (2.19) holds for  $\rho = \pi_1 *_{g,0} \pi_2$  with constants  $c_1$  and  $c_2$  which do not depend on  $(\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ . As in the proof of Proposition 9-(i), we treat  $(\pi_1 *_{g,0} \pi_2)_-$  and  $(\pi_1 *_{g,0} \pi_2)_+$  separately but analogously. That is, after integrating by parts, the term  $(\pi_1 *_{g,0} \pi_2)_+$  can be treated in the same way as  $(\pi_1 *_{g,0} \pi_2)_-$ . Thus we shall only deal with the latter term. We have  $(\pi_1 *_{g,0} \pi_2)_- = \mathcal{I}(\chi_{\pi_1, \pi_2}^-)$  with

$$\chi_{\pi_1, \pi_2}^-(s, t, x, y, \zeta) = \mathbf{1}_{\mathcal{T}_1^-}(s, t) s^{d/2} g(x + \zeta \sqrt{s}) \pi_1(s, x, x + \zeta \sqrt{s}) \pi_2(t - s, x + \zeta \sqrt{s}, y).$$

Then we write  $\partial_x^\alpha (\chi_{\pi_1, \pi_2}^-(s, t, x, x + y\sqrt{t}, \zeta))$  as a weighted sum of terms of the form

$$\begin{aligned} \tilde{\chi}_{\pi_1, \pi_2}^-, \alpha_1, \alpha_2, \alpha_3(s, t, x, y, \zeta) &= \mathbf{1}_{\mathcal{T}_1^-}(s, t) s^{d/2} \partial^{\alpha_1} g(x + \zeta \sqrt{s}) \\ &\quad \partial_x^{\alpha_2} (\pi_1(s, x, x + \zeta \sqrt{s})) \partial_x^{\alpha_3} (\pi_2(t - s, x + \zeta \sqrt{s}, x + y\sqrt{t})), \end{aligned}$$

with  $|\alpha_1| + |\alpha_2| + |\alpha_3| = |\alpha|$ . Then we use (2.16) twice and the same arguments as in the preceding proof to get  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $(\pi_1, \pi_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ ,  $(s, t) \in \mathcal{T}_1$  and  $x, y, \zeta \in \mathbb{R}^d$ ,  $|\tilde{\chi}_{\pi_1, \pi_2}^-, \alpha_1, \alpha_2, \alpha_3(s, t, x, y, \zeta)| \leq c_1 t^{-d/2} \exp(-c_2 \|y\|^2 - c_2 \|\zeta\|^2)$ , and an obvious adaptation of Lemma 19 completes the proof.  $\square$

### 2.3.2 Proof of Theorems 10 and 12

In this section, we assume (B) and (C). We first want to prove Theorem 10. We recall that statement (i) is already known, see [35], theorem 7, page 260. The next lemma is statement (ii).

**Lemma 20.** *Under (B) and (C), for all  $t \in (0, 1]$ ,  $n \geq 1$  and  $x \in \mathbb{R}^d$ ,  $X_t^{n,x}$  has a density  $p_n(t, x, \cdot)$  and  $(p_n, n \geq 1)$  is a bounded sequence in  $\mathcal{G}(\mathbb{R}^d)$ .*

*Proof.* It is known that for all  $n \geq 1$ ,  $k \in \{1, \dots, n\}$  and  $x \in \mathbb{R}^d$ ,  $X_{t_k^n}^{n,x}$  has a density  $p_{n,k}(x, \cdot)$  such that  $p_{n,k}$  is infinitely differentiable and satisfies (2.15)-(2.16) with  $t = t_k^n$  and two constants  $c_1$  and  $c_2$  which do not depend on  $n$  and  $k$  (see the proof of theorem 1.1, page 278, in [61]). Since  $\lfloor nt \rfloor / n \geq t/2$  for all  $t \geq 1/n$ , this shows that the sequence  $(\tilde{p}_n, n \geq 1)$  defined by  $\tilde{p}_n(t, x, y) = \mathbf{1}_{\{nt \geq 1\}} p_{n, \lfloor nt \rfloor}(x, y)$  is bounded in  $\mathcal{G}(\mathbb{R}^d)$ . If we denote by  $\Gamma(t, x, \cdot)$  the density of  $x + b(x)t + \sigma(x)B_t$  ( $t \in (0, 1]$ ), we observe that when  $k \in \{1, \dots, n-1\}$  and  $t \in (t_k^n, t_{k+1}^n)$ ,  $X_t^{n,x}$  has the density  $p_n(t, x, \cdot) = \int_{\mathbb{R}^d} p_{n,k}(x, z) \Gamma(t - t_k^n, z, \cdot) dz = (\tilde{p}_n *_{1,0} \Gamma)(t_k^n, t, x, \cdot)$ . Hence, for all  $t \in (0, 1]$ ,  $n \geq 1$  and  $x \in \mathbb{R}^d$ ,  $X_t^{n,x}$  has the density

$$p_n(t, x, \cdot) = \begin{cases} p_{n,k}(x, \cdot) & \text{if } t = t_k^n, k \in \{1, \dots, n\}, \\ \Gamma(t, x, \cdot) & \text{if } t \in (0, t_1^n), \\ (\tilde{p}_n *_{1,0} \Gamma)(t_k^n, t, x, \cdot) & \text{if } t \in (t_k^n, t_{k+1}^n), k \in \{1, \dots, n-1\}. \end{cases}$$

Observing that  $\Gamma \in \mathcal{G}(\mathbb{R}^d)$  and applying Proposition 9-(ii), we get that  $(p_n, n \geq 1)$  is a bounded sequence in  $\mathcal{G}(\mathbb{R}^d)$ .  $\square$

We shall now prove statement (iii) of Theorem 10. Recall (2.45). We want to make explicit  $C_t$  and  $R_t^n$  as integral operators on  $\mathbb{R}^d$ . To this end, note that, applying recursively Lebesgue's dominated convergence theorem, we have that for all  $t \in (0, 1]$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $x \in \mathbb{R}^d$  and  $\alpha \in \mathbb{N}^d$ ,

$$\partial^\alpha P_t f(x) = \int_{\mathbb{R}^d} f(y) \partial_2^\alpha p(t, x, y) dy. \quad (2.49)$$

The next lemma explicits  $C_t$  as an integral operator. The function  $\pi$  which appears there should be thought of as the kernel of  $C$ .

**Lemma 21.** *Under (B) and (C), there exists  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$ , given by (2.21), such that for all  $t \in (0, 1]$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,*

$$C_t f(x) = \int_{\mathbb{R}^d} f(y) \pi(t, x, y) dy.$$

*Proof.* Using (2.40) for the first equality, (2.29) for the third one and (2.49) for the

fourth one, we have

$$\begin{aligned}
2C_t f(x) &= \int_0^t P_s L_2^* P_{t-s} f(x) ds \\
&= \int_0^t \int_{\mathbb{R}^d} p(s, x, z) L_2^* P_{t-s} f(z) dz ds \\
&= \sum_{1 \leq |\alpha| \leq 3} \int_0^t \int_{\mathbb{R}^d} g_{2,\alpha}^*(z) p(s, x, z) \partial^\alpha P_{t-s} f(z) dz ds \\
&= \sum_{1 \leq |\alpha| \leq 3} \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y) g_{2,\alpha}^*(z) p(s, x, z) \partial_2^\alpha p(t-s, z, y) dy dz ds.
\end{aligned}$$

Using Fubini's theorem, we see that to complete the proof it is enough to show that the function  $\pi$  defined by

$$\pi(t, x, y) = \frac{1}{2} \sum_{1 \leq |\alpha| \leq 3} \int_0^t (p *_{g_{2,\alpha}^*, \alpha} p)(s, t, x, y) ds \quad (2.50)$$

belongs to  $\mathcal{G}_1(\mathbb{R}^d)$ . Now,  $p \in \mathcal{G}(\mathbb{R}^d)$  and, from Remark 15,  $g_{2,\alpha}^* \in C_b^\infty(\mathbb{R}^d)$  so that we can apply Proposition 9-(i):  $p *_{g_{2,\alpha}^*, \alpha} p \in \mathcal{H}_{|\alpha|}(\mathbb{R}^d)$ . In particular,  $\int_0^t (p *_{g_{2,\alpha}^*, \alpha} p)(s, \cdot, \cdot, \cdot) ds \in \mathcal{G}_{|\alpha|-2}(\mathbb{R}^d)$ . Since  $|\alpha| \leq 3$  and by monotonicity of  $(\mathcal{G}_l(\mathbb{R}^d), l \in \mathbb{Z})$ , we finally get that  $\pi \in \mathcal{G}_1(\mathbb{R}^d)$ . To complete the proof, note that (2.50) can be rewritten as (2.21).  $\square$

We have a similar representation for  $A_{1,t}^n$ , recall (2.41). We say that a sequence  $(\pi^n, n \geq 1)$  is  $O(1/n^j)$  in  $\mathcal{G}_l(\mathbb{R}^d)$  if  $(n^j \pi^n, n \geq 1)$  is bounded in  $\mathcal{G}_l(\mathbb{R}^d)$ .

**Lemma 22.** *Under (B) and (C), there exists a  $O(1/n^2)$  sequence  $(\pi_1^n, n \geq 1)$  in  $\mathcal{G}_3(\mathbb{R}^d)$  such that for all  $t \in (0, 1]$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,*

$$A_{1,t}^n f(x) = \int_{\mathbb{R}^d} f(y) \pi_1^n(t, x, y) dy.$$

*Proof.* Recall (2.46). From Remark 15, there is a family  $(g_{3,\alpha}^\#, 1 \leq |\alpha| \leq 4)$  in  $C_b^\infty(\mathbb{R}^d)$  such that  $L_3^\# = \sum_{1 \leq |\alpha| \leq 4} g_{3,\alpha}^\# \partial^\alpha$ , so that, using (2.49), we have

$$\begin{aligned}
&(P_{t_k^n}^n L_2^* P_{t-t_k^n} - P_s L_2^* P_{t-s}) f(x) \\
&= - \sum_{1 \leq |\alpha| \leq 4} \int_{t_k^n}^s \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y) g_{3,\alpha}^\#(z) p(u, x, z) \partial_2^\alpha p(t-u, z, y) dy dz du.
\end{aligned}$$

Using (2.47), we get  $A_{1,t}^n f(x) = \int_{\mathbb{R}^d} f(y) \pi_1^n(t, x, y) dy$  with  $\pi_1^n = \pi_{1,1}^n + \pi_{1,2}^n$  and

$$\begin{aligned}\pi_{1,1}^n(t, x, y) &= -\frac{1}{2n} \sum_{1 \leq |\alpha| \leq 4} \sum_{k=0}^{\lfloor nt \rfloor - 1} \int_{t_k^n}^{t_{k+1}^n} \int_{t_k^n}^s (p *_{g_{3,\alpha}^\#} p)(u, t, x, y) duds, \\ \pi_{1,2}^n(t, x, y) &= -\frac{1}{2n} \sum_{1 \leq |\alpha| \leq 3} \int_{\lfloor nt \rfloor / n}^t (p *_{g_{2,\alpha}^*} p)(s, t, x, y) ds.\end{aligned}$$

Now Proposition 9-(i) states that  $p *_{g_{3,\alpha}^\#} p$  and  $p *_{g_{2,\alpha}^*} p$  belong to  $\mathcal{H}_{|\alpha|}(\mathbb{R}^d)$ . Hence  $(\int_{t_k^n}^{t_{k+1}^n} \int_{t_k^n}^s (p *_{g_{3,\alpha}^\#} p)(u, \cdot, \cdot, \cdot) duds, n \geq 1, k \in \{0, \dots, n-1\})$  is  $O(1/n^2)$  in  $\mathcal{G}_{|\alpha|}(\mathbb{R}^d)$  and  $(\int_{\lfloor n \cdot \rfloor / n}^t (p *_{g_{2,\alpha}^*} p)(s, \cdot, \cdot, \cdot) ds, n \geq 1)$  is  $O(1/n)$  in  $\mathcal{G}_{|\alpha|}(\mathbb{R}^d)$ . As a consequence,  $(\pi_{1,1}^n, n \geq 1)$  is  $O(1/n^2)$  in  $\mathcal{G}_2(\mathbb{R}^d)$  and  $(\pi_{1,2}^n, n \geq 1)$  is  $O(1/n^2)$  in  $\mathcal{G}_3(\mathbb{R}^d)$ . Eventually,  $(\pi_1^n, n \geq 1)$  is  $O(1/n^2)$  in  $\mathcal{G}_3(\mathbb{R}^d)$ .  $\square$

We shall now prove an analogous lemma for  $A_{2,t}^n$ .

**Lemma 23.** *Under (B) and (C), there exists a  $O(1/n^2)$  sequence  $(\pi_2^n, n \geq 1)$  in  $\mathcal{G}_3(\mathbb{R}^d)$  such that for all  $t \in (0, 1]$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,*

$$A_{2,t}^n f(x) = \int_{\mathbb{R}^d} f(y) \pi_2^n(t, x, y) dy.$$

*Proof.* Since  $P_{t_k^n}^n L_2^* P_{t-t_k^n} = P_{t_k^n} L_2^* P_{t-t_k^n}$  when  $k=0$ , (2.42) reads

$$\begin{aligned}2n^2 A_{2,t}^n f(x) &= \sum_{k=1}^{\lfloor nt \rfloor - 1} (P_{t_k^n}^n - P_{t_k^n}) L_2^* P_{t-t_k^n} f(x) \\ &= \sum_{k=1}^{\lfloor nt \rfloor - 1} \int_{\mathbb{R}^d} (p_n - p)(t_k^n, x, z) L_2^* P_{t-t_k^n} f(z) dz \\ &= \sum_{1 \leq |\alpha| \leq 3} \sum_{k=1}^{\lfloor nt \rfloor - 1} \int_{\mathbb{R}^d} (p_n - p)(t_k^n, x, z) g_{2,\alpha}^*(z) \partial^\alpha P_{t-t_k^n} f(z) dz \\ &= \sum_{1 \leq |\alpha| \leq 3} \sum_{k=1}^{\lfloor nt \rfloor - 1} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (p_n - p)(t_k^n, x, z) g_{2,\alpha}^*(z) f(y) \partial_2^\alpha p(t - t_k^n, z, y) dy dz\end{aligned}$$

where we have used (2.29) for the third equality and (2.49) for the fourth one. From Remark 15,  $g_{2,\alpha}^* \in C_b^\infty(\mathbb{R}^d)$  so that to complete the proof it is enough to show that

whenever  $g \in C_b^\infty(\mathbb{R}^d)$  and  $\alpha \in \mathbb{N}^d$ , the sequence  $(\pi^n, n \geq 1)$  defined by

$$\begin{aligned} \pi^n(t, x, y) &= \sum_{k=1}^{\lfloor nt \rfloor - 1} \int_{\mathbb{R}^d} (p_n - p)(t_k^n, x, z) g(z) \partial_2^\alpha p(t - t_k^n, z, y) dz \\ &= \sum_{k=1}^{\lfloor nt \rfloor - 1} ((p_n - p) *_{g, \alpha} p)(t_k^n, t, x, y) \end{aligned}$$

is bounded in  $\mathcal{G}_{|\alpha|}(\mathbb{R}^d)$ . And to do so, it is enough to show that the sequence  $(\rho_{t_k^n}^n, n \geq 2, k \in \{1, \dots, n-1\})$  defined by

$$\rho_{t_k^n}^n(t, x, y) = \mathbf{1}_{\mathcal{T}_1}(t_k^n, t) ((p_n - p) *_{g, \alpha} p)(t_k^n, t, x, y)$$

is  $O(1/n)$  in  $\mathcal{G}_{|\alpha|+2}(\mathbb{R}^d)$ . Let us write  $\rho_{t_k^n}^{n,-}(t, x, y) = \mathbf{1}_{\mathcal{T}_1^-}(t_k^n, t) \rho_{t_k^n}^n(t, x, y)$  and  $\rho_{t_k^n}^{n,+}(t, x, y) = \mathbf{1}_{\mathcal{T}_1^+}(t_k^n, t) \rho_{t_k^n}^n(t, x, y)$  so that  $\rho_{t_k^n}^n = \rho_{t_k^n}^{n,-} + \rho_{t_k^n}^{n,+}$ .

Let us first prove that  $(\rho_{t_k^n}^{n,-}, n \geq 2, k \in \{1, \dots, n-1\})$  is  $O(1/n)$  in  $\mathcal{G}_{|\alpha|+2}(\mathbb{R}^d)$ . The sequence  $(\pi_{t_k^n}, n \geq 2, k \in \{1, \dots, n-1\})$  defined by  $\pi_{t_k^n}(t, x, y) = \mathbf{1}_{\mathcal{T}_1^-}(t_k^n, t) g(x) \partial_2^\alpha p(t - t_k^n, x, y)$  is bounded in  $\mathcal{G}_{|\alpha|}(\mathbb{R}^d)$ , since  $t - t_k^n \geq t/2$  when  $(t_k^n, t) \in \mathcal{T}_1^-$ . Now note that  $\rho_{t_k^n}^{n,-} = P_{t_k^n}^n \pi_{t_k^n} - P_{t_k^n}^n \pi_{t_k^n} = \Delta_{t_k^n}^n \pi_{t_k^n}$  (see (2.51) in the appendix for the definition of  $P_s^n \pi$ ,  $P_s \pi$  and  $\Delta_s^n \pi$  when  $\pi \in \mathcal{G}_l(\mathbb{R}^d)$ ). Thus, from (2.38)-(2.39) and (2.37) applied with  $j = 1$ ,

$$\rho_{t_k^n}^{n,-} = \sum_{m=0}^{k-1} \int_0^{1/n} \int_0^{s_1} P_{t_k^n}^n \Phi_{s_2, 1/n}^{n,2} P_{t_k^n - t_{m+1}^n} \pi_{t_k^n} ds_2 ds_1.$$

Proposition 28 in the appendix states that the family  $(P_{t_k^n}^n \Phi_{s, 1/n}^{n,2} P_{t_k^n - t_{m+1}^n} \pi_{t_k^n}, n \geq 2, k \in \{1, \dots, n-1\}, m \in \{1, \dots, k-1\}, s \in [0, 1/n])$  is bounded in  $\mathcal{G}_{|\alpha|+4}(\mathbb{R}^d)$ . Since  $k \leq \lfloor nt \rfloor$  when  $(t_k^n, t) \in \mathcal{T}_1$ , this implies that  $(\rho_{t_k^n}^{n,-}, n \geq 2, k \in \{1, \dots, n-1\})$  is  $O(1/n)$  in  $\mathcal{G}_{|\alpha|+2}(\mathbb{R}^d)$ .

Let us now prove the same for  $\rho^{n,+}$ . After  $|\alpha|$  integrations by parts and after setting  $z = y - \zeta \sqrt{t - s}$ , we get that  $((p_n - p) *_{g, \alpha} p)_+$  is a weighted sum of terms of the form  $\mathcal{I}(\chi_{\alpha_1, \alpha_2}^{n,+})$  - see Lemma 19 - with

$$\begin{aligned} \chi_{\alpha_1, \alpha_2}^{n,+}(s, t, x, y, \zeta) &= \mathbf{1}_{\mathcal{T}_1^+}(s, t) (t - s)^{d/2} \partial^{\alpha_1} g(y - \zeta \sqrt{t - s}) \\ &\quad \partial_3^{\alpha_2} (p_n - p)(s, x, y - \zeta \sqrt{t - s}) p(t - s, y - \zeta \sqrt{t - s}, y) \end{aligned}$$

and  $|\alpha_1| + |\alpha_2| = |\alpha|$ . Now, from Corollary 26 in the appendix,  $(p_n - p, n \geq 1)$  is  $O(1/n)$  in  $\mathcal{G}_2(\mathbb{R}^d)$  so that, using the same arguments as in Step 2 of the proof of Proposition 9-(i), we get that  $((p_n - p) *_{g, \alpha} p)_+$  is  $O(1/n)$  in  $\mathcal{H}_{|\alpha|+2}(\mathbb{R}^d)$ . Since  $\rho_{t_k^n}^{n,+}(t, x, y) = \mathbf{1}_{\mathcal{T}_1^+}(t_k^n, t) ((p_n - p) *_{g, \alpha} p)_+(t_k^n, t, x, y)$ , we conclude that  $(\rho_{t_k^n}^{n,+}, n \geq 2, k \in \{1, \dots, n-1\})$  is  $O(1/n)$  in  $\mathcal{G}_{|\alpha|+2}(\mathbb{R}^d)$ .  $\square$

Lastly, starting from (2.43), Lemmas 25 and 27 with  $j = 2$  imply

**Lemma 24.** *Under (B) and (C), there exists a  $O(1/n^2)$  sequence  $(\pi_3^n, n \geq 1)$  in  $\mathcal{G}_4(\mathbb{R}^d)$  such that for all  $t \in (0, 1]$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,*

$$A_{3,t}^n f(x) = \int_{\mathbb{R}^d} f(y) \pi_3^n(t, x, y) dy.$$

Statement (iii) of Theorem 10 is now proved: it follows from (2.45), (2.44) and Lemmas 21, 22, 23 and 24.

We now also have all the tools to prove Theorem 12. Indeed, note that (2.38) combined with Lemmas 25 and 27 imply that we have an expansion of arbitrary order  $j$  for  $p_n - p$  :

$$(p_n - p)(t, \cdot, \cdot) = \sum_{i=2}^j \frac{1}{i! n^i} \sum_{k=0}^{\lfloor nt \rfloor - 1} \psi_{t_k}^{n,i}(t, \cdot, \cdot) + r^{n,j}(t, \cdot, \cdot) + \sum_{i=2}^j \frac{(t - \lfloor nt \rfloor / n)^i}{i!} \psi_{\lfloor nt \rfloor / n}^{n,i}(t, \cdot, \cdot).$$

Since  $(r^{n,j}, n \geq 1)$  is  $O(1/n^j)$  in  $\mathcal{G}_{2j}(\mathbb{R}^d)$  and  $(\psi_{t_k}^{n,j}, n \geq 1, k \in \{0, \dots, n\})$  is bounded in  $\mathcal{G}_{2j}(\mathbb{R}^d)$ , this gives (2.23) with  $(\pi_{n,i}, n \geq 1)$  bounded in  $\mathcal{G}_{2i-2}(\mathbb{R}^d)$  and  $(\pi'_{n,i}, n \geq 1)$  and  $(\pi''_{n,i}, n \geq 1)$  bounded in  $\mathcal{G}_{2i}(\mathbb{R}^d)$ .

## 2.4 Appendix

### 2.4.1 Kernel of $R^{n,j}$

Here we make explicit the kernel of the remainder  $R_t^{n,j}$ , recall (2.39):

**Lemma 25.** *Under (B) and (C), for each  $j \in \mathbb{N}^*$ , there exists a  $O(1/n^j)$  sequence  $(r^{n,j}, n \geq 1)$  in  $\mathcal{G}_{2j}(\mathbb{R}^d)$  such that for all  $t \in (0, 1]$ ,  $n \geq 1$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,*

$$R_t^{n,j} f(x) = \int_{\mathbb{R}^d} f(y) r^{n,j}(t, x, y) dy.$$

*Proof.* From (2.39) and (2.37),  $R_t^{n,j} = R_{1,t}^{n,j} + R_{2,t}^{n,j}$  where

$$\begin{aligned} R_{1,t}^{n,j} &= \sum_{k=0}^{\lfloor nt \rfloor - 1} \int_0^{1/n} \int_0^{s_1} \cdots \int_0^{s_j} P_{t_k}^n \Phi_{s_{j+1}, 1/n}^{n,j+1} P_{t-t_{k+1}^n} ds_{j+1} \cdots ds_2 ds_1, \\ R_{2,t}^{n,j} &= \int_0^{t - \lfloor nt \rfloor / n} \int_0^{s_1} \cdots \int_0^{s_j} P_{\lfloor nt \rfloor / n}^n \Phi_{s_{j+1}, t - \lfloor nt \rfloor / n}^{n,j+1} ds_{j+1} \cdots ds_2 ds_1. \end{aligned}$$

Let us first deal with  $R_{1,t}^{n,j}$ . Using the fact that  $k \geq 1$  for the first equality, (2.32) for the second one, the fact that  $P_{1/n-s} P_{t-t_{k+1}^n} = P_{t-t_k^n - s}$  for the third one, and (2.49)

and Fubini's theorem for the last one, we have for all  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ ,  $x \in \mathbb{R}^d$ ,  $t \in [0, 1]$ ,  $n \geq 1$ ,  $k \in \{1, \dots, \lfloor nt \rfloor - 1\}$  and  $s \in (0, 1/n)$ ,

$$\begin{aligned}
& P_{t_k^n}^n \Phi_{s,1/n}^{n,j} P_{t-t_{k+1}^n} f(x) \\
&= \int_{\mathbb{R}^d} p_n(t_k^n, x, z_1) \Phi_{s,1/n}^{n,j} P_{t-t_{k+1}^n} f(z_1) dz_1 \\
&= \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} \int_{\mathbb{R}^d} p_n(t_k^n, x, z_1) g_{j,\alpha,l}(z_1) P_s^n(h_{j,\alpha,l} \partial^\alpha P_{1/n-s}) P_{t-t_{k+1}^n} f(z_1) dz_1 \\
&= \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p_n(t_k^n, x, z_1) g_{j,\alpha,l}(z_1) p_n(s, z_1, z_2) h_{j,\alpha,l}(z_2) \partial^\alpha P_{t-t_k^n-s} f(z_2) dz_2 dz_1 \\
&= \int_{\mathbb{R}^d} f(y) \varphi_{t_k^n}^{n,j}(s, t, x, y) dy
\end{aligned}$$

where  $\varphi_{t_k^n}^{n,j} = \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} \varphi_{t_k^n, \alpha, l}^{n,j}$  with

$$\begin{aligned}
\varphi_{t_k^n, \alpha, l}^{n,j}(s, t, x, y) &= \mathbf{1}_{(0, \frac{1}{n})}(s) \mathbf{1}_{[t_{k+1}^n, 1]}(t) \\
&\quad \int_{\mathbb{R}^d} (p_n *_{g_{j,\alpha,l,0}} p_n)(t_k^n, t_k^n + s, x, z_2) h_{j,\alpha,l}(z_2) \partial_2^\alpha p(t - t_k^n - s, z_2, y) dz_2.
\end{aligned}$$

Now, setting  $q_{t_k^n, \alpha, l}^{n,j}(u, x, z) = \mathbf{1}_{(t_k^n, 1]}(u) (p_n *_{g_{j,\alpha,l,0}} p_n)(t_k^n, u, x, z)$ , it follows from Proposition 9-(ii) that  $(q_{t_k^n, \alpha, l}^{n,j}, n \geq 1, k \in \{1, \dots, n\})$  is a bounded sequence in  $\mathcal{G}(\mathbb{R}^d)$ . Since  $\varphi_{t_k^n, \alpha, l}^{n,j}(s, t, x, y) = \mathbf{1}_{(0, \frac{1}{n})}(s) \mathbf{1}_{[t_{k+1}^n, 1]}(t) (q_{t_k^n, \alpha, l}^{n,j} *_{h_{j,\alpha,l,\alpha}} p)(t_k^n + s, t, x, y)$ , Proposition 9-(i) shows that  $(\varphi_{t_k^n, \alpha, l}^{n,j}, n \geq 1, k \in \{1, \dots, n\})$  is bounded in  $\mathcal{H}_{|\alpha|}(\mathbb{R}^d)$ , so that  $(\varphi_{t_k^n}^{n,j}, n \geq 1, k \in \{1, \dots, n\})$  is bounded in  $\mathcal{H}_{2j}(\mathbb{R}^d)$ .

When  $k = 0$ , we have in the same way for all  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$

$$\Phi_{s,1/n}^{n,j} P_{t-1/n} f(x) = \int_{\mathbb{R}^d} f(y) \varphi_0^{n,j}(s, t, x, y) dy$$

where  $\varphi_0^{n,j} = \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} \varphi_{0,\alpha,l}^{n,j}$  with

$$\varphi_{0,\alpha,l}^{n,j}(s, t, x, y) = \mathbf{1}_{(0, \frac{1}{n})}(s) \mathbf{1}_{[\frac{1}{n}, 1]}(t) g_{j,\alpha,l}(x) (p_n *_{h_{j,\alpha,l,\alpha}} p)(s, t, x, y).$$

Again Proposition 9-(i) imply that  $(\varphi_0^{n,j}, n \geq 1)$  is bounded in  $\mathcal{H}_{2j}(\mathbb{R}^d)$ .

Eventually, for all  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ , we have  $R_{1,t}^{n,j} f(x) = \int_{\mathbb{R}^d} f(y) r_1^{n,j}(t, x, y) dy$  with

$$r_1^{n,j}(t, x, y) = \sum_{k=0}^{\lfloor nt \rfloor - 1} \int_0^{1/n} \int_0^{s_1} \cdots \int_0^{s_j} \varphi_{t_k^n}^{n,j+1}(s_{j+1}, t, x, y) ds_{j+1} \cdots ds_2 ds_1,$$

and since the family  $(\varphi_{t_k^n}^{n,j+1}, n \geq 1, k \in \{0, \dots, n\})$  is bounded in  $\mathcal{H}_{2j+2}(\mathbb{R}^d)$ , the sequence  $(r_1^{n,j}, n \geq 1)$  is  $O(1/n^j)$  in  $\mathcal{G}_{2j}(\mathbb{R}^d)$ .

As for  $R_{2,t}^{n,j}$ , similar arguments lead to

$$P_{[nt]/n}^n \Phi_{s,t-[nt]/n}^{n,j} f(x) = \int_{\mathbb{R}^d} f(y) \phi^{n,j}(s, t, x, y) dy$$

where  $\phi^{n,j} = \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} \phi_{\alpha,l}^{n,j}$  with

$$\begin{aligned} \phi_{\alpha,l}^{n,j}(s, t, x, y) &= \mathbf{1}_{[\frac{1}{n}, 1]}(t) \mathbf{1}_{(0, t - \frac{[nt]}{n})}(s) \int_{\mathbb{R}^d} (p_n * g_{j,\alpha,l,0} p_n) \left( \frac{[nt]}{n}, \frac{[nt]}{n} + s, x, z_2 \right) \\ h_{j,\alpha,l}(z_2) \partial_2^\alpha p \left( t - \frac{[nt]}{n} - s, z_2, y \right) dz_2 &+ \mathbf{1}_{\{0 < s < t < \frac{1}{n}\}} g_{j,\alpha,l}(x) (p_n * h_{j,\alpha,l,\alpha} p) (s, t, x, y). \end{aligned}$$

We can treat  $\phi_{\alpha,l}^{n,j}$  exactly as we have treated  $\varphi_{t_k^n}^{n,j}$ , and get that  $(\phi^{n,j}, n \geq 1)$  is bounded in  $\mathcal{H}_{2j}(\mathbb{R}^d)$ , so that  $R_{2,t}^{n,j}$  has a kernel  $(r_2^{n,j}, n \geq 1)$  defined by

$$r_2^{n,j}(t, x, y) = \int_0^{t-[nt]/n} \int_0^{s_1} \dots \int_0^{s_j} \phi^{n,j+1}(s_{j+1}, t, x, y) ds_{j+1} \dots ds_2 ds_1$$

which is  $O(1/n^j)$  in  $\mathcal{G}_{2j}(\mathbb{R}^d)$ .

Eventually, putting  $r^{n,j} = r_1^{n,j} + r_2^{n,j}$  completes the proof.  $\square$

In particular we have

**Corollary 26.** *Under (B) and (C),  $(p_n - p, n \geq 1)$  is  $O(1/n)$  in  $\mathcal{G}_2(\mathbb{R}^d)$ .*

*Proof.* From (2.38) applied with  $j = 1$  and Lemma 25, we have for all  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} f(y) (p_n - p)(t, x, y) dy = \Delta_t^n f(x) = R_t^{n,1} f(x) = \int_{\mathbb{R}^d} f(y) r^{n,1}(t, x, y) dy$$

so that  $p_n - p = r^{n,1}$ , and Lemma 25 gives the result.  $\square$

Eventually, we have kernels for the operators  $P_{t_k^n}^n L_j^* P_{t-t_k^n}$ :

**Lemma 27.** *Under (B) and (C), for each  $j \in \mathbb{N}^*$ , there exists a bounded sequence  $(\psi_{t_k^n}^{n,j}, n \geq 1, k \in \{0, \dots, n\})$  in  $\mathcal{G}_{2j}(\mathbb{R}^d)$  such that for all  $t \in (0, 1]$ ,  $n \geq 1$ ,  $k \in \{0, \dots, [nt]\}$ ,  $f \in C_{\text{pol}}^\infty(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,*

$$P_{t_k^n}^n L_j^* P_{t-t_k^n} f(x) = \int_{\mathbb{R}^d} f(y) \psi_{t_k^n}^{n,j}(t, x, y) dy.$$

The proof is omitted since it copies the arguments of the proof of Lemma 25 - it is even a bit simpler.



## 2.4.2 Operators on $\mathcal{G}_l(\mathbb{R}^d)$

When  $\pi \in \mathcal{G}_l(\mathbb{R}^d)$ ,  $\pi(t, \cdot, y) \in L^\infty(\mathbb{R}^d)$  so that for  $s \in [0, 1]$  and  $n \geq 1$  we can define two functions  $P_s\pi$  and  $P_s^n\pi$  on  $(0, 1] \times \mathbb{R}^d \times \mathbb{R}^d$  by  $P_s\pi(t, \cdot, y) = \mathbf{1}_{\{s \leq t\}} P_s(\pi(t, \cdot, y))$  and  $P_s^n\pi(t, \cdot, y) = \mathbf{1}_{\{s \leq t\}} P_s^n(\pi(t, \cdot, y))$ , i.e.

$$P_s\pi(t, x, y) = \mathbf{1}_{\{s \leq t\}} \mathbb{E}[\pi(t, X_s^x, y)] \quad \text{and} \quad P_s^n\pi(t, x, y) = \mathbf{1}_{\{s \leq t\}} \mathbb{E}[\pi(t, X_s^{n,x}, y)]. \quad (2.51)$$

We also write  $\Delta_s^n\pi = P_s^n\pi - P_s\pi$ . For  $j \in \mathbb{N}^*$  we denote by  $\Phi^j$  the family  $(\Phi_{s,1/n}^{n,j}, n \geq 1, s \in [0, 1/n])$  of operators on  $\mathcal{G}_l(\mathbb{R}^d)$  defined as in (2.31) by

$$\Phi_{s,1/n}^{n,j}\pi(t, x, y) = \mathbb{E}[L_j^x P_{1/n-s}\pi(t, X_s^{n,x}, y)],$$

i.e., using (2.30),

$$\Phi_{s,1/n}^{n,j} = \sum_{1 \leq |\alpha| \leq 2j} \sum_{l=1}^{m_{j,\alpha}} g_{j,\alpha,l} P_s^n(h_{j,\alpha,l} \partial^\alpha P_{1/n-s}). \quad (2.52)$$

Denoting by  $\mathcal{L}_b(\mathcal{G}_l(\mathbb{R}^d), \mathcal{G}_{l'}(\mathbb{R}^d))$  the space of all morphisms mapping any bounded subset of  $\mathcal{G}_l(\mathbb{R}^d)$  into a bounded subset of  $\mathcal{G}_{l'}(\mathbb{R}^d)$ , we then have

**Proposition 28.** *Under (B) and (C),  $(P_s, s \in [0, 1])$  and  $(P_s^n, s \in [0, 1], n \geq 1)$  are bounded families in  $\mathcal{L}_b(\mathcal{G}_l(\mathbb{R}^d))$ , and  $\Phi^j$  is a bounded family in  $\mathcal{L}_b(\mathcal{G}_l(\mathbb{R}^d), \mathcal{G}_{l+2j}(\mathbb{R}^d))$ .*

*Proof.* Let us first deal with  $(P_s)$ . Let  $\pi \in \mathcal{G}_l(\mathbb{R}^d)$ .  $P_s$  is measurable. Moreover, Lebesgue's dominated convergence theorem shows that  $P_s\pi(t, x, \cdot)$  is infinitely differentiable and that for all  $\beta \in \mathbb{N}^d$

$$\partial_y^\beta P_s\pi(t, x, y) = \mathbf{1}_{\{s \leq t\}} \mathbb{E}[\partial_3^\beta \pi(t, X_s^x, y)].$$

Hypothesis (A) ensures that a version of  $X^x$  can be chosen such that for each  $t \geq 0$ , the map  $x \mapsto X_t^x$  is infinitely differentiable (see, for example, [64]). Since  $\partial_3^\beta \pi(t, \cdot, y) \in C_{\text{pol}}^\infty(\mathbb{R}^d)$ , it follows from Theorem 3.14 page 16 in [67] that  $\partial_y^\beta P_s\pi(t, \cdot, y)$  is infinitely differentiable and that for all  $\alpha \in \mathbb{N}^d$  there exists universal polynomials  $(\Pi_{\alpha,\mu}, |\mu| \leq |\alpha|)$  such that

$$\partial_x^\alpha \partial_y^\beta P_s\pi(t, x, y) = \mathbf{1}_{\{s \leq t\}} \sum_{|\mu| \leq |\alpha|} \mathbb{E}[\partial_2^\mu \partial_3^\beta \pi(t, X_s^x, y) \Pi_{\alpha,\mu}(\partial_x^\nu X_s^x, |\nu| \leq |\alpha|)] \quad (2.53)$$

with

$$\sup_{s \in [0,1], x \in \mathbb{R}^d} \mathbb{E}[\Pi_{\alpha,\mu}(\partial_x^\nu X_s^x, |\nu| \leq |\alpha|)^2] < \infty \quad (2.54)$$

for all  $|\mu| \leq |\alpha|$ . As a consequence,  $P_s\pi(t, \cdot, \cdot)$  is infinitely differentiable and using Cauchy-Schwarz's inequality, (2.14) and (2.54), we see that for all bounded  $\mathcal{B} \subset$

$\mathcal{G}_l(\mathbb{R}^d)$  and  $\alpha, \beta \in \mathbb{N}^d$ , there exists two constants  $c_1 \geq 0$  and  $c_2 > 0$  such that for all  $\pi \in \mathcal{B}$ ,  $s \in [0, 1]$ ,  $t \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,

$$|\partial_x^\alpha \partial_y^\beta P_s \pi(t, x, y)| \leq c_1 \mathbf{1}_{\{s \leq t\}} t^{-(|\alpha| + |\beta| + d + t)/2} \mathbb{E} [\exp(-c_2 \|X_x^s - y\|^2 / t)]^{1/2}. \quad (2.55)$$

Now, partitioning  $\Omega$  into  $\{\|X_x^s - y\| \leq \|x - y\| / 2\}$  and  $\{\|X_x^s - y\| > \|x - y\| / 2\}$ , we have

$$\mathbb{E} [\exp(-c_2 \|X_x^s - y\|^2 / t)] \leq \mathbb{P}(\|X_x^s - y\| \leq \|x - y\| / 2) + \exp(-c_2 \|x - y\|^2 / 4t). \quad (2.56)$$

Using (2.16) for  $p \in \mathcal{G}(\mathbb{R}^d)$  for the fourth inequality, we can find  $c_3, c_5 \geq 0$  and  $c_4, c_6 > 0$  such that for all  $s \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,

$$\begin{aligned} \mathbb{P}(\|X_s^x - y\| \leq \|x - y\| / 2) &\leq \mathbb{P}(\|X_s^x - x\| \geq \|x - y\| / 2) \\ &= \int_{\mathbb{R}^d} \mathbf{1}_{\{\|z-x\| \geq \|x-y\|/2\}} p(s, x, z) dz \\ &= \int_{\mathbb{R}^d} \mathbf{1}_{\{\|\xi\| \geq \|x-y\|/2\sqrt{s}\}} p(s, x, x + \xi\sqrt{s}) s^{d/2} d\xi \\ &\leq c_3 \int_{\mathbb{R}^d} \mathbf{1}_{\{\|\xi\| \geq \|x-y\|/2\sqrt{s}\}} \exp(-c_4 \|\xi\|^2) d\xi \\ &\leq c_5 \exp(-c_6 \|x - y\|^2 / s). \end{aligned} \quad (2.57)$$

Eventually, from (2.56) and (2.57), we can find  $c_7 \geq 0$  and  $c_8 > 0$  such that for all  $s \in [0, 1]$ ,  $t \in (0, 1]$  and  $x, y \in \mathbb{R}^d$ ,

$$\begin{aligned} \mathbf{1}_{\{s \leq t\}} \mathbb{E} [\exp(-c_2 \|X_x^s - y\|^2 / t)] \\ \leq c_5 \exp(-c_6 \|x - y\|^2 / t) + \exp(-c_2 \|x - y\|^2 / 4t) \\ \leq c_7 \exp(-c_8 \|x - y\|^2 / t). \end{aligned} \quad (2.58)$$

It is enough to inject (2.58) into (2.55) to complete the proof for  $(P_s)$ .

This proof naturally extends to the case of  $(P_s^n)$ . Indeed, (2.53) holds with  $(X^n, P^n)$  instead of  $(X, P)$ . Moreover, from Lemma 30, (2.54) holds uniformly in  $n$  with  $X^n$  instead of  $X$ . Eventually, (2.57) holds with  $X^n$  instead of  $X$ , uniformly in  $n$  because  $(p_n, n \geq 1)$  is bounded in  $\mathcal{G}(\mathbb{R}^d)$ .

As for  $\Phi^j$ , it is enough to use (2.52), the boundedness of  $(P_s)$  and  $(P_s^n)$ , Remark 15 and the facts that multiplication by a function in  $\mathcal{B}$  belongs to  $\mathcal{L}_b(\mathcal{G}_l(\mathbb{R}^d), \mathcal{G}_l(\mathbb{R}^d))$  and that  $\partial_2^\alpha \in \mathcal{L}_b(\mathcal{G}_l(\mathbb{R}^d), \mathcal{G}_{l+|\alpha|}(\mathbb{R}^d))$ .  $\square$

### 2.4.3 Moments for the Euler scheme and its derivatives

Let us assume (A). Then it is known that  $X_t^{n,x}$  has bounded moments of any order and that for all  $q \in \mathbb{N}$ , one can find  $c \geq 0$  such that for all  $x \in \mathbb{R}^d$ ,

$$\sup_{t \in [0, 1], n \geq 1} \mathbb{E} [\|X_t^{n,x}\|^q] \leq c(1 + \|x\|^q) \quad (2.59)$$

(see [92]). From (2.4),  $x \mapsto X_t^{n,x}$  is infinitely differentiable and we shall see that analogous upper bounds hold for its derivatives. Following [67], for  $m \geq 1$ , we denote by  $X_t^{(m),n,x}$  the  $m$ -th derivative of  $x \mapsto X_t^{n,x}$  at point  $x$ . It should be thought of as a  $d \times d^m$  matrix. For instance,  $X_t^{(1),n,x}$  is the jacobian matrix of  $x \mapsto X_t^{n,x}$ . Differentiating (2.4), we have

$$X_t^{(1),n,x} = I + \int_0^t b^{(1)}(X_{[ns]/n}^{n,x}) X_{[ns]/n}^{(1),n,x} ds + \sum_{j=1}^r \int_0^t \sigma_j^{(1)}(X_{[ns]/n}^{n,x}) X_{[ns]/n}^{(1),n,x} dB_s^j, \quad (2.60)$$

where  $I$  stands for the identity matrix and  $\sigma_j$  is the  $j$ -th column of  $\sigma$ . Besides, by induction, there are for each  $m \geq 2$  universal polynomials  $P_{m,j}$ ,  $j \in \{0, \dots, r\}$ , such that

$$\begin{aligned} X_t^{(m),n,x} &= \int_0^t b^{(1)}(X_{[ns]/n}^{n,x}) X_{[ns]/n}^{(m),n,x} ds + \sum_{j=1}^r \int_0^t \sigma_j^{(1)}(X_{[ns]/n}^{n,x}) X_{[ns]/n}^{(m),n,x} dB_s^j \\ &\quad + \int_0^t Q_{m,0,[ns]/n}^{n,x} ds + \sum_{j=1}^r \int_0^t Q_{m,j,[ns]/n}^{n,x} dB_s^j, \end{aligned} \quad (2.61)$$

where

$$\begin{cases} Q_{m,0,t}^{n,x} &= P_{m,0}(b^{(2)}(X_t^{n,x}), \dots, b^{(m)}(X_t^{n,x}), X_t^{(1),n,x}, \dots, X_t^{(m-1),n,x}), \\ Q_{m,j,t}^{n,x} &= P_{m,j}(\sigma_j^{(2)}(X_t^{n,x}), \dots, \sigma_j^{(m)}(X_t^{n,x}), X_t^{(1),n,x}, \dots, X_t^{(m-1),n,x}). \end{cases} \quad (2.62)$$

This is analogous to (1.8) page 4 in [67]. Then we have

**Lemma 29.** *Under (A), for all  $m \geq 1$  and  $q \in \mathbb{N}$ , there exists  $c \geq 0$  and  $q' \in \mathbb{N}$  such that for all  $x \in \mathbb{R}^d$ ,*

$$\sup_{t \in [0,1], n \geq 1} \mathbb{E} \left[ \left\| X_t^{(m),n,x} \right\|^q \right] \leq c \left( 1 + \|x\|^{q'} \right). \quad (2.63)$$

*Proof.* We give a proof by induction on  $m$ . Let us first assume that  $m = 1$ . Let  $q \in \mathbb{N}$ . From (2.60), and observing that (A) states that  $b^{(1)}$  and all the  $\sigma_j^{(1)}$  are bounded, Jensen's and Burkholder-Davis-Gundy's inequalities lead to the existence of  $c \geq 0$  such that for all  $t \in [0, 1]$ ,  $n \geq 1$  and  $x \in \mathbb{R}^d$ ,

$$\mathbb{E} \left[ \left\| X_t^{(1),n,x} \right\|^q \right] \leq c \left( 1 + \int_0^t \mathbb{E} \left[ \left\| X_{[ns]/n}^{(1),n,x} \right\|^q \right] ds \right).$$

Taking this inequality at time  $[nt]/n$  and applying Gronwall's lemma, we get that

$$\sup_{t \in [0,1], n \geq 1, x \in \mathbb{R}^d} \mathbb{E} \left[ \left\| X_{[nt]/n}^{(1),n,x} \right\|^q \right] < \infty.$$

From (2.4), one easily checks that the same holds at time  $t$  instead of  $\lfloor nt \rfloor/n$ , so that (2.63) holds for  $m = 1$  with  $q' = 0$ .

