

N° d'ordre : 05/2013-M/MT

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DES SCIENCES ET TECHNOLOGIE HOUARI BOUMEDIENE
FACULTÉ DES MATHÉMATIQUES



MÉMOIRE
PRÉSENTÉ POUR L'OBTENTION DU DIPLOME DE MAGISTER.

EN MATHÉMATIQUES
SPÉCIALITÉ : PROBABILITÉ & STATISTIQUE

Par : Meriem NADJI

Sujet

Quelques Propriétés d'un Estimateur Non
Paramétrique des Dérivées de la Densité pour
des Données Censurées

Soutenu publiquement le 30/11/2013, devant le jury composé de :

Mr. K. BOUKHETALA	Professeur	à l'USTHB	Président.
Mr. A. TATACHAK	Maître de Conférences/A	à l'USTHB	Directeur de mémoire.
Mme. H. GUERBYENNE	Professeur	à l'USTHB	Examinatrice.
Mme. Z. GUESSOUM	Maître de Conférences/A	à l'USTHB	Examinatrice.

Remerciements

Tout d'abord, je tiens à remercier Dieu le tout puissant de m'avoir donné le courage, la morale et la santé pour mener à bien ce travail.

Comme le veut la tradition, je vais tenter de satisfaire au difficile exercice de la page des remerciements, peut être la tâche la plus ardue de ces années du mémoire. Non qu'exprimer ma gratitude envers les personnes en qui j'ai trouvé un soutien soit contre ma nature, bien au contraire. La difficulté tient plutôt dans le fait de n'oublier personne. C'est pourquoi, je remercie par avance ceux dont le nom n'apparaît pas dans cette page et qui m'ont aidé d'une manière ou d'une autre. Ils se reconnaîtront.

Je tiens à remercier en tout premier lieu Mr. A.TATACHAK, pour m'avoir pris sous son aile durant ces trois dernières années et de m'avoir encadré dans ce mémoire, d'aussi bonne grâce, et s'y être intéressé. La disponibilité dont il a fait preuve pour d'intenses et rationnelles discussions, ses riches conseils et son aide m'ont beaucoup apporté en statistiques et m'ont appris à écrire. Au travers de ses qualités, tant humaines que scientifiques, j'ai découvert le monde de la recherche. Qu'il trouve dans ces petits mots l'expression de mes sentiments les plus respectueux, et j'espère avoir été à la hauteur de sa confiance. Être son élève fut, pour moi, un réel plaisir.

Il est très agréable aussi, d'adresser mes vifs remerciements à Mme. O. SADKI d'avoir répondu à mes sollicitations lorsque le besoin s'en faisait sentir. J'éprouve un profond respect pour son travail et son parcours, ainsi que pour ses qualités humaines. Le regard critique, juste et avisé qu'elle a porté sur mon travail ne peut que m'encourager à être encore plus perspicace et engagé dans mes recherches.

J'exprime aussi ma profonde gratitude à Mr. K. BOUKHETALA Professeur à l'USTHB, pour avoir accepté de présider le jury de soutenance.

Je suis très honorée que Mme. H. GUERBYENNE Professeur à l'USTHB, et Mme. Z. GUESSOUM Maître de conférences à l'USTHB aient acceptées de juger ce mémoire. Leur vision m'apportera un nouvel éclairage sur mon travail. Je leur remercie d'avoir fait l'effort de lire mon mémoire et d'examiner mon travail.

Par ailleurs, je souhaite aussi remercier mon ami Mr A.Guidoum d'avoir partagé ses connaissances avec moi et d'être investi autant dans mon travail. Il a réussi à me pousser à l'autonomie sans pression en faisant preuve de dévouement et en étant toujours disponible, pour répondre à mes questions avec une grande pédagogie, ce que j'ai fortement apprécié. Je souhaite que nous aurons toujours l'occasion de travailler ensemble.

Sur un plan plus personnel, je désire remercier mes amis, qui m'ont soutenue et divertie tout au long de l'accomplissement de ce travail.

Le plus fort de mes remerciements est pour ma famille pour son irremplaçable et inconditionnel soutien. Ils ont été présents pour écarter les doutes, soigner les blessures et partager les joies. Ce mémoire est un peu le leur, aussi. Merci papa, maman, soeurs et frère.

J'adresse, évidemment, de tendres remerciements à mon époux, sans qui j'aurais eu du mal à garder l'équilibre. Merci pour tout ce qu'il représente pour moi, et pour cette indicible joie d'être deux. Merci de m'avoir tenu la main jusqu'aux dernières lignes de ce mémoire. En plus de me supporter dans des conditions " normales ", ce qui n'est déjà pas une mince affaire, je le conçois. Et pourtant, quelle présence indispensable. Merci d'être là tous les jours.

A mon cher défunt grand père Abdelkader

Table des matières

Notations et Abréviations	1
Introduction Générale	3
1 Introduction à l'Estimation Non Paramétrique	5
1.1 Estimation non paramétrique d'une fonction de répartition	5
1.1.1 La fonction de répartition empirique	5
1.1.2 La fonction quantile empirique	6
1.2 Estimation non paramétrique d'une fonction densité de probabilité	6
1.2.1 Histogramme de densité	7
1.2.2 Estimateur simple	7
1.2.3 Estimateur à noyau	9
1.2.3.1 Expressions du biais et de la variance	11
1.2.3.2 Expressions asymptotiques du biais et de la variance	13
1.2.3.3 Avantage de l'estimateur à noyau	15
1.2.3.4 Choix théoriques optimaux du paramètre de lissage	16
2 Données Censurées	18
2.1 Introduction	18
2.1.1 Les données tronquées	18
2.1.1.1 La troncature à gauche	19
2.1.1.2 La troncature à droite	19
2.1.1.3 La troncature par intervalle	19
2.1.2 Les données censurées	19
2.1.2.1 Censure à droite	19
2.1.2.2 Censure à gauche	20
2.1.2.3 Censure par intervalle	21
2.1.3 Fonctions associées aux distributions de survie	21

