

N° d'ordre : 3492

THESE

En vue de l'obtention du : **DOCTORAT**

Structure de Recherche : Laboratoire Mathématiques, Informatique et Applications - Sécurité de l'Information

Discipline : Mathématiques appliquées

Spécialité : Statistiques

Présentée et soutenue le 03/07/2021 par :

Soufiane OUAMALICHE

Méthodes de lissage par noyaux et application à la finance de marché

JURY

Zine El Abidine GUENNOUN	PES, Université Mohammed V de Rabat, Faculté des Sciences	Président
Rachid SAADANE	PES, Ecole Hassania des Travaux Publics, Casablanca	Rapporteur/Examinateur
Souad EL BERNOUSSI	PES, Université Mohammed V de Rabat, Faculté des Sciences	Rapporteur/Examinateur
Sidi Mohamed DOURI	PH, Université Mohammed V de Rabat, Faculté des Sciences	Rapporteur/Examinateur
Samir HAKAM	PES, Université Mohammed V de Rabat, Faculté des Sciences	Examinateur
Awatef SAYAH	PES, Université Mohammed V de Rabat, Faculté des Sciences	Directrice de thèse

Année Universitaire : 2020/2021

Remerciements

Ce travail a été effectué au sein du Laboratoire Mathématiques, Informatique et Applications – Sécurité de l’Information de la Faculté des Sciences de l’Université Mohammed V de Rabat sous la direction de Madame Awatef SAYAH.

Je tiens tout d’abord à exprimer toute ma gratitude et ma reconnaissance à ma directrice de thèse Madame Awatef SAYAH, Professeure de l’Enseignement Supérieur à la Faculté des Sciences de l’Université Mohammed V de Rabat, pour la confiance qu’elle m’a accordée, pour son soutien, sa disponibilité et ses encouragements sans lesquels ce travail ne serait jamais arrivé à terme.

J’adresse mes vifs remerciements à Monsieur Zine El Abidine GUENNOUN, Professeur de l’Enseignement Supérieur à la Faculté des Sciences de l’Université Mohammed V de Rabat, de l’honneur qu’il m’a fait en acceptant de présider le jury de thèse.

Je remercie chaleureusement Monsieur Rachid SAADANE, Professeur de l’Enseignement Supérieur à l’Ecole Hassania des Travaux Publics de Casablanca, d’avoir accepté de rapporter et d’examiner ma thèse. Qu’il soit aussi remercié pour sa franchise, sa gentillesse et ses encouragements constants.

Un grand merci à Madame Souad EL BERNOUSSI, Professeure de l’Enseignement Supérieur à la Faculté des Sciences de l’Université Mohammed V de Rabat, d’avoir accepté de rapporter et d’examiner ma thèse. Je la remercie également pour son enthousiasme et sa sympathie ainsi que l’atmosphère amicale et accueillante qu’elle a créée au sein du laboratoire Mathématiques, Informatique et Applications – Sécurité de l’information.

Je tiens également à remercier Monsieur Sidi Mohamed DOUIRI, Professeur habilité à la Faculté des Sciences de l’Université Mohammed V de Rabat, d’avoir accepté de rapporter et d’examiner ma thèse.

Je remercie aussi Monsieur Samir HAKAM, Professeur de l’Enseignement Supérieur à la Faculté des Sciences de l’Université Mohammed V de Rabat, pour avoir accepté d’examiner ma thèse. Je le remercie également pour tout l’enthousiasme qu’il a eu envers mon travail, pour ses commentaires et pour ses remarques pertinentes qui m’ont été d’une très grande utilité.

Je réitère mes remerciements à Ma directrice de thèse, cette fois-ci, pour m’avoir ouvert les portes de l’enseignement et de m’avoir permis de découvrir et de nourrir au fil des années la passion que j’ai pour ce métier.

Un grand merci du fond du cœur à ma chère famille pour leur soutien permanent et inébranlable.

Je tiens à remercier également mes amis de longue et de récente date pour leurs encouragements et je leur souhaite une bonne chance pour l’achèvement de leurs travaux.

Résumé

L'estimation non paramétrique constitue un outil fondamental en statistique permettant de repousser les limites établies par les principes sur lesquels reposent les techniques paramétriques et par conséquent, permet d'aborder des problèmes où l'approche paramétrique soit échoue, soit donne des résultats qui peuvent être significativement améliorés. Le travail conduit au cours de cette thèse s'inscrit dans ce contexte via l'adoption des méthodes non-paramétriques de lissage par noyaux afin de traiter quelques questions en finance quantitative, discipline dans laquelle il est rare que les méthodes paramétriques donnent des résultats satisfaisants. En premier lieu, nous commençons par couvrir les deux volets les plus importants du lissage par noyaux, qui sont l'estimation de densité et la régression, dans les cas des noyaux symétriques et asymétriques. Le but étant d'exposer les propriétés fondamentales et de mettre en relief la différence dans la structure et dans le traitement qu'implique le simple passage d'un cas à l'autre. Par la suite, nous mettons en application le lissage par noyaux, d'abord par la suggestion d'une procédure pour l'estimation de la densité de la volatilité d'actifs financiers et ensuite à travers une formule pour la valorisation de produits financiers.

Mots-clés : Estimation non-paramétrique ; Lissage par noyaux ; Estimation de densité ; Régression ; Polynômes Orthogonaux ; Volatilité stochastique ; modèle de Heston ; Simulation de Monte Carlo ; Valorisation d'option ; Option asiatique.

Abstract

Nonparametric estimation constitutes a fundamental tool in statistics that helps in pushing the boundaries established by the principles on which parametric techniques are based and consequently provides a way to address problems where the parametric approach either fails or gives results that offer substantial room for improvement. The work undertaken in this thesis falls within this context through the adoption of nonparametric kernel smoothing methods in order to deal with some issues in quantitative finance, a field in which parametric methods rarely deliver satisfying results. As a first step, we start by covering the two most important components of kernel smoothing, namely density estimation and regression, in both the symmetric and asymmetric kernel case. The main point being to describe the fundamental properties and to highlight differences in the structure and in the treatment when moving from one case to the other. Afterwards, kernel smoothing is put to the test, firstly by suggesting an estimation procedure for the volatility density of a financial asset and then through a formula for pricing financial derivatives.

Keywords : Nonparametric estimation; Kernel smoothing; Density estimation; Regression; Orthogonal polynomials; Stochastic volatility; Heston model; Monte Carlo simulation; Option valuation; Asian Option.

Table des matières

Introduction générale	10
1 Noyaux continus symétriques	14
1.1 Notions préliminaires	15
1.2 Mesure de l'erreur d'une estimation de densité à noyau	18
1.3 Noyaux d'ordres supérieurs	21
1.3.1 Critères d'erreur d'un estimateur à base d'un noyau d'ordre r	21
1.3.2 Noyau d'ordre r à base de polynômes orthonormés	22
1.4 Sélection du paramètre de lissage	26
1.4.1 Première catégorie	29
1.4.2 Deuxième catégorie	30
1.5 Quelques résultats asymptotiques	33
2 Noyaux continus asymétriques	36
2.1 Éléments de base	37
2.2 Formes fonctionnelles et allures des noyaux gamma et bêta	40
2.2.1 Noyau gamma	40
2.2.2 Noyau bêta	41
2.3 Mesure de l'erreur d'une estimation de densité à noyau asymétriques	43
2.3.1 Estimateur à noyau gamma $K_{x/h+1,h}^G$	44
2.3.2 Estimateur à noyau bêta $K_{x/h+1,(1-x)/h+1}^B$	47
2.4 Sélection du paramètre de lissage	49
2.5 Quelques résultats asymptotiques	50

3	Régression à noyau	53
3.1	Estimateur de Nadaraya-Watson	54
3.2	Estimateur par polynômes locaux	59
3.2.1	Expression de l'estimateur	59
3.2.2	Mesure de l'erreur	61
3.3	Estimateur par noyaux asymétriques	63
3.3.1	Estimateur de Nadaraya-Watson à base du noyau gamma	64
3.3.2	Estimateur de Gasser-Müller à base du noyau beta	65
3.4	Sélection du paramètre de lissage	66
3.4.1	Règle pratique	67
3.4.2	Insertion directe	68
3.4.3	Validation croisée	69
3.4.4	Sélection par plus proche voisin	69
4	Estimation de densité de la volatilité stochastique	70
4.1	La volatilité stochastique	71
4.1.1	Modèle de Heston	71
4.1.2	Schéma de simulation	72
4.2	Estimation de densité	74
4.2.1	Structure de déconvolution	74
4.2.2	Estimateurs	75
4.3	Résultats numériques	76
4.4	Conclusion	80
5	Valorisation d'une option asiatique	82
5.1	Simulations classiques de Monte Carlo pour la valorisation d'options	83
5.1.1	Les simulations de Monte Carlo dans le cadre de Black & Scholes	83
5.1.2	Réduction de la variance	85
5.2	Une approche de Monte Carlo couplée à l'estimation de densité	88
5.3	Résultats numériques	89
5.4	Conclusion	91

Conclusion générale	92
Annexes	94
A1.1	94
A1.2	94
A1.3	95
A1.4	96
A1.5	99
A2.1	100
A3.1	101
A3.2	102
A3.3	104
A4.1	107
A5.1	108
A5.2	109
Bibliographie	112

Table des figures

1.1	Estimation de densité à noyau pour $n = 10$	16
1.2	Quelques noyaux classiques sur le support $[-1, 1]$	17
1.3	Noyaux à base de polynômes de Jacobi (ordres : 1, 3, 5)	26
1.4	Noyaux à base de polynômes de Legendre (ordres : 1, 3, 5)	26
1.5	Noyaux à base de polynômes d'Hermitte (ordres : 1, 3, 5)	26
1.6	Noyaux à base de polynômes de Gegenbauer (ordres : 1, 3, 5)	26
1.7	Différents degrés de lissage pour une estimation de densité à 10 observations	27
1.8	Illustration du compromis biais-variance	28
2.1	Densité gamma pour différentes valeurs des paramètres α et β	41
2.2	Noyau gamma (première variante) pour $h = 0.3$ et x variable	41
2.3	Noyau gamma (deuxième variante) pour $h = 0.3$ et x variable	41
2.4	Noyau gamma (première variante) pour $x = 0.8$ et h variable	42
2.5	Noyau gamma (deuxième variante) pour $x = 0.8$ et h variable	42
2.6	Densité bêta pour différentes valeurs des paramètres α et β	43
2.7	Noyau bêta (première variante) pour $h = 0.3$ et x variable	43
2.8	Noyau bêta (deuxième variante) pour $h = 0.3$ et x variable	43
2.9	Noyau bêta (première variante) pour $x = 0.3$ et h variable	44
2.10	Noyau bêta (deuxième variante) pour $x = 0.3$ et h variable	44
3.1	Lissage par estimateur linéaire local	59
3.2	Lissage par estimateur à polynômes locaux (polynômes de degré 2)	59
4.1	Un échantillon des données simulées	77

4.2	Estimation de densité de la volatilité par déconvolution	78
4.3	Estimation directe de densité de la volatilité par noyau symétrique	79
4.4	Estimation directe de densité de la volatilité par noyau asymétrique	79
4.5	Estimation de densité de la volatilité par moyenne mobile (phase 1) et noyau symétrique (phase 2)	79
4.6	Estimation de densité de la volatilité par moyenne mobile (phase 1) et noyau asymétrique (phase 2)	79
4.7	Estimation de densité de la volatilité par estimateur linéaire local (phase 1) et noyau symétrique (phase 2)	80
4.8	Estimation de densité de la volatilité par estimateur linéaire local (phase 1) et noyau asymétrique (phase 2)	80
4.9	Estimation de densité de la volatilité par estimateur à polynômes locaux de 2nd degré (phase 1) et noyau symétrique (phase 2)	80
4.10	Estimation de densité de la volatilité par estimateur à polynômes locaux de 2nd degré (phase 1) et noyau asymétrique (phase 2)	80
5.1	Valeurs obtenue pour le prix d'un call asiatique arithmétique par MC classique, MC combiné à un noyau gaussien et MC combiné à un noyau gamma avec $r = 0.15$, $\sigma = 0.05$, $K = 95$, $S_0 = 100$	90
5.2	Valeurs obtenue pour le prix d'un call asiatique arithmétique par MC classique, MC combiné à un noyau d'Hermite et MC combiné à un noyau de Laguerre avec $r = 0.15$, $\sigma = 0.05$, $K = 95$, $S_0 = 100$	90
5.3	Valeurs obtenue pour le prix d'un call asiatique arithmétique par MC classique, MC combiné à un noyau gaussien et MC combiné à un noyau gamma avec $r = 0.15$, $\sigma = 0.05$, $K = 105$, $S_0 = 100$	90
5.4	Valeurs obtenue pour le prix d'un call asiatique arithmétique par MC classique, MC combiné à un noyau d'Hermite et MC combiné à un noyau de Laguerre avec $r = 0.15$, $\sigma = 0.05$, $K = 105$, $S_0 = 100$	90
5.5	Valeurs obtenue pour le prix d'un call asiatique arithmétique par MC classique, MC combiné à un noyau gaussien et MC combiné à un noyau gamma avec $r = 0.15$, $\sigma = 0.05$, $K = 100$, $S_0 = 100$	91
5.6	Valeurs obtenue pour le prix d'un call asiatique arithmétique par MC classique, MC combiné à un noyau d'Hermite et MC combiné à un noyau de Laguerre avec $r = 0.15$, $\sigma = 0.05$, $K = 100$, $S_0 = 100$	91

Liste des tableaux

1.1	Quelques exemples de noyaux classiques	17
2.1	Intégrales d'estimateurs à noyau asymétrique pour 3 échantillons aléatoires de taille 5000 issus d'une loi normale	38
2.2	Description de quelques lois asymétriques	39
4.1	Paramètres pour la simulation de volatilité	76
4.2	Description statistique, tests de normalité et de stationnarité.	77
4.3	Valeurs du <i>MISE</i> pour les différentes estimations de densités	81

Introduction générale

Dans un grand nombre de disciplines (ex : Biologie, écologie, astronomie, sciences sociales, ...) on se retrouve face à des plans d'échantillonnage compliqués ainsi qu'à des expériences complexes générant des données qui nécessitent une estimation statistique afin de présenter, analyser ou modéliser certains phénomènes. Ceci se fait généralement à travers une inférence au sujet de la loi produisant ces données. Pour effectuer cette inférence, deux manières différentes se présentent, l'une paramétrique et l'autre non-paramétrique.

L'approche paramétrique, indéniablement considérée comme étant le pilier majeur des statistiques traditionnelles (Kvam [70]), existe depuis les travaux pionniers de de Moivre, Bernoulli et Bayes. Cependant, son aspect moderne, sous lequel on la reconnaît actuellement, fut principalement fondé par Fisher qui a réussi à apporter plusieurs développements en corrigeant et complétant des travaux initiés par Laplace et Gauss. Il est aussi arrivé à résoudre un certain nombre de problèmes qui freinaient le développement et enclavaient l'approche en question dans un état rudimentaire (Hald [54]).

L'idée fondamentale derrière une analyse conduite dans un cadre paramétrique est d'estimer un nombre fini de paramètres. On suppose alors que le modèle, ou l'expérience statistique, produisant les données est une famille connue de lois de probabilité ayant un ou plusieurs paramètres inconnus et devant être estimés. Cette démarche s'avère être utile et très efficace quand on est confiant que l'hypothèse portant sur la distribution des données est plus ou moins vérifiée. Un exemple classique est la théorie des modèles linéaires où l'accent est mis sur la loi normale, un choix justifié par le théorème central limite qui garantit une normalité approximative des données dans le cas d'un échantillon suffisamment large. On retrouve également l'utilisation de la loi de Weibull dans les études de fiabilité, la loi lognormale en finance de marchés et les lois gamma et de Pareto pour l'étude de distribution des richesses.

Néanmoins, il se présente des cas où malgré la grande incertitude vis à vis des hypothèses énoncées des techniques paramétriques sont utilisées, non pas parce qu'elles sont parfaitement appropriées mais seulement parce qu'elles sont commodes. Cette pratique, qui implique forcément des suppositions non cohérentes avec les données disponibles, risque de simplifier à outrance la population ou le processus observé et entraîner une approximation imprécise de sa structure statistique. L'hypothèse de linéarité par exemple, qui est inhérente à la méthodologie d'identification des processus SARIMA, peut engendrer de graves déficiences au niveau des prévisions à long terme pour la simple raison que les modèles expliquant les données d'une manière plus fiable sont souvent plus complexes (Bosq [8]).

Ceci nous amène à la deuxième approche pour l'estimation statistique. L'appellation «non-paramétrique», par laquelle on l'identifie, est attribuée à Wolfowitz en 1942 (Wolfowitz [108]), toutefois, la popularité de cette méthodologie n'a commencé à prendre de l'élan que quelques décennies plus tard, principalement parce que bon nombre de statisticiens commencèrent à pencher vers l'utilisation et le développement de techniques issues de cette approche et donc à repousser ses limites (Hollander, Wolfe et Chicken [59]), chose qui rendit indispensable l'élaboration d'une définition qui puisse traduire l'essence des méthodes non-paramétriques. La plus populaire est la suivante, elle fut énoncée

par Dudewicz et Geller [35] : «les statistiques non-paramétriques peuvent être définies comme étant la science qui fournit des techniques d'inférence statistique pour lesquelles les hypothèses concernant la distribution sous-jacente de la population sont beaucoup plus faibles que les hypothèses énoncées pour des techniques ordinaires traitant un même problème».

Cette définition met en évidence la raison pour laquelle les chercheurs et praticiens ont développé un vif intérêt pour les statistiques non-paramétriques. Celles-ci leur ont permis de se libérer des contraintes de modèles déjà établis et de traiter des données qui peuvent ne satisfaire aucune des hypothèses nécessaires pour les méthodes traditionnelles grâce à des techniques d'inférence applicables à un large éventail de situations, étant donné que l'estimation d'une quantité inconnue se fait exclusivement à partir des observations en émettant des hypothèses beaucoup moins rigides que celles requises pour une inférence paramétrique.

Par ailleurs, si l'on désire établir une comparaison entre les deux approches on devrait différencier deux situations. La première est celle où les données s'alignent avec les hypothèses paramétriques énoncées tandis que la deuxième est la situation où les hypothèses ne tiennent pas. Kvam [70] affirme que pour le premier cas où l'on s'attend normalement à ce que les performances des techniques traditionnelles dépassent de loin les techniques non-paramétriques, la différence réelle n'est pas si choquante que cela, en effet, elle est minimale. Alors que pour le deuxième cas, il serait inconcevable d'utiliser des techniques paramétriques, car seules les non-paramétriques sont valides.

Toujours dans le cadre de la comparaison et lorsque l'on a affaire à des séries chronologiques, on passe impérativement par une phase préliminaire de stationnarisation des données. Il arrive cependant que cette phase puisse perturber les observations à cause des transformations qu'elle leur fait subir. Dans le cadre de l'approche Box & Jenkins, on peut remédier à cela par l'utilisation de transformations qui soient relativement simples et qui permettent de neutraliser les effets de la tendance et de la saisonnalité (Bosq [8]). Par contre, une approche non-paramétrique, offre carrément la possibilité d'éviter toute transformation et d'utiliser directement les données originales pour des fins de modélisation et de prévisions (Poggi [83]).

Les problématiques abordées par les statistiques non-paramétriques ont conduit à ce que celles-ci soient subdivisées de manière informelle en deux segments. L'un classique qui offre des outils d'inférence pour traiter des variables catégorielles, ordinales, etc...Et l'autre qualifié de moderne qui s'intéresse entre autres aux splines, aux ondelettes, aux méthodes de lissage et à l'apprentissage machine.

Dans ce travail, l'intérêt se porte sur les techniques de lissage par noyaux qui sont un volet du second segment et qui offrent un ensemble de procédés pour l'estimation de fonctions applicables à plusieurs problèmes dont les plus fondamentaux sont l'estimation des fonctions de densité et l'estimation des fonctions de régression. En général, la seule hypothèse que l'on adopte est celle que la fonction appartienne à un certain ensemble satisfaisant quelques conditions de régularité (Wasserman [105]). Un exemple simple serait l'ensemble des fonctions continues sur un certain intervalle de \mathbb{R} .

Les points communs entre les méthodes de lissage par noyaux sont, le noyau et le paramètre de lissage appelé aussi fenêtre. Les noyaux, qui interviennent en tant que fonctions de pondération lors de l'estimation, furent introduits par Rosenblatt pour l'étude de la densité, par Parzen pour l'étude du mode de la densité et indépendamment par Nadaraya et Watson pour effectuer une régression non-paramétrique. La construction d'un noyau ou sa sélection peut se faire sur la base de la vitesse de convergence qu'il permet d'atteindre car il est prouvé que plus un noyau a de moments nuls, plus il offre une meilleure vitesse de convergence. Quant au paramètre de lissage, son choix est prépondérant pour le processus d'estimation puisqu'on ne peut atteindre des résultats satisfaisant qu'avec un choix convenable de ce paramètre. Cependant, ce choix pose une réelle difficulté, car en pratique, la quantité de lissage requise est principalement dictée par les besoins de l'utilisateur. Les méthodes automatiques de sélection sont alors utilisées seulement pour fournir une valeur qui

sera considérée comme un point de départ afin de déterminer le lissage nécessaire et adéquat au cas traité.

La richesse de la littérature disponible offre une vision encyclopédique des propriétés et méthodes de construction pour les noyaux classiques ainsi que pour les techniques usuelles destinées à la sélection du paramètre de lissage. On retrouve notamment le classique de Wand et Jones [104] qui représente un recueil des outils mathématiques nécessaires pour aborder les problèmes de lissage en général et les problèmes de lissage par noyaux en particuliers, les ouvrages de Silverman [96] et Takezawa [99] qui couvrent respectivement l'estimation de densité par noyaux et la régression non-paramétrique en en donnant un traitement succinct et accompli, le livre de Simonoff [97] qui couvre de manière exhaustive mais pratique le lissage par noyaux et le livre de Kvam [70] qui relate un certain nombres d'applications en sciences et en ingénierie.

Pour ce qui est des applications pratiques, les méthodes de lissage ont participé efficacement au perfectionnement et à la mise en place de solutions pour des problèmes soulevés par l'étude de plusieurs phénomènes réels. En agriculture par exemple, les États Unis ont mis en place un système de stockage à très long terme pour préserver des graines ayant certaines particularités génétiques. La viabilité ou capacité à germer de ces graines doit être vérifiée périodiquement pour régénérer le stock mort. Seulement, si le test de viabilité est programmé trop tôt, il s'avère qu'une grande partie du stock est toujours viable ce qui constitue un gaspillage de ressources. En revanche si le test est programmé trop tard, on risque de perdre la totalité du stock. Les techniques de lissage ont alors aidé à mettre en place un calendrier de tests efficace (Trap II et al. [101]).

En épidémiologie, une façon d'évaluer le risque de propagation d'une maladie à travers la consommation d'aliments est la réalisation de simulations qui vont modéliser la diffusion des risques aléatoires et épidémiques. Le coût de ces simulations qui est souvent extrêmement élevé en matière de temps et de ressources informatiques fut réduit de manière considérable grâce à l'utilisation des techniques de lissage. Une illustration en est faite par une étude qui évalue la probabilité qu'un enfant ayant mangé des boulettes de viande contenant la bactérie E. coli puisse contracter le syndrome hémolytique et urémique (Trap II [100]; Pouillot et Delignette-Muller [84]).

En apprentissage machine, quand les problèmes en rapport avec la classification et le regroupement sont approchés d'une manière Bayésienne, la connaissance de certaines densités de probabilité constitue souvent, une complication à laquelle on se heurte et s'impose comme un obstacle épineux à surmonter. Dans de multiples cas réels, qu'ils soient univariés ou multivariés, cette complication peut être traitées efficacement en faisant appel à un estimateur de densité par noyau (Di Marzio et Taylor [32]). Notons cependant, que le problème de classification en soi n'est pas un problème d'estimation de densité mais celle-ci ne fait que participer au processus de résolution du problème.

En finance de marché, les outils offerts par les techniques de lissage par noyaux ont favorisé l'élaboration de nouvelles approches qui ont pu apporter des solutions bien plus précises que celles déjà existantes pour bon nombre de problèmes. On peut notamment citer la modélisation du comportement des actifs financiers (actions, obligations, options, swaps, ...) et de leur volatilité (Linton et al. [71]; Ke et al. [67]), la construction de stratégies pour la gestion de portefeuille en temps réel (Wolberg [107]) et la quantification des risques pour des fins d'investissement et d'allocation de capital (Gustafsson et al. [52]) où le lissage par noyaux a permis de saisir les variations extrêmes en rapport avec les sources de risques qui échappent d'habitude aux modèles classiques et ceci avec autant de précision et moins de restrictions théoriques que les modèles développés dans le cadre de la théorie des valeurs extrêmes.

Il est évident que les applications des techniques de lissage par noyaux ne s'arrêtent pas aux exemples cités ci-dessus puisqu'on retrouve également des applications en physique subatomique (Narsky et Porter [77]), en écologie (Carey et al. [22]), en génétique humaine (Deng et Shen [30]), en marketing (Donthu [33]), en astronomie (Wasserman [105]), en médecine sportive (Boecker et al. [5]), en hydrologie (Sharma, Tarboton et Lall [92]), en industrie (Kulczycki [69]) et en intelligence

artificielle (Hastie, Tibshirani et Friedman [57]).

Le grand nombre de disciplines ayant adopté les modèles de lissage par noyaux confirme l'efficacité de ces méthodes à résoudre des problèmes face auxquels les autres approches soit échouent, soient sont rejetées à cause des complications qu'elles posent pour les mettre en œuvre. Cette diversité des champs d'application ouvre une multitude de pistes de recherche dont plusieurs, notamment les plus encourageantes, sont en cours de développement afin d'améliorer la capacité des modèles en question à surmonter les difficultés qui peuvent surgir lors du traitement de certains types de données.

Ce travail a été élaboré dans le but particulier d'adresser certains problèmes en finance de marché à travers le lissage par noyaux. Mais avant cela, il est important que l'on s'intéresse d'abord aux éléments de base de cette approche qui, on le rappelle sont, le paramètre de lissage et les noyaux. Ces derniers, jusque-là négligés en faveur du paramètre de lissage, seront discutés abondamment et leur importance sera justifiée aussi bien théoriquement qu'empiriquement à travers l'évaluation de leur effet sur le comportement des différents estimateurs. Pour le paramètre de lissage, différentes méthodes de sélections, classifiées selon leurs complexités, seront présentées pour chaque contexte de lissage. Il semble toutefois que, la valeur de ce paramètre ne peut être quantifiée objectivement, puisqu'elle dépend principalement du besoin de l'utilisateur, c'est pourquoi, une procédure de sélection, aussi complexe soit-elle risque d'avoir une utilité limitée dans certains cas.

Ainsi, la présente thèse englobe 5 chapitres dont les trois premiers exposent, pour l'estimation de densité et la régression, la théorie du lissage par noyau dans les cadres de noyaux symétriques et asymétriques. On y couvre notamment les différentes définitions, caractéristiques et propriétés fondamentales, quelques techniques pour la sélection du paramètre de lissage ainsi que quelques résultats de convergence. En outre, on profite de l'aspect théorique de cette première partie pour proposer un ensemble de nouveaux noyaux basés dans leur construction sur des polynômes orthogonaux et ayant fait l'objet d'une publication dans *Electronic Journal of Applied Statistical Analysis* Vol.2(2) (2016) «Kernel density smoothing using probability density functions and orthogonal polynomials».

Le 4ème chapitre fait l'objet d'un article accepté pour publication dans *International Journal of Computational Economics and Econometrics* (2020) «A nonparametric estimator for stochastic volatility density», dans lequel on essaie d'améliorer la qualité de l'estimation de densité de la volatilité stochastique en utilisant une combinaison entre la régression et l'estimation de densité à noyau. Les données utilisées sont simulées à travers un modèle d'élasticité constante de variance – volatilité stochastique (CEV-SV), calibré pour reproduire les variations de l'indice *S&P500*, sous la forme d'un processus bidimensionnel dans lequel une seule coordonnées est observable. Une régression par noyau est alors utilisée pour filtrer les données désirées à partir d'une structure de convolution ensuite l'estimation de densité est conduite. Plusieurs estimations impliquant différents noyaux sont effectuées et leur efficacité est évaluée selon une mesure de risque globale.

Le 5ème chapitre fait également l'objet d'une publication dans *International Journal of Scientific and Technology Research* Vol.9(3) (2020) «A KDE-based approach to option pricing», qui s'intéresse cette fois-ci à la valorisation d'un type de contrats échangés sur les marchés financiers appelés options. Le but est alors d'augmenter la précision de l'approche Monte Carlo dans le contexte habituel, dit de Black & Scholes, par l'utilisation d'une nouvelle formule qui intègre des pondérations dont la finalité est d'ajuster l'évaluation numérique de l'espérance donnée par le théorème de Feynman-Kac. Cette espérance, constitue l'élément de base pour évaluer numériquement le prix d'une option. Les poids en question nécessitent pour leur calcul, l'utilisation d'une estimation de densité empirique. Par conséquent, un estimateur de densité couplé à plusieurs noyaux, dont ceux basés sur les polynômes orthogonaux proposés au premier chapitre, est utilisé pour appliquer cette méthode à la valorisation d'un type spécifique d'option, appelé call asiatique. Les prix calculés sont alors comparés à des résultats, issus de la littérature, obtenus par la résolution numérique d'une équation aux dérivées partielles.

Chapitre 1

Noyaux continus symétriques

Dans ce chapitre l'intérêt est porté sur les noyaux continus symétriques qui seront présentés dans le cadre de l'estimation de densité univariée d'un échantillon aléatoire. Ce problème d'estimation est probablement l'une des questions les plus fondamentales en statistiques. En effet, l'estimation de densité permet entre autres d'établir un récapitulatif graphique soulignant la manière dont la variable aléatoire est distribuée à travers son support et rendant possible la conduite d'un examen informel d'un certain nombre de propriétés importantes dont l'asymétrie, l'aplatissement, la multimodalité et la dispersion des données.

Comme il a été souligné dans l'introduction générale, un problème d'estimation statistique peut être abordé soit de manière paramétrique ou non-paramétrique. Pour une estimation de densité paramétrique il est supposé que la fonction à estimer appartient à une famille paramétrique connue dont seuls les paramètres sont inconnus et devront être estimés tandis que pour une estimation non-paramétrique il n'y a aucune hypothèse en rapport avec la forme de la fonction de densité.

Ainsi, le moyen le plus ancien et le plus répandu qui permet d'effectuer une estimation de densité non-paramétrique est l'histogramme qui réalise en quelque sorte un lissage rudimentaire en se basant uniquement sur les données disponibles. En fait, l'histogramme est principalement une fonction en escalier qui commence en une origine x_0 et dont les intervalles de longueur b , appelés classes, sont pondérés selon leurs fréquences relatives dans les données, c'est-à-dire que l'estimation par histogramme en un point x peut être donnée par

$$\hat{f}_H(x) = \frac{\text{nombre d'observations dans l'intervalle contenant } x}{b \times \text{nombre total d'observations}}$$

En outre, il est envisageable de généraliser l'histogramme en permettant aux intervalles d'avoir des longueurs différentes, toutefois cette manœuvre doit être solidement justifiée vu les complications qu'elle produit lors de l'interprétation du graphique en question.

La simplicité et la facilité de mise en place de l'histogramme le rendent un outil pour le moins attrayant et lui garantissent un maintien de sa popularité. Cet outil, souffre cependant de quelques inconvénients qui justifient le besoin d'adopter des méthodes plus sophistiquées. Parmi ces inconvénients, il y a la tendance de l'histogramme à gaspiller l'information à travers la répartition de la masse de probabilité d'un échantillon de manière uniforme tout au long de la classe dans laquelle il est observé malgré qu'en général les observations ne sont pas censées avoir de répartition uniforme dans un intervalle en particulier. L'histogramme est également sujet à une grande sensibilité vis-à-vis de plusieurs choix subjectifs notamment les choix de l'origine et de la longueur des classes. D'autant plus que dans certains cas où l'estimation de densité n'est qu'une phase intermédiaire dans

l'application d'autres méthodes, les dérivées de l'estimation sont également requises, l'histogramme est alors considéré inefficace à cause de sa nature simple et primitive.

Un autre moyen pour estimer une densité de manière non-paramétrique est l'utilisation d'un estimateur à noyau. Celui-ci permet de surmonter un nombre conséquent de difficultés qui résultent de l'utilisation de l'histogramme. En effet, la façon dont l'estimateur à noyau répartit équitablement la masse de probabilité de chaque observation autour d'elle-même offre à l'estimation construite une forte flexibilité et lui donne un aspect beaucoup plus lisse. De plus, cela assure une meilleure extraction de l'information à partir des données disponibles et garantit par conséquent une meilleure convergence vers la densité réelle.

L'objet de ce chapitre est d'abord d'introduire de manière intuitive l'idée de l'estimateur de densité à noyau. Par la suite nous allons présenter des formes asymptotiques pour des mesures de l'erreur globale et locale. Nous nous intéresserons ensuite aux noyaux d'ordre supérieurs pour lesquels nous allons également spécifier les mesures de l'erreur avant de proposer un éventail de noyaux de ce type basés sur les polynômes orthogonaux. Par la suite, nous mettons l'accent sur le paramètre de lissage en présentant plusieurs méthodes de sélections que nous allons classer en deux catégories. On finit le chapitre par quelques résultats de convergence et une discussion par rapport à l'ordre de convergence.

1.1 Notions préliminaires

Soient X_1, X_2, \dots, X_n un échantillon de variables aléatoires indépendantes et identiquement distribuées (*i.i.d*) ayant une densité de probabilité f par rapport à la mesure de Lebesgue sur \mathbb{R} . La fonction de répartition de la variable aléatoire X est alors

$$F_X(x) = \int_{-\infty}^x f(t) dt \quad (1.1)$$

Il s'ensuit donc que la fonction de répartition empirique est de la forme

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x) \quad (1.2)$$

La convergence de $\hat{F}_X(x)$ vers $F_X(x)$ est assurée pour tout x par la loi forte des grands nombres.

Ainsi pour un h suffisamment petit la densité en un point x peut être approchée par

$$f_X(x) \approx \frac{F_X(x+h) - F_X(x-h)}{2h} \quad (1.3)$$

et donc estimée par

$$\hat{f}_X(x) = \frac{\hat{F}_X(x+h) - \hat{F}_X(x-h)}{2h} \quad (1.4)$$

Cet estimateur peut également prendre la forme suivante :

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K_{\mathbb{1}}\left(\frac{x - X_i}{h}\right) \quad (1.5)$$

où

$$K_{\mathbb{1}}(u) = \frac{1}{2} \mathbb{1}(-1 \leq u \leq 1) \quad (1.6)$$

L'estimateur donné par l'équation (1.5) représente le plus simple des estimateurs à noyaux, appelé estimateur de Rosenblatt. Les définitions ci-dessous donnent une généralisation de cet estimateur.

Définition 1.1 On appelle noyau et on note K toute fonction intégrable satisfaisant $\int K(u) du = 1$.

Définition 1.2 Soient K un noyau et h un nombre strictement positif appelé fenêtre ou paramètre de lissage. On définit alors un estimateur de densité à noyau en un point x par

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (1.7)$$

Il convient de noter que cette estimation de densité en un point x est essentiellement une moyenne des valeurs calculées par le noyau en ce même point. Le noyau contrôle donc, à travers la forme qu'il prend, la manière dont la masse de probabilité est assignée. Une illustration de ce concept est donnée dans la figure (1.1) avec un nombre d'observations égal à 10 et un noyau choisit comme étant la fonction de densité de la loi normale centrée réduite.

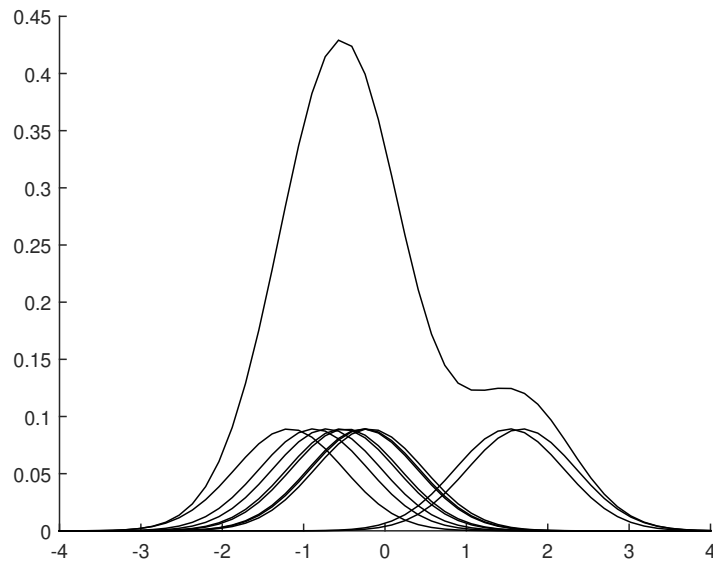


FIGURE 1.1 – Estimation de densité à noyau pour $n = 10$

Remarque 1.1 Selon le besoin et afin qu'il soit mieux adapté au problème traité, un noyau peut se voir attribuer des propriétés complémentaires dont les plus communes sont :

1. la positivité : $K(u) \geq 0 \forall u$
2. symétrie par rapport à 0 : $k(u) = k(-u)$ pour $u > 0$
3. moyenne nulle : $\int uK(u) du = 0$
4. variance finie : $\int u^2K(u) du < \infty$
5. de carré intégrable : $\int k^2(u) du < \infty$

Pour beaucoup d'auteurs il est de coutume d'utiliser comme noyau la fonction de densité d'une loi de probabilité usuelle, la loi normale par exemple, mais il existe également un certain nombre de noyaux classiques qui apparaissent régulièrement dans la littérature, la table (1.1) et la figure (1.2) en donnent un récapitulatif succinct.