Let us now assume that (2.63) holds for the  $m - 1$  first derivatives. Let  $q \in \mathbb{N}$ . From (2.61), and observing again that (A) states that  $b^{(1)}$  and all the  $\sigma_j^{(1)}$  are bounded, Jensen's and Burkholder-Davis-Gundy's inequalities lead to the existence of  $c_1 \geq 0$  such that for all  $t \in [0, 1]$ ,  $n \geq 1$  and  $x \in \mathbb{R}^d$ ,

$$\mathbb{E} \left[ \left\| X_t^{(m),n,x} \right\|^q \right] \leq c_1 \left( \int_0^t \mathbb{E} \left[ \left\| X_{\lfloor ns \rfloor/n}^{(m),n,x} \right\|^q \right] ds + \int_0^t \sum_{j=0}^r \mathbb{E} \left[ \left\| Q_{m,j,\lfloor ns \rfloor/n}^{n,x} \right\|^q \right] ds \right). \quad (2.64)$$

Using (2.62), the induction hypothesis, (A) and (2.59), we find  $c_2 \geq 0$  and  $q' \in \mathbb{N}$  such that for all  $s \in [0, 1]$ ,  $n \geq 1$  and  $x \in \mathbb{R}^d$ ,

$$\sum_{j=0}^r \mathbb{E} \left[ \left\| Q_{m,j,\lfloor ns \rfloor/n}^{n,x} \right\|^q \right] \leq c_2 \left( 1 + \|x\|^{q'} \right).$$

Thus, taking (2.64) at time  $\lfloor nt \rfloor/n$  and applying Gronwall's lemma, we find  $c \geq 0$  such that for all  $x \in \mathbb{R}^d$ ,

$$\sup_{t \in [0,1], n \geq 1} \mathbb{E} \left[ \left\| X_{\lfloor nt \rfloor/n}^{(m),n,x} \right\|^q \right] \leq c \left( 1 + \|x\|^{q'} \right).$$

From (2.4), one easily checks that the same holds at time  $t$  instead of  $\lfloor nt \rfloor/n$ , which completes the proof.  $\square$

Observe that, under (B), the above proof holds with  $q' = 0$  so that we have

**Lemma 30.** *Under (B), for all  $m \geq 1$  and  $q \in \mathbb{N}$ ,*

$$\sup_{t \in [0,1], n \geq 1, x \in \mathbb{R}^d} \mathbb{E} \left[ \left\| X_t^{(m),n,x} \right\|^q \right] < \infty.$$

# Chapitre 3

## Un modèle à volatilité stochastique

Dans ce chapitre, nous étudions un modèle à volatilité stochastique proposé par Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR [29]. Après quelques rappels sur le modèle de Black-Scholes et une rapide présentation des modèles à volatilité stochastique les plus connus, nous résumons le travail théorique des auteurs, concernant en particulier des développements asymptotiques du prix et de la couverture des options vanille lorsque la volatilité retourne vite à sa moyenne. Dans ce cas, on facture au client un risque de volatilité qui dépend peu du strike. Aussi, afin de générer des smiles de grande amplitude ressemblant à ceux souvent observés sur les marchés, nous proposons ensuite des exemples de simulation informatique dans le cas non asymptotique où la constante de temps de retour à la moyenne de la volatilité est comparable à la maturité de l'option étudiée.

Ces travaux ont été menés au sein du Département des marchés de capitaux du Crédit Lyonnais - aujourd'hui Calyon - dans l'équipe de recherche "taux et change" dirigée par Christophe MICHEL, alors secondé par Joe BONNAUD. Une version longue, plus détaillée, de ce texte constitue le chapitre consacré à la volatilité stochastique des notes de cours de Master de Nicole EL KAROUI [28].

### 3.1 Introduction

#### 3.1.1 Notations

Nous étudions le modèle suivant :

$$\begin{cases} dX_t = \mu X_t dt + \sigma_t X_t dW_t, \\ \sigma_t = f(Y_t), \\ dY_t = \alpha(m - Y_t) dt + \beta d\hat{Z}_t. \end{cases} \quad (3.1)$$

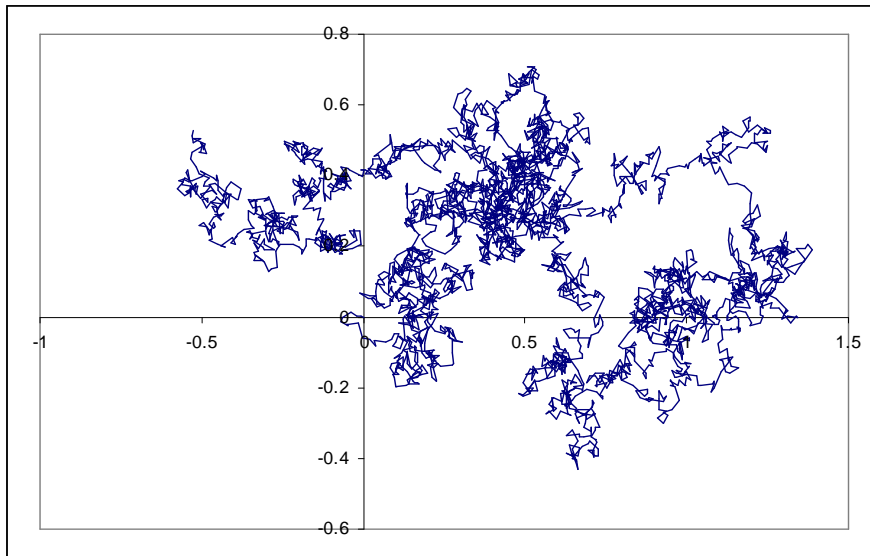


FIG. 3.1 – Une trajectoire de  $(W, Z)$  entre  $t = 0$  et  $t = 1$

Dans cette écriture :

- $X$  représente le sous-jacent,  $X_t$  son cours à la date  $t$ ,
- $\mu$  est le rendement instantané, supposé constant,
- $\sigma_t$  est la valeur à la date  $t$  de la volatilité du cours du sous-jacent ; elle mesure l'intensité du bruit  $\sigma_t X_t dW_t$  auquel est soumis le cours du sous-jacent,
- $W$  est un mouvement brownien standard, la volatilité  $\sigma$  est elle-même un processus stochastique, fonction déterministe du processus  $Y$  ; la fonction  $f$  est définie sur  $\mathbb{R}$  et à valeurs strictement positives,
- $Y$  est un processus d'Ornstein-Uhlenbeck, de moyenne à long terme  $m$  et de variance à long terme  $\frac{\beta^2}{2\alpha}$ ,
- $\hat{Z}$  est un mouvement brownien standard éventuellement corrélé à  $W$  ; nous supposons dans la suite cette corrélation constante et nous la notons  $\rho$ , avec  $\rho \in ]-1, 1[$ , de sorte que  $d\langle W, \hat{Z} \rangle_t = \rho dt$ . Si nous définissons  $Z_t$  par l'égalité  $\hat{Z}_t = \rho W_t + \sqrt{1 - \rho^2} Z_t$ , alors  $W$  et  $Z$  sont deux browniens indépendants.

Les figures 3.1 et 3.2 donnent des exemples de trajectoires de  $(W, Z)$  et de  $(W, \hat{Z})$ , dans le cas où  $\rho = 0.5$  puis dans le cas où  $\rho = -0.5$ . Dans les trois cas, on a utilisé les mêmes séries d'aléas.

Mathématiquement, nous nous plaçons sur un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  muni de la filtration  $\mathcal{F}_t = \sigma(W_s, Z_s, 0 \leq s \leq t)$ . Par exemple, on pourra considérer l'espace  $\Omega = C^0(\mathbb{R}_+, \mathbb{R}^2)$  des fonctions continues à valeurs dans  $\mathbb{R}^2$  muni de sa tribu borélienne et de la mesure de Wiener sur cette tribu. Dans ce cas, il faut voir un événement ponctuel  $\omega$  comme une trajectoire  $t \mapsto (W_t(\omega), Z_t(\omega))$ . Par ailleurs, la filtration  $\mathcal{F}_t$  représente l'information sur les deux mouvements browniens  $W$  et

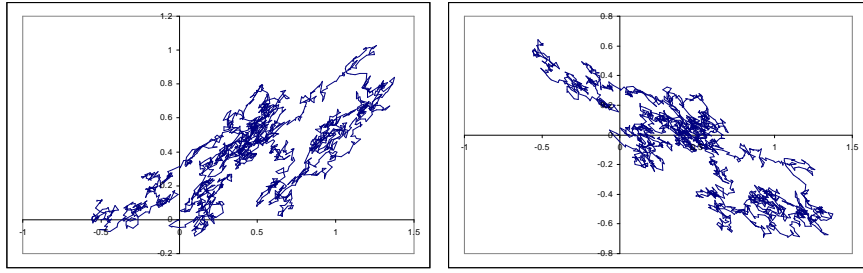


FIG. 3.2 – Une trajectoire de  $(W, \widehat{Z})$  entre  $t = 0$  et  $t = 1$ , avec  $\rho = 0.5$  (gauche) ou  $-0.5$  (droite)

$Z$  jusqu'à la date  $t$ ; c'est l'augmentation habituelle de la tribu engendrée par les ensembles de la forme  $\{\omega \in \Omega \mid W_s \in [a_1, b_1], Z_s \in [a_2, b_2], 0 \leq s \leq t\}$ .

Dans toute la suite, nous nous plaçons sous l'hypothèse d'absence d'opportunité d'arbitrage. Plus précisément, nous supposons l'existence d'une probabilité sur  $(\Omega, \mathcal{F})$  sous laquelle le prix des actifs actualisé est une martingale locale.

### 3.1.2 Interprétation

F. BLACK et M. SCHOLLES ont proposé [8] de modéliser la dynamique du cours  $X_t$  du sous-jacent par l'équation différentielle stochastique

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x. \quad (3.2)$$

Typiquement, le sous-jacent est une action ou un indice boursier. Dans ce modèle,  $\sigma$  est une constante strictement positive, i.e. une quantité supposée indépendante du temps et du hasard, qu'on appelle "volatilité". Nous considérons également un actif sans risque  $X^0$  dont la valeur à la date  $t$  est  $X_t^0 = e^{rt}$ . Ceci revient à supposer le taux d'intérêt à court terme constant égal à  $r$ . L'équation (3.2) a des conséquences importantes :

- Le processus  $X$  est un mouvement brownien géométrique; on dispose d'une expression explicite pour  $X_t$  :

$$X_t = x \exp \left( \sigma W_t + \left( \mu - \frac{\sigma^2}{2} \right) t \right) \quad (3.3)$$

qui prouve que le logarithme du cours  $X_t$  suit une loi gaussienne de moyenne  $\left( \mu - \frac{\sigma^2}{2} \right) t$  et de variance  $\sigma^2 t$ . On peut réécrire (3.3) sous la forme  $X_t = e^{\mu t} M_t$ , où  $M_t = \exp \left( \sigma W_t - \frac{\sigma^2}{2} t \right)$  est une  $\mathbb{P}$ -martingale de moyenne 1, relativement à la filtration  $(\mathcal{G}_t)$  engendrée par  $W$ , i.e.  $\mathcal{G}_t = \sigma(W_s, 0 \leq s \leq t)$ . La figure 3.3 présente quatre trajectoires possibles du processus  $(X_t)_{0 \leq t \leq 1}$  et la trajectoire moyenne.

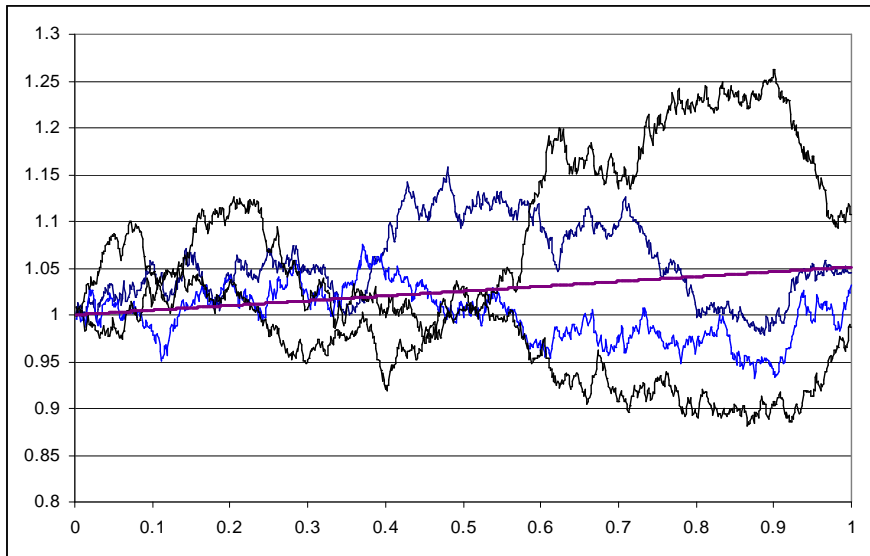


FIG. 3.3 – Quatre trajectoires possibles du sous-jacent sous le modèle de BLACK et SCHOLES;  $\mu = 0.05$ ,  $\sigma = 0.15$

- Le marché est *viable et complet* :
  - Il existe *une et une seule* probabilité  $\mathbb{P}^*$  sous laquelle le processus des prix actualisés  $(e^{-rt}X_t)_{t \geq 0}$  de l'actif risqué est une  $\mathbb{P}^*$ -martingale. Cette probabilité est appelée probabilité risque-neutre.
  - L'évolution du sous-jacent s'écrit

$$dX_t = rX_t dt + \sigma X_t dW_t^*, \quad X_0 = x,$$

où  $W^*$  est un  $\mathbb{P}^*$ -mouvement brownien.

- Toute option européenne de payoff  $H \in L^2(\mathbb{P}^*, \mathcal{G}_T)$ , c'est-à-dire toute option définie par une variable aléatoire  $H$   $\mathcal{G}_T$ -mesurable et de carré intégrable sous la probabilité  $\mathbb{P}^*$ , est simulable : il existe un unique portefeuille admissible, i.e. autofinancé et minoré, ne contenant que de l'actif sans risque et de l'actif risqué, dont la valeur en  $T$  est  $H$ . De plus, la valeur  $V(t)$  de l'option est, sous la probabilité risque-neutre, l'espérance actualisée du flux terminal  $H$  :

$$V(t) = \mathbb{E}^* [e^{-r(T-t)} H \mid \mathcal{G}_t].$$

Ceci est une conséquence du théorème de représentation des martingales browniennes (notons que  $\sigma(W_s, 0 \leq s \leq t) = \sigma(W_s^*, 0 \leq s \leq t)$ ). Autrement dit, on peut se couvrir parfaitement - éliminer le risque - en gérant dynamiquement un portefeuille ne contenant que du liquide et du sous-jacent. Notons que  $V(t)$  est indépendant de la tendance  $\mu$ .



- Dans le cas particulier où  $H = h(X_T)$ , avec  $h$  continue et positive, le prix de l'option se met sous la forme  $P(t, X_t)$  avec

$$P(t, x) = \mathbb{E}^* \left[ e^{-r(T-t)} h \left( x \exp^{\sigma(W_T^* - W_t^*) + (r - \frac{\sigma^2}{2})(T-t)} \right) \right]. \quad (3.4)$$

La fonction  $P$  est solution de l'équation aux dérivées partielles

$$\begin{cases} \mathcal{L}_{BS}(\sigma)P = 0, \\ \forall x > 0, P(T, x) = h(x), \end{cases} \quad (3.5)$$

où

$$\mathcal{L}_{BS}(\sigma) = \frac{\partial}{\partial t} + \frac{\sigma^2}{2} x^2 \frac{\partial^2}{\partial x^2} + r \left( x \frac{\partial}{\partial x} - \cdot \right). \quad (3.6)$$

De plus, la quantité d'actif risqué à détenir à la date  $t$  est

$$a_t = \frac{\partial P}{\partial x}(t, X_t),$$

quantité qu'on appelle le "delta". Par conséquent, le portefeuille de couverture contient

$$b_t = e^{-rt} \left( P(t, X_t) - X_t \frac{\partial P}{\partial x}(t, X_t) \right)$$

unités d'actif sans risque.

- Le cas du call correspond au payoff  $h(x) = (x - K)_+$ , on note alors  $P(t, x) = C_{BS}(t, x; K, T; \sigma)$  et

$$C_{BS}(t, x; K, T; \sigma) = xN(d_+) - Ke^{-r(T-t)}N(d_-) \quad (3.7)$$

où  $N(d) = \int_{-\infty}^d \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$  désigne la fonction de répartition de la loi gaussienne centrée réduite et

$$\begin{cases} d_+ = \frac{\ln\left(\frac{x}{Ke^{-r(T-t)}}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}, \\ d_- = \frac{\ln\left(\frac{x}{Ke^{-r(T-t)}}\right)}{\sigma\sqrt{T-t}} - \frac{1}{2}\sigma\sqrt{T-t}. \end{cases} \quad (3.8)$$

On donne en figure 3.4 le graphe de la fonction  $x \mapsto C_{BS}(t, x; K, T; \sigma)$  pour les valeurs suivantes des paramètres :  $t = 0$ ,  $K = 1.1$ ,  $T = 1$ ,  $\sigma = 0.15$ ,  $r = 0.05$ . On a superposé le graphe donnant le payoff du call, i.e.  $x \mapsto (x - K)_+$ . En figure 3.5, on modifie la valeur de  $\sigma$ , toutes choses égales par ailleurs. On regarde succesivement une volatilité de 5% puis une volatilité de 50%. Les trois graphes sont à la même échelle.

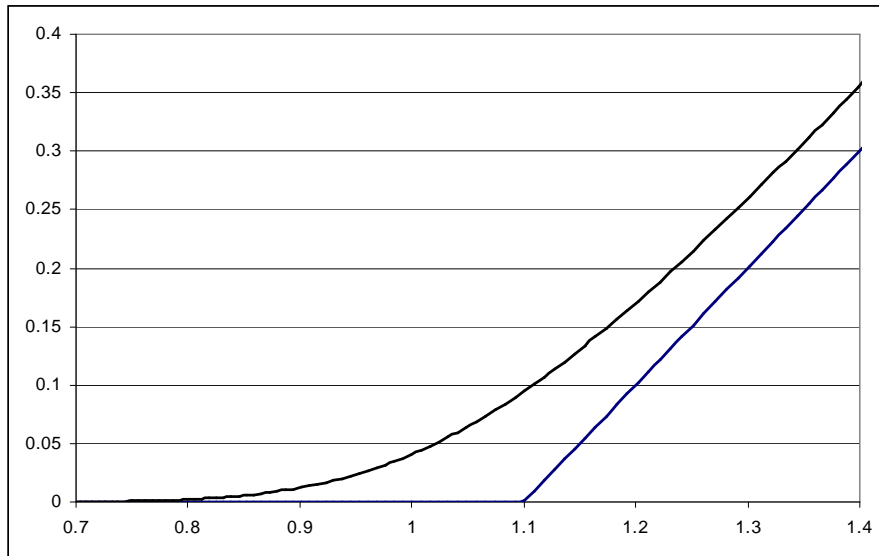


FIG. 3.4 – Prix du call dans le modèle de Black-Scholes;  $t = 0$ ,  $K = 1.1$ ,  $T = 1$ ,  $\sigma = 0.15$ ,  $r = 0.05$

De plus, le portefeuille de couverture contient la quantité

$$a_t = N(d_+)$$

d'actifs risqués. Ceci prouve que l'égalité (3.7) donnant le prix du call sous le modèle de BLACK et SCHOLLES donne aussi la décomposition du portefeuille de couverture en actif risqué et en actif sans risque.

Signalons que ce qui précède reste vrai si l'on autorise le taux d'intérêt à court terme, dit aussi "taux court", et la volatilité à dépendre du temps, *mais pas du hasard*. Il suffit de remplacer  $r$  par

$$\bar{r} = \frac{1}{T-t} \int_t^T r_s ds$$

et  $\sigma$  par  $\bar{\sigma}$  où

$$\bar{\sigma}^2 = \frac{1}{T-t} \int_t^T \sigma_s^2 ds$$

dans les formules (3.4), (3.5), (3.6), (3.7) et (3.8).

Le modèle de Black-Scholes sert de référence à tous ceux qui pratiquent la finance des marchés :

- Il est *simple* : adopter le modèle de Black-Scholes, c'est simplement supposer les cours  $X$  à trajectoires continues et à accroissements relatifs indépendants et stationnaires.

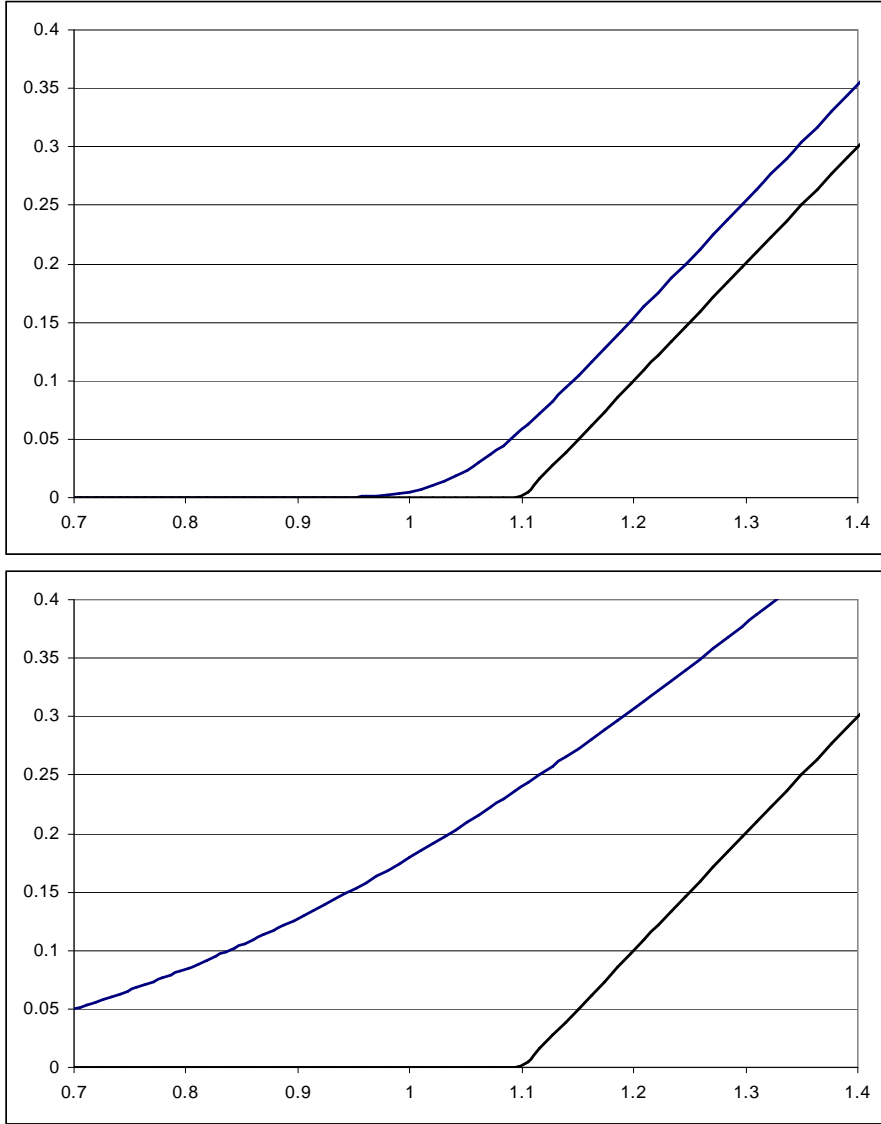


FIG. 3.5 – Prix du call dans le modèle de BLACK et SCHOLLES;  $t = 0$ ,  $K = 1.1$ ,  $T = 1$ ,  $r = 0.05$ ;  $\sigma = 0.05$  (haut) ou  $0.5$  (bas)

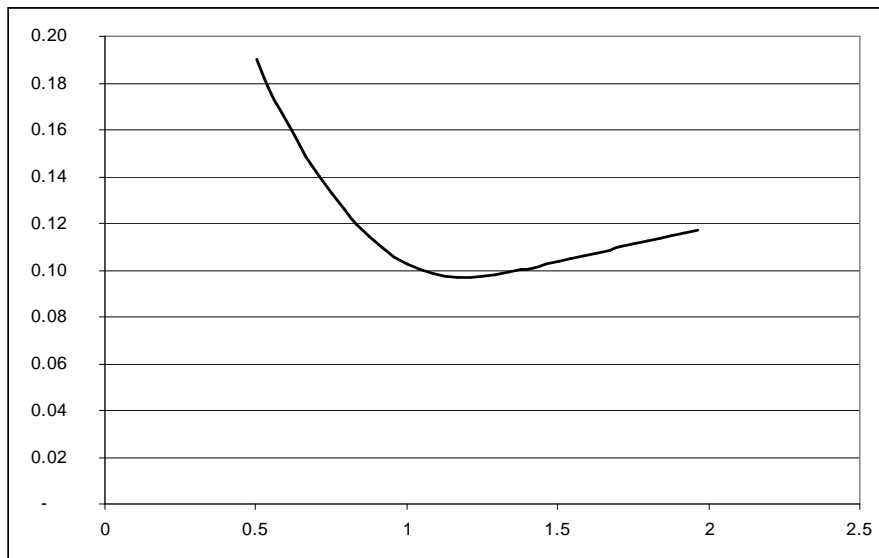


FIG. 3.6 – Un smile observé par un trader du Crédit Lyonnais en traitant des swaptions 5 ans - 10 ans

- Il est *maniab*le : il donne lieu à des formules fermées pour le prix des calls et des puts et pour les deltas correspondants, c'est-à-dire pour les quantités d'actifs risqués que doit contenir le portefeuille de couverture.

Cependant :

- Tous les tests statistiques invalident l'hypothèse log-normale pour le cours du sous-jacent. En réalité, il semble que les queues de distribution soient plus épaisses que ne le prévoit le modèle de BLACK et SCHOLES. De plus, les queues de distribution empiriques de  $\ln(X_t)$  sont souvent asymétriques.
- Nous définissons la volatilité implicite  $I$  par l'égalité

$$C_{BS}(t, x; K, T; I) = C^{obs}$$

où  $C^{obs}$  est le prix observé du call de maturité  $T$  et de strike  $K$ . La définition a un sens puisque l'application  $\sigma \mapsto C_{BS}(t, x; K, T; \sigma)$  est une bijection de  $\mathbb{R}_+^*$  sur  $](x - K)_+, x[$ . Ainsi définie,  $I$  est une fonction de  $t, x, K, T$  et  $C^{obs}$ . Si les prix observés étaient exactement les prix prévus par le modèle de Black-Scholes, la fonction  $K \mapsto I(t, x, K, T, C^{obs})$  serait constante et égale au paramètre  $\sigma$ . Or, les données de marché font apparaître une dépendance en  $K$ . La courbe empirique  $K \mapsto I(t, x, K, T, C^{obs})$  porte souvent le nom de courbe de "smile" en référence à son allure souriante (convexe décroissante puis croissante).

Bien sûr pour expliquer ces phénomènes, il faut raffiner le modèle. Il y a bien des façons de le faire. On peut autoriser les cours  $X_t$  à avoir des sauts, par exemple en remplaçant  $W$  par un processus de Lévy général. On pourra à ce sujet consulter

l'ouvrage [19] de Rama CONT et Peter TANKOV. On peut aussi autoriser la volatilité à dépendre de  $t$  et de  $x$ , la seule source de bruit restant le brownien  $W$ . C'est ce que propose Bruno DUPIRE [26]. On parle alors de modèles à volatilité locale.

Une autre manière naturelle d'étendre le modèle de Black-Scholes est d'autoriser la volatilité à être un processus stochastique gouverné par un deuxième bruit. On parle alors de modèles à volatilité stochastique. Ce deuxième bruit est généralement modélisé par un deuxième brownien  $\hat{Z}$  éventuellement corrélé à  $W$ , mais non parfaitement corrélé, contrairement au cas du modèle de DUPIRE. Dans cette classe de modèle, on conserve donc l'écriture

$$dX_t = \mu X_t dt + \sigma_t X_t dW_t$$

mais  $\sigma$  est désormais un processus aléatoire dépendant du temps et du hasard  $(W, \hat{Z})$ .

C'est donc ici un modèle à volatilité stochastique, initialement proposé par Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR [29], que nous allons étudier. Précisément, nous écrivons  $\sigma_t = f(Y_t)$ , où  $f : \mathbb{R} \rightarrow \mathbb{R}_+^*$  est une fonction déterministe et  $Y$  un processus dit d'Ornstein-Uhlenbeck (O-U) :

$$dY_t = \alpha(m - Y_t) dt + \beta d\hat{Z}_t. \quad (3.9)$$

Le paramètre  $\alpha$  s'appelle le *taux de retour à la moyenne* et le paramètre  $m$  la *moyenne à long terme*. On peut voir  $Y_t$  comme la position à la date  $t$  d'une particule soumise à une force de rappel d'intensité  $\alpha$  qui a tendance à la ramener à sa position d'équilibre (déterministe)  $m$  et à une force aléatoire - par exemple des chocs - modélisée par le bruit  $\beta d\hat{Z}_t$ .

### 3.1.3 D'autres modèles à volatilité stochastique

Un grand nombre de modèles à volatilité stochastique ont été étudiés depuis le milieu des années 1980, générant une littérature extrêmement riche qu'il serait impossible de décrire ici de manière exhaustive. En revanche, nous pouvons mentionner quelques modèles "repères" qui permettent de situer le modèle (3.1) dans le développement des modèles à volatilité stochastique.

#### Les modèles d'Hull et White, de Stein et Stein et d'Heston

HULL et WHITE [53], STEIN et STEIN [88], et HESTON [44] proposent différents modèles pour la volatilité  $(\sigma_t, 0 \leq t \leq T)$ . Le tableau suivant, inspiré de [23], les décrit :

Modèle	Processus de volatilité
Hull et White	$d\sigma_t = \sigma_t \{ (2\mu + v^2) dt + 2v d\hat{Z}_t \}$
Stein et Stein	$d\sigma_t = (\alpha - \lambda\sigma_t) dt + v d\hat{Z}_t$
Heston	$d\sigma_t = (v^2 - 2\lambda\sigma_t) dt + 2v\sqrt{\sigma_t} d\hat{Z}_t$

Le modèle d'Hull et White suppose une dynamique Black-Scholes sur le processus de volatilité, qui est donc un mouvement brownien géométrique. Les modèles de Stein et Stein et d'Heston sont voisins du modèle de Fouque, Papanicolaou et Sircar. Comme lui, ils capturent, par leur coefficient de dérive, le phénomène de retour à la moyenne. Le modèle de Stein et Stein est en fait un sous-modèle du modèle (3.1) dans lequel on a choisi  $f(x) = x$ . Il autorise donc des volatilités négatives,  $(\sigma_t, 0 \leq t \leq T)$  étant directement un processus d'O-U. Quant au modèle d'Heston, il ne diffère du modèle de Stein et Stein que par le terme de diffusion en  $\sqrt{\sigma_t}$  qui empêche  $\sigma_t$  de prendre des valeurs négatives. La volatilité est alors un processus de Cox-Ingersoll-Ross.

En notant

$$\begin{aligned} U_t &= \alpha \int_0^t e^{-\lambda(t-s)} ds, \\ u_t &= v \sqrt{\frac{1 - e^{-2\lambda t}}{2\lambda}}, \\ w_t &= \frac{v}{u_t} \int_0^t e^{-\lambda(t-s)} d\hat{Z}_s, \end{aligned}$$

on a les solutions explicites suivantes :

Modèle	Solution	Terme aléatoire
Hull et White	$\sigma_t = \sigma_0 \exp((2\mu - v^2)t + 2v\tilde{Z}_t)$	$\tilde{Z}_t \sim \mathcal{N}(0, t)$
Stein et Stein	$\sigma_t = \sigma_0 e^{-\lambda t} + U_t + u_t w_t$	$w_t \sim \mathcal{N}(0, 1)$
Heston	$\sigma_t = (\sqrt{\sigma_0} e^{-\lambda t} + u_t w_t)^2$	$w_t \sim \mathcal{N}(0, 1)$

En particulier, dans le modèle d'Heston, la volatilité normalisée  $\sigma_t/u_t^2$  suit une loi du  $\chi^2$  décentrée à un degré de liberté, de paramètre de décentrage

$$\gamma_t = \frac{2\lambda e^{-2\lambda t} \sigma_0}{v^2(1 - e^{-2\lambda t})}.$$

Les modèles de Stein et Stein et d'Heston ont la propriété remarquable et très appréciée des praticiens de la finance de déboucher sur des formules analytiques (on dit aussi "formules fermées") d'estimation de produits optionnels, au moyen d'une transformation de Fourier inverse.

### Le modèle SABR

Le modèle SABR, pour "Stochastic  $\alpha\beta\rho$ ", a été analysé par P. HAGAN *et al.* [43]. Il fournit une formule fermée - en réalité un développement limité en "vol de

vol” petite - pour les prix des options vanille, ce qui lui vaut une grande popularité chez les praticiens de la finance. Il suppose la dynamique suivante :

$$\begin{cases} dX_t = \alpha_t X_t^\beta dW_t, \\ d\alpha_t = v\alpha_t d\hat{Z}_t. \end{cases} \quad (3.10)$$

Le processus  $\alpha$  ressemble à une volatilité, il est lié à la volatilité implicite du call à la monnaie, c’est-à-dire du call de prix d’exercice la valeur actualisée en  $T$  de l’actif sous-jacent, appelée volatilité à la monnaie. Les paramètres sont  $v$ ,  $\rho$  et  $\beta$  :

- $v \in \mathbb{R}_+^*$ , la “vol de la vol”, décrit les variations de la simili-volatilité  $\alpha$ , en particulier le fait que celle-ci “a des hauts et des bas”,
- $\rho \in ]-1, 1[$  est la corrélation entre les browniens  $W$  et  $\hat{Z}$ ,
- $\beta \in [0, 1]$  décrit le lien entre le sous-jacent et la volatilité à la monnaie. Lorsque  $\beta \simeq 1$ , la volatilité à la monnaie reste stable quand le cours du sous-jacent varie. Lorsque  $\beta \ll 1$ , la volatilité à la monnaie bouge en direction opposée au marché.

P. HAGAN *et al.* [43] montrent que les prix des options vanille de strike  $K$  et de maturité  $T$  peuvent être calculées approximativement à partir de la formule de Black-Scholes à condition de l’utiliser avec le paramètre de volatilité suivant :

$$\sigma(K, x) = \frac{\alpha_0 \left( 1 + \left( \frac{\alpha_0^2(1-\beta)^2}{24(Kx)^{1-\beta}} + \frac{\rho\beta v\alpha_0}{4(Kx)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}v^2 \right) T \right)}{(Kx)^{(1-\beta)/2} \left( 1 + \frac{(1-\beta)^2}{24} \ln^2 \frac{x}{K} + \frac{(1-\beta)^4}{1920} \ln^4 \frac{x}{K} \right)} \chi(z)$$

où  $x = X_0$  et

$$\begin{cases} z = \frac{v}{\alpha_0} (Kx)^{(1-\beta)/2} \ln \frac{x}{K}, \\ \chi(z) = \ln \left( \frac{\sqrt{1-2\rho z+z^2+z-\rho}}{1-\rho} \right). \end{cases}$$

Les prix peuvent donc être calculés très rapidement. De plus, cette formule de volatilité implicite offre un moyen de calibration des paramètres. Ces caractéristiques rendent ce modèle très populaire chez les “quants”.

### 3.1.4 A quoi s’attendre ?

Il y a une bonne raison *a priori* de considérer la volatilité comme une quantité aléatoire : des études empiriques sur les rendements du cours du sous-jacent permettent d’estimer la volatilité et celle-ci semble présenter un comportement stochastique. Mais modéliser la volatilité par un processus stochastique, c’est en fait reconnaître que quantifier le risque à travers un paramètre de volatilité *constant* est aujourd’hui insuffisant pour expliquer certains phénomènes de marché. En particulier pour expliquer la courbe de smile. Et c’est une modification profonde et puissante qui permet de décrire un marché bien plus complexe que le marché de BLACK et SCHOLLES :

- Nous pouvons reproduire des lois plus réalistes pour les rendements. En particulier, les queues de ces distributions sont plus épaisses que celles des lois lognormales.
- Nous pouvons faire apparaître du smile. J. HULL et A. WHITE [53] furent les premiers à faire cette observation, en 1987. Eric RENAULT et Nizar TOUZI [85] l'ont montré dans un cadre général.
- Nous pouvons rendre les distributions des rendements asymétriques en corrélant les deux bruits  $W$  et  $\hat{Z}$ . Cette asymétrie, souvent appelée "skew", se répercute sur le smile. Typiquement, une chute des cours boursiers s'accompagne d'une hausse de l'incertitude, donc de la volatilité.  $W$  et  $\hat{Z}$  sont dans ce cas négativement corrélés et la pente du smile à la monnaie est alors négative.

Evidemment rien n'est gratuit - surtout dans le monde de la finance des marchés - et il faut bien payer quelque part le prix de ces nettes améliorations :

- On ne peut pas observer directement la volatilité. Estimer les paramètres  $\alpha, m, \beta$  du modèle et le niveau actuel de la volatilité sont donc des problèmes difficiles.
- Le marché ainsi modélisé est *incomplet* : lorsqu'on traite une option, on ne peut pas éliminer le risque en gérant un portefeuille contenant du liquide et du sous-jacent. En effet, la variation infinitésimale de la valeur d'un tel portefeuille contient des termes en  $dW_t$  et en  $dZ_t$  que l'on ne peut annuler simultanément.

## 3.2 Prix d'une option européenne

Dans ce paragraphe, on explique comment Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR définissent le prix d'une option européenne dans un tel contexte. On résume successivement l'approche par équation aux dérivées partielles (EDP) et l'approche probabiliste.

### 3.2.1 EDP d'évaluation

Dans cette section, on considère la dynamique (3.1) et on s'intéresse au prix d'une option européenne d'échéance  $T_1$  et de payoff  $h$  ; l'acheteur d'une telle option reçoit  $h(X_{T_1})$  en  $T_1$ . On suppose qu'il existe une fonction  $P^{(T_1)} : \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R} \longrightarrow \mathbb{R}_+$  suffisamment régulière telle que le prix de cette option à la date  $t \in [0, T_1]$  s'écrit  $P^{(T_1)}(t, X_t, Y_t)$ . Il est impossible d'annuler le risque avec seulement l'actif sous-jacent. Aussi allons nous considérer un portefeuille qui contient  $a_t$  unités d'actif risqué,  $b_t$  unités d'actif sans risque et  $c_t$  options européennes d'échéance  $T_2 > T_1$  et de même payoff  $h$ . Nous cherchons  $a_t, b_t$  et  $c_t$  tels que le portefeuille réplique l'option et soit autofinancé. L'hypothèse de réplication correspond à l'égalité

$$P^{(T_1)}(T_1, X_{T_1}, Y_{T_1}) = a_{T_1} X_{T_1} + b_{T_1} e^{rT_1} + c_{T_1} P^{(T_2)}(T_1, X_{T_1}, Y_{T_1}) \quad \mathbb{P} - p.s. \quad (3.11)$$



et l'hypothèse d'autofinancement à

$$dP^{(T_1)}(t, X_t, Y_t) = a_t dX_t + b_t d(e^{rt}) + c_t dP^{(T_2)}(t, X_t, Y_t). \quad (3.12)$$

Par absence d'opportunité d'arbitrage, (3.11) implique que à toute date  $t \leq T_1$  la valeur du portefeuille est égale au prix de l'option, c'est-à-dire que

$$P^{(T_1)}(t, X_t, Y_t) = a_t X_t + b_t e^{rt} + c_t P^{(T_2)}(t, X_t, Y_t) \quad \mathbb{P} - p.s. \quad (3.13)$$

Par la formule d'Itô, on a<sup>1</sup>

$$\begin{aligned} dP^{(T_1)}(t, X_t, Y_t) &= \frac{\partial P^{(T_1)}}{\partial t}(t, X_t, Y_t) dt + \frac{\partial P^{(T_1)}}{\partial x}(t, X_t, Y_t) dX_t \\ &\quad + \frac{\partial P^{(T_1)}}{\partial y}(t, X_t, Y_t) dY_t + \frac{1}{2} \frac{\partial^2 P^{(T_1)}}{\partial x^2}(t, X_t, Y_t) d\langle X \rangle_t \\ &\quad + \frac{\partial^2 P^{(T_1)}}{\partial x \partial y}(t, X_t, Y_t) d\langle X, Y \rangle_t + \frac{1}{2} \frac{\partial^2 P^{(T_1)}}{\partial y^2}(t, X_t, Y_t) d\langle Y \rangle_t \\ &= \mathcal{A}_1 P^{(T_1)}(t, X_t, Y_t) dt \\ &\quad + \frac{\partial P^{(T_1)}}{\partial x}(t, X_t, Y_t) dX_t + \frac{\partial P^{(T_1)}}{\partial y}(t, X_t, Y_t) dY_t \end{aligned}$$

où  $\mathcal{A}_1$  est l'opérateur défini par

$$\mathcal{A}_1 = \frac{\partial}{\partial t} + \frac{1}{2} x^2 f(y)^2 \frac{\partial^2}{\partial x^2} + \rho \beta x f(y) \frac{\partial^2}{\partial x \partial y} + \frac{1}{2} \beta^2 \frac{\partial^2}{\partial y^2}.$$

Mais, d'après l'hypothèse d'autofinancement, la variation infinitésimale de la valeur du portefeuille est aussi égale à

$$\begin{aligned} &dP^{(T_1)}(t, X_t, Y_t) \\ &= a_t dX_t + b_t d(e^{rt}) + c_t dP^{(T_2)}(t, X_t, Y_t) \\ &= a_t dX_t + r b_t e^{rt} dt \\ &\quad + c_t \left( \mathcal{A}_1 P^{(T_2)}(t, X_t, Y_t) dt + \frac{\partial P^{(T_2)}}{\partial x}(t, X_t, Y_t) dX_t + \frac{\partial P^{(T_2)}}{\partial y}(t, X_t, Y_t) dY_t \right) \\ &= (c_t \mathcal{A}_1 P^{(T_2)}(t, X_t, Y_t) + r b_t e^{rt}) dt \\ &\quad + \left( a_t + c_t \frac{\partial P^{(T_2)}}{\partial x}(t, X_t, Y_t) \right) dX_t \\ &\quad + c_t \frac{\partial P^{(T_2)}}{\partial y}(t, X_t, Y_t) dY_t. \end{aligned}$$

---

<sup>1</sup>les égalités faisant intervenir  $X_t$  et  $Y_t$  sont à prendre au sens p.s.

Il n'y a de termes en  $dZ_t$  que dans  $dY_t$ , si bien que l'identification des termes en  $dZ_t$  donne

$$c_t = \frac{\frac{\partial P^{(T_1)}}{\partial y}(t, X_t, Y_t)}{\frac{\partial P^{(T_2)}}{\partial y}(t, X_t, Y_t)}.$$

L'identification des termes en  $dW_t$  donne

$$a_t = \frac{\partial P^{(T_1)}}{\partial x}(t, X_t, Y_t) - c_t \frac{\partial P^{(T_2)}}{\partial x}(t, X_t, Y_t)$$

d'où on déduit

$$b_t = e^{-rt} (P^{(T_1)}(t, X_t, Y_t) - a_t X_t - c_t P^{(T_2)}(t, X_t, Y_t)).$$

Enfin l'identification des termes en  $dt$  donne

$$\begin{aligned} \mathcal{A}_1 P^{(T_1)}(t, X_t, Y_t) + \mu X_t \frac{\partial P^{(T_1)}}{\partial x}(t, X_t, Y_t) + \alpha (m - Y_t) \frac{\partial P^{(T_1)}}{\partial y}(t, X_t, Y_t) \\ = c_t \mathcal{A}_1 P^{(T_2)}(t, X_t, Y_t) + r b_t e^{rt} + \mu \left( a_t + c_t \frac{\partial P^{(T_2)}}{\partial x}(t, X_t, Y_t) \right) X_t \\ + \alpha (m - Y_t) c_t \frac{\partial P^{(T_2)}}{\partial y}(t, X_t, Y_t), \end{aligned}$$

qui s'écrit aussi en remplaçant  $a_t$ ,  $b_t$  et  $c_t$  par leurs expressions :

$$\left( \frac{\partial P^{(T_1)}}{\partial y}(t, X_t, Y_t) \right)^{-1} \mathcal{A}_2 P^{(T_1)}(t, X_t, Y_t) = \left( \frac{\partial P^{(T_2)}}{\partial y}(t, X_t, Y_t) \right)^{-1} \mathcal{A}_2 P^{(T_2)}(t, X_t, Y_t)$$

où

$$\mathcal{A}_2 = \mathcal{A}_1 + r \left( x \frac{\partial}{\partial x} - \cdot \right).$$

Autrement dit, si on définit l'opérateur  $\mathcal{U}$  par

$$\mathcal{U} = \left( \frac{\partial}{\partial y} \right)^{-1} \mathcal{A}_2$$

alors

$$\mathcal{U} P^{(T_1)}(t, X_t, Y_t) = \mathcal{U} P^{(T_2)}(t, X_t, Y_t).$$

Comme le membre de gauche dépend de  $T_1$  mais pas de  $T_2$  et que le membre de droite, lui, dépend de  $T_2$  mais pas de  $T_1$ , les deux membres sont en fait indépendants de  $T_1$  et de  $T_2$ . Moralement, l'opérateur  $\mathcal{U}$  annule la dépendance en l'échéance. Il existe donc une fonction  $\psi : \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R} \rightarrow \mathbb{R}$  telle que quelle que soit son échéance  $T > 0$ , une option de payoff  $h$  a un prix  $P^{(T)}(t, X_t, Y_t)$  qui vérifie

$$\mathcal{U} P^{(T)}(t, X_t, Y_t) = \psi(t, X_t, Y_t).$$

Désormais on considère l'option de payoff  $h$  et d'échéance  $T$ . Son prix  $P(t, X_t, Y_t)$  vérifie l'équation

$$\mathcal{A}_2 P(t, X_t, Y_t) - \psi(t, X_t, Y_t) \frac{\partial P}{\partial y}(t, X_t, Y_t) = 0. \quad (3.14)$$

Pour des raisons qui apparaitront claires bientôt, on introduit les fonctions  $\Lambda$  et  $\gamma$  définies par

$$\psi(t, x, y) = \beta \Lambda(t, x, y) - \alpha(m - y) \quad (3.15)$$

et

$$\Lambda(t, x, y) = \rho \frac{\mu - r}{f(y)} + \sqrt{1 - \rho^2} \gamma(t, x, y). \quad (3.16)$$

Avec ces notations, (3.14) se réécrit, en omettant la dépendance de  $P$  et de ses dérivées en  $(t, X_t, Y_t)$ ,

$$\begin{aligned} \frac{\partial P}{\partial t} + \frac{1}{2} X_t^2 f(Y_t)^2 \frac{\partial^2 P}{\partial x^2} + \rho \beta X_t f(Y_t) \frac{\partial^2 P}{\partial x \partial y} + \frac{1}{2} \beta^2 \frac{\partial^2 P}{\partial y^2} \\ + r \left( X_t \frac{\partial P}{\partial x} - P \right) + \alpha(m - Y_t) \frac{\partial P}{\partial y} - \beta \Lambda \frac{\partial P}{\partial y} = 0. \end{aligned} \quad (3.17)$$

Cette égalité étant vraie p.s., le prix  $P$  est solution de l'équation aux dérivées partielles

$$(\mathcal{L}_{BS}(f(y)) + \mathcal{L}_{OU} + \mathcal{L}_1) P(t, x, y) = 0 \quad (3.18)$$

avec la condition terminale

$$P(T, x, y) = h(x),$$

où

$$\mathcal{L}_{BS}(f(y)) = \frac{\partial}{\partial t} + \frac{1}{2} x^2 f(y)^2 \frac{\partial^2}{\partial x^2} + r \left( x \frac{\partial}{\partial x} - \cdot \right)$$

est l'opérateur Black-Scholes de paramètre de volatilité  $f(y)$ ,

$$\mathcal{L}_{OU} = \frac{1}{2} \beta^2 \frac{\partial^2}{\partial y^2} + \alpha(m - y) \frac{\partial}{\partial y}$$

est le générateur infinitésimal du processus d'Ornstein-Uhlenbeck, et

$$\mathcal{L}_1 = \rho \beta x f(y) \frac{\partial^2}{\partial x \partial y} - \beta \Lambda(t, x, y) \frac{\partial}{\partial y}$$

est un opérateur faisant intervenir la corrélation  $\rho$  d'une part, en facteur de la dérivée croisée, et la fonction  $\Lambda$  d'autre part. Cette dernière est appelée "*prime de risque de volatilité*". Plus précisément, la fonction  $\frac{\mu - r}{f}$  est exactement la prime de risque liée à la première source de bruit  $W$  et  $\gamma$  est exactement la prime de risque liée à la

deuxième source de bruit  $Z$ . En effet, une variation infinitésimale du prix de l'option s'écrit, en utilisant (3.18) et la formule d'Itô :

$$\begin{aligned} & dP(t, X_t, Y_t) \\ &= \left\{ rP + \frac{\mu - r}{f(Y_t)} \left( X_t f(Y_t) \frac{\partial P}{\partial x} + \beta \rho \frac{\partial P}{\partial y} \right) + \gamma(t, X_t, Y_t) \beta \sqrt{1 - \rho^2} \frac{\partial P}{\partial y} \right\} dt \\ &+ \left\{ X_t f(Y_t) \frac{\partial P}{\partial x} + \beta \rho \frac{\partial P}{\partial y} \right\} dW_t + \left\{ \beta \sqrt{1 - \rho^2} \frac{\partial P}{\partial y} \right\} dZ_t. \end{aligned}$$

La fonction  $\Lambda$  agrège les primes de risque liées aux deux sources indépendantes de hasard, à travers (3.16). On comprend maintenant pourquoi il est bienvenu d'écrire la fonction  $\psi$  sous la forme (3.15)-(3.16).