2.2	Estimation de la fonction de survie dans un modèle complet	23
2.3	Estimation de la fonction de survie dans un modèle censuré	24
2.3.1	Problème d'identifiabilité dans le modèle de censure	24
2.3.2	Estimateur de Kaplan-Meier	25
2.4	Propriétés de l'estimateur de Kaplan-Meier	27
2.4.1	Cohérence	27
2.4.2	Loi fortes	27
2.4.3	Convergence en loi	29
2.5	Estimation de la fonction de densité de probabilité dans un modèle censuré . .	30
2.5.1	Etude du biais de l'estimateur \hat{f}_n	32
2.5.2	consistance de $\hat{f}_n(t)$	39
3	Vitesse de Convergence de l'Estimateur des Dérivées de la Densité dans un	
	Modèle Complet	42
3.1	Introduction	42
3.2	Convergence uniforme des $f_n^{(r)}$	43
3.3	Expressions du biais et de la variance pour l'estimateur à noyau des dérivées	
	de la densité :	49
3.3.1	Biais	50
3.3.2	Variance	51
3.3.3	MISE	52
3.4	Expressions asymptotique du biais et de la variance pour l'estimateur à noyau	
	des dérivées de la densité :	53
3.4.1	Biais	54
3.4.2	Variance	55
3.4.3	MSE	57
4	Estimateur à noyau des dérivées de la densité pour le modèle censuré	58
4.1	Introduction :	58
4.2	Convergence uniforme de $f_n^{(r)}$:	59
5	Simulations	67
5.1	Estimation du paramètre de lissage	67
5.1.1	Maximum de vraisemblance avec validation croisée	67
5.1.2	Validation croisée non biaisée	68
5.1.3	Validation croisée biaisée	70

5.1.4	Critère du AMISE	72
5.2	Simulations et résultats	72
5.2.1	Simulations	72
5.2.2	Résultats	74
	Conclusion et Perspectives	84
	Annexe	86
	Bibliographie	92

Table des figures

5.1	MLCV pour la loi Bimodale, $n=500$	79
5.2	MLCV pour la loi $N(0,1)$, $n=500$	79
5.3	UCV pour la loi Bimodale, $r=0$, $n=500$	79
5.4	UCV pour la loi Bimodale, $r=2$, $n=500$	80
5.5	BCV pour la loi Bimodale, $r=1$, $n=500$	80
5.6	BCV pour la loi Bimodale, $r=2$, $n=500$	80
5.7	AMISE pour la loi Bimodale, $r=0$, $n=500$	81
5.8	AMISE pour la loi Bimodale, $r=3$, $n=500$	81
5.9	Estimation de la loi Bimodale par les méthodes UCV, BCV, AMISE respectivement, pour noyau gaussien, $r=1$, $n=500$	81
5.10	Estimation de la loi $N(0,1)$ par les méthodes UCV, BCV, AMISE respectivement, pour noyau gaussien, $r=1$, $n=500$	82
5.11	Impact du choix du noyau, avec la méthode UCV, loi Bimodale, $r=0$, $n=500$	82
5.12	Impact du choix du noyau, avec la méthode UCV, loi Bimodale, $r=1$, $n=500$	82
5.13	Impact du choix du noyau, avec la méthode BCV, loi $N(0,1)$, $r=1$, $n=500$	83
5.14	$X : N(0,1)$, $C : N(0,1)$, $r=0$, $n=500$	83
5.15	$X : N(0,1)$, $C : N(0,1)$, $r=1$, 3 respectivement, $n=500$	83

Liste des tableaux

5.1	Résultats des simulations de la loi Bimodale pour $n=500$	76
5.2	Résultats des simulations de la loi Normale centrée réduite pour $n=500$	78

Notations et Abréviations

- (Ω, A, \mathbb{P}) : espace probabilisé.
- \mathbb{P} : mesure de probabilité.
- $\mathbb{P}(A)$: probabilité de l'évènement A .
- \mathbb{N} : ensemble de tous les entiers.
- \mathbb{R} : ensemble de tous les nombres réels.
- \mathbb{R}_+ : ensemble de tous les nombres réels non négatifs.
- \mathbb{R}_+^* : ensemble de tous les nombres réels positifs.
- $v.a$: variable aléatoire.
- $f.d.r$: fonction de répartition.
- $i.i.d$: indépendantes et identiquement distribuées.
- $T.C.L$: théorème central limite.
- \mathbb{I}_A : fonction indicatrice de l'ensemble A .
- N : loi normale.
- $|x|$: valeur absolue de x .
- $x \wedge y$: minimum de x et y .
- \xrightarrow{L} : convergence en loi.

- \xrightarrow{P} : convergence en probabilité.
- *p.s* : presque sûrement.
- $O(\cdot)$: grand O .
- $o(\cdot)$: petit o .

Introduction Générale

L'estimation non paramétrique est une méthode très attrayante et efficace pour obtenir des propriétés ayant des aspects généraux d'une courbe (régression, densité, etc.). L'une des principales méthodes utilisées dans ce domaine est celle de l'estimation à noyau (voir Silverman [34] sur l'estimation non paramétrique de la densité, ou Härdle [15] et Müller [26] au sujet de la régression). L'estimation des fonctions de densité de probabilité avec la méthode du noyau a connu un succès remarquable en raison de sa facilité d'interprétation et de visualisation, alors que l'estimation des dérivées de la fonction de densité a reçu moins d'attention. En effet, il s'agit d'un problème théorique plus difficile qui possède des applications intéressantes.

Dans le domaine de l'estimation d'une densité, il y a certains aspects importants tels que l'approximation du mode, des points d'inflexion et la détermination du paramètre de lissage optimal (celui-ci dépend de la dérivée seconde). Pour cette raison, il est d'un grand intérêt d'estimer non seulement la densité elle-même, mais aussi les dérivées successives de la fonction de densité. Il en est ainsi pour des raisons théoriques et pratiques (voir, par exemple, Hildenbrand et Hildenbrand [13]). L'estimation des dérivées de la fonction de densité est délaissée dans la littérature, c'est ainsi qu'il y a eu très peu de tentatives pour explorer les applications des estimateurs des dérivées de la densité à des problèmes statistiques (voir, par exemple, Singh [36]). Rosenblatt [29], introduisant le concept de l'estimation par la méthode du noyau, a souligné qu'un estimateur de la densité ne peut pas être sans biais. Parzen [28] et Bartlett [2] suggèrent que le biais de ces estimateurs peut être réduit. Singh [35] expose explicitement les estimateurs de la r ème dérivée d'une densité qui permettent d'atteindre un taux minimal asymptotique de l'erreur quadratique moyenne, tout en améliorant des estimateurs proposés par Bhattacharya [3], Schwartz [42] et Schuster [31]. Pour toutes ces raisons, nous proposons un estimateur des dérivées de la fonction de densité dans le cas de données incomplètes (censure droite), pour lequel, nous établissons des résultats de convergence uniforme presque sûre avec vitesse.

Le mémoire est organisé comme suit :

Chapitre 1 : Dans ce chapitre, nous rappelons les notions de base et les concepts mathématiques nécessaires à l'estimation non paramétriques d'une fonction de densité de probabilité dans un modèle complet, avec des variables aléatoires indépendantes et identiquement distribuées.

Chapitre 2 : Ce chapitre traite les notions de l'analyse des données de survie et les données censurées, où nous nous sommes intéressées au cas de censure aléatoire à droite avec l'hypothèse d'indépendance des durées de survie et les variables de censure. Nous étudions l'estimateur non paramétrique de la fonction de survie, en particulier l'estimateur de Kaplan-Meier, et nous abordons l'estimation non paramétrique d'une densité de probabilité dans un modèle censuré en utilisant la théorie des martingales.