Noyau	Fonction noyau	Support
Silverman	$(1/2) \exp(- u /\sqrt{2}) \sin(u /\sqrt{2}) + (\pi/4)$	\mathbb{R}
Epanechnikov	$(3/4)(1-u^2)$	$[-1, 1]$
Biweight	$(15/16)(1-u^2)^2$	$[-1, 1]$
Triweight	$(35/32)(1-u^2)^3$	$[-1, 1]$
Triangulaire	$1- u $	$[-1, 1]$

TABLE 1.1 – Quelques exemples de noyaux classiques

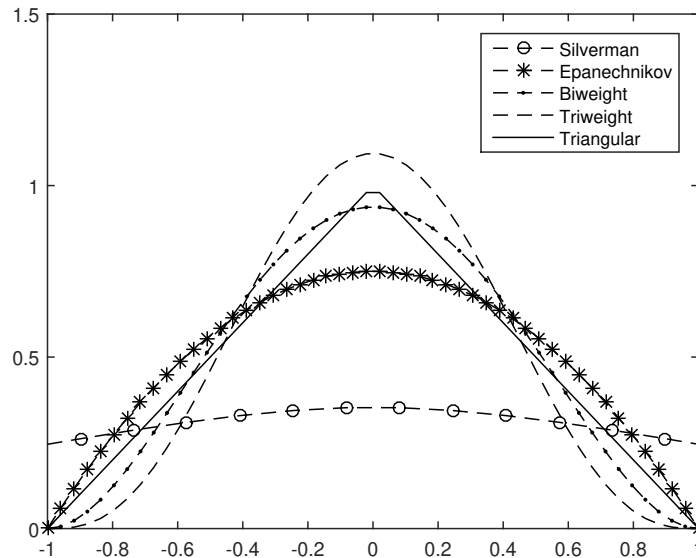


FIGURE 1.2 – Quelques noyaux classiques sur le support $[-1, 1]$

1.2 Mesure de l'erreur d'une estimation de densité à noyau

Que l'estimation soit ponctuelle ou sur tout le support de la densité, il est important de pouvoir évaluer la performance de l'estimateur à travers une mesure de l'écart entre la densité estimée et la densité réelle. Une mesure commune aux méthodes paramétriques et non-paramétriques est l'erreur moyenne quadratique MSE , ce critère d'évaluation se distingue par rapport aux autres mesures traitées dans la littérature par sa flexibilité vis-à-vis des manipulations mathématiques. Ainsi pour un point x quelconque appartenant au support de la densité on a

$$MSE_x [\hat{f}_n] = E [\hat{f}_n(x) - f(x)]^2 \quad (1.8)$$

Une propriété intéressante de l'erreur moyenne quadratique est qu'elle peut faire l'objet d'une décomposition comportant le biais et la variance, une telle forme permet alors une meilleure étude de la qualité de l'estimateur. Pour retrouver cette décomposition il suffit de développer la formule donnée par l'équation (1.8) comme suit :

$$\begin{aligned} MSE_x [\hat{f}_n] &= E [\hat{f}_n^2(x)] - 2f(x) E [\hat{f}_n(x)] + f^2(x) \\ &= E [\hat{f}_n^2(x)] - E^2 [\hat{f}_n(x)] + \left\{ E [\hat{f}_n(x)] - f(x) \right\}^2 \\ &= Var [\hat{f}_n(x)] + Biais^2 [\hat{f}_n(x)] \end{aligned} \quad (1.9)$$

puisque

$$Var [\hat{f}_n(x)] = [\hat{f}_n^2(x)] - E^2 [\hat{f}_n(x)] \text{ et } Biais^2 [\hat{f}_n(x)] = \left\{ E [\hat{f}_n(x)] - f(x) \right\}^2$$

Calculer l'erreur ponctuelle MSE de l'estimateur revient alors à établir des expressions en fonction du noyau pour le biais et la variance. En commençant par l'espérance on retrouve

$$\begin{aligned} E [\hat{f}_n(x)] &= E \left[\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) \right] \\ &= E \left[\frac{1}{h} K \left(\frac{x - t}{h} \right) \right] \\ &= \frac{1}{h} \int K \left(\frac{x - t}{h} \right) f(t) dt \end{aligned}$$

En effectuant le changement de variable $t = x - uh$ l'espérance devient

$$E [\hat{f}_n(x)] = \int K(u) f(x - uh) du$$

Le biais s'exprime donc ainsi :

$$Biais \left[\hat{f}_n(x) \right] = \int K(u) f(x - uh) du - f(x) \quad (1.10)$$

Cette expression expose clairement le lien étroit qui existe entre le noyau, le paramètre de lissage et la valeur que peut prendre le biais. Il s'ensuit donc que la réduction efficace de cette dernière ne pourra pas se faire à travers une simple augmentation de la taille de l'échantillon mais plutôt à travers un choix correct du noyau et du paramètre de lissage.

Concernant la variance, elle prend la forme suivante :

$$\begin{aligned} Var \left[\hat{f}_n(x) \right] &= Var \left[\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) \right] \\ &= \frac{1}{nh^2} Var \left[K \left(\frac{x - t}{h} \right) \right] \\ &= \frac{1}{nh^2} \left\{ E \left[K^2 \left(\frac{x - t}{h} \right) \right] - E^2 \left[K \left(\frac{x - t}{h} \right) \right] \right\} \\ &= \frac{1}{nh^2} \left\{ \int K^2 \left(\frac{x - t}{h} \right) f(t) dt - \left[\int K \left(\frac{x - t}{h} \right) f(t) dt \right]^2 \right\} \end{aligned}$$

et en posant encore une fois $t = x - uh$ on a

$$Var \left[\hat{f}_n(x) \right] = \frac{1}{nh} \int K^2(u) f(x - uh) du - \frac{1}{n} \left[\int K(u) f(x - uh) du \right]^2 \quad (1.11)$$

Il est donc possible d'associer les équations (1.10) et (1.11) pour avoir une formule plus explicite de l'erreur moyenne quadratique. L'utilité de ce critère est toutefois limitée, vu qu'il n'arrive à mesure l'erreur qu'en un point. Une mesure plus commode et qui s'adapterait plus à l'intérêt de l'estimation de densité serait une mesure globale qui permette d'évaluer l'erreur sur tout le support de la densité. Un critère convenable qui découle de l'erreur quadratique intégrée *ISE*

$$ISE \left[\hat{f}_n \right] = \int \left[\hat{f}_n(x) - f(x) \right]^2 dx$$

et qui s'apparente à l'erreur moyenne quadratique, est l'erreur moyenne quadratique intégrée *MISE*

$$MISE \left[\hat{f}_n \right] = \int MSE_x \left[\hat{f}_n \right] dx \quad (1.12)$$

Notons que l'utilisation de l'erreur quadratique intégrée aurait été suffisante si nous nous intéressions qu'à un unique échantillon issu de la densité f . Seulement, nous nous intéressons à un quelconque échantillon issu de cette densité et ceci rend l'usage de l'erreur moyenne quadratique intégrée plus adéquat.

Maintenant, il est clair qu'en reportant directement les formules déjà établies de la variance et du biais dans l'expression suivante :

$$MISE \left[\hat{f}_n \right] = \int Var \left[\hat{f}_n(x) \right] dx + \int Biais^2 \left[\hat{f}_n(x) \right] dx \quad (1.13)$$

La formule de l'erreur *MISE* deviendra trop compliquée pour qu'elle soit d'une quelconque utilité pratique. Il serait alors désirable d'essayer de cerner l'expression de l'erreur globale afin d'en tirer quelques conclusions concernant le comportement de la méthode sous certaines conditions. Un moyen usuel qui permet de simplifier de telles expressions est la dérivation d'approximation à l'aide de la formule de Taylor.

Avant d'effectuer les calculs, on supposera que le noyau est une fonction symétrique par rapport à zéro satisfaisant les trois conditions suivantes :

$$\int K(u) du = 1 ; \int uK(u) du = 0 ; \int u^2 K(u) du < \infty$$

On utilisera donc

$$f(x - uh) \approx f(x) - uhf'(x) + \frac{u^2 h^2}{2} f''(x)$$

pour avoir l'approximation suivante de l'espérance :

$$E[\hat{f}_n(x)] \approx f(x) + \frac{1}{2} h^2 f''(x) \int u^2 K(u) du$$

et ainsi retrouver des approximations asymptotiques pour le biais et la variance. En utilisant les notations $\mu_2(g) = \int u^2 g(u) du$ et $R(g) = \int g^2(u) du$ cela donne

$$\text{Biais}[\hat{f}_n(x)] \approx \frac{1}{2} h^2 f''(x) \mu_2(K) \tag{1.14}$$

et

$$\begin{aligned} \text{Var}[\hat{f}_n(x)] &\approx \frac{1}{nh} \int K^2(u) \left[f(x) - hu f'(x) + \frac{h^2 u^2}{2} f''(x) \right] du - \frac{1}{n} f(x) \\ \text{Var}[\hat{f}_n(x)] &\approx \frac{1}{nh} f(x) R(K) \end{aligned} \tag{1.15}$$

À leur tour, les équations (1.14) et (1.15) permettront de donner des formes asymptotiques relativement simples et maniables à l'erreur moyenne quadratique et à l'erreur moyenne quadratique intégrée :

$$\text{AMSE}_x[\hat{f}_n] = \frac{1}{nh} f(x) R(K) + \frac{1}{4} h^4 [f''(x)]^2 \mu_2^2(K) \tag{1.16}$$

$$\text{AMISE}[\hat{f}_n] = \frac{1}{nh} R(K) + \frac{1}{4} h^4 \mu_2^2(K) R(f'') \tag{1.17}$$

1.3 Noyaux d'ordres supérieurs

Un noyau d'ordre supérieur est principalement un moyen qui va permettre d'améliorer quelques aspects de l'estimation en évitant d'adhérer à une pratique courante qui préconise un peu de simplicité et qui veuille que le noyau utilisé soit une densité de probabilité, chose qui implique que tous les noyaux admissibles aient un second moment μ_2 non nul et impose un ordre de $o(h^2)$ pour l'approximation asymptotique du biais.

En fait, permettre au noyau d'avoir n moments nuls entraîne une réduction du biais sans pour autant causer une augmentation de la variance et donc garantit une meilleure convergence en contournant efficacement le dilemme biais-variance.

Définition 1.3 Soit r un entier strictement positif. On dit qu'un noyau K est d'ordre r si les conditions suivantes sont vérifiées :

$$\int K(x) dx = 1 ; \int x^j K(x) dx = 0 \text{ pour } j = 1, \dots, r-1 ; \int x^r K(x) dx \neq 0$$

Il existe également une définition alternative moins répandue mais qui est tout de même utilisée par quelques auteurs.

Définition 1.4 Soit r un entier strictement positif. On dit qu'un noyau K est d'ordre r si les conditions suivantes sont vérifiées :

$$\int K(x) dx = 1 ; \int x^j K(x) dx = 0 \text{ pour } j = 1, \dots, r$$

Remarque 1.2 La différence entre ces deux définitions engendre quelques changements minimes au niveau du calcul du biais mais cela n'affecte en rien les propriétés des noyaux.

Notons que du moment que nous sommes dans le cadre symétrique, les moments d'ordres impairs seront toujours nuls (annexe A1.1). Il est donc clair que tous les noyaux d'ordres impairs peuvent être assimilés à des noyaux d'ordres pairs.

1.3.1 Critères d'erreur d'un estimateur à base d'un noyau d'ordre r

Nous traitons dans cette partie les noyaux d'ordre r définis selon la définition (1.3). Le calcul du biais qui correspond aux noyaux de la définition (1.4) est donné dans l'annexe (A1.2).

Lors de l'approximation asymptotique du critère de l'erreur d'une estimation effectuée à base d'un noyau d'ordre r , les moments nuls dictent quelques modifications qui engendrent de nouvelles formules pour les critères *AMSE* et *AMISE*. Néanmoins, La méthode pour retrouver ces formules reste la même.

En réutilisant le développement de Taylor mais cette fois-ci à l'ordre r , on obtient

$$f(x - uh) \approx f(x) - uhf'(x) + \frac{u^2 h^2}{2} f''(x) + \dots + (-1)^r \frac{u^r h^r}{r!} f^{(r)}(x)$$

En posant $\mu_r(K) = \int u^r K(u) du$ un calcul du biais à partir de ce développement donne

$$\text{Biais} \left[\hat{f}_n(x) \right] \approx (-1)^r \frac{h^r}{r!} f^{(r)}(x) \mu_r(K) \quad (1.18)$$

La variance quant à elle ne changera pas et gardera la forme donnée par l'équation (1.15). Ainsi on retrouve

$$\text{AMSE}_x \left[\hat{f}_n \right] = \frac{1}{nh} f(x) R(K) + \left(\frac{h^r}{r!} f^{(r)}(x) \mu_r(K) \right)^2 \quad (1.19)$$

$$\text{AMISE} \left[\hat{f}_n \right] = \frac{1}{nh} R(K) + \left(\frac{h^r}{r!} \right)^2 R \left(f^{(r)} \right) \mu_r^2(K) \quad (1.20)$$

1.3.2 Noyau d'ordre r à base de polynômes orthonormés

Classiquement, un noyau d'ordre $r > 2$ selon la définition (1.3) est construit selon des règles bien établies à partir de noyaux d'ordre inférieur à r . Un noyau d'ordre 4 par exemple peut être obtenu par

$$K_4(u) = \frac{3}{2} K_2(u) + \frac{1}{2} u K_2'(u) \quad (1.21)$$

où le noyau K_2 est souvent une densité de probabilité symétrique.

Toutefois, quand ils sont comparés à des noyaux d'ordre 2, l'efficacité pratique des noyaux d'ordre supérieurs construits selon l'équation (1.21) et d'autres règles similaires reste incertaine et peine à justifier la complexité et le coût de calcul qui en résultent.

Une autre approche intéressante qui aide à la construction de noyaux d'ordre r , cette fois-ci selon la définition (1.4), est l'utilisation de polynômes orthogonaux. En effet, la normalisation de ces polynômes leur offre une structure favorable à la construction d'un noyau d'une forme simple et d'un d'ordre quelconque. Mais avant de proposer quelques noyaux à base de polynômes orthonormés, donnons-en d'abord une définition (quelques autres généralités sont données dans l'annexe (A1.3)).

Définition 1.5 Soit I un intervalle sur \mathbb{R} et ω une fonction strictement positive et intégrable sur cet intervalle. On appelle polynômes orthonormés sur I par rapport à la fonction poids ω , une suite de polynômes $\Pi_0, \Pi_1, \dots, \Pi_n$ (où Π_m est de degré m) tels que

$$\langle \Pi_i | \Pi_j \rangle = \int_I \Pi_i(u) \Pi_j(u) \omega(u) du = \delta_{ij}$$

$\delta_{i,j}$ étant le symbole de Kronecker.

Les noyaux à base de polynômes orthonormés sont alors construits comme suit :

Soit $\Pi_i(\xi)_{i=0}^{\infty}$ une base orthonormée. On a donc pour tout entier naturel $m \in \{0, 1, \dots, r\}$

$$u^m = \sum_{d=0}^m a_{dm} \Pi_d(u) ; a_{dm} \in \mathbb{R}$$

et soit K_{Π} une fonction de la forme

$$K_{\Pi}(u) = \sum_{i=0}^r \Pi_i(0) \Pi_i(u) \omega(u) \quad (1.22)$$

on trouve alors que

$$\begin{aligned} \mu_m(K_{\Pi}) &= \int u^m K_{\Pi}(u) du \\ &= \int \sum_{d=0}^m \sum_{i=0}^r a_{dm} \Pi_d(u) \Pi_i(0) \Pi_i(u) \omega(u) \\ &= \begin{cases} 1 & \text{si } m = 0 \\ 0 & \text{si } m = 1, 2, \dots, r \end{cases} \end{aligned}$$

d'où K_{Π} est un noyau d'ordre r selon la définition (1.4).

En combinant différents polynômes orthonormés et leurs poids avec l'équation (1.22), on peut facilement construire un large choix de noyaux de différents ordres offrant plusieurs niveaux de flexibilité, en voici quelques exemples :

Notons d'abord que dans tout ce qui suit, on considère J^N, L^N, G^N, \dots comme les formes normalisées des polynômes J, L, G, \dots

Polynômes de Jacobi (limites d'orthogonalité $[-1, 1]$)

Formule de Rodrigues :

$$J(k, \alpha, \beta, u) = \frac{(1-u)^{-\alpha} (1+u)^{-\beta}}{(-1)^k 2^k k!} \frac{d^k}{du^k} \left[(1-u)^{\alpha} (1+u)^{\beta} (1-u^2)^k \right]$$

Poids :

$$\omega(\alpha, \beta, u) = (1-u)^{\alpha} (1+u)^{\beta}$$

Norme :

$$\eta(k, \alpha, \beta) = \sqrt{\frac{2^{\alpha+\beta+1} \Gamma(k+\alpha+1) \Gamma(k+\beta+1)}{k! (2k+\alpha+\beta+1) \Gamma(k+\alpha+\beta+1)}} ; \alpha, \beta > -1$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r J^N(k, \alpha, \beta, 0) J^N(k, \alpha, \beta, u) \omega(\alpha, \beta, u) \quad (1.23)$$

Polynômes de Legendre (limites d'orthogonalité $[-1, 1]$)

Formule de Rodrigues :

$$L(k, u) = \frac{1}{(-1)^k 2^k k!} \frac{d^k}{du^k} \left[(1-u^2)^k \right]$$

Poids :

$$\omega = 1$$

Norme :

$$\eta(k) = \sqrt{\frac{2}{2k+1}}$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r L^N(k, 0) L^N(k, u) \quad (1.24)$$

Polynômes de Gegenbauer (limites d'orthogonalité $[-1, 1]$)

Les polynômes de Gegenbauer sont donnés en termes de polynômes de Jacobi avec la spécification $\alpha = \beta = \lambda - \frac{1}{2}$.

Formule de Rodrigues :

$$G(k, \lambda, u) = \frac{\Gamma(2\lambda + k) \Gamma(\lambda + \frac{1}{2})}{\Gamma(2\lambda) \Gamma(\lambda + k + \frac{1}{2})} J\left(k, \lambda - \frac{1}{2}, \lambda - \frac{1}{2}, u\right) ; \lambda > \left(-\frac{1}{2}\right)$$

Poids :

$$\omega(\lambda, u) = (1 - u^2)^{\lambda - \frac{1}{2}}$$

Norme :

$$\eta(k, \lambda) = \sqrt{\frac{\pi 2^{1-2\lambda} \Gamma(2\lambda + k)}{(\lambda + k) k! \Gamma(\lambda)^2}} ; \lambda \neq 0$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r G^N(k, \lambda, 0) G^N(k, \lambda, u) \omega(\lambda, u) \quad (1.25)$$

Polynômes de Chebyshev de première espèce (limites d'orthogonalité $[-1, 1]$)

Les polynômes de Chebyshev de première espèce sont donnés en termes de polynômes de Jacobi avec la spécification $\alpha = \beta = -\frac{1}{2}$.

Formule de Rodrigues :

$$T(k, u) = \frac{2^{2k} (k!)^2}{(2k)!} J\left(k, -\frac{1}{2}, -\frac{1}{2}, u\right)$$

Poids :

$$\omega(u) = (1 - u^2)^{-1/2}$$

Norme :

$$\eta = \begin{cases} \sqrt{\frac{\pi}{2}} & \text{si } k \neq 0 \\ \sqrt{\pi} & \text{si } k = 0 \end{cases}$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r T^N(k, 0) T^N(k, u) \omega(u) \quad (1.26)$$

Polynômes de Chebyshev de deuxième espèce (limites d'orthogonalité $[-1, 1]$)

Les polynômes de Chebyshev de deuxième espèce sont donnés en termes de polynômes de Jacobi avec la spécification $\alpha = \beta = \frac{1}{2}$.

Formule de Rodrigues :

$$U(k, u) = \frac{2^{2k} (k!) (k+1)!}{(2k+1)!} J\left(k, \frac{1}{2}, \frac{1}{2}, u\right)$$

Poids :

$$\omega(u) = (1 - u^2)^{-1/2}$$

Norme :

$$\eta = \sqrt{\frac{\pi}{2}}$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r U^N(k, 0) U^N(k, u) \omega(u) \quad (1.27)$$

Polynômes d'Hermite (limites d'orthogonalité \mathbb{R})

Formule de Rodrigues :

$$H(k, u) = (-1)^k \exp(u^2) \frac{d^k}{du^k} [\exp(-u^2)]$$

Poids :

$$\omega(u) = \exp(-u^2)$$

Norme :

$$\eta(k) = \sqrt{2^k k! \sqrt{\pi}}$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r H^N(k, 0) H^N(k, u) \omega(u) \quad (1.28)$$

Polynômes de Laguerre (limites d'orthogonalité \mathbb{R}^+)

Formule de Rodrigues :

$$Lag(k, \alpha, u) = \frac{e^u u^{-\alpha}}{k!} \frac{d^k}{du^k} [e^{-u} u^{k+\alpha}] ; \alpha > -1$$

Poids :

$$\omega(\alpha, u) = u^\alpha e^{-u}$$

Norme :

$$\eta(k) = \sqrt{\frac{\Gamma(k + \alpha + 1)}{k!}}$$

Noyau correspondant :

$$K(u) = \sum_{k=0}^r \text{Lag}^N(k, \alpha, 0) \text{Lag}^N(k, \alpha, u) \omega(\alpha, u) \quad (1.29)$$

Les figures (1.3) à (1.6) illustrent les allures que prennent quelques uns de ces noyaux à différents ordres sur l'intervalle $] -1, 1[$.

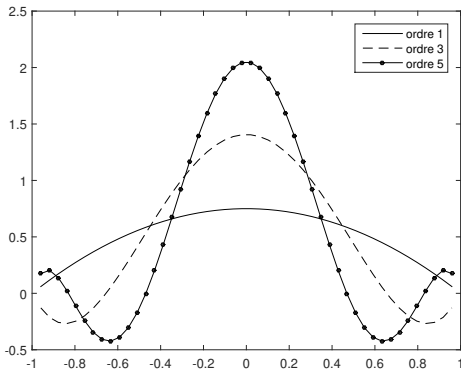


FIGURE 1.3 – Noyaux à base de polynômes de Jacobi (ordres : 1, 3, 5)

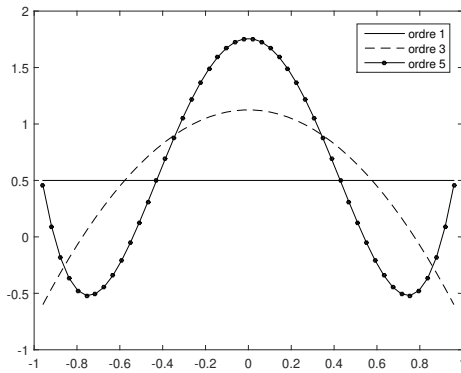


FIGURE 1.4 – Noyaux à base de polynômes de Legendre (ordres : 1, 3, 5)

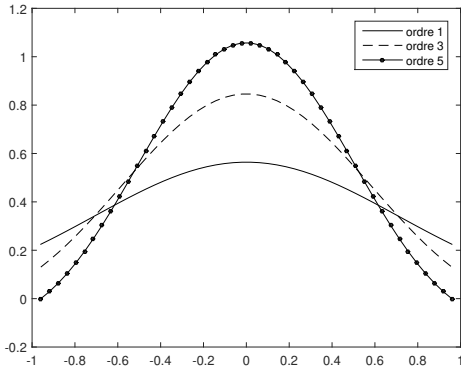


FIGURE 1.5 – Noyaux à base de polynômes d'Hermite (ordres : 1, 3, 5)

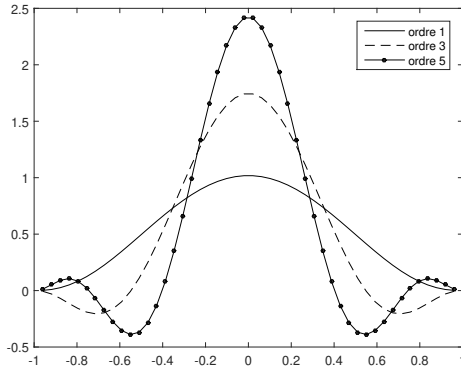


FIGURE 1.6 – Noyaux à base de polynômes de Gegenbauer (ordres : 1, 3, 5)

1.4 Sélection du paramètre de lissage

Sélectionner le paramètre de lissage constitue une préoccupation majeure lors de l'implémentation des techniques de lissage par noyaux. D'ailleurs, dans le cas de l'estimation de densité, la qualité du résultat dépend étroitement de la quantité de lissage à appliquer, en raison du fait que le paramètre

h qui contrôle cette quantité dicte la façon dont le noyau est déployé sur le support. Une petite valeur de h conduit donc à un sous lissage qui se traduit par une densité à l'aspect volatile et rugueux reproduisant des traits fallacieux causés principalement par le processus d'échantillonnage. Néanmoins, il se peut qu'un faible lissage, appliqué de manière modérée, soit utile dans la mesure où il permet de faire ressortir des nuances imperceptibles mais au coût d'une queue plutôt bosselée. Dans le cas où une valeur élevée est donnée à h , le lissage appliqué est élevé aussi et risque de mener vers un sur-lissage qui, même s'il offre un traitement correct au niveau des queues, n'arrive pas à capturer des caractéristiques importantes de la densité à estimer.

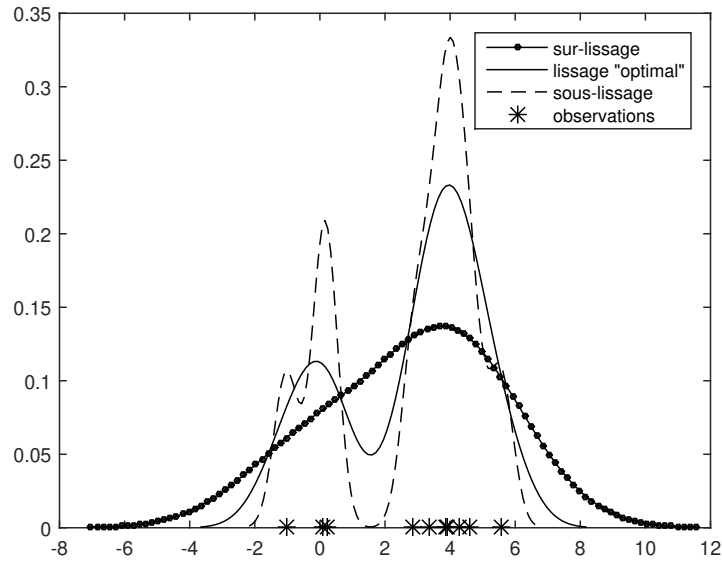


FIGURE 1.7 – Différents degrés de lissage pour une estimation de densité à 10 observations

Ces phénomènes sont principalement une manifestation de l'impact que peut avoir la valeur de h sur l'estimation. En examinant, par exemple, les équations (1.14) et (1.15) qui correspondent respectivement au biais et à la variance on remarque que le cas du sur-lissage causé par un h élevé est le résultat d'un trop grand biais et d'une faible variance alors que le sous-lissage causé par un petit h est le résultat d'un faible biais et d'une grande variance. On en déduit donc, qu'en théorie, le choix du paramètre de lissage doit prendre en compte cette relation qui existe entre le biais et la variance de sorte à garantir un certain équilibre entre ces deux sources d'erreur (1.8).

Ainsi, le dilemme engendré par l'effet du paramètre de lissage sur l'estimation a conduit à ce que les techniques de sélection automatique aient en général un même point de départ qui repose sur la minimisation d'une mesure de l'erreur souvent choisie comme le critère *AMISE*. Pourtant, il n'est pas garanti qu'une valeur du paramètre de lissage obtenue par une méthode de sélection automatique soit adéquate aux fins pour lesquelles l'estimation est conduite, étant donné que ces dernières influencent largement ce que peut être un choix convenable.

Une sélection automatique est toutefois utile, dans le sens où elle permet d'offrir une première valeur raisonnable à partir de laquelle l'utilisateur peut se diriger vers un lissage qui soit, par exemple, plus approprié pour mettre en relief des aspects importants de la structure de données ou pour mener des analyses exploratoires. Ce type de sélection est également nécessaire lorsque l'on doit recourir à des méthodes qui exigent une application répétitive de lissages et où une sélection interactive risque de devenir lourde et onéreuse en termes de temps.

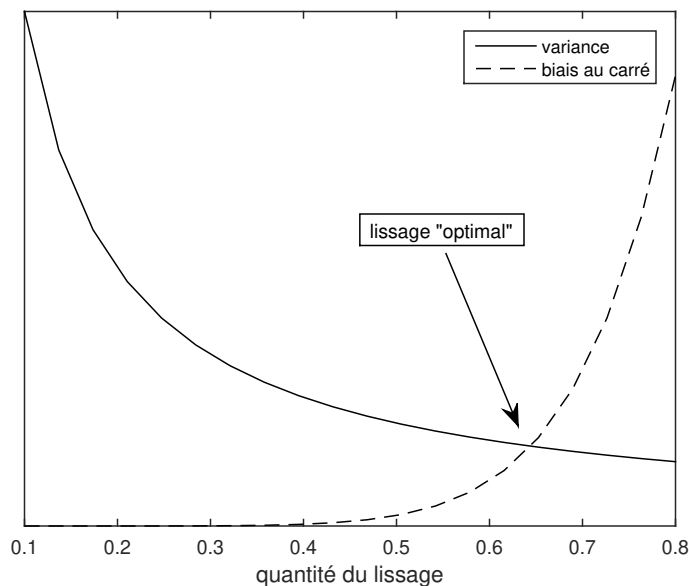


FIGURE 1.8 – Illustration du compromis biais-variance

Par ailleurs, malgré l'existence de plusieurs méthodes de sélection automatique, il n'y a aucun accord sur laquelle est la meilleure ou la plus fiable. Mais au moins, il existe une sorte de consensus qui vise à classer ces méthodes en deux catégories :

- Une première catégorie qui inclut des techniques, telles que quelques approches basées sur la validation croisée, qui sont plutôt simples et dont le but est de procurer une valeur acceptable du paramètre de lissage mais sans offrir de garantie quant à son optimalité.
- Une deuxième catégorie qui se veut plus rigoureuse et pointue et dont les méthodes englobent principalement les techniques de bootstrap et d'insertion (plug-in).

Nous présentons alors dans ce qui suit, quelques-unes de ces méthodes classées selon la catégorie à laquelle elles appartiennent.

Comme nous l'avons souligné plus tôt, afin de fournir une valeur pour le paramètre de lissage, plusieurs approches partent d'un point commun qui est la minimisation du critère *AMISE*. Donc, dans le cas d'un noyau d'ordre 2 (Définition 1.3), en dérivant l'équation (1.17) par rapport à h et en égalisant à zéro, on trouve la valeur optimale du paramètre de lissage (équation (1.30)) qui minimise l'erreur.

$$h_{(2)}^{opt} = \left[\frac{R(K)}{n\mu_2^2(K) R(f'')} \right]^{1/5} \quad (1.30)$$

Pour un noyau d'ordre r quelconque et en utilisant cette fois-ci l'équation (1.20), on trouve la valeur optimale suivante :

$$h_{(r)}^{opt} = \left[\frac{(r!)^2 R(K)}{2nr\mu_r^2(K) R(f^{(r)})} \right]^{1/(2r+1)} \quad (1.31)$$

Une utilisation immédiate de ces formules (équations (1.30) et (1.31)) n'est pourtant pas possible parce qu'elles-mêmes dépendent de la densité inconnue f à travers les termes $R(f'')$ et $R(f^{(r)})$ et c'est principalement cette difficulté qui est à l'origine de la diversité des méthodes de sélection.

Toutes les approches illustrées dans ce qui suit correspondent au cas d'un noyau d'ordre 2 et peuvent être facilement adaptées aux noyaux d'ordres supérieurs.

1.4.1 Première catégorie

Règle pratique (Référence à une famille paramétrique)

Cette méthode doit sa popularité à Silverman [96] et repose sur un principe simple et direct qui vise à remplacer le terme inconnu de la formule du paramètre de lissage par une estimation se référant à la densité d'une loi paramétrique généralement choisie comme la loi normale dont le paramètre d'échelle est estimé à partir des données. Ainsi, l'évaluation du paramètre de lissage peut se faire aisément à travers la formule

$$\hat{h}_N = \left[\frac{8\sqrt{\pi}R(K)}{3n\mu_2^2(K)} \right]^{1/5} \hat{\sigma} \quad (1.32)$$

où $\hat{\sigma}$ est une estimation du paramètre d'échelle.

Or, autant cette méthode paraît simple et attrayante, autant elle risque d'entraîner des erreurs dans l'estimation, puisqu'elle tend à attribuer une valeur élevée au paramètre h et par conséquent causer un sur-lissage qui masquera des modes majeurs dans le cas de densités multimodales ou des propriétés importantes dans le cas des densités unimodales lorsqu'elles diffèrent significativement de la densité normale. En contrepartie, le sur-lissage dont il est question a l'avantage de décourager toute interprétation excessive d'aspects superflus.

Un autre inconvénient de cette technique est sa sensibilité aux valeurs aberrantes, causée principalement par le terme $\hat{\sigma}$ lorsqu'il est considéré comme l'écart-type de l'échantillon. Toutefois, ce problème peut être résolu en adoptant des estimateurs plus robustes pour le paramètre d'échelle tel que l'écart inter-quartile ou l'écart médian absolu.

Validation croisée par moindres carrés

Comme son nom l'indique, cette méthode est fondée sur le principe de validation croisée. Proposée d'abord par Rudemo [89] ensuite par Bowman [14], elle part du développement de l'erreur quadratique intégrée

$$\begin{aligned} ISE[\hat{f}_n] &= \int [\hat{f}_n(x) - f(x)]^2 dx \\ &= \int \hat{f}_n^2(x) dx - 2 \int f(x) \hat{f}_n(x) dx + \int f^2(x) dx \end{aligned}$$

le but est alors de choisir un paramètre de lissage qui permette de minimiser cette erreur. Le dernier terme, ne dépendant pas de h , conduit à ce que la forme à minimiser soit réécrite sous la forme

$$ISE[\hat{f}_n] - \int f^2(x) dx = \int \hat{f}_n^2(x) dx - 2 \int f(x) \hat{f}_n(x) dx$$

De plus, les deux autres termes peuvent être calculés à partir des données ce qui donne la forme suivante à la fonction de validation croisée par moindres carrés :

$$LSCV(h) = \int \hat{f}_n^2(x) dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{n,-i}(X_i) \quad (1.33)$$

où $\hat{f}_{n,-i}$ est un estimateur à noyau de la densité par exclusion de l'observation X_i qui correspond à

$$\hat{f}_{n,-i}(X_i) = \frac{1}{h(n-1)} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h}\right)$$

Il s'ensuit alors que le paramètre de lissage optimal soit la quantité qui minimise l'équation (1.33)

$$\hat{h}_{LSCV} = \arg \min_{h \in \mathbb{R}^+} LSCV(h)$$

À première vue, cette méthode, très populaire, paraît convenable et le paramètre qui peut en résulter semble être une estimation raisonnablement précise du paramètre de lissage. Il existe cependant, deux points qui constituent des faiblesses majeures dont souffre cette technique. Le premier point est la tendance de la fonction de validation croisée à avoir plusieurs minima locaux, chose qui risque d'engendrer un blocage de l'algorithme de minimisation au niveau de l'un d'eux. Un moyen d'éviter ce problème serait alors de tracer la fonction afin d'avoir un repère visuel pouvant aider à l'identification d'un minimiseur global de la fonction, ce minimiseur n'est pourtant pas la valeur idéale pour \hat{h}_{LSCV} puisqu'il cause un sous-lissage excessif de la densité. Le paramètre de lissage peut donc être choisi comme étant le plus grand des minima locaux ce qui réduit, mais n'annule pas, l'effet du sous lissage causé par cette méthode. Le deuxième point reproché à cette méthode est la grande variabilité des résultats qu'elle retourne, c'est à dire que pour des jeux de données issus d'une même distribution, les paramètres de lissage calculés diffèrent considérablement les uns des autres.

1.4.2 Deuxième catégorie

Insertion par résolution d'équation

L'insertion par résolution d'équation fait partie de la classe des méthodes d'insertion (Sheater et Jones [93] ; Engel, Hermann et Gasser [38]) qui, au lieu de faire appel à une famille paramétrique, recourent à l'estimation de fonctionnelles de la densité pour estimer la quantité inconnue dans la formule du paramètre de lissage (équation (1.30)). Ces techniques sont actuellement les plus utilisées en pratique du fait de leur très faible variabilité et de leur rapidité relative en comparaison avec la validation croisée par moindres carrés.

La variante d'insertion dont il est question dans ce paragraphe, rajoute la résolution d'une équation, que l'on détaillera plus tard, à la phase d'estimation du terme inconnu $R(f'')$. Cette dernière repose principalement sur une estimation de la fonctionnelle Φ_r , telle que

$$\Phi_r = \int f^{(r)}(x) f(x) dx$$

avec r pair.

Notons alors que

$$\Phi_r = E \left[f^{(r)}(x) \right]$$

et rappelons que

$$R(f^{(s)}) = \int [f^{(s)}(x)]^2 dx$$

Une intégration par partie donne

$$R(f^{(s)}) = (-1)^s \int f^{(2s)}(x) f(x) dx$$

Donc pour s pair et $r = 2s$ on a

$$R(f^{(s)}) = \Phi_r$$

Un estimateur de $R(f^{(s)})$ peut alors être donné par

$$\hat{\Phi}_r = \frac{1}{n} \sum_{i=1}^n \hat{f}_n^{(r)}(X_i)$$

avec

$$\hat{f}_n^{(r)}(x) = \frac{1}{n\beta^{r+1}} \sum_{i=1}^n L^{(r)}\left(\frac{x - X_i}{\beta}\right) \quad (1.34)$$

et β et L sont respectivement le paramètre de lissage pilote et le noyau pilote pour l'estimateur $\hat{\Phi}_r$. Le paramètre β est alors choisi de sorte à ce qu'il minimise l'erreur moyenne quadratique asymptotique $AMSE$ de $\hat{\Phi}_r$. Les calculs (annexe A1.4) donnent

$$\beta_{AMSE} = \left[\frac{-k!L^{(r)}(0)}{\mu_k(L)\Phi_{r+k}n} \right]^{1/(r+k+1)}$$

où k est l'ordre du noyau L selon la définition (1.3).