### 3.2.2 Interprétation probabiliste

On cherche ici à donner une interprétation probabiliste du prix  $P(t, x, y)$ . Posons

$$\begin{cases} \theta_t^W = \frac{\mu - r}{f(Y_t)}, \\ \theta_t^Z = \gamma(t, X_t, Y_t), \end{cases}$$

et

$$M_t = \exp \left( - \int_0^t \theta_s^W dW_s - \int_0^t \theta_s^Z dZ_s - \frac{1}{2} \int_0^t \left( (\theta_s^W)^2 + (\theta_s^Z)^2 \right) ds \right).$$

Sous certaines conditions techniques, par exemple sous la condition de Novikov :

$$\mathbb{E} \left[ \exp \left( \frac{1}{2} \int_0^T \left( (\theta_s^W)^2 + (\theta_s^Z)^2 \right) ds \right) \right] < \infty, \quad (3.19)$$

$M$  est une  $\mathbb{P}$ -martingale. On définit alors bien une mesure de probabilité  $\mathbb{P}^{*(\gamma)}$  sur  $(\Omega, \mathcal{F}_T)$  en posant

$$d\mathbb{P}^{*(\gamma)}(\omega) = M_T(\omega) d\mathbb{P}(\omega).$$

Cette nouvelle probabilité est équivalente à  $\mathbb{P}$  et, d'après le théorème de Girsanov, les processus

$$W_t^* = W_t + \int_0^t \theta_s^W ds$$

et

$$Z_t^* = Z_t + \int_0^t \theta_s^Z ds$$

sont deux  $\mathbb{P}^{*(\gamma)}$ -mouvements browniens indépendants. On a

$$\begin{aligned} dX_t &= \mu X_t dt + f(Y_t) X_t dW_t \\ &= \mu X_t dt + f(Y_t) X_t \left( dW_t^* - \frac{\mu - r}{f(Y_t)} dt \right) \\ &= r X_t dt + f(Y_t) X_t dW_t^* \end{aligned}$$

et

$$\begin{aligned}
dY_t &= \alpha(m - Y_t) dt + \beta d\hat{Z}_t \\
&= \alpha(m - Y_t) dt + \beta \left( \rho dW_t + \sqrt{1 - \rho^2} dZ_t \right) \\
&= \alpha(m - Y_t) dt + \beta \rho \left( dW_t^* - \frac{\mu - r}{f(Y_t)} dt \right) \\
&\quad + \beta \sqrt{1 - \rho^2} (dZ_t^* - \gamma(t, X_t, Y_t) dt) \\
&= (\alpha(m - Y_t) - \beta \Lambda(t, X_t, Y_t)) dt + \beta \left( \rho dW_t^* + \sqrt{1 - \rho^2} dZ_t^* \right).
\end{aligned}$$

Si on pose

$$\hat{Z}_t^* = \rho W_t^* + \sqrt{1 - \rho^2} Z_t^*$$

on définit un  $\mathbb{P}^{*(\gamma)}$ -mouvement brownien et on peut réécrire la dynamique (3.1) sous la forme

$$\begin{cases} dX_t = rX_t dt + \sigma_t X_t dW_t^*, \\ \sigma_t = f(Y_t), \\ dY_t = \{\alpha(m - Y_t) - \beta \Lambda(t, X_t, Y_t)\} dt + \beta d\hat{Z}_t^*. \end{cases} \quad (3.20)$$

Sous la probabilité  $\mathbb{P}^{*(\gamma)}$ , le processus des prix actualisés  $(\tilde{X}_t)_{0 \leq t \leq T}$  défini par

$$\tilde{X}_t = e^{-rt} X_t$$

est une martingale locale. On se placera sous les hypothèses qui assurent que c'est en fait une vraie  $\mathbb{P}^{*(\gamma)}$ -martingale. On pourra par exemple supposer que

$$\mathbb{E}^{*(\gamma)} \left[ \int_0^T f(Y_t)^2 X_t^2 dt \right] < \infty$$

ou même seulement

$$\mathbb{E}^{*(\gamma)} \left[ \sqrt{\int_0^T f(Y_t)^2 X_t^2 dt} \right] < \infty.$$

Alors si on évalue l'option européenne de maturité  $T$  et de payoff  $H = h(X_T)$  par

$$V_t = \mathbb{E}^{*(\gamma)} [e^{-r(T-t)} h(X_T) \mid \mathcal{F}_t]$$

on supprime toute possibilité d'arbitrage. L'énorme inconvénient de ce modèle à volatilité stochastique, lorsqu'on le compare au modèle de Black-Scholes, c'est qu'à chaque fonction  $\gamma(t, x, y)$  correspond une probabilité risque neutre  $\mathbb{P}^{*(\gamma)}$ . On peut adopter le point de vue suivant : le marché sélectionne naturellement une prime de risque de volatilité  $\gamma$  qu'il s'agit de mesurer en étudiant l'historique des données et/ou l'ensemble des prix d'options cotées sur le marché. Pour ce faire, il sera sans

doute raisonnable de supposer d'abord que  $\gamma$  est une constante, puis de considérer que  $\gamma$  ne dépend que de  $y$ , ou éventuellement de  $t$  et de  $y$ . Dans ces derniers cas en effet, la dynamique de  $Y$  reste autonome, au sens où la dynamique de  $X$  n'interfère pas avec celle de  $Y$ . C'est sans doute un problème difficile, et qui ne sera pas traité dans la suite de cette étude. Retenons qu'il existe *a priori* une infinité de  $\gamma$  possibles, auxquels correspondent une infinité de probabilités risque-neutre équivalentes  $\mathbb{P}^{*(\gamma)}$ . Cette propriété est caractéristique de l'*incomplétude* du marché.

Il semble cependant raisonnable, étant donnée l'interprétation financière de  $\Lambda$ , de ne considérer que des fonctions  $y \mapsto \Lambda(y)$  bornées. Au vu de (3.20), la prime de risque  $\Lambda$  n'intervient que dans le terme de drift de  $Y$  où elle s'ajoute au terme  $\alpha(m - Y_t)$  qui lui n'est pas borné. Il est donc légitime de penser que  $\Lambda$  ne joue qu'au "second ordre". Ainsi, lors des simulations numériques, on ne considérera que le cas où  $\Lambda = 0$ .

### 3.3 Analyse asymptotique

L'idée principale de Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR [29] est de considérer :

- d'une part que la volatilité possède la propriété de retour à la moyenne, qu'on modélise par la force de rappel déterministe  $\alpha(m - Y_t) dt$ ,
- d'autre part que ce retour à la moyenne est *rapide*. On suppose donc que l'intensité  $\alpha$  de la force de rappel est grande. Grande devant quoi?  $\alpha$  est l'inverse d'un temps. Il s'agit donc de comparer  $\varepsilon = 1/\alpha$  - temps caractéristique de retour à la moyenne - à l'échelle de temps du problème :  $T - t$ . Aussi on considérera que  $\varepsilon \ll T - t$  ou, de manière équivalente, que  $\alpha \gg (T - t)^{-1}$ .

Les auteurs proposent alors un développement limité en  $\sqrt{\varepsilon}$  du prix de l'option, dans la situation asymptotique où  $\varepsilon \rightarrow 0$  et  $\nu^2$  est constant. Dans cette section, nous décrivons brièvement leur démarche et rappelons leurs résultats, en particulier les développements limités qu'ils obtiennent du prix (paragraphe 3.3.2) et de différentes stratégies de couverture (paragraphe 3.3.3). Revenons d'abord sur la propriété de retour à la moyenne à travers l'exemple du processus d'Ornstein-Uhlenbeck.

#### 3.3.1 Retour à la moyenne

On se propose ici d'étudier plus en détail le processus d'Ornstein-Uhlenbeck  $Y$  de dynamique

$$dY_t = \alpha(m - Y_t) dt + \beta d\hat{Z}_t, \quad Y_0 = y.$$

On a une expression explicite pour  $Y_t$  :

$$Y_t = m + (y - m)e^{-\alpha t} + \beta \int_0^t e^{-\alpha(t-s)} d\hat{Z}_s$$

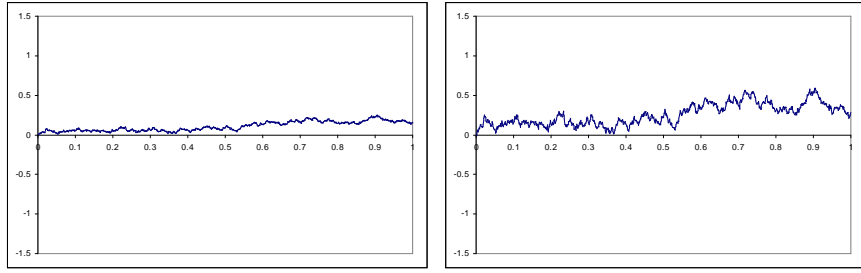


FIG. 3.7 – Une trajectoire du processus  $Y$  ;  $\alpha = 0.1$  (gauche) ou  $1$  (droite)

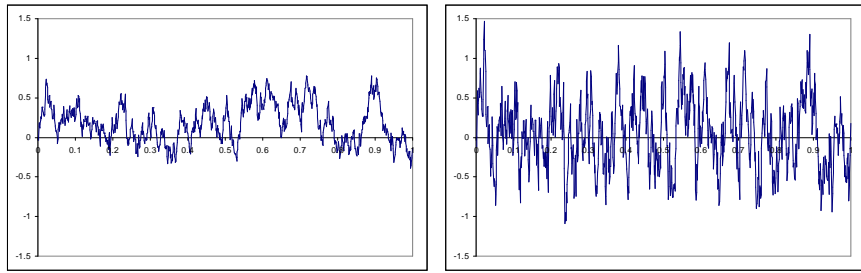


FIG. 3.8 – Une trajectoire du processus  $Y$  ;  $\alpha = 10$  (gauche) ou  $100$  (droite)

qui prouve que  $Y_t$  suit la loi gaussienne de moyenne  $m + (y - m)e^{-\alpha t}$  et de variance  $\nu^2(1 - e^{-2\alpha t})$ , où  $\nu^2 = \frac{\beta^2}{2\alpha}$  ;  $Y_t$  converge en loi lorsque  $t \rightarrow +\infty$  vers  $\pi = \mathcal{N}(m, \nu^2)$ , la loi gaussienne de moyenne  $m$  et de variance  $\nu^2$ . Les figures 3.7 et 3.8 représentent des trajectoires du processus  $(Y_t)_{0 \leq t \leq 1}$  pour  $m = 0$ ,  $\nu = 0.5$ ,  $y = 0$ . On teste plusieurs valeurs de  $\alpha$  :  $0.1$ ,  $1$ ,  $10$ ,  $100$ .

À  $\nu^2$  constant, comme la loi de  $Y_t$  ne dépend de la force de rappel  $\alpha$  et du temps  $t$  que via le produit  $\alpha t$ , sa limite en loi lorsque  $t \rightarrow +\infty$  s'interprète aussi, à horizon  $T$  fini, comme une limite lorsque la force de rappel croît indéfiniment.

### Probabilité stationnaire

Cette loi limite  $\pi$  est aussi la loi stationnaire du processus  $Y$  : si  $Y_0$  suit la loi  $\mathcal{N}(m, \nu^2)$ , alors à toute date  $t \geq 0$  la variable aléatoire  $Y_t$  suit aussi la loi  $\mathcal{N}(m, \nu^2)$ . Sa densité  $\Phi$  est donnée par

$$\Phi(u) = \frac{1}{\sqrt{2\pi\nu}} \exp\left(-\frac{(u - m)^2}{2\nu^2}\right)$$

et vérifie l'équation

$$\mathcal{L}_{OU}^* \Phi = 0$$

où

$$\mathcal{L}_{OU}^* = -\alpha \frac{\partial}{\partial y} ((m - y) \cdot) + \frac{1}{2} \beta^2 \frac{\partial^2}{\partial y^2}$$

est l'adjoint de  $\mathcal{L}_{OU}$ , le générateur infinitésimal de la diffusion  $Y$ .

### Propriété de décorrélation, théorème ergodique

L'inverse  $\varepsilon$  de l'intensité  $\alpha$  de la force de rappel s'interprète aussi comme le temps caractéristique de décorrélation du processus d'Ornstein-Uhlenbeck, puisque, si  $s \leq t$ ,

$$\begin{aligned}
\text{cov}(Y_s, Y_t) &= \text{cov} \left( \beta \int_0^s e^{-\alpha(s-u)} d\hat{Z}_u, \beta \int_0^t e^{-\alpha(t-v)} d\hat{Z}_v \right) \\
&= \beta^2 e^{-\alpha(t+s)} \mathbb{E} \left[ \int_0^s e^{\alpha u} d\hat{Z}_u \int_0^t e^{\alpha v} d\hat{Z}_v \right] \\
&= \beta^2 e^{-\alpha(t+s)} \mathbb{E} \left[ \int_0^s e^{\alpha u} d\hat{Z}_u \int_0^s e^{\alpha v} d\hat{Z}_v \right] \\
&= \beta^2 e^{-\alpha(t+s)} \int_0^s e^{2\alpha u} du \\
&= \nu^2 (e^{-\alpha(t-s)} - e^{-\alpha(t+s)}).
\end{aligned}$$

Par conséquent, si  $s$  et  $t$  tendent vers  $+\infty$  de sorte que  $\Delta = |t - s|$  reste constant, la covariance limite de  $Y_s$  et  $Y_t$  vaut  $\nu^2 e^{-\alpha\Delta}$ . Notons que c'est exactement la covariance de  $Y_s$  et  $Y_t$  sous la loi stationnaire. A la limite, lorsque ce temps typique de décorrélation  $\varepsilon$  est infiniment petit, i.e. lorsque  $\alpha$  est infiniment grand, les valeurs  $Y_s$  et  $Y_t$ , même pour des temps voisins  $s$  et  $t$ , sont indépendantes (leur covariance est nulle et le processus  $Y$  est gaussien). C'est pour cette raison qu'on va disposer de théorèmes ergodiques du type loi forte des grands nombres. Plus précisément, pour toute fonction  $g$  intégrable contre la mesure stationnaire  $\mathcal{N}(m, \nu^2)$ , p.s.

$$\lim_{\alpha \rightarrow +\infty} \frac{1}{T-t} \int_t^T g(Y_s) ds = \langle g \rangle \tag{3.21}$$

où, par définition,  $\langle g \rangle$  est l'espérance de la fonction  $g$  contre la mesure stationnaire, i.e.

$$\langle g \rangle = \int_{-\infty}^{+\infty} g(u) \Phi(u) du.$$

En pratique, l'approximation

$$\frac{1}{T-t} \int_t^T g(Y_s) ds \approx \langle g \rangle$$

ne sera valable que si  $\alpha \gg \frac{1}{T-t}$ . Dans le contexte des marchés financiers, cela signifie que nous ne pourrons faire cette approximation que si nous sommes suffisamment loin de l'échéance  $T$  de l'option traitée.



### 3.3.2 Le prix Black-Scholes corrigé

#### Notations

On se place ici sous l'hypothèse  $\varepsilon \ll T - t$ . Moralement, sous cette hypothèse,  $Y_t$  atteint sa loi limite  $\mathcal{N}(m, \nu^2)$  en temps fini. On se placera donc dans l'asymptotique

$$\varepsilon \longrightarrow 0, \quad \nu^2 = \text{cte.}$$

Comme  $\nu^2 = \frac{\beta^2}{2\alpha}$ , cela signifie que  $\beta = \nu\sqrt{2\alpha} = \frac{\nu\sqrt{2}}{\sqrt{\varepsilon}}$  tend vers  $+\infty$ . On peut réécrire la dynamique (3.20) en faisant apparaître le paramètre infiniment petit  $\varepsilon$  :

$$\begin{cases} dX_t^\varepsilon = rX_t^\varepsilon dt + \sigma_t X_t^\varepsilon dW_t^*, \\ \sigma_t = f(Y_t^\varepsilon), \\ dY_t^\varepsilon = \left\{ \frac{1}{\varepsilon}(m - Y_t^\varepsilon) - \frac{\nu\sqrt{2}}{\sqrt{\varepsilon}}\Lambda(t, X_t^\varepsilon, Y_t^\varepsilon) \right\} dt + \frac{\nu\sqrt{2}}{\sqrt{\varepsilon}} d\hat{Z}_t^*. \end{cases}$$

Le prix  $P^\varepsilon$  de l'option européenne de maturité  $T$  et de payoff  $H = h(X_T^\varepsilon) \in L^2(\mathbb{P}^*, \mathcal{F}_T)$  vérifie alors (cf (3.17))

$$\begin{aligned} \frac{\partial P^\varepsilon}{\partial t} + \frac{1}{2}(X_t^\varepsilon)^2 f(Y_t^\varepsilon)^2 \frac{\partial^2 P^\varepsilon}{\partial x^2} + \rho \frac{\nu\sqrt{2}}{\sqrt{\varepsilon}} X_t^\varepsilon f(Y_t^\varepsilon) \frac{\partial^2 P^\varepsilon}{\partial x \partial y} + \frac{\nu^2}{\varepsilon} \frac{\partial^2 P^\varepsilon}{\partial y^2} \\ + r \left( X_t^\varepsilon \frac{\partial P^\varepsilon}{\partial x} - P^\varepsilon \right) + \frac{1}{\varepsilon}(m - Y_t^\varepsilon) \frac{\partial P^\varepsilon}{\partial y} - \frac{\nu\sqrt{2}}{\sqrt{\varepsilon}} \Lambda \frac{\partial P^\varepsilon}{\partial y} = 0 \end{aligned}$$

et est donc solution de l'EDP

$$\left( \frac{1}{\varepsilon} \mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}} \mathcal{L}_1 + \mathcal{L}_2 \right) P^\varepsilon(t, x, y) = 0 \quad (3.22)$$

avec la condition terminale

$$P^\varepsilon(T, x, y) = h(x),$$

où

$$\mathcal{L}_2 = \mathcal{L}_{BS}(f(y)) = \frac{\partial}{\partial t} + \frac{1}{2}x^2 f(y)^2 \frac{\partial^2}{\partial x^2} + r \left( x \frac{\partial}{\partial x} - \cdot \right)$$

est l'opérateur Black-Scholes de paramètre de volatilité  $f(y)$ ,

$$\mathcal{L}_0 = \varepsilon \mathcal{L}_{OU} = \nu^2 \frac{\partial^2}{\partial y^2} + (m - y) \frac{\partial}{\partial y}$$

est le générateur infinitésimal du processus d'Ornstein-Uhlenbeck multiplié par  $\varepsilon$ , et

$$\mathcal{L}_1 = \rho\nu\sqrt{2}xf(y) \frac{\partial^2}{\partial x \partial y} - \nu\sqrt{2}\Lambda \frac{\partial}{\partial y}$$

fait intervenir la corrélation entre volatilité et sous-jacent d'une part et le risque de volatilité d'autre part. *Dans toute la suite, on fait l'hypothèse que la prime de risque  $\Lambda$  est une fonction continue **bornée ne dépendant que de  $y$** .*

## Le problème à résoudre

Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR se placent dans les conditions où, pour tout  $\varepsilon > 0$ , l'EDP (3.22) a une unique solution, montrent que cette solution  $P^\varepsilon$  a une limite quand  $\varepsilon$  tend vers 0 et calculent la correction d'ordre 1. Pour ce faire, ils supposent l'existence d'un développement en série de la forme

$$P^\varepsilon = P_0 + \sqrt{\varepsilon}P_1 + \varepsilon P_2 + \varepsilon\sqrt{\varepsilon}P_3 + \varepsilon^2 P_4 + \dots$$

(3.22) se réécrit alors formellement

$$\begin{aligned} 0 &= \frac{1}{\varepsilon}\mathcal{L}_0 P_0 \\ &+ \frac{1}{\sqrt{\varepsilon}}(\mathcal{L}_0 P_1 + \mathcal{L}_1 P_0) \\ &+ (\mathcal{L}_0 P_2 + \mathcal{L}_1 P_1 + \mathcal{L}_2 P_0) \\ &+ \sqrt{\varepsilon}(\mathcal{L}_0 P_3 + \mathcal{L}_1 P_2 + \mathcal{L}_2 P_1) \\ &+ \dots \end{aligned}$$

Il s'agit donc de chercher des fonctions  $P_i : \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R} \rightarrow \mathbb{R}$  suffisamment régulières qui vérifient :

$$\begin{cases} \mathcal{L}_0 P_0 = 0, & P_0(T, x, y) = h(x), \\ \mathcal{L}_0 P_1 + \mathcal{L}_1 P_0 = 0, & P_1(T, x, y) = 0, \\ \mathcal{L}_0 P_n + \mathcal{L}_1 P_{n-1} + \mathcal{L}_2 P_{n-2} = 0, & P_n(T, x, y) = 0, \quad \forall n \geq 2. \end{cases}$$

En fait, on se limitera à la recherche des fonctions  $P_0$  et  $P_1$ . Elles permettent d'obtenir le prix corrigé à l'ordre 1 :  $P_0 + \sqrt{\varepsilon}P_1$ .

## La solution

Les équations  $\mathcal{L}_0 P_0 = 0$  et  $\mathcal{L}_0 P_1 + \mathcal{L}_1 P_0 = 0$  impliquent successivement que les fonctions  $P_0$  et  $P_1$ , et donc le prix corrigé à l'ordre 1,  $P_0 + \sqrt{\varepsilon}P_1$ , ne dépendent pas de  $y$ . C'est une propriété remarquable : à l'ordre 1, il n'est pas besoin de connaître le niveau actuel de la volatilité pour déterminer le prix de l'option ; la date  $t$  de la transaction et le niveau actuel  $x$  du cours du sous-jacent suffisent. C'est appréciable car la volatilité n'est pas une quantité directement observable.

Comme  $P_1$  ne dépend pas de  $y$ ,  $\mathcal{L}_1 P_1 = 0$  ; l'équation  $\mathcal{L}_0 P_2 + \mathcal{L}_1 P_1 + \mathcal{L}_2 P_0 = 0$  se réduit donc à  $\mathcal{L}_0 P_2 + \mathcal{L}_2 P_0 = 0$ . Or, en notant  $\pi = \mathcal{N}(m, \nu^2) = \Phi(u)$  du la probabilité

invariante de  $Y$ , comme  $\mathcal{L}_0^* \Phi = 0$ , on a

$$\begin{aligned} \langle \mathcal{L}_0 P_2 \rangle(t, x) &= \int_{-\infty}^{+\infty} (\mathcal{L}_0 P_2)(t, x, u) \Phi(u) du \\ &= \int_{-\infty}^{+\infty} P_2(t, x, u) (\mathcal{L}_0^* \Phi(u)) du \\ &= 0. \end{aligned}$$

Par conséquent,  $\langle \mathcal{L}_2 P_0 \rangle = 0$ , i.e.  $\mathcal{L}_{BS}(\bar{\sigma}) P_0 = 0$ , où

$$\bar{\sigma}^2 = \int_{-\infty}^{+\infty} f^2(u) \Phi(u) du = \langle f^2 \rangle.$$

A l'ordre 0, le prix de l'option est donc le prix donné par le modèle de Black-Scholes, utilisé avec le paramètre de volatilité constant  $\bar{\sigma}$ , moyenne ergodique de la volatilité stochastique.

Le calcul de  $P_1$ , plus complexe, utilise la solution  $P_2$  de l'équation dite de Poisson

$$\mathcal{L}_0 P_2 = -\mathcal{L}_2 P_0,$$

puis la moyenne stationnaire de l'équation  $\mathcal{L}_0 P_3 + \mathcal{L}_1 P_2 + \mathcal{L}_2 P_1 = 0$ , qui s'écrit  $\mathcal{L}_{BS}(\bar{\sigma}) P_1 = -\langle \mathcal{L}_1 P_2 \rangle$ . Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR obtiennent finalement la correction du prix  $\tilde{P}_1 = \sqrt{\varepsilon} P_1$  à l'ordre 1 sous la forme

$$\tilde{P}_1(t, x) = -(T - t) \left( V_2 x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x) + V_3 x^3 \frac{\partial^3 P_0}{\partial x^3}(t, x) \right), \quad (3.23)$$

où

$$\begin{aligned} V_2 &= \frac{\nu}{\sqrt{2\alpha}} (2\rho \langle f \phi' \rangle - \langle \Lambda \phi' \rangle), \\ V_3 &= \frac{\rho \nu}{\sqrt{2\alpha}} \langle f \phi' \rangle, \\ \phi'(y) &= \frac{1}{\nu^2 \Phi(y)} \left( \int_{-\infty}^y (f(z)^2 - \langle f^2 \rangle) \Phi(z) dz \right). \end{aligned} \quad (3.24)$$

Notons que, si on note  $L_2(t, x) = x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x)$  et  $L_3(t, x) = x^3 \frac{\partial^3 P_0}{\partial x^3}(t, x)$ , on peut décomposer la correction du prix au premier ordre en

- une partie de *corrélation* pure, en facteur de la combinaison linéaire  $2L_2 + L_3$ ,
- et une partie de *prime de risque de volatilité* pure, en facteur de  $L_2$ .

En effet,

$$\tilde{P}_1(t, x) = -(T - t) \left( \frac{\rho \nu}{\sqrt{2\alpha}} \langle f \phi' \rangle (2L_2 + L_3)(t, x) - \frac{\nu}{\sqrt{2\alpha}} \langle \Lambda \phi' \rangle L_2(t, x) \right).$$

On s'aperçoit que pour calculer le prix corrigé à l'ordre 1, il suffit de connaître les trois quantités  $\bar{\sigma}$ ,  $V_2$  et  $V_3$ . Dans la situation où le retour à la moyenne est très fort, il suffit donc de calibrer ces trois paramètres qui agrègent  $\alpha, m, \beta, \rho, f, \gamma$  et  $\mu$ . Nous traiterons ce point plus loin au paragraphe 3.4.

Dans le cas du call, sous hypothèse d'ellipticité ( $f(y) \geq \eta > 0$ ) et de bornitude de  $f$  et de la prime de risque  $\gamma$ , Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR [31] prouvent, par une technique de régularisation du payoff, que  $\tilde{P}_1$  est en effet une approximation du prix d'ordre "presque"  $\varepsilon$ . Précisément, ils montrent que pour tous  $t < T$ ,  $x > 0$  et  $y \in \mathbb{R}$ ,  $P^\varepsilon(t, x, y) = \tilde{P}_1(t, x) + o(\varepsilon |\log \varepsilon|^{1+p})$  pour tout  $p > 0$ .

### 3.3.3 Stratégies de couverture

Rappelons que dans le modèle de Black-Scholes, pour la dynamique suivante du sous-jacent :

$$dX_t = \mu X_t dt + \bar{\sigma} X_t dW_t,$$

une option européenne qui paye  $h(X_T) \in L^2(\mathbb{P}^*)$  vaut  $P_0(t, X_t)$  à la date  $t$  et est parfaitement couverte par le portefeuille autofinancé qui contient

$$a_t = \frac{\partial P_0}{\partial x}(t, X_t)$$

unités de sous-jacent et

$$b_t = e^{-rt} \left( P_0(t, X_t) - X_t \frac{\partial P_0}{\partial x}(t, X_t) \right)$$

unités monétaires (on parle alors de couverture en delta neutre). En effet :

- Ce portefeuille *réplique* l'option : à toute date  $t$ , la valeur de ce portefeuille est  $P_0(t, X_t)$ . En particulier, à maturité, sa valeur est  $P_0(T, X_T) = h(X_T)$ .
- Ce portefeuille est *autofinancé* : la formule d'Itô donne

$$dP_0(t, X_t) = \frac{\partial P_0}{\partial x}(t, X_t) dX_t + \left( \frac{\partial P_0}{\partial t}(t, X_t) + \frac{1}{2} \bar{\sigma}^2 X_t^2 \frac{\partial^2 P_0}{\partial x^2}(t, X_t) \right) dt,$$

et comme  $P_0$  est solution de l'EDP Black-Scholes

$$\frac{\partial P_0}{\partial t} + \frac{\sigma^2}{2} x^2 \frac{\partial^2 P_0}{\partial x^2} + r \left( x \frac{\partial P_0}{\partial x} - P_0 \right) = 0, \quad (3.25)$$

on a

$$dP_0(t, X_t) = \frac{\partial P_0}{\partial x}(t, X_t) dX_t + r \left( P_0(t, X_t) - X_t \frac{\partial P_0}{\partial x}(t, X_t) \right) dt,$$

i.e.

$$dP_0(t, X_t) = a_t dX_t + b_t d(e^{rt}).$$

Autrement dit, la variation infinitésimale  $dP_0(t, X_t)$  de la valeur de ce portefeuille est exactement la variation due au marché.

Dans le cas du modèle à volatilité stochastique (3.1)

$$\begin{cases} dX_t = \mu X_t dt + f(Y_t) X_t dW_t, \\ dY_t = \alpha(m - Y_t) dt + \beta d\hat{Z}_t, \end{cases}$$

le marché est incomplet, il n'est pas possible d'éliminer le risque en gérant un portefeuille ne contenant que du liquide et du sous-jacent, et il s'agit de transiger entre les pertes éventuelles dues à une mauvaise couverture et le coût de la couverture. On mesure sous la probabilité subjective  $\mathbb{P}$  les performances statistiques d'une stratégie. Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR proposent trois stratégies de couverture dont on détaille brièvement les propriétés ci-dessous.

### Une couverture en delta neutre

On décide de suivre la même stratégie de gestion de portefeuille que dans le cas Black-Scholes avec volatilité  $\bar{\sigma} = \sqrt{\langle f^2 \rangle}$ , c'est-à-dire qu'on choisit

$$\begin{cases} a_t^{(1)} = \frac{\partial P_0}{\partial x}(t, X_t), \\ b_t^{(1)} = e^{-rt} \left( P_0(t, X_t) - X_t \frac{\partial P_0}{\partial x}(t, X_t) \right). \end{cases}$$

**Coût exact** A tout instant, la valeur de ce portefeuille est  $a_t^{(1)} X_t + b_t^{(1)} e^{rt} = P_0(t, X_t)$ . Comme  $P_0(T, x) = h(x)$ , la valeur finale est  $h(X_T)$  donc ce portefeuille *réplique* l'option. En revanche, il *n'est pas* autofinancé. En effet,

$$dP_0(t, X_t) = a_t^{(1)} dX_t + \left( \frac{\partial P_0}{\partial t}(t, X_t) + \frac{1}{2} f(Y_t)^2 X_t^2 \frac{\partial^2 P_0}{\partial x^2}(t, X_t) \right) dt$$

tandis que la variation due au marché vaut

$$a_t^{(1)} dX_t + b_t^{(1)} d(e^{rt}) = a_t^{(1)} dX_t + r \left( P_0(t, X_t) - X_t \frac{\partial P_0}{\partial x}(t, X_t) \right) dt,$$

si bien que la différence, qu'on interprète comme le *coût infinitésimal* de la couverture, n'est pas nulle et vaut exactement, en utilisant (3.25) :

$$\begin{aligned} dC_t^{(1)} & \tag{3.26} \\ &= dP_0(t, X_t) - \left( a_t^{(1)} dX_t + b_t^{(1)} d(e^{rt}) \right) \\ &= \left( \frac{\partial P_0}{\partial t}(t, X_t) + \frac{1}{2} f(Y_t)^2 X_t^2 \frac{\partial^2 P_0}{\partial x^2}(t, X_t) - r \left( P_0(t, X_t) - X_t \frac{\partial P_0}{\partial x}(t, X_t) \right) \right) dt \\ &= \frac{1}{2} (f(Y_t)^2 - \bar{\sigma}^2) X_t^2 \frac{\partial^2 P_0}{\partial x^2}(t, X_t) dt. \end{aligned}$$

Si  $dC_t^{(1)} > 0$ , on doit injecter de l'argent dans le portefeuille; si  $dC_t^{(1)} < 0$ , on doit en retirer. La formule (3.26) est remarquable car *elle est valable trajectoire par trajectoire*; c'est une formule exacte, valable  $\omega$  par  $\omega$ . Remarquons tout de suite que le coût de gestion du portefeuille est :

- un processus stochastique à *variation finie*,
- d'autant plus petit que
  - la volatilité moyenne  $\bar{\sigma}$  est proche de la vraie volatilité  $f(Y_t)$ ,
  - la convexité de  $P_0$  est petite, i.e.  $x \mapsto P_0(t, x)$  a un profil plat.

Le vendeur de l'option touche  $P_0(0, X_0) + \tilde{P}_1(0, X_0)$  à la date  $t = 0$ ;  $\tilde{P}_1(0, X_0)$  est de signe quelconque, comme le montre (3.23). Le vendeur investit  $P_0(0, X_0)$  dans le portefeuille, répartis en  $a_0$  unités d'actif risqué et  $b_0$  unités monétaires; après cet investissement initial, la gestion dynamique du portefeuille de réplication  $(a_t^{(1)}, b_t^{(1)})$  a un coût aléatoire

$$C_t^{(1)} = \int_0^t dC_s^{(1)} = \frac{1}{2} \int_0^t (f(Y_s)^2 - \bar{\sigma}^2) X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) ds.$$

De plus, la richesse  $\tilde{P}_1(0, X_0)$  est placée - si elle est positive - ou empruntée - si elle est négative - à la banque.

**Effet de moyenne** Heuristiquement, l'ergodicité du processus  $Y$  traduit un effet de moyenne et permet d'identifier, dans la limite où  $\alpha$  est grand, l'intégrale

$$\int_0^t f(Y_s)^2 X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) ds$$

et l'intégrale

$$\int_0^t \bar{\sigma}^2 X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) ds$$

et par conséquent le coût  $C_t^{(1)}$ , qui est la différence des deux intégrales, est petit. Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR en donnent même un équivalent quand  $\alpha$  croît :

$$C_t^{(1)} \simeq \frac{1}{\sqrt{\alpha}} (B_t + M_t)$$

où  $(B_t)_{t \geq 0}$  est un processus à variation finie et  $(M_t)_{t \geq 0}$  une  $\mathbb{P}$ -martingale de moyenne nulle, tous deux p.s. bornés en  $\alpha$ . Précisément,

$$\begin{aligned} M_t &= -\frac{\nu}{\sqrt{2}} \int_0^t X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) \phi'(Y_s) d\hat{Z}_s, \\ B_t &= -\frac{\rho\nu}{\sqrt{2}} \int_0^t f(Y_s) \phi'(Y_s) \left( 2X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) + X_s^3 \frac{\partial^3 P_0}{\partial x^3}(s, X_s) \right) ds. \end{aligned} \quad (3.27)$$

La preuve est basée sur le fait que toute primitive  $\phi$  de (3.24) vérifie  $\mathcal{L}_0\phi = f^2 - \bar{\sigma}^2$ . Ainsi,  $(f(Y_s)^2 - \bar{\sigma}^2) ds = (\mathcal{L}_0\phi)(Y_s) ds$  et comme, par la formule d'Itô,

$$d\phi(Y_s) = \alpha (\mathcal{L}_0\phi)(Y_s) ds + \nu\sqrt{2\alpha}\phi'(Y_s) d\hat{Z}_s,$$

on a

$$(f(Y_s)^2 - \bar{\sigma}^2) ds = \frac{1}{\alpha} \left( d\phi(Y_s) - \nu\sqrt{2\alpha}\phi'(Y_s) d\hat{Z}_s \right),$$

de sorte que le coût de la couverture s'écrit

$$C_t^{(1)} = \frac{1}{2\alpha} \left\{ \int_0^t X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) d\phi(Y_s) - \nu\sqrt{2\alpha} \int_0^t X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) \phi'(Y_s) d\hat{Z}_s \right\}.$$

Le reste de la démonstration consiste en une application judicieusement répétée de la formule d'Itô, aboutissant à l'égalité

$$\begin{aligned} C_t^{(1)} &= \frac{1}{\sqrt{\alpha}}(B_t + M_t) \\ &+ \frac{1}{2\alpha} \int_0^t d \left( \phi(Y_s) X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) \right) \\ &- \frac{1}{2\alpha} \int_0^t \phi(Y_s) d \left( X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) \right). \end{aligned} \quad (3.28)$$

qui fournit l'équivalent souhaité.

Notons que (3.27) ne permet pas de décider du signe du biais  $B$ . Une bonne stratégie de couverture devrait, à l'ordre 1 au moins, ne rien coûter *en moyenne*. A cette fin, les auteurs proposent de corriger légèrement le portefeuille Black-Scholes.

### Une couverture autofinancée en moyenne

Pour qu'une stratégie ne coûte rien en moyenne au premier ordre, il suffit de "faire apparaître" le terme

$$\frac{\rho\nu}{\sqrt{2\alpha}} \langle f\phi' \rangle \int_0^t \left( 2X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) + X_s^3 \frac{\partial^3 P_0}{\partial x^3}(s, X_s) \right) ds \quad (3.29)$$

qui viendrait s'ajouter aux trois termes (3.28) composant  $C_t^{(1)}$ . En effet, ajouté au terme  $\frac{1}{\sqrt{\alpha}}B_t$ , il génère un terme d'ordre  $\frac{1}{\alpha}$ , à cause de l'effet de moyenne décrit ci-dessus.

Une idée naturelle, au vu de ce qui précède, serait de choisir

$$a_t = \frac{\partial (P_0 + \tilde{P}_1)}{\partial x}(t, X_t),$$

à la place de  $\frac{\partial P_0}{\partial x}(t, X_t)$ . En fait,  $\tilde{P}_1$  est solution de l'EDP

$$\begin{aligned} \mathcal{L}_{BS}(\bar{\sigma})\tilde{P}_1 &= \frac{\rho\nu}{\sqrt{2\alpha}}\langle f\phi' \rangle \left( 2x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x) + x^3 \frac{\partial^3 P_0}{\partial x^3}(t, x) \right) \\ &\quad - \frac{\nu}{\sqrt{2\alpha}}\langle \Lambda\phi' \rangle x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x). \end{aligned} \quad (3.30)$$

Le terme  $\frac{\rho\nu}{\sqrt{2\alpha}}\langle f\phi' \rangle \left( 2x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x) + x^3 \frac{\partial^3 P_0}{\partial x^3}(t, x) \right)$  ressemble beaucoup à ce qu'on voudrait voir apparaître, mais le terme  $-\frac{\nu}{\sqrt{2\alpha}}\langle \Lambda\phi' \rangle x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x)$  est de trop : c'est en réalité un terme de prime de risque de volatilité pure, alors que le biais  $B$  que l'on cherche à compenser est un terme de corrélation uniquement. Aussi est-il utile de définir la fonction  $\tilde{Q}_1 : \mathbb{R}_+ \times \mathbb{R}_+^* \rightarrow \mathbb{R}$  solution de l'EDP

$$\mathcal{L}_{BS}(\bar{\sigma})\tilde{Q}_1 = \frac{\rho\nu}{\sqrt{2\alpha}}\langle f\phi' \rangle \left( 2x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x) + x^3 \frac{\partial^3 P_0}{\partial x^3}(t, x) \right) \quad (3.31)$$

avec la condition terminale

$$\tilde{Q}_1(T, x) = 0 \quad \forall x > 0$$

et de considérer le portefeuille défini par

$$\begin{cases} a_t^{(2)} = \frac{\partial(P_0 + \tilde{Q}_1)}{\partial x}(t, X_t), \\ b_t^{(2)} = e^{-rt} \left( (P_0 + \tilde{Q}_1)(t, X_t) - X_t \frac{\partial(P_0 + \tilde{Q}_1)}{\partial x}(t, X_t) \right). \end{cases}$$

Ce portefeuille vaut  $(P_0 + \tilde{Q}_1)(t, X_t)$  à la date  $t$ ; sa variation infinitésimale est

$$\begin{aligned} &d(P_0 + \tilde{Q}_1)(t, X_t) \\ &= a_t^{(2)} dX_t + \left( \frac{\partial(P_0 + \tilde{Q}_1)}{\partial t}(t, X_t) + \frac{1}{2} f(Y_t)^2 X_t^2 \frac{\partial^2(P_0 + \tilde{Q}_1)}{\partial x^2}(t, X_t) \right) dt, \end{aligned}$$

tandis que la variation due au marché est

$$\begin{aligned} &a_t^{(2)} dX_t + b_t^{(2)} d(e^{rt}) \\ &= a_t^{(2)} dX_t + r \left( (P_0 + \tilde{Q}_1)(t, X_t) - X_t \frac{\partial(P_0 + \tilde{Q}_1)}{\partial x}(t, X_t) \right) dt, \end{aligned}$$



si bien que le coût infinitésimal vérifie

$$\begin{aligned} \frac{dC_t^{(2)}}{dt} &= \frac{\partial (P_0 + \tilde{Q}_1)}{\partial t}(t, X_t) + \frac{1}{2}f(Y_t)^2 X_t^2 \frac{\partial^2 (P_0 + \tilde{Q}_1)}{\partial x^2}(t, X_t) \\ &\quad - r \left( (P_0 + \tilde{Q}_1)(t, X_t) - X_t \frac{\partial (P_0 + \tilde{Q}_1)}{\partial x}(t, X_t) \right) \\ &= \mathcal{L}_{BS}(\bar{\sigma})(P_0 + \tilde{Q}_1)(t, X_t) + \frac{1}{2}(f(Y_t)^2 - \bar{\sigma}^2) X_t^2 \frac{\partial^2 (P_0 + \tilde{Q}_1)}{\partial x^2}(t, X_t). \end{aligned}$$

Etant donné (3.31), et vu que  $\mathcal{L}_{BS}(\bar{\sigma})P_0 = 0$ , on a finalement

$$\begin{aligned} C_t^{(2)} &= C_t^{(1)} + \frac{\rho\nu}{\sqrt{2\alpha}} \langle f\phi' \rangle \int_0^t \left( 2X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) + X_s^3 \frac{\partial^3 P_0}{\partial x^3}(s, X_s) \right) ds \\ &\quad + \frac{1}{2} \int_0^t (f(Y_s)^2 - \bar{\sigma}^2) X_s^2 \frac{\partial^2 \tilde{Q}_1}{\partial x^2}(s, X_s) ds. \end{aligned}$$

Le dernier terme est d'ordre  $\frac{1}{\alpha}$ , puisque l'effet de moyenne joue sur  $X_s^2 \frac{\partial^2 \tilde{Q}_1}{\partial x^2}(s, X_s)$  qui est déjà d'ordre  $\frac{1}{\sqrt{\alpha}}$ , comme  $\tilde{Q}_1$ . Le second terme est précisément le terme voulu (3.29) et finalement, au premier ordre,  $C_t^{(2)} \simeq \frac{1}{\sqrt{\alpha}} M_t$ .

### Une couverture qui colle au prix de l'option

Notons que si l'on suit la stratégie de gestion précédente, les valeurs du portefeuille et de l'option diffèrent dès l'ordre  $\frac{1}{\sqrt{\alpha}}$ , car cette différence est précisément  $\tilde{Q}_1 - \tilde{P}_1$ . Si l'on souhaite que les valeurs du portefeuille et de l'option ne diffèrent qu'à l'ordre  $\frac{1}{\alpha}$ , il suffit d'adopter la stratégie suivante, déjà évoquée plus haut :

$$\begin{cases} a_t^{(3)} = \frac{\partial (P_0 + \tilde{P}_1)}{\partial x}(t, X_t), \\ b_t^{(3)} = e^{-rt} \left( (P_0 + \tilde{P}_1)(t, X_t) - X_t \frac{\partial (P_0 + \tilde{P}_1)}{\partial x}(t, X_t) \right). \end{cases}$$

En effet, à la date  $t$  ce portefeuille vaut  $(P_0 + \tilde{P}_1)(t, X_t)$ , c'est-à-dire le prix de l'option, à  $\frac{1}{\alpha}$  près. La méthode présentée dans la section précédente montre que le coût de la couverture vaut alors

$$\begin{aligned} C_t^{(3)} &= C_t^{(2)} - \frac{\nu}{\sqrt{2\alpha}} \langle \Lambda\phi' \rangle \int_0^t X_s^2 \frac{\partial^2 P_0}{\partial x^2}(s, X_s) ds \\ &\quad + \frac{1}{2} \int_0^t (f(Y_s)^2 - \bar{\sigma}^2) X_s^2 \frac{\partial^2 (\tilde{P}_1 - \tilde{Q}_1)}{\partial x^2}(s, X_s) ds. \end{aligned}$$

Il est d'ordre  $\frac{1}{\sqrt{\alpha}}$ . Le dernier terme est d'ordre  $\frac{1}{\alpha}$ . Le deuxième terme, d'ordre  $\frac{1}{\sqrt{\alpha}}$ , est le prix à fournir pour retrouver, à l'ordre  $\frac{1}{\sqrt{\alpha}}$  du moins, le prix de l'option à partir du coût de la stratégie autofinancée en moyenne. C'est donc le prix que le vendeur de l'option attache à l'incomplétude du marché. Celle-ci est due au caractère stochastique de la volatilité. Le vendeur subit donc un risque de volatilité et le quantifie via  $-\frac{\nu}{\sqrt{2\alpha}}\langle\Lambda\phi'\rangle\int_0^t X_s^2\frac{\partial^2 P_0}{\partial x^2}(s, X_s) ds$  qui est un terme de risque de volatilité uniquement.