Chapitre 3 : Ce chapitre est consacré à l'étude de l'estimateur à noyau (de type Parzen-Rosenblatt ; Parzen (1962), Rosenblatt (1956)) pour les dérivées de la densité, où nous fournissons les résultats de la convergence uniforme ainsi que les expressions du biais et de la variance pour cet estimateur.

Chapitre 4 : Nous présentons dans cet avant dernier chapitre, des résultats portant sur un estimateur non paramétrique des dérivées de la densité pour des données censurées.

Chapitre 5 : Ce chapitre comporte deux parties. La première partie présente brièvement les techniques dites de sélection du paramètre de lissage et expose les fondements des méthodes de validation croisée. La seconde partie est consacrée aux simulations numériques et à l'analyse des résultats obtenues. .

Nous terminons le mémoire par un rappel des résultats obtenues sous forme de conclusion, suivie de quelques perspectives.

Mots-clefs : *consistance forte, estimateur à noyau, dérivées de la densité, données censurées, censure à droite, estimateur de Kaplan Meier, estimation non paramétrique, analyse de survie, taux de convergence, validation croisée, paramètre de lissage.*

Chapitre 1

Introduction à l'Estimation Non Paramétrique

1.1 Estimation non paramétrique d'une fonction de répartition

1.1.1 La fonction de répartition empirique

Soit (Ω, A, \mathbb{P}) un espace de probabilité. Soit X une *v.a* positive de *f.d.r* F supposée inconnue, avec $F(x) = \mathbb{P}(X \leq x)$. Soit X_1, X_2, \dots, X_n un échantillon *i.i.d.* de F et $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ les observations ordonnées. Un bon estimateur pour F est la fonction de répartition empirique, notée F_n , et définie par

$$\begin{aligned} F_n(x) &= \frac{\text{nombre d'observations} \leq x}{n} \\ &= \frac{\#\{i : X_i \leq x\}}{n} \\ &= \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x\} \\ &= \frac{1}{n} \sum_{i=1}^n I\{X_{(i)} \leq x\} \\ &= \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{k}{n} & \text{si } X_{(k)} \leq x < X_{(k+1)} \quad k = 1, \dots, n-1 \\ 1 & \text{si } x \geq X_{(n)}. \end{cases} \end{aligned}$$

Propriétés 1.1. – Biais de l'estimateur $F_n(x)$

$$E \{F_n(x)\} = \frac{1}{n} \sum_{i=1}^n E \{I \{X_i \leq x\}\} = \mathbb{P} \{X \leq x\} = F(x)$$

Donc, pour tout point x , $F_n(x)$ est un estimateur sans biais de $F(x)$.

– Variance de l'estimateur $F_n(x)$

Pour tout x , la variance de l'estimateur $F_n(x)$ est donnée par :

$$\text{Var} \{F_n(x)\} = F(x)(1 - F(x)).$$

– La loi des grands nombres donne

$$\forall x \in \mathbb{R} : F_n(x) \xrightarrow{P} F(x), \text{ si } n \rightarrow \infty.$$

– Le théorème central-limite donne

$$\frac{nF_n(x) - nF(x)}{\sqrt{nF(x)(1 - F(x))}} \xrightarrow{L} N(0, 1)$$

$$\Rightarrow \sqrt{n}(F_n(x) - F(x)) \xrightarrow{L} N(0, F(x)(1 - F(x)))$$

– La distance de Kolmogorov-Smirnov est définie par

$$\sup_x |F_n(x) - F(x)|$$

1.1.2 La fonction quantile empirique

Le p ème quantile (ou quantile d'ordre p) de la population $F^{-1}(p) = \inf \{x : F(x) \geq p\}$ $0 < p < 1$ peut être estimé par $F_n^{-1}(p) = \inf \{x : F_n(x) \geq p\}$, le p ème quantile de la fonction de répartition empirique.

1.2 Estimation non paramétrique d'une fonction densité de probabilité

Nous allons décrire dans cette section quelques méthodes importantes d'estimation non-paramétrique d'une densité, en se basant sur les observations X_1, \dots, X_n .

1.2.1 Histogramme de densité

Il s'agit de l'estimateur le plus simple. Cet estimateur n'est pas conçu pour les variables aléatoires continues mais discrètes.

On choisit un point d'origine t_0 et une longueur de classe h ($h > 0$). Les classes sont définies par :

$$B_k = [t_k, t_{k+1}[, \quad k \in \mathbb{Z} \quad (\text{la } k \text{ ème classe})$$

Avec

$$t_{k+1} = t_k + h, \quad k \in \mathbb{Z}.$$

Un estimateur de f est donné par

$$\hat{f}_H(x) = \frac{1}{nh} \# \{i : X_i \text{ est dans la classe qui contient } x\}$$

Si nous notons le nombre d'observations dans une classe B_k par v_k , l'estimateur du type histogramme de densité s'écrit de la façon suivante :

$$\hat{f}_H(x) = \frac{v_k}{nh} = \frac{1}{nh} \sum_{i=1}^n I_{[t_k, t_{k+1}[}(X_i) \quad \text{pour } x \in B_k.$$

- L'histogramme de densité est un estimateur très élémentaire, il s'agit d'une fréquence relative locale selon une fenêtre rectangulaire, mais peut quand même déjà donner une première idée assez bonne de la forme de la densité f . Par contre, si on voulait utiliser cet estimateur dans d'autres analyses statistiques (comme par exemple l'analyse discriminante, l'estimation d'un taux de hasard, etc) il vaudrait mieux démarrer avec un estimateur plus précis.
- L'histogramme de densité est une fonction étagée, et donc discontinue.
- L'estimateur \hat{f}_H dépend de deux paramètres : le point d'origine t_0 et la largeur de classe h . Ces deux paramètres peuvent avoir une influence importante sur l'histogramme.

1.2.2 Estimateur simple

Rappelons que la densité de probabilité f est égale à la dérivée de la fonction de répartition F (si cette dérivée existe). On peut donc écrire

$$\begin{aligned} f(x) &= \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} \\ &= \lim_{h \rightarrow 0} \frac{P\{x-h < X \leq x+h\}}{2h} \end{aligned}$$

Un estimateur de $f(x)$ est alors

$$\begin{aligned}\widehat{f}(x) &= \frac{1}{2h} \frac{\#\{i : x-h < X_i \leq x+h\}}{n} \\ &= \frac{1}{2hn} \sum_{i=1}^n I\{x-h < X_i \leq x+h\} \\ &= \frac{1}{2hn} \sum_{i=1}^n I\left\{-1 \leq \frac{x-X_i}{h} < 1\right\}.\end{aligned}$$

Notons que cet estimateur peut encore s'écrire comme

$$\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x-X_i}{h}\right) \quad (1.1)$$

Où

$$w(y) = \begin{cases} 1/2 & \text{si } y \in [-1, 1[\\ 0 & \text{sinon} \end{cases}, \text{ la densité de probabilité uniforme sur } [-1, 1[.$$

Propriétés 1.2. Remarquons que

$$\widehat{f}(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}$$

Avec F_n la fonction de répartition empirique. Le paramètre de lissage h dépend de la taille de l'échantillon n , c'est-à-dire $h = h_n$.