Donc, pour $L = K$ un noyau d'ordre 2 (définition (1.3)) et $\hat{\Phi}_4$ un estimateur de $R(f'')$ on aura

$$\beta = \left[\frac{-2K^{(4)}(0)}{n\mu_2(K)\hat{\Phi}_6} \right]^{1/7} \quad (1.35)$$

En remplaçant par n de l'équation (1.30) dans l'équation (1.35) on trouve

$$\beta = h_{AMISE}^{5/7} \left[\frac{-2K^{(4)}(0)\mu_2(K)\hat{\Phi}_4(b_1)}{R(K)\hat{\Phi}_6(b_2)} \right]^{1/7} \quad (1.36)$$

tels que b_1 et b_2 sont respectivement les paramètres de lissage optimaux des estimateurs $\hat{\Phi}_4$ et $\hat{\Phi}_6$ en terme d'erreur moyenne quadratique asymptotique. Bien entendu, cette approche peut s'étendre indéfiniment puisque le paramètre de lissage optimal de $\hat{\Phi}_r$ dépend toujours de $\hat{\Phi}_{r+2}$. Un moyen commun de résoudre ce problème est de s'arrêter à un certain stade du développement et calculer des valeurs exactes de Φ_r et Φ_{r+2} en se référant à une loi paramétrique.

Ceci rendra possible la résolution de l'équation

$$h_{AMISE} - \left[\frac{R(K)}{n\mu_2^2(K)\hat{\Phi}_4(\beta)} \right]^{1/5} = 0$$

dont l'inconnu est le paramètre de lissage optimal spécifiée par l'équation (1.30).

Remarque 1.3 *Il convient de noter que pour obtenir rapidement un estimateur de $R(f^{(s)})$ et éviter ainsi la complexité de l'insertion par résolution d'équation, une insertion immédiate à été adoptée dans Ouamaliche et Sayah [79] à travers l'utilisation d'une référence à la loi normale pour la sélection du paramètre β et par l'évaluation de l'équation (1.34) par la r -ième dérivée d'un noyau Gaussien*

$$L^{(r)}(u) = (-1)^r H(r, u) L(u)$$

où $H(r, \cdot)$ est un polynôme d'Hermite d'ordre r et $L(u) = (1/\sqrt{2\pi}) \exp[-(u^2/2)]$.

Validation croisée lissée

La sélection du paramètre h par validation croisée lissée permet de réduire la variabilité qui se répercute directement sur l'estimation de la densité et dont souffrent les méthodes classiques de validation croisée. Cette technique, présente une certaine similarité avec les méthodes d'insertion dans le sens où elle fait appel à un paramètre de lissage pilote qui va servir à estimer le biais quadratique intégré du critère *MISE* de l'estimateur de densité à noyau. En effet, cette technique adopte une approximation asymptotique pour la variance uniquement, tandis que pour le biais, la forme exacte est conservée. On se retrouve alors avec la fonction suivante à minimiser :

$$SCV(h) = \frac{1}{nh} R(K) + ISB(h)$$

où *ISB* est le biais quadratique intégré qui, écrit de manière compacte, donne

$$ISB(h) = \int (K_h * f - f)^2(x) dx$$

avec $K_h(u) = (1/h)K(u/h)$ et $(f * g)(x) = \int f(x-y)g(y)dy$.

Une estimation du ISB peut alors être donnée par

$$I\hat{S}B(h) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n (K_h * K_h * L_g * L_g - 2 * K_h * L_g * L_g + L_g * L_g)(X_i - X_j)$$

avec g le paramètre de lissage pilote et L le noyau pilote de l'estimateur à noyau de la densité f . Bien évidemment, Ces deux quantités sont inconnues, néanmoins elles peuvent être identifiées sans grande peine, le noyau pilote L par exemple peut-être choisit comme étant égal au noyau K alors que le paramètre g est choisit selon une approche similaire à celle utilisée dans le cas de la sélection par insertion. Par ailleurs, il est important de noter que cette méthode a elle aussi son lot de défauts, notamment la difficulté d'implémentation et le sous-lissage excessif. Toujours est-il qu'il est possible de surmonter ces problèmes, d'une part, en utilisant des noyaux gaussiens pour simplifier l'implémentation et d'autre part, en choisissant le plus grand des minima locaux comme paramètre h pour réduire le sous-lissage, comme c'est le cas pour la validation croisée par moindres carrés.

1.5 Quelques résultats asymptotiques

L'intérêt des résultats asymptotiques provient du fait que lorsqu'il s'agit d'un échantillon ayant une taille particulièrement importante, ils vont permettre de décrire le comportement de l'estimateur dans un cadre général bien défini. D'habitude, K et f sont fixées et satisfont quelques conditions de régularité, l'estimation est construite à partir de n observations indépendantes et identiquement distribuées où $n \rightarrow \infty$ et le paramètre de lissage h dépend de n de sorte à ce que h diminue lorsque n augmente, toutefois h ne doit pas converger vers 0 aussi rapidement que n^{-1} .

Quelques-uns des aspects importants que l'on peut traiter dans le cadre asymptotique sont la consistance ponctuelle, la consistance sur le support de la densité, l'ordre de convergence et la qualité de l'optimalité asymptotique. Nous intégrons dans cette partie quelques résultats en rapport avec la consistance ainsi qu'une courte discussion concernant l'ordre de convergence. L'optimalité asymptotique, quant à elle, sera traitée dans l'annexe (A1.5).

Soient les hypothèses suivantes :

$$H1 : \int K(u) du < \infty ; \int K(u) du = 1 ; |uK(u)| \xrightarrow{u \rightarrow \infty} 0$$

$$H2 : h \rightarrow 0 ; nh \xrightarrow{n \rightarrow \infty} \infty$$

$$H3 : h \rightarrow 0 ; nh [\log(n)]^{-1} \xrightarrow{n \rightarrow \infty} \infty$$

Consistance ponctuelle

La consistance d'un estimateur de densité est sa capacité à converger selon un certain mode vers la densité réelle.

Théorème 1.1 (Parzen [81]) *Soit f une densité continue sur son support et \hat{f}_n son estimateur à noyau classique dont le noyau vérifie les hypothèses H1 et le paramètre de lissage vérifie les hypothèses H2. On a alors*

$$\hat{f}_n(x) \xrightarrow{P} f(x) \text{ lorsque } n \rightarrow \infty$$

Consistance sur le support

Théorème 1.2 (Devroye et Györfi [31]) *Soit f une densité continue sur son support et \hat{f}_n son estimateur à noyau classique dont le noyau vérifie les hypothèses H1 et le paramètre de lissage vérifie les hypothèses H2. On a alors*

$$\int |\hat{f}_n(x) - f(x)| dx \rightarrow 0 \text{ lorsque } n \rightarrow \infty$$

Théorème 1.3 (Silverman [95] ; Bertrand-Retali [4]) *Soit f une densité continue uniformément sur son support et \hat{f}_n son estimateur à noyau classique dont le noyau vérifie les hypothèses H1 et le paramètre de lissage vérifie les hypothèses H3. On a alors*

$$\sup_x |\hat{f}_n(x) - f(x)| \rightarrow 0 \text{ lorsque } n \rightarrow \infty$$

Ordre de convergence

L'ordre de convergence est étroitement lié au critère selon lequel il est considéré. Pour le retrouver au sens de l'erreur *MISE* par exemple, il est nécessaire d'utiliser les approximations asymptotiques données par les équations (1.17) et (1.20) ainsi que les paramètres de lissage donnés par les équations (1.30) et (1.31) qui minimisent respectivement ces approximations.

En effet, le fait de remplacer dans chacune des équations (1.17) et (1.20) par le paramètre de lissage 'optimal' convenable, va produire les critères *AMISE* suivants :

$$AMISE_{opt} [\hat{f}_n] = \frac{5}{4} [\mu_2^2(K) R^4(K) R(f'')]^{1/5} n^{-4/5} \quad (1.37)$$

$$AMISE_{opt} [\hat{f}_n] = \frac{2r+1}{2r} [2r(r!)^{-2} \mu_r^2(K) R^{2r}(K) R(f^{(r)})]^{1/(2r+1)} n^{(-2r)/(2r+1)} \quad (1.38)$$

Selon la définition (1.3), l'équation (1.37) concernera alors les noyaux d'ordre 2, tandis que l'équation (1.38) concernera les noyaux d'ordre $r > 2$.

On remarque donc que pour un noyau d'ordre 2, la valeur *AMISE* tend vers 0 avec un ordre de $n^{-4/5}$ alors que pour un noyau d'ordre supérieur on peut s'attendre, dès lors que r est suffisamment élevé, à ce que l'ordre de convergence soit particulièrement proche de n^{-1} qui est typiquement l'ordre retrouvé pour les problèmes paramétriques. Par conséquent, ce résultat encourage l'utilisation de noyaux d'ordres supérieurs lorsque les données sont abondantes.

Il faudra néanmoins insister sur le fait que ces ordres de convergence élevés par lesquels se distinguent les estimateurs à noyaux sont réalisables uniquement quand le paramètre de lissage est

choisit correctement, surtout que les hypothèses de consistance d'un estimateur à noyau peuvent être trompeuses. Autrement dit, il existe bien des valeurs qui peuvent satisfaire les hypothèses mais en même temps mener vers des ordres de convergence décevants.

Chapitre 2

Noyaux continus asymétriques

Bien que les noyaux continus symétriques aient prouvé un degré élevé d'efficacité dans les cas les plus classiques, leur utilisation continue de se heurter aux effets de bords qui sont un inconvénient majeur lorsque le support de la densité à estimer est borné. Ce problème surgit principalement quand le support du noyau dépasse l'intervalle dans lequel les observations sont enregistrées. L'efficacité des noyaux symétriques est alors préservée dans les régions intérieures du support tandis qu'elle se voit réduite quand l'estimation se fait auprès des bords à cause des poids qui sont assignés à l'extérieur du support de la densité à estimer et qui risquent d'augmenter la valeur du biais de l'estimation. Cette complication est d'autant plus grave lorsqu'une partie importante des données se retrouve aux bords du support.

Afin de remédier à ce problème de biais, plusieurs approches furent proposées. Leur point commun, en plus du fait qu'elles se basent généralement sur des noyaux classiques symétriques, est qu'elles vont diviser le support en région intérieure et régions de bord pour ensuite appliquer pour chacune de ces régions un lissage différent. Par exemple, La méthode des données reflétées (Silverman [96]) propose d'isoler les observations qui causeront une attribution de poids hors du support pour ensuite rajouter aux données disponibles leurs réflexions par rapport aux bords. La méthode des noyaux de bords (Müller [74]) utilise pour chaque point situé à proximité des bords un noyau différent afin de déplacer à l'intérieur du support la majorité du poids. La méthode des pseudo-données (Cowling et Hall [27]) qui ressemble en quelque sorte à la méthode des données reflétées, recourt à la création, au-delà des bords, de nouvelles données qui sont des combinaisons linéaires des observations existantes. Ces méthodes sont toutefois sujettes à quelques limitations parmi lesquelles les restrictions d'application à certaines classes de densités et la génération d'estimations négatives aux bords.

Outre les méthodes hybrides et plutôt lourdes (Gustafsson et al. [51]; Karunamuni et Alberts [66]) qui furent développées afin de contourner ces limitations en combinant plusieurs techniques, un autre moyen de résoudre ces problèmes sans pour autant créer de complications superflues est de faire coïncider les supports du noyau et de la densité à estimer (Abadir et Lawford [1]).

Cette dernière idée fut à la base de l'émergence des noyaux asymétriques où les noyaux utilisés vont changer de forme en fonction du point dans lequel l'estimation est effectuée. Le noyau est alors symétrique uniquement au milieu du support et plus on s'éloigne de ce point, plus l'asymétrie est prononcée jusqu'à ce qu'elle devienne maximale lorsqu'on atteint les bords. Par conséquent, ceci évite les biais de bord dans le sens où aucun poids n'est assigné en dehors du support.

Dans ce chapitre l'intérêt est porté sur l'estimation de densités à base de noyaux asymétriques. On commence par donner une définition générale pour ce type de noyau avant d'énoncer les formes

fonctionnelles des noyaux les plus importants et de passer à la mesure de l'erreur spécifique des estimateurs de densité à base de chacun de ces noyaux. Par la suite on aborde le sujet de la sélection du paramètre de lissage et on finit par quelques résultats de consistance et une discussion de l'ordre de convergence.

2.1 Éléments de base

Afin de garder le lien avec les noyaux continus symétriques, on note l'estimateur de densité à noyau continu asymétriques à l'aide de la fonction de poids général. Cette fonction permet de réécrire l'estimateur donné par l'équation (1.7) sous la forme suivante

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \quad (2.1)$$

où

$$K_{x,h}(u) = \frac{1}{h} K\left(\frac{x-u}{h}\right)$$

$K_{x,h}$ donne alors une forme plus générale au noyau de sorte à ce que l'ensemble des noyaux symétriques devienne un cas particulier d'un éventail plus large de fonctions, dont les noyaux asymétriques. L'estimateur de densité à base de ces derniers est alors donné par l'équation (2.1). Concernant la fonction de densité à estimer, il convient de préciser qu'elle a un support $[a, b]$ où $a \in \mathbb{R}$ et $b \in \mathbb{R}$.

Dans l'absence d'une définition formelle des noyaux asymétriques (Hirukawa [58]), on peut toujours établir une définition à travers des propriétés qui regroupent les aspects particuliers qui caractérisent ces noyaux. D'abord, ils permettent de produire des estimations qui sont positives et sans effets de bords à l'image des noyaux suggérés dans les travaux de Guillamón, Navarro et Ruiz [50] et Abadir et Lawford [1]. Ces derniers sont toutefois conçus selon le cadre classique qui n'accorde pas d'attention particulière à la manière dont les observations sont dispersées, ils appliquent donc, sur la totalité du support, une quantité constante de lissage déterminée par l'unique paramètre h . Il est alors important de rajouter que les noyaux asymétriques utilisent une forme de lissage adapté afin de capter les différentes formes que peuvent prendre les courbes estimées. En effet, la quantité de lissage appliqué dépend, en plus du paramètre h , de l'endroit du support dans lequel l'estimation est conduite. Il s'ensuit un lissage différent pour chaque point et par conséquent une allure du noyau qui va être variable tout au long du support. En termes libres, on peut donc définir les noyaux asymétriques par ce qui suit :

Définition 2.1 *On appelle noyau asymétrique toute fonction de poids $K_{x,h}$ ayant les deux propriétés suivantes :*

- *La fonction de poids est une fonction de densité de probabilité de support $[a, b]$ avec $a \in \mathbb{R}$ et $b \in \mathbb{R}$.*
- *Les paramètres de forme et de position de la fonction $K_{x,h}$ sont tous deux fonctions du point d'estimation x et du paramètre h .*

Notons que puisque la fonction $K_{x,h}$ est une densité de probabilité, elle peut posséder un paramètre d'échelle qui régit son aplatissement, un paramètre de position qui régit sa localisation et

un paramètre de forme qui régit son allure. La loi normale $N(\mu, \sigma^2)$, par exemple, a comme paramètre de position la moyenne μ et comme paramètre d'échelle l'écart-type σ tandis que la loi bêta $Beta(a, b)$ possède deux paramètres a et b qui sont tous les deux des paramètres de forme.

Remarque 2.1 *Étant destiné à surmonter le problème des effets de bords, le noyau asymétrique doit être choisi en tenant compte du support de la densité f à estimer. En effet, Si l'on considère $\mathbb{S}_{x,h}$ le support du noyau $K_{x,h}$ et \mathbb{S} le support de la densité f on doit alors avoir $\mathbb{S} \subseteq \cup_x \mathbb{S}_{x,h}$.*

Remarque 2.2 *L'aptitude des noyaux asymétriques à changer de forme selon le point d'estimation x est intrinsèquement liée à la particularité qu'ont ces noyaux de converger vers ce même point de sorte que l'on ait $E(K_{x,h}) \rightarrow x$ lorsque $h \rightarrow 0$ avec $K_{x,h}$ une variable aléatoire de densité $K_{x,h}$.*

Contrairement au cas vu au chapitre précédent où les noyaux pouvaient très bien ne pas découler d'une densité de probabilité, les composantes de bases des noyaux continus asymétriques sont nécessairement des fonctions de densité asymétriques. On pourra citer par exemple les densités des loi de Pareto, Loi log-normale, loi inverse gaussienne, loi gamma et loi bêta. La table (2.2) donne un récapitulatif de ces densités. Précisons que les fonctions gamma et bêta qui apparaissent respectivement dans les fonctions de densités de mêmes noms désignent les intégrales suivantes :

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

Par ailleurs, bien que les noyaux asymétriques soient fondamentalement de simples fonctions de densités, les estimateurs à base de ces noyaux ne sont pas forcément des densités de probabilité (Gourieroux et Monfort [49]). La table (2.1) présente, pour différents échantillons, les valeurs des intégrales de quelques estimateurs à noyaux asymétriques sur leur support respectifs.

	Échantillon 1	Échantillon 2	Échantillon 3
Noyau gamma	1.0375	1.0078	0.9953
Noyau bêta	1.0444	1.0592	1.0473

TABLE 2.1 – Intégrales d'estimateurs à noyau asymétrique pour 3 échantillons aléatoires de taille 5000 issus d'une loi normale

On peut donc présumer que la variété des noyaux asymétriques est plutôt limitée, étant donné la nature du lien qu'ont ces noyaux avec les fonctions de densité asymétriques. Cet inconvénient peut néanmoins être adressé à travers la manipulation des formes fonctionnelles des paramètres des noyaux asymétriques qui peut permettre d'obtenir divers noyau issu d'une même loi (Igarashi et Kakizawa [62]).

En ce qui nous concerne, nous nous intéressons exclusivement aux noyaux issus de densités bêta et gamma. D'une part parce que les supports de ces densités qui sont $[0, 1]$ et \mathbb{R}^+ peuvent être considérés comme représentatifs des supports qui risquent de causer des effets de bords lors d'une estimation à noyau et d'autre part parce que ces noyaux semble bien se prêter aux applications qui impliquent des données financières telles que les rendements, les taux d'intérêt et les manques à gagner qui, en général, disposent de distributions se distinguant par l'existence d'une borne inférieure, souvent l'origine, auprès de laquelle s'accumulent les observations créant ainsi une asymétrie positive et une distribution de données sporadique tout au long de la queue de droite.

Loi	Fonction de densité	Support	Espérance	Variance
Pareto : $\mathcal{P}(\alpha, \beta)$	$f(x) = \beta \frac{\alpha^\beta}{x^{\beta+1}}$	$[\alpha, +\infty[$	$\frac{\alpha\beta}{\beta-1}$	$\frac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}$
Log-normale : $LN(\alpha, \beta^2)$	$f(x) = \frac{1}{x\beta\sqrt{2\pi}} \exp\left(-\frac{(\log(x) - \alpha)^2}{2\beta^2}\right)$	$]0, +\infty[$	$\exp\left(\alpha + \frac{\beta^2}{2}\right)$	$(\exp(\beta^2) - 1) \exp(2\alpha + \beta^2)$
Inverse Gaussienne : $IG(\alpha, \beta)$	$f(x) = \left(\frac{\beta}{2\pi x^3}\right)^{1/2} \exp\left(\frac{-\beta(x - \alpha)^2}{2\alpha^2 x}\right)$	$]0, +\infty[$	α	$\frac{\alpha^3}{\beta}$
Gamma : $Gamma(\alpha, \beta)$	$f(x) = \frac{x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha) \beta^\alpha}$	$[0, +\infty[$	$\alpha\beta$	$\alpha\beta^2$
Bêta : $Beta(\alpha, \beta)$	$f(x) = \frac{1}{\beta(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$[0, 1]$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$

TABLE 2.2 – Description de quelques lois asymétriques

De surcroît, contrairement aux estimateurs à noyaux continus symétriques dont les comportements asymptotiques des propriétés locales et globales peuvent être analysés à partir d'un nombre restreint de formules se basant sur un ensemble de propriétés et de conditions communes, l'analyse du comportement des estimateurs à noyaux asymétriques requiert des développements asymptotiques qui tendent à varier en fonction du noyau utilisé. Une étude de tous les noyaux risque donc d'être pesante et peu utile.

2.2 Formes fonctionnelles et allures des noyaux gamma et bêta

2.2.1 Noyau gamma

Le noyau gamma apparait dans la littérature sous différentes formes. Initialement, Chen [24] se sert de deux variantes de ce noyau lesquelles sont évidemment des fonctions de densité gamma qui diffèrent par leurs paramétrisations. La première variante utilise les paramètres $(\alpha = x/h + 1, \beta = h)$

$$K_{x/h+1,h}^G(t) = \frac{t^{x/h} e^{-t/h}}{h^{x/h+1} \Gamma(x/h + 1)} ; t \geq 0 \quad (2.2)$$

tandis que la deuxième utilise $(\alpha = \rho_h(x), \beta = h)$

$$K_{\rho_h(x),h}^G(t) = \frac{t^{\rho_h(x)-1} e^{-t/h}}{h^{\rho_h(x)} \Gamma(\rho_h(x))} ; t \geq 0 \quad (2.3)$$

avec

$$\rho_h(x) = \begin{cases} (x/h) & \text{si } x \geq 2h \\ (x/(2h))^2 + 1 & \text{si } x \in [0, 2h[\end{cases}$$

Les quelques travaux qui mentionnent les noyaux asymétriques dans le traitement de données sur $[0, +\infty[$ font intervenir majoritairement ces deux variantes dans leur formes originales ou en leur appliquant quelques légères modifications. Par exemple, Bouezmarni, El Ghouch et Mesfioui [9] utilisent la deuxième variante telle quelle afin de traiter des données censurées à droite, Jeon et Kim [63] adoptent la première variante, pour étudier des pertes en assurance, en modifiant ses paramètre de sorte à ce que le point d'estimation x est remplacé par les observations X_i ($1 \leq i \leq n$) utilisées pour le calcul des poids, Malec et Schienle [73] prennent en compte les deux variantes et arrivent à la conclusion que la première forme devient plus efficace que la deuxième dès que $x \geq 2h$, ils proposent également de rajouter au noyau un paramètre constant dont la valeur sera déterminée en fonction des données disponibles.

La tendance des auteurs à conserver les noyaux issus de la loi gamma peut être justifiée par la flexibilité prononcée de la fonction de densité de cette loi tout au long de son support. La figure (2.1) montre les formes de la densité gamma sur l'intervalle $[0, 20]$ pour différentes valeurs des paramètres α et β .

Une autre indication de l'adaptabilité chez les noyaux gamma est l'allure changeante qu'ils présentent lorsque l'on fait varier le point d'estimation x en fixant le paramètre de lissage h ou au contraire lorsque l'on fixe le point d'estimation et l'on fait varier le paramètre de lissage qui est également paramètre d'échelle des noyaux gamma. Une illustration de ce comportement est donnée dans les figures (2.2) à (2.5) dans le cas des deux variantes données par les équations (2.2) et (2.3).

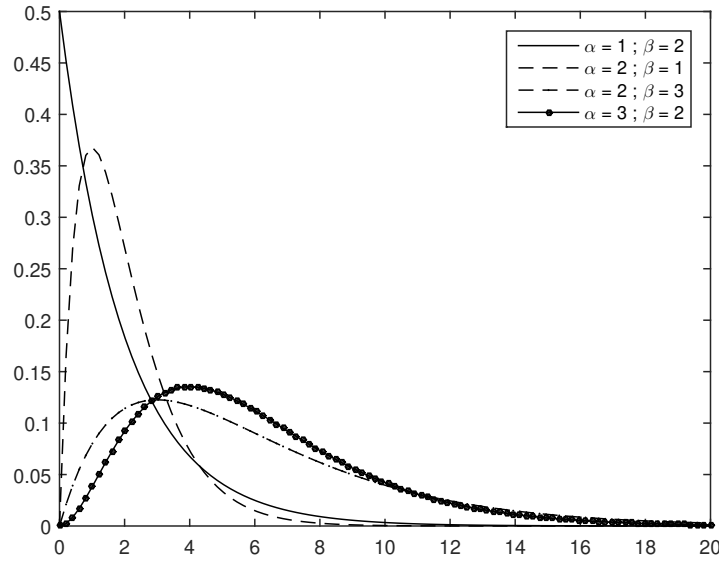


FIGURE 2.1 – Densité gamma pour différentes valeurs des paramètres α et β

2.2.2 Noyau bêta

À l'instar des noyaux gamma, les noyaux bêta peuvent également prendre différentes formes qui se distinguent les unes des autres par leurs paramètres. Principalement, on rencontre dans la littérature deux variantes de ce noyau qui, elles aussi, furent proposées par Chen [23]. Bien entendu, ces variantes sont des densités bêta dont les paramètres sont $(\alpha = x/h + 1, \beta = (1 - x)/h + 1)$ pour la première (équation (2.4)) et $(\alpha = \varrho_{(h,0)}(x), \beta = \varrho_{(h,1)}(x))$ pour la deuxième (équation (2.5))

$$K_{x/h+1, (1-x)/h+1}^B(t) = \frac{t^{x/h} (1-t)^{(1-x)/h}}{B(x/h+1, (1-x)/h+1)} ; t \in [0, 1] \quad (2.4)$$

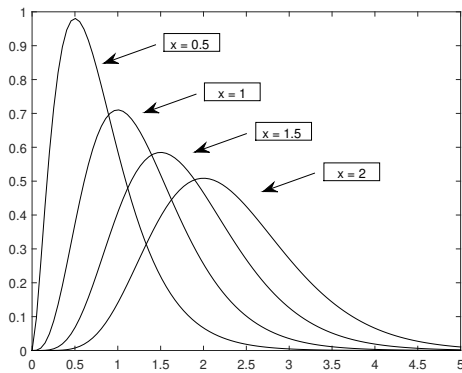


FIGURE 2.2 – Noyau gamma (première variante) pour $h = 0.3$ et x variable

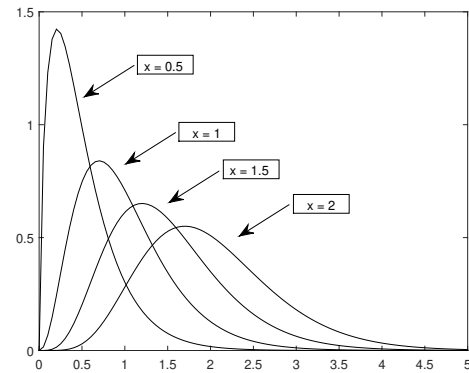


FIGURE 2.3 – Noyau gamma (deuxième variante) pour $h = 0.3$ et x variable

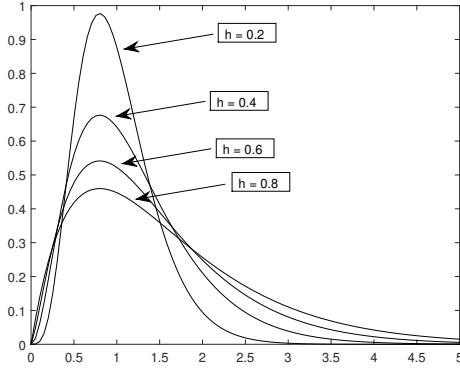


FIGURE 2.4 – Noyau gamma (première variante) pour $x = 0.8$ et h variable

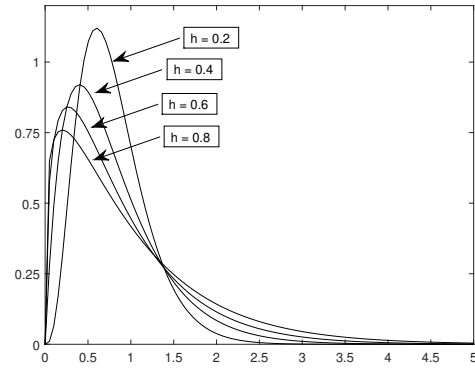


FIGURE 2.5 – Noyau gamma (deuxième variante) pour $x = 0.8$ et h variable

$$K_{\varrho_{h,0}(x), \varrho_{h,1}(x)}^B(t) = \frac{t^{\varrho_{h,0}(x)-1} (1-t)^{\varrho_{h,1}(x)-1}}{B(\varrho_{h,0}(x), \varrho_{h,1}(x))} ; t \in [0, 1] \quad (2.5)$$

avec

$$\varrho_{h,0}(x) = \begin{cases} \varrho_h(x) & \text{si } x \in [0, 2h[\\ x/h & \text{si } x \in [2h, 1] \end{cases}$$

$$\varrho_{h,1}(x) = \begin{cases} (1-x)/h & \text{si } x \in [0, 1-2h] \\ \varrho_h(1-x) & \text{si } x \in]1-2h, 1] \end{cases}$$

où

$$\varrho_h(x) = 2h^2 + \frac{5}{2} - \sqrt{4h^2 + 6h^2 + \frac{9}{4} - x^2 - \frac{x}{h}}$$

On peut remarquer que, tout comme pour le cas des noyaux gamma, la deuxième variante réserve un traitement particulier pour les bords de l'intervalle $[0, 1]$ en attribuant différents paramètres au noyau selon la zone dans laquelle se trouve le point d'estimation. L'objectif principal de cette alternative est de produire un estimateur avec une expression de biais qui n'inclut pas la dérivée première de la fonction à estimer. On constate cependant que les auteurs ayant utilisé une estimation à noyau bêta pour traiter des cas pratiques, préfèrent, en général, opter pour la première variante. On peut citer par exemple des applications, en pertes d'assurances (Gustafsson et al. [51]), en recouvrement (Renault et Scaillet [85]) et en volatilité (Kristensen [68]), dans lesquelles l'accent est mis uniquement sur la première forme du noyau bêta sans aucune allusion à la deuxième.

Deux autres points de ressemblance avec les noyaux gamma sont, premièrement, la souplesse illustrée dans la figure (2.6) par la diversité des formes que prend la densité bêta sur tout son support et qui correspondent chacune à des valeurs particulières pour les paramètres (α, β) , et deuxièmement, le mécanisme qui relie la quantité du lissage appliquée par l'estimateur à un certain point, non pas à la simple valeur du paramètre de lissage h comme dans le cas d'un lissage adaptatif classique, mais à l'allure du noyau, sachant que celle-ci dépend étroitement du point d'estimation x

et du paramètre h . Les figures (2.7) à (2.10) représentent clairement cette dépendance sur tout le support pour les deux variantes du noyau bêta.

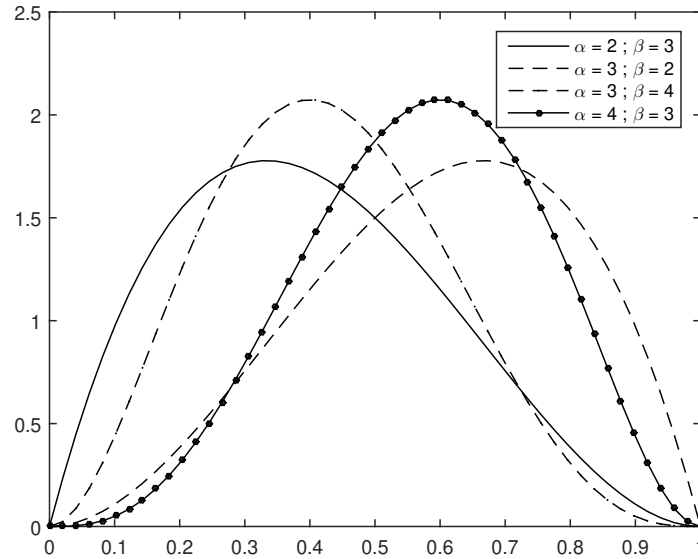


FIGURE 2.6 – Densité bêta pour différentes valeurs des paramètres α et β

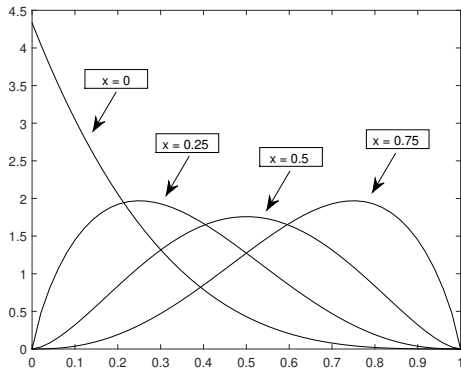


FIGURE 2.7 – Noyau bêta (première variante) pour $h = 0.3$ et x variable

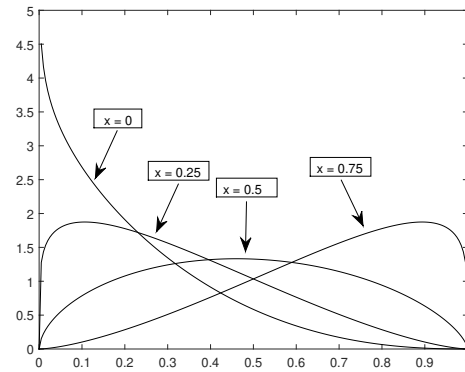


FIGURE 2.8 – Noyau bêta (deuxième variante) pour $h = 0.3$ et x variable

2.3 Mesure de l'erreur d'une estimation de densité à noyau asymétriques

Afin de juger de la qualité d'un estimateur de densité à noyaux asymétriques nous reprenons les mêmes critères adoptés dans le cas symétrique pour l'évaluation de la différence entre la densité réelle et la densité estimée, à savoir l'erreur moyenne quadratique MSE pour une mesure locale et l'erreur moyenne quadratique intégrée $MISE$ pour une mesure globale.

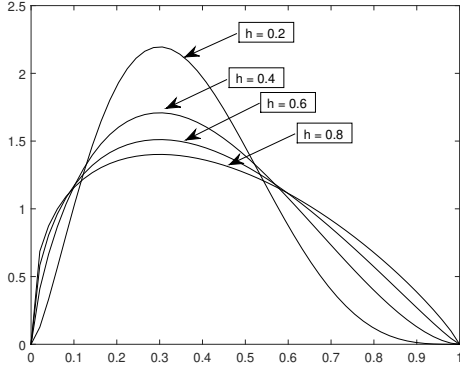


FIGURE 2.9 – Noyau bêta (première variante) pour $x = 0.3$ et h variable

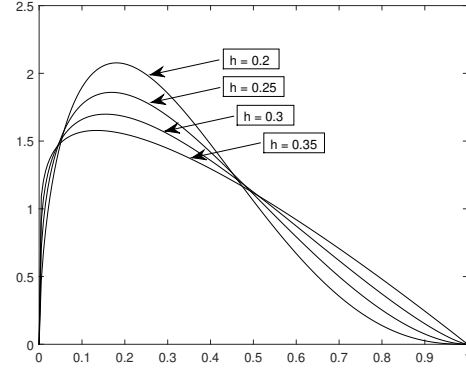


FIGURE 2.10 – Noyau bêta (deuxième variante) pour $x = 0.3$ et h variable

On rappelle que ces deux mesures peuvent être décomposées en termes de biais et de variance, néanmoins, malgré cette propriété attrayante et comme dans le cas des estimateurs à noyaux symétriques, une forme claire et exacte qui permette d'avoir une évaluation précise de l'erreur reste inaccessible. On recourt alors à des développements asymptotiques pour obtenir des approximations dont la forme renseigne quant à l'effet que peuvent avoir certains paramètres sur la qualité de l'estimation.

Comme il a été précisé plus haut, les noyaux asymétriques peuvent avoir des structures très différentes. Ceci conduit à ce que les biais et variance de chaque noyau soient décrits de manière différente, chose qui contraste clairement avec le cas symétrique où chaque propriété est décrite par une formule unique.

Nous présentons donc des développements qui mènent à des approximations des biais et variance pour les estimateurs à base des premières variantes des noyaux gamma et bêta. Pour les estimateurs à base des deuxièmes variantes, on peut adopter ces mêmes développements, les formules de la variance qui en résultent sont pratiquement similaires à ce qui sera présenté dans ce qui suit, Les formules du biais quand à elles seront données dans l'annexe (A2.1).

2.3.1 Estimateur à noyau gamma $K_{x/h+1,h}^G$

Biais :

Avant de calculer le biais, observons que

$$\begin{aligned}
 E[\hat{f}_n(x)] &= E[K_{x/h+1,h}^G(X_1)] \\
 &= \int K_{x/h+1,h}^G(t) f(t) dt \\
 &= E[f(\mathcal{K}_{x/h+1,h}^G)]
 \end{aligned} \tag{2.6}$$

et rappelons que $\mathcal{K}_{x/h+1,h}^G$ est une variable aléatoire de densité $K_{x/h+1,h}^G$.