### 3.4 Calibration

Le modèle introduit beaucoup de paramètres  $(\alpha, \beta, m, \rho, \gamma, f)$  mais seuls trois suffisent à déterminer le prix corrigé et les stratégies de couverture étudiées plus haut, dans le cas où la force de rappel de la volatilité à sa moyenne est grande :  $\bar{\sigma}, V_2$  et  $V_3$ . Ils sont bien sûr fonctions de  $(\alpha, \beta, m, \rho, \gamma, f)$ . Afin de calibrer le modèle, on pourrait chercher les  $\bar{\sigma}, V_2$  et  $V_3$  tels que le prix corrigé

$$\left(P_0 + \tilde{P}_1\right)(t, x) = P_0(t, x) - (T - t) \left( V_2 x^2 \frac{\partial^2 P_0}{\partial x^2}(t, x) + V_3 x^3 \frac{\partial^3 P_0}{\partial x^3}(t, x) \right)$$

soit “le plus proche possible” des prix observés, par exemple pour une option d'achat classique. On préférera travailler avec les volatilités implicites, qui ne sont qu'une manière de représenter les prix à travers le prisme Black-Scholes.

#### 3.4.1 Le prix corrigé de l'option d'achat

Dans ce paragraphe, on étudie le cas particulier du call d'échéance  $T$  et de prix d'exercice  $K$  :  $h(x) = (x - K)_+$ . On souhaite donner le prix corrigé  $C_0 + \tilde{C}_1$  du call.  $C_0$  est le prix du call Black-Scholes pour la volatilité  $\bar{\sigma}$  et

$$\tilde{C}_1(t, x) = -(T - t) \left( V_2 x^2 \frac{\partial^2 C_0}{\partial x^2}(t, x) + V_3 x^3 \frac{\partial^3 C_0}{\partial x^3}(t, x) \right) \quad (3.32)$$

Par conséquent,

$$C_0(t, x) = C_{BS}(t, x; K, T; \bar{\sigma}) = xN(d_+) - Ke^{-r(T-t)}N(d_-) \quad (3.33)$$

où

$$\begin{cases} d_+ = \frac{\ln\left(\frac{x}{Ke^{-r(T-t)}}\right)}{\bar{\sigma}\sqrt{T-t}} + \frac{1}{2}\bar{\sigma}\sqrt{T-t}, \\ d_- = \frac{\ln\left(\frac{x}{Ke^{-r(T-t)}}\right)}{\bar{\sigma}\sqrt{T-t}} - \frac{1}{2}\bar{\sigma}\sqrt{T-t}, \end{cases} \quad (3.34)$$

et  $N(d) = \int_{-\infty}^d g(u) du$ , avec  $g(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$ . Et comme

$$\begin{aligned}\frac{\partial d_{\pm}}{\partial x}(t, x) &= \frac{1}{x\bar{\sigma}\sqrt{T-t}} \\ \frac{\partial C_0}{\partial x}(t, x) &= N(d_+) \\ \frac{\partial^2 C_0}{\partial x^2}(t, x) &= \frac{1}{x\bar{\sigma}\sqrt{T-t}} g(d_+) \\ \frac{\partial^3 C_0}{\partial x^3}(t, x) &= -\frac{g(d_+)}{x^2\bar{\sigma}^2(T-t)} \left(d_+ + \bar{\sigma}\sqrt{T-t}\right),\end{aligned}$$

en utilisant (3.32), le prix corrigé au premier ordre de l'option d'achat s'écrit finalement

$$\begin{aligned}(C_0 + \tilde{C}_1)(t, x) &= xN(d_+) - Ke^{-r(T-t)}N(d_-) \\ &\quad - \frac{xg(d_+)}{\bar{\sigma}^2} \left((V_2 - V_3)\bar{\sigma}\sqrt{T-t} - V_3d_+\right).\end{aligned}$$

### 3.4.2 Surface de volatilité implicite

On cherche à développer la volatilité implicite sous la forme

$$I(t, x) = I_0(t, x) + \sqrt{\varepsilon}I_1(t, x) + \varepsilon I_2(t, x) + \varepsilon\sqrt{\varepsilon}I_3(t, x) + \dots$$

et on souhaite calculer les deux premiers coefficients  $I_0$  et  $I_1$ . Si nous croyons à notre modèle de marché à volatilité stochastique, on a

$$C_{BS}(t, x; K, T; I(t, x)) = C_0(t, x) + \sqrt{\varepsilon}C_1(t, x) + \varepsilon C_2(t, x) + \dots$$

On développe le terme de gauche :

$$C_{BS}(t, x; K, T; I(t, x)) = C_{BS}(t, x; K, T; I_0(t, x)) + \sqrt{\varepsilon}I_1(t, x) \frac{\partial C_{BS}}{\partial \sigma}(t, x; K, T; I_0(t, x)) + \dots$$

L'identification des termes montre que le modèle à volatilité stochastique ergodique génère donc un smile :

$$I(t, x) = \bar{\sigma} - \frac{1}{\bar{\sigma}} \left( V_2 - V_3 \left( \frac{r}{\bar{\sigma}^2} + \frac{3}{2} \right) + V_3 \frac{\ln\left(\frac{K}{x}\right)}{\bar{\sigma}^2(T-t)} \right) + O(\varepsilon).$$

Etudier la surface de volatilité implicite, c'est s'intéresser à la dépendance en  $(K, T)$  de cette expression. Elle est du type

$$(K, T) \mapsto a \frac{\ln\left(\frac{K}{x}\right)}{T-t} + b + O(\varepsilon), \quad (3.35)$$

avec

$$\begin{cases} a = -\frac{V_3}{\bar{\sigma}^3}, \\ b = \bar{\sigma} - \frac{1}{\bar{\sigma}} (V_2 - V_3 (\frac{r}{\bar{\sigma}^2} + \frac{3}{2})). \end{cases} \quad (3.36)$$

$V_2$  et  $V_3$  sont petits, d'ordre  $\sqrt{\varepsilon}$ , donc  $a$  est petit et  $b$  est proche de  $\bar{\sigma}$ .

En pratique, on observe la “surface” de volatilité implicite empirique (on ne dispose que de quelques points en réalité), on cherche  $a$  et  $b$  de sorte que (3.35) colle au mieux à la surface de volatilité implicite empirique et on inverse le système (3.36) :

$$\begin{cases} V_3 = -a\bar{\sigma}^3, \\ V_2 = \bar{\sigma} (\bar{\sigma} - b - a (r + \frac{3}{2}\bar{\sigma}^2)). \end{cases} \quad (3.37)$$

Il faut avoir auparavant mesuré la volatilité moyenne  $\bar{\sigma}$ , par exemple en étudiant statistiquement la variance des rendements empiriques  $\frac{\Delta X}{X}$ .

On note que si on s'arrête au premier ordre, la volatilité implicite prévue par le modèle à volatilité stochastique est une fonction monotone de  $K$ , proportionnelle à  $\ln(K)$ . L'approximation au premier ordre ne permet pas de générer un sourire, i.e. une courbe convexe décroissante puis croissante. La volatilité implicite (3.35) est convexe en  $K$  si et seulement si  $a < 0$ , c'est-à-dire  $V_3 > 0$ . Dans ce cas, l'approximation au premier ordre de la volatilité implicite théorique est une fonction convexe décroissante qui tend vers l'infini quand  $K$  tend vers 0. La décroissance traduit un “skew”, une asymétrie de la volatilité implicite, due à la corrélation entre volatilité et sous-jacent. Notons en effet que si cette corrélation était nulle,  $V_3$ , et donc  $a$ , le seraient également, et la volatilité implicite serait constante, donc symétrique, au premier ordre. Pour voir apparaître un sourire, il faut pousser le développement limité au second ordre.

On note surtout que l'asymptotique considérée n'explique que des smiles de faible amplitude, d'ordre  $1/\sqrt{\alpha}$ , loin des ampleurs couramment observées sur les marchés. Pour expliquer de tels smiles dans le modèle (3.1)<sup>2</sup>, on doit alors considérer des valeurs de  $\alpha$  de l'ordre de  $\frac{1}{T-t}$ . On ne peut alors plus utiliser le développement limité (3.35) et on ne dispose évidemment pas de formule fermée pour le prix du call. Dans le paragraphe suivant, qui clôt ce chapitre, on se propose donc effectuer des simulations numériques originales afin d'évaluer numériquement des prix d'options et des volatilités implicites par méthode de Monte-Carlo.

---

<sup>2</sup>Une autre solution consiste à enrichir le modèle en faisant dépendre la volatilité  $\sigma_t$  d'un facteur supplémentaire, à variations lentes. C'est ce que proposent Jean-Pierre FOUQUE, George PAPANICOLAOU, K. Ronnie SIRCAR et Knut SOLNA dans leur article [30] paru en 2003.

## 3.5 Simulations numériques

### 3.5.1 Schéma d'Euler

**Définition** Comme expliqué à la fin du paragraphe 3.2.2, on suppose dans la suite que  $\Lambda = 0$ . Quitte à remplacer  $f(y)$  par  $f(y+m)$ , on peut choisir  $m = 0$ . L'équation différentielle stochastique (3.20) s'écrit dans ce cas

$$\begin{cases} dX_t = rX_t dt + \sigma_t X_t dW_t^*, \\ \sigma_t = f(Y_t), \\ dY_t = -\alpha Y_t dt + \beta d\hat{Z}_t^*, \end{cases} \quad (3.38)$$

avec les conditions initiales  $X_0 = x, Y_0 = y$ . On se donne  $M \geq 1$  pas de temps et on pose  $\Delta t = T/M$  et  $t_n = n\Delta t$ . Autrement dit, on considère la subdivision régulière

$$0 = t_0 < t_1 < t_2 < \dots < t_{M-1} < t_M = T$$

de l'intervalle  $[0, T]$ . On considère le schéma d'Euler  $(\bar{X}_{t_n}, \bar{Y}_{t_n})_{0 \leq n \leq M}$  défini par  $\bar{X}_0 = x, \bar{Y}_0 = y$  et

$$\begin{cases} \bar{X}_{t_{n+1}} - \bar{X}_{t_n} = r\bar{X}_{t_n} \Delta t + f(\bar{Y}_{t_n}) \bar{X}_{t_n} \sqrt{\Delta t} \hat{G}_n \\ \bar{Y}_{t_{n+1}} - \bar{Y}_{t_n} = -\alpha \bar{Y}_{t_n} \Delta t + \beta \sqrt{\Delta t} (\rho \hat{G}_n + \sqrt{1 - \rho^2} \check{G}_n) \end{cases} \quad (3.39)$$

où  $(\hat{G}_n)_{0 \leq n \leq M-1}$  et  $(\check{G}_n)_{0 \leq n \leq M-1}$  sont deux suites indépendantes de variables aléatoires indépendantes de même loi gaussienne centrée réduite. Notons que la dynamique de  $\bar{Y}$  est autonome.

**Erreurs statistique et trajectorielle** On lance  $N$  simulations de trajectoires (3.39) indépendantes, générant  $N$  variables aléatoires  $\bar{X}_T^{(i)}$  indépendantes, et on estime la valeur en  $t = 0$  de l'option de payoff  $H = h(X_T)$  par la moyenne empirique  $e^{-rT} \frac{1}{N} \sum_{i=1}^N h(\bar{X}_T^{(i)})$ . Ce faisant, on commet deux erreurs :

- une erreur de nature statistique, d'ordre  $1/\sqrt{N}$  :

$$\frac{\sqrt{N}}{\sigma_N} \left( \frac{1}{N} \sum_{i=1}^N h(\bar{X}_T^{(i)}) - \mathbb{E}^* [h(\bar{X}_T)] \right) \longrightarrow \mathcal{N}(0, 1)$$

en loi lorsque  $N$  tend vers l'infini, où

$$\sigma_N^2 = \frac{1}{N} \sum_{i=1}^N h(\bar{X}_T^i)^2 - \left( \frac{1}{N} \sum_{i=1}^N h(\bar{X}_T^i) \right)^2.$$

est un estimateur de la variance de  $h(\bar{X}_T)$ . On constate en pratique que  $\sigma_N^2$  converge vite vers une limite  $\sigma^2$  qui semble *dépendre peu* de la valeur du pas de temps  $\Delta t$ . Ceci permet de bien contrôler le "rayon"  $\frac{2\sigma}{\sqrt{N}}$  de l'erreur statistique, uniformément en  $\Delta t$ .

- une erreur de discrétisation temporelle  $|\mathbb{E}^* [h(\bar{X}_T)] - \mathbb{E}^* [h(X_T)]|$ . Les résultats présentés au chapitre 2, même si leurs hypothèses ne sont pas ici complètement respectées, indiquent que, même pour des payoffs  $h$  irréguliers, cette erreur est d'ordre  $\Delta t = T/M$ , ce que confirme la simulation numérique. On a dans ce cas

$$\mathbb{E}^* [h(\bar{X}_T)] - \mathbb{E}^* [h(X_T)] = C\Delta t + O(\Delta t^2) \quad (3.40)$$

où  $C$  est une constante dépendant des coefficients de la diffusion, des valeurs initiales  $x$  et  $y$  de  $X$  et  $Y$ , du payoff  $h$ , de  $T$ , mais indépendante de  $\Delta t$ . Avec les notations du chapitre 2 dans le cas où  $T = 1$ ,

$$C = \frac{1}{2} \int_0^1 P_s L_2^* P_{1-s} \mathbf{h}(x, y) ds.$$

où  $\mathbf{h}(x, y) = h(x)$  et  $(P_t, t \geq 0)$  est le semigroupe du processus de Markov  $(X, Y) = ((X_t, Y_t), t \geq 0)$ .

Il est donc naturel de choisir  $N$  et  $M$  de sorte que  $N \approx M^2$ , i.e. pour que les deux erreurs soient du même ordre<sup>3</sup>.

**Réduction de variance** On améliore la vitesse de calcul en proposant des méthodes de réduction de variance :

Tout d'abord, l'utilisation de variables antithétiques divise par quatre les appels à la fonction random : on lance  $N/4$  simulations de quatre jeux de trajectoires obtenues via les deux suites indépendantes de gaussiennes centrées réduites indépendantes  $(\hat{G}_n)_{0 \leq n \leq M-1}$  et  $(\check{G}_n)_{0 \leq n \leq M-1}$  en utilisant respectivement les suites  $(\hat{G}_n, \check{G}_n)_{0 \leq n \leq M-1}$ ,  $(\hat{G}_n, -\check{G}_n)_{0 \leq n \leq M-1}$ ,  $(-\hat{G}_n, \check{G}_n)_{0 \leq n \leq M-1}$  et  $(-\hat{G}_n, -\check{G}_n)_{0 \leq n \leq M-1}$  qui ont toutes même loi. L'utilisation de variables antithétiques réduit automatiquement la variance dans un certain nombre de cas connus. Par exemple, lorsque les variables aléatoires  $(J_i, i \geq 1)$  sont indépendantes et de même loi symétrique par rapport à 0 ( $J_1$  a même loi que  $-J_1$ ), alors l'estimateur  $B_N = \frac{1}{N} \sum_{i=1}^N \frac{f(J_i) + f(-J_i)}{2}$  de  $\mathbb{E}[f(J_1)]$  a une variance inférieure à celle de l'estimateur  $A_N = \frac{1}{2N} \sum_{i=1}^{2N} f(J_i)$  dès que  $f$  est monotone. Ici, la réduction de variance est empirique : les simulations montrent, dans le cas du put ( $h(x) = (K - x)_+$ ), que la variance est divisée par un facteur de l'ordre de 2.

Ensuite, dans le cas où  $\alpha$  est grand devant  $\frac{1}{T-t}$ , on sait que le prix de l'option est proche de la solution  $P_0$  de l'EDP  $\mathcal{L}_{BS}(\bar{\sigma})P_0 = 0$  avec la condition terminale  $P_0(T, x) = h(x)$ . Si on dispose d'une formule pour  $P_0$ , on peut espérer réduire la variance en introduisant le mouvement brownien géométrique  $X^{bs}$  :

$$dX_t^{bs} = rX_t^{bs} dt + \bar{\sigma}X_t^{bs} dW_t^*, \quad X_0^{bs} = x$$

<sup>3</sup>Reste à estimer la constante  $C_{\Delta t}$  en facteur de  $\Delta t$ . Voir ci-dessous le paragraphe sur l'extrapolation de Romberg.

et en notant que

$$\mathbb{E}^* [h(\bar{X}_T)] = \mathbb{E}^* [h(\bar{X}_T) - h(X_T^{bs})] + P_0(0, x).$$

On calcule  $X_T^{bs} = x \exp(\bar{\sigma}W_T^* + (r - \bar{\sigma}^2/2)T)$  en réutilisant les simulations exactes  $(\sqrt{\Delta t}\hat{G}_n, 0 \leq n \leq M-1)$  des accroissements du brownien  $W^* : W_T^*$  a même loi que  $\sqrt{\Delta t} \sum_{n=0}^{M-1} \hat{G}_n$ .

On espère que  $\text{Var}(h(\bar{X}_T) - h(X_T^{bs})) < \text{Var}(h(\bar{X}_T))$ . On peut introduire un paramètre  $q$  qui minimise  $\text{Var}(h(\bar{X}_T) - qh(X_T^{bs}))$  et écrire

$$\mathbb{E}^* [h(\bar{X}_T)] = \mathbb{E}^* [h(\bar{X}_T) - qh(X_T^{bs})] + qP_0(0, x).$$

Le meilleur  $q$  possible est

$$q = \frac{\text{cov}(h(\bar{X}_T), h(X_T^{bs}))}{\text{Var}(h(X_T^{bs}))}, \quad (3.41)$$

on peut l'estimer au cours de la simulation. On parle alors d'algorithme adaptatif. Attention! Si on estime  $q$  au cours de la simulation, disons par  $\hat{q}_N$ , ce dernier est fonction de tous les tirages de gaussiennes, donc corrélé à  $\bar{X}_T$  et  $X_T^{bs}$ . Par conséquent, les variables aléatoires  $\hat{q}_N$  et  $X_T^{bs,(i)}$  ne sont pas forcément indépendantes, l'estimateur

$$\frac{1}{N} \sum_{i=1}^N \left( h(\bar{X}_T^{(i)}) - \hat{q}_N h(X_T^{bs,(i)}) \right) \quad (3.42)$$

de  $\mathbb{E}^* [h(\bar{X}_T) - qh(X_T^{bs})]$  peut être biaisé, de même que l'estimateur de la variance, et si l'on dispose d'un théorème central limite (TCL) pour (3.42), l'écart-type limite est certainement plus large que celui obtenu en remplaçant  $\hat{q}_N$  par le  $q$  déterministe correspondant (3.41) dans (3.42).

On peut estimer (3.41) d'abord en appelant  $p$  fois ( $p$  "petit") la fonction `random`, ce qui nous fournit un estimateur  $\hat{q}_p$  de  $q$ , puis en utilisant  $N$  nouveaux tirages pour estimer  $\mathbb{E}^* [h(\bar{X}_T) - qh(X_T^{bs})]$  via

$$\frac{1}{N} \sum_{i=1}^N \left( h(\bar{X}_T^{(i)}) - \hat{q}_p h(X_T^{bs,(i)}) \right).$$

Alors  $\hat{q}_p$  est une variable aléatoire indépendante de  $X_T^{bs,(i)}$ , mais il faut tenir compte de sa loi pour déterminer un intervalle de confiance relatif à l'estimateur ci-dessus. On pourra consulter l'ouvrage [7] pour l'étude de certains algorithmes adaptatifs. On pourra également trouver des preuves de convergence d'algorithme adaptatifs, et des TCL correspondants, dans les travaux récents de Christophe ANDRIEU et

Eric MOULINES [1] sur les méthodes de simulation “Monte Carlo Markov Chain” adaptatives.

Enfin, on peut choisir *a priori* un  $q$  déterministe (par exemple  $q = 1$ ). Dans ce cas, l’estimation de l’intervalle de confiance est aisé - mais on peut être “loin” du meilleur  $q$  possible.

Même pour de petites valeurs de  $\alpha$ , on peut se donner une valeur de  $\bar{\sigma}$  et appliquer la méthode. En pratique, pour  $q = 1$  et  $\alpha = 1$ , on observe que la variance est réduite d’un facteur 8 environ, et pour une large gamme de  $\bar{\sigma}$ .

**Extrapolation de Romberg** Sous réserve que l’erreur de discrétisation est d’ordre  $\Delta t$ , c’est-à-dire si l’on a bien le développement (3.40), alors on peut mettre en œuvre la méthode d’extrapolation de Romberg - que les analystes appellent aussi méthode d’extrapolation de Richardson : on estime  $\mathbb{E}^* [h(\bar{X}_T)]$  pour deux valeurs  $\Delta t$  et  $\Delta t/2$  du pas de temps, et on en déduit une estimation de

$$Z_T^{\Delta t} = 2\mathbb{E}^* \left[ h \left( \bar{X}_T^{\Delta t/2} \right) \right] - \mathbb{E}^* \left[ h \left( \bar{X}_T^{\Delta t} \right) \right].$$

$Z_T^{\Delta t}$  est une approximation d’ordre 2 en temps de  $\mathbb{E}^* [h(X_T)]$  car (3.40) implique

$$Z_T^{\Delta t} - \mathbb{E}^* [h(X_T^{\Delta t})] = O(\Delta t^2).$$

Pour mesurer numériquement cette erreur, il est essentiel de pouvoir estimer dans le code informatique la constante en facteur du  $\Delta t^2$ . Pour ce faire, *supposons* - l’expérience numérique le confirme - qu’on dispose d’un développement limité

$$\mathbb{E}^* \left[ h \left( \bar{X}_T^{\Delta t} \right) \right] - \mathbb{E}^* [h(X_T)] = C'_1 \Delta t + C'_2 \Delta t^2 + o(\Delta t^2). \quad (3.43)$$

Alors il existe  $C_2$  tel que

$$Z_T^{\Delta t} - \mathbb{E}^* [h(X_T)] = C_2 \Delta t^2 + o(\Delta t^2),$$

et

$$Z_T^{\Delta t} - Z_T^{\Delta t/2} = \frac{3}{4} C_2 \Delta t^2 + o(\Delta t^2).$$

Ceci permet d’estimer *a posteriori* l’erreur due à la discrétisation temporelle  $C_2 \Delta t^2$  par

$$C_2^{\Delta t} \Delta t^2 = \frac{4}{3} \left( Z_T^{\Delta t} - Z_T^{\Delta t/2} \right).$$

En pratique, par exemple pour estimer le put européen d’échéance  $T$ ,

- on fixe un grand nombre de tirages  $N$ ,
- on se donne un pas de temps minimal  $\delta > 0$ ,
- on initialise  $\Delta t$  à  $T$ ,



- pour cette valeur de  $\Delta t$ , on estime le put par la moyenne empirique  $Z_N^{\Delta t}$  qui estime  $Z_T^{\Delta t}$  et on estime le rayon

$$\varepsilon_s^{\Delta t} = 2\sqrt{\frac{\text{Var}(Z_N^{\Delta t})}{N}} \quad (3.44)$$

de l'erreur statistique,

- on divise  $\Delta t$  par 2, on recommence les opérations de la ligne précédente, et on calcule

$$C_{2,N}^{\Delta t} \Delta t^2 = \frac{4}{3} \left( Z_N^{\Delta t} - Z_N^{\Delta t/2} \right); \quad (3.45)$$

on vérifie que cette quantité est de l'ordre du carré du pas de temps, ce qui justifie numériquement la validité du développement (3.43); en pratique, on réalisera cet item et le précédent au cours d'une même étape où les tirages de gaussiennes pour le pas de temps  $\Delta t/2$  seront ensuite réutilisées pour la simulation de pas de temps  $\Delta t$  de telle sorte que les trajectoires browniennes simulées pour les deux pas de temps coïncident aux dates  $n\Delta t$ ; ce faisant, par rapport à la situation où l'on générerait deux trajectoires browniennes indépendantes, on vérifie numériquement que l'on réduit à la fois la variance et le temps de calcul,

- on s'arrête dès que<sup>4</sup>  $C_{2,N}^{\Delta t} \Delta t^2 \leq \varepsilon_s^{\Delta t}$  ou quand  $\Delta t < \delta$ . En effet, des valeurs de  $C_{2,N}^{\Delta t} \Delta t^2$  inférieures à  $\varepsilon_s^{\Delta t}$  n'ont *aucune signification* puisqu'on ne calcule la valeur de  $Z_N^{\Delta t}$  (resp.  $Z_N^{\Delta t/2}$ ) qu'avec une précision  $\pm \varepsilon_s^{\Delta t}$  (resp.  $\pm \varepsilon_s^{\Delta t/2}$ ).

Notons que numériquement, sauf pour quelques grandes valeurs de  $\Delta t$ , on a  $\varepsilon_s^{\Delta t} \approx \varepsilon_s^{\Delta t/2}$ . On obtient donc une erreur statistique uniforme en  $\Delta t$ .

L'algorithme détermine donc un pas de temps  $\Delta t$  tel qu'on a confiance en l'estimation

$$Z_N^{\Delta t} \pm 2\varepsilon_s^{\Delta t}$$

pour le put.

### 3.5.2 Amélioration du schéma

Le schéma d'Euler a deux grands avantages : il est universel et facile à implémenter. Mais, dans le cas de la diffusion (3.38) vue à travers les deux browniens indépendants  $W^1$  et  $W^2$  :

$$\begin{cases} dX_t = rX_t dt + f(Y_t)X_t \left( \rho dW_t^2 + \sqrt{1 - \rho^2} dW_t^1 \right), & X_0 = x, \\ dY_t = -\alpha Y_t dt + \beta dW_t^2, & Y_0 = y, \end{cases} \quad (3.46)$$

on peut espérer être plus précis en exploitant les propriétés suivantes :

<sup>4</sup> *A priori*,  $C_{2,N}^{\Delta t} \Delta t^2$  décroît avec  $\Delta t$ . On peut recommencer une ou deux fois les opérations pour s'en assurer.

- la dynamique de  $Y$  est autonome - elle ne dépend pas de la dynamique de  $X$  ; on peut donc simuler d'abord  $Y$ , puis  $X$  sachant  $Y$ ,
- on sait simuler exactement la loi de  $Y$ , car  $Y$  est un processus gaussien ; en vue de simuler  $X$ , il est en fait nécessaire<sup>5</sup> de simuler la loi du couple  $(Y, W^2)$  ; ça n'est pas plus difficile car  $(Y, W^2)$  est également un processus gaussien,
- pour simuler efficacement la loi de  $(Y, W^2)$ , on peut exploiter le caractère markovien de ce processus,
- on peut implémenter un schéma d'Euler pour simuler  $X$  sachant  $Y$  et  $W^2$ . Mais on dispose d'une expression exacte de  $X_T$  en fonction des trajectoires de  $Y$ ,  $W^2$  et  $W^1$  :

$$X_T = xe^{rT} \exp \left( \rho \int_0^T f(Y_t) dW_t^2 + \sqrt{1 - \rho^2} \int_0^T f(Y_t) dW_t^1 - \frac{1}{2} \int_0^T f(Y_t)^2 dt \right).$$

On peut espérer être plus précis en simulant directement l'argument dans l'exponentielle. Cela revient à simuler par un schéma d'Euler  $\log(X_T)$  plutôt que  $X_T$ .

Revenons point par point sur ces améliorations.

**Simulation du couple  $(Y, W^2)$**  Nous proposons deux méthodes pour simuler la loi du couple  $(Y, W^2)$ . Elles sont toutes deux basées sur le résultat suivant : si  $M \geq 1$  et si  $0 < t_1 < t_2 < \dots < t_M = T$ , alors le vecteur

$$V = (Y_{t_1}, W_{t_1}^2, Y_{t_2}, W_{t_2}^2, Y_{t_3}, W_{t_3}^2, \dots, Y_{t_M}, W_{t_M}^2)$$

est un vecteur gaussien.

**Méthode 1** On génère la matrice de variance-covariance  $\Gamma$  du vecteur gaussien  $V$ , son vecteur des moyennes  $m$ , on calcule une racine carrée  $A$  de  $\Gamma$  par la méthode de Cholevsky - i.e. on cherche l'unique  $A$  triangulaire inférieure à diagonale strictement positive telle que  $\Gamma = AA^t$  - et on génère un vecteur  $G$  de  $2M$  variables aléatoires gaussiennes centrées réduites indépendantes. Alors le vecteur  $AG + m$  a même loi que  $V$ .

Pour calculer  $\Gamma$ , il suffit de connaître, pour  $1 \leq i \leq j \leq M$  :

$$\begin{aligned} \text{cov}(Y_{t_i}, Y_{t_j}) &= \nu^2 (e^{-\alpha(t_i+t_j)} - e^{-\alpha(t_j-t_i)}), \\ \text{cov}(W_{t_i}^2, W_{t_j}^2) &= t_i, \\ \text{cov}(Y_{t_i}, W_{t_j}^2) &= \frac{\beta}{\alpha} (1 - e^{-\alpha t_i}), \\ \text{cov}(Y_{t_j}, W_{t_i}^2) &= e^{-\alpha(t_j-t_i)} \text{cov}(Y_{t_i}, W_{t_j}^2). \end{aligned}$$

---

<sup>5</sup>sauf dans le cas où  $\rho = 0$ .

Pour calculer  $m$ , il suffit de connaître

$$\begin{aligned}\mathbb{E}^* [Y_{t_i}] &= ye^{-\alpha t_i}, \\ \mathbb{E}^* [W_{t_i}^2] &= 0.\end{aligned}$$

L'inconvénient de cette méthode est que l'on utilise des matrices de taille  $2M$  - deux fois le nombre de pas de temps - donc éventuellement de grande taille.

**Méthode 2** Ici on exploite le caractère markovien - *en plus* du caractère gaussien - du couple  $(Y_t, W_t^2)_{0 \leq t \leq T}$ . Pour cela, on se donne une subdivision régulière  $0 = t_0 < t_1 < t_2 < \dots < t_M = T$  de  $[0, T]$  de pas  $\Delta t$  et on remarque que pour  $1 \leq n \leq M$ ,

$$\begin{aligned}Y_{t_n} &= e^{-\alpha \Delta t} \left( Y_{t_{n-1}} + \beta e^{-\alpha t_{n-1}} \int_{t_{n-1}}^{t_n} e^{\alpha s} dW_s^2 \right), \\ W_{t_n}^2 &= W_{t_{n-1}}^2 + (W_{t_n}^2 - W_{t_{n-1}}^2),\end{aligned}$$

si bien que si l'on pose :

$$\begin{aligned}V_n &= (Y_{t_n}, W_{t_n}^2)^t \\ U_n &= \left( \beta e^{-\alpha t_{n-1}} \int_{t_{n-1}}^{t_n} e^{\alpha s} dW_s^2, W_{t_n}^2 - W_{t_{n-1}}^2 \right)^t \\ g \left( \begin{pmatrix} y \\ w \end{pmatrix}, \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} \right) &= \begin{pmatrix} e^{-\alpha \Delta t} (y + u^1) \\ w + u^2 \end{pmatrix}\end{aligned}$$

on a pour pour  $n \geq 1$

$$V_n = g(V_{n-1}, U_n).$$

Or on note que  $(U_n)_{1 \leq n \leq M}$  est une suite de vecteurs aléatoires gaussiens indépendants de dimension 2 de même loi  $\mathcal{N}_2(0, \Gamma)$ , où

$$\Gamma = \begin{pmatrix} \nu^2 (e^{2\alpha \Delta t} - 1) & \frac{\beta}{\alpha} (e^{\alpha \Delta t} - 1) \\ \frac{\beta}{\alpha} (e^{\alpha \Delta t} - 1) & \Delta t \end{pmatrix},$$

et indépendante de  $V_0$ . Par conséquent,  $(V_n)_{0 \leq n \leq M}$  est une chaîne de Markov qu'on peut simuler facilement en générant  $M$  lois  $\mathcal{N}_2(0, \Gamma)$  indépendantes. Pour ce faire, on applique la méthode 1 - mais en dimension 2 seulement! Cette méthode est plus rapide, car la méthode de Cholevsky en dimension  $d$  nécessite de l'ordre de  $d^3$  opérations.

**Simulation de  $X_T$**  On rappelle que

$$X_T = xe^{rT} \exp \left( \rho \int_0^T f(Y_t) dW_t^2 + \sqrt{1-\rho^2} \int_0^T f(Y_t) dW_t^1 - \frac{1}{2} \int_0^T f(Y_t)^2 dt \right)$$

et que l'on a déjà simulé le couple  $(Y, W^2)$  aux dates  $t_n = \frac{nT}{M}$ . On décide d'approximer respectivement  $I_0 = \int_0^T f(Y_t)^2 dt$ ,  $I_1 = \int_0^T f(Y_t) dW_t^1$  et  $I_2 = \int_0^T f(Y_t) dW_t^2$  par

$$\begin{aligned} I_0^{\Delta t} &= \Delta t \sum_{i=0}^{M-1} f(Y_{t_i})^2, \\ I_1^{\Delta t} &= \sum_{i=0}^{M-1} f(Y_{t_i}) \left( W_{t_{i+1}}^1 - W_{t_i}^1 \right), \\ I_2^{\Delta t} &= \sum_{i=0}^{M-1} f(Y_{t_i}) \left( W_{t_{i+1}}^2 - W_{t_i}^2 \right). \end{aligned}$$

Les valeurs des  $Y_{t_i}$  et  $W_{t_i}^2$  sont connues, il suffit donc à cette étape de générer une suite  $W_{t_{i+1}}^1 - W_{t_i}^1$  de gaussiennes indépendantes centrées et de variance  $\Delta t$ . On approxime alors  $X_T$  par

$$\bar{X}_T^{\Delta t} = xe^{rT} \exp \left( \rho I_2^{\Delta t} + \sqrt{1-\rho^2} I_1^{\Delta t} - \frac{1}{2} I_0^{\Delta t} \right). \quad (3.47)$$

### 3.5.3 Mise en œuvre

Détaillons la mise en œuvre informatique de l'évaluation du put européen dont le payoff  $h(x) = (K - x)_+$  a l'avantage d'être borné. Par la relation de parité call-put, on en déduit une évaluation du prix du call, puis une évaluation de la volatilité implicite :

1. on fixe `NB_TIRAGES`  $\equiv N$  le nombre d'appels à la fonction `random`,
2. on entre les paramètres du modèle :
  - pour le sous-jacent :
    - la valeur `SPOT_TODAY`  $\equiv x$  du spot aujourd'hui ; par défaut `SPOT_TODAY` vaut 1, c'est-à-dire qu'on raisonne avec un spot unité,
    - le taux court `TAUX`  $\equiv r$ ,
    - la fonction de volatilité  $f$  ; on considère souvent  $f(y) = me^y$  ou  $f(y) = m' |m + y|$ ,
  - pour le processus  $Y$  qui gouverne la volatilité :
    - la valeur initiale `Y_TODAY`  $\equiv y$ ,
    - la valeur `ALPHA`  $\equiv \alpha$  de la force de rappel,

- la valeur  $\text{NU} \equiv \nu$  de l'écart-type à long terme de  $Y$ ,
  - la valeur  $\text{RHO} \equiv \rho$  de la corrélation entre les deux browniens qui gouvernent  $X$  et  $Y$ ,
3. on fixe  $\text{ECHEANCE} \equiv T$ , la maturité de l'option; par défaut elle vaut 1, c'est-à-dire qu'on raisonne sur une année,
  4. on fixe  $\text{M}_{\min}$  et  $\text{M}_{\max}$  les valeurs minimale et maximale autorisées du nombre  $M$  de pas de temps,
  5. afin de tracer la courbe de volatilité implicite  $K \mapsto I(K)$ , on envisage différentes valeurs du strike; on fixe  $\text{strike}_{\min}$ ,  $\text{strike}_{\max}$  et  $\text{nb\_strike}$  qui sont respectivement le plus petit strike, le plus grand strike et le nombre de strikes considérés; on en déduit  $\text{pas\_strike}$ , le pas constant entre deux strikes de la grille,
  6. on fixe une volatilité Black-Scholes  $\text{sigmabs}$ ,
  7. on fixe le coefficient  $q$  - déterministe - qui intervient dans la réduction de variance,
  8. l'exécution commence : on lance un chronomètre,
  9. on crée deux matrices et huit vecteurs :
    - $\text{prix}[j][0]$  contient le prix du put pour le strike  $\text{strike}_{\min} + j \text{pas\_strike}$ ,
    - $\text{prix}[j][1]$  contient l'erreur statistique sur ce prix,
    - $\text{prixfin}[j][0]$  et  $\text{prixfin}[j][1]$  sont les analogues pour un pas de temps moitié,
    - $E$  est un échantillon de  $\text{NB\_TIRAGES}$  spots à la date  $T$ ,
    - $e$  est son analogue pour un pas de temps moitié,
    - $E_{\text{bs}}$  est l'échantillon des  $\text{NB\_TIRAGES}$  spots Black-Scholes construits à partir de  $\text{sigmabs}$ ,  $W_T^1$  et  $W_T^2$ ,
    - $\text{erreurstat}[j] \equiv \text{prixfin}[j][1]$  et  $\text{erreurtemps}[j]$  contiennent respectivement le rayon de l'erreur statistique (3.44) et le rayon de l'erreur (3.45) due à la discrétisation temporelle correspondants au strike  $\text{strike}_{\min} + j \text{pas\_strike}$ ,
    - $\text{call}[j]$  contient le prix du call calculé à partir du prix du put par la relation de parité,
    - $\text{vol}[j]$  contient la volatilité implicite calculée à partir de  $\text{call}[j]$  en inversant la formule de Black-Scholes,
    - $\text{volsup}[j]$  est celle calculée à partir de  $\text{call}[j] + \text{erreurstat}[j]$ ,
  10. on initialise à zéro l'entier  $\text{arret}$ ; l'exécution cessera quand  $\text{arret}$  vaudra  $\text{nb\_strike}$ , signifiant que pour tout  $j$ ,  $\text{erreurtemps}[j] < \text{erreurstat}[j]$ ,
  11. on initialise  $M$  à  $\text{M}_{\min}$ ,
  12. tant que  $\text{arret} < \text{nb\_strike}$  et  $M \leq \text{M}_{\max}$ ,

- NB\_PAS\_DE\_TEMPS reçoit  $M$ ,
- on met `arret` à zéro,
- on remplit `e` pour la valeur  $\Delta t/2 = \frac{1}{2M}$  du pas de temps, `E` pour le pas de temps  $\Delta t$  et `Ebs`. Plus précisément, on répète NB\_TIRAGES/4 fois l'opération suivante :
  - on simule le vecteur  $V^{\Delta t/2} = (Y_{t_1}, W_{t_1}^2, Y_{t_2}, W_{t_2}^2, Y_{t_3}, W_{t_3}^2, \dots, Y_{t_{2M}}, W_{t_{2M}}^2)$ , où  $t_n = n \frac{\Delta t}{2}$ ; ceci nécessite le tirage d'un vecteur  $G^{\Delta t/2}$  de  $4M$  gaussiennes centrées réduites; on en déduit  $I_0^{\Delta t/2}$  et  $I_2^{\Delta t/2}$ ,
  - on simule  $B^{\Delta t/2} = (W_{t_1}^1, W_{t_2}^1, W_{t_3}^1, \dots, W_{t_{2M}}^1)$ ; ceci nécessite  $2M$  tirages de gaussiennes; on en déduit  $I_1^{\Delta t/2}$ ,
  - on en déduit  $\bar{X}_T^{\Delta t/2}$  via (3.47); on remplit ainsi une nouvelle case de `e`,
  - en utilisant  $G^{\Delta t/2}$  avec  $-B^{\Delta t/2}$ , puis  $-G^{\Delta t/2}$  avec  $B^{\Delta t/2}$  et enfin  $-G^{\Delta t/2}$  avec  $-B^{\Delta t/2}$ , on génère trois autres valeurs  $\bar{X}$ ; au total on a donc rempli quatre cases de `e`,
  - on utilise *les mêmes valeurs* de  $V^{\Delta t} = (Y_{t_2}, W_{t_2}^2, Y_{t_4}, W_{t_4}^2, \dots, Y_{t_{2M}}, W_{t_{2M}}^2)$  et de  $B^{\Delta t} = (W_{t_2}^1, W_{t_4}^1, \dots, W_{t_{2M}}^1)$  pour calculer  $I_0^{\Delta t}$ ,  $I_1^{\Delta t}$ ,  $I_2^{\Delta t}$  et  $\bar{X}_T^{\Delta t}$ ; ceci se fait sans appel à la fonction `random`; par le même jeu sur les signes de  $G^{\Delta t}$  et de  $B^{\Delta t}$ , on remplit quatre cases de `E`,
  - enfin on utilise  $\pm W_T^1$  et  $\pm W_T^2$  pour remplir quatre cases de `Ebs`,
- pour chaque  $j \in \{0, 1, \dots, \text{nb\_strike} - 1\}$ ,
  - STRIKE reçoit `strike_min + j pas_strike`,
  - `prix` reçoit `prixfin`, et on réactualise `prixfin`, c'est-à-dire qu'on affecte à `prixfin[j][0]` la quantité

$$qP_{\text{sigmabs}}^{bs} + \frac{1}{\text{NB\_TIRAGES}} \sum_{i=0}^{\text{NB\_TIRAGES}-1} (2h(e[i]) - h(E[i]) - qh(Ebs[i]))$$

où  $P_{\text{sigmabs}}^{bs}$  est le prix que donnent BLACK et SCHOLLES au put pour la volatilité `sigmabs`, et à `prixfin[j][1]` deux fois l'écart-type correspondant; on place cette dernière valeur dans `erreurstat[j]`,

- si  $M \neq M_{\text{min}}$ , on calcule l'erreur temporelle `erreurtemps[j]` par la formule  $\frac{4}{3}(\text{prix}[j][0] - \text{prixfin}[j][0])$  et si `erreurtemps[j] < erreurstat[j]`, on incrémente `arret` de 1,
  - on multiplie  $M$  par 2,
13. si on est sortis de la boucle précédente sans que `arret=nb_strike`, c'est que  $M_{\text{max}}$  n'a pas été choisi suffisamment grand; on affiche un message en ce sens,
  14. sinon on écrit dans un fichier la valeur des paramètres entrés, la dernière valeur prise par le pas de temps  $\Delta t$  puis, pour chaque  $j$ , la valeur du strike correspondant, la valeur du put `prixfin[j][0]`, l'erreur statistique `erreurstat[j]`,

l'erreur due à la discrétisation temporelle `erreurtemps [j]` (plus petite), la volatilité implicite `vol [j]` et l'erreur `volsup [j]-vol [j]` sur cette dernière,

15. enfin, on affiche le temps total d'exécution.

### 3.5.4 Résultats

**Un exemple de tableau des résultats** On commence par donner l'ensemble des résultats pour les paramètres suivants :

$N$	250 000
$x$	1
$r$	0
$f(y)$	$\sigma_f e^y$
$\sigma_f$	0.2
$Y_0$	0
$\alpha$	1
$\nu$	0.5
$\rho$	0
$T$	1

On souhaite représenter graphiquement la volatilité implicite en fonction de la quantité  $\log\left(\frac{K}{x}\right)$ . Pour obtenir des points régulièrement espacés sur ce graphe, on considère en fait les strikes

$$K_j = \exp(l_{\min} + j h_t)$$

pour  $j \in \{0, 1, \dots, 12\}$ ,  $l_{\min} = -0.6$  et  $h_t = 0.1$ .

L'exécution prend fin en moins d'une seconde - sur une machine équipée d'un processeur courant - avec le pas de temps  $\Delta t = 1/8$  et donne les résultats suivants :

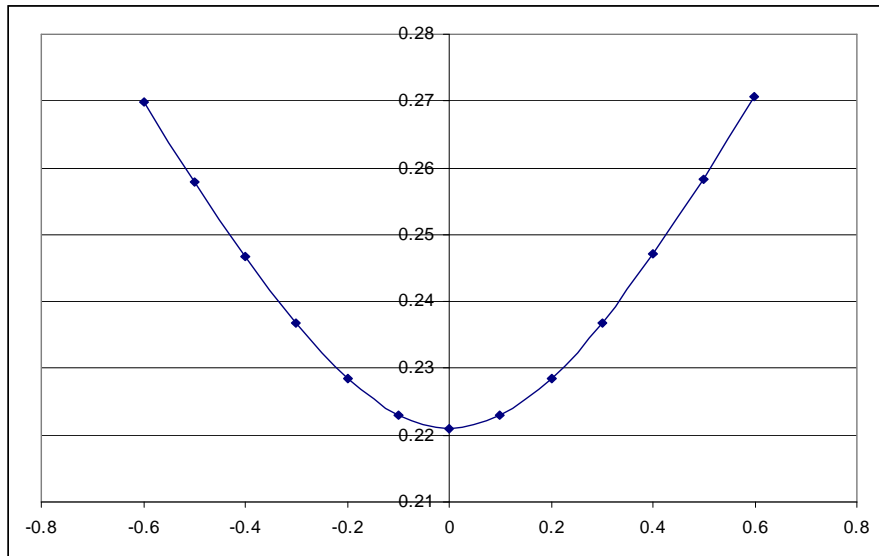


FIG. 3.9 – Volatilité implicite  $I$  en fonction de  $\log \frac{K}{x}$

$\log \frac{K}{x}$	put	erreurstat	erreurtemps	vol	erreurvol
-0.6	0.000906	0.000046	-0.000022	0.269799	0.003624
-0.5	0.001997	0.000069	-0.000038	0.257973	0.002861
-0.4	0.004440	0.000106	-0.000026	0.24672	0.002408
-0.3	0.009887	0.000161	0.000017	0.236683	0.002098
-0.2	0.021635	0.000234	0.000114	0.228493	0.001907
-0.1	0.045294	0.000292	0.000181	0.223003	0.001711
0	0.088006	0.000309	0.000116	0.221048	0.001559
0.1	0.155243	0.000343	0.000166	0.223039	0.001821
0.2	0.247806	0.000409	0.000075	0.228421	0.002724
0.3	0.363213	0.000473	-0.000073	0.23673	0.004518
0.4	0.498493	0.000523	-0.000186	0.247078	0.007701
0.5	0.652046	0.000558	-0.000108	0.258379	0.013041
0.6	0.82381	0.000580	-0.000109	0.270705	0.021148

La figure 3.9 représente la volatilité implicite  $I$  en fonction de  $\log \frac{K}{x}$ .