Nous savons que

$$nF_n(x) = \sum_{i=1}^n I\{X_i \leq x\} \mapsto \text{Bin}(n, F(x))$$

Et

$$\begin{aligned}2nh_n\widehat{f}(x) &= nF_n(x+h_n) - nF_n(x-h_n) \mapsto \text{Bin}(n, F(x+h_n) - F(x-h_n)) \\ &\Rightarrow E\{2nh_n\widehat{f}(x)\} = n[F(x+h_n) - F(x-h_n)] \\ &\Rightarrow E\{\widehat{f}(x)\} = \frac{1}{2h_n}[F(x+h_n) - F(x-h_n)].\end{aligned}$$

Pour la variance nous trouvons :

$$\text{Var}\{2nh_n\widehat{f}(x)\} = n[F(x+h_n) - F(x-h_n)][1 - F(x+h_n) + F(x-h_n)]$$

$$\Rightarrow \text{Var} \left\{ \widehat{f}(x) \right\} = \frac{1}{4nh_n^2} [F(x+h_n) - F(x-h_n)] [1 - F(x+h_n) + F(x-h_n)].$$

Remarquons que si $n \rightarrow \infty$ et $h_n \rightarrow 0$, alors

$$E \left\{ \widehat{f}(x) \right\} \rightarrow f(x) \text{ et } nh_n \cdot \text{Var} \left\{ \widehat{f}(x) \right\} \rightarrow \frac{1}{2} f(x).$$

Le risque quadratique moyen de l'estimateur $\widehat{f}(x)$ de $f(x)$ est donné par

$$\begin{aligned} E \left\{ \widehat{f}(x) - f(x) \right\}^2 &= E \left\{ \widehat{f}(x) - E \left\{ \widehat{f}(x) \right\} + E \left\{ \widehat{f}(x) \right\} - f(x) \right\}^2 \\ &= \text{Var} \left\{ \widehat{f}(x) \right\} + \left[E \left\{ \widehat{f}(x) \right\} - f(x) \right]^2 \\ &= \text{Var} \left\{ \widehat{f}(x) \right\} + \left[\text{Biais} \left\{ \widehat{f}(x) \right\} \right]^2. \end{aligned}$$

Donc, si $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, on a

$$E \left\{ \widehat{f}(x) - f(x) \right\}^2 \rightarrow 0 \text{ pour tout point } x.$$

L'estimateur simple $\widehat{f}(x)$ est alors un estimateur consistant de $f(x)$.

Remarque 1.1. – On n'a plus le problème du choix d'un point d'origine (un point t_0) comme dans le cas d'un histogramme de densité.

– L'estimateur

$$\widehat{f}(x) = \frac{1}{2hn} \sum_{i=1}^n I \{x-h < X_i \leq x+h\} = \frac{1}{2hn} \sum_{i=1}^n I \{X_i - h \leq x < X_i + h\}$$

est une fonction discontinue, avec des discontinuités aux points $X_i h$, et constante entre ces points.

1.2.3 Estimateur à noyau

Il s'agit de l'estimateur le plus populaire. Il est adapté aux variables aléatoires continues (mais surtout pas aux *v.a* discrètes). L'estimateur simple (1.1) peut être généralisé en remplaçant la fonction de poids $w(\cdot)$ (la densité de probabilité uniforme) par une fonction de poids plus générale K (par exemple une densité de probabilité quelconque). Ceci résulte en l'estimateur de Parzen-Rosenblatt suivant :

$$\begin{aligned}\widehat{f}(x) &= \frac{1}{h} \int K\left(\frac{x-u}{h}\right) dF_n(u) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-X_i}{h}\right),\end{aligned}\tag{1.2}$$

où K est appelé fonction de poids (weight function) ou noyau (kernel function), et h est appelé paramètre de lissage (smoothing parameter) ou fenêtre (window width).

Selon cette définition, toute fonction K peut servir comme noyau d'estimation d'une densité f .

Les noyaux les plus couramment utilisés en pratique sont :

- Le noyau rectangulaire (noyau de Rosenblatt) :

$$K(u) = \frac{1}{2} 1_{[-1,1]}(u)$$

- Le noyau triangulaire :

$$K(u) = (1 - |u|) 1_{[-1,1]}(u)$$

- Le noyau d'Epanechnikov :

$$K(u) = \frac{3}{4} (1 - u^2) 1_{[-1,1]}(u)$$

- Le noyau gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-u^2/2}$$

Propriétés 1.3. – Les fonctions noyaux sont symétriques par rapport à l'axe O_y ,

$$\int_{-\infty}^{+\infty} K(u) du = 1 \text{ et } K(u) \geq 0.$$

– L'estimateur par noyau est une fonction de densité.

– \widehat{f} a les mêmes propriétés de continuité et de différentiabilité que K :

Si K est continue, \widehat{f} sera une fonction continue.

Si K est différentiable, \widehat{f} sera une fonction différentiable.

Si K peut prendre des valeurs négatives, alors \widehat{f} pourra aussi prendre des valeurs négatives.

– Au contraire de l'histogramme, l'estimateur par noyaux ne dépend pas d'un origine.

– Si la largeur de bande h et la fonction noyau sont connues préalablement, l'estimateur par noyau appliqué à un jeu de données est unique, et dépend de h et K .

1.2.3.1 Expressions du biais et de la variance

Considérons l'estimateur à noyau

$$\begin{aligned}\widehat{f}(x) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \\ &= \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),\end{aligned}\tag{1.3}$$

où nous avons introduit la notation $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ pour une version transformée de K . La moyenne de l'estimateur à noyau peut être écrite comme suit :

$$\begin{aligned}E\left\{\widehat{f}(x)\right\} &= E\left\{K_h(x - X)\right\} \\ &= \int K_h(x - y)f(y)dy,\end{aligned}\tag{1.4}$$

car les X_i sont identiquement distribuées.