En utilisant l'équation (2.6), on peut écrire

$$\text{Biais} [\hat{f}_n(x)] = E \left[f \left(\mathcal{K}_{x/h+1,h}^G \right) \right] - f(x) \quad (2.7)$$

Un développement de Taylor à l'ordre 2 au voisinage de $E \left[\mathcal{K}_{x/h+1,h}^G \right]$ donne

$$\begin{aligned} f \left(\mathcal{K}_{x/h+1,h}^G \right) &\approx f \left(E \left[\mathcal{K}_{x/h+1,h}^G \right] \right) + \left(\mathcal{K}_{x/h+1,h}^G - E \left[\mathcal{K}_{x/h+1,h}^G \right] \right) f' \left(E \left[\mathcal{K}_{x/h+1,h}^G \right] \right) \\ &\quad + \frac{1}{2} \left(\mathcal{K}_{x/h+1,h}^G - E \left[\mathcal{K}_{x/h+1,h}^G \right] \right)^2 f'' \left(E \left[\mathcal{K}_{x/h+1,h}^G \right] \right) \end{aligned}$$

on en déduit alors que

$$E \left[f \left(\mathcal{K}_{x/h+1,h}^G \right) \right] \approx f \left(E \left[\mathcal{K}_{x/h+1,h}^G \right] \right) + \frac{1}{2} \text{Var} \left(\mathcal{K}_{x/h+1,h}^G \right) f'' \left(E \left[\mathcal{K}_{x/h+1,h}^G \right] \right) \quad (2.8)$$

et comme on a $E \left[\mathcal{K}_{x/h+1,h}^G \right] = x+h$ et $\text{Var} \left[\mathcal{K}_{x/h+1,h}^G \right] = xh+h^2$, on peut donner l'approximation suivante du biais :

$$\text{Biais} [\hat{f}_n(x)] \approx hf'(x) + \frac{1}{2}xhf''(x) \quad (2.9)$$

Variance :

La variance peut être calculée à partir de la forme suivante :

$$\begin{aligned} \text{Var} [\hat{f}_n(x)] &= \frac{1}{n} \text{Var} \left[K_{x/h+1,h}^G(X_1) \right] \\ &= \frac{1}{n} E \left[\left(K_{x/h+1,h}^G(X_1) \right)^2 \right] - \frac{1}{n} E^2 \left[K_{x/h+1,h}^G(X_1) \right] \\ &= \frac{1}{n} \int \left[K_{x/h+1,h}^G(t) \right]^2 f(t) dt - \frac{1}{n} E^2 \left[K_{x/h+1,h}^G(X_1) \right] \end{aligned}$$

Posons $I = \frac{1}{n} \int \left[K_{x/h+1,h}^G(t) \right]^2 f(t) dt$ et $J = \frac{1}{n} E^2 \left[K_{x/h+1,h}^G(X_1) \right]$.

Un développement de Taylor pour f à l'ordre 2 au voisinage de x donne

$$\begin{aligned} I &\approx \frac{1}{n} \int \left[K_{x/h+1,h}^G(t) \right]^2 f(x) dt \\ &\quad + \frac{1}{n} \int \left[K_{x/h+1,h}^G(t) \right]^2 \left[(t-x)f'(x) + \frac{1}{2}(t-x)^2 f''(x) \right] dt \end{aligned}$$

de plus, si f est bornée et sous l'hypothèse que

$$\int_{\mathbb{S}^*} \left[K_{x/h+1,h}^G(t) \right]^2 dt < \frac{H(x)}{hr} ; \mathbb{S}^* = \mathbb{S} \cap \mathbb{S}_{x/h+1,h}$$

avec $0 < H(x) < \infty$ on trouve que

$$I \approx \frac{1}{n} f(x) \int \left[K_{x/h+1,h}^G(t) \right]^2 dt$$

le calcul de $\int \left[K_{x/h+1,h}^G(t) \right]^2 dt$ donne

$$\begin{aligned} \int \left[K_{x/h+1,h}^G(t) \right]^2 dt &= \int \frac{t^{2x/h} e^{-2t/h}}{h^{2x/h+2} \Gamma^2(x/h+1)} dt \\ &= \frac{h^{-1}}{2^{2x/h+1} \Gamma^2(x/h+1)} \int \left(\frac{2t}{h} \right)^{2x/h} e^{-2t/h} d\left(\frac{2t}{h} \right) \\ &= h^{-1} \frac{\Gamma(2x/h+1)}{\Gamma^2(x/h+1) 2^{2x/h+1}} \end{aligned}$$

Le deuxième terme J est dominé quant à lui par $\frac{1}{n}$, en effet

$$\begin{aligned} J &= \frac{1}{n} E^2 \left[K_{x/h+1,h}^G(X_1) \right] \\ &\approx \frac{1}{n} E^2 \left[\hat{f}_n(x) \right] \\ &\approx \frac{1}{n} f^2(x) \\ J &\approx O\left(\frac{1}{n}\right) \end{aligned}$$

On a donc

$$Var \left[\hat{f}_n(x) \right] \approx \frac{1}{n} G_h f(x)$$

avec

$$G_h = \frac{h^{-1} \Gamma(2x/h+1)}{2^{2x/h+1} \Gamma^2(x/h+1)}$$

Considérons maintenant la fonction R , croissante et majorée par 1 (Brown et Chen [21])

$$R(z) = \frac{1}{\Gamma(z+1)} \left(\frac{z}{e} \right)^z \sqrt{2\pi z}; \quad z \geq 0 \quad (2.10)$$

En exprimant G_h en termes de $R^2(x/h)$ et $R(2x/h)$ on peut avoir

$$G_h = \frac{h^{-1/2} x^{-1/2} R^2(x/h)}{2\sqrt{\pi} R(2x/h)} \leq \frac{h^{-1/2} x^{-1/2}}{2\sqrt{\pi}}$$

l'inégalité est justifiée par le fait que $R^2(z)/R(2z) < 1$.

L'approximation de la variance peut alors prendre deux formes différentes selon que $x/h \rightarrow \infty$ ou $x/h \rightarrow k$ où k est une constante positive.

$$\text{Var} \left[\hat{f}_n(x) \right] \approx \begin{cases} \frac{h^{-1/2} x^{-1/2}}{2n\sqrt{\pi}} f(x) & \text{si } x/h \rightarrow \infty \\ \frac{h^{-1} n^{-1} \Gamma(2x/h+1)}{2^{2x/h+1} \Gamma^2(x/h+1)} f(x) & \text{si } x/h \rightarrow k \end{cases} \quad (2.11)$$

L'effet de la variance et de la variance intégrée au bord étant négligeable (Chen [24]; Bouezmarni, EL Ghouch et Mesfioui [9]), les approximations du *MSE* et du *MISE* prennent en compte uniquement l'expression de la variance dans la région intérieure. Donc en remplaçant par les valeurs convenables dans les équations (1.9) et (1.13) on retrouve les formes asymptotiques *AMSE* et *AMISE*

$$\text{AMSE}_x \left[\hat{f}_n \right] = \frac{h^{-1/2} x^{-1/2}}{2n\sqrt{\pi}} f(x) + h^2 \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 \quad (2.12)$$

$$\text{AMISE} \left[\hat{f}_n \right] = \frac{h^{-1/2}}{2n\sqrt{\pi}} \int x^{-1/2} f(x) dx + h^2 \int \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx \quad (2.13)$$

2.3.2 Estimateur à noyau bêta $K_{x/h+1, (1-x)/h+1}^B$

L'obtention d'approximations pour le biais et la variance de l'estimateur à noyau bêta se fait en gardant les mêmes démarches suivie pour le cas du noyau gamma. Toutefois, les spécificités de chaque noyau conduisent à des résultats différents.

Biais :

En partant de l'équation (2.7) et en prenant en considération que (Johnson, Kotz et Balakrishnan [64])

$$E \left[\mathcal{K}_{x/h+1, (1-x)/h+1}^B \right] \approx x + h(1-2x)$$

$$\text{Var} \left[\mathcal{K}_{x/h+1, (1-x)/h+1}^B \right] \approx hx(1-x)$$

un développement de Taylor de $f \left(\mathcal{K}_{x/h+1, (1-x)/h+1}^B \right)$, mènera directement à une approximation du biais

$$\text{Biais} \left[\hat{f}_n(x) \right] \approx h(1-2x) f'(x) + \frac{1}{2} hx(1-x) f''(x) \quad (2.14)$$

Variance :

En procédant de la même manière que pour l'estimateur à base de noyau gamma, on se retrouve avec

$$\begin{aligned}
\text{Var} \left[\hat{f}_n(x) \right] &\approx \frac{1}{n} f(x) \int \left[K_{x/h+1, (1-x)/h+1}^B(t) \right]^2 dt \\
&\approx \frac{1}{n} f(x) \frac{1}{B^2(x/h+1, (1-x)/h+1)} \int t^{2x/h} (1-t)^{2(1-x)/h} dt \\
&\approx \frac{1}{n} f(x) \frac{B(2x/h+1, 2(1-x)/h+1)}{B^2(x/h+1, (1-x)/h+1)}
\end{aligned}$$

Posons

$$B_h = \frac{B(2x/h+1, 2(1-x)/h+1)}{B^2(x/h+1, (1-x)/h+1)}$$

L'expression de B_h sous forme de fonctions gamma donne

$$B_h = \frac{\Gamma(2x/h+1) \Gamma(2(1-x)/h+1) \Gamma^2(1/h+2)}{\Gamma(2/h+2) \Gamma^2(x/h+1) \Gamma^2((1-x)/h+1)}$$

Cette même formule exprimée en termes de la fonction R (équation (2.10)) donne cette fois-ci

$$B_h = \frac{h(h^{-1}+1)^{3/2} \left[1 + (2h^{-1} + 5/2)^{-1} \right]^{2/h+3}}{2e\sqrt{\pi}\sqrt{x(1-x)}} \rho_{x,h}$$

avec

$$\rho_{x,h} = \frac{R^2(x/h) R^2((1-x)/h) R(2/h+1)}{R(2x/h) R(2(1-x)/h) R^2(1/h+1)}$$

En faisant tendre h vers 0 on trouve que

$$\rho_{x,h} \longrightarrow 1 \text{ et } \left[1 + (2/h + 5/2)^{-1} \right]^{2/h+3} \longrightarrow e$$

on aura alors

$$B_h \approx \begin{cases} \frac{h^{-1/2}}{2\sqrt{\pi}\sqrt{x(1-x)}} & \text{si } x/h \longrightarrow \infty \text{ et } (1-x)/h \longrightarrow \infty \\ \frac{\Gamma(2x/h+1)}{2^{2x/h+1}\Gamma^2(x/h+1)} h^{-1} & \text{si } x/h \longrightarrow k \text{ ou } (1-x)/h \longrightarrow k \end{cases}$$

D'où l'approximation de la variance

$$\text{Var} \left[\hat{f}_n(x) \right] \approx \begin{cases} \frac{n^{-1}h^{-1/2}}{2\sqrt{\pi}\sqrt{x(1-x)}} f(x) & \text{si } x/h \longrightarrow \infty \text{ et } (1-x)/h \longrightarrow \infty \\ \frac{\Gamma(2x/h+1)}{2^{2x/h+1}\Gamma^2(x/h+1)} n^{-1}h^{-1} f(x) & \text{si } x/h \longrightarrow k \text{ ou } (1-x)/h \longrightarrow k \end{cases} \quad (2.15)$$

Donc, en tenant encore une fois compte du fait que la variance est négligeable aux bords, des approximations $AMSE$ et $AMISE$ de l'estimateur peuvent être données par

$$AMSE_x(\hat{f}_n) \approx \frac{n^{-1}h^{-1/2}}{2\sqrt{\pi}\sqrt{x(1-x)}}f(x) + h^2 \left[(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right]^2 \quad (2.16)$$

$$AMISE(\hat{f}_n) \approx \frac{n^{-1}h^{-1/2}}{2\sqrt{\pi}} \int \frac{f(x)}{\sqrt{x(1-x)}} dx + h^2 \int \left[(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right]^2 dx \quad (2.17)$$

2.4 Sélection du paramètre de lissage

L'importance dont jouit le paramètre de lissage n'est plus à démontrer. En effet, de même que pour le cas des noyaux symétriques, le choix du paramètre de lissage dans le cas asymétrique revêt un intérêt majeur quant à la qualité et par conséquent l'utilité de l'estimation conduite. Rappelons néanmoins qu'il existe une différence notable concernant l'utilisation du paramètre de lissage par chacun de ces types de noyaux. En fait, malgré que l'accent soit mis sur l'estimation d'un paramètre unique, l'effet constant qu'il exerce dans une estimation à noyaux symétriques ne s'étend pas aux noyaux asymétriques puisque la nature de ces derniers leur offre l'avantage de pouvoir appliquer un lissage adaptif à chaque point où l'estimation est effectuée. Toutefois, cette nature a l'inconvénient d'engendrer un certain nombre de difficultés qui empêchent le développement des méthodes générales de sélection et compliquent l'ajustement de méthodes classiques au cas asymétrique, chose qui impose un traitement au cas par cas.

Par exemple, l'approche d'insertion élémentaire proposée par Silverman [96] ne peut être reproduite telle quelle, compte tenu du fait que chaque estimateur, dépendant du noyau utilisé, aura un critère d'erreur différent et donc un minimiseur différent aussi. Ainsi pour des estimateurs à base de noyaux gamma $K_{x/h+1,h}^G$ et bêta $K_{x/h+1,(1-x)/h+1}^B$ les paramètres de lissage qui minimisent les critères données par les équations (2.13) et (2.17) sont respectivement

$$h_{(G)}^{opt} = \left(\frac{1}{8\sqrt{\pi}} \right)^{2/5} \frac{\left[\int x^{-1/2} f(x) dx \right]^{2/5}}{\left[\int \left(f'(x) + \frac{1}{2}x f''(x) \right)^2 dx \right]^{2/5}} n^{-2/5} \quad (2.18)$$

$$h_{(B)}^{opt} = \left(\frac{1}{8\sqrt{\pi}} \right)^{2/5} \frac{\left[\int (x(1-x))^{-1/2} f(x) dx \right]^{2/5}}{\left[\int \left((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right)^2 dx \right]^{2/5}} n^{-2/5} \quad (2.19)$$

Donc, comme pour le cas de l'estimation à base de noyaux symétriques, la présence de la densité inconnue et de ses dérivées dans la formule du paramètre optimal empêche un calcul direct de ce paramètre. Ceci, renvoie alors vers la règle pratique vue au chapitre précédent, Cependant la densité

de référence, ne faisant l'objet d'aucun accord, n'est pas fixée, contrairement au cas classique où l'on se rapporte dans la plupart du temps à la densité normale.

La qualité de cette méthode n'étant pas toujours au rendez-vous, plusieurs auteurs préfèrent s'en tenir aux méthodes de validation croisée par exclusion et par m-block (Bouezmarni et Rombouts [12]).

La fonction de validation croisée par exclusion reprend exactement la forme donnée dans le chapitre précédent avec en toute évidence la différence au niveau de l'estimateur de densité. La fonction de validation par m-block, quand à elle, utilise un estimateur de densité qui exclut $2m + 1$ observations qui comprennent le point X_i et m points de chaque cotés de X_i .

2.5 Quelques résultats asymptotiques

De manière semblable au chapitre précédent, nous exposons dans cette partie quelques résultats de convergence et d'ordre de convergence selon le critère *AMISE*. Ces résultats concernent principalement la consistance des estimateurs de densité à base de noyaux gamma et bêta et plus précisément la convergence en probabilité et la convergence presque-sûre de ces estimateurs vers la densité réelle.

Noyau gamma : (Bouezmarni et Scaillet [13]; Bouezmarni et Rombouts [11])

Théorème 2.1 (Convergence en probabilité) *Soient, f une fonction de densité de probabilité continue et bornée sur $[0, +\infty[$ et \hat{f}_n un estimateur de densité à base de noyau gamma. Pour I un compact de $[0, +\infty[$ et sous les conditions*

$$\lim_{n \rightarrow \infty} h = 0 ; \lim_{n \rightarrow \infty} nh^2 = \infty$$

on a

$$\sup_{x \in I} |\hat{f}_n(x) - f(x)| \xrightarrow{P} 0 \text{ quand } n \rightarrow \infty$$

Théorème 2.2 (Convergence presque sûre) *Soient, f une fonction de densité de probabilité continue et bornée sur $[0, +\infty[$ et \hat{f}_n un estimateur de densité à base de noyau gamma. Pour I un compact de $[0, +\infty[$ et sous les conditions*

$$\lim_{n \rightarrow \infty} h = 0 ; \lim_{n \rightarrow \infty} \frac{nh^2}{\log n} = \infty$$

on a

$$\sup_{x \in I} |\hat{f}_n(x) - f(x)| \xrightarrow{p.s.} 0 \text{ quand } n \rightarrow \infty$$

Théorème 2.3 (Convergence en probabilité dans L_1) *Soient, f une fonction de densité de probabilité sur $[0, +\infty[$ et \hat{f}_n un estimateur de densité à base de noyau gamma. Sous les conditions*

$$\lim_{n \rightarrow \infty} h = 0 ; \lim_{n \rightarrow \infty} nh^2 = \infty$$

on a

$$\int_0^\infty |\hat{f}_n(x) - f(x)| \xrightarrow{P} 0 \text{ quand } n \rightarrow \infty$$

Noyau bêta : (Bouezmarni et Rolin [10])

Théorème 2.4 (Convergence en probabilité) *Soient, f une fonction de densité de probabilité continue et bornée sur $[0, 1]$ et \hat{f}_n un estimateur de densité à base de noyau bêta. Sous les conditions*

$$\lim_{n \rightarrow \infty} h = 0 ; \lim_{n \rightarrow \infty} nh^2 = \infty$$

on a

$$\sup_{x \in [0,1]} |\hat{f}_n(x) - f(x)| \xrightarrow{P} 0 \text{ quand } n \rightarrow \infty$$

Théorème 2.5 (Convergence presque sûre) *Soient, f une fonction de densité de probabilité continue et bornée sur $[0, 1]$ et \hat{f}_n un estimateur de densité à base de noyau bêta. Sous les conditions*

$$\lim_{n \rightarrow \infty} h = 0 ; \lim_{n \rightarrow \infty} \frac{nh^2}{\log n} = \infty$$

on a

$$\sup_{x \in [0,1]} |\hat{f}_n(x) - f(x)| \xrightarrow{p.s.} 0 \text{ quand } n \rightarrow \infty$$

Ordre de convergence

Au sens du critère *AMISE*, les ordres de convergence de l'estimateur de densité à base des noyaux asymétriques gamma et bêta s'obtiennent de la même manière que pour les estimateurs à base de noyaux symétriques. Bien entendu, il est nécessaire de souligner que, contrairement à ces derniers, où les ordres de convergences retrouvés correspondent à toute la classe des estimateurs, dans le cas asymétriques ces ordres correspondent uniquement aux noyaux impliqués dans l'estimation.

On utilisera donc les formules des paramètres de lissage données par les équations (2.18) et (2.19) afin de trouver un critère *AMISE* optimal qui corresponde à chaque noyau en remplaçant dans les équations (2.13) et (2.17).

Pour le noyau gamma on trouve

$$AMISE_{(G)}^{opt} [\hat{f}_n] = 5 \left(\frac{1}{8\sqrt{\pi}} \right)^{4/5} \left[\int (f'(x) + \frac{x}{2} f''(x))^2 dx \right]^{1/5} \left[\int \frac{1}{\sqrt{x}} f(x) dx \right]^{4/5} n^{-4/5} \quad (2.20)$$

Alors que pour le noyau bêta on a

$$\begin{aligned}
 AMISE_{(B)}^{opt} [\hat{f}_n] &= 5 \left(\frac{1}{8\sqrt{\pi}} \right)^{4/5} \left[\int \left((1-2x) f'(x) + \frac{x}{2} (1-x) f''(x) \right)^2 dx \right]^{1/5} \\
 &\quad \times \left[\int \frac{1}{\sqrt{x(1-x)}} f(x) dx \right]^{4/5} n^{-4/5} \quad (2.21)
 \end{aligned}$$

Ainsi, on remarque d'abord que les paramètres de lissage optimaux donnés par les équations (2.18) et (2.19) atteignent un ordre de $n^{-2/5}$ alors que ces mêmes paramètres dans le cas symétrique ont un ordre de $n^{-1/5}$. Ensuite, en termes du critère *AMISE* les estimateurs à noyaux asymétriques enregistrent un ordre de $n^{-4/5}$ qui correspond à l'ordre de convergence optimal, selon le sens de Stone [98], atteint également lors de l'utilisation des noyaux symétriques.

Chapitre 3

Régression à noyau

Etant donnée la disponibilité d'une quantité suffisante d'observations d'un certain nombre de variables, la régression peut constituer un outil statistique puissant pour modéliser la relation entre l'une de ces variables, appelé variable de réponse, et les autres variables appelées variables d'entrées ou prédicteurs. Cette modélisation permettra, entre autres, de faire des prévisions, construire des tests et des intervalles de confiance...

Le cas de figure le plus simple et le mieux compris de ce type de régression est la régression linéaire laquelle établi une simple relation linéaire entre la variable de réponse et les prédicteurs. Toutefois, malgré tous les développements et les mises au point que ces modèles ont pu connaître, il s'est avéré, à plusieurs reprises, qu'ils sont incapables de décrire certaines relations du fait qu'il existe une non linéarité intrinsèque aux données. Ce dernier point représente une difficulté majeure rencontrée lors de l'utilisation des modèles de régression paramétriques où la fonction recherchée se caractérise par une certaine rigidité imposée par la forme supposée de la relation qui lie la variable de réponse aux prédicteurs. Une régression non paramétrique permettra alors de contourner cette rigidité en relaxant un certain nombre d'hypothèses restrictives en rapport avec la relation variable de réponse/prédicteur et de proposer une alternative générale au modèle paramétrique, qui soit de la forme

$$y_i = m(x_i) + \epsilon_i ; i = 1, 2, \dots, n \quad (3.1)$$

où la fonction m n'obéit à aucune forme prédéfinie.

Il est vrai qu'en général, l'alternative non-paramétrique, offrant une flexibilité plutôt élevée, est considérée comme étant une rivale de la régression paramétrique, ceci n'est cependant pas toujours vrai, car dans plusieurs cas la régression non paramétrique constitue une première étape dans la réduction de la complexité des modèles paramétriques. Cette idée de rivalité vient également du fait que la souplesse de l'approche non paramétrique n'implique pas de complications superflues au niveau des calculs ou de l'analyse d'autant plus que l'approche qui nous intéresse, à savoir la régression à noyau, a d'une part l'avantage d'être beaucoup plus intuitive que les autres méthodes et d'autre part, a le même contexte théorique et nécessite les mêmes techniques utilisées pour l'estimation de densité à noyau, chose qui la rend beaucoup plus abordable et accessible.

L'estimateur de régression à noyau le plus commun est l'estimateur de Nadaraya-Watson (Nadaraya [75]; Watson [106]), sa popularité est principalement due à son aspect simple et élémentaire puisqu'en effet, il correspond à l'ajustement de constantes locales pondérées par un noyau. Cette méthode souffre pourtant d'un problème de variance qui conduit souvent à un sous lissage.

Par ailleurs, le fait qu'une constante corresponde à un polynôme de degré 0, évoque la possibilité d'une généralisation de la méthode en ajustant, toujours à l'aide de moindres carrés pondérés, des polynômes de degré $p > 0$. L'estimateur est alors appelé estimateur polynomial local, toutefois le cas particulier de $p = 1$ est appelé estimateur linéaire local. Ces estimateurs présentent d'excellentes propriétés asymptotiques et un comportement aux bords très correct (Fan [40]).

Les techniques citées ci-dessus pour l'estimation de la fonction m (équation (3.1)) peuvent être étudiées dans deux contextes différents, le premier est celui d'un dispositif expérimental à effet fixe où les x_i sont ordonnés et non-aléatoires tandis que le deuxième est un dispositif expérimental à effet aléatoire qui correspond à des X_i strictement aléatoire.

Ce chapitre commence par l'introduction de l'estimateur de régression le plus élémentaire qui est celui de Nadaraya-Watson, ensuite on passe à une généralisation de cet estimateur à travers le passage aux estimateurs par polynôme locaux de la fonction de régression et des estimateurs des dérivées de cette dernière. Finalement, on aborde les estimateurs de régression à base de noyaux asymétriques. Par ailleurs, pour chacune des méthodes présentées, la mesure de l'erreur et la sélection du paramètre de lissage seront traitées.

3.1 Estimateur de Nadaraya-Watson

En reprenant le modèle donné par l'équation (3.1) et en considérant un couple de variables aléatoires réelles (X, Y) on peut définir la fonction de régression m par l'espérance conditionnelle de Y sachant $X = x$

$$\begin{aligned} m(x) &= E[Y|X = x] = \int y f_{y|x}(y|x) dy \\ &= \int y \frac{f_{X,Y}(x,y)}{f_X(x)} dy \end{aligned} \quad (3.2)$$

avec $f_{y|x}$ la densité conditionnelle de Y sachant X , $f_{X,Y}$ la densité jointe du couple (X, Y) et f_X la densité marginale de X .

Il faudra également noter que l'espérance du résidu conditionnel est nulle $E[\epsilon|X = x] = 0$ et sa variance n'est pas nécessairement constante $Var[\epsilon|X = x] = \sigma^2(x)$.

Remarque 3.1 Si le modèle général adopté est de la forme

$$Y_i = m(X_i) + \nu^{1/2}(X_i) \epsilon_i ; i = 1, 2, \dots, n$$

on a alors $E[\epsilon] = 0$, $Var[\epsilon] = 1$ et $E[Y|X = x] = m(x)$, $Var[Y|X = x] = \nu(x)$.

Le problème d'une estimation non paramétrique de m peut-être approché à l'aide d'estimateurs à noyaux des densités inconnues dans l'équation (3.2). Ceux-ci permettront de dériver des estimateurs qui appartiennent à la classe des estimateurs linéaires définie ci-dessous et qui se présentent principalement sous forme de moyennes locales pondérées.

Définition 3.1 On appelle \hat{m} un estimateur linéaire de la fonction de régression non paramétrique m , si on peut écrire

$$\hat{m}(x) = \sum_{i=1}^n Y_i W_i(x)$$

où la fonction de poids W_i est indépendante des observations Y_i .

Remarque 3.2 Généralement, les poids d'un estimateur linéaire de régression satisfont l'égalité

$$\sum_{i=1}^n W_i = 1$$

Donc, en considérant pour les densités $f_{X,Y}$ et f_X des estimateurs à noyaux symétriques tels qu'ils sont définis dans le premier chapitre

$$\hat{f}_{X,Y}(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^n K_x\left(\frac{x-X_i}{h_x}\right) K_y\left(\frac{y-Y_i}{h_y}\right)$$

$$\hat{f}_X(x) = \frac{1}{nh_x} \sum_{i=1}^n K_x\left(\frac{x-X_i}{h_x}\right)$$

et en remplaçant dans l'équation (3.2), on retrouve, à l'aide d'un simple calcul, l'estimateur suivant pour la fonction de régression m :

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} \quad (3.3)$$

on appelle cet estimateur, Estimateur de Nadaraya-Watson et on le note \hat{m}_{NW} . On peut également écrire

$$\hat{m}_{NW}(x) = \sum_{i=1}^n Y_i W_i(x)$$

avec

$$W_i(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h}\right)}$$

Remarque 3.3 L'effet que peuvent exercer le noyau K et le paramètre de lissage h sur l'estimation de la densité est conservé lors de l'estimation de la fonction de régression. On rappelle que le noyau contrôle la manière dont la masse de probabilité est assigné au voisinage de x , tandis que le paramètre de lissage détermine la taille de ce voisinage.

Qu'il s'agisse d'un dispositif expérimental aléatoire ou fixe, l'estimateur de Nadaraya-Watson constitue toujours un choix adéquat. N'empêche, il existe des estimateurs alternatifs qui sont dédiés uniquement au dispositif expérimental fixe. Parmi les plus populaire, l'estimateur de Priestley-Chao

$$\hat{m}_{PC}(x) = h^{-1} \sum_{i=1}^n (X_i - X_{i-1}) K\left(\frac{x - X_i}{h}\right) Y_i \quad (3.4)$$

Et celui de Gasser-Müller

$$\hat{m}_{GM}(x) = h^{-1} \sum_{i=1}^n \left[\int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h}\right) du \right] Y_i \quad (3.5)$$

avec $s_{i-1} = (x_{i-1} + x_i) / 2$.

Mesure de l'erreur pour l'estimateur de Nadaraya-Watson

Les critères de mesure des erreurs locale et globale adoptés pour le cas de l'estimateur de Nadaraya-Watson sont identiques à ceux utilisés pour les estimateurs de densités. La décomposition de cette mesure en biais-variance est encore une fois utilisée comme un point de départ pour obtenir des expressions pour l'erreur moyenne quadratique et l'erreur moyenne quadratique intégrée. De manière similaire à l'estimation de densité, le développement de Taylor est adopté afin de retrouver des formes asymptotiques qui permettent d'exprimer de manière claire la relation entre le noyau, le paramètre de lissage et l'erreur.

Par ailleurs, l'estimateur de Nadaraya-Watson peut être exprimé sous la forme

$$\hat{m}_{NW}(x) = \frac{\hat{r}_n(x)}{\hat{f}_n(x)} \quad (3.6)$$

avec

$$\hat{r}_n(x) = \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)$$

Cette écriture permet de faciliter les développements, menant au calcul du biais et de la variance.

Biais :

Énonçons d'abord le théorème (3.1) (Nadaraya [76]) dont une démonstration est donnée dans l'annexe (A3.1).

Théorème 3.1 *soit le terme*

$$\hat{E}[\hat{m}_{NW}(x)] = \frac{E[\hat{r}_n(x)]}{E[\hat{f}_n(x)]}$$

1. Si $|Y| \leq c \leq \infty$ et $nh \rightarrow \infty$ on a alors

$$E[\hat{m}_{NW}(x)] = \hat{E}[\hat{m}_{NW}(x)] + O\left((nh)^{-1}\right)$$

2. Si $E[Y^2] \leq \infty$ et $nh^2 \rightarrow \infty$ alors

$$E[\hat{m}_{NW}(x)] = \hat{E}[\hat{m}_{MW}(x)] + O\left(\left(n^{1/2}h\right)^{-1}\right)$$

Dès lors, on a le biais qui peut être approximé par

$$\text{Biais}[\hat{m}_{NW}(x)] \approx \hat{E}[\hat{m}_{NW}(x)] - m(x) \quad (3.7)$$

$$\text{Biais}[\hat{m}_{NW}(x)] \approx \frac{E[\hat{r}_n(x)]}{E[\hat{f}_n(x)]} - m(x)$$

$$\begin{aligned} \text{Biais}[\hat{m}_{NW}(x)] &\approx \frac{1}{E\left[K\left(\frac{x-X}{h}\right)\right]} \left[\int \frac{1}{h} K\left(\frac{x-t}{h}\right) r(t) dt \right. \\ &\quad \left. - r(x) + r(x) - m(x) \int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t) dt \right] \\ &\approx \frac{h^2}{2} \left(\frac{1}{f(x)} \right) [r''(x) - m(x) f''(x)] \mu_2(K) \\ &\approx \frac{h^2}{2} \left[m''(x) + \frac{2m'(x) f'(x)}{f(x)} \right] \mu_2(K) \end{aligned} \quad (3.8)$$

Variance :

Le calcul de la variance se fera également à travers l'utilisation de la forme donnée par l'équation (3.6).

D'après l'équation (1.15) une approximation asymptotique de la variance de \hat{f}_n est donnée par

$$\text{Var}[\hat{f}_n(x)] \approx \frac{1}{nh} f(x) R(K)$$

un calcul similaire à celui menant à l'expression de $\text{Var}[\hat{f}_n(x)]$ donne

$$\text{Var}[\hat{r}_n(x)] \approx \frac{1}{nh} s(x) R(K)$$

avec $s(x) = \int y^2 f_{X,Y}(x,y) dy$.

On trouve aussi que

$$E\left\{ \left[\hat{f}_n(x) - E[\hat{f}_n(x)] \right] \left[\hat{r}_n(x) - E[\hat{r}_n(x)] \right] \right\} \approx \frac{1}{nh} r(x) R(K)$$

De plus, soit

$$\Sigma \approx \frac{1}{nh} \begin{pmatrix} f(x) & r(x) \\ r(x) & s(x) \end{pmatrix} R(K)$$

la matrice de variance-covariance du vecteur $V_n(x) = \begin{pmatrix} \hat{f}_n(x) & \hat{r}_n(x) \end{pmatrix}^T$.

Il s'en suit que

$$Var[\hat{m}_{NW}(x)] \approx \frac{1}{nh} \left[\frac{s(x)}{(f(x))^2} - \frac{(r(x))^2}{(f(x))^3} \right] R(K) \quad (3.9)$$

Les critères d'erreurs asymptotiques seront donc donnés par

$$AMSE_x[\hat{m}_{NW}] = \frac{h^4}{4} \left[m''(x) + \frac{2m'(x)f'(x)}{f(x)} \right]^2 \mu_2^2(K) + \frac{1}{nh} \left[\frac{s(x)}{(f(x))^2} - \frac{(r(x))^2}{(f(x))^3} \right] R(K) \quad (3.10)$$

$$AMISE[\hat{m}_{NW}] = \frac{h^4}{4} \int \left[m''(x) + \frac{2m'(x)f'(x)}{f(x)} \right]^2 \mu_2^2(K) dx + \frac{1}{nh} \int \left[\frac{s(x)}{(f(x))^2} - \frac{(r(x))^2}{(f(x))^3} \right] R(K) dx \quad (3.11)$$

Critères d'erreur pour un estimateur N-W à base d'un noyau d'ordre q

Les résultats précédents supposent que le noyau utilisé est un noyau d'ordre 2 selon la définition (1.3). Dans ce qui suit et toujours selon la même définition le noyau utilisé est supposé être un noyau d'ordre q quelconque.

De même que pour l'estimation de densité, l'expression de la variance reste inchangée, quant au biais le changement figure au niveau du calcul des espérances $E[\hat{r}_n(x)]$ et $E[\hat{f}_n(x)]$ pour lesquelles on trouve

$$E[\hat{r}_n(x)] \approx r(x) + \frac{h^q}{q!} r^{(q)}(x) \mu_q(K)$$

$$E[\hat{f}_n(x)] \approx f(x) + \frac{h^q}{q!} f^{(q)}(x) \mu_q(K)$$

Ainsi, en suivant des développements identiques au cas des noyaux d'ordre 2 on trouve

$$Biais[\hat{m}_{NW}(x)] \approx \frac{h^q}{q!} \left[m^{(q)}(x) + q \frac{m^{(q-1)}(x)f'(x)}{f(x)} \right] \mu_q(K)$$

et par conséquent les critères $AMSE$ et $AMISE$ prennent la forme

$$AMSE_x [\hat{m}_{NW}] = \frac{h^{2q}}{(q!)^2} \left[m^{(q)}(x) + q \frac{m^{(q-1)}(x) f'(x)}{f(x)} \right]^2 \mu_q^2(K) + \frac{1}{nh} \left[\frac{s(x)}{(f(x))^2} - \frac{(r(x))^2}{(f(x))^3} \right] R(K) \quad (3.12)$$

$$AMISE[\hat{m}_{NW}] = \frac{h^{2q}}{(q!)^2} \int \left[m^{(q)}(x) + q \frac{m^{(q-1)}(x) f'(x)}{f(x)} \right]^2 \mu_q^2(K) dx + \frac{1}{nh} \int \left[\frac{s(x)}{(f(x))^2} - \frac{(r(x))^2}{(f(x))^3} \right] R(K) dx \quad (3.13)$$

3.2 Estimateur par polynômes locaux

Cette technique (figure (3.1) et (3.2)) repose sur le fait d'ajuster pour chaque point x de l'estimation un polynôme local à l'aide de moindres carrés pondérés où le poids assigné pour chaque Y_i correspond à $K_h(x - X_i) = (1/h) K((x - X_i)/h)$ où K est un noyau symétrique d'un ordre quelconque. Cette pondération a comme effet d'augmenter l'impact des observations proches du point x . La quantité de cet impact, comme c'est le cas pour l'estimation de densité, est réglée par le paramètre de lissage h .

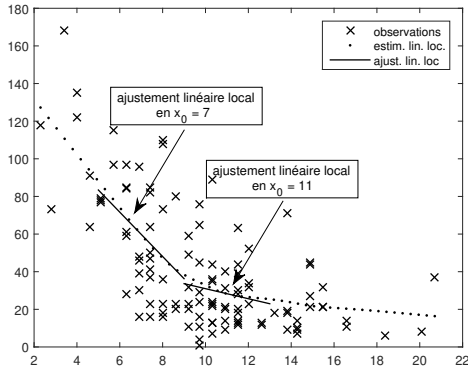


FIGURE 3.1 – Lissage par estimateur linéaire local

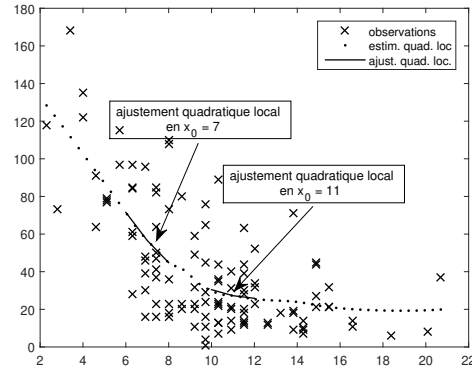


FIGURE 3.2 – Lissage par estimateur à polynômes locaux (polynômes de degré 2)

En effet, une valeur de h qui s'approche excessivement de 0 renforce la dépendance de l'ajustement aux observations, l'estimation commence alors à s'apparenter à une interpolation. Dans le cas contraire, lorsque la valeur de h est élevée, l'effet est que les observations tendent à avoir une même influence sur l'estimation, chose qui conduit à ce que cette dernière soit assimilée à une simple régression linéaire.

3.2.1 Expression de l'estimateur

La forme donnée par l'équation (3.6) pour l'estimateur de Nadaraya-Watson implique que cet estimateur constitue une solution à un problème des moindres carrés pondérés, puisqu'il minimise

$$\sum_{i=1}^n (Y_i - \beta_0)^2 K_h(x - X_i)$$

Cet estimateur correspond alors à une approximation locale de la fonction m par une constante. En outre, comme il a été mentionné au début du chapitre, puisque la constante correspond à un polynôme de degré 0, Il est donc naturel de chercher à ajuster aux données des polynômes locaux d'un degré ≥ 1 afin d'améliorer l'estimation en lui accordant la capacité de mieux capturer les variations au niveau des régions qui peuvent présenter des courbures prononcées.

En considérant p le degré du polynôme local à ajuster, l'estimateur \hat{m} sera alors la composante $\hat{\beta}_0$ du vecteur $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$ qui minimise

$$\sum_{i=1}^n [Y_i - \beta_0 - \beta_1(x - X_i) - \dots - \beta_p(x - X_i)^p]^2 K_h(x - X_i)$$

Ceci découle du fait qu'un développement de Taylor de la fonction m autour d'un point a donne

$$m(x) \approx \sum_{j=0}^p \frac{m^{(j)}(a)}{j!} (x - a)^j = \sum_{j=0}^p \beta_j (x - a)^j \quad (3.14)$$

Notons que la fonction m est supposée être dérivable autant de fois qu'il le faut.