On constate que :

- le smile est convexe, décroissant puis croissant, avec un minimum autour de la monnaie ( $K = x$  ici car  $r = 0$ ),
- le smile *vu comme fonction de  $\log \frac{K}{x}$*  est remarquablement symétrique,
- la valeur minimale  $I_{\min} \approx 0.221$  est supérieure à  $\sigma_f \exp(\mathbb{E}[Y_t]) = \sigma_f = 0.2$



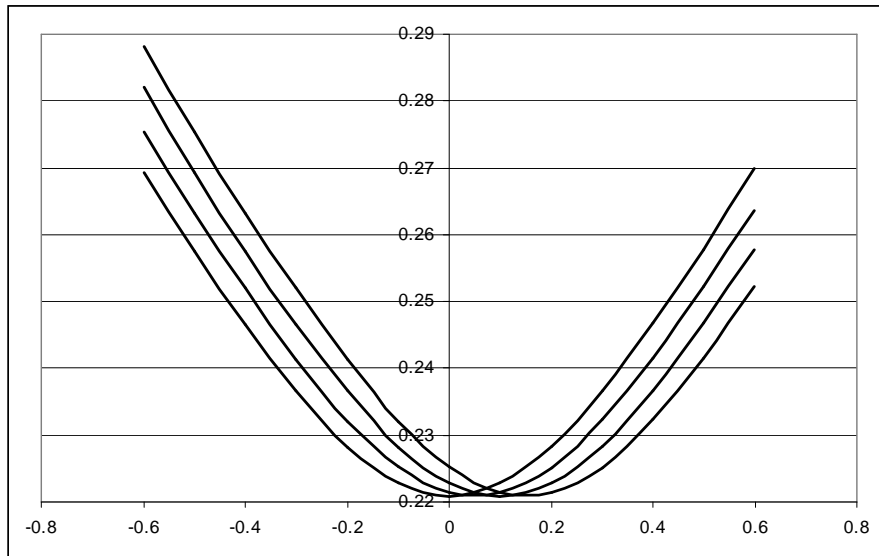


FIG. 3.10 – De gauche à droite :  $r = 0, 0.05, 0.1, 0.15$

mais inférieure à  $\bar{\sigma} = \sqrt{\langle f^2 \rangle}$ , i.e. si  $G \sim \mathcal{N}(0, 1)$ ,

$$\bar{\sigma}^2 = \mathbb{E}[f(\nu G)^2] = \sigma_f^2 \mathbb{E}[e^{2\nu G}] = \sigma_f^2 e^{2\nu^2}$$

d'où  $\bar{\sigma} \approx 0.2568$ .

A partir de cette courbe prise comme référence, nous allons étudier l'influence de chacun des paramètres sur le smile. On fait successivement varier  $r, \sigma_f, y, \alpha, \nu, \rho$  et  $T$ , toutes choses égales par ailleurs.

#### Influence de $r$ (figure 3.10)

Augmenter  $r$  revient à translater de  $rT$  le smile sur la droite. En effet,

$$X_T = e^{rT} \times x \exp \left( \rho \int_0^T f(Y_t) dW_t^2 + \sqrt{1 - \rho^2} \int_0^T f(Y_t) dW_t^1 - \frac{1}{2} \int_0^T f(Y_t)^2 dt \right).$$

#### Influence de $\sigma_f$ (figure 3.11)

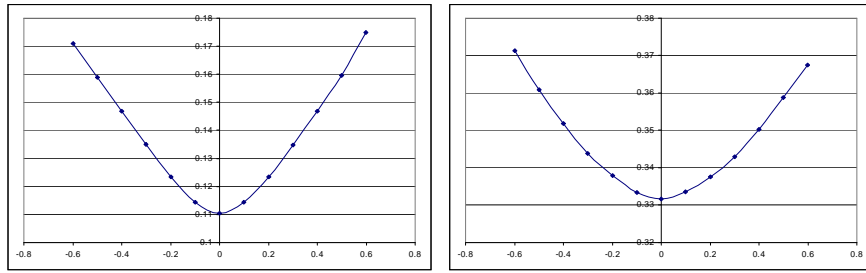


FIG. 3.11 –  $\sigma_f = 0.1$  (gauche) ou  $0.3$  (droite)

On donne ci-dessous les valeurs de  $I$  pour les trois valeurs testées de  $\sigma_f$  :

$\log \frac{K}{x}$	$\sigma_f = 0.1$	$\sigma_f = 0.2$	$\sigma_f = 0.3$
-0.6	0.1710	0.2700	0.3714
-0.5	0.1573	0.2583	0.3609
-0.4	0.1454	0.2478	0.3518
-0.3	0.1345	0.2377	0.3439
-0.2	0.1237	0.2295	0.3378
-0.1	0.1143	0.2238	0.3334
0	0.1104	0.2213	0.3317
0.1	0.1144	0.2237	0.3335
0.2	0.1234	0.2292	0.3375
0.3	0.1335	0.2370	0.3430
0.4	0.1441	0.2459	0.3503
0.5	0.1535	0.2555	0.3587
0.6	0.1618	0.2645	0.3674

Autour de la monnaie,  $\sigma_f$  semble jouer proportionnellement ; loin de la monnaie,  $\sigma_f$  semble jouer plutôt additivement.

### Influence de $y$ (figure 3.12)

Pour des grandes valeurs de  $\alpha$ , l'effet de moyenne masque  $y = Y_0$ . En revanche, pour des valeurs de  $\alpha T$  proches de 1, comme c'est le cas ici ( $\alpha = T = 1$ ), on n'oublie pas la condition initiale du processus  $Y$ .

### Influence de $\alpha$ (figure 3.13)

On peut faire les observations suivantes :

- quand  $\alpha$  est proche de 0, le smile s'applatit sur la valeur  $\sigma_f = 0.2$  ; on rappelle qu'ici  $y = 0$  ; dans ce cas en effet, les trajectoires sur  $[0, 1]$  de  $Y$  restent proches de l'axe des abscisses (cf. figure 3.7), donc  $\sigma_t = f(Y_t) = \sigma_f e^{Y_t}$  reste proche de  $\sigma_f$  ; quantitativement, la variance de  $Y_t$  vaut  $\nu^2 (1 - e^{-2\alpha t})$  est petite,

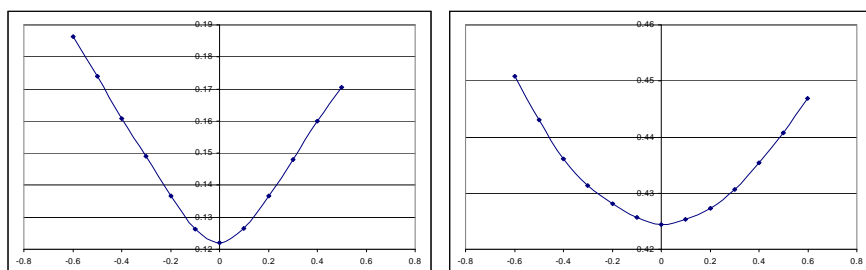


FIG. 3.12 -  $y = -1$  (gauche) ou  $1$  (droite)

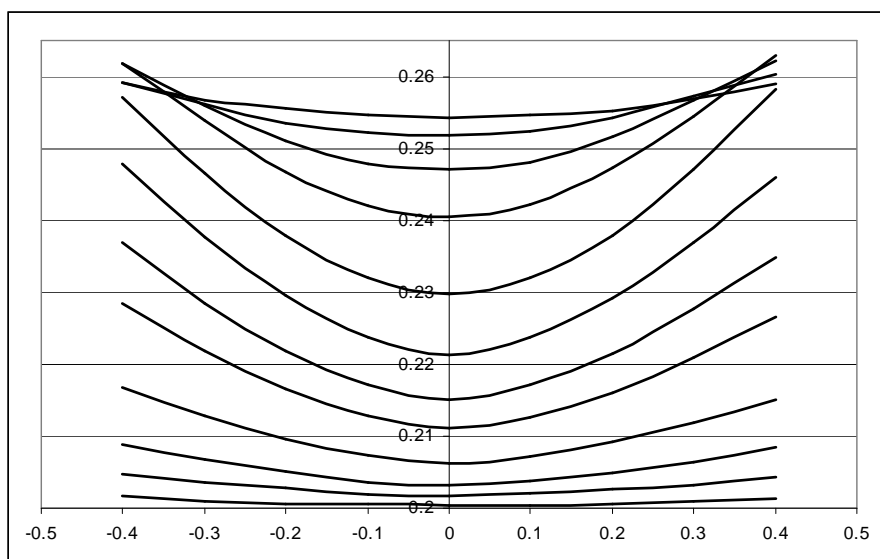


FIG. 3.13 - De bas en haut pour l'abscisse  $0$  :  $\alpha = 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 1, 2, 5, 10, 20, 40$

- quand  $\alpha$  augmente, la variance de  $Y_t$  aussi, la volatilité de la volatilité augmente, ce qui tend à augmenter la courbure - ou convexité - du smile,
- quand  $\alpha$  tend vers  $+\infty$ , le smile s'applatit sur la valeur  $\bar{\sigma} \approx 0.2568$ , conformément à l'analyse asymptotique développée par Jean-Pierre FOUQUE, George PAPANICOLAOU et K. Ronnie SIRCAR ; c'est l'effet de moyenne : moralement, toutes les trajectoires de  $Y$  sur  $[0, 1]$  visitent tous les états possibles selon la probabilité stationnaire ; pour le sous-jacent  $X$ , il y a donc *trajectoire par trajectoire* un effet de moyenne sur les valeurs possibles de  $Y$  ; de ce fait, quelle que soit la trajectoire  $(X_t, Y_t)_{t \in [0, 1]}$ , le variable aléatoire  $X_T$  est proche d'une variable log-normale pour la valeur  $\bar{\sigma}$  qui est à la fois :
  - la moyenne spatiale des volatilités  $f(Y)$  contre la probabilité stationnaire de  $Y$ ,
  - la moyenne temporelle de  $\sqrt{\frac{1}{T} \int_0^T f(Y_t)^2 dt}$  sur chaque trajectoire - la même pour chaque trajectoire.

C'est caractéristique de la propriété d'*ergodicité*.

Il existe donc une valeur critique  $\alpha_c$  de  $\alpha$  qui maximise la courbure du smile à la monnaie ; sur notre exemple où  $T = 1$ ,  $\alpha_c \approx 1$ . Rappelons que  $\alpha$  est l'inverse d'un temps, aussi est-il raisonnable de rechercher  $\alpha_c$  de la forme  $\frac{k\alpha}{T}$ . Cette valeur de  $\alpha$  marque l'équilibre entre deux mécanismes :

- **Mécanisme 1** : quand  $\alpha$  augmente,  $\text{Var}[Y_t] = \nu^2 (1 - e^{-2\alpha t})$  augmente et la courbure du smile aussi : l'incertitude sur la volatilité est plus grande et la distribution de  $X_T$  est d'autant plus loin d'une log-normale.
- **Mécanisme 2** : quand  $\alpha$  augmente, les trajectoires de  $Y$  tendent à visiter *chacune en particulier* tous les états possibles de  $Y$  suivant la probabilité  $\pi = \mathcal{N}(0, \nu^2)$ , si bien que la quantité  $\frac{1}{T} \int_0^T f(Y_t) dt$  tend à être la même pour toutes les trajectoires, de même pour  $\sqrt{\frac{1}{T} \int_0^T f(Y_t)^2 dt}$  qui tend à être  $\bar{\sigma}$  sur toutes les trajectoires, et  $X_T$  est proche d'une distribution Black-Scholes pour la valeur  $\bar{\sigma}$ .

Notons aussi que la volatilité implicite  $I^{ATM}(\alpha)$  à la monnaie<sup>6</sup> semble une fonction croissante de  $\alpha$ , telle que  $I^{ATM}(0^+) = \sigma_f$  et  $I^{ATM}(+\infty) = \bar{\sigma}$ . Nous verrons en outre plus loin que  $I^{ATM}$  semble ne dépendre de  $\alpha$  et  $T$  que *via* le produit  $\alpha T$ . Elle est donc aussi d'autant plus grande que  $T$  est grand.

### Influence de $\nu$ (figure 3.14)

On remarque que :

- lorsque  $\nu$  tend vers 0, le smile s'applatit sur la valeur  $\sigma_f = 0.2$  ; en effet, à la limite où  $\nu$  tend vers 0, vu que  $Y_0 = 0$ ,  $Y$  vaut identiquement 0,
- lorsque  $\nu$  augmente, la volatilité implicite augmente et la courbure du smile à

---

<sup>6</sup> *ATM* pour "at the money"

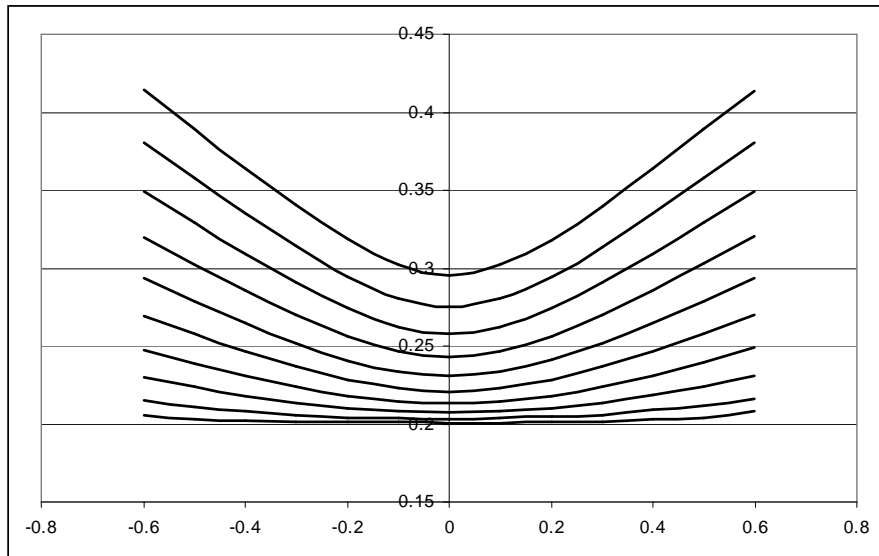


FIG. 3.14 – De bas en haut :  $\nu = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$

la monnaie aussi ; les smiles semblent avoir des asymptotes quand  $\log \frac{K}{x}$  tend vers  $-\infty$  ou  $+\infty$ , dont les pentes augmentent - en valeur absolue - avec  $\nu$ .

Ceci peut se comprendre de la façon suivante : lorsque  $\nu$  augmente, c'est la volatilité de la volatilité qui augmente, on fait payer un risque de volatilité, les prix d'option - et donc la volatilité implicite - sont donc plus élevés ; par ailleurs, la distribution de  $X_T$  s'éloigne de plus en plus d'une log-normale car on donne de plus en plus d'importance au bruit dans  $Y$ ,

- numériquement, il semble que la pente asymptotique soit proportionnelle à  $\nu$ . En effet, si pour la fonction  $I(\log \frac{K}{x})$  on note  $\Delta_g = I(-0.6) - I(-0.5)$  et  $\Delta_d = I(0.6) - I(0.5)$ , on a :

$\nu$	$\frac{\Delta_g}{\nu}$	$\frac{\Delta_d}{\nu}$
0.2	0.0192	0.0232
0.3	0.0206	0.0223
0.4	0.0217	0.0242
0.5	0.0235	0.0243
0.6	0.0239	0.0244
0.7	0.0248	0.0248
0.8	0.0250	0.0251
0.9	0.0247	0.0251
1.0	0.0250	0.0251

**Influence de  $\rho$**  (figure 3.15)

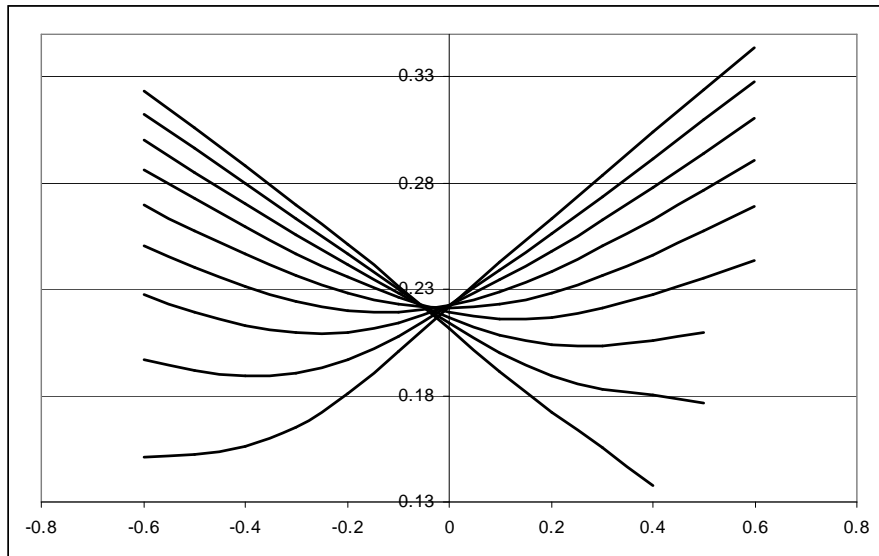


FIG. 3.15 – De haut en bas, à gauche :  $\rho = -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$

On note que :

- les volatilités à la monnaie sont à peu près toutes les mêmes,
- lorsque  $\rho \neq 0$ , la courbe de smile n'est plus symétrique par rapport à la monnaie; ceci signifie que la distribution de  $X_T$  n'est plus symétrique autour de  $x$ . On peut décrire ce phénomène par la caricature suivante. Supposons  $\rho < 0$ , et  $Y_{t_n} = 0$ . Imaginons un bruit discret  $\Delta W_{t_n}^2 > 0$ . Alors sur l'intervalle  $[t_n, t_{n+1}]$   $Y$  augmente, la volatilité  $f(Y)$  aussi (à la date  $t_{n+1}$ , elle est alors strictement supérieure à  $\sigma_f$ ). Par ailleurs, le bruit  $\rho \Delta W_{t_n}^2 + \sqrt{1 - \rho^2} \Delta W_{t_n}^1$  sur le spot a alors tendance à être négatif, donc le spot (actualisé) a tendance à être “beaucoup” plus petit en  $t_{n+1}$  qu'en  $t_n$ , “beaucoup” car le bruit négatif sur le spot est amplifié par la grande valeur de la volatilité. Inversement, si  $\Delta W_{t_n}^2 < 0$ , alors sur l'intervalle  $[t_n, t_{n+1}]$   $Y$  diminue, la volatilité  $f(Y)$  aussi (à la date  $t_{n+1}$ , elle est alors strictement inférieure à  $\sigma_f$ ). Par ailleurs, le bruit  $\rho \Delta W_{t_n}^2 + \sqrt{1 - \rho^2} \Delta W_{t_n}^1$  sur le spot a alors tendance à être positif, donc le spot (actualisé) a tendance à être “un peu” plus grand en  $t_{n+1}$  qu'en  $t_n$ , “un peu” car le bruit positif sur le spot est atténué par la faible valeur de la volatilité. Ainsi, les queues de distribution de  $\log X_T$  sont dissymétriques : la queue de distribution de gauche est plus large que celle de droite; par conséquent, par rapport à la situation où  $\rho = 0$ , on donne plus de poids aux événements du type  $X_T \in [x_1, x_2]$ , où  $x_1 < x_2 < x$ . Pour les strikes en-dessous de  $x$ , les calls - et la volatilité implicite - sont donc plus bas. C'est l'inverse pour les strikes au-dessus de  $x$ . Financièrement, à une volatilité plus grande correspond des prix de sous-jacents plus bas. On peut bien sûr faire un raisonnement analogue

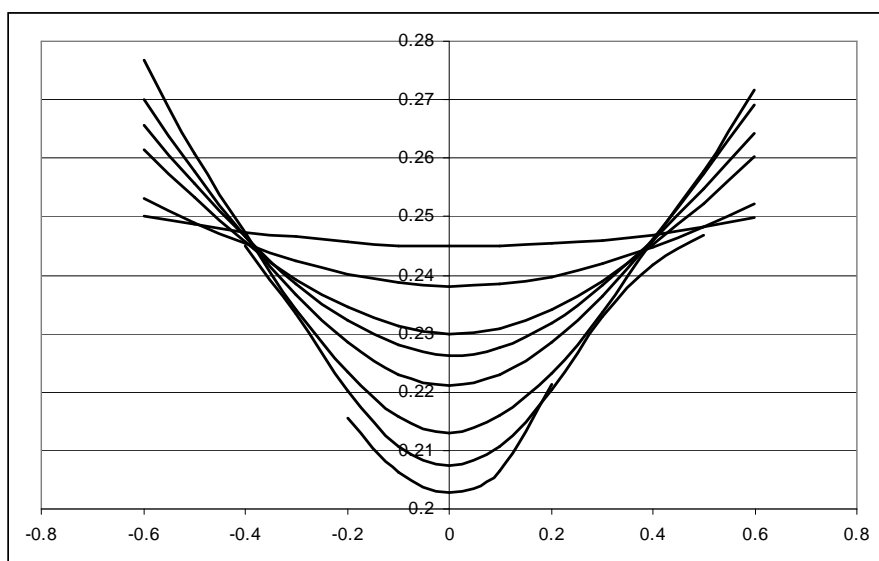


FIG. 3.16 – Du plus courbé au plus plat :  $T = 0.1, 0.25, 0.5, 1, 2, 4, 8$

dans le cas où  $\rho > 0$ .

#### Influence de $T$ (figure 3.16)

On remarque que :

- lorsque  $T$  tend vers  $+\infty$ , le smile tend à s'aplatir sur la valeur  $\bar{\sigma} \approx 0.2568$  ; de ce point de vue, il y a équivalence formelle entre les limites  $\alpha \rightarrow +\infty$  et  $T \rightarrow +\infty$  ; l'interprétation est la même que dans l'asymptotique  $\alpha \rightarrow +\infty$ ,
- lorsque  $T$  tend vers 0, en revanche, le smile semble se courber de plus en plus.

Certes la loi de  $Y_t$  ne dépend de  $t$  que *via*  $\alpha T$ , mais la loi de  $X_T$  dépend de toute la trajectoire de  $Y$  jusqu'à  $T$ , qui elle dépend de  $\alpha$  et  $T$  de manière plus complexe que simplement *via* le produit  $\alpha T$ . En revanche, moralement, aussi bien dans la limite où  $T \rightarrow +\infty$  que dans la limite où  $\alpha \rightarrow +\infty$ , on peut remplacer  $(f(Y_t))_{0 \leq t \leq T}$  par  $\bar{\sigma}$ .

#### Influence de $\alpha$ et $T$ à $\alpha T$ constant (figure 3.17)

Il semble que :

- le prix de l'option à la monnaie - et donc la volatilité implicite  $I^{ATM}$  - ne dépend de  $\alpha$  et  $T$  que *via* le produit  $\alpha T$ ,
- à  $\alpha T$  constant, la courbure est d'autant plus grande que  $T$  est petit, donc que  $\alpha$  est grand.

**Cas où  $f(y) = \sigma_- \mathbf{1}_{\{y \leq 0\}} + \sigma_+ \mathbf{1}_{\{y > 0\}}$**  Nous reprenons en partie l'étude ci-dessus dans le cas où  $f(y) = \sigma_- \mathbf{1}_{\{y \leq 0\}} + \sigma_+ \mathbf{1}_{\{y > 0\}}$ . On prend  $\sigma_- = 0.1, \sigma_+ = 0.2$ . La fonction

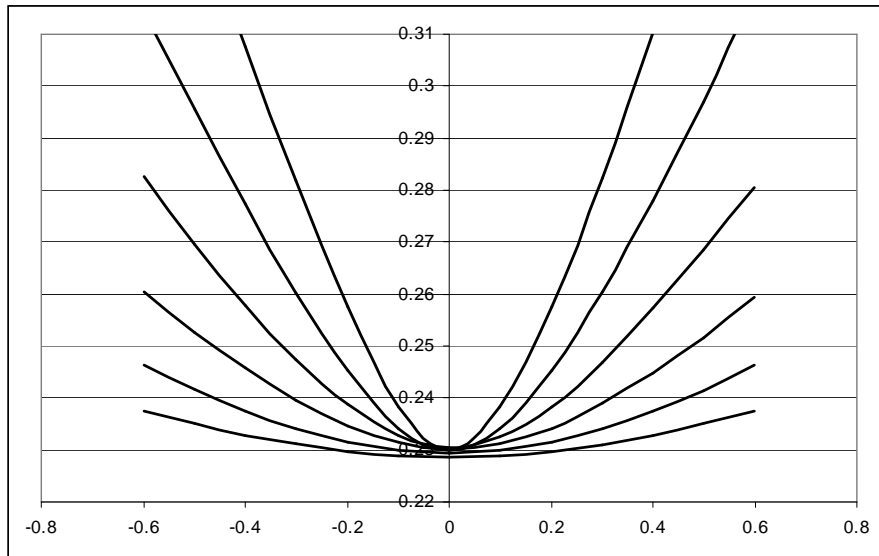


FIG. 3.17 -  $\alpha T = 2$ ; de bas en haut :  $\alpha = 0.25, 0.5, 1, 2, 4, 8$

$f$  est dans ce cas discontinue en 0; pour cette raison on distinguera le cas  $Y_0 = 0$  du cas  $Y_0 \neq 0$ .

On commence par traiter ce dernier cas, en choisissant  $Y_0 = 0.1$ . Dans ce cas,  $f(Y)$  commence par valoir  $\sigma_+$ . Si la force de rappel  $\alpha$  est petite devant  $\frac{1}{T}$ , les trajectoires  $(Y_t)_{0 \leq t \leq T}$  ont de bonnes chances de rester dans les positifs et on s'attend à voir un smile qui s'applatit sur  $\sigma_+$  quand  $\alpha \rightarrow 0^+$ . En revanche, si  $\alpha$  est grand devant  $\frac{1}{T}$ , les trajectoires  $(Y_t)_{0 \leq t \leq T}$  perdent la mémoire de la condition initiale et, quand  $\alpha \rightarrow +\infty$ , pour tout  $t$ , la loi de  $Y_t$  tend vers la loi normale centrée autour de 0 et de variance  $\nu^2$ . Au vu de la théorie ergodique détaillée plus haut, on s'attend donc à voir dans cette limite un smile qui s'applatit sur

$$\bar{\sigma} = \sqrt{\frac{\sigma_-^2 + \sigma_+^2}{2}} < \sigma_+.$$

On vérifie ces prédictions sur la figure 3.18, construite avec les valeurs suivantes des



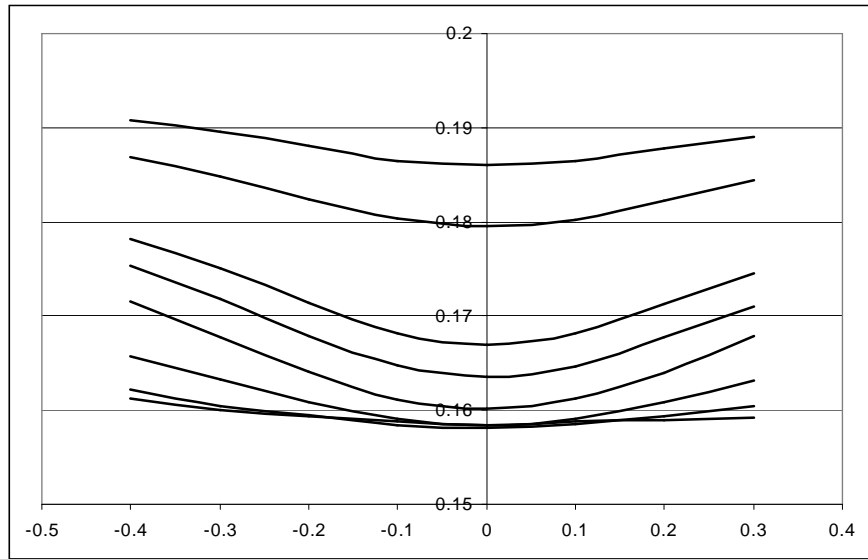


FIG. 3.18 – De bas en haut :  $\alpha = 20, 10, 5, 2, 1, 0.5, 0.1, 0.05$

paramètres :

$N$	250 000
$x$	1
$r$	0
$f(y)$	$\sigma_- \mathbf{1}_{\{y \leq 0\}} + \sigma_+ \mathbf{1}_{\{y > 0\}}$
$\sigma_+$	0.2
$\sigma_-$	0.1
$Y_0$	0.1
$\nu$	0.5
$\rho$	0
$T$	1

De nouveau il existe une valeur critique  $\alpha_c$  pour laquelle la courbure à la monnaie est maximale.

D'après la figure 3.19, la volatilité implicite à la monnaie semble encore un invariant de  $\alpha T$  dans ce cas.

Dans le cas où  $Y_0 = 0$ , même pour des  $\alpha$  très petits, les trajectoires  $(Y_t)_{0 \leq t \leq T}$  quittent l'état 0, soit par le haut, soit par le bas, ni  $\sigma_+$  ni  $\sigma_-$  n'est privilégiée ; il n'y a donc aucune raison pour que le smile s'applatisse quand  $\alpha \rightarrow 0^+$ . Au contraire, plus  $\alpha$  est petit, plus les trajectoires qui commencent par grimper dans les  $y > 0$  vont avoir tendance à y rester, de même pour celles qui commencent par descendre dans les  $y < 0$ . Ce phénomène a tendance à accroître l'écart de la loi de  $X_T$  à une log-normale, même si la variance de  $Y_t$  est de plus en plus petite. On observe bien ceci sur la figure 3.20.

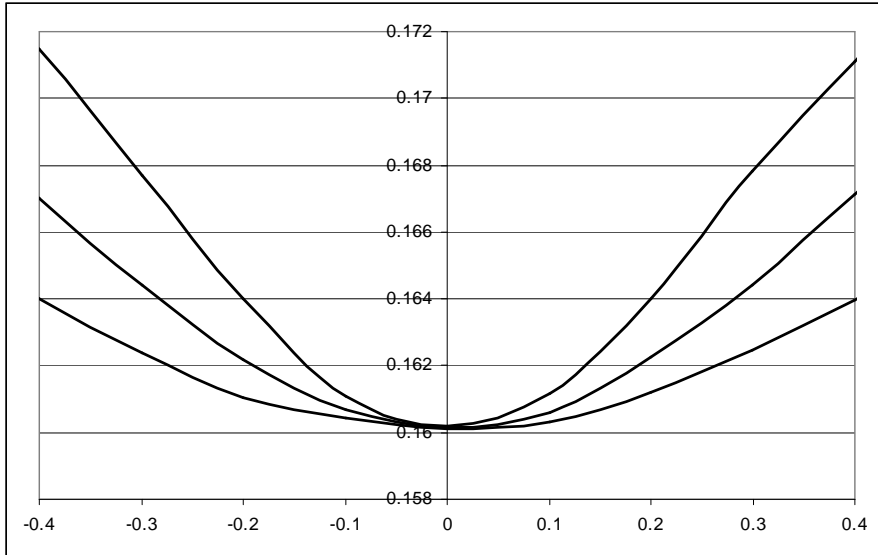


FIG. 3.19 –  $\alpha T = 2$  ; de bas en haut :  $T = 4, 2, 1$

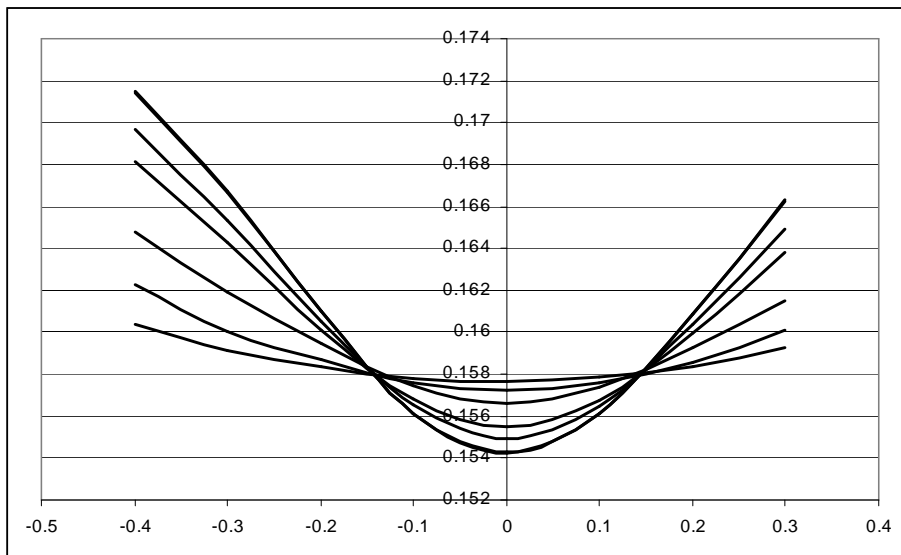


FIG. 3.20 – A la monnaie, de bas en haut :  $\alpha = 0.05, 0.1, 1, 2, 5, 10, 20$  ; les deux courbes pour  $\alpha = 0.05$  et  $\alpha = 0.1$  sont presque confondues

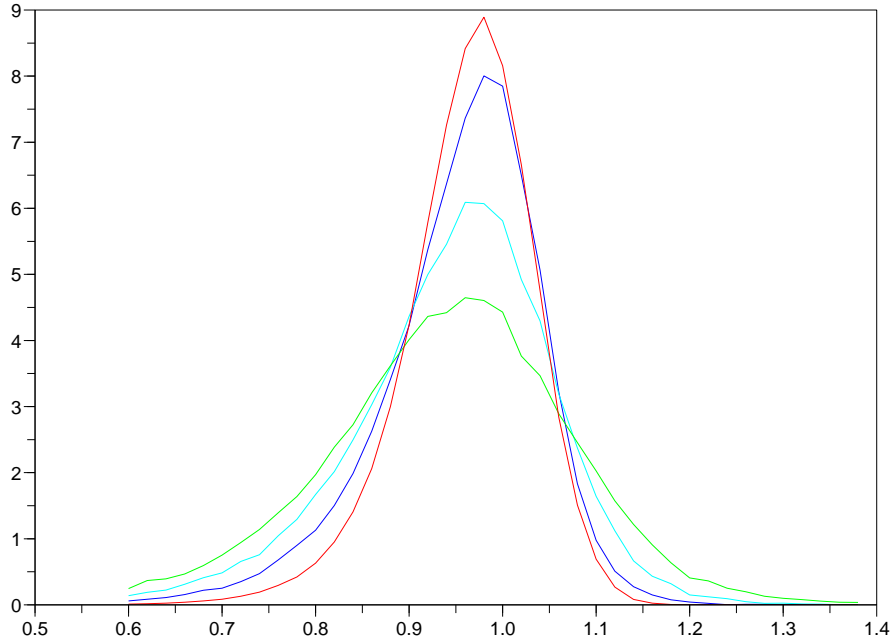


FIG. 3.21 – Du plus au moins piqué :  $\alpha = 0.01, 1, 4, 10$

On semble avoir atteint le smile limite quand  $\alpha \rightarrow 0^+$  puisque les deux smiles générés pour les valeurs  $\alpha = 0.1$  et  $\alpha = 0.05$  sont confondus; ce sont ceux qui ont la plus grande courbure; on peut dire ici que  $\alpha_c = 0^+$ .

**Simulation de la densité de  $\bar{X}_T$**  Enfin, avec Majd CHEIKH-ALI, élève de l'ENSTA dont j'ai encadré le stage de 2ème année au CERMICS, nous avons simulé la densité de  $\bar{X}_T$ , en utilisant directement le schéma d'Euler (3.39) et les poids de Malliavin proposés par A. KOHATSU-HIGA et R. PETERSON [60]. La figure 3.21 montre l'étalement de la densité de  $\bar{X}_T$  lorsque  $\alpha$  croît.

### 3.6 Conclusion

Les simulations numériques montrent que le modèle (3.1) est capable de reproduire les smiles observés sur les marchés. En particulier, dans le cas où  $f(y) = \sigma_f e^y$  :

- L'amplitude du smile est gouvernée par les paramètres de force de rappel,  $\alpha$ , et de "vol de vol"  $\nu$ . Elle croît avec la "vol de vol" et est maximale quand la

force de rappel est de l'ordre de l'inverse de l'échelle de temps du problème.

- Le "skew", c'est-à-dire la dissymétrie du smile par rapport à la monnaie (où la pente est négative), s'obtient en corrélant négativement les deux browniens  $W$  et  $\hat{Z}$ .

Les simulations indiquent aussi une instabilité de la volatilité implicite lorsqu'on s'approche de la maturité de l'option ( $T \rightarrow 0$ ), le smile semblant dans ce cas avoir une convexité de plus en plus grande (cf figure 3.16). Dans cette situation où on s'approche de la maturité de l'option, lorsqu'on fait croître artificiellement la force de rappel  $\alpha$  de sorte que  $\alpha T$  reste constant, on observe encore ce phénomène de courbure du smile, mais la volatilité implicite à la monnaie, elle, reste stable, si bien que le modèle (3.1) semble définir un prix des calls à la monnaie qui ne dépend de la force de rappel et de la maturité que par leur produit (cf figure 3.17).

D'un point de vue pratique, après avoir supposé  $\Lambda = 0$  par exemple, il convient de fixer un certain nombre de paramètres raisonnables, par exemple  $\nu$  et  $f$  - donc  $m$  -, et chercher les trois paramètres  $\alpha$ ,  $\rho$  et  $Y_0$  qui permettent d'approcher au mieux - en un sens à définir - le smile lu aujourd'hui. Les formules de calibration (3.37) ne permettent de traiter que les cas où l'on observe un smile de très faible amplitude. Dans tous les autres cas, fréquents, il faudra utiliser des méthodes numériques, par exemple la méthode de Monte-Carlo que nous avons détaillée, ou bien des schémas de discrétisations spatiale et temporelle de l'EDP d'évaluation (3.17), et implémenter des méthodes d'estimation. Les paramètres étant estimés, le modèle est dit "calibré" aux données de marché et permet d'estimer le prix des options exotiques, c'est-à-dire des options échangées de gré à gré, dont la valeur n'est pas fixée par le jeu de l'offre et de la demande, comme c'est le cas pour les options classiques (on dit aussi "vanille"). Ainsi les modèles à volatilité stochastique ont l'avantage sur le modèle de Black-Scholes d'être cohérent avec les smiles de marché et d'offrir un paramétrage de l'incertitude sur la volatilité qui permet d'évaluer, pour les options exotiques, la surfacturation due à ce risque de volatilité.

## Chapitre 4

# Théorèmes limites pour les chaînes de Markov bifurcantes. Application à la détection du vieillissement cellulaire

Dans la dernière partie de cette thèse, on s'intéresse aux chaînes de Markov indexées par un arbre binaire et à leur utilisation en biologie. On reproduit ci-dessous un article soumis pour publication, augmenté de quelques graphiques et des commentaires afférents.

We propose a general method to study dependent data in a binary tree, where an individual in one generation gives rise to two different offspring, one of type 0 and one of type 1, in the next generation. For any specific characteristic of these individuals, we assume that the characteristic is stochastic and depends on its ancestors' only through the mother's characteristic. The dependency structure may be described by a transition probability  $P(x, dydz)$  which gives the probability that the pair of daughters' characteristics is around  $(y, z)$  given that the mother's characteristic is  $x$ . Note that  $y$ , the characteristic of the daughter of type 0, and  $z$ , that of the daughter of type 1, may be conditionally dependent given  $x$ , and their respective conditional distributions may differ. We then speak of bifurcating Markov chains.

We derive laws of large numbers and central limit theorems for such stochastic processes. We then apply these results to detect cellular aging in *Escherichia Coli*, using the data of E. J. STEWART *et al.* and a bifurcating autoregressive model.

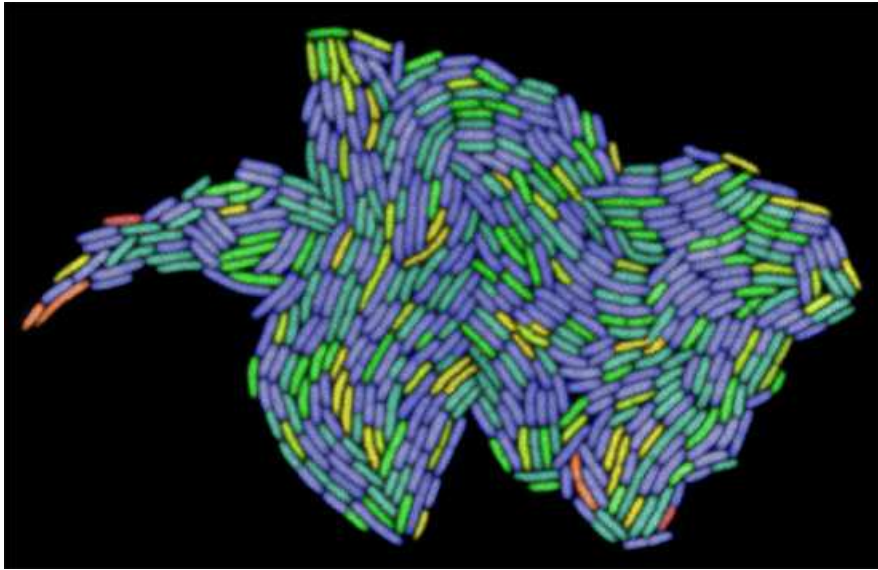


FIG. 4.1 – "What is true for *E. coli* is true for the elephant" As aging requires a differential distribution of aged and young components between parent and offspring, it has been postulated that organisms that reproduce by symmetrical division do not age. It has now been shown, however, that in the symmetrically dividing bacterium *Escherichia coli* aging can occur even in the absence of both a visible asymmetry and a juvenile phase. This false-color image shows a growing microcolony of *E. coli*, where the cells are colored by age (in numbers of divisions). The oldest cells are red, and the youngest blue. Original quote attributed to Jacques Monod. Cover credit: Photograph provided by Eric Stewart and Stefanie Timmermann, Inserm U571, Faculté de Médecine Necker Enfants-Malades, Paris, France.

## 4.1 Introduction

### 4.1.1 Motivation

This study has been motivated by a collaboration [42] with biologists from the Laboratoire de Génétique moléculaire, évolutive et médicale (INSERM U571, Faculté de Médecine Necker, Paris). F. TADDÉI, E. J. STEWART, A. LINDNER and G. PAUL, together with R. MADDEN from the Institut des Hautes Études Scientifiques, have been working on *Escherichia Coli*'s aging. *E. Coli* is a single-celled, model organism. It has been widely studied by the biologists who have gathered a large amount of information on its physiology. Whereas aging is obvious in macroscopic organisms, it is not in single-celled ones, where nevertheless one has the best chances of describing and quantifying the molecular process involved. It is especially hard to identify in *E. Coli*, which reproduces without a juvenile phase and with an apparently symmetric division. E. J. STEWART *et al.* [89] have designed an experimental protocol which brings evidence of aging in *E. Coli* and we propose a statistical study of the data they collected.

In this section, we describe the biological experiment and present the data (Subsection 4.1.2). Inter-experiment averaging shows a clear segregation between the new and old poles (see Subsection 4.1.2), whereas single-experiment data does not. In Subsection 4.1.3, we propose a linear Gaussian model that allows to study the populations of old and new poles experiment-wise. The model consists of a bifurcating markovian dynamics. This motivates Section 4.2, where we give a detailed study of such stochastic processes. We pay special attention to limit theorems such as laws of large numbers (Theorems 35, 38 and 41) and central limit theorems (Theorem 46 and its corollaries). Eventually, in Section 4.3, we apply these results to the model, proving strong laws of large numbers and a central limit theorem (see Propositions 54 and 55), and derive rigorous estimation and test procedures which are performed on the data in order to detect cellular aging.

### 4.1.2 The biology

Here we briefly describe *E. Coli*'s life cycle, the experiment designed by E. J. STEWART *et al.* and the data they get. Figure 4.2 is taken from [89] where one can find further information.

#### The experiment

*E. Coli* is a rod-shaped bacterium. It reproduces by dividing in the middle, producing a new end per progeny cell (see Figure 4.2). This new end is called the new pole, whereas the other end is pre-existing and is called the old pole.

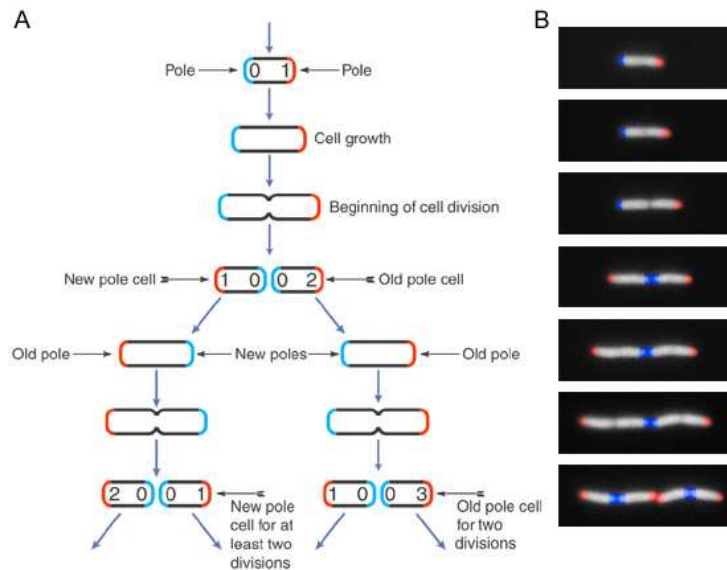


FIG. 4.2 – The life cycle of *E. coli*, from E. J. STEWART *et al.* [89]

This defines an age in divisions for each pole, and hence for each cell. One expects any cell component formed in the poles and with limited diffusion to accumulate at the old pole, so that there might be a physiological asymmetry between the old and new poles. To determine if *E. Coli* experiences aging related to the inheritance of the old pole, E. J. STEWART *et al.* followed 95 individual exponentially growing cells through up to nine generations in an automated fluorescence microscopy system which allowed them to determine the complete lineage, the identity of each pole and, among other physical parameters, the growth rate of each cell. Let us now present their results.

### Original data

Each of the 95 films gives rise to a genealogical tree such as the ones in Figure 4.3. The new poles are the solid lines and the old poles the dashed lines. On the  $y$ -axis appears the growth rate, whereas the  $x$ -axis displays time in divisions. There is no striking evidence for reproductive asymmetry between the progeny cells visible to the naked eye. Note that generally the data is not regular: some generations are not completely observed, and in few cases a cell's growth rate might be measured whereas her sister's is not.