La convolution entre deux fonctions f et g est définie par

$$(f * g)(x) = \int f(x - y)g(y)dy.$$

$$E\left\{\widehat{f}(x) - f(x)\right\} = \underbrace{(K_h * f)(x)}_{\text{version lissée de } f} - f(x).\tag{1.5}$$

La variance de l'estimateur à noyau peut être écrite comme suit :

$$\begin{aligned}Var\left\{\widehat{f}(x)\right\} &= \frac{1}{n}Var\left\{K_h(x - X)\right\} \\ &= \frac{1}{n}\left\{EK_h^2(x - X) - [EK_h(x - X)]^2\right\}\end{aligned}\tag{1.6}$$

En utilisant l'équation (1.4)

$$Var\left\{\widehat{f}(x)\right\} = \frac{1}{n}\left\{(K_h^2 * f)(x) - (K_h * f)^2(x)\right\}.\tag{1.7}$$

L'erreur quadratique moyenne (Mean squared error, MSE) de l'estimateur à noyau est donnée par :

$$\begin{aligned}
MSE \left\{ \widehat{f}(x) \right\} &= E \left\{ \widehat{f}(x) - f(x) \right\}^2 \\
&= Var \left\{ \widehat{f}(x) \right\} + \left\{ E[\widehat{f}(x)] - f(x) \right\}^2 \\
&= Var \left\{ \widehat{f}(x) \right\} + \left[Biais(\widehat{f}(x)) \right]^2.
\end{aligned} \tag{1.8}$$

En utilisant les équations (1.5) et (1.7), nous avons

$$\begin{aligned}
MSE \left\{ \widehat{f}(x) \right\} &= \frac{1}{n} \left\{ (K_h^2 * f)(x) - (K_h * f)^2(x) \right\} + \left\{ (K_h * f)(x) - f(x) \right\}^2 \\
&= \frac{1}{n} (K_h^2 * f)(x) + \left(1 - \frac{1}{n} \right) (K_h * f)^2(x) - 2(K_h * f)(x)f(x) + f^2(x).
\end{aligned} \tag{1.9}$$

L'expression exacte de l'erreur quadratique moyenne intégrée (Mean Integrated Squared Error, $MISE = IMSE$) peut être obtenue à partir de

$$MISE \left\{ \widehat{f} \right\} = \int MSE \left\{ \widehat{f}(x) \right\} dx$$

et est égale à

$$MISE \left\{ \widehat{f}(x) \right\} = \frac{1}{n} \int \left\{ (K_h^2 * f)(x) - (K_h * f)^2(x) \right\} dx + \int \left\{ (K_h * f)(x) - f(x) \right\}^2 dx \tag{1.10}$$

$$\begin{aligned}
MISE \left\{ \widehat{f}(x) \right\} &= \frac{1}{n} \int (K_h^2 * f)(x) dx + \left(1 - \frac{1}{n} \right) \int (K_h * f)^2(x) dx - 2 \int (K_h * f)(x)f(x) dx \\
&\quad + \int f^2(x) dx
\end{aligned}$$

Comme

$$\begin{aligned}
\int (K_h^2 * f)(x) dx &= \int \frac{1}{h^2} \left\{ \int K^2 \left(\frac{x-y}{h} \right) f(y) dy \right\} dx \\
&= \frac{1}{h} \int \int K^2(u) f(x - uh) du dx, \text{ avec } u = \frac{x-y}{h} \\
&= \frac{1}{h} \int K^2(u) \left\{ \int f(x - uh) dx \right\} du \\
&= \frac{1}{h} \int K^2(u) du,
\end{aligned}$$

Nous trouvons

$$MISE \left\{ \widehat{f}(x) \right\} = \frac{1}{nh} \int K^2(u) du + \left(1 - \frac{1}{n} \right) \int (K_h * f)^2(x) dx - 2 \int (K_h * f)(x) f(x) dx + \int f^2(x) dx.$$

Malgré le fait qu'on ait des expressions exactes pour $MSE \left\{ \widehat{f}(x) \right\}$ et $MISE \left\{ \widehat{f}(x) \right\}$, ces expressions ne sont pas très attrayantes, car elles dépendent de manière très complexe du paramètre de lissage h . Pour cette raison on cherche des expressions asymptotiques qui pourraient dépendre de h de manière plus simple.

1.2.3.2 Expressions asymptotiques du biais et de la variance

Une approximation asymptotique de l'espérance de l'estimateur $\widehat{f}(x)$ est donnée sous les conditions suivantes sur f , h et K

1. La dérivée seconde $f''(x)$ est continue, de carré intégrable.
2. $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} nh = 0$;
3. Pour que $\widehat{f}(x)$ soit une densité, on suppose que $K(u) \geq 0$ et $\int K(u) du = 1$. La fonction noyau est supposée être symétrique autour de zéro, i.e, $\int uk(u) du = 0$, et possède un moment d'ordre 2 fini, i.e, $\int u^2 k(u) du < \infty$.

$$E \left\{ \widehat{f}(x) \right\} = \int K_h(x - y) f(y) dy$$

En utilisant l'équation (1.4) et un changement de variable, nous avons

$$E \left\{ \widehat{f}(x) \right\} = \int K(u) f(x - uh) du, \text{ avec } u = \frac{x - y}{h}, du = -\frac{1}{h} dy$$

En utilisant un développement de Taylor, nous avons

$$E \left\{ \widehat{f}(x) \right\} = \int K(u) \left[f(x) - f'(x)uh + \frac{1}{2} f''(x)u^2 h^2 + \dots \right] du \quad (1.11)$$

$$E \left\{ \widehat{f}(x) \right\} = f(x) \int K(u) du - f'(x)h \int K(u)u du + \frac{1}{2} f''(x)h^2 \int K(u)u^2 du + o(h^2). \quad (1.12)$$

D'après l'hypothèse (3), le noyau K satisfait

$$K \geq 0, \int K(u) du = 1, \int K(u)u du = 0, 0 < \mu_2 = \int K(u)u^2 du < \infty.$$

Alors

$$\begin{aligned} E \left\{ \widehat{f}(x) \right\} - f(x) &= \frac{1}{2} f''(x) h^2 \int K(u) u^2 du + o(h^2) \\ &= \frac{1}{2} f''(x) h^2 \mu_2 + o(h^2). \end{aligned} \quad (1.13)$$

Comme

$$Var \left\{ \widehat{f}(x) \right\} = \frac{1}{n} \left\{ EK_h^2(x - X) - [EK_h(x - X)]^2 \right\}$$

Et

$$\begin{aligned} EK_h^2(x - X) &= \frac{1}{h^2} \int K^2 \left(\frac{x - y}{h} \right) f(y) dy \\ &= \frac{1}{h} \int K^2(u) f(x - uh) du, \text{ avec } u = \frac{x - y}{h} \\ &= \frac{1}{h} \int K^2(u) [f(x) - f'(x)hu + \dots] du, \quad \text{par le développement de Taylor} \\ &= \frac{1}{h} f(x) \int K^2(u) du - f'(x) \int K^2(u) u du + o(1) \end{aligned}$$

Nous trouvons que

$$Var \left\{ \widehat{f}(x) \right\} = \frac{1}{nh} f(x) \int K^2(u) du + o \left(\frac{1}{nh} \right). \quad (1.14)$$

Nous avons donc établi que

$$\begin{aligned} \text{Biais} \left\{ \widehat{f}(x) \right\} &= \frac{1}{2} f''(x) \mu_2 h^2 + o(h^2) \\ Var \left\{ \widehat{f}(x) \right\} &= \frac{1}{n} f(x) R(K) + o \left(\frac{1}{nh} \right) \end{aligned} \quad (1.15)$$

Où $\mu_2 = \int K(u) u^2 du$ et $R(g) = \int g^2(u) du$ pour une fonction g de carré intégrable.