La méthode des moindres carrés conduit ainsi au résultat suivant

$$\hat{\beta} = (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \mathbb{Y} \quad (3.15)$$

qui est solution de la formulation matricielle du problème de minimisation, donnée par

$$\min_{\beta \in \mathbb{R}^{p+1}} (\mathbb{Y} - \mathbb{X}_x \beta)^T W_x (\mathbb{Y} - \mathbb{X}_x \beta) \quad (3.16)$$

avec \mathbb{X}_x la matrice associée au dispositif expérimental

$$\mathbb{X}_x = \begin{pmatrix} 1 & x - X_1 & \dots & (x - X_1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x - X_n & \dots & (x - X_n)^p \end{pmatrix}$$

W_x la matrice diagonale des poids

$$W_x = \text{diag}\{K_h(x - X_1), \dots, K_h(x - X_n)\}$$

et \mathbb{Y} le vecteur de réponses

$$\mathbb{Y} = (Y_1, \dots, Y_n)^T$$

Ceci, étant donné bien évidemment, l'inversibilité de $\mathbb{X}_x^T W_x \mathbb{X}_x$. Du reste, il est maintenant possible de donner une définition simple de l'estimateur par polynôme locaux qui constitue une généralisation de l'estimateur de Nadaraya-Watson.

Définition 3.2 On appelle estimateur à polynômes locaux de $m(x)$, la quantité définie par

$$\hat{m}_{PL}(x) = e_1^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \mathbb{Y} \quad (3.17)$$

e_1^T étant un vecteur de taille $(p+1) \times 1$ dont le premier élément correspond à 1 et les autres éléments correspondent à 0.

Il faudra aussi noter qu'en plus de l'estimation de $m(x)$, le développement de Taylor donné par l'équation (3.14) permet d'estimer les dérivées $m^{(j)}(x)$ pour $j = 1, \dots, p$ de la régression $m(x)$ à partir de la solution β . En effet, il est clair que $j! \hat{\beta}_j$ est un estimateur de $m^{(j)}(x)$. Ainsi la formule que l'on retrouve dans la définition (3.2) peut être généralisée.

Définition 3.3 De manière générale, un estimateur par polynômes locaux de la dérivée $m^{(j)}(x)$ est défini par

$$\hat{m}_{PL}^{(j)}(x) = j! e_{j+1}^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \mathbb{Y} \quad (3.18)$$

e_{j+1}^T étant un vecteur de taille $(p+1) \times 1$ dont le $(j+1)$ -ième élément correspond à 1 et les autres éléments correspondent à 0.

3.2.2 Mesure de l'erreur

Encore une fois, dans cette partie seront donnés le biais et la variance d'une estimation par polynômes locaux, vu que ce sont les deux quantités principales qui participent au calcul de l'erreur moyenne quadratique et l'erreur moyenne quadratique intégrée qui continuent à être les mesures de l'erreur adoptées pour évaluer la qualité de l'estimation et le comportement de l'estimateur.

À partir de ce point, les formes des critères MSE et $MISE$ ne seront plus présentées, car elles peuvent être facilement retrouvées à l'aide des équations (1.9) et (1.13).

La solution donnée par l'équation (3.15) au problème de minimisation (équation (3.16)) implique directement que

$$\begin{aligned} E[\hat{\beta} | X_1, \dots, X_n] &= (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x m \\ &= \beta + (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x r \end{aligned}$$

avec $m = (m(X_1), \dots, m(X_n))^T$, $\beta = (m(x), \dots, m^{(p)}(x)/p!)^T$ et $r = m - \mathbb{X}_x \beta$ le vecteur des résidus.

On aura donc

$$Biais \left[\hat{\beta} | X_1, \dots, X_n \right] = (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x r \quad (3.19)$$

La variance quant à elle sera

$$Var \left[\hat{\beta} | X_1, \dots, X_n \right] = (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} (\mathbb{X}_x^T W_x V W_x \mathbb{X}_x) (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \quad (3.20)$$

avec $V = \text{diag} \left\{ Var [Y_1 | X_1], \dots, Var [Y_n | X_n] \right\}$.

Pour clarifier un peu plus l'idée, une illustration des calculs à partir de ces formules est donnée dans l'annexe (A3.2) dans le cas d'un dispositif expérimental fixe et pour $p = 1$ qui correspond à l'estimateur linéaire local. Des calculs détaillés davantage pour la dérivation des biais et variance asymptotiques sont donnés dans l'annexe (A3.3).

Bien que ces formules soient claires et simples, elles ne sont toutefois pas exploitables vu qu'elles dépendent de quantités inconnues, notamment le vecteur r et la matrice V . Des développements asymptotiques tels qu'adoptés dans Ruppert et Wand [90] permettent d'avoir les approximations des biais et variance conditionnels de l'estimateur $\hat{m}_{PL}(x)$ données par les équations (3.21), (3.22) et (3.23). Mais avant de présenter ces résultats, donnons d'abord quelques notations essentielles.

- $\mathbb{N}_p = ([\mu_{i+j-2}(K)])_{i,j} \in \mathcal{M}_{p+1}(\mathbb{R})$
- $\mathbb{M}_p(u)$ une matrice similaire à \mathbb{N}_p mais dont la première colonne est égale à $(1, u, \dots, u^p)^T$.
- K_p le noyau défini par

$$K_p(u) = \left[\frac{|\mathbb{M}_p(u)|}{|\mathbb{N}_p(u)|} \right] K(u)$$

Pour p impair, le biais conditionnel est

$$Biais [\hat{m}_{PL}(x) | X_1, \dots, X_n] \approx \frac{h^{p+1}}{(p+1)!} m^{(p+1)}(x) \mu_{p+1}(K_{(p)}) \quad (3.21)$$

tandis que pour p pair

$$Biais [\hat{m}_{PL}(x) | X_1, \dots, X_n] \approx h^{p+2} \left[\frac{m^{(p+1)}(x) f'(x)}{(p+1)! f(x)} + \frac{m^{(p+2)}(x)}{(p+2)!} \right] \mu_{p+2}(K_{(p)}) \quad (3.22)$$

Pour les deux cas, la variance est donnée par

$$Var [\hat{m}_{PL}(x) | X_1, \dots, X_n] \approx \left[\frac{(1/nh) R(K_{(p)})}{f(x)} \right] \sigma^2(x) \quad (3.23)$$

La différence paraît évidente entre les biais, lorsque l'ajustement est réalisé à l'aide de polynômes de degré pair ou impair. Dans le cas impair, l'équation (3.21) dépend de x uniquement à travers le terme $m^{(p+1)}(x)$ alors que pour le cas pair la dépendance est beaucoup plus complexe, chose qui rend l'interprétation plus difficile. Cette dépendance compliquée peut également se répercuter négativement sur la sélection du paramètre de lissage au cas où le critère à minimiser est l'erreur

moyenne quadratique intégrée c'est à dire le même critère utilisé pour les méthodes d'insertions dans le cas de l'estimation de densité.

Par ailleurs, la même logique qui permet l'estimation des dérivées d'ordre k de la régression conduit à une extension des approximations des biais et variance conditionnels de ces dérivées là. Dans ce cas, on définit

- $\mathbb{M}_{k,p}(u)$ une matrice similaire à \mathbb{N}_p dont la $(k+1)$ -ième colonne est égale à $(1, u, \dots, u^p)^T$
- $K_{(k,p)}(u)$ le noyau défini par

$$K_{(k,p)}(u) = k! \left[\frac{|\mathbb{M}_{k,p}(u)|}{|\mathbb{N}_p|} \right] K(u)$$

Pour le biais, on distingue encore une fois le cas de $(p-k)$ impair, pour lequel on trouve

$$\text{Biais} \left[\hat{m}_{PL}^{(k)}(x) | X_1, \dots, X_n \right] \approx \frac{1}{(p+1)!} \mu_{p+1}(K_{(k,p)}) m^{(p+1)}(x) h^{p-r+1} \quad (3.24)$$

et le cas de $(p-k)$ pair pour lequel on trouve

$$\begin{aligned} \text{Biais} \left[\hat{m}_{PL}^{(k)}(x) | X_1, \dots, X_n \right] \approx & \left[\frac{1}{(p+2)!} \mu_{p+2}(K_{(k,p)}) m^{(p+2)}(x) + \frac{1}{(p+1)!} \right. \\ & \left. \times (\mu_{p+2}(K_{(k,p)}) - k \mu_{p+1}(K_{(k-1,p)})) \frac{m^{(p+1)}(x) f'(x)}{f(x)} \right] h^{p-k+2} \quad (3.25) \end{aligned}$$

la variance est quant à elle, donnée par

$$\text{Var} \left[\hat{m}_{PL}^{(k)}(x) | X_1, \dots, X_n \right] \approx n^{-1} h^{-2k-1} R(K_{(k,p)}) \sigma^2(x) f(x) \quad (3.26)$$

Le biais du cas $(p-k)$ impair continue à se distinguer par sa simplicité à l'opposé du cas pair qui se présente sous une forme plutôt complexe. Les développements nécessaires pour obtenir les équations (3.21) à (3.26) sont donnés en notation matricielle dans l'annexe (A3.3).

3.3 Estimateur par noyaux asymétriques

Contrairement au cas des noyaux symétriques où l'on arrive à identifier des formes générales pour le biais et la variance de l'estimateur. L'utilisation des noyaux asymétriques se heurte à la difficulté, déjà rencontrée lors de l'estimation de densité, posée par la nature de ces noyaux qui font que lorsque le noyau change, la manière dont les propriétés de l'estimateur se calculent et s'expriment changent également.

De surcroît, la diversité est d'autant plus grande lors de l'estimation de régression par noyau, d'une part à cause de la possibilité d'ajuster des polynômes de degrés différents et d'autre part à cause du type de l'estimateur lui-même, car comme il a été précisé au début du chapitre en plus de l'estimateur de Nadaraya-Watson (équation (3.3)), on retrouve notamment les estimateurs de Priestley-Chao (équation (3.4)) et Gasser-Muller (équation (3.5)).

Pourtant, en dépit de la difficulté exprimée ci-dessus, il semble dans plusieurs cas, tels que ceux que nous allons exposer, que les techniques et les développements nécessaires pour le calcul des

propriétés -biais et variance- de l'estimateur de la fonction de régression se ramènent aux approches utilisées dans le cas de l'estimation de densité. Afin d'illustrer cela, nous présentons les biais et variance de l'estimateur de Nadaraya-Watson à base d'un noyau gamma et de l'estimateur de Gasser-Müller à base d'un noyau bêta.

3.3.1 Estimateur de Nadaraya-Watson à base du noyau gamma

L'estimateur à base du noyau gamma, s'écrit de manière similaire à l'estimateur donné par l'équation (3.3), il suffit simplement de remplacer le noyau symétrique par le noyau gamma $K_{x/h+1,h}^G$ donné par l'équation (2.2). On trouve alors

$$\hat{m}_{NW}^G(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} \quad (3.27)$$

À l'aide d'un simple calcul, on trouve aussi que

$$\hat{m}_{NW}^G(x) = \frac{\sum_{i=1}^n X_i^{n/h} \exp(-X_i/h) Y_i}{\sum_{i=1}^n X_i^{n/h} \exp(-X_i/h)} \quad (3.28)$$

Mesure de l'erreur

Avant d'énoncer le biais et la variance de cet estimateur, donnons le résultat suivant qui est une proposition nécessaire aux calculs (Shi et Song [94]).

Proposition 3.1 *Soit g_{α_k, β_k} la fonction de densité de la loi Gamma (α_k, β_k) avec $\alpha_k = kx/h + 1$ et $\beta_k = h/k$*

$$g_{\alpha_k, \beta_k}(u) = \frac{u^{\alpha_k-1} \exp(-u/\beta_k)}{\Gamma(\alpha_k) \beta_k^{\alpha_k}} ; u > 0$$

et soit l une fonction deux fois dérivable, continue et bornée sur $]0, +\infty[$. On a alors

$$\int g_{\alpha_k, \beta_k}(u) l(u) du = l(x) + \frac{[2l'(x) + xl''(x)]h}{2k} + o(h)$$

Biais :

Le biais étant

$$\text{Biais} [\hat{m}_{NW}^G(x) | X_1, \dots, X_n] = E [\hat{m}_{NW}^G(x) | X_1, \dots, X_n] - m(x)$$

En considérant l'équation (3.27), il peut être réécrit sous la forme

$$\text{Biais} [\hat{m}_{NW}^G(x) | X_1, \dots, X_n] = \frac{\sum_{i=1}^n K_{x/h+1,h}(X_i) (m(X_i) - m(x))}{n \hat{f}_n(x)} \quad (3.29)$$

avec \hat{f}_n l'estimateur de densité à base du noyau gamma (équation (2.2)).

Si l'on pose

$$N(x) = \frac{1}{n} \sum_{i=1}^n K_{x/h+1,h}(X_i) [m(X_i) - m(x)]$$

on aura

$$E[N(x)] = E[K_{x/h+1,h}(X) m(X)] - m(x) E[K_{x/h+1,h}(X)]$$

ensuite, une application directe de la proposition (3.1) pour $E_1 = E[K_{x/h+1,h}(X) m(X)]$ et $E_2 = E[K_{x/h+1,h}(X)]$ avec $\varphi = fm$ donne

$$E_1 = \varphi(x) + \frac{[2\varphi'(x) + x\varphi''(x)]h}{2} + o(h)$$

$$E_2 = f(x) + \frac{[2f'(x) + xf''(x)]h}{2} + o(h)$$

il s'ensuit que

$$\text{Biais} [\hat{m}_{NW}^G(x) | X_1, \dots, X_n] \approx hm'(x) + \frac{1}{2}hxm''(x) + hxm'(x) \frac{f'(x)}{f(x)} \quad (3.30)$$

Variance :

Pour la variance, il suffit de reprendre des calculs tout à fait similaire à ceux effectués pour le calcul de la variance de l'estimateur de densité à noyau gamma, on retrouve alors directement

$$\text{Var} [\hat{m}_{NW}^G(x) | X_1, \dots, X_n] \approx \frac{x^{-1/2}h^{-1/2}}{2nf(x)\sqrt{\pi}}\sigma^2(x) \quad (3.31)$$

3.3.2 Estimateur de Gasser-Müller à base du noyau beta

L'estimateur de Gasser-Müller à base d'un noyau symétrique est donné par l'équation (3.5), afin d'obtenir l'estimateur que l'on désire considérer, il faudra remplacer le noyau symétrique par le noyau bêta donné par l'équation (2.4). On aura alors

$$\hat{m}_{GM}^B(x) = \sum_{i=1}^n y_i \int_{s_{i-1}}^{s_i} K_{x/h+1,(1-x)/h+1}(t) dt \quad (3.32)$$

avec $s_i = (1/2)(x_i + x_{i+1})$, $s_0 = 0$, $s_n = 1$.

Mesure de l'erreur

Biais :

Une approche identique à celle adoptée pour retrouver l'équation (2.7) lors de l'estimation de densité à base des noyaux asymétriques mène à ce que le biais ait cette forme

$$\text{Biais} [\hat{m}_{GM}^B(x)] = E \left[m \left(\mathcal{K}_{x/h+1, (1-x)/h+1}^B \right) \right] - m(x) \quad (3.33)$$

avec $\mathcal{K}_{x/h+1, (1-x)/h+1}^B$ une variable aléatoire de densité $K_{x/h+1, (1-x)/h+1}^B$.

Une fois de plus, un développement similaire à celui qui a conduit à l'équation (2.14) donne

$$\text{Biais} [\hat{m}_{GM}^B(x)] \approx h(1-2x)m'(x) + \frac{1}{2}hx(1-x)m''(x) \quad (3.34)$$

Variance :

Ici également, l'approche reproduit exactement les étapes suivies pour calculer la variance de l'estimateur de densité à base du noyau bêta. Nous trouvons alors que

$$\begin{aligned} \text{Var} [\hat{m}_{GM}^B(x)] &= \sum_{i=1}^n \sigma^2(x_i) \left[\int_{s_{i-1}}^{s_i} K_{x/h+1, (1-x)/h+1}^B(t) dt \right]^2 \\ &\approx \frac{B_h(x)}{nf(x)} \sigma^2(x) \end{aligned} \quad (3.35)$$

et on rappelle que

$$B_h \approx \begin{cases} \frac{h^{-1/2}}{2\sqrt{\pi}\sqrt{x(1-x)}} & \text{si } x/h \rightarrow \infty \text{ et } (1-x)/h \rightarrow \infty \\ \frac{\Gamma(2x/h+1)}{2^{2x/h+1}\Gamma^2(x/h+1)} h^{-1} & \text{si } x/h \rightarrow k \text{ ou } (1-x)/h \rightarrow k \end{cases}$$

d'où

$$\text{Var} [\hat{m}_{GM}^B(x)] \approx \begin{cases} \frac{h^{-1/2}n^{-1}}{2f(x)\sqrt{\pi}\sqrt{x(1-x)}} \sigma^2(x) & \text{si } x/h \rightarrow \infty \text{ et } (1-x)/h \rightarrow \infty \\ \frac{\Gamma(2x/h+1)n^{-1}h^{-1}}{2^{2x/h+1}f(x)\Gamma^2(x/h+1)} \sigma^2(x) & \text{si } x/h \rightarrow k \text{ ou } (1-x)/h \rightarrow k \end{cases} \quad (3.36)$$

3.4 Sélection du paramètre de lissage

Comme c'est le cas pour l'estimation de densité, le choix du paramètre de lissage constitue un élément clé pour obtenir des résultats corrects dans une régression à noyau en raison du lien étroit qui continue d'exister entre ce paramètre et la manière dont la méthode interagit avec les données.

Bien évidemment, les considérations dans le rapport entre la quantité du lissage, le biais et la variance sont toujours valables. Le paramètre h est également perçu comme un moyen de régler la complexité du modèle, ainsi sous l'effet du lissage, l'estimation effectuée peut osciller entre une interpolation où le biais est minime et la régression linéaire où le biais est beaucoup plus important.

Par ailleurs, le choix initial du paramètre du lissage se fait à travers des méthodes qui sont à la base des adaptations de méthodes utilisées lors de l'estimation de densité, notamment la règle pratique, la validation croisée, le plus proche voisin et les méthodes d'insertions.

La règle pratique et les méthodes d'insertion, se basent sur des critères *MISE* pour un paramètre de lissage global et *MSE* pour un paramètre local. Ce dernier, bien entendu, permet d'exercer un meilleur contrôle sur l'équilibre entre le biais et la variance, principalement dans les zones où les données sont soit trop denses ou trop éparées. Ceci offre alors différents degrés de lissage et améliore la flexibilité de l'ajustement en contrepartie d'un temps de calcul élevé.

En théorie, les critères *AMSE* conditionnelle et *AMISE* conditionnelle pondérée de l'estimateur par polynômes locaux dans le cas de p impair sont données par

$$AMSE_x[\hat{m}_{PL}|X_1, X_2, \dots, X_n] = \left[\frac{h^{p+1}}{(p+1)!} \right]^2 \left(m^{(p+1)}(x) \right)^2 \mu_{p+1}^2(K_{(p)}) + \frac{R(K_{(p)})}{nhf(x)} \sigma^2(x) \quad (3.37)$$

$$AMISE[\hat{m}_{PL}|X_1, X_2, \dots, X_n] = \left[\frac{h^{p+1} \mu_{p+1}(K_{(p)})}{(p+1)!} \right]^2 \times \int (m^{(p+1)}(u))^2 f(u) du + \frac{R(K_{(p)})}{nh} \sigma^2(x) \quad (3.38)$$

et, dans le cas homoscedastique, sont respectivement minimisés par

$$h_{AMSE}(x) = \left[\frac{(p+1)(p!)^2 R(K_{(p)}) \sigma^2}{2n \mu_{p+1}^2(K_{(p)}) \left(m^{(p+1)}(x) \right)^2 f(x)} \right]^{1/(2p+3)} \quad (3.39)$$

$$h_{AMISE} = \left[\frac{(p+1)(p!)^2 R(K_{(p)}) \sigma^2}{2n \mu_{p+1}^2(K_{(p)}) \int \left(m^{(p+1)}(u) \right)^2 f(u) du} \right]^{1/(2p+3)} \quad (3.40)$$

Un certain nombre d'éléments inconnus, notamment la variance σ^2 , la fonction $m^{(p+1)}$ et la densité f empêchent une application directe de ces formules. La règle pratique entend donner une estimation simple et commode mais raisonnable du paramètre de lissage tandis que la méthode d'insertion est une approche plus élaborée visant à augmenter le degré de précision de la valeur calculée.

3.4.1 Règle pratique

Dans le cas de la régression, la règle pratique suggère des moyens relativement rapides pour estimer convenablement les valeurs inconnues. Cette approche est particulièrement utile lorsqu'il

s'agit d'estimer des paramètres pilotes pour des méthodes de sélection plus élaborées ou lorsqu'il est nécessaire d'estimer ce paramètre à plusieurs reprises.

Pour la densité inconnue la règle adopte une référence à une loi paramétrique ou une estimation de densité à noyau. Pour la dérivée $m^{(p+1)}$, il est suggéré d'effectuer un ajustement polynomial simple, de préférence de degré $(p + 3)$ afin de donner une certaine flexibilité à la dérivée. La variance quant à elle est estimée par la somme des carrés des résidus standardisés de l'ajustement polynomial.

3.4.2 Insertion directe

Cette approche s'adresse principalement au cas où p est impair (équation (3.40)) et plus particulièrement à l'estimateur linéaire local en raison de sa relative simplicité. Sous l'hypothèse d'un modèle homoscedastique, le paramètre de lissage qui minimise le critère *AMISE* de l'estimateur linéaire local est donné par

$$h_{AMISE} = \left[\frac{R(K) \sigma^2}{n \mu_2^2(K) \int \left(m^{(2)}(u)\right)^2 f(u) du} \right]^{1/5} \quad (3.41)$$

Soit

$$I_2 = \int \left(m^{(2)}(u)\right)^2 f(u) du$$

un estimateur de I_2 peut être donné par

$$\hat{I}_2(g) = n^{-1} \sum_{i=1}^n \left(\hat{m}^{(2)}(X_i)\right)^2 \quad (3.42)$$

alors, qu'un estimateur de σ^2 est donné par

$$\hat{\sigma}^2(\lambda) = \nu^{-1} \sum_{i=1}^n (Y_i - \hat{m}(X_i))^2 \quad (3.43)$$

avec

$$\nu = n - 2 \sum_{i=1}^n w_{ii} + \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2$$

et

$$w_{ij} = e_1^T (\mathbb{X}_{X_i} W_{X_i} \mathbb{X}_{X_i})^{-1} \mathbb{X}_{X_i}^T W_{X_i} e_j$$

ces deux estimateurs sont respectivement basés sur deux paramètres de lissage pilote g et λ .

Le paramètre \hat{h}_{AMISE} estimé par insertion directe prend donc la forme

$$\hat{h}_{AMISE} = \left[\frac{R(K) \hat{\sigma}^2(\lambda)}{n\mu_2^2(K) \hat{I}_2(g)} \right]^{1/5} \quad (3.44)$$

Il est nécessaire maintenant de retrouver les paramètres pilotes g et λ requis pour le calcul de l'estimateur \hat{h}_{AMISE} . Comme c'est le cas pour les techniques d'insertion, un choix doit se faire à ce niveau, il faut alors, soit se fier à la règle pratique pour le calcul des paramètres pilotes, soit réutiliser un critère d'erreur pour le minimiser et se retrouver avec d'autres paramètres de lissage à estimer.

3.4.3 Validation croisée

C'est dans le contexte de la régression que le principe de la validation croisée est le plus facile à présenter et à appliquer. En effet, le critère à minimiser s'exprime de manière très simple

$$CV(h) = n^{-1} \sum_{i=1}^n (Y_i - \hat{m}_{PL,-i}(X_i))^2 \quad (3.45)$$

où $\hat{m}_{PL,-i}$ est l'estimateur de régression par exclusion du couple (X_i, Y_i) .

Le paramètre de lissage selon cette approche est alors le minimiseur de l'équation (3.45). Il est toutefois important de rappeler que les inconvénients de la sélection par validation croisée souligné dans le cas de l'estimation par densité, notamment les minima locaux et le problème de sous lissage, s'étendent au cas de la régression ce qui cause des fluctuations très prononcées et risque d'affecter gravement toute prévision à base de ce modèle.

3.4.4 Sélection par plus proche voisin

Cette approche hybride entre le plus proche voisin et les paramètres de lissage variables permet d'effectuer une sélection adaptative simple. Elle est particulièrement utile dans le cas d'un schéma aléatoire qui nécessiterait un réglage du paramètre selon que les observations sont nombreuses dans certaines zones ou pas.

Le paramètre de lissage est exprimé de la façon suivante :

$$\hat{h}(x) = (X_{(l+k)} - X_{(l-k)}) / 2 \quad (3.46)$$

où X_l est l'observation la plus proche du point d'estimation x et k un entier naturel qui détermine l'étendue des points à couvrir.

Chapitre 4

Estimation de densité de la volatilité stochastique

Dans ce chapitre et le suivant nous nous intéresserons à des applications du lissage par noyaux, dans lesquels nous abordons quelques questions qui surviennent lorsque l'on traite des aspects quantitatifs des marchés financiers.

En premier lieu, l'attention est portée sur la volatilité qui est l'un des paramètres les plus importants impactant le comportement des intervenants, tellement que sur certains marchés les options¹ sont listées en termes de valeur de volatilité en raison de son rôle décisif dans la détermination de la prime de risque². La volatilité des actifs financiers est également considérée comme la variable d'état sous-jacente, à la place du prix, pour différents types d'options, vu qu'elle offre une meilleure visibilité sur l'état de santé de l'actif. Ceci implique qu'une bonne connaissance de la volatilité est un élément clé pour évaluer correctement l'état et la stabilité du marché si l'on désire élaborer une politique d'investissement efficace et optimale en termes de gestion du risque.

Par ailleurs, acquérir une compréhension exhaustive de la volatilité implique l'examen d'un certain nombre d'aspects dont plusieurs ont été largement étudiés. Par exemple le problème de la modélisation de la volatilité a été approché à partir de plusieurs angles différents. Hormis le modèle EWMA qui est une variante non stationnaire du populaire GARCH(1,1) (Bollerslev [6]) et les nombreux modèles de types ARCH et GARCH examinés dans la littérature (Bollerslev, Engle et Nelson[7]; Franses et Van dijk [42]; Satchell et Knight [91]), d'autres pistes plus exotiques impliquant les réseaux de neurones (Hajizadeh et al.[53]) et la programmation génétique (Yin, O'Sullivan et Brabazon [109]) ont été explorées.

De plus, quelques problèmes pratiques sur les marchés tels que la valorisation d'option et le besoin pour un modèle qui soit en temps continu et qui adopte une dynamique réaliste pour le sous-jacent ont engendré les modèles de volatilité stochastique qui affichent un niveau de flexibilité élevé grâce à leur capacité à reproduire des caractéristiques importantes de la volatilité telles que l'asymétrie (Hull et White [60]) et le 'smile' (Gatheral [44]).

Un autre problème associé à la volatilité est le problème d'estimation de la densité. Ce sujet a été traité de manière efficace dans le cadre non paramétrique par un modèle de déconvolution dans lequel les observations sont de la forme

1. Une option est un contrat qui donne à l'acheteur le choix d'acheter ou de vendre un actif financier appelé, actif sous-jacent.

2. La prime de risque est le prix de marché d'une option. Le chapitre suivant offrira plus de détails sur les options et leurs spécificités.

$$Y_i = X_i + \epsilon_i \quad (4.1)$$

Où (X_i) est le signal qui nous intéresse et (ϵ_i) le bruit. Sous l'hypothèse que (X_i) et (ϵ_i) sont *i.i.d* Fan [39] a étudié l'ordre de convergence atteint par une estimation de densité à noyau alors que Pensky et Vidakovic [82] ont utilisé un estimateur adaptif basé sur les ondelettes. Cependant, l'hypothèse de l'indépendance n'est pas toujours vérifiée pour le signal en question qui, par conséquent, peut être modélisé comme une variable aléatoire stationnaire présentant une faible forme de dépendance exprimée par des conditions de mélange. Des idées générales sur comment employer les estimateurs à noyaux et les estimateurs par projection sous de telles conditions sont données par Rio [86]. Un travail plus complet employant la transformée de Fourier et le théorème d'inversion de Fourier est présenté dans Van Es, Spreij et Van Zanten [102].

Dans ce chapitre d'application, nous commençons par présenter le modèle bidimensionnel adopté pour simuler l'évolution de volatilité stochastique ensuite nous passons au schéma de discrétisation qui nous servira à implémenter le modèle de manière optimale. Par la suite, nous introduisons la structure de convolution et les estimateurs qui permettront d'estimer la volatilité à partir de la coordonnée observable avant de passer aux expérimentations numériques durant lesquels on effectue une comparaison entre les résultats obtenus par différentes méthodes. On termine par une discussion des résultats et une conclusion.

4.1 La volatilité stochastique

Dans le cadre de la théorie classique du portefeuille, le *log*-prix d'un actif financier peut être représenté par l'équation différentielle stochastique suivante :

$$dY_t = b_t dt + v_t dW_t \quad (4.2)$$

où (W_t) est un mouvement Brownien Standard et (b_t) et (v_t) satisfont quelques conditions habituelles de régularité (Karatzas et Shreve [65]). De plus, le processus de volatilité (v_t) qui entraîne le coefficient de diffusion est supposé indépendant de (W_t) , strictement stationnaire ergodique et sa distribution admet une densité par rapport à la mesure de Lebesgue.

Par ailleurs, le fait que l'on désire estimer la densité à partir de données simulées, accentue le besoin pour un modèle qui permette à la volatilité d'évoluer d'une manière réaliste et libre et qui, en même temps, puisse préserver la relation avec le processus de prix. Pour la simple raison que, contrairement au prix des actifs, leur volatilité n'est pas une caractéristique visible sur les marchés. Par conséquent, elle peut être considérée, dans le cadre d'un processus bidimensionnel, comme une coordonnée cachées qui ne peut être estimée qu'à travers la coordonnée observable.

4.1.1 Modèle de Heston

Une approche populaire pour reproduire de manière vraisemblable les fluctuations de la volatilité est l'utilisation de modèles appartenant à la classe CEV-SV pour "Constant Elasticity of Variance - Stochastic Volatility" (Andersen et Piterbarg [3]). Ces modèles offrent un moyen de décrire un processus de diffusion bidimensionnel à travers les EDS suivantes sous la mesure risque neutre

$$\begin{cases} dS_t = \mu S_t dt + \lambda \sqrt{V_t} S_t^\beta dW_t^S \\ dV_t = k(\theta - V_t) dt + \omega V_t^\alpha dW_t^V \end{cases} \quad (4.3)$$

Dans un contexte financier, (S_t) représente le prix de l'actif et (V_t) représente sa variance. Pour ce qui est des paramètres du modèle, μ est la dérive risque neutre du prix de l'actif, λ est une constante d'échelle, k est la vitesse de retour à la moyenne pour la variance, θ la moyenne à long terme pour la variance et ω la volatilité de la variance. Finalement, (W_t^S) et (W_t^V) sont deux mouvements Brownien Standards tels que $\langle dW^S | dW^V \rangle = \rho dt$.

En prenant $\alpha = 1/2$, $\beta = 1$ et $\lambda = 1$ on retrouve le modèle de Heston qui suppose pour le prix de l'actif un mouvement Brownien géométrique conformément au cadre classique de Black & Scholes et pour la variance, un processus de retour à la moyenne inspiré du modèle de Cox, Ingersoll et Ross pour les taux d'intérêts (Cox, Ingersoll et Ross [28]). Ces équations se sont montrées particulièrement efficaces dans l'élimination de l'hypothèse de la volatilité constante et dans l'intégration de la volatilité stochastique dans les modèles de valorisation traditionnels en lui attribuant une dynamique spécifique (Gatheral [44]).

Dans ce qui suit, le système ci-dessous sera utilisé pour modéliser les mouvements du log-prix (Y_t) et de la variance (V_t) .

$$\begin{cases} dY_t &= \sqrt{V_t} dW_t^Y \\ dV_t &= k(\theta - V_t) dt + \omega \sqrt{V_t} dW_t^V \end{cases} \quad (4.4)$$

Les résultats donnés dans les propositions (4.1) et (4.2) renseignent sur la manière dont la dynamique de ce système fonctionne.

Proposition 4.1 *Soit $T > t$, l'espérance et variance conditionnelles de V_T par rapport à V_t sont données par*

$$E[V_T | V_t] = V_t e^{-k(T-t)} + \theta \left(1 - e^{-k(T-t)}\right) \quad (4.5)$$

$$Var[V_T | V_t] = \frac{V_t \omega^2}{k} \left(e^{-k(T-t)} - e^{-2k(T-t)}\right) + \frac{\theta \omega^2}{2k} \left(1 - e^{-k(T-t)}\right)^2 \quad (4.6)$$

Proposition 4.2 (Condition de Feller) *Étant donné une valeur initiale V_0 strictement positive, si $k\theta > \omega^2/2$ alors le processus (V_t) n'atteint jamais zéro.*

4.1.2 Schéma de simulation

Maintenant, il est nécessaire de pouvoir mettre en place une discrétisation adéquate du modèle adoptée. Plusieurs variantes simples et directes du Schéma d'Euler existent (Lord, Koekkoek et Van Dijk [72]), toutefois ces variantes partagent toutes un même inconvénient qui est leur incapacité à empêcher le processus de variance d'atteindre des valeurs négatives. En effet, malgré que le processus racine-carrée est connu pour conserver sa positivité étant donné une valeur initiale strictement positive, sa discrétisation ne dispose pas de cette propriété (Feller [41]). Par exemple, le schéma direct d'Euler suivant pour (V_t) :

$$V_{t+\Delta t} = V_t (1 - k\Delta t) + k\theta\Delta t + \omega \sqrt{V_t} \Delta W_t^V \quad (4.7)$$

a la probabilité suivante d'atteindre des valeurs négatives :

$$P(V_{t+\Delta t} < 0) = \Phi\left(\frac{-V_t - k(\theta - V_t)\Delta t}{\omega\sqrt{V_t\Delta t}}\right) \quad (4.8)$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

Des moyens classiques pour contourner ce problème sont les techniques d'absorption et de réflexion. La première consiste en le remplacement des valeurs négatives avec des zéros alors que la deuxième utilise $|V_t|$ au lieu de V_t (Gatheral [44]). Bien que ces approches soient commodes et très pratiques, elles présentent des défauts qui réduisent leur efficacité. D'une part, l'absorption implique une absence de volatilité durant certains laps de temps. D'autre part, la réflexion transforme les faibles volatilités en de fortes volatilités.

Par ailleurs, la conception d'un schéma qui ne va produire aucune valeur négative est un autre moyen d'adresser ce problème. Alfonsi [2] montre qu'un schéma de Milstein implicite réduit significativement l'apparition de valeurs négatives. Broadie et Kaya [20] montrent comment un échantillonnage à partir de la loi de transition exacte empêche complètement le processus de passer en dessous de zéro aux dépens d'un temps de calcul raisonnable. L'utilisation d'un schéma de volatilité transformée, préserve également la positivité des valeurs produites par la simulation, pourtant une implémentation brute entraîne un comportement incohérent de la simulation. Cette complication peut, toutefois, être surmontée en intégrant au schéma une approximation de la moyenne à long terme donnée par Zhu [110]. Cette dernière approche est décrite dans ce qui suit.

D'abord, la volatilité (V_t) est simulée au lieu de la variance. Par le lemme d'Itô, dont une présentation intuitive est donnée dans l'annexe (A4.1) et pour lequel une démonstration rigoureuse peut être retrouvée dans Karatzas et Shreve [65], on trouve

$$dv_t = \frac{k}{2} \left[\frac{1}{v_t} \left(\theta - \frac{\omega^2}{4k} \right) - v_t \right] dt + \frac{1}{2} \omega dW_t^V \quad (4.9)$$

on peut remarquer que la moyenne à long terme

$$\theta_v = \frac{1}{v_t} \left(\theta - \frac{\omega^2}{4k} \right)$$

a une nature stochastique, chose qui affecte négativement la qualité de la simulation (Rouah [88]). Ce problème peut être résolu par l'utilisation d'une approximation robuste $\hat{\theta}_t$, suggérée par Zhu [110]

$$\hat{\theta}_t = \frac{\beta - v_t \exp(-kdt/2)}{1 - \exp(-kdt/2)} \quad (4.10)$$

où

$$\beta = \sqrt{\left(\theta + (v_t - \theta) e^{-kdt} - \frac{\omega^2}{4k} (1 - e^{-kdt}) \right)^+}$$

une discrétisation d'Euler du processus de la volatilité transformée, donne alors

$$v_{t+\Delta t} = v_t + \frac{k}{2} (\hat{\theta}_t - v_t) \Delta t + \frac{1}{2} \omega \Delta W_t^V \quad (4.11)$$

L'équation (4.11) sera alors la forme utilisée pour simuler la volatilité accompagnée d'une discrétisation d'Euler naïve pour simuler le log-prix de l'actif. Finalement le schéma de simulation globale est le suivant :

$$\begin{cases} Y_{t+\Delta t} &= Y_t + \sqrt{V_t \Delta t} Z^Y \\ v_{t+\Delta t} &= v_t + \frac{k}{2} (\hat{\theta}_t - v_t) \Delta t + \frac{1}{2} \omega \sqrt{\Delta t} Z^V \end{cases} \quad (4.12)$$

avec $Z^V = Z_1$, $Z^Y = \rho Z^V + \sqrt{1 - \rho^2} Z_2$ et Z_1, Z_2 deux variables aléatoires *i.i.d.*, suivant une loi normale centrée réduite.