### Averaged data

In order to eliminate the random effects which appear in Figure 4.3, E. J. STEWART *et al.* have averaged the 95 lineages by each unique cell position within the



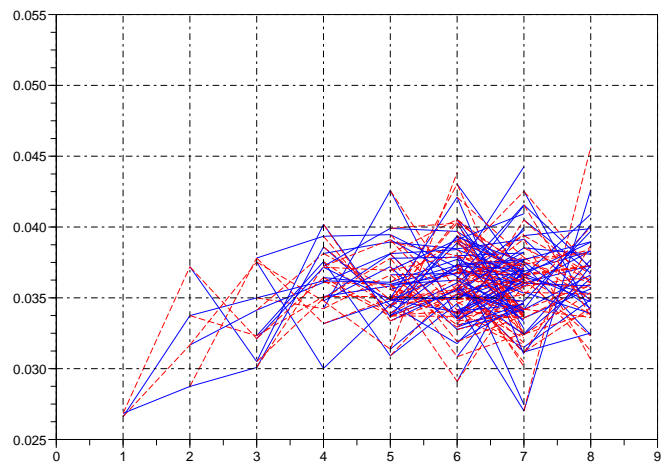
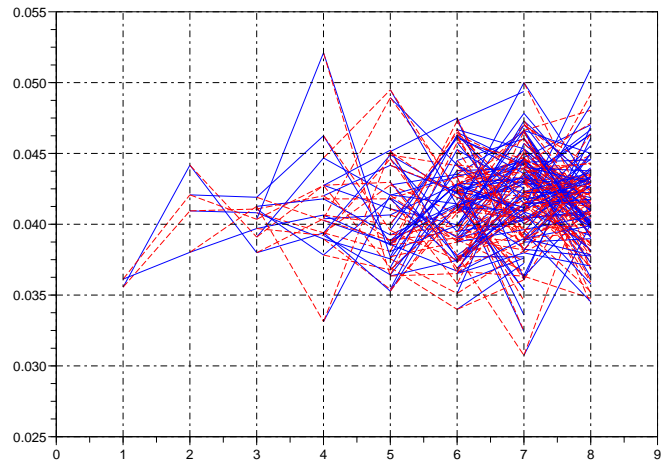


FIG. 4.3 – Two single-experiment data trees (two films)

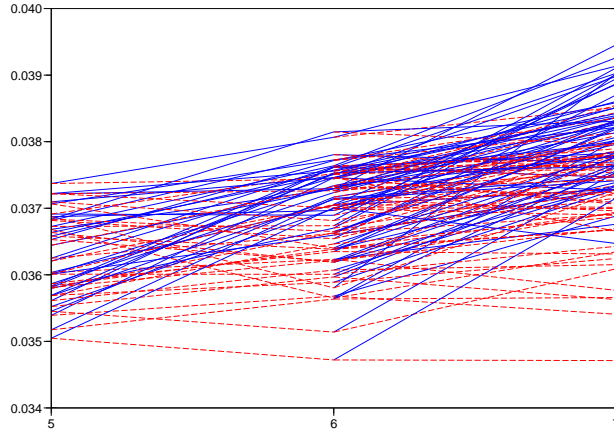


FIG. 4.4 – The average data tree

lineage. Figure 4.4 is the average tree thus produced for generations 5, 6 and 7. It clearly shows a segregation between the new and old poles. The old poles grow slower than the old poles, which is evidence for aging.

In Figure 4.5, each branch point is an average cell for that position in the lineage, and the length of the lines connecting cells to their progeny are proportional to the growth rate of the cell. At each division, the cell inheriting the old pole of the progenitor cell is on the right branch, in red. Again there is striking evidence for cumulatively slowed growth in the old poles.

Eventually, E. J. STEWART *et al.* performed a pair-wise comparison of the 7,953 sets of sister cells that were produced at the eighth generation in each of the 95 films, see Figure 4.6. Their analysis indicates that the average growth rate of the old pole progeny cell is 2.2% slower than that of the new pole cell, see [89] for details.

However, we would prefer to study each experiment separately, since we do not know if the experiments are independent and/or identically distributed. Indeed, two initial cells giving rise to two different films are actually taken from the same macrocolony, so that there might be a correlation between the experiments. Moreover, as shown in Figure 4.3, the range of values of the growth rate changes from film to film, probably due to a slight change in the experimental conditions. In the next section, we propose a statistical model that allows us to study the populations of old and new poles experiment-wise. It also has the advantage of taking into account the structure of the dependencies within a lineage. To be precise, contrary to E. J. STEWART *et al.*, we take the effect of the mother into account, and we will prove that this effect is important (see Remark 66).

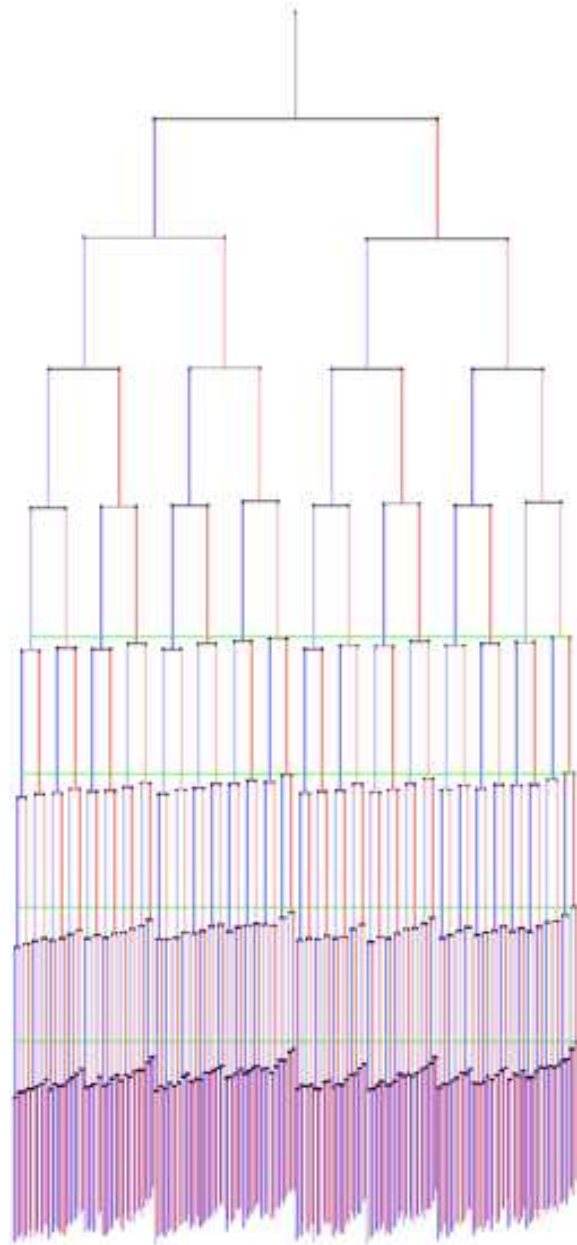


FIG. 4.5 – Average lineage showing old pole effect on growth rate, from E. J. STEWART *et al.* [89]

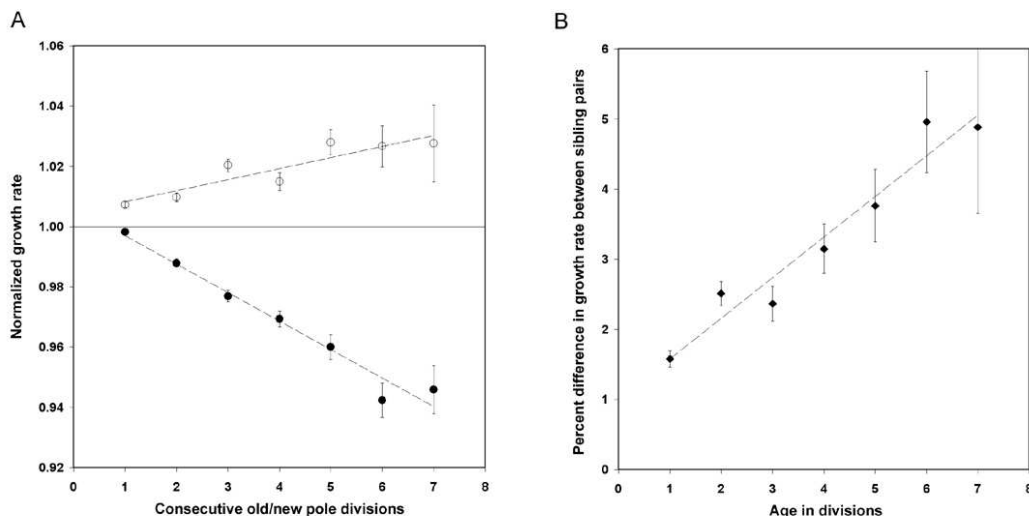


FIG. 4.6 – The effects of consecutive divisions as an old or new pole on growth rate, from E. J. STEWART *et al.* [89]

### 4.1.3 The mathematical model

In order to describe the dynamics of the growth rate, let  $X_i$  denote the growth rate of individual  $i$  and  $n$  denote the mother of  $2n$  - the new pole progeny cell - and  $2n + 1$  - the old pole progeny cell, see Figure 4.7. We propose the following markovian model with memory one:  $X_1$ , the ancestor's growth rate, has distribution  $\nu$  and for all  $n \geq 1$ ,

$$\begin{cases} X_{2n} &= \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} &= \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \quad (4.1)$$

where  $\alpha_0, \alpha_1 \in (-1, 1)$ ,  $\beta_0, \beta_1 \in \mathbb{R}$  and  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$  forms a sequence of independent and identically distributed (i.i.d.) centered bivariate Gaussian random variables (r.v.), say  $(\varepsilon_{2n}, \varepsilon_{2n+1}) \sim \mathcal{N}_2(0, \Gamma)$  with

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1)$$

( $\varepsilon_{2n}$  and  $\varepsilon_{2n+1}$  are thus supposed to have common variance  $\sigma^2$ ). We speak of memory one because a cell's growth rate is explained by its mother's. For instance, a markovian model with memory two would also take into account the grandmother's growth rate. We regard model (4.1) as "the simpler" reasonable model which describes a dependency within the colony.

**Remark 31.** Since a Gaussian r.v. may take arbitrarily big negative values, here we allow the growth rate to take negative values. However, provided we correctly estimate the parameters, this should happen with extremely small probability.

We aim at

- (1) estimating the 4-dimensional parameter  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ ,  $\rho$  and  $\sigma^2$ ,
- (2) testing the null hypothesis  $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$  against its alternative  $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$ .

In view of the biological question addressed here, point (2) is crucial: rejecting  $H_0$  comes down to accepting that the dynamics of the growth rate of the old pole offspring is different from that of the new pole offspring. We shall actually see the old pole progeny cell experiences slowed growth rate and hence should be considered an aging parent repeatedly producing rejuvenated offspring.

Bifurcating autoregressive (BAR) models, such as model (4.1), have already been studied. R. COWAN and R. STAUDTE [20] were pioneers and studied model (4.1) in the special case when  $(\alpha_0, \beta_0) = (\alpha_1, \beta_1)$ , i.e. under  $H_0$ . Several extensions [5, 16, 17, 49, 50, 51, 87] followed, improving inference results or/and generalizing the model, but no distinction was never made between new and old poles. In mathematical terms, in all these papers, the distribution of  $X_{2n}$  given  $X_n$  has always been assumed to be the same as the distribution of  $X_{2n+1}$  given  $X_n$ . Now, detecting a discrepancy between these distributions is the central question addressed here. Hence, model (4.1) generalizes existing BAR models and allows us to detect dissymmetry between sisters. Such a generalization is a source of mathematical difficulties. For instance, there is no stationary distribution in the sense of [20], that is to say a distribution common to all cells in the clone. This leads us to develop a new theory (see Section 4.2).

We shall call  $X = (X_n, n \geq 1)$  a bifurcating Markov chain. The next section is devoted to the study of this family of stochastic processes. Establishing laws of large numbers and central limit theorems will be crucial in achieving the two above objectives. That is the reason why we will pay special attention to such limit theorems.

## 4.2 Bifurcating Markov chains. Limit theorems

### 4.2.1 Definitions

Markov chains (MCs) are usually indexed by the integers. Here we give a definition of a MC when the index set is the (regular) binary tree  $\mathbb{T}$ . We then speak of bifurcating Markov chain or  $\mathbb{T}$ -Markov chain, which we often write  $\mathbb{T}$ -MC.  $\mathbb{T}$ -MCs are well adapted to modeling data on the descent of an initial individual, where each individual in one generation gives rise to two offspring in the next one. Cell lineage data, such as the one presented in Section 4.1.2, are typically of this kind.

Let us introduce some notation about the binary tree  $\mathbb{T}$ , see Figure 4.7. Each vertex  $n \in \mathbb{T}$  is seen as a positive integer  $n \in \mathbb{N}^*$ . It should be thought of as an individual or a cell. It has exactly two daughters,  $2n$  and  $2n + 1$ , and we label

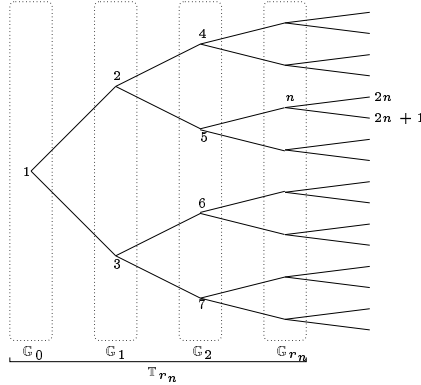


FIG. 4.7 – The binary tree  $\mathbb{T}$

the root 1. We denote by  $\mathbb{G}_q = \{2^q, 2^q + 1, \dots, 2^{q+1} - 1\}$  the  $q$ -th generation and by  $\mathbb{T}_r = \cup_{q=0}^r \mathbb{G}_q$  the subtree consisting of the first  $r+1$  generations. With this notation,  $\mathbb{G}_0 = \{1\}$  and,  $|\cdot|$  standing for the cardinality,  $|\mathbb{G}_q| = 2^q$  and  $|\mathbb{T}_r| = 2^{r+1} - 1$ . We also denote by  $r_n = \lfloor \log_2 n \rfloor$  the generation of individual  $n$ , i.e.  $n \in \mathbb{G}_{r_n}$ . In terms of labelling the vertices,  $\mathbb{T}$  is assimilated to  $\mathbb{N}^*$ , but the topology is different: within  $\mathbb{T}$ ,  $n$  and  $2n$  (resp.  $n$  and  $2n+1$ ) should be seen as neighbours.

Let  $(S, \mathcal{S})$  be a metric space endowed with its Borel  $\sigma$ -field, and think of it as the state space. For instance, in the BAR model (4.1),  $S = \mathbb{R}$ . For any integer  $p \geq 2$  we equip  $S^p$  with the product  $\sigma$ -field, say  $\mathcal{S}^p$ .

**Definition.** We call  $\mathbb{T}$ -transition probability any mapping  $P : S \times \mathcal{S}^2 \rightarrow [0, 1]$  such that

- $P(\cdot, A)$  is measurable for all  $A \in \mathcal{S}^2$ ,
- $P(x, \cdot)$  is a probability measure on  $(S^2, \mathcal{S}^2)$  for all  $x \in S$ .

We also define, for  $x \in S$  and  $B \in \mathcal{S}$ ,  $P_0(x, B) = P(x, B \times S)$  and  $P_1(x, B) = P(x, S \times B)$ .  $P_0$  and  $P_1$  are transition probabilities on  $(S, \mathcal{S})$ . In the BAR model (4.1), they respectively correspond to the transition probabilities of the new poles and of the old poles.

For  $p \geq 1$ , we denote by  $\mathcal{B}(S^p)$  (resp.  $\mathcal{B}_b(S^p)$ ,  $\mathcal{C}(S^p)$ ,  $\mathcal{C}_b(S^p)$ ) the set of all  $\mathcal{S}^p$ -measurable (resp.  $\mathcal{S}^p$ -measurable and bounded, continuous, continuous and bounded) mappings  $f : S^p \rightarrow \mathbb{R}$ . If  $p \in \{2, 3\}$  and  $f \in \mathcal{B}(S^p)$ , when it is defined, we denote by  $Pf \in \mathcal{B}(S)$  the function

$$x \mapsto Pf(x) = \begin{cases} \int_{S^2} f(y, z) P(x, dydz) & \text{if } p = 2, \\ \int_{S^2} f(x, y, z) P(x, dydz) & \text{if } p = 3. \end{cases}$$

Let  $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$  be a filtered probability space and, defined on it,  $(X_n, n \in \mathbb{T})$  be a family of  $S$ -valued random variables. Let  $\nu$  be a probability on  $(S, \mathcal{S})$  and  $P$  be a  $\mathbb{T}$ -transition probability.

**Definition.** We say that  $(X_n, n \in \mathbb{T})$  is a  $(\mathcal{F}_r)$ -bifurcating Markov chain, or  $(\mathcal{F}_r)$ - $\mathbb{T}$ -MC (with initial distribution  $\nu$  and  $\mathbb{T}$ -transition probability  $P$ ), if

- $X_n$  is  $\mathcal{F}_{r_n}$ -measurable for all  $n \in \mathbb{T}$ ,
- $X_1$  has distribution  $\nu$ ,
- for all  $q \in \mathbb{N}$  and for all family  $(f_n, n \in \mathbb{G}_q)$  in  $\mathcal{B}_b(S^2)$ ,

$$\mathbb{E} \left[ \prod_{n \in \mathbb{G}_q} f_n(X_{2n}, X_{2n+1}) \middle| \mathcal{F}_q \right] = \prod_{n \in \mathbb{G}_q} P f_n(X_n).$$

This means that, given generations 0 to  $q$ ,  $\mathbb{T}_q$ , one builds generation  $\mathbb{G}_{q+1}$  by drawing  $2^q$  *independent* couples  $(X_{2n}, X_{2n+1})$  according to  $P(X_n, \cdot)$  ( $n \in \mathbb{G}_q$ ). A  $(\mathcal{F}_r)$ - $\mathbb{T}$ -MC is also a  $(\mathcal{F}_r^X)$ - $\mathbb{T}$ -MC, where  $\mathcal{F}_r^X = \sigma(X_i, i \in \mathbb{T}_r)$ . When unstated, the filtration implicitly used is the latter. Note that for  $f \in \mathcal{B}_b(S^2)$ ,  $\mathbb{E}[f(X_{2n}, X_{2n+1}) | \mathcal{F}_{r_n}]$  factorizes through  $X_n$ , so that  $\mathbb{E}[f(X_{2n}, X_{2n+1}) | \mathcal{F}_{r_n}] = \mathbb{E}[f(X_{2n}, X_{2n+1}) | X_n]$ . This means that any  $X_n$  depends on past generations only through her mother. This explains why we speak of a Markov chain (with memory one).

Last but not least, note that, contrary to much of the (still sparse) literature on the subject, we allow conditional dependency between sisters. Conditional independance corresponds to the case when  $P$  factorizes as a product  $P_0 \otimes P_1$  of two transition probabilities on  $(S, \mathcal{S})$ , that is  $P(x, dydz) = P_0(x, dy) \otimes P_1(x, dz)$  for all  $x \in S$ . [2, 6, 70, 98, 71, 73, 91] deal with more general than binary -and possibly random- trees but all assume that, conditionally on their mother's type, the daughters have independant and identically distributed types. In our case, this corresponds to conditional independency with  $P_0 = P_1$ . As said in Section 4.1.3, to our best knowledge, BAR models, although they allow for conditional dependence, have always been studied until now under the assumption that  $P_0 = P_1$ . Now, detecting that  $P_0 \neq P_1$  will be the central question when we study cellular aging (see Section 4.3). Moreover, like in the second part of [2], we consider general state spaces whereas the first part of [2] and [6] deal only with countable ones and [70, 98, 71, 73, 91] only with finite ones. Note that in the latter case, one may regard a  $\mathbb{T}$ -MC  $X$  as a multitype branching process and apply F. MAËOUÏA and A. TOUATI's identification techniques [73].

## 4.2.2 Weak law of large numbers

### Introduction

A first natural question is to know whether a  $\mathbb{T}$ -Markov chain  $(X_n)$  obeys laws of large numbers (LLN), that is convergence of empirical means.

Given  $f \in \mathcal{B}(S)$  and a finite subset  $I \subset \mathbb{T}$ , let us write  $M_I(f) = \sum_{i \in I} f(X_i)$  and  $\overline{M}_I(f) = |I|^{-1} M_I(f)$ . Several empirical averages can be considered:

- One may average over the  $q$ -th generation, i.e. compute  $\overline{M}_{\mathbb{G}_q}(f)$ .
- One may prefer to average over the first  $r+1$  generations, i.e. compute  $\overline{M}_{\mathbb{T}_r}(f)$ . This is meaningful because  $\mathbb{G}_q$  naturally precedes  $\mathbb{G}_{q+1}$ , since one cannot draw the whole  $(q+1)$ -th generation without having completely drawn the  $q$ -th one.
- One may also average over the “first”  $n$  individuals, i.e. compute  $n^{-1} \sum_{i=1}^n f(X_i)$ . However, there is no natural order within a generation  $\mathbb{G}_q$ : all the individuals  $(X_n, n \in \mathbb{G}_q)$  of the  $q$ -th level can be generated simultaneously. That is why we introduce the set  $\mathfrak{S}$  of all permutations of  $\mathbb{N}^*$  that leave each  $\mathbb{G}_q$  invariant and, for  $f \in \mathcal{B}(S)$  and  $\pi \in \mathfrak{S}$ , consider the sums

$$M_n^\pi(f) = \sum_{i=1}^n f(X_{\pi(i)}).$$

As far as the asymptotic behaviour of  $\overline{M}_n^\pi(f) \equiv n^{-1}M_n^\pi(f)$  is concerned, the choice of  $\pi$  matters. To illustrate this, consider the following example.

**Example 32.** Assume that  $S = \{0, 1\}$ ,  $f = \text{id}_S$  and, whatever the mother’s type,  $X_{2n} = 1$  and  $X_{2n+1} = 0$  - in other words,  $P(x, dydz) = \delta_1(dy)\delta_0(dz)$  for all  $x \in S$ , where  $\delta_x$  stands for the Dirac mass at point  $x$ . If  $\pi \in \mathfrak{S}$  sends the “first half” of each  $\mathbb{G}_q$ , i.e.  $\{2^q, 2^q + 1, \dots, 3 \cdot 2^{q-1} - 1\}$ , onto the even elements in  $\mathbb{G}_q$ , i.e.  $\mathbb{G}_q \cap (2\mathbb{N})$ , then  $\liminf_{n \rightarrow \infty} \overline{M}_n^\pi(f) = 1/2$  and  $\limsup_{n \rightarrow \infty} \overline{M}_n^\pi(f) = 2/3$ : the empirical average  $\overline{M}_n^\pi(f)$  oscillates between  $1/2$  and  $2/3$ . Conversely, if  $\pi$  sends the “first half” of each  $\mathbb{G}_q$  onto the odd elements in  $\mathbb{G}_q$ , i.e.  $\mathbb{G}_q \cap (2\mathbb{N} + 1)$ , then  $\liminf_{n \rightarrow \infty} \overline{M}_n^\pi(f) = 1/3$  and  $\limsup_{n \rightarrow \infty} \overline{M}_n^\pi(f) = 1/2$ . But for  $\pi = \text{id}_{\mathbb{N}^*}$ ,  $\overline{M}_n^\pi(f)$  converges to  $1/2$ .

- A natural answer to this issue is to explore each new generation “by chance”, i.e. to draw a permutation  $\Pi$  “uniformly” on  $\mathfrak{S}$ , independently on  $X = (X_n, n \in \mathbb{T})$ . Drawing  $\Pi$  “uniformly” on  $\mathfrak{S}$  means drawing the restriction of  $\Pi$  on  $\mathbb{G}_q$  uniformly among the  $(2^q)!$  permutations of  $\mathbb{G}_q$ , independently for each  $q$ . Then we consider the empirical average  $\overline{M}_n^\Pi(f) \equiv n^{-1}M_n^\Pi(f)$ , where

$$M_n^\Pi(f) = \sum_{i=1}^n f(X_{\Pi(i)}).$$

Thus we introduce extra randomness, but this will allow us to get through Liapunov’s condition when we try to derive a central limit theorem for  $X$ .

**Remark 33.** Note that for all  $\pi \in \mathfrak{S}$  and  $r \geq 0$ ,  $M_{|\mathbb{T}_r|}^\pi(f) = M_{\mathbb{T}_r}(f)$ . Besides, for all  $r \geq 0$ ,

$$\overline{M}_{\mathbb{T}_r}(f) = \sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \overline{M}_{\mathbb{G}_q}(f) \tag{4.2}$$



and, for all  $\pi \in \mathfrak{S}$  and  $n \geq 1$ ,

$$\frac{1}{n} M_n^\Pi(f) = \sum_{q=0}^{r_n-1} \frac{|\mathbb{G}_q|}{n} \overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n} \sum_{i=2^{rn}}^n f(X_{\Pi(i)}) \quad (4.3)$$

(we systematically use the convention that a sum over an empty set is zero).

Because of the branching, empirical averages of  $\mathbb{T}$ -MCs may not behave like corresponding MCs' ones. Precisely, given a transition probability  $R$ , a LLN may hold for MCs with transition probability  $R$  and fail for  $\mathbb{T}$ -MCs with  $\mathbb{T}$ -transition probability  $R \otimes R$ . A very simple but crucial illustration of this is Example 34. At least for the case of finite state spaces, let us keep in mind that periodicity is problematic and that there is more to ask than irreducibility and recurrence for a  $\mathbb{T}$ -MC to obey a LLN.

**Example 34.** Consider the two-state MC, say  $S = \{0, 1\}$ , with  $R(0, \cdot) = \delta_1$  and  $R(1, \cdot) = \delta_0$ . For a MC  $Y$ , a LLN holds true: whatever the initial state, the empirical average  $\frac{1}{n} \sum_{i=1}^n Y_i$  converges to  $1/2$  when  $n$  grows to infinity. But for the corresponding  $\mathbb{T}$ -MC  $X$  with  $\mathbb{T}$ -transition probability  $P = R \otimes R$ , i.e.  $P(0, \cdot) = \delta_1 \otimes \delta_1$  and  $P(1, \cdot) = \delta_0 \otimes \delta_0$ , it endlessly fluctuates from  $1/3$  to  $2/3$ . Indeed, assume that  $X_1 = 0$ . Then  $\overline{M}_{\mathbb{G}_q}(\text{id})$  is 0 when  $q$  is even and 1 when  $q$  is odd, so that (4.2) shows that

$$\overline{M}_{\mathbb{T}_{2^s}}(\text{id}) = \sum_{q=0}^{2^s} \frac{2^q}{2^{2s+1} - 1} \mathbf{1}_{2^{\mathbb{N}+1}}(q) = \sum_{l=0}^{s-1} \frac{2^{2l+1}}{2^{2s+1} - 1} = \frac{1}{3} \frac{2^{2s+1} - 2}{2^{2s+1} - 1}$$

tends to  $1/3$ , whereas

$$\overline{M}_{\mathbb{T}_{2^{s+1}}}(\text{id}) = \sum_{q=0}^{2^{s+1}} \frac{2^q}{2^{2s+2} - 1} \mathbf{1}_{2^{\mathbb{N}+1}}(q) = \sum_{l=0}^s \frac{2^{2l+1}}{2^{2s+2} - 1} = \frac{2}{3}.$$

Conversely, if  $X_1 = 1$ , then  $\lim_{s \rightarrow \infty} \overline{M}_{\mathbb{T}_{2^s}}(\text{id}) = 2/3$  and  $\lim_{s \rightarrow \infty} \overline{M}_{\mathbb{T}_{2^{s+1}}}(\text{id}) = 1/3$ .

## Results

Here we ask ourselves whether the various empirical averages introduced in Section 4.2.2 converge, in quadratic mean, when the size of the tree grows to infinity and we then speak of weak LLN. A sufficient condition for a weak LLN to hold appears to be the ergodicity, see Definition 4.2.2, of the induced MC  $(Y_r, r \in \mathbb{N})$  defined as follows. Start from the root and recursively choose one of the two daughters tossing a balanced coin, independently on the  $\mathbb{T}$ -MC  $X$ . In mathematical terms,  $Y_0 = X_1$  and if  $Y_r = X_n$  then  $Y_{r+1} = X_{2n+\zeta_{r+1}}$  for a sequence of independent balanced Bernoulli r.v.  $(\zeta_q, q \in \mathbb{N}^*)$  independent on  $(X, \Pi)$ . Here 'balanced' means that

$\mathbb{P}(\zeta_q = 0) = \mathbb{P}(\zeta_q = 1) = 1/2$ . It is easy to check that  $(Y_r, r \in \mathbb{N})$  is a MC with initial distribution  $\nu$  and transition probability

$$Q = \frac{P_0 + P_1}{2}.$$

**Definition.** We say that a MC  $Y$  is ergodic if there exists a probability  $\mu$  on  $(S, \mathcal{S})$  such that  $\lim_{r \rightarrow \infty} \mathbb{E}_x[f(Y_r)] = \int_S f d\mu$  for all  $x \in S$  and  $f \in \mathcal{C}_b(S)$ .

Then  $\mu$  is the unique stationary distribution of  $Y$ , and the sequence  $(Y_r, r \in \mathbb{N})$  converges in distribution to  $\mu$ . Sufficient conditions for ergodicity may be found in [9, 76]. We are now in the position to state the main theorem of this section:

**Theorem 35.** *Assume that the induced MC  $(Y_r, r \in \mathbb{N})$  is ergodic, with stationary distribution  $\mu$ . Then, for any  $f \in \mathcal{C}_b(S)$ , the three empirical averages  $\overline{M}_{\mathbb{G}_q}(f)$ ,  $\overline{M}_{\mathbb{T}_r}(f)$  and  $\overline{M}_n^{\Pi}(f)$  converge to  $(\mu, f)$  in  $L^2$ .*

**Remark 36.** It is noteworthy that the asymptotic behaviour of the three above empirical averages depends on the  $\mathbb{T}$ -transition probability  $P$  only through  $Q = (P_0 + P_1)/2$ .

**Remark 37.** K. ATHREYA and H.-J. KANG, in the second part of their article [2], use an analogous ergodicity hypothesis to get laws of large numbers. Namely, their results hold for Galton-Watson trees in which particles move according to a Markov chain  $R$  on  $(S, \mathcal{S})$ , and they assume  $\lim_{m \rightarrow \infty} R^m(x, \cdot) = \mu$ . If this happens uniformly in  $x$  on the compact subsets of  $S$ , they get a weak LLN; if this happens uniformly in  $x$  on  $S$ , they get a strong LLN. Observe that we do not assume any uniformity in  $x$  (but our tree is deterministic).

In the application (Section 4.3), the function  $f$  will typically be unbounded so that we shall actually prove an extended version of Theorem 35. To this end, let us first introduce some notation. We denote by:

- $i \wedge j$  the most recent common ancestor of  $i, j \in \mathbb{T}$ ,
- $f \otimes g$  the mapping  $(x, y) \mapsto f(x)g(y)$ ,
- $Q^p$  the  $p$ -th iterated of  $Q$ , recursively defined by the formulas  $Q^0(x, \cdot) = \delta_x$  and  $Q^{p+1}(x, B) = \int_S Q(x, dy)Q^p(y, B)$  for all  $B \in \mathcal{S}$ ;  $Q^p$  is a transition probability on  $(S, \mathcal{S})$ ,
- $\nu Q$  the distribution on  $(S, \mathcal{S})$  defined by  $\nu Q(B) = \int_S \nu(dx)Q(x, B)$ ;  $\nu Q^p$  is the law of  $Y_p$ ,
- $(Qf)(x) = \int_S f(y)Q(x, dy)$  when it is defined,
- $\nu(f)$  or  $(\nu, f)$  the integral  $\int_S f d\nu$  when it is defined.

With such a notation, for any distribution  $\lambda$ , transition probability  $Q$  and function  $f \in \mathcal{B}(S)$  such that  $\lambda Q(|f|) < \infty$ , we have  $\lambda Q(f) = \lambda(Qf)$  which is hence simply written  $\lambda Qf$ .

Now, let  $F$  denote a subspace of  $\mathcal{B}(S)$  such that

- (i)  $F$  contains the constants,
- (ii)  $F^2 \subset F$ ,
- (iii)  $F \otimes F \subset L^1(P(x, \cdot))$  for all  $x \in S$ , and  $P(F \otimes F) \subset F$ ,
- (iv) there exists a probability measure  $\mu$  on  $(S, \mathcal{S})$  such that  $F \subset L^1(\mu)$  and  $\lim_{r \rightarrow \infty} \mathbb{E}_x[f(Y_r)] = (\mu, f)$  for all  $x \in S$  and  $f \in F$ ,
- (v) for all  $f \in F$ , there exists  $g \in F$  such that for all  $r \in \mathbb{N}$ ,  $|Q^r f| \leq g$ ,
- (vi)  $F \subset L^1(\nu)$ ,

where we have used the notations  $F^2 = \{f^2 | f \in F\}$ ,  $F \otimes F = \{f \otimes g | f, g \in F\}$  and  $PE = \{Pf | f \in E\}$  whenever an operator  $P$  acts on a set  $E$ . Note that (i) and (iii) imply the condition

- (iii') for  $z = 0, 1$ ,  $F \subset L^1(P_z(x, \cdot))$  for all  $x \in S$  and  $P_z F \subset F$ ,

since  $P_0 f = P(f \otimes \mathbf{1})$  and  $P_1 f = P(\mathbf{1} \otimes f)$ . This in its turn implies

- (iii'')  $F \subset L^1(Q(x, \cdot))$  for all  $x \in S$  and  $QF \subset F$ ,

so that in (iv) and (v)  $\mathbb{E}_x[f(Y_r)] = Q^r f(x)$  is well defined. Note also that if  $F$  contains enough functions, *e.g.* if it contains the set of all bounded lipschitz functions, then  $\mu$  is the unique stationary distribution of  $Y$ , *i.e.*  $\mu Q = \mu$ . The next theorem states that the result in Theorem 35 remains true for  $f$ 's in such a  $F$ :

**Theorem 38.** *Let  $F$  satisfy conditions (i)-(vi) above. Then, for any  $f \in F$ , the three empirical averages  $\overline{M}_{\mathbb{G}_q}(f)$ ,  $\overline{M}_{\mathbb{T}_r}(f)$  and  $\overline{M}_n^\Pi(f)$  converge to  $(\mu, f)$  in  $L^2$ .*

Obviously  $F = \mathcal{C}_b(S)$  fulfils conditions (i)-(vi) as soon as  $Y$  is ergodic, so that Theorem 38 implies Theorem 35. In Section 4.3 we shall take  $F$  to be the set of all continuous and polynomially growing functions.

We shall also need an easy extension of Theorem 38 to the case when  $f$  does not only depend on an individual  $X_i$ , but on the mother-daughters triangle  $(X_i, X_{2i}, X_{2i+1})$ . This will be useful in the application (Section 4.3). Let us denote  $\Delta_n = (X_n, X_{2n}, X_{2n+1})$  and, for  $f \in \mathcal{B}(S^3)$  and  $I$  a finite subset of  $\mathbb{T}$ ,

$$M_I(f) = \sum_{i \in I} f(\Delta_i) \quad \text{and} \quad M_n^\Pi(f) = \sum_{i=1}^n f(\Delta_{\Pi(i)}).$$

Then we have

**Theorem 39.** *Let  $F$  satisfy conditions (i)-(vi). Let  $f \in \mathcal{B}(S^3)$  such that  $Pf$  and  $Pf^2$  exist and belong to  $F$ . Then the three empirical averages  $\overline{M}_{\mathbb{G}_q}(f)$ ,  $\overline{M}_{\mathbb{T}_r}(f)$  and  $\overline{M}_n^\Pi(f)$  converge to  $(\mu, Pf)$  in  $L^2$ .*

## Proofs

This section is devoted to the proofs of Theorems 38 and 39.

*Proof of Theorem 38.* Considering the function  $f - (\mu, f)$  leaves us with the case when  $(\mu, f) = 0$ . Then condition (iv) implies that

$$\forall x \in S, \quad \lim_{r \rightarrow \infty} Q^r f(x) = 0. \quad (4.4)$$

We shall study the three empirical averages  $\overline{M}_{\mathbb{G}_q}(f)$ ,  $\overline{M}_{\mathbb{T}_r}(f)$  and  $\overline{M}_n^{\text{II}}(f)$  successively.

**Step 1.** Let us first deal with  $\overline{M}_{\mathbb{G}_q}(f)$ . First note that  $f(X_i) \in L^2$  for all  $i \in \mathbb{G}_q$ . Indeed, there is a unique path  $(z_1, \dots, z_q) \in \{0, 1\}^q$  in the binary tree from the root 1 to  $i$ ; here  $(z_1, \dots, z_q)$  should be seen as the realisation of the coin toss r.v.  $(\zeta_1, \dots, \zeta_q)$  that joins 1 to  $i$ . For instance,  $(1, 0, 0, 1)$  is the path from 1 to 25. Thus

$$\mathbb{E}[f(X_i)^2] = \nu P_{z_1} \cdots P_{z_q} f^2,$$

which, from (ii), (iii') and (vi), is finite.

Independently on  $X$ , let us now draw two independent indices  $I_q$  and  $J_q$  uniformly from  $\mathbb{G}_q$ . Then  $f(X_{I_q})f(X_{J_q}) \in L^1$  and we have

$$\mathbb{E}[\overline{M}_{\mathbb{G}_q}(f)^2] = \frac{1}{|\mathbb{G}_q|^2} \sum_{i, j \in \mathbb{G}_q} \mathbb{E}[f(X_i)f(X_j)] = \mathbb{E}[f(X_{I_q})f(X_{J_q})].$$

Let us fix  $p \in \{0, \dots, q\}$  and reason conditionally on the event  $\{I_q \wedge J_q \in \mathbb{G}_p\}$ . Then  $I_q \wedge J_q$  is uniformly distributed on  $\mathbb{G}_p$ , so that  $X_{I_q \wedge J_q}$  has the same distribution as  $Y_p$ , i.e. has distribution  $\nu Q^p$ . Besides, for  $p < q$ , conditionally on the states  $(X_{2(I_q \wedge J_q)}, X_{2(I_q \wedge J_q)+1})$  of the two daughters of  $I_q \wedge J_q$ ,  $X_{I_q}$  and  $X_{J_q}$  are independent and have the same distribution as  $Y_{q-p-1}$  with respective initial conditions  $X_{2(I_q \wedge J_q)}$  and  $X_{2(I_q \wedge J_q)+1}$ . Provided we use the convention that  $P(Q^{-1}f \otimes Q^{-1}f) = f^2$ , we then have

$$\mathbb{E}[f(X_{I_q})f(X_{J_q}) | I_q \wedge J_q \in \mathbb{G}_p] = \nu Q^p P(Q^{q-p-1}f \otimes Q^{q-p-1}f). \quad (4.5)$$

Now,  $\mathbb{P}(I_q \wedge J_q \in \mathbb{G}_q) = \mathbb{P}(I_q = J_q) = 2^{-q}$  and, for  $p \in \{0, \dots, q-1\}$ ,  $\mathbb{P}(I_q \wedge J_q \in \mathbb{G}_p) = 2^{-p-1}$ . Indeed, since  $I_q$  and  $J_q$  are independent, the paths  $(\zeta_1^I, \dots, \zeta_q^I)$  from 1 to  $I_q$  and  $(\zeta_1^J, \dots, \zeta_q^J)$  from 1 to  $J_q$  are independent so that for  $p \in \{0, \dots, q-1\}$ ,

$$\mathbb{P}(I_q \wedge J_q \in \mathbb{G}_p) = \mathbb{P}(\zeta_1^I = \zeta_1^J, \dots, \zeta_p^I = \zeta_p^J, \zeta_{p+1}^I \neq \zeta_{p+1}^J) = 2^{-p-1}.$$

In short, we write  $\mathbb{P}(I_q \wedge J_q \in \mathbb{G}_p) = 2^{-p-1} \mathbf{1}_{\{p < q\}}$ . Combined with (4.5), this finally gives

$$\mathbb{E}[\overline{M}_{\mathbb{G}_q}(f)^2] = \mathbb{E}[f(X_{I_q})f(X_{J_q})] = \sum_{p=0}^q 2^{-p-1} \mathbf{1}_{\{p < q\}} \nu Q^p P(Q^{q-p-1}f \otimes Q^{q-p-1}f). \quad (4.6)$$

Let us now fix  $\varepsilon > 0$  and choose  $p_\varepsilon \in \mathbb{N}$  such that  $2^{-p_\varepsilon} \leq \varepsilon$ . Then  $\sum_{p > p_\varepsilon} 2^{-p} \leq \varepsilon$ . Besides, from (iii), (v) and (vi), there is a  $c_f \geq 0$  such that

$$|\nu Q^p P(Q^{q-p-1} f \otimes Q^{q-p-1} f)| \leq c_f \quad (4.7)$$

for all  $0 \leq p \leq q$ . Hence for  $q > p_\varepsilon$

$$\mathbb{E}[\overline{M}_{\mathbb{G}_q}(f)^2] \leq \varepsilon c_f + \sum_{p=0}^{p_\varepsilon} |\nu Q^p P(Q^{q-p-1} f \otimes Q^{q-p-1} f)|. \quad (4.8)$$

From (v), there exists  $g \in F$  such that for all  $r \in \mathbb{N}$ ,  $|Q^r f \otimes Q^r f| \leq g \otimes g$ . From (iii),  $P(g \otimes g) \in F$  so that, using (v) and (vi),  $P(g \otimes g) \in \cap_{p \in \mathbb{N}} L^1(\nu Q^p)$ . This shows that the sequence of functions  $(Q^r f \otimes Q^r f, r \in \mathbb{N})$  is dominated by  $g \otimes g \in \cap_{p \in \mathbb{N}} L^1(\nu Q^p P)$ . Then (4.4) and Lebesgue's dominated convergence theorem imply that

$$\forall p \in \mathbb{N}, \quad \lim_{r \rightarrow \infty} \nu Q^p P(Q^r f \otimes Q^r f) = 0. \quad (4.9)$$

As a consequence, the r.h.s. of (4.8) converges to  $\varepsilon c_f$  as  $q$  grows to infinity. Since  $\varepsilon$  is arbitrary, the proof is complete for  $\overline{M}_{\mathbb{G}_q}(f)$ .

Convergence results for  $\overline{M}_{\mathbb{T}_r}(f)$  or  $\overline{M}_n^\Pi(f)$  may be easily deduced from those for  $\overline{M}_{\mathbb{G}_q}(f)$  by using (4.2) or (4.3) and the following lemma.

**Lemma 40.** *Let  $(u_r, r \in \mathbb{N})$  be a sequence of nonnegative real numbers converging to 0. Let*

$$v_r = \sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} u_q \quad \text{and} \quad a_n = \sum_{q=0}^{r_n-1} \frac{|\mathbb{G}_q|}{n} u_q.$$

*Then  $(v_r, r \in \mathbb{N})$  and  $(a_n, n \in \mathbb{N}^*)$  converge to 0.*

*Proof.* Let us fix  $\varepsilon > 0$ . We can find  $q_\varepsilon \in \mathbb{N}^*$  such that  $u_q \leq \varepsilon$  for all  $q \geq q_\varepsilon$ . Letting  $M = \sup_{q \in \mathbb{N}} u_q$ , we then have, for all  $r \geq q_\varepsilon$ ,  $v_r \leq \varepsilon + M \sum_{q=0}^{q_\varepsilon-1} \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|}$ . The r.h.s. tends to  $\varepsilon$  as  $r$  grows to infinity, so that  $\lim_{r \rightarrow \infty} v_r = 0$ .

As for  $(a_n, n \in \mathbb{N}^*)$ , it is enough to notice that, since  $|\mathbb{T}_{r_n-1}| \leq n$ ,  $a_n = v_{r_n-1} |\mathbb{T}_{r_n-1}| / n \leq v_{r_n-1}$  and to apply the result for  $(v_r, r \in \mathbb{N})$ .  $\square$

**Step 2.** Let us now treat  $\overline{M}_{\mathbb{T}_r}(f)$ . From (4.2), we have by the triangle inequality  $\|\overline{M}_{\mathbb{T}_r}(f)\|_{L^2} \leq \sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \|\overline{M}_{\mathbb{G}_q}(f)\|_{L^2}$ . From Step 1,  $\overline{M}_{\mathbb{G}_q}(f)$  converges to 0 in quadratic mean. Lemma 40 implies that the r.h.s. tends to 0, which ends the proof for  $\overline{M}_{\mathbb{T}_r}(f)$ .