En se basant sur l'hypothèse (2), nous avons

Si $h = h_n \rightarrow 0$ quand $n \rightarrow \infty$, alors

$$\text{Biais} \left\{ \widehat{f}(x) \right\} \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Si $h = h_n \rightarrow 0$ et $nh_n \rightarrow \infty$, alors

$$Var \left\{ \widehat{f}(x) \right\} \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Remarquons que

Si h décroît alors le $(\text{biais})^2$ décroît et la variance croît.

Si h augmente alors le $(\text{biais})^2$ croît et la variance décroît.

Il faut donc essayer de choisir un h qui fasse un compromis entre le $(\text{biais})^2$ et la variance.

Les expressions asymptotiques du biais et de la variance nous permettent de trouver des expressions asymptotiques pour la MSE et la $MISE$.

Ces expressions ont été obtenues sous la condition (3) sur K et en supposant que la densité de probabilité f avait toutes les dérivées (continues) nécessaires.

A partir de l'équation (1.15) on peut obtenir facilement les approximations asymptotiques suivantes pour la MSE et la $MISE$.

$$MSE \left\{ \widehat{f}_n(x) \right\} = \frac{1}{4} h^4 \mu_2^2 \{f''(x)\}^2 + \frac{1}{nh} f(x) R(K) + o\left(h^4 + \frac{1}{nh}\right) \quad (1.16)$$

$$MISE \left\{ \widehat{f}_n(x) \right\} = \frac{1}{4} h^4 \mu_2^2 \int \{f''(x)\}^2 dx + \frac{1}{nh} R(K) + o\left(h^4 + \frac{1}{nh}\right), \quad (1.17)$$

sous des conditions appropriées d'intégrabilité de f et ses dérivées.

On note l'approximation asymptotique de la MSE par

$$AMSE \left\{ \widehat{f}_n(x) \right\} = \frac{1}{4} h^4 \mu_2^2 \{f''(x)\}^2 + \frac{1}{nh} f(x) R(K) \quad (1.18)$$

et l'approximation asymptotique de la $MISE$ par

$$AMISE \left\{ \widehat{f}_n(\cdot) \right\} = \frac{1}{4} h^4 \mu_2^2 R(f'') + \frac{1}{nh} R(K) \quad (1.19)$$

1.2.3.3 Avantage de l'estimateur à noyau

L'estimateur à noyau :

- A une forme flexible
- Réduit les inconvénients de l'histogramme (disparité des observations, tracé de la courbe...)
- S'écrit de manière simple et conviviale
- Ne dépend que de la largeur de bande et de la fonction noyau
- Meilleure prise en compte de la contribution des observations situées dans le voisinage d'un point x
- Possède un biais qui converge plus rapidement que celui de l'histogramme

1.2.3.4 Choix théoriques optimaux du paramètre de lissage

Pour le paramètre de lissage on fait la distinction entre h paramètre de lissage constant (ou global), et $h(x)$ paramètre de lissage variable (local).

Ces choix différents du paramètre de lissage résultent en les estimateurs à noyau suivants :

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

$$\widehat{f}_{n,L}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h(x)} K\left(\frac{x - X_i}{h(x)}\right)$$

Nous allons décrire des choix théoriques optimaux des paramètres de lissage h et $h(x)$.

Un critère approprié pour sélectionner un paramètre de lissage constant h est la *MISE*. Le paramètre de lissage optimal est la valeur de h qui minimise la *MISE*. Notons cette valeur par h_{MISE} .

Une approximation asymptotique de h_{MISE} est donnée par h_{AMISE} , la valeur de h qui minimise $AMISE\{\widehat{f}_n(\cdot)\}$.

Il est facile de vérifier à partir de (1.5) que

$$h_{AMISE} = \left\{ \frac{R(K)}{\mu_2^2 R(f'')} \right\}^{1/5} n^{-1/5}$$

Et

$$h_{MISE} \approx \left\{ \frac{R(K)}{\mu_2^2 R(f'')} \right\}^{1/5} n^{-1/5},$$

c'est-à-dire

$$\lim_{n \rightarrow \infty} \frac{h_{MISE}}{h_{AMISE}} = 1$$

.

Remarquons que si f montre des changements rapides, alors $R(f'')$ sera grand, et h_{AMISE} sera petit.

Un critère approprié pour sélectionner un paramètre de lissage variable (local) $h(x)$ est la mesure de performance locale $MSE \widehat{f}_{n,L}(x)$.

Nous introduisons les notations suivantes :

$$h_{MSE} = \arg \min_h MSE \widehat{f}_{n,L}(x)$$

Et

$$h_{AMSE} = \arg \min_h AMSE \widehat{f}_{n,L}(x).$$

sous condition que $f''(x) \neq 0$.

Les choix h_{AMISE} et $h_{AMSE}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités inconnues f et f'' .

En substituant h_{AMISE} dans l'expression de l'AMISE, on montre que pour l'estimateur à noyau

$$n^{\frac{4}{5}} AMISE \hat{f}_{h_{AMISE}} = O(1)$$

Alors que pour l'histogramme, nous avons :

$$n^{\frac{2}{3}} AMISE \hat{f}_{h_{AMISE}} = O(1)$$

Nous remarquons ainsi que la vitesse de convergence de l'estimateur à noyau est supérieur à celle de l'histogramme.

Chapitre 2

Données Censurées

2.1 Introduction

L'analyse de survie est un domaine de la statistique qui a pour objet l'étude de la durée qui s'écoule entre un temps d'origine et un certain évènement d'intérêt. L'évènement typique –qui donne son nom à la discipline– est le décès.

Toutefois, les méthodes qui se regroupent sous le terme de " analyse de survie " peuvent être (et ont été) appliquées à l'étude de toute une variété d'évènements, comme par exemple l'occurrence d'une maladie et sa récurrence, le décès après une intervention chirurgicale (biostatistique), la naissance d'un enfant (sciences sociales), la panne d'une machine (ingénierie).