4.2 Estimation de densité

4.2.1 Structure de déconvolution

La simulation du modèle de Heston implique que le processus (Y_t) sera observé en des instants de temps t , discrets et réguliers. L'intervalle d'échantillonnage Δt est supposé tendre vers 0 et le nombre d'observations est considéré comme infini. Maintenant, Comme c'est le cas dans Genon-Catalot, Jeantheau, et Laredo [45] et Van Es, Spreij et Van Zanten [102], on définit les incréments suivants :

$$X_t = \frac{1}{\sqrt{\Delta t}} (Y_t - Y_{t-1}) \quad (4.13)$$

Lorsque $\Delta t \rightarrow 0$, une approximation de (X_t) peut aussi être donnée par

$$\begin{aligned} X_t &= \frac{1}{\sqrt{\Delta t}} \int_{t-1}^t v_u dW_u \\ &\approx \frac{1}{\sqrt{\Delta t}} v_{t-1} (W_t - W_{t-1}) \end{aligned} \quad (4.14)$$

ceci permet d'écrire

$$\log(X_t^2) = \log(V_{t-1}) + \log(\xi_t^2) \quad (4.15)$$

avec

$$\xi_t = \frac{1}{\sqrt{\Delta t}} (W_t - W_{t-1}) \sim N(0, 1)$$

étant donné l'indépendance et la stationnarité des incréments browniens.

Pour simplifier la notation, on pose $\alpha_t = \log(X_t^2)$, $\beta_t = \log(V_{t-1})$, $\epsilon_t = \log(\xi_t^2)$. On retrouve alors la forme de convolution suivante :

$$\alpha_t = \beta_t + \epsilon_t \quad (4.16)$$

4.2.2 Estimateurs

D'après l'équation (4.16), on voit bien que la densité de (α_t) est une convolution de deux densités. La première est celle qui nous intéresse, à savoir (β_t) et la deuxième est celle relative à (ϵ_t) . Un moyen d'extraire la densité recherchée est d'utiliser une idée populaire issue de la théorie de déconvolution statistique (Van Es, Spreij, et Van Zanten [103]), une autre approche serait de filtrer les données nécessaires à partir de l'équation (4.16) en utilisant une régression non paramétrique avant d'effectuer une estimation de densité à noyau (Ouamaliche et Sayah [80]). Ces deux approches sont présentées dans ce qui suit.

Soit l'équation (4.16) une représentation de données bruitées, le bruit étant le terme ϵ_t . Habituellement, (ϵ_t) est considéré comme une suite *i.i.d* homoscélastique de densité connue.

Approche par déconvolution

La densité f_α étant une convolution entre f_β et f_ϵ , on peut écrire

$$f_\alpha = f_\beta * f_\epsilon$$

le passage à la fonction caractéristique par la transformée de Fourier donne alors

$$\begin{aligned} \varphi_{f_\alpha}(t) &= E[e^{it\alpha}] \\ &= \varphi_{f_\beta}(t) \varphi_{f_\epsilon}(t) \end{aligned}$$

ensuite une application directe du théorème d'inversion de Fourier donne

$$f_\beta = \frac{1}{2\pi} \int e^{-itx} \frac{\varphi_{f_\alpha}(t)}{\varphi_{f_\epsilon}(t)} dt$$

la densité f_α étant inconnue, elle peut être remplacée par un estimateur à noyau. On aura alors

$$\hat{f}_\beta(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\varphi_{\hat{f}_\alpha}(t)}{\varphi_{f_\epsilon}(t)} dt \quad (4.17)$$

Approche par régression à noyau

Comme il a déjà été indiqué, cette approche se fait sur deux étapes. La première est l'application d'une régression unidimensionnelle à noyau par l'ajustement d'un polynôme de degré p aux observations $\{(t/n, \alpha_t)\}_{t=1}^n$ de sorte à ce qu'il minimise

$$\sum_{i=1}^n \left\{ \alpha_i - b_0 - \dots - b_p \left(x - \frac{i}{n} \right)^p \right\}^2 K \left(\frac{x - (i/n)}{h_1} \right) \quad (4.18)$$

Une estimation de β_t est alors obtenue en retranchant la moyenne du bruit, supposée différente de 0, aux $\hat{\alpha}_t$ estimés

$$\hat{\beta}_t = \hat{m}(t/n) - E(\epsilon) \quad (4.19)$$

par la suite la densité de $\log(V)$ peut être estimée par

$$\hat{f}_n(x) = \frac{1}{nh_2} \sum_{i=1}^n K\left(\frac{x - \hat{\beta}_i}{h_2}\right) \quad (4.20)$$

4.3 Résultats numériques

Avant de passer aux résultats des simulations, il est d'abord nécessaire de spécifier un repère qui pourrait être utilisé pour évaluer la qualité des estimations proposées.

En reprenant l'équation (4.4) et en considérant les hypothèses

$$n \rightarrow \infty ; \Delta t \rightarrow 0 ; n\Delta t \rightarrow \infty$$

ainsi que la condition de Feller qui exige que l'inégalité $2k\theta \geq \omega^2$ soit vérifiée pour que le processus n'atteigne jamais zéro. La fonction de densité (équation (4.21)) suggérée dans Genon-Catalot, Jeantheau et Laredo [45] semble être un choix convenable pour être une densité de référence

$$\pi(v) = \frac{\mu^a}{\Gamma(a)} v^{a-1} e^{-\mu v} \mathbf{1}(v > 0) \quad (4.21)$$

avec $a = 2k\theta/\omega^2 \geq 1$ et $\mu = 2k/\omega^2$. Il est clair alors que c'est une densité gamma avec a et μ les paramètres respectifs de forme et d'intensité.

En outre, il faut aussi faire attention au fait que la densité à estimer correspond à $\log(V)$ et par conséquent, la forme correcte de la densité de référence sera

$$f_Y(y) = \pi(e^y) e^y \quad (4.22)$$

Maintenant que la densité de référence est fixée, les paramètres du schéma de simulation (table (4.1)) sont sélectionnés de sorte à reproduire les variations de l'indice *S&P500* (Broadie et Kaya [20]). Ce choix est principalement motivé par des fins de comparaison.

S_0	V_0	k	θ	ω
100	0.007569	3.46	0.008	0.14

TABLE 4.1 – Paramètres pour la simulation de volatilité

La figure (4.1), représente une simulation d'un échantillon de données recueillies à la tête de chaque heure durant 30 années. Par ailleurs, le grand nombre de réalisations simulées et le pas de temps relativement petits sont, d'un côté, supposés assurer que les approximations tiennent et d'un autre côté permettre la comparaison avec les résultats existants.

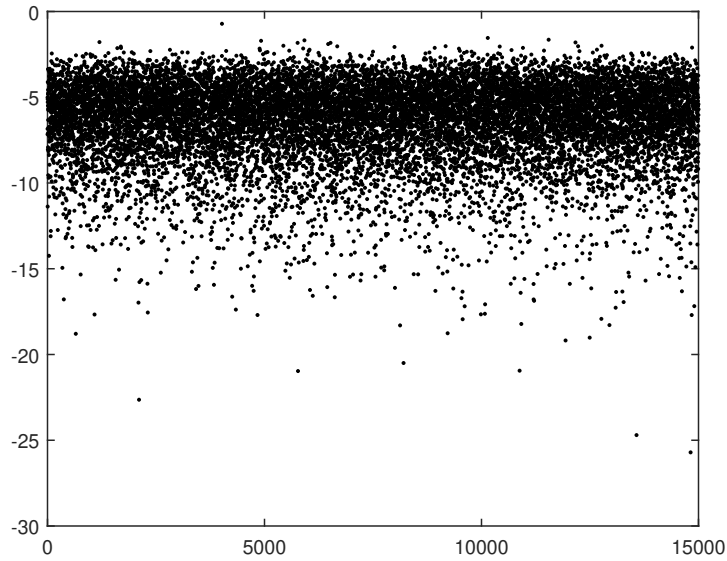


FIGURE 4.1 – Un échantillon des données simulées

La partie 1 du tableau (4.2) représente une description statistique des données simulées, tandis que la partie 2 du même tableau présente le test de normalité de Jarque-Bera en plus de quelques tests de racines unitaires afin de vérifier la normalité.

Partie 1	
Moyenne	-6.3243
Médiane	-5.8882
Maximum	-0.7201
Minimum	-25.7069
Écart-type	2.3090
Asymétrie	-1.3568
aplatissement	6.2151
Partie 2	
Jarque-Bera	0.001
Dickey Fuller Augmenté	0.001
Phillips-Perron	0.001
KPSS	0.01

TABLE 4.2 – Description statistique, tests de normalité et de stationnarité.

Des conclusions importantes concernant la distribution des données peuvent être tirées à partir de la table (4.2). Le coefficient d'asymétrie négatif indique que la distribution est asymétrique à gauche, le coefficient d'aplatissement supérieur à 3 indique, quant à lui, que la distribution est leptokurtique avec des queues épaisses. Ces deux points, couplés au résultat du test de Jarque-Bera rejettent l'hypothèse de normalité de la distribution des données. Les trois autres tests sont dédiés à la stationnarité, les tests de Dickey-Fuller augmenté et Phillips-Perron rejettent l'existence d'une racine unitaire alors que le KPSS rejette l'hypothèse de données stationnaire. Ce dernier est toutefois connu pour induire une erreur de type I et tend à rejeter trop souvent une hypothèse nulle qui est

vraie. On se fierait alors aux tests de racines unitaires pour accepter l'hypothèse de stationnarité.

À présent, on passe à l'estimation de densité selon l'approche par régression à noyau indiquée dans la section précédente. Pour la première phase qui consiste en l'estimation des α_t (équation (4.16)), on utilise d'abord une simple moyenne mobile, puis des estimateurs à polynômes locaux de 1er et de 2nd degré à noyau symétrique tricube (équation (4.23)) qui produit des résultats satisfaisants, sans être très éprouvant en terme de calculs (Takezawa [99]).

$$K_{tric}(u) = (1 - |u|^3)^3 \mathbb{1}(|u| \leq 1) \quad (4.23)$$

pour la deuxième phase, après avoir retranché la moyenne du bruit, la densité est estimée à l'aide d'un noyau symétrique d'Epanechnikov (Table (1.1)) et d'un noyau asymétrique gamma.

Remarque 4.1 *Il convient de noter que les estimations faisant appel au noyau gamma dont le support est $[0, +\infty[$, ont été conduites avec les opposées des données simulées.*

En outre, on présente également les estimateurs obtenus par la méthode de déconvolution (équation (4.17)) ainsi que par une application directe de l'estimation de densité à noyau des α_t diminués de $\bar{\epsilon}$.

Hormis la méthode où l'on retranche uniquement $\bar{\epsilon}$, toutes les autres estimations de β_t ont mené à des estimations de densité qui ont démontré une meilleure performance que celle enregistrée par la méthode de déconvolution. Les figures (4.2) à (4.10) donnent une illustration des estimations obtenues superposées à la densité de référence donnée par l'équation (4.22).

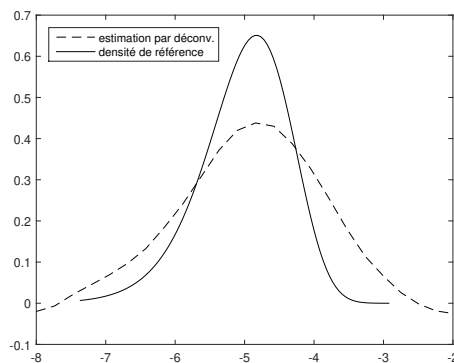


FIGURE 4.2 – Estimation de densité de la volatilité par déconvolution

A partir des graphiques, on peut d'abord observer que contrairement aux autres approches qui donnent des estimations raisonnables et plus précises que la méthode de déconvolution, l'utilisation directe d'un estimateur de densité à noyau et l'omission de la phase de filtration produit une estimation plutôt inutile qui risque de constituer une source de conclusions erronées concernant les aspects les plus élémentaires de la densité en question. Dans ce cas de figure, l'utilisation d'un noyau asymétrique n'a aucun impact positif sur les résultats.

Ensuite, on peut remarquer que lorsqu'on utilise une phase de filtration la qualité de l'estimation est meilleure après le passage du symétrique à l'asymétrique. L'impact est cependant plus prononcé dans le cas du lissage par moyenne mobile. En effet, vu que pour ce dernier l'estimation de densité par noyau symétrique laisse une large marge de perfectionnement, l'utilisation du noyau gamma

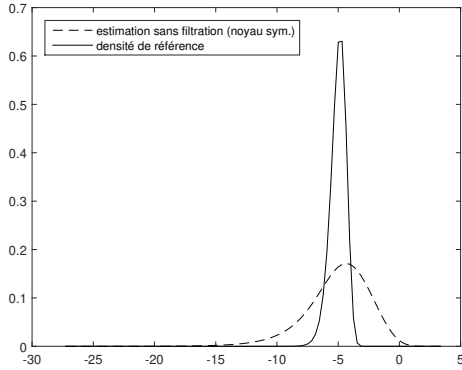


FIGURE 4.3 – Estimation directe de densité de la volatilité par noyau symétrique

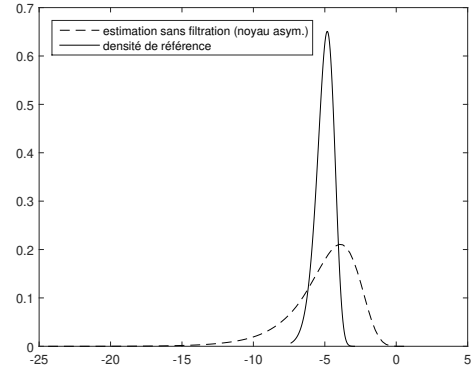


FIGURE 4.4 – Estimation directe de densité de la volatilité par noyau asymétrique

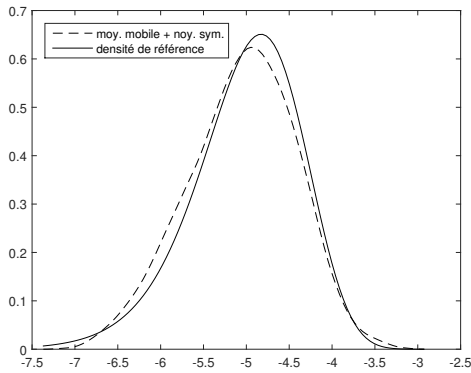


FIGURE 4.5 – Estimation de densité de la volatilité par moyenne mobile (phase 1) et noyau symétrique (phase 2)

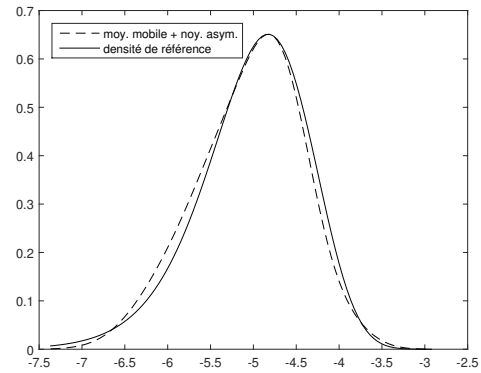


FIGURE 4.6 – Estimation de densité de la volatilité par moyenne mobile (phase 1) et noyau asymétrique (phase 2)

permet d'améliorer de façon très visible les résultats obtenus de sorte à combler la différence causée dans le cas symétrique par les méthodes de filtration.

Afin de confirmer davantage les observations basées sur les figures et évaluer de manière précise la qualité des estimations réalisées, la table (2.1) présente le critère *MISE* calculé, par rapport à la densité de référence, pour chacun des estimateurs selon la formule

$$MISE \approx \frac{1}{2m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} (\hat{f}_j(x_i) - f(x_i))^2 \quad (4.24)$$

avec \hat{f}_j étant l'estimation de densité basée sur la j -ème simulation de donnée.

Les valeurs calculées pour le critère *MISE* confirment les conclusions tirées à partir des figures. En particulier, l'ajustement de polynômes de degré élevé pour la régression à noyau et l'utilisation d'un noyau asymétrique offre la meilleure performance en permettant de mieux capturer la densité et par conséquent réduire l'erreur de manière significative.

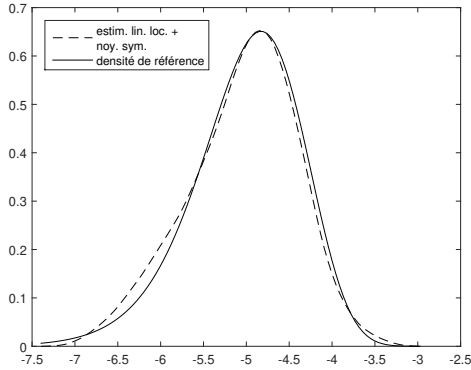


FIGURE 4.7 – Estimation de densité de la volatilité par estimateur linéaire local (phase 1) et noyau symétrique (phase 2)

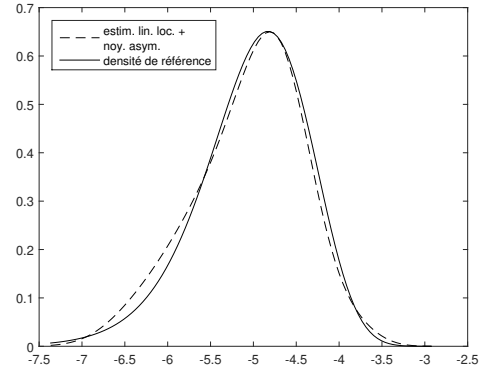


FIGURE 4.8 – Estimation de densité de la volatilité par estimateur linéaire local (phase 1) et noyau asymétrique (phase 2)

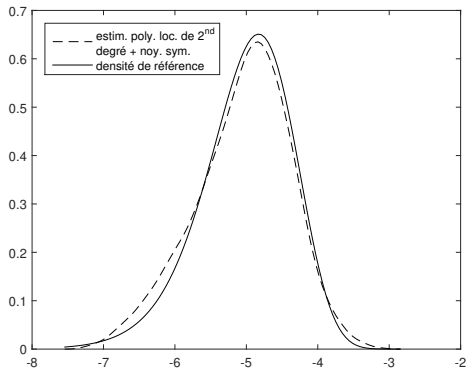


FIGURE 4.9 – Estimation de densité de la volatilité par estimateur à polynômes locaux de 2nd degré (phase 1) et noyau symétrique (phase 2)

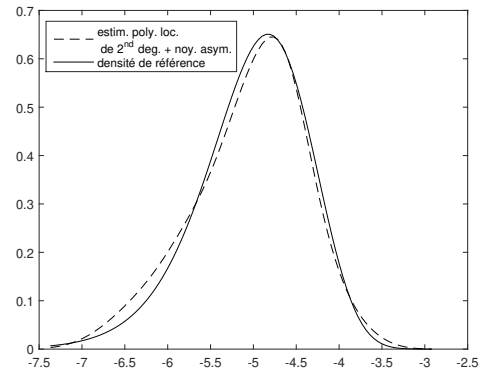


FIGURE 4.10 – Estimation de densité de la volatilité par estimateur à polynômes locaux de 2nd degré (phase 1) et noyau asymétrique (phase 2)

4.4 Conclusion

Dans ce chapitre l'intérêt principal était d'estimer la densité de la coordonnée inobservable (V_t) dans un processus de diffusion bidimensionnel (Y_t, V_t) dans le cas d'observations à haute fréquence simulées à partir du modèle de Heston avec un schéma de volatilité transformée intégrant une approximation robuste de moyenne à long terme.

La procédure proposée pour l'estimation de densité recourt à une première phase durant laquelle les données sont filtrées d'abord à travers une moyenne mobile ensuite par l'ajustement d'un polynôme de degré $p \geq 1$ par moindres carrés pondérées. Cette première phase permet de capturer efficacement le comportement réel du processus. Pour la deuxième phase une estimation de densité à noyau est utilisée.

Concernant la qualité des estimations obtenues, les principales observations se font sur la base des deux éléments suivants : le type de filtration adoptée et le type du noyau utilisé dans l'estimateur de densité. D'une part, lorsque la méthode de filtration est fixée, le passage du noyau symétrique au noyau asymétrique dans l'estimateur de densité réduit sensiblement l'erreur. D'une autre part,

	$MISE(m_1 = m_2 = 100)$
Déconvolution	0.1217
Estimation directe avec noyau symétrique	0.3782
Estimation directe avec noyau asymétrique	0.3696
Moyenne mobile + noyau symétrique	0.0912
Moyenne mobile + noyau asymétrique	0.0871
Poly de degré 1 + noyau symétrique	0.0855
Poly de degré 1 + noyau asymétrique	0.0849
Poly de degré 2 + noyau symétrique	0.0781
Poly de degré 2 + noyau asymétrique	0.0711

TABLE 4.3 – Valeurs du $MISE$ pour les différentes estimations de densités

lorsque c'est le type du noyau de l'estimateur de densité qui est fixé, c'est la méthode de filtration qui agit sur la qualité des résultats. On trouve également que plus le degré du polynôme ajusté est élevé, plus la différence, en terme du critère $MISE$, engendrée par le type du noyau est réduite.

Il semblerait donc qu'une combinaison entre l'ajustement d'un polynôme de degré élevé pour la filtration et un noyau asymétrique pour l'estimation de densité peut certes, mener vers une estimation à erreur très réduite mais coûteuse en terme de temps de calcul.

Chapitre 5

Valorisation d'une option asiatique

Sur les marchés financiers modernes il est possible d'investir dans différents types d'actifs standardisés tels que les actions, les obligations et les produits dérivés. Les plus communs parmi ces derniers sont les Forwards, les Futures et les options.

Notre travail s'intéresse principalement aux options, lesquelles sont des contrats qui donnent le droit mais pas l'obligation d'acheter ou de vendre un certain actif pour un prix préétabli. Les deux options les plus simples sont les options call et put européennes.

Lorsqu'on achète un call européen, on acquiert le droit d'acheter un actif sous-jacent à un prix K que l'on appelle le prix d'exercice (ou strike) en une date T que l'on appelle échéance (ou maturité). Si à l'échéance, le prix S_T de l'actif sous-jacent est supérieur au strike, l'exercice de l'option offre un gain de $S_T - K$. Si au contraire le prix S_T de l'actif sous-jacent est inférieur au strike, l'option n'est pas exercée. Le gain de cette option (ou payoff) est alors $\max(S_T - K, 0)$. Un put européen donne quant à lui le droit de vendre à l'échéance l'actif sous-jacent. Dans ce cas, il sera exercé uniquement si le strike est supérieur au prix de l'actif. Le payoff est alors $\max(K - S_T, 0)$.

Ce type d'option, appelé communément options vanilles, a l'avantage de se soustraire au problème de valorisation (ou pricing) qui est l'une des complications majeures rencontrées sur les marchés pour ces types de contrats. En effet, le formalisme de Black & Scholes permet d'établir et de résoudre analytiquement, pour les options vanilles, une équation qui décrit la dynamique du prix et par conséquent permet aussi d'identifier une valeur, du moins théorique, pour ce dernier (Annexe (A5.2)).

De plus, le cadre instauré par le modèle de Black & Scholes a offert un terrain sur lequel se construisent la majorité des approches de valorisation pour tous types d'options. Toutefois, bien que ces approches conduisent elles aussi vers des équations aux dérivées partielles, elles ne disposent pas de solutions analytiques.

En tant qu'exemples, on peut considérer les options américaines (Brennan et Schwartz [18]) qui donnent la possibilité d'exercer l'option à tout moment de la durée de vie du contrat, les options Bermudéennes (Egloff, Kohler et Torodovic [37]) qui peuvent être exercées uniquement en certaines dates prédéterminées entre la date d'acquisition et la date d'expiration du contrat, les options barrières (Zvan, Vetzal et Forsyth [111]) dont le payoff dépend du prix de l'actif sous-jacent franchissant ou dépassant un niveau de barrière prédéfini, les options lookback (Conze [26]) dont le payoff dépend de la valeur optimale atteinte par le prix de l'actif sous-jacent durant la durée de vie du contrat et les options asiatiques (Rogers et Shi [87]) dont le payoff dépend du chemin pris par le prix de l'actif sous-jacent durant une période de temps prédéfinie.

En outre, l'absence de formules fermées et la difficulté de trouver celles-ci pour le prix d'une

option, à l'exception d'une option vanille sur un sous-jacent qui ne paie pas de dividende, a poussé théoriciens et praticiens à envisager l'utilisation de simulations et à considérer des approximations numériques pour la résolution de ces problèmes. Hormis le modèle binomial, qui adresse la valorisation d'une option vanille à sous-jacent unique, proposé par Cox, Ross et Rubinstein [29], les premières tentatives sérieuses abordant par des méthodes numériques la valorisation d'options furent conduites durant les années quatre-vingt-dix. D'abord, il y a eu une méthode binomiale pour les processus de diffusion autres que le mouvement Brownien géométrique suggéré par Nelson et Ramaswamy [78] ensuite, il y a eu des méthodes trinomiales introduites par Hull et White [61]. Concernant les simulations, elles furent utilisées en premier à la fin des années soixante-dix (Boyle [15]) et depuis elles sont appliquées de façon efficace dans la valorisation de différents types d'options. Une vue d'ensemble sur l'utilisation des simulations pour la valorisation d'actifs financier est donnée dans Boyle, Broadie et Glasserman [16].

Dans ce qui suit, nous nous intéressons à la construction d'une combinaison entre la simulation de Monte Carlo et l'estimation de densité par noyau dans le but de générer des approximations raisonnables pour les prix de différentes options. L'adoption de l'approche Monte Carlo est surtout motivée par le fait bien établi que, depuis son introduction par Boyle [15], elle est devenu un outil numérique majeur dans l'univers de la finance quantitative, essentiellement parce qu'elle réussit à produire des résultats fiables là où les autres méthodes échouent (Fusai et Roncoroni [43]). L'utilisation des simulations de Monte Carlo ne s'arrête pas uniquement à la valorisation d'option, qui est l'objectif de ce chapitre, mais s'étend à la simulation de techniques de couvertures, à l'évaluation des règles de gestion de portefeuilles et à l'estimation de la VaR (Value at Risk). Toutefois, un inconvénient potentiel de cette approche est la lourdeur des calculs, c'est-à-dire, qu'il est nécessaire de réaliser un grand nombre de répliques afin d'avoir des approximations qui soient satisfaisantes.

Ce chapitre débute par une présentation de la méthode classique de Monte Carlo pour la valorisation d'une option quelconque. Cette présentation est suivie d'une description de quelques méthodes de réduction de la variance dont chacune est associée à une application utile sur les marchés financiers. Par la suite, nous présentons la formule pour l'approche de Monte Carlo ajustée par poids et nous mettons en relief l'importance de l'estimation de densité pour l'implémentation de cette technique pour ensuite la tester sur la valorisation d'options asiatiques et finalement discuter les résultats et conclure.

5.1 Simulations classiques de Monte Carlo pour la valorisation d'options

Les simulations de Monte Carlo semblent être un choix convenable lorsqu'il s'agit de la valorisation d'options, en particulier, quand il est question d'options exotiques qui disposent de quelques caractéristiques compliquées auxquelles il faut faire face. Elles offrent par exemple la possibilité de traiter de manière efficace des problèmes en dimension supérieure qu'elle soit en termes de paramètre de temps, lors de la valorisation d'option 'path-dependent', ou en termes de variables d'états (sous-jacent) lors de la valorisation de diverses variantes d'options panier... De plus, il y a une relative simplicité dans l'implémentation en comparaison avec les autres approches ainsi qu'un ordre de convergence qui est indépendant de la dimension du problème traité et qui est égal à $O(n^{-1/2})$ où n est le nombre de simulations effectuées.

5.1.1 Les simulations de Monte Carlo dans le cadre de Black & Scholes

Le cadre de Black & Scholes qui met les fondements pour des modèles plus complexes se concentre sur une option d'achat standard pour une seule variable d'état. Ce cadre suppose que le marché a

un taux d'intérêt sans risque constant, se comporte sans imperfections, sans tensions, est ouvert en continu, et dans lequel il n'existe ni taxes, ni coûts de transactions. Il est également basé sur le principe d'absence d'arbitrage qui stipule que deux portefeuilles d'actifs financiers ayant un même payoff doivent avoir un même prix à tout moment et dans tous les scénarios possibles. Dans le cas contraire, un bénéfice sûr peut être dégagé sans aucun investissement.

Dans le modèle de Black & Scholes, l'actif sous-jacent est représenté par le processus $\{X_t, 0 \leq t \leq T\}$ décrit par l'EDS suivante

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (5.1)$$

avec (W_t) un mouvement brownien standard et les coefficients μ et σ deux constantes qui satisfont quelques conditions habituelles assurant l'existence et l'unicité d'une solution continue et adaptée pour l'équation (5.1) (Karatzas et Shreve [65]).

Selon le théorème de Feynman-Kac (annexe (A5.1)) et sous des conditions très générales, le prix sans arbitrage de la majorité des options vanilles ou exotiques (Européennes, américaines, asiatiques, ...) peut être exprimé comme l'espérance actualisée au taux sans risque du payoff, associé à l'option, (Glasserman [46]; Harrison et Kreps [55]; Harrison et Pliska [56]) souvent défini comme une fonctionnelle du processus de l'actif sous-jacent. Les prix actualisés étant sans dérives, sont alors des martingales (Duffie [36]; Karatzas et Shreve [65]).

Afin de formaliser ceci, le prix $p(t)$ d'une option dont le payoff est défini par la fonction Φ et impliquant les instants (t_1, t_2, \dots, t_n) est donné par

$$p(t) = E^Q [\Phi(X_{t_1}, X_{t_2}, \dots, X_{t_n}) | X(0) = x] \quad (5.2)$$

notons que $0 < t_1 \leq t_2 \leq \dots \leq t_n = T$ où T est le temps d'arrêt qui correspond à la date d'expiration de l'option.

Maintenant, à partir de l'équation (5.2) il est clair que l'utilisation d'une approche Monte Carlo est un choix naturel si l'on veut calculer une valeur pour $p(t)$ et la manière la plus directe de le faire se présente comme suit :

1. Simuler des trajectoires jusqu'à l'échéance pour le prix de l'actif sous-jacent.
2. Evaluer le payoff actualisé de l'option pour chaque trajectoire, tel que c'est spécifié par la nature du produit en question.
3. Calculer la moyenne de tous les payoffs actualisés par rapport au nombre des trajectoires simulées.

Par ailleurs, cette approche visiblement simple permet l'évaluation d'intégrales multidimensionnelles, ainsi elle peut très bien être adaptée et appliquée à un grand éventail de produits dérivés, par exemple, le call asiatique, le call américain et le call panier dont les prix respectifs sont donnés par $E \left[e^{-rT} \left(\frac{1}{T} \int_0^T S_t dt - K \right)^+ \right]$, $\max_{\tau} E \left[e^{-rT} (S_t - K)^+ \right]$ et $E \left[e^{-rT} (\sum \omega_i S_i - K)^+ \right]$ où r est le taux d'intérêt sans risque, T est la maturité de l'option, τ est la durée de temps jusqu'à la date d'exercice de l'option américaine, S_t est le prix de l'actif sous-jacent au temps t et ω_i est le poids du sous-jacent i dans le panier considéré.

Cependant, comme il a été indiqué plus haut, l'approche Monte Carlo manifeste un grand besoin pour un nombre élevé de simulations. Cet inconvénient peut, toutefois, être adressé à travers les techniques de réduction de la variance dont quelques-unes seront discutées ci-dessous.

5.1.2 Réduction de la variance

La réduction de la variance est un élément important lorsqu'on désire effectuer des simulations de Monte Carlo étant donné qu'elle améliore considérablement l'efficacité des calculs (Glynn et Whitt [48]). Dans ce qui suit, on présente quelques techniques de réduction de la variance en indiquant comment elles pourraient être utilisées dans un contexte financier.

Variabes antithétiques

La méthode des variables antithétiques est probablement la méthode la plus directe à adopter quand on désire réduire la variance. De plus, elle est relativement simple à implémenter et se combine facilement aux autres techniques.

L'idée de base est la suivante : si l'on considère deux variables aléatoires X_1 et X_2 négativement corrélées, alors la variance de leur somme $Var(X_1 + X_2)$ serait plus petite que si elles avaient été indépendantes.

Afin d'illustrer l'utilisation de cette technique en finance, considérons la formule

$$X(t) = X(0) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right) \quad (5.3)$$

qui est la solution analytique de l'EDS donnée par l'équation (5.1).

Ceci suggère que pour l'implémentation de chaque scénario du prix de l'actif, il faudra faire appel aux deux équations suivantes :

$$X(t + \Delta t) = X(t) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\epsilon\sqrt{\Delta t}\right) \quad (5.4)$$

$$X(t + \Delta t) = X(t) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta t - \sigma\epsilon\sqrt{\Delta t}\right) \quad (5.5)$$

où $\epsilon \sim N(0, 1)$.

Les $X(t + \Delta t)$ calculés à partir des équations (5.4) et (5.5) sont considérées comme des variables antithétiques puisque, clairement, elles sont négativement corrélées. C'est pourquoi, la valorisation d'une option à travers ces variables peut se faire en considérant le payoff de chaque réplication comme la moyenne des payoff des deux scénarios. Ceci accélère la convergence vers le prix réel de l'option. Plus de détails sur cette méthode sont données dans Clewlow et Carverhill [25].

Variabes de contrôle

Cette méthode largement utilisée est basée sur l'introduction d'une variable de contrôle positivement corrélée avec la quantité que l'on désire estimer.

Soient X et Y deux variables aléatoires corrélées, avec Y appelée variable de contrôle de X . En considérant que l'espérance de Y est connue, une nouvelle variable \tilde{X} est donnée par

$$\tilde{X} = X + \beta(Y - E[Y]) \quad (5.6)$$

et comme $E[\tilde{X}] = E[X]$, une estimation de X peut être effectuée à partir de \tilde{X} .

La technique est principalement centrée sur le choix d'un paramètre β qui minimise la variance de \tilde{X}

$$Var[\tilde{X}] = Var[x] + \beta^2 Var[Y] - 2\beta Cov[X, Y] \quad (5.7)$$

un β qui minimise l'équation (5.7) est alors donnée par

$$\beta^* = \frac{Cov[X, Y]}{Var[Y]} \quad (5.8)$$

et la variance de \tilde{X} devient alors

$$Var[\tilde{X}] = Var[X] - \frac{Cov[X, Y]^2}{Var[Y]} \quad (5.9)$$

ce qui implique directement que $Var[\tilde{X}] \leq Var[X]$.

Notons également que cette technique peut facilement être étendue à plusieurs variables de contrôle. Ce qui donne

$$\tilde{X} = X + \underline{\beta}^T (\underline{Y} - E_Y) \quad (5.10)$$

avec $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_n)^T$ le vecteur des coefficients, $\underline{Y} = (Y_1, Y_2, \dots, Y_n)^T$ le vecteur des variables de contrôle et $E_Y = (E[Y_1], E[Y_2], \dots, E[Y_n])^T$ le vecteur des espérances des Y_i , avec $i = 1, 2, \dots, n$.

Une application en finance peut être présentée à travers l'exemple de la valorisation d'une option asiatique. Soit V_{art} le prix d'une option asiatique arithmétique et V_{geo} le prix d'une option asiatique géométrique. Ces deux options sont supposées avoir les mêmes spécifications mis à part la façon dont le payoff est calculé. En pratique, la majorité des payoffs des options asiatiques sont calculées par moyenne arithmétique. Cependant, les prix d'options asiatiques arithmétiques ne peuvent pas être évalués par une formule fermée, en revanche les prix d'options asiatiques géométriques le peuvent. Ce contraste peut être exploité par la méthode des variables de contrôle pour calculer V_{art} . On peut donc écrire

$$\tilde{V}_{art} = \hat{V}_{art} - \beta (\hat{V}_{geo} - V_{geo}) \quad (5.11)$$

où \tilde{V}_{art} est le prix simulé de l'option asiatique arithmétique considérée, \hat{V}_{geo} est le prix simulé de l'option asiatique géométrique considérée et V_{geo} est le prix donnée par la solution analytique.

Et comme la corrélation entre \hat{V}_{art} et sa variable de contrôle est forte, le paramètre β est considéré égal à 1 Ce qui entraîne

$$\tilde{V}_{art} = V_{geo} + (\hat{V}_{art} - \hat{V}_{geo}) \quad (5.12)$$

L'équation (5.12) permet alors de calculer le prix d'une option asiatique en réduisant la variance de la simulation. Une analyse en profondeur de cette méthode est donnée dans Broadie et Glasserman [19] et Glasserman [46].

Échantillonnage préférentiel

L'échantillonnage préférentiel, s'appuie sur une idée connue en mathématiques financières qui est l'utilisation du 'ratio de vraisemblance' aussi connu comme 'la dérivée de Radon Nikodym'. Cette idée permet de changer la mesure de probabilité dans le but de représenter l'espérance sous une certaine mesure en tant qu'une espérance sous une autre mesure qui pourra faire ressortir les événements pertinents. En particulier, cette technique permet d'éviter le besoin d'avoir un très grand nombre de réplifications pour simuler des événements rares ou inhabituels.

Soit une variable aléatoire X de densité f , on désire estimer

$$\theta = E_f [h(x)] = \int h(x) f(x) dx \quad (5.13)$$

Supposons que l'on dispose d'une autre fonction de densité g , l'équation (5.13) peut alors être réécrite sous la forme

$$\theta = \int \frac{h(x)}{g(x)} f(x) g(x) dx = E_g \left[\frac{h(x) f(x)}{g(x)} \right] \quad (5.14)$$

le ratio de vraisemblance est alors la quantité $(f(x)/g(x))$ qui aide à corriger le changement dans la mesure de probabilité.

Pour expliquer comment ce changement de mesure pourrait être utile dans la réduction de la variance, il faut d'abord remarquer que les deux estimateurs de θ , donnés par les équations (5.13) et (5.14) ont la même espérance, leurs variances quant à elles sont

$$Var_f [h(x)] = \int h^2(x) f(x) dx - \theta^2 \quad (5.15)$$

$$Var_g [h^*(x)] = \int h^{*2}(x) g(x) dx - \theta^2 \quad (5.16)$$

avec

$$h^*(x) = \frac{h(x) f(x)}{g(x)} \quad (5.17)$$

Le choix de g doit alors être fait de sorte à réduire la variance. L'équation (5.16) indique clairement que si

$$g(x) = \frac{h(x)f(x)}{\theta} \quad (5.18)$$

alors on trouve que $Var_g[h^*(x)] = 0$.