**Step 3.** Eventually, let us look at  $\overline{M}_n^\Pi(f)$ . From (4.3) and the triangle inequality,  $\|\overline{M}_n^\Pi(f)\|_{L^2} \leq a_n + b_n$  where

$$a_n = \sum_{q=0}^{r_n-1} \frac{|\mathbb{G}_q|}{n} \|\overline{M}_{\mathbb{G}_q}(f)\|_{L^2} \quad \text{and} \quad b_n = \left\| \frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right\|_{L^2}.$$

Since  $\overline{M}_{\mathbb{G}_q}(f)$  converges to 0 in quadratic mean, Lemma 40 implies that  $\lim_{n \rightarrow \infty} a_n = 0$ . As for  $b_n$ , first note that since each  $f(X_i)$  belongs to  $L^2$ ,  $f(X_{\Pi(i)})f(X_{\Pi(j)}) \in L^1$  for all  $i, j \in \{2^{r_n}, \dots, n\}$  so that

$$b_n^2 = \frac{1}{n^2} \sum_{i, j=2^{r_n}}^n \mathbb{E}[f(X_{\Pi(i)})f(X_{\Pi(j)})].$$

Let us compute the latter expectation, depending on  $i = j$  or  $i \neq j$ . For all  $i \in \{2^{r_n}, \dots, n\}$ ,  $\Pi(i)$  has the uniform distribution on  $\mathbb{G}_{r_n}$  so that when  $i = j$ ,  $\mathbb{E}[f(X_{\Pi(i)})f(X_{\Pi(j)})] = \mathbb{E}[f^2(X_{\Pi(i)})] = \nu Q^{r_n} f^2$ . Let us now treat the case when  $i \neq j$ . Then  $r_n \geq 1$ . Independently on  $(X, \Pi)$ , draw two independent indices  $I_{r_n}$  and  $J_{r_n}$  uniformly from  $\mathbb{G}_{r_n}$ . Then since  $i \neq j$ , the law of  $(\Pi(i), \Pi(j))$  is the conditional law of  $(I_{r_n}, J_{r_n})$  given  $\{I_{r_n} \neq J_{r_n}\}$  so that

$$\begin{aligned} \mathbb{E}[f(X_{\Pi(i)})f(X_{\Pi(j)})] &= \mathbb{E}[f(X_{I_{r_n}})f(X_{J_{r_n}})\mathbf{1}_{\{I_{r_n} \neq J_{r_n}\}}] / \mathbb{P}(I_{r_n} \neq J_{r_n}) \\ &= (1 - 2^{-r_n})^{-1} \mathbb{E}[f(X_{I_{r_n}})f(X_{J_{r_n}})\mathbf{1}_{\{I_{r_n} \neq J_{r_n}\}}] \\ &= (1 - 2^{-r_n})^{-1} (\mathbb{E}[f(X_{I_{r_n}})f(X_{J_{r_n}})] - \mathbb{E}[f^2(X_{I_{r_n}})\mathbf{1}_{\{I_{r_n}=J_{r_n}\}}]) \\ &= (1 - 2^{-r_n})^{-1} (\mathbb{E}[f(X_{I_{r_n}})f(X_{J_{r_n}})] - \mathbb{E}[f^2(X_{I_{r_n}})]\mathbb{P}(I_{r_n} = J_{r_n})) \\ &= (1 - 2^{-r_n})^{-1} (\mathbb{E}[f(X_{I_{r_n}})f(X_{J_{r_n}})] - 2^{-r_n} \nu Q^{r_n} f^2) \\ &= (1 - 2^{-r_n})^{-1} \sum_{p=0}^{r_n-1} 2^{-p-1} \nu Q^p P(Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f), \end{aligned}$$

where we have used  $\mathbb{P}(I_{r_n} = J_{r_n}) = 2^{-r_n}$  in the second and fifth equalities, the independence of  $(X, I_{r_n})$  and  $\mathbf{1}_{\{I_{r_n}=J_{r_n}\}}$  in the fourth one, the fact that  $X_{I_{r_n}}$  has the same distribution as  $Y_{r_n}$  in the fifth one and (4.6) with  $q = r_n$  in the last one. Eventually, we have proved that  $b_n^2 = b'_n + b''_n$  with

$$\begin{aligned} b'_n &= \frac{n - 2^{r_n} + 1}{n^2} \nu Q^{r_n} f^2, \\ b''_n &= \frac{(n - 2^{r_n})(n - 2^{r_n} + 1)}{n^2(1 - 2^{-r_n})} \sum_{p=0}^{r_n-1} 2^{-p-1} \nu Q^p P(Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f). \end{aligned}$$

Since  $(n - 2^{r_n} + 1)/n^2 \leq 1/n$ , and using (ii), (v) and (vi),  $\lim_{n \rightarrow \infty} b'_n = 0$ . As for  $b''_n$ , let us fix  $\varepsilon > 0$  and choose  $p_\varepsilon \in \mathbb{N}^*$  such that  $2^{-p_\varepsilon} \leq \varepsilon$ . From (4.7), there is a  $c_f \geq 0$  such that  $|\nu Q^p P(Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f)| \leq c_f$  for all  $p$  and  $n$  such that  $r_n \geq p$ . Since  $(n - 2^{r_n})(n - 2^{r_n} + 1)/n^2(1 - 2^{-r_n}) \leq 1$ , we then have

$$b''_n \leq \varepsilon c_f + \sum_{p=0}^{p_\varepsilon-1} |\nu Q^p P(Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f)|.$$

Now, using (4.9), we get that each term of the latter finite sum tends to 0 as  $n$  tends to infinity, so that finally  $\lim_{n \rightarrow \infty} b''_n = 0$ , which completes the proof.  $\square$

*Proof of Theorem 39.* Considering the function  $f - (\mu, Pf)$  leaves us with the case when  $(\mu, Pf) = 0$ . Let us treat the case of  $\overline{M}_{\mathbb{G}_q}(f)$ . Observe that  $f(\Delta_i) \in L^2$  for all  $i \in \mathbb{G}_q$ . Indeed,  $(z_1, \dots, z_q)$  denoting the path from the root 1 to  $i$  in the tree,  $\mathbb{E}[f(\Delta_i)^2] = \nu P_{z_1} \cdots P_{z_q} Pf^2$ , which is finite from (iii') and (vi), since  $Pf^2 \in F$ . Thus, by conditioning on  $\mathcal{F}_q$ ,  $\mathbb{E}[f(\Delta_i)f(\Delta_j)] = \mathbb{E}[Pf(X_i)Pf(X_j)]$  for all  $i \neq j \in \mathbb{G}_q$ , and  $\mathbb{E}[f^2(\Delta_i)] = \mathbb{E}[Pf^2(X_i)]$  for all  $i \in \mathbb{G}_q$ . Hence

$$\begin{aligned} \mathbb{E}[M_{\mathbb{G}_q}(f)^2] &= \sum_{i,j \in \mathbb{G}_q} \mathbb{E}[f(\Delta_i)f(\Delta_j)] \\ &= \sum_{i \in \mathbb{G}_q} \mathbb{E}[Pf^2(X_i)] + \sum_{i \neq j \in \mathbb{G}_q} \mathbb{E}[Pf(X_i)Pf(X_j)] \\ &= \mathbb{E}[M_{\mathbb{G}_q}(Pf)^2] + \mathbb{E}[M_{\mathbb{G}_q}(Pf^2 - (Pf)^2)], \end{aligned}$$

so that

$$\mathbb{E}[\overline{M}_{\mathbb{G}_q}(f)^2] = \mathbb{E}[\overline{M}_{\mathbb{G}_q}(Pf)^2] + \frac{\mathbb{E}[\overline{M}_{\mathbb{G}_q}(Pf^2 - (Pf)^2)]}{|\mathbb{G}_q|}.$$

We can apply Theorem 38 twice:  $\lim_{q \rightarrow \infty} \mathbb{E}[\overline{M}_{\mathbb{G}_q}(Pf)^2] = 0$  and  $\lim_{q \rightarrow \infty} \mathbb{E}[\overline{M}_{\mathbb{G}_q}(Pf^2 - (Pf)^2)] = (\mu, Pf^2 - (Pf)^2)$ , so that  $\lim_{q \rightarrow \infty} \mathbb{E}[\overline{M}_{\mathbb{G}_q}(f)^2] = 0$ , i.e.  $\overline{M}_{\mathbb{G}_q}(f)$  converges to 0 in  $L^2$ . Using Lemma 40 we extend this result to  $\overline{M}_{\mathbb{T}_r}(f)$ . The proof for  $\overline{M}_n^{\text{II}}(f)$  is similar to Step 3 of the proof of Theorem 38, with the same extra conditioning argument as above.  $\square$

### 4.2.3 Strong law of large numbers

So far, we have proved weak LLN, that is convergence in quadratic mean for empirical averages. We now seek for strong LLN. Theorem 41 gives sufficient conditions under which the empirical averages over the  $q$ -th generation and over the first  $r + 1$  generations converge to a constant with probability one.

**Theorem 41.** *Let  $F$  satisfy conditions (i)-(vi). Let  $f \in F$  such that  $(\mu, f) = 0$ . Assume that there exists  $h \in F$  such that*

$$P \left( \sum_{r \in \mathbb{N}} |Q^r f \otimes Q^r f| \right) \leq h.$$

*Then  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  almost surely converge to 0 as  $q \rightarrow \infty$ .*

*Proof. Step 1.* Let us first treat  $\overline{M}_{\mathbb{G}_q}(f)$ . Let us write  $\eta_q = \mathbb{E}[\overline{M}_{\mathbb{G}_q}(f)^2]$ . It is enough to check that  $\sum_{q \in \mathbb{N}} \eta_q < \infty$ . Now, using (4.6) and Fubini's theorem, we have

$$\begin{aligned} \sum_{q \in \mathbb{N}} \eta_q &= \sum_{q \in \mathbb{N}} \sum_{p=0}^q 2^{-p-1\{p < q\}} \nu Q^p P(Q^{q-p-1} f \otimes Q^{q-p-1} f) \\ &\leq \sum_{p \in \mathbb{N}} \sum_{q=p}^{+\infty} 2^{-p-1\{p < q\}} |\nu Q^p P(Q^{q-p-1} f \otimes Q^{q-p-1} f)| \\ &\leq \sum_{p \in \mathbb{N}} 2^{-p} \nu Q^p \left( f^2 + P \left( \sum_{r \in \mathbb{N}} |Q^r f \otimes Q^r f| \right) \right) \\ &\leq \sum_{p \in \mathbb{N}} 2^{-p} \nu Q^p (f^2 + h) \end{aligned}$$

which, from (v) and (vi), is finite, since  $f^2 + h \in F$ .

**Step 2.** Let us now deal with  $\overline{M}_{\mathbb{T}_r}(f)$ . From (4.2),  $|\overline{M}_{\mathbb{T}_r}(f)| \leq \sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} |\overline{M}_{\mathbb{G}_q}(f)|$ . From Step 1, a.s.  $\lim_{q \rightarrow \infty} |\overline{M}_{\mathbb{G}_q}(f)| = 0$ . It is enough to apply Lemma 40 to get that  $\overline{M}_{\mathbb{T}_r}(f)$  a.s. converges to 0.  $\square$

In particular, we have

**Corollary 42.** *Let  $F$  satisfy conditions (i)-(vi). Let  $f \in F$  such that  $(\mu, f) = 0$ . Assume there exists  $c \in F$  and a nonnegative sequence  $(\kappa_r, r \in \mathbb{N})$  such that  $\sum_{r \in \mathbb{N}} \kappa_r < \infty$  and*

$$\forall x \in S, \quad \forall r \in \mathbb{N}, \quad |Q^r f(x)| \leq c(x) \kappa_r.$$

*Then  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  almost surely converge to 0.*

*Proof.* From (iii),  $h = (\sum_{r \in \mathbb{N}} \kappa_r) P(c \otimes c) \in F$  and  $P(\sum_{r \in \mathbb{N}} |Q^r f \otimes Q^r f|) \leq h$ , so that Theorem 41 gives the result.  $\square$

**Remark 43.** In the case when  $\kappa_r = \kappa^r$  for some  $\kappa \in (0, 1)$ , we speak of geometric ergodicity. Geometric ergodicity implies the almost sure convergence of  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$ .

**Remark 44.** Assume that the state space  $S$  is finite, and that the induced MC  $Y$  is irreducible and aperiodic. Then  $Y$  is ergodic and,  $\mu$  standing for its stationary distribution, the sequence of functions  $(Q^r g, r \in \mathbb{N})$  uniformly converges to  $(\mu, g)$  with exponential speed. Taking  $F = \mathcal{B}(S)$  and  $f = g - (\mu, g)$ , Corollary 42 applies:  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  almost surely converge to 0, i.e.  $\overline{M}_{\mathbb{G}_q}(g)$  and  $\overline{M}_{\mathbb{T}_r}(g)$  almost surely converge to  $(\mu, g)$ . This covers the main result in [98, 71, 91] when applied to the binary tree.



In the case when  $f$  depends on the mother-daughters triangle  $\Delta_n = (X_n, X_{2n}, X_{2n+1})$ , we can prove as well

**Theorem 45.** *Let  $F$  satisfy conditions (i)-(vi). Let  $f \in \mathcal{B}(S^3)$  such that  $Pf$  and  $Pf^2$  exist and belong to  $F$ , with  $(\mu, Pf) = 0$ .*

(i) *Assume that there exists  $h \in F$  such that  $P(\sum_{r \in \mathbb{N}} |Q^r Pf \otimes Q^r Pf|) \leq h$ .*

*Then  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  almost surely converge to 0.*

(ii) *In particular, if there exists  $c \in F$  and a nonnegative sequence  $(\kappa_r, r \in \mathbb{N})$  such that  $\sum_{r \in \mathbb{N}} \kappa_r < \infty$  and*

$$\forall x \in S, \quad \forall r \in \mathbb{N}, \quad |Q^r Pf(x)| \leq c(x)\kappa_r,$$

*then  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  almost surely converge to 0.*

#### 4.2.4 Central limit theorem

We are now interested in proving a central limit theorem (CLT) for the  $\mathbb{T}$ -MC  $(X_n)$ . This will be done by using a CLT for martingales.

**Theorem 46.** *Let  $F$  satisfy (i)-(vi). Let  $f \in \mathcal{B}(S^3)$  such that  $Pf^2$  and  $Pf^4$  exist and belong to  $F$ . Assume that  $Pf = 0$ . Then  $n^{-1/2}M_n^\Pi(f)$  converges in distribution to the Gaussian law  $\mathcal{N}(0, s^2)$ , where  $s^2 = (\mu, Pf^2)$ .*

*Proof.* Let  $M_0^\Pi(f) = 0$ ,  $\mathcal{H}_0 = \sigma(X_1)$  and  $\mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, 1 \leq i \leq n)$  for all  $n \geq 1$ . Note that  $X_{\Pi(i)}$  is  $\mathcal{H}_{i-1}$ -measurable and that, conditionally on  $\mathcal{H}_{i-1}$ ,  $\Delta_{\Pi(i)}$  has distribution  $\delta_{X_{\Pi(i)}} \otimes P(X_{\Pi(i)}, \cdot)$ . Since  $Pf = 0$ ,  $(M_n^\Pi(f), n \geq 0)$  is a  $(\mathcal{H}_n)$ -martingale. It has bracket

$$\langle M^\Pi(f) \rangle_n = \sum_{i=1}^n \mathbb{E}[f^2(\Delta_{\Pi(i)}) | \mathcal{H}_{i-1}] = \sum_{i=1}^n Pf^2(X_{\Pi(i)}) = M_n^\Pi(Pf^2).$$

According to Theorem 38, since  $Pf^2 \in F$ ,  $n^{-1}\langle M^\Pi(f) \rangle_n$  converges to  $(\mu, Pf^2) = s^2$  in  $L^2$ , and thus in probability. It remains to check Liapunov's condition, say for the fourth moment, that is to prove that the sequence of positive r.v.  $(L_n, n \geq 1)$  defined by

$$L_n = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}[f^4(\Delta_{\Pi(i)}) | \mathcal{H}_{i-1}]$$

tends in probability to 0 (see for instance [25]). But  $L_n = \overline{M}_n^\Pi(Pf^4)/n$  and  $\overline{M}_n^\Pi(Pf^4)$  converges to  $(\mu, Pf^4)$  in quadratic mean, so that  $L_n$  converges to 0 in probability.  $\square$

In the general case when  $Pf \neq 0$ , we have

**Corollary 47.** *Let  $F$  satisfy (i)-(vi). Let  $f \in \mathcal{B}(S^3)$  such that  $Pf$ ,  $Pf^2$  and  $Pf^4$  exist and belong to  $F$ . Then  $n^{-1/2}(M_n^\Pi(f) - M_n^\Pi(Pf))$  converges in distribution to  $\mathcal{N}(0, s^2)$ , where  $s^2 = (\mu, Pf^2) - (\mu, (Pf)^2)$ .*

*Proof.* It is enough to apply Theorem 46 to the function  $g$  defined by  $g(x, y, z) = f(x, y, z) - Pf(x)$ .  $\square$

Considering the subsequence of indices  $n = |\mathbb{T}_r|$ ,  $r \in \mathbb{N}$ , we can state

**Corollary 48.** *Let  $F$  satisfy (i)-(vi). Let  $f \in \mathcal{B}(S^3)$  such that  $Pf$ ,  $Pf^2$  and  $Pf^4$  exist and belong to  $F$ . Then  $|\mathbb{T}_r|^{1/2}(\overline{M}_{\mathbb{T}_r}(f) - \overline{M}_{\mathbb{T}_r}(Pf))$  converges in distribution to  $\mathcal{N}(0, s^2)$ , where  $s^2 = (\mu, Pf^2) - (\mu, (Pf)^2)$ .*

If we take  $F$  to be  $\mathcal{C}_b(S)$ , we get

**Corollary 49.** *Assume that the induced MC  $(Y_r, r \in \mathbb{N})$  is ergodic, with stationary distribution  $\mu$  (see Definition 4.2.2). Then, for any  $f \in \mathcal{C}_b(S^3)$ ,  $n^{-1/2}(M_n^\Pi(f) - M_n^\Pi(Pf))$  converges in distribution to  $\mathcal{N}(0, s^2)$ , where  $s^2 = (\mu, Pf^2) - (\mu, (Pf)^2)$ .*

**Remark 50.** Note that the normalizing factor is the square root of the number of individuals,  $n$  or  $|\mathbb{T}_r|$ , and not the square root of the number of generations,  $r_n$  or  $r$ , as one might have thought. Convergence is fast with  $r$ : with 20 generations ( $r = 19$ ), the normalizing factor  $|\mathbb{T}_r|^{1/2}$  is approximately  $10^3$ .

Using characteristic functions, it is easy to generalize Corollary 49 to the case when  $f$  is vector-valued:

**Corollary 51.** *Let  $F$  satisfy (i)-(vi). Let  $f_1, \dots, f_d \in \mathcal{B}(S^3)$  such that  $Pf_i$ ,  $P(f_i f_j)$  and  $P(f_i f_j f_k f_l)$  exist and belong to  $F$  for all  $i, j, k, l$ . Let  $f = (f_1, \dots, f_d)$ . Then  $n^{-1/2}(M_n^\Pi(f) - M_n^\Pi(Pf))$  converges in distribution to the  $d$ -dimensional Gaussian law  $\mathcal{N}_d(0, \Sigma)$ , where  $\Sigma_{ij} = (\mu, P(f_i f_j)) - (\mu, Pf_i Pf_j)$ .*

*Proof.* Let us denote  $R^n(f) = n^{-1/2}(M_n^\Pi(f) - M_n^\Pi(Pf))$ . Let  $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ . The characteristic function of  $R^n(f)$  is

$$\varphi(u) = \mathbb{E}[\exp(i(u_1 R_1^n(f) + \dots + u_d R_d^n(f)))] = \mathbb{E}[\exp(iR^n(\langle u, f \rangle))].$$

where  $\langle u, f \rangle = u_1 f_1 + \dots + u_d f_d$ . Now, since  $P(\langle u, f \rangle)$ ,  $P(\langle u, f \rangle)^2$  and  $P(\langle u, f \rangle)^4 \in F$ , Corollary 47 implies that the latter term tends to

$$\exp\left(-\frac{1}{2}((\mu, P\langle u, f \rangle^2) - (\mu, (P\langle u, f \rangle)^2))\right) = \exp\left(-\frac{1}{2}u^t \Sigma u\right),$$

which completes the proof.  $\square$

## 4.3 Detection of cellular aging

### 4.3.1 Limit theorems in the BAR model

Here we seek to apply the results in Section 4.2 to the BAR model (4.1).

#### Weak law of large numbers and central limit theorem

In this section, we take  $F$  to be the set  $C_{\text{poi}}(\mathbb{R})$  of continuous and polynomially growing functions, i.e. the set of all continuous functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that there exists  $c \geq 0$  and  $m \in \mathbb{N}$  such that for all  $x \in \mathbb{R}$

$$|f(x)| \leq c(1 + |x|^m).$$

In order to apply Theorems 38 and 39 and Corollary 51, we need to check conditions (i)-(vi). Conditions (i) and (ii) are obvious. The next lemma states that condition (iii) is fulfilled too.

**Lemma 52.** *Let  $f, g \in C_{\text{poi}}(\mathbb{R})$ . Then  $f \otimes g \in L^1(P(x, \cdot))$  and  $P(f \otimes g) \in C_{\text{poi}}(\mathbb{R})$ .*

*Proof.* Let  $G_0, G_1$  be two independent standard Gaussian variables. Let

$$G = \begin{pmatrix} G_0 \\ G_1 \end{pmatrix}, \quad M(x) = \begin{pmatrix} \alpha_0 x + \beta_0 \\ \alpha_1 x + \beta_1 \end{pmatrix} \quad \text{and} \quad \Lambda = \sigma \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{pmatrix}.$$

Then  $G(x) \equiv M(x) + \Lambda G$  has distribution  $P(x, \cdot)$ , so that we have  $P(|f \otimes g|)(x) = \mathbb{E}[|f \otimes g(G(x))|]$ . Now,  $\|\cdot\|$  denoting the euclidian norm in  $\mathbb{R}^2$ , we can find  $c \geq 0$  and  $m \in \mathbb{N}$  such that  $|f \otimes g(G(x))| \leq c(1 + \|G(x)\|^m)$ . Since  $\|G(x)\| \leq c(1 + |x| + \|G\|)$  for a constant  $c$  and  $\mathbb{E}[\|G\|^m] < \infty$ , we may eventually find a  $c \geq 0$  such that for all  $x \in \mathbb{R}$

$$P(|f \otimes g|)(x) \leq c(1 + |x|^m),$$

which completes the proof.  $\square$

**Lemma 53.**  $C_{\text{poi}}(\mathbb{R})$  fulfils conditions (iv)-(v);  $\mu$  is the stationary distribution of  $Y$ .

*Proof.* In the BAR model (4.1), the induced MC has the stochastic dynamics

$$Y_{r+1} = a_{r+1}Y_r + b_{r+1}, \tag{4.10}$$

where  $((a_r, b_r), r \in \mathbb{N}^*)$  is a sequence of i.i.d. r.v., independent of  $Y_0$ . Precisely,  $a_{r+1} = \alpha_{\zeta_{r+1}}$ ,  $b_{r+1} = \beta_{\zeta_{r+1}} + \varepsilon'_{r+1}$  where  $(\varepsilon'_q, q \in \mathbb{N}^*)$  and  $(\zeta_q, q \in \mathbb{N}^*)$  are independent sequences of i.i.d. r.v., independent of  $Y_0$ , each  $\varepsilon'_q$  has law  $\mathcal{N}(0, \sigma^2)$  and each  $\zeta_q$  is a balanced Bernoulli r.v., i.e.  $\mathbb{P}(\zeta_q = 0) = \mathbb{P}(\zeta_q = 1) = 1/2$ . Ph. BOUGEROL and N.

PICARD [13] call the sequence  $(Y_r, r \in \mathbb{N})$  a generalized autoregressive process. It is sometimes called AR(1), for AutoRegressive of order 1, in the literature. We have

$$Y_r = a_r a_{r-1} \cdots a_2 a_1 Y_0 + \sum_{k=1}^r a_r a_{r-1} \cdots a_{k+1} b_k.$$

Since the r.v.  $((a_r, b_r), r \in \mathbb{N}^*)$  are i.i.d.,  $Y_r$  has the same distribution as

$$Z_r = a_1 a_2 \cdots a_{r-1} a_r Y_0 + \sum_{k=1}^r a_1 a_2 \cdots a_{k-1} b_k. \quad (4.11)$$

Let us first prove (v). Let  $f \in C_{\text{pol}}(\mathbb{R})$ ,  $x \in \mathbb{R}$  and  $S = \sum_{k=1}^{\infty} |a_1 a_2 \cdots a_{k-1} b_k|$ . From (4.11),  $|Z_r| \leq |Y_0| + S$  for all  $r \in \mathbb{N}$ , so that we can find  $c \geq 0$  and  $m \in \mathbb{N}^*$  such that  $\mathbb{E}_x[|f(Z_r)|] \leq c(|x|^m + \mathbb{E}[S^m])$ . Now, let us denote  $\alpha = \max\{|\alpha_0|, |\alpha_1|\} < 1$ . Using the triangle inequality in the first line and the fact that the  $L^m$ -norm of  $b_k$ ,  $\|b_k\|_{L^m} \equiv c_m$ , does not depend on  $k$  in the last one,

$$\|S\|_{L^m} \leq \sum_{k=1}^{\infty} \|a_1 a_2 \cdots a_{k-1} b_k\|_{L^m} \leq \sum_{k=1}^{\infty} \alpha^{k-1} \|b_k\|_{L^m} = c_m \sum_{k=1}^{\infty} \alpha^{k-1} < \infty. \quad (4.12)$$

Eventually,  $|Q^r f(x)| \leq \mathbb{E}_x[|f(Y_r)|] = \mathbb{E}_x[|f(Z_r)|] \leq c'_m(1 + |x|^m)$  for some  $c'_m$  which does not depend on  $r$ , which proves (v).

Let us now prove (iv). Since  $|a_1 a_2 \cdots a_{r-1} a_r Y_0| \leq \alpha^r |Y_0|$ ,  $\lim_{r \rightarrow \infty} a_1 a_2 \cdots a_{r-1} a_r Y_0 = 0$  a.s.. Besides, the sum in (4.11) a.s. converges as  $r$  grows to infinity as  $\mathbb{E}[|S|^m] < \infty$ . Eventually, the sequence  $(Z_r, r \in \mathbb{N})$  almost surely converges to

$$Z_{\infty} = \sum_{k=1}^{\infty} a_1 a_2 \cdots a_{k-1} b_k. \quad (4.13)$$

Let  $\mu$  denote the distribution of  $Z_{\infty}$ . Then  $C_{\text{pol}}(\mathbb{R}) \subset L^1(\mu)$ . Indeed,  $\|Z_{\infty}\|_{L^m} \leq \|S\|_{L^m} < \infty$  for all  $m \in \mathbb{N}$ . Let us eventually prove that  $\lim_{r \rightarrow \infty} \mathbb{E}_x[f(Y_r)] = (\mu, f)$ . Since  $|f(Z_r)| \leq c(|Y_0|^m + S^m) \in L^1(\mathbb{P}_x)$  and  $(Z_r, r \in \mathbb{N})$  almost surely converges to  $Z_{\infty}$ , we can apply Lebesgue's dominated convergence theorem and get that

$$\lim_{r \rightarrow \infty} \mathbb{E}_x[f(Y_r)] = \lim_{r \rightarrow \infty} \mathbb{E}_x[f(Z_r)] = \mathbb{E}_x[f(Z_{\infty})] = (\mu, f).$$

Condition (iv) is now fully checked, and  $\mu$  is the unique stationary distribution of  $Y$ .  $\square$

Let us denote by  $C_{\text{pol}}(\mathbb{R}^3)$  the set of all continuous and polynomially growing functions  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ . Since  $C_{\text{pol}}(\mathbb{R}^3)^2 \subset C_{\text{pol}}(\mathbb{R}^3)$  and  $P(C_{\text{pol}}(\mathbb{R}^3)) \subset C_{\text{pol}}(\mathbb{R})$ , Theorems 38 and 39 and Corollary 51 imply

**Proposition 54.** *In the BAR model (4.1), assume that the distribution of the ancestor  $X_1, \nu$ , has finite moments of all orders. Let  $\mu$  be the unique stationary distribution of the induced MC  $(Y_r, r \in \mathbb{N})$ . Then,*

- (i) *for all  $f \in C_{\text{pol}}(\mathbb{R})$ ,  $\overline{M}_{\mathbb{G}_q}(f)$ ,  $\overline{M}_{\mathbb{T}_r}(f)$  and  $\overline{M}_n^\Pi(f)$  converge to  $(\mu, f)$  in  $L^2$ ,*
- (ii) *for all  $f \in C_{\text{pol}}(\mathbb{R}^3)$ ,  $\overline{M}_{\mathbb{G}_q}(f)$ ,  $\overline{M}_{\mathbb{T}_r}(f)$  and  $\overline{M}_n^\Pi(f)$  converge to  $(\mu, Pf)$  in  $L^2$ ,*
- (iii) *for all  $f_1, \dots, f_d \in C_{\text{pol}}(\mathbb{R}^3)$ ,  $n^{-1/2}(M_n^\Pi(f) - M_n^\Pi(Pf))$  converges in distribution to  $\mathcal{N}_d(0, \Sigma)$ , where  $f = (f_1, \dots, f_d)$  and  $\Sigma_{ij} = (\mu, P(f_i f_j)) - (\mu, P f_i P f_j)$ .*

### Strong law of large numbers

We can also derive almost sure convergence results:

**Proposition 55.** *With the assumptions of Proposition 54,*

- (i) *almost surely, for any  $f \in C_{\text{pol}}(\mathbb{R})$ ,  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  converge to  $(\mu, f)$ ,*
- (ii) *almost surely, for any  $f \in C_{\text{pol}}(\mathbb{R}^3)$ ,  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  converge to  $(\mu, Pf)$ .*

*Proof.* Let us take  $F$  to be the set  $C_{\text{pol}}^1(\mathbb{R})$  of all  $C^1$  functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  such  $|f| + |f'|$  is bounded above by a polynomial. One can easily check that  $C_{\text{pol}}^1(\mathbb{R})$  satisfies (i)-(v).

**Step 1.** Let us first prove that

$$\forall f \in C_{\text{pol}}^1(\mathbb{R}), \quad \mathbb{P} \left( \lim_{q \rightarrow \infty} \overline{M}_{\mathbb{G}_q}(f) = \lim_{r \rightarrow \infty} \overline{M}_{\mathbb{T}_r}(f) = (\mu, f) \right) = 1. \quad (4.14)$$

Let  $f \in C_{\text{pol}}^1(\mathbb{R})$ . We want to apply Corollary 42 with  $F = C_{\text{pol}}^1(\mathbb{R})$  and to the function  $g = f - (\mu, f) \in C_{\text{pol}}^1(\mathbb{R})$ . First note that  $Q^r g(x) = Q^r f(x) - (\mu, f) = \mathbb{E}_x[f(Z_r) - f(Z_\infty)]$ , so that using Cauchy-Schwarz's inequality

$$|Q^r g(x)| \leq \mathbb{E}_x[W_r |Z_r - Z_\infty|] \leq (\mathbb{E}_x[W_r^2] \mathbb{E}_x[(Z_r - Z_\infty)^2])^{1/2}$$

where  $W_r = \sup_{z \in [Z_r, Z_\infty]} |f'(z)|$ . We can find  $c_1 \geq 0$  and  $m \in \mathbb{N}$  such that for all  $z \in \mathbb{R}$ ,  $|f'(z)|^2 \leq c_1(1 + |z|^m)$  so that, using (4.11) and (4.12), there is a  $c'_m \geq 0$  such that for all  $x \in \mathbb{R}$  and  $r \in \mathbb{N}$ ,

$$\mathbb{E}_x[W_r^2] \leq \mathbb{E}_x \left[ \sup_{z \in [Z_r, Z_\infty]} c_1(1 + |z|^m) \right] \leq c_1(1 + \mathbb{E}_x[|Z_r|^m] + \mathbb{E}_x[|Z_\infty|^m]) \leq c'_m(1 + |x|^m).$$

Moreover,  $Z_r - Z_\infty = a_1 a_2 \cdots a_{r-1} a_r Y_0 - \sum_{k=r+1}^{+\infty} a_1 a_2 \cdots a_{k-1} b_k$  so that, recalling  $\alpha = \max\{|\alpha_0|, |\alpha_1|\}$ ,  $\|Z_r - Z_\infty\|_{L^2(\mathbb{P}_x)} \leq \alpha^r x + \left\| \sum_{k=r+1}^{+\infty} a_1 a_2 \cdots a_{k-1} b_k \right\|_{L^2}$ . Now we have  $\left\| \sum_{k=r+1}^{+\infty} a_1 a_2 \cdots a_{k-1} b_k \right\|_{L^2} \leq c_2 \sum_{k=r+1}^{+\infty} \alpha^{k-1} = c_2 \alpha^r / (1 - \alpha)$  where  $c_2 = \|b_k\|_{L^2}$  does not depend on  $k$ . Thus we can find  $c_3 \geq 0$  such that for all  $x \in \mathbb{R}$  and  $r \in \mathbb{N}$

$$\mathbb{E}_x[(Z_r - Z_\infty)^2] \leq c_3 \alpha^{2r} (1 + x^2).$$

Eventually,

$$|Q^r g(x)| \leq (c'_m c_3 \alpha^{2r} (1 + |x|^m)(1 + x^2))^{1/2} \leq c(x) \alpha_r \quad (4.15)$$

with  $\alpha_r = \alpha^r$  and a function  $c \in C^1_{\text{pol}}(\mathbb{R})$ . Corollary 42 implies that  $\overline{M}_{\mathbb{G}_q}(g)$  and  $\overline{M}_{\mathbb{T}_r}(g)$  almost surely converge to 0, i.e.  $\overline{M}_{\mathbb{G}_q}(f)$  and  $\overline{M}_{\mathbb{T}_r}(f)$  almost surely converge to  $(\mu, f)$ , which proves (4.14).

**Step 2.** Let us now prove that almost surely, the empirical distributions  $\overline{M}_{\mathbb{G}_q}$  and  $\overline{M}_{\mathbb{T}_r}$  weakly converge to  $\mu$ . There exists a sequence  $(f_p, p \in \mathbb{N})$  of  $C^\infty$  functions with compact support which characterizes convergence in distribution. Hence it is enough to show that almost surely, for all  $p \in \mathbb{N}$ ,  $\lim_{q \rightarrow \infty} \overline{M}_{\mathbb{G}_q}(f_p) = \lim_{r \rightarrow \infty} \overline{M}_{\mathbb{T}_r}(f_p) = (\mu, f_p)$ . But this immediately follows from Step 1, since  $f_p \in C^1_{\text{pol}}(\mathbb{R})$ .

**Step 3.** Let us now prove assertion (i). Let us deal with  $\overline{M}_{\mathbb{G}_q}$  (the proof for  $\overline{M}_{\mathbb{T}_r}$  is similar). For  $k, l \in \mathbb{N}$ , let us write  $f_{k,l}(x) = k(1 + x^{2l})$ . Since  $f_{k,l} \in C^1_{\text{pol}}(\mathbb{R})$ , from Step 1, almost surely,

$$\forall k, l \in \mathbb{N}, \quad \lim_{q \rightarrow \infty} \overline{M}_{\mathbb{G}_q}(f_{k,l}) = (\mu, f_{k,l}). \quad (4.16)$$

From Step 2, the empirical distributions  $\overline{M}_{\mathbb{G}_q}$  a.s. weakly converge to  $\mu$ . Besides, for all  $f \in C_{\text{pol}}(\mathbb{R})$ , there exists  $k, l \in \mathbb{N}$  such that  $f^2 \leq f_{k,l}$ . Thus from (4.16), a.s. for all  $f \in C_{\text{pol}}(\mathbb{R})$  the sequence  $(\overline{M}_{\mathbb{G}_q}(f^2), q \in \mathbb{N})$  is bounded, which proves that a.s. every  $f \in C_{\text{pol}}(\mathbb{R})$  is  $(\overline{M}_{\mathbb{G}_q}, q \in \mathbb{N})$ -uniformly integrable. Hence a.s., for all  $f \in C_{\text{pol}}(\mathbb{R})$ ,  $\overline{M}_{\mathbb{G}_q}(f)$  converges to  $(\mu, f)$ .

**Step 4.** Lastly, let us prove assertion (ii). We denote by  $C^1_{\text{pol}}(\mathbb{R}^3)$  the set of all  $C^1$  functions  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  such  $|f| + |\partial_1 f| + |\partial_2 f| + |\partial_3 f|$  is bounded above by a polynomial. As above, we first prove that

$$\forall f \in C^1_{\text{pol}}(\mathbb{R}^3), \quad \mathbb{P} \left( \lim_{q \rightarrow \infty} \overline{M}_{\mathbb{G}_q}(f) = \lim_{r \rightarrow \infty} \overline{M}_{\mathbb{T}_r}(f) = (\mu, Pf) \right) = 1, \quad (4.17)$$

and then get that, a.s., the empirical distributions  $\overline{M}_{\mathbb{G}_q}$  and  $\overline{M}_{\mathbb{T}_r}$  on  $(S^3, \mathcal{S}^3)$  weakly converge to the distribution  $\mu \otimes P$  defined by  $\mu \otimes P(f) = (\mu, Pf)$ . Applying an analogous argument of uniform integrability as in Step 3, we get assertion (ii). Let  $f \in C^1_{\text{pol}}(\mathbb{R}^3)$ . Then  $g = f - (\mu, Pf) \in C^1_{\text{pol}}(\mathbb{R}^3)$  satisfies  $Pg, Pg^2 \in C^1_{\text{pol}}(\mathbb{R})$ ,  $(\mu, Pg) = 0$  and, in the same way as we prove (4.15), we have  $|Q^r Pg(x)| \leq c(x) \alpha^r$  for some  $c \in C^1_{\text{pol}}(\mathbb{R})$ . Hence, from Theorem 45-(ii), the empirical averages  $\overline{M}_{\mathbb{G}_q}(g)$  and  $\overline{M}_{\mathbb{T}_r}(g)$  a.s. converge to 0, which completes the proof.  $\square$

**Remark 56.** A natural choice for  $\nu$  is the stationary distribution  $\mu$ . Indeed, the ancestor  $X_1$  is picked from a metacolony that has evolved for a long time, so that in the BAR model (4.1) its distribution should be close to  $\mu$ . With this particular choice, we can apply Propositions 54 and 55. Indeed,  $\mu$  has finite moments of all orders, since  $C_{\text{pol}}(\mathbb{R}) \subset L^1(\mu)$ .

### 4.3.2 Estimation of the parameters

We seek to estimate the 4-dimensional parameter  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ , as well as  $\sigma^2$  and  $\rho$ . Assume we observe a complete subtree  $\mathbb{T}_{r+1}$ . Then, since the couples  $(\varepsilon_{2i}, \varepsilon_{2i+1})$  are i.i.d. bivariate Gaussian vectors, the maximum likelihood estimator  $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$  of  $\theta$  is also the least squares one: for  $\varepsilon \in \{0, 1\}$ ,

$$\begin{cases} \hat{\alpha}_\varepsilon^r = \frac{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i X_{2i+\varepsilon} - \left( |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i \right) \left( |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\varepsilon} \right)}{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i^2 - \left( |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i \right)^2}, \\ \hat{\beta}_\varepsilon^r = |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\varepsilon} - \hat{\alpha}_\varepsilon^r |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i. \end{cases} \quad (4.18)$$

Hence  $\hat{\alpha}_0^r$  (resp.  $\hat{\alpha}_1^r$ ) is the empirical correlation between new (resp. old) pole daughters and their mothers. We shall denote by  $\mathbf{xy}$  (resp.  $\mathbf{x}, \mathbf{y}, \mathbf{x}^2$ ) the element of  $C_{\text{pol}}(\mathbb{R}^3)$  defined by  $(x, y, z) \mapsto xy$  (resp.  $x, y, x^2$ ).

**Remark 57.** Note that  $(\mu, \mathbf{x}^2) - (\mu, \mathbf{x})^2 > 0$ . Indeed, it is nonnegative, and if it were 0,  $\mu$  would be a Dirac mass. Now a Dirac mass cannot be stationary for  $Y$ , because  $\sigma^2 > 0$ .

**Proposition 58.** *In the BAR model (4.1), assume that the distribution of the ancestor  $X_1, \nu$ , has finite moments of all orders. Then  $(\hat{\theta}^r, r \in \mathbb{N})$  is a strongly consistent estimator of  $\theta$ .*

*Proof.* Let us treat  $\hat{\alpha}_0^r$ . Convergence of  $\hat{\beta}_0^r, \hat{\alpha}_1^r$  and  $\hat{\beta}_1^r$  may be treated in a similar way. Note that  $\hat{\alpha}_0^r = C_r/B_r$  with

$$C_r = \overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})\overline{M}_{\mathbb{T}_r}(\mathbf{y}) \quad \text{and} \quad B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2.$$

Since  $P(\mathbf{xy})(x) = x(\alpha_0 x + \beta_0)$  and  $P(\mathbf{y})(x) = \alpha_0 x + \beta_0$ , Proposition 55 implies that  $C_r$  a.s. converges to  $(\mu, \mathbf{x}(\alpha_0 \mathbf{x} + \beta_0)) - (\mu, \mathbf{x})(\mu, \alpha_0 \mathbf{x} + \beta_0) = \alpha_0((\mu, \mathbf{x}^2) - (\mu, \mathbf{x})^2)$  and  $B_r$  a.s. converges to  $(\mu, \mathbf{x}^2) - (\mu, \mathbf{x})^2$ , which from Remark 57 is positive, so that  $\hat{\alpha}_0^r$  a.s. converges to  $\alpha_0$ .  $\square$

**Remark 59.** Let us denote  $\bar{\alpha} = (\alpha_0 + \alpha_1)/2$ ,  $\bar{\beta} = (\beta_0 + \beta_1)/2$  and so on. Then

$$(\mu, \mathbf{x}) = \frac{\bar{\beta}}{1 - \bar{\alpha}} \quad \text{and} \quad (\mu, \mathbf{x}^2) = \frac{2\bar{\alpha}\bar{\beta}\bar{\beta}/(1 - \bar{\alpha}) + \bar{\beta}^2 + \sigma^2}{1 - \bar{\alpha}^2}. \quad (4.19)$$

Indeed, recalling (4.10) and (4.13),  $Z_\infty$  has the same law as  $a_1 Z_\infty + b_1$  where the pair  $(a_1, b_1)$  is independent of  $Z_\infty$  and takes values  $(\alpha_0, \beta_0)$  and  $(\alpha_1, \beta_1)$  with probability

1/2. Hence  $(\mu, \mathbf{x}) = \mathbb{E}[Z_\infty] = \mathbb{E}[a_1 Z_\infty + b_1] = \bar{\alpha}(\mu, \mathbf{x}) + \bar{\beta}$ , as announced. Likewise,

$$\begin{aligned} (\mu, \mathbf{x}^2) &= \mathbb{E}[Z_\infty^2] = \mathbb{E}[(a_1 Z_\infty + b_1)^2] = \mathbb{E}[a_1^2 Z_\infty^2] + 2\mathbb{E}[a_1 b_1 Z_\infty] + \mathbb{E}[b_1^2] \\ &= \bar{\alpha}^2(\mu, \mathbf{x}^2) + 2\bar{\alpha}\bar{\beta}(\mu, \mathbf{x}) + \bar{\beta}^2 + \sigma^2, \end{aligned}$$

from which we deduce the second equality in (4.19).

From the preceding remark, we define two continuous functions  $\mu_1 : \Theta \rightarrow \mathbb{R}$  and  $\mu_2 : \Theta \times \mathbb{R}_+^* \rightarrow \mathbb{R}$  by writing

$$(\mu, \mathbf{x}) = \mu_1(\theta) \quad \text{and} \quad (\mu, \mathbf{x}^2) = \mu_2(\theta, \sigma^2). \quad (4.20)$$

where  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1) \in \Theta = (-1, 1) \times \mathbb{R} \times (-1, 1) \times \mathbb{R}$ . Let us now build a confidence region for  $\theta$ .

**Proposition 60.** *In the BAR model (4.1), assume that the distribution  $\nu$  of the ancestor  $X_1$  has finite moments of all orders. Let  $\mu$  be the unique stationary distribution of the induced MC  $(Y_r, r \in \mathbb{N})$ . Then  $|\mathbb{T}_r|^{1/2}(\hat{\theta}^r - \theta)$  converges in law to  $\mathcal{N}_4(0, \Sigma')$  where*

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix} \quad \text{with} \quad K = \frac{1}{\mu_2(\theta, \sigma^2) - \mu_1(\theta)^2} \begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta, \sigma^2) \end{pmatrix}. \quad (4.21)$$

*Proof.* For  $f_1, \dots, f_d \in C_{\text{pol}}(\mathbb{R}^3)$ , we denote  $f = (f_1, \dots, f_d)$  and  $U^r(f) = |\mathbb{T}_r|^{1/2}(M_{\mathbb{T}_r}(f) - M_{\mathbb{T}_r}(Pf))$ . Let us denote  $\zeta_r = |\mathbb{T}_r|^{1/2}(\hat{\theta}^r - \theta)$ . We first observe that  $\zeta_r = \varphi(U^r(f), A_r, B_r)$  with  $f = (\mathbf{xy}, \mathbf{y}, \mathbf{xz}, \mathbf{z})$ ,  $\varphi(u, a, b) = M(a, b)u$ ,

$$M(a, b) = \begin{pmatrix} 1/b & -a/b & 0 & 0 \\ -a/b & (b+a^2)/b & 0 & 0 \\ 0 & 0 & 1/b & -a/b \\ 0 & 0 & -a/b & (b+a^2)/b \end{pmatrix},$$

$A_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x})$  and  $B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2$ . From Proposition 54-(iii),  $U^r(f)$  converges in distribution to  $G \sim \mathcal{N}_4(0, \Sigma)$  with

$$\Sigma = \sigma^2 \begin{pmatrix} \mu_2(\theta, \sigma^2) & \mu_1(\theta) & \rho\mu_2(\theta, \sigma^2) & \rho\mu_1(\theta) \\ \mu_1(\theta) & 1 & \rho\mu_1(\theta) & \rho \\ \rho\mu_2(\theta, \sigma^2) & \rho\mu_1(\theta) & \mu_2(\theta, \sigma^2) & \mu_1(\theta) \\ \rho\mu_1(\theta) & \rho & \mu_1(\theta) & 1 \end{pmatrix}.$$

Besides, Proposition 54-(i) implies that  $(A_r, B_r)$  converges in law to the constant  $(a, b) = (\mu_1(\theta), \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2)$ . Thus Slutsky's theorem states that  $(U^r(f), A_r, B_r)$



converges in law to  $(G, a, b)$ . Then, by continuity of  $\varphi$  on  $\mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^*$  and recalling from Remark 57 that  $b > 0$ ,  $\zeta_r = \varphi(U^r(f), A_r, B_r)$  converges in law to  $\varphi(G, a, b) = M(a, b)G$ , which is a centered Gaussian vector with covariance matrix  $\Sigma' = M(a, b)\Sigma M(a, b)^t$ . Now, we have

$$\Sigma = \sigma^2 \begin{pmatrix} L & \rho L \\ \rho L & L \end{pmatrix} \text{ with } L = \begin{pmatrix} \mu_2(\theta, \sigma^2) & \mu_1(\theta) \\ \mu_1(\theta) & 1 \end{pmatrix}, \text{ and } M(a, b) = \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}.$$

Since  $LK = I_2$ , the  $2 \times 2$  identity matrix, we get

$$\Sigma' = \sigma^2 \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix} \begin{pmatrix} L & \rho L \\ \rho L & L \end{pmatrix} \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix} = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix},$$

which completes the proof.  $\square$

We also need to estimate the conditional variance,  $\sigma^2$ , and the conditional sister-sister correlation,  $\rho$ . Since  $\sigma^2$  is the common expectation of the i.i.d. r.v.  $(\varepsilon_i^2, i \geq 2)$ , it is naturally estimated, given a complete observation  $(X_i, i \in \mathbb{T}_{r+1})$ , by

$$\hat{\sigma}_r^2 = \frac{1}{2|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} (\hat{\varepsilon}_{2i}^2 + \hat{\varepsilon}_{2i+1}^2),$$

where

$$\begin{cases} \hat{\varepsilon}_{2n} &= X_{2n} - \hat{\alpha}_0^r X_n - \hat{\beta}_0^r, \\ \hat{\varepsilon}_{2n+1} &= X_{2n+1} - \hat{\alpha}_1^r X_n - \hat{\beta}_1^r \end{cases}$$

are the residues. Likewise, since  $\rho = \text{Cov}(\varepsilon_{2i}, \varepsilon_{2i+1})/\sigma^2$ , it is naturally estimated by

$$\hat{\rho}_r = \frac{1}{\hat{\sigma}_r^2 |\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} \hat{\varepsilon}_{2i} \hat{\varepsilon}_{2i+1}.$$

We have checked that  $(\hat{\sigma}_r^2, \hat{\rho}_r)$  is the maximum likelihood estimator of  $(\sigma^2, \rho)$ .