La variable d'intérêt en analyse de survie est donc le temps. Il s'agit d'une variable toujours positive et d'habitude continue.

La principale source de difficulté dans l'analyse des durées de vie, et pour diverses raisons, est la présence de données incomplètes pour lesquelles la variable d'intérêt n'est pas complètement observée pour toutes les données de l'échantillon. Nous présentons dans ce qui suit deux cas de données incomplètes : celles censurées et celle tronquées.

2.1.1 Les données tronquées

Il existe plusieurs types de troncature dont la troncature à gauche , la troncature à droite et la troncature par intervalle.

2.1.1.1 La troncature à gauche

Soit X la variable d'intérêt, et soit T une variable aléatoire indépendante de X . On dit qu'il y a troncature à gauche lorsque X n'est observable que si $X > T$. On observe le couple (X, T) , avec $X > T$.

2.1.1.2 La troncature à droite

De même, il y a troncature à droite lorsque X n'est observable que si $X < T$.

2.1.1.3 La troncature par intervalle

Quand une durée est tronquée à droite et à gauche, on dit qu'elle est tronquée par intervalle.

Lyndell-Bell [23] proposa une estimation non paramétrique de la *f.d.r* de X et Woodroffe [44] posa le cadre de l'estimation dans le modèle de troncature.

2.1.2 Les données censurées

Par définition, le temps de survie d'un individu est dit censuré, lorsque sa valeur exacte n'est pas observée ; seules des bornes supérieures ou inférieures pour cette valeur sont disponibles.

La censure peut se manifester pour différentes raisons : l'évènement d'intérêt n'est pas survenu au moment de l'analyse, un sujet peut être perdu de vue avant d'avoir expérimenté l'évènement d'intérêt, un évènement concurrent peut être survenu avant l'évènement d'intérêt, un sujet peut être exclu de l'étude sans avoir expérimenté l'évènement d'intérêt, etc.

Il existe plusieurs types de censure dont la censure à droite, la censure à gauche et la censure par intervalle.

2.1.2.1 Censure à droite

Une durée de vie est dite censurée à droite si l'individu n'a pas encore subi l'évènement lors de sa dernière observation. Dans ce cas, les durées de vie ne sont pas toutes observées. La seule information dont on dispose est que ces durées sont supérieures à une certaine valeur. En d'autres mots, il ya une borne inférieur pour la durée de vie.

On distingue trois types de censure à droite :

Censure de type I : fixée

Soit C une valeur fixée, au lieu d'observer les variables X_1, X_2, \dots, X_n qui nous intéressent, on n'observe X_i uniquement que lorsque $X_i \leq C$, sinon on sait uniquement que $X_i > C$. On observe donc une variable Z_i telle que $Z_i = \min(X_i, C)$.

Censure de type II : Attente (jusqu'au k-ième décès)

On observe les durées de vie de n patients jusqu'à ce que k d'entre eux soient décédés et d'arrêter l'étude à ce moment. Si on ordonne les durées Y_1, Y_2, \dots, Y_n , on obtient les statistiques d'ordre $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. La date de la censure est alors $X_{(k)}$ et on observe $Z_{(1)} = X_{(1)}, Z_{(2)} = X_{(2)}, \dots, Z_{(k)} = X_{(k)}, Z_{(k+1)} = X_{(k)}, \dots, Z_{(n)} = X_{(k)}$.

Censure de type III : aléatoire

On associe à chaque patient i , non seulement son temps de survie X_i mais aussi son temps de censure C_i . Ce temps de censure sera propre à chaque patient et peut-être dû à plusieurs raisons dont :

- Perte de vue : le patient ne revient plus dans l'étude.
- Arrêt du traitement : il peut y avoir des effets secondaires si désastreux pour le patient qu'on est obligé d'arrêter le traitement.
- Fin de l'étude : l'étude se termine alors que certains patients sont toujours vivants.

Ce qu'on observe alors c'est le couple (Z_i, δ_i) où $Z_i = \min(X_i, C_i)$ et $\delta_i = \mathbb{I}_{\{X_i \leq C_i\}}$ (l'indicatrice de non censure).

2.1.2.2 Censure à gauche

La censure à gauche correspond au cas où l'individu a déjà subi l'évènement avant qu'on ne l'observe. Dans ce cas, la seule information dont on dispose est que ces durées sont inférieures à une certaine valeur. En d'autres mots, il ya une borne supérieur pour la durée de vie. Pour chaque individu, on peut associer un couple de variables aléatoires (Z, δ) :

$$Z = X \vee C = \max(X, C), \quad \delta = \mathbb{I}_{\{X > C\}}.$$

On suppose que la censure C est indépendante de X .

2.1.2.3 Censure par intervalle

Une date est censurée par intervalle si au lieu d'observer avec certitude le temps de l'évènement, la seule information connue est un intervalle de temps contenant la durée de survie. En général, la censure par intervalle intervient lorsque les sujets sont suivis de manière discontinue.

Seule la censure à droite peut être considérée pour l'application des méthodes présentées dans ce mémoire.

2.1.3 Fonctions associées aux distributions de survie

On indique avec X ($X \geq 0$), la variable aléatoire qui représente la durée de vie d'une unité dans une expérience. On se limite au cas d'une variable X continue.

La loi de probabilité de X peut être caractérisée par plusieurs fonctions qui sont, en fait, interreliées, car la connaissance d'une seule est suffisante pour dériver les autres.

a) La fonction de densité

La fonction de densité de probabilité de X , notée $f(t)$, est définie par :

$$f(t) = \lim_{dt \rightarrow 0^+} \frac{\mathbb{P}(t \leq X < t + dt)}{dt}$$

b) La fonction de répartition

La fonction de répartition, notée $F(t)$, est définie comme la probabilité de décéder entre 0 et t :

$$F(t) = \mathbb{P}(X \leq t) = \int_0^t f(u) du$$

La fonction de répartition $F(t)$ est croissante et telle que $\lim_{t \rightarrow 0^+} F(t) = 0$ et $\lim_{t \rightarrow \infty} F(t) = 1$.

c) La fonction de survie

La fonction de survie, notée $S(t)$, est définie comme la probabilité de survie jusqu'à un certain temps t :

$$S(t) = \mathbb{P}(X > t) = 1 - F(t)$$

De plus, si la dérivée de $S(t)$ existe, la densité $f(t)$ de la variable aléatoire X est estimée comme le rapport entre les individus décédés dans l'intervalle et la largeur de l'intervalle. Ainsi, la fonction de survie est l'intégrale de la fonction de densité $f(t)$:

$$S(t) = \int_0^{\infty} f(u) du$$

Donc, $f(t) = \frac{-dS(t)}{dt}$.