Pourtant, l'utilisation de l'équation (5.18) nécessite la connaissance de θ qui nous est inconnue. Mais, tout de même, une approximation de la densité idéale g peut conduire vers une réduction optimale de la variance. Plus de détails sont donnés dans Brandimarte [17].

Sur les marchés financiers, cette approche peut être adoptée lorsqu'il est question de valoriser des options hors de la monnaie (dont l'exercice ne rapporte aucun gain). En effet, si l'on utilise une simulation de Monte Carlo directe, pour une option hors de la monnaie, alors une grande majorité des trajectoires simulées ne seraient par efficaces pour la valorisation de l'option puisqu'elles vont conduire vers une valeur nulle. Par conséquent, il semble beaucoup plus pratique d'utiliser des trajectoires utiles, c'est à dire celles qui offrent un payoff strictement positif.

Par exemple, si l'on connaît f la densité du sous-jacent à l'échéance et l'on peut analytiquement identifier g la probabilité que le prix du sous-jacent soit supérieur au strike K à l'instant T , on pourra alors simuler des trajectoires à partir de $g = f/q$ au lieu de f . Ceci permettra de mettre l'accent sur les scénarios qui génèrent un prix du sous-jacent supérieur au prix d'exercice. Un exemple plus détaillé est donné dans Glasserman, Heidelberger et Shahabuddin [47].

5.2 Une approche de Monte Carlo couplée à l'estimation de densité

Dans cette section, nous présentons une modification de l'approche Monte Carlo classique dont la finalité est d'améliorer la précision et d'augmenter la vitesse de convergence. L'approche suggérée se base sur le théorème de représentation de Feynman-Kac et s'inspire de l'échantillonnage préférentiel dans le sens où elle recourt principalement à un rapport de densités de probabilités pour introduire des pondérations d'ajustement qui seront calculées par une estimation de densité.

Une formule générale pour la valorisation d'options

La description présentée dans ce qui suit, constitue une approche générale qui pourrait être adaptée pour la valorisation d'une large variété d'options puisqu'elle ne considère aucun payoff spécifique, mais essaye plutôt d'évaluer l'espérance d'une fonctionnelle quelconque dépendante d'un processus connu, chose qui permet de couvrir divers types de payoffs. Par ailleurs, malgré que l'on se concentre sur les options à sous-jacent unique dans la présentation, la méthode peut facilement être généralisée à des options à multiples sous-jacents.

Soient X_i ($1 \leq i \leq n$) des tirages indépendants effectués selon une loi de probabilité inconnue de support positif $[a, b]$. Le but est d'utiliser ces tirages pour calculer sur $[a, b]$ l'espérance $E[\Phi(Y)]$ où Φ est une fonction continue sur $[a, b]$ et Y une variable aléatoire de densité de probabilité connue. En théorie, une première approximation peut-être donnée par

$$E[\Phi(Y)] \leftarrow \frac{1}{n} \sum_{i=1}^n \Phi(X_i) g(X_i) \frac{1}{f(X_i)} ; n \rightarrow \infty \quad (5.19)$$

où g et f sont respectivement les fonctions de densités de probabilités des variables aléatoires Y et X . Comme mentionné plus haut, la fonction f étant inconnue rend l'équation (5.19) inexploi-

table et soulève un problème d'estimation de densité qui sera adressé par un estimateur à noyau. L'introduction de ce dernier donne une autre approximation de $E[\Phi(Y)]$

$$E[\Phi(Y)] \leftarrow \sum_{i=1}^n \frac{\Phi(X_i)g(X_i)}{\sum_{j=1}^n K_h(X_i - X_j)} ; n \rightarrow \infty \quad (5.20)$$

avec $K_h(X_i - X_j) = (1/h)K[(X_i - X_j)/h]$. En outre, pour rester dans le cadre de Black & Scholes, la fonction g peut être considérée comme la fonction de densité de la loi log-normale de paramètres μ et σ pouvant être estimés à partir des données.

$$g(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\text{Log}(x) - \mu)^2}{2\sigma^2}\right) \quad (5.21)$$

5.3 Résultats numériques

L'objectif principal des expériences numériques menées ci-après est d'évaluer l'efficacité de la formule proposée lors de la valorisation d'options, ainsi que d'établir s'il existe une quelconque amélioration par rapport à l'approche Monte Carlo classique en terme de précision et de vitesse de convergence.

L'intérêt est porté sur la valorisation d'un call asiatique, la fonction Φ sera alors le payoff de cette option qui est égal à $\left(\frac{1}{T} \int_0^T S_t dt - K\right)^+$. De plus, l'estimation de densité se fera à travers divers noyaux afin d'identifier l'effet de chacun d'eux sur les performances de la formule. Les différents noyaux utilisés seront alors :

- Le noyau gaussien qui correspond à la fonction de densité de la loi normale centrée réduite.
- Le noyau gamma (équation (2.2)).
- Le noyau d'Hermite (équation (1.28)).
- Le noyau de Laguerre (équation (1.29)).

Il convient également de noter que le noyau gaussien peut être considéré comme une référence pour évaluer la qualité des résultats obtenus par l'utilisation du noyau gamma et des noyaux à base polynômes orthogonaux.

Valorisation

A ce stade, on cherche à valoriser trois options asiatiques ayant les mêmes volatilités et échéances mais des prix d'exercice différents. Les prix calculés seront comparés à un prix cible qui est le résultat obtenu par Dubois et Lelièvre [34] lorsque l'option a été valorisée par une solution numérique de l'EDP proposé par Rogers et Shi [87].

Les résultats donnés par l'équation (5.20) combinée aux différents noyaux ainsi que le prix obtenu par la méthode Monte Carlo classique sont présentés dans les figures (5.1) à (5.6).

De manière générale, on constate que pour l'ensemble des cas considérés, la méthode Monte Carlo classique montre une surévaluation du prix cible et ne donne aucun signe d'amélioration malgré une augmentation considérable du nombre de simulations. Ensuite, pour la méthode proposée (équation 130), on remarque que les approximations obtenues commencent loin du résultat attendu et présentent un comportement instable et des fluctuations élevées pour un nombre de simulations inférieur à 5000. Néanmoins, avec l'augmentation de ce dernier le résultat commence à montrer une

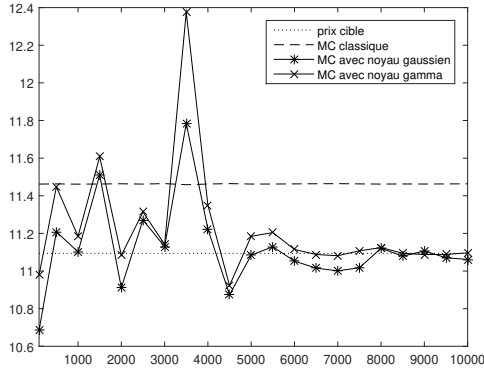


FIGURE 5.1 – Valeurs obtenue pour le prix d’un call asiatique arithmétique par MC classique, MC combiné à un noyau gaussien et MC combiné à un noyau gamma avec $r = 0.15$, $\sigma = 0.05$, $K = 95$, $S_0 = 100$.

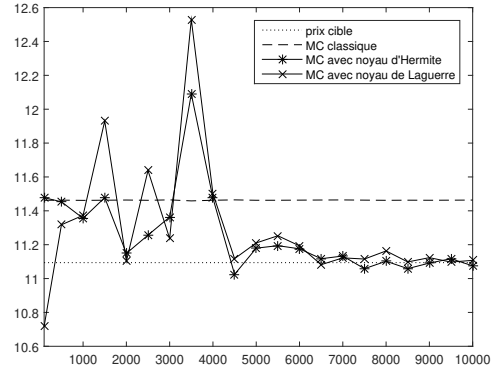


FIGURE 5.2 – Valeurs obtenue pour le prix d’un call asiatique arithmétique par MC classique, MC combiné à un noyau d’Hermite et MC combiné à un noyau de Laguerre avec $r = 0.15$, $\sigma = 0.05$, $K = 95$, $S_0 = 100$.

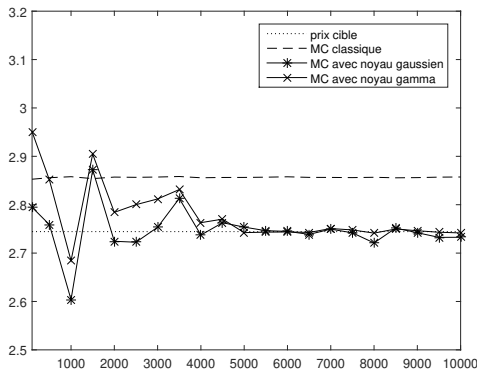


FIGURE 5.3 – Valeurs obtenue pour le prix d’un call asiatique arithmétique par MC classique, MC combiné à un noyau gaussien et MC combiné à un noyau gamma avec $r = 0.15$, $\sigma = 0.05$, $K = 105$, $S_0 = 100$.

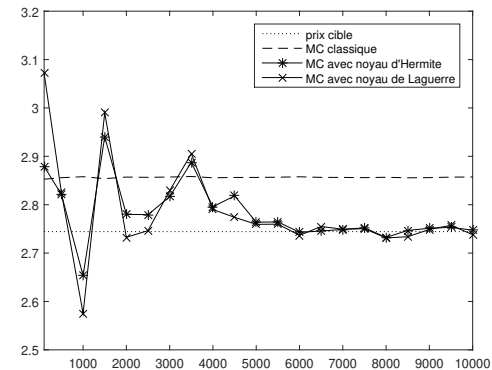


FIGURE 5.4 – Valeurs obtenue pour le prix d’un call asiatique arithmétique par MC classique, MC combiné à un noyau d’Hermite et MC combiné à un noyau de Laguerre avec $r = 0.15$, $\sigma = 0.05$, $K = 105$, $S_0 = 100$.

convergence graduelle et relativement stable vers le prix cible pour finalement s’aligner sur ce dernier mais avec un coût de calcul élevé.

Concernant les comportements, il n’existe pas de différences frappantes. Il semble en fait que les approximations à base de différents noyaux reproduisent à peu près les mêmes variations mais avec différents degrés. En effet, en dessous des 5000 simulations les fluctuations des approximations impliquant des noyaux à base de polynômes orthogonaux ont des magnitudes beaucoup plus importantes que pour celles avec les noyaux gaussien et gamma. Par contre, lorsque le nombre de simulations passe au-dessus de 5000, ces magnitudes se réduisent et les approximations à base des noyaux Hermite, Laguerre et gamma montrent, autour de la valeur cible, moins de mouvements que le noyau gaussien.

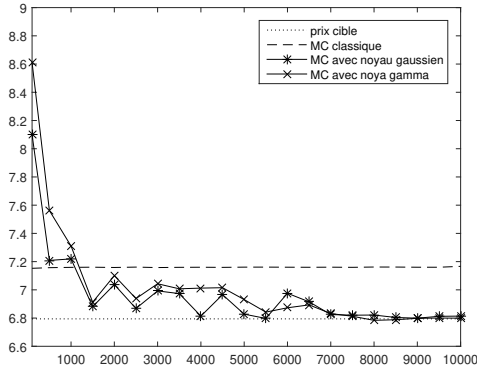


FIGURE 5.5 – Valeurs obtenue pour le prix d’un call asiatique arithmétique par MC classique, MC combiné à un noyau gaussien et MC combiné à un noyau gamma avec $r = 0.15$, $\sigma = 0.05$, $K = 100$, $S_0 = 100$.

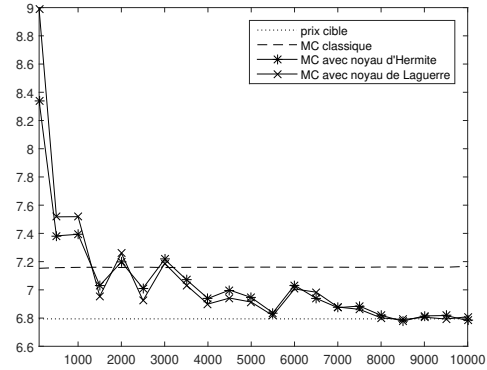


FIGURE 5.6 – Valeurs obtenue pour le prix d’un call asiatique arithmétique par MC classique, MC combiné à un noyau d’Hermite et MC combiné à un noyau de Laguerre avec $r = 0.15$, $\sigma = 0.05$, $K = 100$, $S_0 = 100$.

5.4 Conclusion

Ce chapitre s’est intéressé à l’amélioration de l’approche classique de Monte Carlo pour la valorisation d’options dans le cadre du formalisme de Black & Scholes. La modification suggérée s’inspire de l’échantillonnage préférentiel et est d’ordre général dans le sens où elle reste toujours dans le contexte du théorème de représentation de Feynman-Kac par lequel on cherche à calculer l’espérance du payoff et peut donc être appliquée à un large éventail d’option sans nécessiter d’attention particulière.

La méthode cherche principalement à ajuster le calcul rudimentaire de la moyenne des payoffs effectué par MC Classique par l’introduction de pondérations qui nécessitent l’utilisation de deux densités. L’une analytique connue et l’autre inconnue mais estimée empiriquement à travers un estimateur à noyau. Lors de cette estimation on a fait appel à des noyaux à base de densités de probabilités ainsi qu’à d’autres construits à base de polynômes orthogonaux pour des fins de comparaisons. En effet, un autre but était d’observer le comportement de l’approximation pour des noyaux de formes différentes.

Cette nouvelle approche a été testée sur une option d’achat asiatique, laquelle a été d’abord valorisée par une méthode de Monte Carlo classique avec une réduction de la variance. Pour un nombre réduit de simulations, et contrairement à la méthode classique, la formule proposée a affiché un comportement instable et incohérent quelque soit le noyau utilisé, néanmoins, la qualité de l’approximation s’est améliorée et a commencé à converger vers la cible en augmentant le nombre de simulation tandis que le prix calculé par la méthode classique est resté insensible à ce changement, chose qui l’a empêché d’atteindre le prix recherché.

En termes de qualité, la formule proposée montre donc une meilleure performance par rapport à l’approche Monte Carlo classique à condition d’avoir un nombre de simulations élevé (plus de 10000 répliquations). Concernant, la comparaison entre les différents noyaux utilisés et leur effet sur la valorisation de l’option, on peut remarquer que l’utilisation du noyau gamma et des noyaux à base de polynômes orthogonaux, paraît légèrement préférable puisqu’ils montrent une certaine consistance dans la convergence vers la cible. Ceci, suggère également que des ordres plus élevés pour les noyaux d’Hermite et de Laguerre pourraient entraîner une convergence plus rapide et réduire les fluctuations.

Conclusion générale

Hormis les cas les plus simples, très peu de problèmes en finance quantitatives disposent de solutions exactes et efficaces, chose qui constitue un sérieux obstacle lors de l'étude de marchés financiers sophistiqués. Cette complication est principalement due à bon nombre de paramètres qui stimulent la nature aléatoire des marchés, dont la diversité des intervenants et la complexité des produits et de leurs sous-jacents.

Deux des problèmes les plus intéressants sur les marchés sont, la compréhension de la volatilité vue l'influence qu'elle exerce sur le comportement des intervenants sur les marchés et la valorisation d'options exotiques vue la part du marché qu'elles occupent et l'importance que représente la détermination du prix pour les spéculateurs et les opérateurs de couverture afin de prendre les bonnes décisions et garantir un certain équilibre dans leurs portefeuilles. Un point commun à ces problématiques, est la difficulté d'identifier une loi de probabilité qui puisse régir l'évolution des processus de prix ou de volatilité.

Dans cette thèse, nous nous sommes intéressés aux méthodes de lissage par noyaux, en particulier, l'estimation de densité et la régression, le but étant de tenter d'exploiter la puissance offerte par ces méthodes qui réside, entre autres, dans la faiblesse des hypothèses qui leurs sont inhérentes afin d'essayer d'aborder, par des approches non classiques, le problème de valorisation d'option et celui de l'estimation de la densité de la volatilité. Nous avons également porté une certaine attention aux noyaux en proposant la construction de noyaux non conventionnels d'un ordre quelconque à base de polynômes orthogonaux dont l'utilité principale est de contourner le dilemme biais-variance en réduisant le biais mais sans avoir d'effet sur la variance. Un autre moyen de réduire le biais de bords cette fois-ci, est l'utilisation des estimateurs à base de noyaux asymétrique lesquels ont été abordés avec grand détail pour mettre en valeur leur utilité lorsqu'ils sont confrontés aux noyaux symétriques classiques.

L'introduction de noyaux qui permettent d'adresser le problème du biais sans affecter la variance allège un peu la tension exercée sur le paramètre de lissage vu que lors de l'utilisation de noyaux classiques il est le seul responsable sur l'équilibre entre ces deux sources d'erreur. Toutefois cet équilibre n'est jamais absolu, car d'une part, la valeur du paramètre 'optimal' va souvent changer d'une méthode de sélection à une autre et d'autre part, il existe un sérieux problème d'optimalité asymptotique dans la logique de sélection actuelle.

Par ailleurs, les travaux menés dans le cadre de cette thèse ont permis de proposer un certain nombre de contributions dont la première est au niveau des méthodes de lissage par noyau en proposant une famille de noyaux d'ordre quelconque en plus d'une méthode d'insertion simplifiée pour la sélection du paramètre de lissage. La deuxième est la suggestion d'une approche qui remplace efficacement la méthode actuelle de référence basée sur la théorie de déconvolution pour l'estimation de la densité de la volatilité et qui s'aligne presque parfaitement sur le benchmark théorique. La troisième, est la suggestion d'une nouvelle formule à base de la méthode Monte Carlo pour la valorisation d'options qui améliore de façon significative les résultats obtenus par la méthode Monte Carlo classique. La quatrième contribution est d'ordre comparatif, en effet elle a permis de comparer les performances de plusieurs noyaux pour aboutir à la conclusion que pour nos applications, les noyaux usuels sont moins efficaces que les noyaux symétriques d'ordre supérieurs tandis que ceux-ci sont aussi bons que les noyaux asymétriques. Il s'ensuit que puisque la théorie des noyaux asymétriques n'est pas aussi développées que celle des noyaux symétriques, il serait préférable, pour les applications immédiates, d'utiliser ces derniers que l'on comprend particulièrement bien mais en gardant toutefois un ordre qui ne soit pas trop élevé afin de préserver un certain équilibre entre les performances et le coût en terme de temps de calculs.

Pour conclure, cette thèse ne constitue qu'un premier pas dans l'exploration du potentiel énorme offert par la souplesse et la flexibilité des méthodes de lissage pour la résolution de problèmes tant en

finance quantitative qu'en d'autres domaines. De plus, avec la grande popularité que commencent à prendre les statistiques dans différentes disciplines avec l'avènement des sciences de données et des méthodes d'apprentissage machine, les applications potentielles n'en sont que plus nombreuses et intéressantes dont quelques unes qui sont en cours et d'autres que nous projetons d'entamer. En effet, toujours en finance quantitative, nous nous penchons actuellement sur des extensions immédiates de ce travail qui sont la valorisation d'options panier et d'options sur volatilité, ensuite on désire aborder la problématique de la valorisation des options climatiques pour lesquels la connaissance des distributions de probabilité des phénomènes naturels est primordiale. En apprentissage machine, les forêts d'arbres décisionnels semblent être un terrain propice au développement à travers l'intégration des estimateurs de densité et de régression. Sur le plan théorique, il semble intéressant d'étudier les propriétés minimax des estimateurs ainsi que de conduire une analyse de Fourier pour identifier une forme différente De l'erreur globale qui pourrait donner davantage de renseignement sur le comportement de l'estimateur en question.

Annexes

Les notations de chaque annexe correspondent aux notations du chapitre auquel elle se réfère.

A1.1

Sous réserve d'existence, le moment d'ordre r d'une distribution de probabilité de densité f est donné par

$$m_r(f) = \int x^r f(x) dx$$

en l'occurrence, l'existence d'un moment d'ordre r est équivalent à l'existence des deux intégrales finies $\int_{-\infty}^0 x^r f(x) dx$ et $\int_0^{\infty} x^r f(x) dx$.

Si l'on suppose que pour r impair on a $\int_0^{\infty} x^r f(x) dx = A$ et que l'on pose $x = -u$ on trouve

$$\int_{-\infty}^0 x^r f(x) dx = \int_{\infty}^0 (-u)^r f(-u) (-du) = (-1)^r \int_0^{\infty} (u^r) f(-u) du = - \int_0^{\infty} (u^r) f(u) du = -A$$

d'où

$$\int x^r f(x) dx = 0$$

A1.2

Soient $\rho, \lambda \in \mathbb{R}$ et $r = \lfloor \rho \rfloor$.

On suppose que la densité f appartient à $\mathcal{H}(\rho, \lambda)$ qui est l'ensemble des fonctions définies sur un intervalle $I \subset \mathbb{R}$, r fois dérivable et dont la dérivée satisfait la condition de Hölder

$$|f^{(r)}(x) - f^{(r)}(y)| \leq \lambda |x - y|^{\rho - r} ; \forall x, y \in I$$

Maintenant de manière similaire à ce qui a été fait aux sections (1.2) et (1.3), on a

$$\text{Biais} [\hat{f}_n(x)] = \int K(u) [f(x - uh) - f(x)] du$$

par Taylor-Lagrange cette fois-ci (au Lieu de Taylor-Young) on trouve

$$\text{Biais} [\hat{f}_n(x)] = \int K(u) \frac{(uh)^r}{r!} f^{(r)}(x - \theta uh) du ; \theta \in]0, 1[$$

la définition (1.4) permet d'écrire

$$\text{Biais} [\hat{f}_n(x)] = \int K(u) \frac{(uh)^r}{r!} [f^r(x - \theta uh) - f^r(x)] du$$

et la condition de Hölder donne

$$|\text{Biais} [\hat{f}_n(x)]| \leq \lambda \int |K(u)| \frac{|uh|^r}{r!} |-\theta uh|^{\rho-r} du$$

A1.3

Soient

- H un espace de Hilbert de produit scalaire $\langle \cdot | \cdot \rangle$
- $\{g_1, g_2, \dots, g_n\}$ un système d'éléments linéairement indépendants de H

Quelques définitions

- Un système $\{g_1, g_2, \dots, g_n\}$ fini ou non, est dit orthonormé si et seulement si

$$\langle g_i | g_j \rangle = \delta_{i,j} ; 1 \leq i, j \leq n$$

- Étant donné un intervalle $I \subset \mathbb{R}$, une fonction poids ω sur I , on définit le produit scalaire

$$\langle f | g \rangle = \int_I f(x) g(x) \omega(x) dx$$

- On peut également définir un produit scalaire par une intégrale de Stieltjes

$$\langle f | g \rangle = \int_I f(x) g(x) d\alpha(x)$$

où α est une fonction croissante possédant une infinité de points de croissance.

- Pour l'étude de fonctions à valeurs discrètes, le produit scalaire devient

$$\langle f | g \rangle = \sum_{i=1}^n \omega_i f(x_i) g(x_i)$$

Quelques formules

Formule de Christoffel

- Soit $\{P_n(x)\}_{n \geq 1}$ une famille orthonormée de polynômes associée à la distribution $w(x) dx$ sur $[a, b]$.
- Soit $\rho(x) = c(x - x_1) \dots (x - x_l)$ un polynôme de degré inférieur ou égal à l sur $[a, b]$ avec $c \neq 0$ et $x_i \neq x_j, \forall i \neq j$.

Alors les polynômes orthogonaux $\{q_n(x)\}_{n \in \mathbb{N}}$ associés à la distribution $\rho(x) w(x) dx$ peuvent se représenter comme suit :

$$q_n(x) = \frac{1}{\rho(x)} \begin{vmatrix} P_n(x) & P_{n+1}(x) & \dots & P_{n+l}(x) \\ P_n(x_1) & P_{n+1}(x_1) & \dots & P_{n+l}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ P_n(x_l) & P_{n+1}(x_l) & \dots & P_{n+l}(x_l) \end{vmatrix}$$

Formule de récurrence

Soit $\{P_n(x)\}_{n \geq 0}$ un système de polynômes orthogonaux. On a alors la formule de récurrence qui permet de calculer P_{n+1} à partir de P_n et P_{n-1}

$$P_{n+1}(x) = (A_n x + B_n) P_n(x) - C_n P_{n-1}(x) ; n \geq 1$$

où A_n, B_n, C_n sont les constantes suivantes

$$A_n = \frac{k_{n+1}}{k_n} ; B_n = -A_n \frac{\langle x P_n | P_n \rangle}{\langle P_n | P_n \rangle} ; C_n = A_n \frac{\langle x P_n | P_{n-1} \rangle}{\langle P_{n-1} | P_{n-1} \rangle}$$

avec k_n le coefficient directeur de P_n .

Formule de Christoffel-Darboux

Pour un système orthogonal $\{P_n(x)\}_{n \geq 0}$ de polynômes, on a les formules pour $x \neq y$ dans $[a, b]$

$$\sum_{\nu=0}^n \frac{P_\nu(x) P_\nu(y)}{h_\nu} = \frac{k_n}{k_{n+1} h_n} \frac{P_{n+1}(x) P_n(y) - P_n(x) P_{n+1}(y)}{x - y}$$

$$\sum_{\nu=0}^n \frac{P_\nu^2(x)}{h_\nu} = \frac{k_n}{k_{n+1} h_n} (P_{n+1}'(x) P_n(x) - P_n'(x) P_{n+1}(x))$$

avec k_n le coefficient directeur de P_n et $h_n = \langle P_n | P_n \rangle$.

A1.4

Dans ce qui suit, on utilise la notation compacte $L_\beta(u) = (1/\beta) L(u/\beta)$.

On désire donner une approximation asymptotique pour l'erreur quadratique moyenne de l'estimateur

$$\begin{aligned}\hat{\Phi}_r &= \frac{1}{n} \sum_{i=1}^n \hat{f}_n^{(r)}(x_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n L_\beta^{(r)}(X_i - X_j)\end{aligned}$$

L'approximation se dérive sous les hypothèses suivantes

— L est un noyau symétrique d'ordre k pair, r fois dérivable tel que

$$(-1)^{(r+k)/2+1} L^{(r)}(0) \mu_k(L) > 0$$

— La densité f dispose de p dérivées continues et absolument monotone, avec $p > k$.

— $\beta = \beta_n$ une suite positive de paramètres de lissage telle que

$$\lim_{n \rightarrow \infty} \beta = 0 ; \quad \lim_{n \rightarrow \infty} n\beta^{2r+1} = \infty$$

Biais de $\hat{\Phi}_r$

Il est clair que l'estimateur s'écrit sous la forme

$$\hat{\Phi}_r = \frac{1}{n} L_\beta^{(r)}(0) + \frac{1}{n^2} \sum_i \sum_{j \neq i} L_\beta^{(r)}(X_i - X_j)$$

son espérance est donc

$$E[\hat{\Phi}_r] = \frac{1}{n} L_\beta^{(r)}(0) + \left(1 - \frac{1}{n}\right) E[L_\beta^{(r)}(X_1 - X_2)]$$

par la formule de Taylor et un changement de variable similaire à celui effectué aux sections (1.2) et (1.3), on trouve

$$\begin{aligned}E[L_\beta^{(r)}(X_1 - X_2)] &= \iint L_\beta^{(r)}(x - y) f(x) f(y) dx dy \\ &= \iint L_\beta(x - y) f(x) f^{(r)}(y) dx dy \\ &= \iint L(u) f(y + \beta u) f^{(r)}(y) du dy \\ &\approx \iint L(u) f^{(r)}(y) \left\{ \sum_{l=0}^k (l!)^{-1} (u\beta)^l f^{(l)}(y) \right\} du dy \\ &\approx \Phi_r + (k!)^{-1} \mu_k(L) \beta^k \Phi_{r+k}\end{aligned}$$

une approximation du biais est alors donné par

$$\text{Biais} [\hat{\Phi}_r] \approx \frac{1}{n\beta^{r+1}} L^{(r)}(0) + (k!)^{-1} \beta^k \mu_k(L) \Phi_{r+k}$$

Variance de $\hat{\Phi}_r$

Soient X_1, X_2, \dots, X_n des v.a.r *i.i.d* et S une fonction symétrique par rapport à zéro. Si

$$U = \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n S(X_i - X_j)$$

on a alors

$$\text{Var}[U] = \frac{2}{n^3} (n-1) \text{Var}[S(X_1 - X_2)] + \frac{4}{n^3} (n-1)(n-2) \text{Cov}[S(X_1 - X_2), S(X_2 - X_3)]$$

ceci implique que

$$\begin{aligned} \text{Var}[\hat{\Phi}_r] &= \frac{2}{n^3} (n-1) \text{Var}[L_\beta^{(r)}(X_1 - X_2)] + \frac{4}{n^3} (n-1)(n-2) \\ &\quad \times \text{Cov}[L_\beta^{(r)}(X_1 - X_2), L_\beta^{(r)}(X_2 - X_3)] \end{aligned}$$

Calculons d'abord $E\left[\left(L_\beta^{(r)}(X_1 - X_2)\right)^2\right]$

$$\begin{aligned} E\left[\left(L_\beta^{(r)}(X_1 - X_2)\right)^2\right] &= \iint \left[L_\beta^{(r)}(x - y)\right]^2 f(x) f(y) dx dy \\ &= \frac{1}{\beta^{2r+1}} \iint L^{(r)}(u)^2 f(y + \beta u) f(y) du dy \\ &\approx \frac{1}{\beta^{2r+1}} \Phi_0 R(L^{(r)}) \end{aligned}$$

ensuite $E\left[L_\beta^{(r)}(X_1 - X_2) L_\beta^{(r)}(X_2 - X_3)\right]$

$$\begin{aligned} E\left[L_\beta^{(r)}(X_1 - X_2) L_\beta^{(r)}(X_2 - X_3)\right] &= \iiint L_\beta^{(r)}(x - y) L_\beta^{(r)}(y - z) f(x) f(y) f(z) dx dy dz \\ &= \iiint L_\beta(x - y) L_\beta(y - z) f^{(r)}(x) f(y) f^{(r)}(z) dx dy dz \\ &= \iiint L(u) L(v) f^{(r)}(y + gu) f(y) f^{(r)}(y - gv) du dv dy \\ &\approx \int f^{(r)}(y)^2 f(y) dy \end{aligned}$$

on a aussi

$$E \left[L_{\beta}^{(r)} (X_1 - X_2) \right] \approx \Phi_r$$

par conséquent, on trouve

$$Var [\Phi_r] \approx \frac{2}{n^2 \beta^{2r+1}} \Phi_0 R \left(L^{(r)} \right) + \frac{4}{n} \left(\int f^{(r)}(x)^2 f(x) dx - \Phi_r^2 \right)$$

de là on peut obtenir facilement une expression pour le *AMSE* et la minimiser de la même manière que pour le *AMISE*.

A1.5

Le concept d'optimalité a été construit principalement autour du critère *MISE* qui s'avère être une mesure souple et utile lorsqu'il s'agit de choisir un type de noyau à utiliser ou de sélectionner un paramètre de lissage.

Cette notion d'optimalité n'est cependant pas absolue, dans le sens où l'approche qui consiste en le choix d'un paramètre de lissage à travers la minimisation de ce critère ne mène pas forcément à la construction de l'estimateur idéal.

Par exemple, si l'on considère \hat{f}_{epan} un estimateur de densité à base du noyau d'Epanechnikov, on trouve que le paramètre de lissage optimal selon le critère *MISE* est donné par

$$h_{epan}^* = \left(\frac{15}{nR(f'')} \right)^{1/5}$$

on trouve aussi que

$$\lim_{n \rightarrow \infty} n^{4/5} MISE \left[\hat{f}_{epan} \right] = \frac{3^{4/5}}{5^{1/5} 4} [R(f'')]^{1/5}$$

Toutefois, si \hat{f}_n est un estimateur dont le noyau K satisfait

$$\mu_2(K) = 0 ; \int K^2(u) du < \infty$$

et dont le paramètre de lissage est

$$h_{\epsilon} = n^{-1/5} \epsilon^{-1} R(K)$$

alors $\forall \epsilon > 0$ on a

$$\lim_{n \rightarrow \infty} \sup \lim_{n \rightarrow \infty} n^{4/5} MISE \left[\hat{f}_n \right] \leq \epsilon$$

ceci implique que quel que soit le noyau utilisé pour l'estimation, il existera toujours un autre estimateur à noyau avec un critère *MISE* meilleur pourvu que le paramètre de lissage utilisé corresponde à h_ϵ .

A2.1

Rappel

Noyau gamma I (première variante)

$$K_{x/h+1,h}^G(t) = \frac{t^{x/h} e^{-t/h}}{h^{x/h+1} \Gamma(x/h+1)} ; t \geq 0$$

Noyau gamma II (deuxième variante)

$$K_{\rho_h(x),h}^G(t) = \frac{t^{\rho_h(x)-1} e^{-t/h}}{h^{\rho_h(x)} \Gamma(\rho_h(x))} ; t \geq 0$$

avec

$$\rho_h(x) = \begin{cases} (x/h) & \text{si } x \geq 2h \\ (x/(2h))^2 + 1 & \text{si } x \in [0, 2h[\end{cases}$$

Noyau bêta I (première variante)

$$K_{x/h+1,(1-x)/h+1}^B(t) = \frac{t^{x/h} (1-t)^{(1-x)/h}}{B(x/h+1, (1-x)/h+1)} ; t \in [0, 1]$$

Noyau bêta II (deuxième variante)

$$K_{\varrho_{h,0}(x),\varrho_{h,1}(x)}^B(t) = \frac{t^{\varrho_{h,0}(x)-1} (1-t)^{\varrho_{h,1}(x)-1}}{B(\varrho_{h,0}(x), \varrho_{h,1}(x))} ; t \in [0, 1]$$

avec

$$\varrho_{h,0}(x) = \begin{cases} \varrho_h(x) & \text{si } x \in [0, 2h[\\ x/h & \text{si } x \in [2h, 1] \end{cases}$$

$$\varrho_{h,1}(x) = \begin{cases} (1-x)/h & \text{si } x \in [0, 1-2h] \\ \varrho_h(1-x) & \text{si } x \in]1-2h, 1] \end{cases}$$

où

$$\varrho_h(x) = 2h^2 + \frac{5}{2} - \sqrt{4h^2 + 6h^2 + \frac{9}{4} - x^2 - \frac{x}{h}}$$

Biais

On définit d'abord les deux quantités

$$\varsigma_h(x) = \frac{(1-x)(\varrho(x) - x/h)}{[1 + h(\varrho(x) - x/h)]} ; \xi_h(x) = \left[\frac{x}{2h} - 1 \right]^2$$

on a alors les approximations suivantes

Noyau et support		Biais asymptotique $o(h)$ de l'estimateur
gamma I	$x \geq 0$	$h(f'(x) + (x/2)f''(x))$
gamma II	$x \in [0, 2h[$	$h(\xi_h(x)f'(x))$
	$x \geq 2h$	$h((x/2)f''(x))$
bêta I	$x \in [0, 1]$	$h((1-2x)f'(x) + (x/2)(1-x)f''(x))$
bêta II	$x \in [0, 2h[$	$h(\varsigma_h(x)f'(x))$
	$x \in [2h, 1-2h]$	$h((x/2)(1-x)f''(x))$
	$x \in]1-2h, 1]$	$h(-\varsigma_h(1-x)f'(x))$

Biais asymptotiques des variantes des noyaux gamma et bêta

A3.1

Soit l'identité

$$\frac{1}{\hat{f}_n(x)} = \frac{1}{E[\hat{f}_n(x)]} - \frac{\hat{f}_n(x) - E[\hat{f}_n(x)]}{E^2[\hat{f}_n(x)]} + \frac{(\hat{f}_n(x) - E[\hat{f}_n(x)])^2}{\hat{f}_n(x) E^2[\hat{f}_n(x)]}$$

En passant à l'espérance de $\hat{m}_{NW}(x)$ on trouve

$$\begin{aligned} E[\hat{m}_{NW}(x)] &= \hat{E}[\hat{m}_{NW}(x)] - \frac{E\left[(\hat{r}_n(x) - E[\hat{r}_n(x)])(\hat{f}_n(x) - E[\hat{f}_n(x)])\right]}{E^2[\hat{f}_n(x)]} \\ &\quad + E\left[\frac{\hat{r}_n(x)(\hat{f}_n(x) - E[\hat{f}_n(x)])^2}{\hat{f}_n(x) E^2[\hat{f}_n(x)]}\right] \\ &\triangleq \hat{E}[\hat{m}_{NW}(x)] + \frac{a_n(x) + b_n(x)}{E^2[\hat{f}_n(x)]} \end{aligned}$$

On rappelle que des approximations asymptotiques pour la variance de $\hat{f}_n(x)$ et $\hat{r}_n(x)$ sont données par

$$Var[\hat{f}_n(x)] \approx \frac{1}{nh} f(x) R(K)$$

$$\text{Var} [\hat{r}_n(x)] \approx \frac{1}{nh} s(x) R(K)$$

avec $s(x) = \int y^2 f(x, y) dy$.

L'inégalité de Cauchy-Schwartz donne alors

$$a_n(x) = O\left(\frac{1}{nh}\right)$$

ensuite, si $|Y| \leq M$ on trouve que

$$\frac{\hat{r}_n(x)}{\hat{f}_n(x)} = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \leq \frac{\sum_{i=1}^n MK\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} = M$$

le terme $b_n(x)$ sera alors borné par

$$\begin{aligned} b_n(x) &\leq M \times E \left[\left(\hat{f}_n(x) - E \left[\hat{f}_n(x) \right] \right)^2 \right] \\ &\approx \frac{M}{nh} f(x) R(K) \\ &\approx O\left(\frac{1}{nh}\right) \end{aligned}$$

maintenant, si $E[Y^2] < \infty$

$$\begin{aligned} b_n(x) &\leq E \left[\max_{1 \leq i \leq n} |Y_i| \left(\hat{f}_n(x) - E \left[\hat{f}_n(x) \right] \right)^2 \right] \\ &\approx \sqrt{nE[Y^2]} \\ &\approx O\left(\frac{1}{n^{1/2}h}\right) \end{aligned}$$

A3.2

Le traitement qui suit concerne un estimateur linéaire local dans le cas d'un dispositif expérimental à effet fixe.