**Proposition 61.** *In the BAR model (4.1), assume that the distribution of the ancestor  $X_1, \nu$ , has finite moments of all orders. Then  $((\hat{\sigma}_r^2, \hat{\rho}_r), r \in \mathbb{N})$  is a strongly consistent estimator of  $(\sigma^2, \rho)$ .*

*Proof.* Let us first deal with  $\hat{\sigma}_r^2$ . Observe that

$$\hat{\sigma}_r^2 = \frac{1}{2|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} \left( (X_{2i} - \hat{\alpha}_0^r X_i - \hat{\beta}_0^r)^2 + (X_{2n+1} - \hat{\alpha}_1^r X_n - \hat{\beta}_1^r)^2 \right) = \frac{1}{2|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} f(\Delta_i, \hat{\theta}^r)$$

where  $f(\Delta, \theta) = (y - \alpha_0 x - \beta_0)^2 + (z - \alpha_1 x - \beta_1)^2$ , with  $\Delta = (x, y, z)$  and  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ . Thus we have  $\hat{\sigma}_r^2 = \overline{M}_{\mathbb{T}_r}(f(\cdot, \theta))/2 + D_r$  with

$$D_r = \frac{1}{2|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} \left( f(\Delta_i, \hat{\theta}^r) - f(\Delta_i, \theta) \right).$$

Since  $f(\cdot, \theta) \in C_{\text{pol}}(\mathbb{R}^3)$ , we can apply Proposition 55-(ii):  $\overline{M}_{\mathbb{T}_r}(f(\cdot, \theta))$  a.s. converges to  $(\mu, P(f(\cdot, \theta)))$ . Now,  $P(f(\cdot, \theta))(x) = \mathbb{E}_x[(X_2 - \alpha_0 X_1 - \beta_0)^2 + (X_3 - \alpha_1 X_1 - \beta_1)^2] = \mathbb{E}[\varepsilon_2^2 + \varepsilon_3^2] = 2\sigma^2$ . Hence,  $\overline{M}_{\mathbb{T}_r}(f(\cdot, \theta))$  a.s. converges to  $2\sigma^2$ . Thus it is enough to prove that  $D_r$  a.s. tends to 0. Let us write  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) \in \Theta = (-1, 1) \times \mathbb{R} \times (-1, 1) \times \mathbb{R}$ . From the Taylor-Lagrange formula, for any  $\Delta \in \mathbb{R}^3$  and  $\theta, \theta' \in \Theta$ , we can find  $\lambda \in (0, 1)$  such that

$$f(\Delta, \theta') - f(\Delta, \theta) = \sum_{j=1}^4 (\theta'_j - \theta_j) \partial_{\theta_j} f(\Delta, \theta + \lambda(\theta' - \theta)).$$

Now, observing that  $f$  is a polynomial of global degree 4 and of degree 2 in each  $\theta_j$ , we can find  $g \in C_{\text{pol}}(\mathbb{R}^3)$  such that for all  $j \in \{1, 2, 3, 4\}$ ,  $\Delta \in \mathbb{R}^3$  and  $\theta \in \Theta$ ,  $|\partial_{\theta_j} f(\Delta, \theta)| \leq g(\Delta)(1 + \|\theta\|)$ . Therefore, for all  $r \in \mathbb{N}$ ,

$$\begin{aligned} |D_r| &\leq \frac{1}{2} \|\hat{\theta}^r - \theta\| \frac{1}{|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} g(\Delta_i) \left(1 + \|\theta\| + \|\hat{\theta}^r - \theta\|\right) \\ &= \frac{1}{2} \|\hat{\theta}^r - \theta\| \left(1 + \|\theta\| + \|\hat{\theta}^r - \theta\|\right) \overline{M}_{\mathbb{T}_r}(g). \end{aligned}$$

From Proposition 55-(ii),  $\overline{M}_{\mathbb{T}_r}(g)$  a.s. converges. Besides, Proposition 58 states that  $\|\hat{\theta}^r - \theta\|$  a.s. tends to 0. As a consequence, so does  $D_r$ . This completes the proof for  $\sigma^2$ . The proof for  $\rho$  is very similar.  $\square$

### 4.3.3 Detection of cellular aging

As explained in [42], detecting cellular aging boils down, in the BAR model (4.1), to rejecting hypothesis  $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ . Let us now build a statistical test that allows us to segregate between  $H_0$  and its alternative  $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$ . Wald's test is well adapted to the situation. We write  $\hat{\mu}_{1,r} = \mu_1(\hat{\theta}_r)$  and  $\hat{\mu}_{2,r} = \mu_2(\hat{\theta}_r, \hat{\sigma}_r)$  (recall (4.20)).

**Proposition 62.** *In the BAR model (4.1), assume that the distribution of the ancestor  $X_1$ ,  $\nu$ , has finite moments of all orders. Then the test statistic*

$$\chi_r^{(1)} = \frac{|\mathbb{T}_r|}{2\hat{\sigma}_r^2(1 - \hat{\rho}_r)} \left\{ (\hat{\alpha}_0^r - \hat{\alpha}_1^r)^2 (\hat{\mu}_{2,r} - \hat{\mu}_{1,r}^2) + \left( (\hat{\alpha}_0^r - \hat{\alpha}_1^r) \hat{\mu}_{1,r} + \hat{\beta}_0^r - \hat{\beta}_1^r \right)^2 \right\}$$

converges in distribution to  $\chi^2(2)$ , the  $\chi^2$  distribution with two degrees of freedom, under  $H_0$ , and almost surely diverges to  $+\infty$  under  $H_1$ .

*Proof.* Recall that  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ . Let us set  $g(\theta) = (\alpha_0 - \alpha_1, \beta_0 - \beta_1)^t$ . Then  $H_0 = \{g(\theta) = 0\}$ . From Proposition 60,  $|\mathbb{T}_r|^{1/2}(\hat{\theta}^r - \theta)$  converges in law to  $\mathcal{N}_4(0, \Sigma')$

so that  $|\mathbb{T}_r|^{1/2}(g(\hat{\theta}^r) - g(\theta))$  converges in law to  $\mathcal{N}_2(0, \Sigma'')$  with

$$\Sigma'' = dg(\theta)\Sigma'dg(\theta)^t = \sigma^2(I_2 \quad -I_2) \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix} \begin{pmatrix} I_2 \\ -I_2 \end{pmatrix} = 2\sigma^2(1 - \rho)K.$$

Under  $H_0$ ,  $g(\theta) = 0$  so that  $|\mathbb{T}_r|^{1/2}g(\hat{\theta}^r)$  converges in law to  $G \sim \mathcal{N}_2(0, \Sigma'')$ . Now, from (4.19),  $K = K(\theta, \sigma)$  is a continuous function of  $(\theta, \sigma) \in \Theta \times \mathbb{R}_+^*$  so that, letting  $\hat{K}_r = K(\hat{\theta}^r, \hat{\sigma}_r)$ , Propositions 58 and 61 imply that  $\hat{\Sigma}_r'' \equiv 2\hat{\sigma}_r^2(1 - \hat{\rho}_r)\hat{K}_r$  converges in probability to  $\Sigma''$ . By continuity of  $G \mapsto \Sigma''^{-1/2}G$ , Slutsky's theorem shows that  $|\mathbb{T}_r|^{1/2}\hat{\Sigma}_r''^{-1/2}g(\hat{\theta}^r)$  converges in law to  $\mathcal{N}_2(0, I_2)$ . In particular,

$$\left\| |\mathbb{T}_r|^{1/2}\hat{\Sigma}_r''^{-1/2}g(\hat{\theta}^r) \right\|^2 = |\mathbb{T}_r| g(\hat{\theta}^r)^t \hat{\Sigma}_r''^{-1} g(\hat{\theta}^r) = \frac{|\mathbb{T}_r|}{2\hat{\sigma}_r^2(1 - \hat{\rho}_r)} g(\hat{\theta}^r)^t \hat{K}_r^{-1} g(\hat{\theta}^r) = \chi_r^{(1)}$$

converges to the  $\chi^2$  distribution with two degrees of freedom.

Under  $H_1$ ,  $\chi_r/|\mathbb{T}_r| = g(\hat{\theta}^r)^t \hat{\Sigma}_r''^{-1} g(\hat{\theta}^r)$  a.s. converges to  $g(\theta)^t \Sigma''^{-1} g(\theta) > 0$  so that  $\chi_r^{(1)}$  a.s. diverges to  $+\infty$ .  $\square$

The same technique may be used to test  $\{\alpha_0 = 0\}$ ,  $\{\alpha_1 = 0\}$ ,  $\{\beta_0 = 0\}$ ,  $\{\beta_1 = 0\}$ ,  $\{\alpha_0 = \alpha_1\}$  and  $\{\beta_0 = \beta_1\}$ . For instance, we have

**Proposition 63.** *In the BAR model (4.1), assume that the distribution of the ancestor  $X_1, \nu$ , has finite moments of all orders. Then*

(i) *the test statistic*

$$\chi_r^{(2)} = |\mathbb{T}_r| (\hat{\alpha}_0^r)^2 \frac{\hat{\mu}_{2,r} - \hat{\mu}_{1,r}^2}{\hat{\sigma}_r^2}$$

*converges in distribution to  $\chi^2(1)$  under  $\{\alpha_0 = 0\}$  and diverges in probability under  $\{\alpha_0 \neq 0\}$ ,*

(ii) *the test statistic*

$$\chi_r^{(3)} = |\mathbb{T}_r| \left( \hat{\beta}_0^r \right)^2 \frac{\hat{\mu}_{2,r} - \hat{\mu}_{1,r}^2}{\hat{\mu}_{2,r} \hat{\sigma}_r^2}$$

*converges in distribution to  $\chi^2(1)$  under  $\{\beta_0 = 0\}$  and diverges in probability under  $\{\beta_0 \neq 0\}$ ,*

(iii) *the test statistic*

$$\chi_r^{(4)} = |\mathbb{T}_r| (\hat{\alpha}_0^r - \hat{\alpha}_1^r)^2 \frac{\hat{\mu}_{2,r} - \hat{\mu}_{1,r}^2}{2\hat{\sigma}_r^2(1 - \hat{\rho}_r)}$$

*converges in distribution to  $\chi^2(1)$  under  $\{\alpha_0 = \alpha_1\}$  and diverges in probability under  $\{\alpha_0 \neq \alpha_1\}$ .*

(iv) the test statistic

$$\chi_r^{(5)} = |\mathbb{T}_r| \left( \hat{\beta}_0^r - \hat{\beta}_1^r \right)^2 \frac{\hat{\mu}_{2,r} - \hat{\mu}_{1,r}^2}{2\hat{\mu}_{2,r}\hat{\sigma}_r^2(1 - \hat{\rho}_r)}$$

converges in distribution to  $\chi^2(1)$  under  $\{\beta_0 = \beta_1\}$  and diverges in probability under  $\{\beta_0 \neq \beta_1\}$ .

We now present a statistical test that allows to segregate between  $H'_0 = \{\beta_0/(1 - \alpha_0) = \beta_1/(1 - \alpha_1)\}$  and its alternative  $H'_1$ . This allows to test if the two fixed points corresponding to the two affine regressions of the BAR model (4.1) are equal. This may happen even if  $(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)$ . Rejecting  $H'_0$  means accepting that the new pole and the old pole populations are even distinct in mean. Again we use Wald's test, since  $H'_0 = \{g(\theta) = 0\}$  with  $g(\theta) = \beta_0/(1 - \alpha_0) - \beta_1/(1 - \alpha_1)$ . The proof is obvious and not detailed here.

**Proposition 64.** *In the BAR model (4.1), assume that the distribution of the ancestor  $X_1, \nu$ , has finite moments of all orders. Let*

$$dg(\theta) = \left( \frac{\beta_0}{(1 - \alpha_0)^2}, \quad \frac{1}{1 - \alpha_0}, \quad \frac{-\beta_1}{(1 - \alpha_1)^2}, \quad \frac{1}{1 - \alpha_1} \right)$$

be the gradient of  $g$  and  $\hat{s}_r^2 = dg(\hat{\theta}_r) \hat{\Sigma}'_r dg(\hat{\theta}_r)^t$ , where  $\hat{\Sigma}'_r$  is  $\Sigma'$  evaluated in  $(\hat{\theta}_r, \hat{\sigma}_r, \hat{\rho}_r)$ , and  $\Sigma'$  is defined in (4.21). Then the test statistic

$$\chi_r^{(6)} = \frac{|\mathbb{T}_r|}{\hat{s}_r^2} \left( \frac{\hat{\beta}_0^r}{1 - \hat{\alpha}_0^r} - \frac{\hat{\beta}_1^r}{1 - \hat{\alpha}_1^r} \right)^2$$

converges in distribution to  $\chi^2(1)$  under  $H'_0$  and a.s. diverges to  $+\infty$  under  $H'_1$ .

In the case when  $\alpha_0 = \alpha_1 = 0$ , testing  $H_0$  or  $H'_0$  boils down to testing  $\{\beta_0 = \beta_1\}$ :

**Proposition 65.** *In the BAR model (4.1), assume that  $\alpha_0 = \alpha_1 = 0$  and that the distribution of the ancestor  $X_1, \nu$ , has finite moments of all orders. Then the test statistic*

$$\xi_r = \frac{1}{\hat{\sigma}_r \sqrt{2|\mathbb{T}_r|(1 - \hat{\rho}_r)}} \sum_{i \in \mathbb{T}_r} (X_{2i} - X_{2i+1})$$

converges in distribution to  $\mathcal{N}(0, 1)$  under  $\{\beta_0 = \beta_1\}$  and almost surely tends to  $+\infty$  (resp.  $-\infty$ ) under  $\{\beta_0 > \beta_1\}$  (resp.  $\{\beta_0 < \beta_1\}$ ).

*Proof.* Let  $f(x, y, z) = y - z$ . Observe that  $f \in C_{\text{pol}}(\mathbb{R}^3)$  and that  $Pf(x) = \beta_0 - \beta_1$ , since  $\alpha_0 = \alpha_1 = 0$ .

Let us assume that  $\beta_0 = \beta_1$ . Then  $Pf = 0$  and Proposition 54-(iii) implies that  $\hat{\sigma}_r \sqrt{2(1 - \hat{\rho}_r)} \xi_r$  converges in distribution to  $\mathcal{N}(0, s^2)$ , where  $s^2 = (\mu, Pf^2) =$

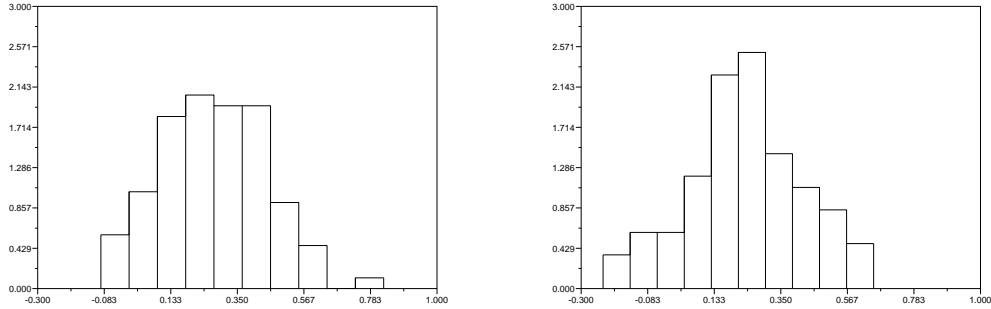


FIG. 4.8 – Histograms of  $\hat{\alpha}_0$  (left) and  $\hat{\alpha}_1$  (right)

$2\sigma^2(1-\rho)$ . Now,  $(\hat{\sigma}_r^2, \hat{\rho}_r)$  a.s. converges to the constant  $(\sigma^2, \rho)$  so that, with Slutsky's theorem,  $\xi_r$  converges in distribution to  $\mathcal{N}(0, 1)$ .

Let us now assume that  $\beta_0 \neq \beta_1$ . Proposition 55 states that  $|\mathbb{T}_r|^{-1/2} \hat{\sigma}_r \sqrt{2(1-\hat{\rho}_r)} \xi_r = |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} f(\Delta_i)$  a.s. converges to  $(\mu, Pf) = \beta_0 - \beta_1 \neq 0$ . Since  $\hat{\sigma}_r \sqrt{2(1-\hat{\rho}_r)}$  converges in probability to  $\sigma \sqrt{2(1-\rho)} > 0$ , we conclude that  $\xi_r$  a.s. diverges as  $|\mathbb{T}_r|^{1/2}$ , to  $+\infty$  if  $\beta_0 > \beta_1$ , and to  $-\infty$  if  $\beta_0 < \beta_1$ .  $\square$

**Remark 66.** Note that in the model where  $\alpha_0 = \alpha_1 = 0$ ,  $\chi_r^{(1)}$  would read

$$\chi_r^{(1)} = \frac{|\mathbb{T}_r|}{2\hat{\sigma}_r^2(1-\hat{\rho}_r)} \left( \hat{\beta}_0^r - \hat{\beta}_1^r \right)^2$$

and is thus equal to  $\xi_r^2$ . The latter test looks like the ones E. J. STEWART *et al.* performed in [89]: it focuses on the differences  $X_{2i} - X_{2i+1}$  between sisters. But it is relevant only in the case when the correlation parameters  $\alpha_0$  and  $\alpha_1$  are zero, that is in a dynamics with no memory. Now, the data analysis strongly rejects this assumption, as we shall see.

### 4.3.4 Data numerical analysis

We now perform the estimation and test procedures on E. J. STEWART *et al.*'s data. It consists of 95 films, and each film should be seen as an incomplete binary tree of growth rates. How do we compute the estimators and test statistics? According to the above presentation, we should restrict the observation to the bigger complete subtree  $\mathbb{T}_{r+1}$ . We actually take into account all the observations, noting that

- very few cells are observed in a generation, say  $r$ , when generation  $r-1$  is not completely observed,

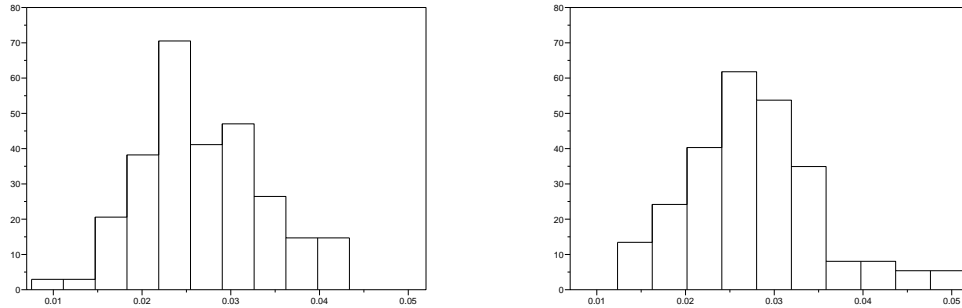


FIG. 4.9 – Histograms of  $\hat{\beta}_0$  (left) and  $\hat{\beta}_1$  (right)

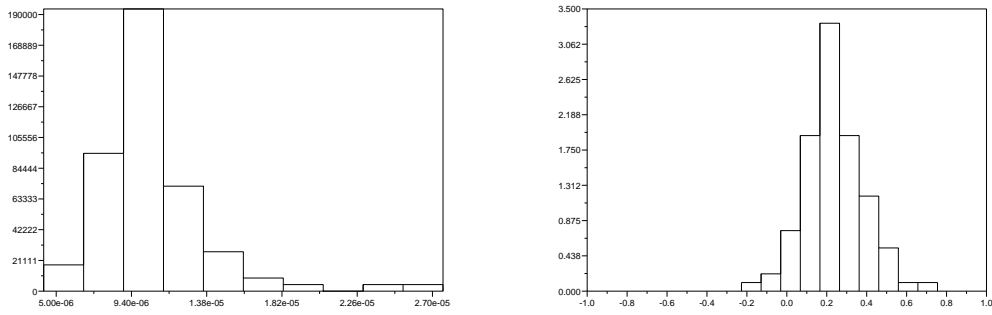


FIG. 4.10 – Histograms of  $\hat{\sigma}^2$  (left) and  $\hat{\rho}$  (right)

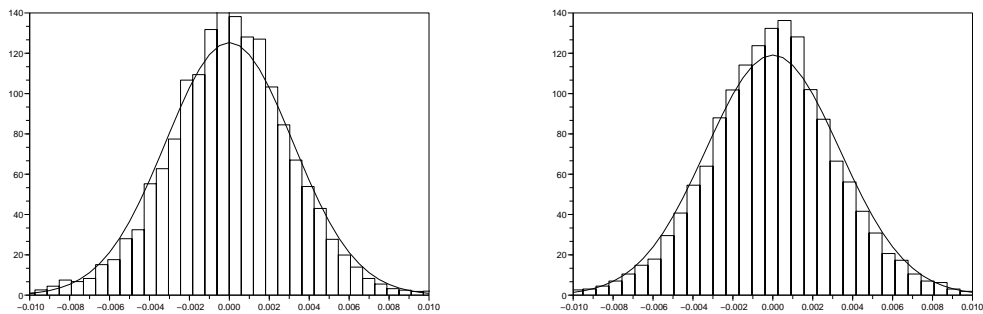


FIG. 4.11 – Histograms of the residues  $\hat{\varepsilon}_{2n}$  (new poles, left) and  $\hat{\varepsilon}_{2n+1}$  (old poles, right), and their fit to Gaussian distributions

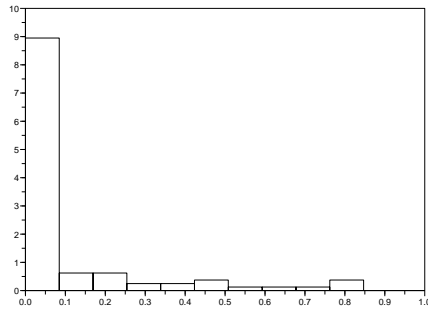


FIG. 4.12 –  $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ . Histogram of the  $p$ -values  $\mathbb{P}(\chi^2(2) \geq \chi_{obs}^{(1)})$

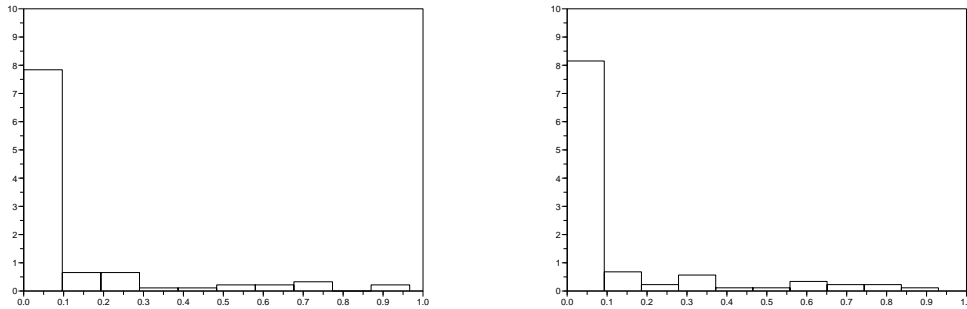


FIG. 4.13 – Left:  $H_0 = \{\alpha_0 = 0\}$ . Histogram of the  $p$ -values. Right:  $H_0 = \{\alpha_1 = 0\}$ . Histogram of the  $p$ -values

- cells observed in the last generation are assumed to be the result of a random permutation  $\Pi$ , independent of  $X$  ; this should be correct as a first approximation.

Figures 4.8, 4.9 and 4.10 synthetize the 95 estimations. We see that the Gaussian nature is not perfectly checked, this may be because the law of the estimators depend on the experiment - a slight change in the experimental conditions may involve a change in the parameters  $\theta$ ,  $\sigma^2$  and  $\rho$ . Nevertheless, the six histograms in these figures are qualitatively not very far from Gaussian.

Figure 4.11 gives the global empirical distribution of the residues  $\hat{\varepsilon}$  over the 95 films. We have separated new poles' residues (left) from old poles' ones (right), and fitted to normal distributions. Both histograms are close to Gaussian laws.

Figure 4.12 shows that  $H_0$  can be strongly rejected. This indicates that the dynamics of the growth rate of the old pole offspring is different from that of the new pole offspring. The nullity of any parameter ( $\alpha_0$ ,  $\beta_0$ ,  $\alpha_1$  or  $\beta_1$ ) can be strongly rejected as well: he 95  $p$ -values are extremely small for  $\beta_0$  and  $\beta_1$  (they are all

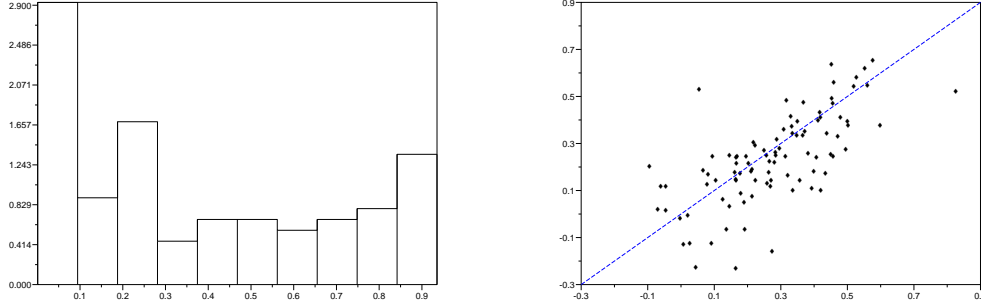


FIG. 4.14 –  $H_0 = \{\alpha_0 = \alpha_1\}$ ; left: histogram of the  $p$ -values  $\mathbb{P}(\chi^2(1) \geq \chi_{obs}^{(4)})$ ; right:  $\hat{\alpha}_0$  on the  $x$ -axis,  $\hat{\alpha}_1$  on the  $y$ -axis; the dashed line is the diagonal

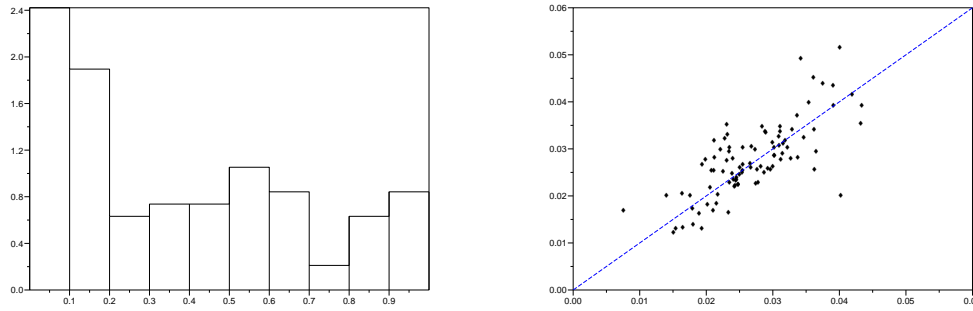


FIG. 4.15 –  $H_0 = \{\beta_0 = \beta_1\}$ ; left: histogram of the  $p$ -values  $\mathbb{P}(\chi^2(1) \geq \chi_{obs}^{(5)})$ ; right:  $\hat{\beta}_0$  on the  $x$ -axis,  $\hat{\beta}_1$  on the  $y$ -axis; the dashed line is the diagonal

less than  $10^{-2}$ ), and Figure 4.13 shows the  $p$ -values for the  $\alpha$ 's. . This enlightens the relevance of a Markovian modelisation with memory one: the mother cell is a significant predictor of offspring growth rate in general.

Besides, from Figures 4.14 and 4.15 we cannot reject the hypothesis that both  $\alpha$ 's are equal on the one hand, and that both  $\beta$ 's are equal on the other hand. This is confirmed by looking at the bivariate scatter plots of  $\alpha_1$  versus  $\alpha_0$  and  $\beta_1$  versus  $\beta_0$ : they lie on both sides of the diagonal. But we strongly reject that both fixed points, namely  $\gamma_0 = \beta_0/(1 - \alpha_0)$  and  $\gamma_1 = \beta_1/(1 - \alpha_1)$ , are equal, see Figures 4.16 and 4.17. Hence the parametrization  $(\alpha, \gamma)$ , which makes more physical sense than the parametrization  $(\alpha, \beta)$ , has the following advantage: with no assumption on the  $\alpha$ 's, we can detect aging by looking only at the  $\gamma$ 's, which we cannot do with the  $\beta$ 's. It also means that the new poles and the old poles are not only different in distribution, but also in mean.



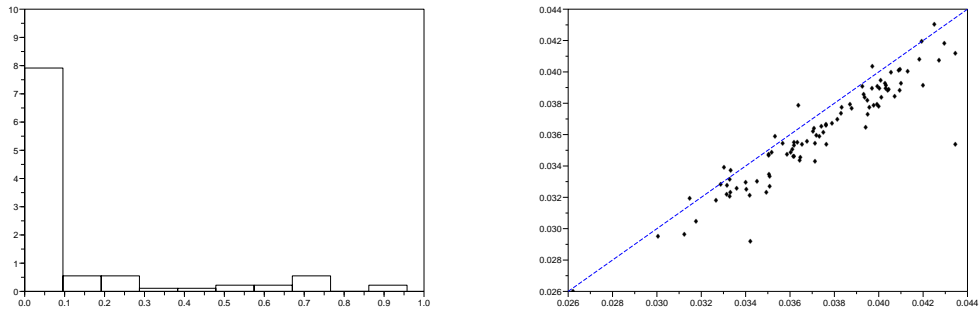


FIG. 4.16 -  $H'_0 = \{\beta_0/(1 - \alpha_0) = \beta_1/(1 - \alpha_1)\}$ ; left: histogram of the  $p$ -values  $\mathbb{P}(\chi^2(1) \geq \chi_{obs}^{(6)})$ ; right:  $\hat{\beta}_0/(1 - \hat{\alpha}_0)$  on the  $x$ -axis,  $\hat{\beta}_1/(1 - \hat{\alpha}_1)$  on the  $y$ -axis; the dashed line is the diagonal

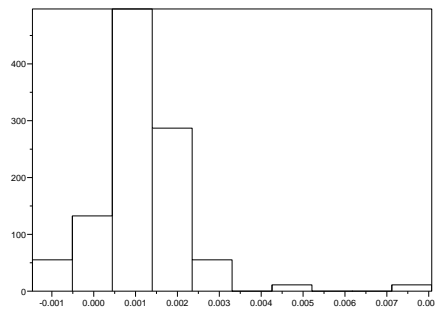


FIG. 4.17 - Histogram of  $\hat{\beta}_0/(1 - \hat{\alpha}_0) - \hat{\beta}_1/(1 - \hat{\alpha}_1)$

The scatter plot in Figure 4.16 indicates that  $\gamma_0 > \gamma_1$ . More precisely, the line  $\gamma_0 = \gamma_1 + \delta$  fits well the data with  $\delta$  significantly positive. Numerically,  $\delta \approx 0.0012 \pm 0.0011$ , or  $0.0011 \pm 0.0008$  if we delete the two aberrant points in Figure 4.16 (right). This may be seen as statistical evidence of aging in *E. Coli*, since on average old pole cells grow slower than the new pole cells, which is characteristic of aged individuals. Quantitatively, they seem to grow 3% slower (we may speak in terms of percentage since the range of values of  $\gamma$ 's is narrow). This result is close to E. J. STEWART *et al.*'s original calculations, since in [89] they estimated this ratio to be around 2%.

# Bibliographie

- [1] Christophe Andrieu and Eric Moulines. On the ergodicity properties of some adaptive mcmc algorithms. *To appear in Ann. Appl. Probab.*, 2006.
- [2] Krishna B. Athreya and Hye-Jeong Kang. Some limit theorems for positive recurrent branching Markov chains. I, II. *Adv. in Appl. Probab.*, 30(3) :693–710, 711–722, 1998.
- [3] Vlad Bally and Denis Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1) :43–60, 1996.
- [4] Vlad Bally and Denis Talay. The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods Appl.*, 2(2) :93–128, 1996.
- [5] Ishwar Basawa and Jin Zhou. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.*, 41A :55–64, 2004. Stochastic methods and their applications.
- [6] Itai Benjamini and Yuval Peres. Markov chains indexed by trees. *Ann. Probab.*, 22(1) :219–243, 1994.
- [7] Albert Benveniste, Michel Métivier, and Pierre Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990.
- [8] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81 :635–654, 1973.
- [9] Alexandr Alekseevich Borovkov. *Ergodicity and stability of stochastic processes*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1998.
- [10] Mireille Bossy and Denis Talay. Convergence rate for the approximation of the limit law of weakly interacting particles : application to the Burgers equation. *Ann. Appl. Probab.*, 6(3) :818–861, 1996.
- [11] Mireille Bossy and Denis Talay. A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Math. Comp.*, 66(217) :157–192, 1997.

- [12] Bruno Bouchard, Ivar Ekeland, and Nizar Touzi. On the Malliavin approach to Monte Carlo approximation of conditional expectations. *Finance Stoch.*, 8(1) :45–71, 2004.
- [13] Philippe Bougerol and Nico Picard. Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20(4) :1714–1730, 1992.
- [14] Mark Broadie and Paul Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2) :269–285, 1996.
- [15] Evelyn Buckwar and Tony Shardlow. Weak approximation of stochastic differential delay equations. *IMA J. Numer. Anal.*, 25(1) :57–86, 2005.
- [16] Quang Bui and Richard Huggins. Robust inference for the bivariate bifurcating autoregressive model. *Aust. N. Z. J. Stat.*, 40(2) :151–163, 1998.
- [17] Quang Bui and Richard Huggins. Inference for the random coefficients bifurcating autoregressive model for cell lineage studies. *J. Statist. Plann. Inference*, 81(2) :253–262, 1999.
- [18] Emmanuelle Clément, Arturo Kohatsu-Higa, and Damien Lamberton. A duality approach for the weak approximation of stochastic differential equations,. *to appear in Ann. Appl. Probab.*, 2006.
- [19] Rama Cont and Peter Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.
- [20] Richard Cowan and Robert Staudte. The bifurcating autoregression model in cell lineage studies. *Biometrics*, 42 :769–783, 1986.
- [21] Ana Bela Cruzeiro, Paul Malliavin, and Anton Thalmaier. Geometrization of Monte-Carlo numerical analysis of an elliptic operator : strong approximation. *C. R. Math. Acad. Sci. Paris*, 338(6) :481–486, 2004.
- [22] Amir Dembo, Peter Mörters, and Scott Sheffield. Large deviations of Markov chains indexed by random trees. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(6) :971–996, 2005.
- [23] Jérôme B. Detemple and Carlton Osakwe. The valuation of volatility options. CIRANO Working Papers 99s-43, CIRANO, November 1999. available at <http://ideas.repec.org/p/cir/cirwor/99s-43.html>.
- [24] Michelangelo Diligenti, Paolo Frasconi, and Marco Gori. Hidden tree markov models for document image classification. *IEEE Transactions on pattern analysis and machine intelligence*, 25(4) :519–523, 2003.
- [25] Marie Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.
- [26] Bruno Dupire. Pricing and hedging with smiles. In *Mathematics of derivative securities (Cambridge, 1995)*, volume 15 of *Publ. Newton Inst.*, pages 103–111. Cambridge Univ. Press, Cambridge, 1997.

- [27] Jean-Baptiste Durand, Yann Guédon, Yves Caraglio, and Evelyne Costes. Analysis of the plant architecture via tree-structured statistical models : The hidden markov tree models. *New Phytologist*, 166(3) :813–825, 2005.
- [28] Nicole El Karoui. Couverture des risques dans les marchés financiers. *Notes de cours de Master, Université Paris VI*, 2004.
- [29] Jean-Pierre Fouque, George Papanicolaou, and Ronnie Sircar. *Derivatives in financial markets with stochastic volatility*. Cambridge University Press, Cambridge, 2000.
- [30] Jean-Pierre Fouque, George Papanicolaou, Ronnie Sircar, and Knut Solna. Multiscale stochastic volatility asymptotics. *Multiscale Model. Simul.*, 2(1) :22–42 (electronic), 2003.
- [31] Jean-Pierre Fouque, George Papanicolaou, Ronnie Sircar, and Knut Solna. Singular perturbations in option pricing. *SIAM J. Appl. Math.*, 63(5) :1648–1665 (electronic), 2003.
- [32] Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, and Pierre-Louis Lions. Applications of Malliavin calculus to Monte-Carlo methods in finance. II. *Finance Stoch.*, 5(2) :201–236, 2001.
- [33] Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions, and Nizar Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, 3(4) :391–412, 1999.
- [34] Nicolas Fournier and Jean-Sébastien Giet. Existence of densities for jumping SDE. *To appear in Stoch. Proc. Appl.*, pages –, 2006.
- [35] Avner Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, New Jersey, 1964.
- [36] Paul Glasserman and David Yao. Some guidelines and guarantees of common random numbers. *Management Science*, 38 :884–908, 1992.
- [37] Peter Glynn. Optimization of stochastic systems via simulation. *Proceedings of the 1989 Winter Simulation Conference*, pages 90–105, 1989.
- [38] Emmanuel Gobet and Stéphane Menozzi. Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme. *Stochastic Process. Appl.*, 112(2) :201–223, 2004.
- [39] Emmanuel Gobet and Rémi Munos. Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control. *SIAM J. Control Optim.*, 43(5) :1676–1713 (electronic), 2005.
- [40] Julien Guyon. Euler scheme and tempered distributions. *Stoch. Proc. Appl.*, 116(6) :877–904, 2006.
- [41] Julien Guyon. Limit theorems for bifurcating markov chains. application to the detection of cellular aging. CERMICS Preprints 301, CERMICS, February

2006. available at <http://cermics.enpc.fr/reports/CERMICS-2006/CERMICS-2006-301.pdf>.
- [42] Julien Guyon, Ariane Bize, Grégory Paul, Eric Stewart, Jean-François Delmas, and François Taddéi. Statistical study of cellular aging. *ESAIM Proceedings, CEMRACS 2004 Proceedings*, 14 :100–114, 2005.
  - [43] Patrick Hagan, Deep Kumar, Andrew Lesniewski, and Diana Woodward. Managing smile risk. *WILMOTT Magazine*, September :84–108, 2002.
  - [44] Steven Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies, Oxford University Press for Society for Financial Studies*, 6(2) :327–343, 1993.
  - [45] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. Optimal approximation of stochastic differential equations by adaptive step-size control. *Math. Comp.*, 69(231) :1017–1034, 2000.
  - [46] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. The optimal discretization of stochastic differential equations. *J. Complexity*, 17(1) :117–153, 2001.
  - [47] Yaozhong Hu, Salah-Eldin A. Mohammed, and Feng Yan. Discrete-time approximations of stochastic delay equations : the Milstein scheme. *Ann. Probab.*, 32(1A) :265–314, 2004.
  - [48] Yaozhong Hu and Shinzo Watanabe. Donsker’s delta functions and approximation of heat kernels by the time discretization methods. *J. Math. Kyoto Univ.*, 36(3) :499–518, 1996.
  - [49] Richard Huggins. Robust inference for variance components models for single trees of cell lineage data. *Ann. Statist.*, 24(3) :1145–1160, 1996.
  - [50] Richard Huggins and Ishwar Basawa. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.*, 36(4) :1225–1233, 1999.
  - [51] Richard Huggins and Ishwar Basawa. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.*, 42(4) :423–432, 2000.
  - [52] Richard Huggins and Robert Staudte. Variance components models for dependent cell populations. *J. Amer. Statist. Assoc.*, 89 :19–29, 1994.
  - [53] John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42 :281–300, 1987.
  - [54] Jean Jacod. The Euler scheme for Lévy driven stochastic differential equations : limit theorems. *Ann. Probab.*, 32(3A) :1830–1872, 2004.
  - [55] Jean Jacod, Thomas G. Kurtz, Sylvie Méléard, and Philip Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3) :523–558, 2005.

- [56] Jean Jacod, Sylvie Méléard, and Philip Protter. Explicit form and robustness of martingale representations. *Ann. Probab.*, 28(4) :1747–1780, 2000.
- [57] Jean Jacod and Philip Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, 26(1) :267–307, 1998.
- [58] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [59] Arturo Kohatsu-Higa and Miquel Montero. Malliavin calculus in finance. In *Handbook of computational and numerical methods in finance*, pages 111–174. Birkhäuser Boston, Boston, MA, 2004.
- [60] Arturo Kohatsu-Higa and Roger Pettersson. Variance reduction methods for simulation of densities on Wiener space. *SIAM J. Numer. Anal.*, 40(2), 2002.
- [61] Valentin Konakov and Enno Mammen. Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods Appl.*, 8(3) :271–285, 2002.
- [62] Uwe Küchler and Eckhard Platen. Strong discrete time approximation of stochastic differential equations with time delay. *Math. Comput. Simulation*, 54(1-3) :189–205, 2000.
- [63] Uwe Küchler and Eckhard Platen. Weak discrete time approximation of stochastic differential equations with time delay. *Math. Comput. Simulation*, 59(6) :497–507, 2002.
- [64] Hiroshi Kunita. On the decomposition of solutions of stochastic differential equations. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, volume 851 of *Lecture Notes in Math.*, pages 213–255. Springer, Berlin, 1981.
- [65] Shigeo Kusuoka. Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara '98*, volume 31 of *Adv. Stud. Pure Math.*, pages 147–165. Math. Soc. Japan, Tokyo, 2001.
- [66] Shigeo Kusuoka and Daniel Stroock. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, volume 32 of *North-Holland Math. Library*, pages 271–306. North-Holland, Amsterdam, 1984.
- [67] Shigeo Kusuoka and Daniel Stroock. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 32(1) :1–76, 1985.
- [68] Shigeo Kusuoka and Daniel Stroock. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 34(2) :391–442, 1987.
- [69] Pierre L’Ecuyer and Gaétan Perron. On the convergence rates of IPA and FDC derivative estimators. *Oper. Res.*, 42(4) :643–656, 1994.

- [70] Wen Liu and Liying Wang. The Markov approximation of the random fields on Cayley trees and a class of small deviation theorems. *Statist. Probab. Lett.*, 63(2) :113–121, 2003.
- [71] Wen Liu and Weiguo Yang. Some strong limit theorems for Markov chain fields on trees. *Probab. Engrg. Inform. Sci.*, 18(3) :411–422, 2004.
- [72] Terry Lyons and Nicolas Victoir. Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 460(2041) :169–198, 2004. Stochastic analysis with applications to mathematical finance.
- [73] Faïza Maâouia and Abderrahmen Touati. Identification of multitype branching processes. *Ann. Stat.*, 18(3) :2655–2694, 2005.
- [74] Vigirdas Mackevičius. On the convergence rate of Euler scheme for SDE with Lipschitz drift and constant diffusion. In *Proceedings of the Eighth Vilnius Conference on Probability Theory and Mathematical Statistics, Part I (2002)*, volume 78, pages 301–310, 2003.
- [75] Paul Malliavin and Anton Thalmaier. Numerical error for SDE : asymptotic expansion and hyperdistributions. *C. R. Math. Acad. Sci. Paris*, 336(10) :851–856, 2003.
- [76] Sean Meyn and Richard Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.
- [77] Grigori N. Milstein, John G. M. Schoenmakers, and Vladimir Spokoiny. Transition density estimation for stochastic differential equations via forward-reverse representations. *Bernoulli*, 10(2) :281–312, 2004.
- [78] Kyoung-Sook Moon, Anders Szepessy, Raúl Tempone, and Georgios Zouraris. Convergence rates for adaptive weak approximation of stochastic differential equations. *Stoch. Anal. Appl.*, 23(3) :511–558, 2005.
- [79] Syoiti Ninomiya. A new simulation scheme of diffusion processes : application of the Kusuoka approximation to finance problems. *Math. Comput. Simulation*, 62(3-6) :479–486, 2003. 3rd IMACS Seminar on Monte Carlo Methods—MCM 2001 (Salzburg).
- [80] Syoiti Ninomiya. A partial sampling method applied to the Kusuoka approximation. *Monte Carlo Methods Appl.*, 9(1) :27–38, 2003.
- [81] Ivan Nourdin and Andreas Neuenkirch. Exact rate of convergence of some approximation schemes associated to SDEs driven by a fractional brownian motion. *Preprint PMA-1043*, 2006.
- [82] David Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.



- [83] Bernt Øksendal. *Stochastic differential equations, 5th edition*. Springer-Verlag, Berlin, 1998.
- [84] Philip Protter and Denis Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1) :393–423, 1997.
- [85] Eric Renault and Nizar Touzi. Option pricing and implied volatilities in a stochastic volatility model. *Mathematical Finance*, 6 :279–302, 1996.
- [86] Chris Rogers. Smooth transition densities for one-dimensional diffusions. *Bull. London Math. Soc.*, 17 :157–161, 1985.
- [87] Robert Staudte, Richard Huggins, J. Zhang, David Axelrod, and Marek Kimmel. Estimating clonal heterogeneity and interexperiment variability with the bifurcating autoregressive model for cell lineage data. *Math. Biosci.*, 143(2) :103–121, 1997.
- [88] Elias Stein and Jeremy Stein. Stock-price distributions with stochastic volatility - an analytic approach. *Review of Financial Studies*, 4 :727–752, 1991.
- [89] Eric Stewart, Richard Madden, Grégory Paul, and François Taddéi. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol*, 3(2) :e45, 2005.
- [90] Anders Szepessy, Raúl Tempone, and Georgios E. Zouraris. Adaptive weak approximation of stochastic differential equations. *Comm. Pure Appl. Math.*, 54(10) :1169–1214, 2001.
- [91] Christiane Takacs. Strong law of large numbers for branching Markov chains. *Markov Process. Related Fields*, 8(1) :107–116, 2002.
- [92] Denis Talay. Probabilistic numerical methods for partial differential equations : Elements of analysis. In *Graham, C. (ed.) et al., Probabilistic Models for Non-linear Partial Differential Equations. Lectures given at the 1st session of the Centro Internazionale Matematico Estivo, Montecatini Terme, Italy, May 22-30, 1995*. Berlin : Springer. *Lect. Notes Math.*, 1627 :148–196, 1996.
- [93] Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4) :483–509 (1991), 1990.
- [94] Denis Talay and Ziyu Zheng. Approximation of quantiles of components of diffusion processes. *Stochastic Process. Appl.*, 109(1) :23–46, 2004.
- [95] Etienne Tanré. *Étude probabiliste des équations de Smoluchowski ; Schéma d'Euler pour des fonctionnelles ; Amplitude du mouvement brownien avec dérive*. PhD thesis, Université Henri Poincaré Nancy 1, 2001.
- [96] YongHong Tian, Tiejun Huang, and Wen Gao. Two-phase web site classification based on hidden markov tree models. *International Journal : Web Intelligence and Agent System*, 4(2) :249–264, 2004.

- [97] Liqing Yan. The Euler scheme with irregular coefficients. *Ann. Probab.*, 30(3) :1172–1194, 2002.
- [98] Weiguo Yang and Wen Liu. Strong law of large numbers and Shannon-McMillan theorem for Markov chain fields on trees. *IEEE Trans. Inform. Theory*, 48(1) :313–318, 2002.