Biais

On rappelle que

$$E[\hat{m}(x)] = e_1^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \mathbf{m}$$

avec $\mathbf{m} = (m(x_1), \dots, m(x_n))^T$.

La symétrie du noyau permet d'écrire

$$\mathbb{X}_x = \begin{pmatrix} 1 & x_1 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{pmatrix}$$

ensuite, par la formule de Taylor on obtient

$$M \approx \mathbb{X}_x \begin{pmatrix} m(x) \\ m'(x) \end{pmatrix} + \frac{1}{2} m''(x) \begin{pmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{pmatrix}$$

et en développant l'espérance de $\hat{m}(x)$, on trouve le premier terme

$$e_1^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} (\mathbb{X}_x^T W_x \mathbb{X}_x) \begin{pmatrix} m(x) \\ m'(x) \end{pmatrix} = e_1^T \begin{pmatrix} m(x) \\ m'(x) \end{pmatrix} = m(x)$$

le Biais est donc

$$Biais[\hat{m}(x)] \approx \frac{1}{2} m''(x) e_1^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \begin{pmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{pmatrix}_{1 \leq i \leq n}$$

maintenant, en remarquant que

$$n^{-1} \mathbb{X}_x^T W_x \mathbb{X}_x = \begin{pmatrix} \hat{s}_0(x; h) & \hat{s}_1(x; h) \\ \hat{s}_1(x; h) & \hat{s}_2(x; h) \end{pmatrix}$$

et que

$$n^{-1} \mathbb{X}_x^T W_x \begin{pmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{pmatrix} = \begin{pmatrix} \hat{s}_2(x; h) \\ \hat{s}_3(x; h) \end{pmatrix}$$

et sous les conditions usuelles et pour les points intérieurs, une approximation des éléments $s_l(x; h)$ peut être donnée par

$$\begin{aligned}
\hat{s}_l(x; h) &\approx \int_0^1 \frac{(y-x)^l}{h} K\left(\frac{y-x}{h}\right) dy \\
&\approx h^l \int_{-x/h}^{(1-x)/h} u^l K(u) du \\
&\approx h^l \int_{-1}^1 u^l K(u) du
\end{aligned}$$

on se retrouve donc avec

$$n^{-1} \mathbb{X}_x^T W_x \mathbb{X}_x \approx \begin{pmatrix} 1 & O(n^{-1}) \\ O(n^{-1}) & h^2 \mu_2(K) \end{pmatrix}$$

et

$$\mathbb{X}_x^T W_x \begin{pmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{pmatrix} \approx \begin{pmatrix} h^2 \mu_2(K) \\ O(n^{-1}) \end{pmatrix}$$

après calcul, le biais est alors

$$Biais[\hat{m}(x)] \approx \frac{1}{2} h^2 m''(x) \mu_2(K)$$

Variance

On rappelle que

$$Var[\hat{m}(x)] = e_1^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} (\mathbb{X}_x^T W_x V W_x \mathbb{X}_x) (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} e_1$$

En adoptant des approximations similaires au cas du biais, on trouve que

$$\begin{aligned}
n^{-1} \mathbb{X}_x^T W_x V W_x \mathbb{X}_x &= \frac{1}{nh} \sum_{i=1}^n K^2\left(\frac{x_i - x}{h}\right) \sigma^2(x_i) \begin{pmatrix} 1 & x_i - x \\ x_i - x & (x_i - x)^2 \end{pmatrix} \\
&\approx \begin{pmatrix} h^{-1} R(K) \sigma^2(x) & O(n^{-1}) \\ O(n^{-1}) & h \mu_2(K^2) \sigma^2(x) \end{pmatrix}
\end{aligned}$$

et donc

$$Var[\hat{m}(x)] \approx \frac{1}{nh} R(K) \sigma^2(x)$$

A3.3

Pour simplifier l'écriture des équations, posons

$$\mu_j = \int u^j K(u) du ; \nu_j = \int u^j K^2(u) du ; S = (\mu_{j+l})_{0 \leq j, l \leq p}$$

$$S^* = (\nu_{j+l})_{0 \leq j, l \leq p} ; S_n = \mathbb{X}_x^T W_x \mathbb{X}_x ; S_{n,j} = \sum_{i=1}^n K_h(X_i - x) (X_i - x)^j$$

$$\Sigma = W V W ; S_n^* = \mathbb{X}_x^T \Sigma \mathbb{X}_x ; S_{n,j}^* = \sum_{i=1}^n K_h(X_i - x) (X_i - x)^j \sigma^2(X_i)$$

$$H = \text{diag}\{1, h, \dots, h^p\} ; c_n = (S_{n,p+1}, \dots, S_{n,2p+1})^T ; c_p = (\mu_{p+1}, \dots, \mu_{2p+1})^T$$

$$\tilde{S} = (\mu_{j+l+1})_{0 \leq j, l \leq p} ; \tilde{c}_n = (S_{n,p+2}, \dots, S_{n,2p+2}) ; \tilde{c}_p = (\mu_{p+2}, \dots, \mu_{2p+2})^T$$

Variance

Commençons par rappeler que

$$\text{Var} [\hat{\beta} | X_1, \dots, X_n] = (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} (\mathbb{X}_x^T W_x V W_x \mathbb{X}_x) (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1}$$

la variance peut donc être réécrite sous la forme

$$\text{Var} [\hat{\beta} | X_1, \dots, X_n] = S_n^{-1} S_n^* S_n^{-1}$$

il est nécessaire maintenant de trouver des approximations pour S_n et S_n^* .

Notons que

$$\begin{aligned} S_{n,j} &\approx E[S_{n,j}] \\ &\approx nh^j f(x) \mu_j \end{aligned} \tag{A3.E1}$$

et de là, on trouve que

$$S_n \approx nf(x) H S H \tag{A3.E2}$$

De la même manière on a

$$S_{n,j}^* \approx nh^{j-1} f(x) \sigma^2(x) \nu_j$$

et

$$S_n^* \approx nh^{-1} f(x) \sigma^2(x) H S^* H$$

d'où

$$\text{Var} \left[\hat{\beta} | X_1, \dots, X_n \right] \approx \frac{\sigma^2(x)}{f(x) nh} H^{-1} S^{-1} S^* S^{-1} H^{-1}$$

et comme $j! \hat{\beta}_j$ est un estimateur de $m^j(x)$, une approximation de sa variance peut être immédiatement obtenue.

Biais

Il est nécessaire de distinguer entre les cas $(p-k)$ pair et $(p-k)$ impair. p étant le degré du polynôme local ajusté et k l'ordre de la dérivée de la fonction de régression à estimer.

On rappelle que

$$\text{Biais} \left[\hat{\beta} | X_1, \dots, X_n \right] = (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_r$$

on peut donc écrire

$$\text{Biais} \left[\hat{\beta} | X_1, \dots, X_n \right] = S_n^{-1} \mathbb{X}_x^T W_r$$

et par développement de Taylor

$$\begin{aligned} \text{Biais} \left[\hat{\beta} | X_1, \dots, X_n \right] &\approx S_n^{-1} \mathbb{X}_x^T W \left(\beta_{p+1} (X_i - x)^{p+1} \right)_{1 \leq i \leq n} \\ &\approx S_n^{-1} (\beta_{p+1} c_n) \end{aligned}$$

ensuite, à l'aide des équation (A3.E1) et (A3.E2) on trouve

$$\text{Biais} \left[\hat{\beta} | X_1, \dots, X_n \right] \approx H^{-1} S^{-1} c_p \beta_{p+1} h^{p+1} \quad (\text{A3.E3})$$

Cette formule n'est toutefois pas valable pour les cas $(p-k)$ pair, vu que les moments impairs des noyaux symétriques sont nuls. Il faudra donc procéder à des développements d'ordres plus élevés.

$S_{n,j}$ devient alors

$$S_{n,j} = nh^j (f(x) \mu_j + h f'(x) \mu_{j+1}) \quad (\text{A3.E4})$$

et donc

$$S_n \approx nH \left(f(x) S + h f'(x) \tilde{S} \right) H \quad (\text{A3.E5})$$

encore une fois, par le développement de Taylor, on peut écrire

$$\begin{aligned} \text{Biais} [\beta|X_1, \dots, X_n] &\approx S_n^{-1} \mathbb{X}_x^T W \left(\beta_{p+1} (X_i - x)^{p+1} + \beta_{p+2} (X_i - x)^{p+2} \right)_{1 \leq i \leq n} \\ &\approx S_n^{-1} (\beta_{p+1} c_n + \beta_{p+1} \tilde{c}_n) \end{aligned}$$

ensuite, en remplaçant par (A3.E4) et (A3.E5) on trouve

$$\text{Biais} [\beta|X_1, \dots, X_n] \approx h^{p+1} H^{-1} (\beta_{p+1} S^{-1} c_p + h b^*(x)) \quad (\text{A3.E6})$$

avec

$$b^*(x) = \frac{f'(x) \beta_{p+1} + \beta_{p+2} f(x)}{f(x)} s^{-1} \tilde{c}_p - \frac{f'(x)}{f(x)} \beta_{p+1} S^{-1} \tilde{S} S^{-1} c_p$$

Les formules (A3.E3) et (A3.E6) mènent alors directement vers une approximation du biais de l'estimateur dans les cas $(p - k)$ impair et pair.

A4.1

Soit l'équation différentielle

$$dx_t = a(x_t, t) dt + b(x_t, t) dW_t$$

qui, dans un certain sens peut être considérée comme la limite de

$$\delta x_t = a(x_t, t) \delta t + b(x_t, t) \epsilon_t \sqrt{\delta t} ; \epsilon \sim N(0, 1) \quad (\text{A4.E1})$$

Le but est alors de dériver une EDS pour une fonction F de (x_t) et t . Un élément essentiel est la formule de différentiation d'une fonction à deux variables G

$$dG = \frac{\partial G}{\partial x} dx + \frac{\partial G}{\partial y} dy$$

obtenue à partir du développement de Taylor

$$\delta G = \frac{\partial G}{\partial x} \delta x + \frac{\partial G}{\partial y} \delta y + \frac{1}{2} \frac{\partial^2 G}{\partial x^2} (\delta x)^2 + \frac{1}{2} \frac{\partial^2 G}{\partial y^2} (\delta y)^2 + \frac{\partial^2 G}{\partial x \partial y} \delta x \delta y + \dots$$

quand δx et δy tendent vers 0.

Il est donc possible d'appliquer ce même développement à F en s'arrêtant aux termes dominants. Il faudra toutefois, faire attention au terme $\sqrt{\delta t}$ de l'équation (A4.E1) qui ne peut être négligé lorsqu'on passe au carré.

Par ailleurs, un élément important à considérer, est le fait que le terme $\epsilon^2 \delta t$ peut-être traité comme non stochastique lorsque $\delta t \rightarrow 0$. Ainsi, dans le développement de Taylor on trouve

$$(\delta x_t)^2 \longrightarrow b^2 dt$$

En omettant maintenant les termes d'ordre supérieur et en prenant $\delta x_t, \delta t \rightarrow 0$ on obtient

$$dF = \frac{\partial F}{\partial x_t} dx_t + \frac{\partial F}{\partial t} dt + \frac{1}{2} \frac{\partial^2 F}{\partial x_t^2} b^2 dt$$

Finalement le lemme d'Itô est obtenu en remplaçant dx_t

$$dF = \left(a \frac{\partial F}{\partial x_t} + \frac{\partial F}{\partial t} + \frac{1}{2} b^2 \frac{\partial^2 F}{\partial x_t^2} \right) dt + b \frac{\partial F}{\partial x_t} dW$$

avec $a = a(x_t, t)$ et $b = b(x_t, t)$.

A5.1

Soit l'EDS

$$dx_t = \mu(x_t, t) dt + \sigma(x_t, t) dW_t^Q$$

avec (W_t^Q) un mouvement brownien sous la mesure Q , et soit la fonction dérivable $V = V(x_t, t)$ telle que

$$\frac{\partial V}{\partial t} + \mu(x_t, t) \frac{\partial V}{\partial x_t} + \frac{1}{2} \sigma^2(x_t, t) \frac{\partial^2 V}{\partial x_t^2} - r(x_t, t) V(x_t, t) = 0 \quad (\text{A5.E1})$$

avec la condition aux limites $V(x_T, T)$.

Le théorème de Feynman-Kac stipule que

$$V(x_t, t) = E^Q \left[\exp \left(- \int_t^T r(x_u, u) du \right) V(x_T, T) \mid \mathcal{F}_t \right]$$

Pour appliquer ce résultat à l'équation de Black & Scholes, il faut noter que la valeur $V_t = V(S_t, t)$ à l'instant t , lorsque le taux d'intérêt r est constant, d'un call européen ayant un prix d'exercice K suit l'EDP de Black & Scholes suivante

$$\frac{\partial V}{\partial t} + r S_t \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2} - r V_t = 0$$

avec la condition aux limites $V(S_T, T) = (S_T - K)^+$. En faisant une simple analogie avec l'équation (A5.E1), on trouve que $x_t = S_t$, $\mu(x_t, t) = rS_t$ et $\sigma(x_t, t) = \sigma S_t$. Ainsi par application du théorème de Feynman-Kac, on trouve que la valeur d'un call européen est donnée par

$$V(S_t, t) = e^{-r(T-t)} E^Q \left[(S_T - K)^+ | \mathcal{F}_t \right]$$

A5.2

Dans cette annexe nous donnons une dérivation de l'équation de Black & Scholes et la dérivation du prix du call décrit par cette équation à l'aide d'une terminologie purement financière.

Définitions

- Position longue : un investisseur est en position longue par rapport à un actif signifie que l'investisseur a acheté et possède l'actif en question.
- Position courte : un investisseur en position courte par rapport à un actif signifie que l'investisseur a vendu cet actif sans le posséder et se voit dans l'obligation de l'acheter pour le livrer.

Construction de l'équation de Black & Scholes pour un call européen

Soit C le prix d'un call européen sur une action ne versant pas de dividende. C dépend alors de deux variables t et S , qui sont respectivement, le temps et le prix de l'action sous-jacente, dont les variations sont représentées par un mouvement brownien géométrique.

Par le lemme d'Ito, on peut écrire

$$dC = \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial S} dS + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} dt \quad (\text{A5.E2})$$

Maintenant, on suppose que l'on dispose d'un portefeuille se composant de

- une position longue sur un call
- une position courte sur une fraction Δ de l'action sous-jacente au call.

La valeur Π de ce portefeuille est alors

$$\Pi = C - \Delta S ; (\Delta < 1)$$

et sa variation dans le temps est

$$d\Pi = dC - \Delta dS$$

En remplaçant par l'équation (A5.E2) on trouve

$$d\Pi = \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial S} dS + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} dt - \Delta dS \quad (\text{A5.E3})$$

Si l'on prend $\Delta = \partial C / \partial S$ on arrivera à éliminer dS et par conséquent la partie stochastique. On se retrouvera alors avec un portefeuille sans risque

$$d\Pi = \left(\frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \right) dt$$

Ici, intervient le principe d'absence d'arbitrage lequel spécifie qu'un portefeuille sans risque doit avoir un taux de rendement égal au taux sans risque r , donc

$$d\Pi = r\Pi dt$$

Finalement, l'utilisation dans cette égalité des valeurs respectives de $d\Pi$ et Π donnent l'équation de Black & Scholes

$$\frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0$$

Prix d'un call européen sur une action ne versant pas de dividende

Le payoff final du call en question est donné par $(S_T - K)^+$ avec S_T le prix de l'action à l'échéance T et K le prix d'exercice de l'option. On s'intéresse donc au calcul de $E^Q \left[(S - K)^+ | \mathcal{F}_t \right]$.

Soit f_S la fonction de densité de S . En simplifiant les notation, on aura

$$E \left[(S - K)^+ \right] = \int_K^\infty (\xi - K) f_S(\xi) d\xi$$

sachant que S suit une loi Lognormale et que μ et ς^2 sont respectivement la moyenne et la variance de $\log(S)$, on définit la variable centrée réduite $z = (\log(S) - \mu) / \varsigma$ de densité f_Z . L'espérance peut alors être réécrite sous la forme

$$E \left[(S - K)^+ \right] = \int_u^\infty \left[e^{(z\varsigma + \mu)} - K \right] f_Z(z) dz$$

avec $u = (\log(K) - \mu) / \varsigma$. On se retrouve alors, avec deux intégrales I_1 et I_2

$$I_1 = \int_u^\infty \left[e^{z\varsigma + \mu} f_Z(z) dz \right] ; I_2 = K \int_u^\infty f_Z(z) dz$$

ensuite, par un simple développement on trouve que

$$I_1 = e^{\mu + \varsigma^2/2} \int_u^\infty f(z - \varsigma) dz$$

Maintenant, si Φ est la fonction de répartition de la loi normale centrée réduite, on aura

$$1 - \Phi(f(z - \varsigma)) = 1 - \Phi\left(\frac{\log(K) - \mu}{\varsigma} - \varsigma\right)$$

donc

$$1 - \Phi\left(\frac{\log(K) - \mu}{\varsigma} - \varsigma\right) = \Phi\left(\frac{-\log(K) + \mu}{\varsigma} + \varsigma\right)$$

et comme μ est la moyenne de $\log(S)$ on peut l'écrire sous la forme $\mu = \log(E[S]) - (\varsigma^2/2)$ pour obtenir

$$\Phi\left(\frac{\log(E[S]/K) + \varsigma^2/2}{\varsigma}\right) = \Phi(d_1)$$

avec

$$d_1 = \frac{\log(E[S]/K) + \varsigma^2/2}{\varsigma}$$

Des développements similaires pour l'intégrale I_2 permettent de trouver $\Phi(d_2)$ avec

$$d_2 = \frac{\log(E[S]/K) - \varsigma^2/2}{\varsigma}$$

et par conséquent

$$E^Q[(S_T - K)^+ | \mathcal{F}_t] = E^Q[S_T | S_0] \Phi(d_1) - K \Phi(d_2)$$

Le calcul de l'espérance sous une mesure martingale Q donne

$$E^Q[S_T | S_0] = S_0 e^{rT}$$

avec r le taux sans risque. En prenant également $\varsigma = \sigma\sqrt{T}$, on trouve que

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} ; d_2 = d_1 - \sigma\sqrt{T}$$

Le prix du call est donc donnée par

$$C = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2)$$

qui est la formule de Black & Scholes.

Bibliographie

- [1] Abadir, K.M., et Lawford, S. : Optimal asymmetric kernels. *Economics Letters*. **83**(1), 61-68 (2004)
- [2] Alfonsi, A. : On the discretization schemes for the CIR (and Bessel Squared) processes. *Monte Carlo Methods and Applications*. **11**(4), 355-384 (2005)
- [3] Andersen, L.B. et Piterbarg, V.V. : Moment explosions in stochastic volatility models. *Finance and Stochastics*. **11**(1), 29-50 (2007)
- [4] Bertrand-Retali, M. : Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Revue Roumaine des Mathématiques Pures et Appliquées*. **23**(3), 361-385 (1978)
- [5] Boecker, H., Hillman, C.H., Scheef, L. et Strüder, H.K. (Eds.) : *Functional Neuroimaging in Exercise and Sport Sciences*. Springer Science & Business Media, New York (2012)
- [6] Bollerslev, T. : Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. **31**(3), 307-327 (1986)
- [7] Bollerslev, T., Engle, R.F. et Nelson, D.B. : ARCH models. In : Engle, R., McFadden, D.(eds.) *Handbook of Econometrics*, pp.2959-3038, **4**(1). Elsevier Science (1994)
- [8] Bosq, D. : *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag, New York (1998)
- [9] Bouezmarni, T., El Ghouch, A. et Mesfioui, M. : Gamma kernel estimators for density and hazard rate of right-censored data. *Journal of Probability and Statistics*. **2011** (2011)
- [10] Bouezmarni, T. et Rolin, J.M. : Consistency of the beta kernel density function estimator. *Canadian Journal of Statistics*. **31**(1), 89-98 (2003)
- [11] Bouezmarni, T. et Rombouts, J.V.K. : Nonparametric density estimation for multivariate bounded data. *Journal of Statistical Planning and Inference*. **140**(1), 139-152 (2010a)
- [12] Bouezmarni, T. et Rombouts, J.V.K. : Nonparametric density estimation for positive time series. *Computational Statistics and Data Analysis*. **54**(2), 245-261 (2010b)
- [13] Bouezmarni, T. et Scaillet, O. : Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*. **21**(2), 390-412 (2005)
- [14] Bowman, A.W. : An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*. **71**(2), 353-360 (1984)
- [15] Boyle, P.P. : Options : A Monte Carlo approach. *Journal of Financial Economics*. **4**(3), 323-338 (1977)
- [16] Boyle, P.P., Broadie, M. et Glasserman, P. : Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*. **21**(8-9), 1267-1321 (1997)
- [17] Brandimarte, P. : *Numerical Methods in Finance and Economics : a MATLAB-Based Introduction*. John Wiley & Sons, Hoboken, New Jersey (2013)
- [18] Brennan, M.J. et Schwartz, E.S. : The valuation of American put options. *The Journal of Finance*. **32**(2), 449-462 (1977)
- [19] Broadie, M. et Glasserman, P. : Estimating security price derivatives using simulation. *Management Science*. **42**(2), 269-285 (1996)

- [20] Broadie, M. et Kaya, Ö. : Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*. **54**(2), 217-231 (2006)
- [21] Brown, B.M. et Chen, S.X. : Beta-Bernstein smoothing for regression curves with compact support. *Scandinavian Journal of Statistics*. **26**(1), 47-59 (1999)
- [22] Carey, J.R., Liedo, P., Müller, H.G., Wang, J.L. et Chiou, J.M. : Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of mediterranean fruit fly females. *The Journals of Gerontology Series A : Biological Sciences and Medical Sciences*. **53**(4), B245-B251 (1998)
- [23] Chen, S.X. : Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*. **31**(2), 131-145 (1999)
- [24] Chen, S. X. : Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*. **52**(3), 471-480 (2000)
- [25] Clewlow, L. et Carverhill, A. : On the simulation of contingent claims. *The Journal of Derivatives*. **2**(2), 66-74 (1994)
- [26] Conze, A. : Path dependent options : The case of lookback options. *The Journal of Finance*. **46**(5), 1893-1907 (1991)
- [27] Cowling, A. et Hall, P. : On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society : Series B (Methodological)*. **58**(3), 551-563 (1996)
- [28] Cox, J.C., Ingersoll, J. et Ross, S. : A theory of the term structure of interest rates. *Econometrica*. **53**(2), 385-407 (1985)
- [29] Cox, J.C., Ross, S.A. et Rubinstein, M. : Option pricing : A simplified approach. *Journal of Financial Economics*. **7**(3), 229-263 (1979)
- [30] Deng, H.W. et Shen, H. (Eds.) : *Current Topics in Human Genetics : Studies in Complex Diseases*. World Scientific, Singapore (2007)
- [31] Devroye, L. et Györfi, L. : *Density Estimation : The L1 View*. John Wiley & Sons, New York (1985)
- [32] Di Marzio, M. et Taylor C.C. : Kernel density classification and boosting : an L2 analysis. *Statistics and Computing*. **15** (2), 113-123 (2005)
- [33] Donthu, N. : Comparing market areas using kernel density estimation. *Journal of the Academy of Marketing Science*. **19** (4), 323-332 (1991)
- [34] Dubois, F. et Lelièvre, T. : Efficient pricing of Asian options by the PDE approach. *Journal of Computational Finance*. **8**(2), 55-64 (2004)
- [35] Dudewicz, E. J. et Geller, N. L. : Book Review of "Nonparametric Statistical Inference" by Jean Dickinson Gibbons. *Journal of the American Statistical Association*. **67**, 962-963 (1972)
- [36] Duffie, D. : *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, New Jersey (2010)
- [37] Egloff, D., Kohler, M. et Todorovic, N. : A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *The Annals of Applied Probability*. **17**(4), 1138-1171 (2007)
- [38] Engel, J., Herrmann, E. et Gasser, T. : An iterative bandwidth selector for kernel estimation of densities and their derivatives. *Journal of Nonparametric Statistics*. **4**(1), 21-34 (1994)
- [39] Fan, J. : On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*. **19**(3), 1257-1272 (1991)
- [40] Fan, J. : Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*. **21**(1), 196-216 (1993)
- [41] Feller, W. : Two singular diffusion problems. *Annals of Mathematics*. **54**(1), 173-182 (1951)
- [42] Franses, P.H. et Van Dijk, D. : *Non-Linear Time Series Models in Empirical Finance*. Cambridge university press, New York (2000)

- [43] Fusai, G. et Roncoroni, A. : Implementing Models in Quantitative Finance : Methods and Cases. Springer-Verlag, Berlin Heidelberg (2007)
- [44] Gatheral, J. : The Volatility Surface : A Practitioner's Guide. John Wiley & Sons, Hoboken, New Jersey (2006)
- [45] Genon-Catalot, V., Jeantheau, T. et Laredo, C. : Limit theorems for discretely observed stochastic volatility models. *Bernoulli*. **4** (3), 283-303 (1998)
- [46] Glasserman, P. : Monte Carlo Methods in Financial Engineering. Springer Science & Business Media, New York (2013)
- [47] Glasserman, P., Heidelberger, P. et Shahabuddin, N.P. : Pricing of options using importance sampling and stratification/Quasi-Monte Carlo. U.S. Patent No. 6,381,586. Washington, DC : U.S. Patent and Trademark Office (2002)
- [48] Glynn, P.W. et Whitt, W. : The asymptotic efficiency of simulation estimators. *Operations Research*. **40**(3), 505-520 (1992)
- [49] Gouriéroux, C. et Monfort, A. : (Non) Consistency of the beta kernel estimator for recovery rate distribution. Preprint of Centre de Recherche en Economie et Statistique (No. 2006-31) (2006)
- [50] Guillaón, A., Navarro, J., et Ruiz, J.M. : A note on kernel estimators for positive valued random variables. *Sankhyā : The Indian Journal of Statistics, Series A*. **61**(2), 276-281 (1999)
- [51] Gustafsson, J., Hagmann, M., Nielsen, J.P. et Scaillet, O. : Local transformation kernel density estimation of loss distributions. *Journal of Business & Economic Statistics*. **27**(2), 161-175 (2009)
- [52] Gustafsson, J., Nielsen, J.P., Pritchard, P. et Roberts, D. : Quantifying operational risk guided by kernel smoothing and continuous credibility : a practitioners view. Available at SSRN 880361 (2006)
- [53] Hajizadeh, E., Seifi, A., Zarandi, M.F. et Turksen, I.B. : A hybrid modeling approach for forecasting the volatility of S&P 500 index return. *Expert Systems with Applications*. **39**(1), 431-436 (2012)
- [54] Hald, A. : A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935. Springer-Verlag, New York (2007)
- [55] Harrison, J.M. et Kreps, D.M. : Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*. **20**(3), 381-408 (1979)
- [56] Harrison, J.M. et Pliska, S.R. : Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*. **11**(3), 215-260 (1981)
- [57] Hastie, T., Tibshirani, R. et Friedman, J. : The Elements of Statistical Learning. Springer series in statistics, New York (2009)
- [58] Hirukawa, M. : Asymmetric Kernel Smoothing : Theory and Applications in Economics and Finance. Springer, Singapore (2018)
- [59] Hollander, M., Wolfe D.A. et Chicken, E. : Nonparametric Statistical Methods (2nd Ed.). John Wiley & Sons, New York (1999)
- [60] Hull, J. et White, A. : The pricing of options on assets with stochastic volatilities. *The journal of Finance*. **42**(2), 281-300 (1987)
- [61] Hull, J. et White, A. : One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis*. **28**(2), 235-254 (1993)
- [62] Igarashi, G. et Kakizawa, Y. : Re-formulation of inverse Gaussian, reciprocal inverse Gaussian, and Birnbaum–Saunders kernel estimators. *Statistics & Probability Letters*. **84**, 235-246 (2014)
- [63] Jeon, Y. et Kim, J.H. : A gamma kernel density estimation for insurance loss data. *Insurance : Mathematics and Economics*. **53**(3), 569-579 (2013)
- [64] Johnson, N.L., Kotz, S.I. et Balakrishnan, N. : Continuous Univariate Distributions. John Wiley & Sons, New York (1994)
- [65] Karatzas, I. et Shreve, S. : Brownian Motion and Stochastic Calculus. Springer, New York (1991)

- [66] Karunamuni, R.J. et Alberts, T. : A generalized reflection method of boundary correction in kernel density estimation. *Canadian Journal of Statistics*. **33**(4), 497-509 (2005)
- [67] Ke, J., Wang, L., Murray, L. et Xiong, H. : Research on volatility spillover effects in financial markets based on copula and kernel function. In : 2010 International Conference on Computational and Information Sciences, pp. 402-405. IEEE (2010)
- [68] Kristensen, D. : Nonparametric filtering of the realized spot volatility : A kernel-based approach. *Econometric Theory*. **26**(1), 60-93 (2010)
- [69] Kulczycki, P. : Kernel estimators in industrial applications. In : Prasad, B.(ed.) *Soft Computing Applications in Industry*, pp. 69-91. Springer, Berlin, Heidelberg (2008)
- [70] Kvam, P.H. : *Nonparametric Statistics with Applications to Science and Engineering*. John Wiley & Sons, Hoboken, New Jersey (2007)
- [71] Linton, O., Mammen, E., Nielsen, J. et Taanggard, C. : Yield curve estimation by kernel smoothing. *Financial Markets Group, London School of Economics and Political Science* (2004)
- [72] Lord, R., Koekkoek, R. et Van Dijk, D. : A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*. **10**(2), 177-194 (2010)
- [73] Malec, P. et Schienle, M. : Nonparametric kernel density estimation near the boundary. *Computational Statistics & Data Analysis*. **72**, 57-76 (2014)
- [74] Müller, H.G. : Smooth optimum kernel estimators near endpoints. *Biometrika*. **78**(3), 521-530 (1991)
- [75] Nadaraya, E.A. : On estimating regression. *Theory of Probability & Its Applications*. **9**(1), 141-142 (1964)
- [76] Nadaraya, E.A. : *Nonparametric Estimation of Probability Densities and Regression Curves*. Springer, Netherlands (1989)
- [77] Narsky, I. et Porter, F.C. : *Statistical Analysis Techniques in Particle Physics*. Wiley-VCH, Weinheim (2014)
- [78] Nelson, D.B. et Ramaswamy, K. : Simple binomial processes as diffusion approximations in financial models. *The Review of Financial Studies*. **3**(3), 393-430 (1990)
- [79] Ouamaliche, S. et Sayah, A. : Kernel density smoothing using probability density functions and orthogonal polynomials. *Electronic Journal of Applied Statistical Analysis*. **9**(2), 400-416 (2016)
- [80] Ouamaliche, S. et Sayah, A. : A nonparametric estimator for stochastic volatility density. Accepted for publication in *International Journal of Computational Economics and Econometrics* (2020).
- [81] Parzen, E. : On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*. **33**(3), 1065-1076 (1962)
- [82] Pensky, M. et Vidakovic, B. : Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*. **27**(6), 2033-2053 (1999)
- [83] Poggi, J.M. : Préviation non paramétrique de la consommation électrique. *Revue de Statistiques Appliquées*. **42** (4), 83-98 (1994)
- [84] Pouillot, R. et Delignette-Muller, M.L. : Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology*. **142** (3), 330-340 (2010)
- [85] Renault, O. et Scaillet, O. : On the way to recovery : a nonparametric bias free estimation of recovery rate densities. *Journal of Banking & Finance*. **28**(12), 2915-2931 (2004)
- [86] Rio, E. : *Inequalities and Limit Theorems for Weakly Dependent Sequences* (2013)
- [87] Rogers, L.C.G. et Shi, Z. : The value of an Asian option. *Journal of Applied Probability*. **32**(4), 1077-1088 (1995)
- [88] Rouah, F.D. : *The Heston Model and Its Extensions in Matlab and C*. John Wiley & Sons, Hoboken, New Jersey (2013)

- [89] Rudemo, M. : Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*. **9**(2), 65-78 (1982)
- [90] Ruppert, D. et Wand, M.P. : Multivariate locally weighted least squares regression. *The Annals of Statistics*. **22**(3), 1346-1370 (1994)
- [91] Satchell, S. et Knight, J. : *Forecasting Volatility in the Financial Markets*. Elsevier, Netherlands (2011).
- [92] Sharma, A., Tarboton, D.G. et Lall, U. : Streamflow simulation : a nonparametric approach. *Water Resources Research*. **33** (2), 291-308 (1997)
- [93] Sheather, S.J. et Jones, M.C. : A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society : Series B (Methodological)*. **53**(3), 683-690 (1991)
- [94] Shi, J. et Song, W. : Asymptotic results in gamma kernel regression. *Communications in Statistics-Theory and Methods*. **45**(12), 3489-3509 (2016)
- [95] Silverman, B.W. : Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*. **6**, 177-184 (1978)
- [96] Silverman, B.W. : *Density Estimation for Statistics and Data Analysis*. . Chapman & Hall, London - New York (1986)
- [97] Simonoff, J.S. : *Smoothing Methods in Statistics*. Springer-Verlag, New York (1996)
- [98] Stone, C.J. : Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*. **8**(6), 1348-1360 (1980)
- [99] Takezawa, K. : *Introduction to Nonparametric Regression*. John Wiley & Sons, Hoboken, New Jersey (2006)
- [100] Trapp II, A.F. : Applications of nonparametric kernel smoothing estimators in Monte Carlo risk assessments. Digital Repository @ Iowa State University (2012)
- [101] Trapp II, A.F., Dixon, P., Widrechner, M.P. et Kovach, D.A. : Scheduling viability tests for seeds in long-term storage based on a bayesian multi-level model. *Journal of Agricultural, Biological, and Environmental Statistics*. **17** (2), 192-208 (2012)
- [102] Van Es, B., Spreij, P. et Van Zanten, H. : Nonparametric volatility density estimation. *Bernoulli*. **9**(3), 451-465 (2003)
- [103] Van Es, B., Spreij, P. et Van Zanten, H. : Nonparametric methods for volatility density estimation. In : Di Nunno, G., Øksendal, B.(eds.) *Advanced Mathematical Methods for Finance*, pp. 293-312. Springer, Heidelberg (2011)
- [104] Wand, M.P. et Jones, M.C. : *Kernel Smoothing*. Chapman & Hall, London - New York (1995)
- [105] Wasserman, L. : *All of Nonparametric Statistics*. Springer-Verlag, New York (2006)
- [106] Watson, G.S. : Smooth regression analysis. *Sankhyā : The Indian Journal of Statistics, Series A*. **26**(2), 359-372 (1964)
- [107] Wolberg, J.R. : *Expert Trading Systems : Modeling Financial Markets with Kernel Regression*. John Wiley & Sons, New York (2000)
- [108] Wolfowitz, J. : Additive partition functions and a class of statistical hypotheses. *The Annals of Mathematical Statistics*. **13** (3), 247-279 (1942)
- [109] Yin, Z., O'Sullivan, C. et Brabazon, A. : An analysis of the performance of genetic programming for realised volatility forecasting. *Journal of Artificial Intelligence and Soft Computing Research*. **6**(3), 155-172 (2016)
- [110] Zhu, J. : *Applications of Fourier Transform to Smile Modeling : Theory and Implementation*. Springer-Verlag, Heidelberg (2010)
- [111] Zvan, R., Vetzal, K.R. et Forsyth, P.A. : PDE methods for pricing barrier options. *Journal of Economic Dynamics and Control*. **24**(11-12), 1563-1590 (2000)

Résumé

L'estimation non paramétrique constitue un outil fondamental en statistique permettant de repousser les limites établies par les principes sur lesquels reposent les techniques paramétriques et par conséquent, permet d'aborder des problèmes où l'approche paramétrique soit échoue, soit donne des résultats qui peuvent être significativement améliorés. Le travail conduit au cours de cette thèse s'inscrit dans ce contexte via l'adoption des méthodes non-paramétriques de lissage par noyaux afin de traiter quelques questions en finance quantitative, discipline dans laquelle il est rare que les méthodes paramétriques donnent des résultats satisfaisants. En premier lieu, nous commençons par couvrir les deux volets les plus importants du lissage par noyaux, qui sont l'estimation de densité et la régression, dans les cas des noyaux symétriques et asymétriques. Le but étant d'exposer les propriétés fondamentales et de mettre en relief la différence dans la structure et dans le traitement qu'implique le simple passage d'un cas à l'autre. Par la suite, nous mettons en application le lissage par noyaux, d'abord par la suggestion d'une procédure pour l'estimation de la densité de la volatilité d'actifs financiers et ensuite à travers une formule pour la valorisation de produits financiers.

Mots-clés : Estimation non-paramétrique, Lissage par noyaux, Estimation de densité, Régression, Polynômes orthogonaux, Volatilité stochastique, Valorisation d'option.

Abstract

Nonparametric estimation constitutes a fundamental tool in statistics that helps in pushing the boundaries established by the principles on which parametric techniques are based and consequently provides a way to address problems where the parametric approach either fails or gives results that offer substantial room for improvement. The work undertaken in this thesis falls within this context through the adoption of nonparametric kernel smoothing methods in order to deal with some issues in quantitative finance, a field in which parametric methods rarely deliver satisfying results. As a first step, we start by covering the two most important components of kernel smoothing, namely density estimation and regression, in both the symmetric and asymmetric kernel case. The main point being to describe the fundamental properties and to highlight differences in the structure and in the treatment when moving from one case to the other. Afterwards, kernel smoothing is put to the test, firstly by suggesting an estimation procedure for the volatility density of a financial asset and then through a formula for pricing financial derivatives.

Keywords : Nonparametric estimation, Kernel smoothing, Density estimation, Regression, Orthogonal polynomials, Stochastic volatility, Option valuation.