La fonction de survie $S(t)$ est décroissante et telle que $\lim_{t \rightarrow 0^+} S(t) = 1$ et $\lim_{t \rightarrow \infty} S(t) = 0$.

d) La fonction de risque

La fonction de risque, notée $\lambda(t)$, est définie comme la probabilité pour un sujet de décéder au temps t sachant que ce sujet est encore vivant juste avant t :

$$\lambda(t) = \lim_{dt \rightarrow 0^+} \frac{\mathbb{P}(t \leq X < t + dt / X \geq t)}{dt}$$

Cette fonction est aussi appelée fonction de risque instantané de décès ou taux de panne, taux de décès conditionnel ou force de mortalité.

Elle peut aussi s'écrire :

$$\lambda(t) = \frac{f(t)}{S(t)}$$

et on en déduit que

$$S(t) = \exp\left(-\int_0^t \lambda(u) du\right)$$

La fonction de risque donne habituellement plus d'informations sur le mécanisme de décès que la fonction de survie. C'est pourquoi cette fonction est souvent employée pour résumer les données de survie.

e) La fonction de risque cumulé

La fonction de risque cumulé ou de hasard intégré, notée $\Lambda(t)$, est définie par :

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln[S(t)]$$

et son domaine se situe entre 0 et l'infini.

On peut déduire la fonction de survie du taux de hasard cumulé grâce à la relation

$$S(t) = \exp\{-\Lambda(t)\}$$

Quelques quantités associées à la loi de survie

1. Les quantiles de la durée de survie

Pour $0 < p < 1$, on définit le quantile t_p et la fonction $q(p)$ $p \in (0, 1)$ comme

$$t_p \equiv q(p) = \inf \{t : F(t) \geq p\}.$$

Quand $F(t)$ est strictement croissante et continue alors

$$t_p = q(p) = F^{-1}(p), 0 < p < 1.$$

Pour p fixé, le quantile t_p est le temps auquel une proportion p de la population a disparu.

2. Moyenne et variance de la durée de survie

Le temps moyen de survie $E(X)$ ainsi que sa variance $V(X)$ sont des quantités importantes :

$$E(X) = \int_0^\infty S(t)dt, V(X) = 2 \int_0^\infty tS(t)dt - \{E(X)\}^2.$$

La moyenne et la variance peuvent être déduites de n'importe laquelle des cinq fonctions ci-dessus ($F, S, f, \lambda, \Lambda$), mais pas vice versa.

2.2 Estimation de la fonction de survie dans un modèle complet

Nous avons déjà vu, dans le chapitre précédent, que $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{T_i \leq t\}} \forall t \geq 0$ et comme $S(t) = 1 - F(t)$, alors :

$$\hat{S}_n(t) = 1 - \hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{T_i > t\}}$$

Propriétés 2.1. *Absence de biais :*

$$E(\hat{S}_n(t)) = S(t)$$

Convergence uniforme presque sûre (Glivenko-Cantelli) :

$$\sup_{t \in \mathbb{R}} \left| \hat{S}_n(t) - S(t) \right| \xrightarrow[n \rightarrow \infty]{} 0$$

Normalité asymptotique :

$$\sqrt{n} \left(\hat{S}_n - S \right) \xrightarrow{L} W$$

Où W est un processus gaussien centré (pont Brownien) de fonction de var-cov :

$$\rho(s, t) = F(s) \wedge F(t) - F(s)F(t).$$

2.3 Estimation de la fonction de survie dans un modèle censuré

Soit (Ω, A, \mathbb{P}) un espace de probabilité. Nous disposons de copies (X, C) où X désigne la *v.a* d'intérêt. Celle-ci étant interprétée comme une durée de survie, il est naturel de supposer qu'elle ne prend que des valeurs positives. Ici, on désigne par C le temps (aléatoire) de censure.

Nous supposons l'indépendance entre X et C . L'échantillon observé est alors (Z_i, δ_i) , $i = 1, \dots, n$ pour une taille d'échantillon $n \geq 1$, où pour $i = 1, \dots, n$,

$$\begin{cases} Z_i = X_i \wedge C_i \\ \delta_i = \mathbb{I}_{\{X_i \leq C_i\}} \end{cases} \quad (2.1)$$

$\mathbb{I}_{(E)}$ désigne la fonction indicatrice de E .

Au vu de (2.1), Z_i s'interprète comme le temps de participation (ou le temps de suivi) du patient dans l'étude.

L'indicatrice δ_i vaut 1 en l'absence de censure, lorsque $Z_i = X_i$, et 0 autrement.

Pour $t \in \mathbb{R}$, soient $F(t) := \mathbb{P}(X \leq t)$ et $G(t) := \mathbb{P}(C \leq t)$ les (versions continues à droite des) fonctions de répartition de X et C .

L'hypothèse d'indépendance entre X et C implique alors que

$$H(t) = \mathbb{P}(Z \leq t) = 1 - (1 - F(t))(1 - G(t)).$$

2.3.1 Problème d'identifiabilité dans le modèle de censure

Considérons le modèle de censure défini par (2.1). Si on connaît la loi de l'observation de Z , cette dernière permet-elle d'identifier la loi exacte de la variable d'intérêt X , en d'autre terme, si l'on connaissait parfaitement la loi des observations, pourrait-on en déduire de manière unique la loi de X (par exemple, sa densité ou sa survie)? La réponse n'est pas toujours positive, car le fait de disposer de données incomplètes entraîne une perte d'information.

Dans le cas général, si durées et censures ne sont pas indépendantes, il n'est pas possible d'identifier la loi de X à partir des observations. Par contre, lorsque les durées et les censures sont indépendantes, l'identification est possible.

Pour toute fonction de répartition L , nous noterons $\tau_L = \sup \{t : L(t) < 1\}$.

D'autre part $L_n(\cdot)$ désignera un estimateur fonctionnel de $L(\cdot)$. Enfin, nous poserons $\bar{L}(\cdot) = 1 - L(\cdot)$.

Proposition 2.1. *Dans le cas d'une censure à droite aléatoire C , si X et C sont indépendantes de f.d.r F et G respectives et si $\tau_F \leq \tau_G$, alors la loi de X est identifiable à partir de la loi des observations (Z_i, δ_i) .*

Preuve. *Supposons que l'on connaisse la loi de (Z_i, δ_i) . Notons \bar{F} et \bar{G} les fonctions de survie respectives de X et C . Pour tout t on a*

$$\mathbb{P}(Z > t) = \mathbb{P}(X > t) \cdot \mathbb{P}(C > t) = \bar{F}(t) \cdot \bar{G}(t)$$

On identifie (et donc on connaît) le produit $\bar{F} \cdot \bar{G}$ en tout t , par ailleurs, on connaît la fonction

$$\mathbb{P}(Z > t, \delta = 1) = \mathbb{P}(X > t, X < C) = \int_t^\infty (1 - G(u)) f(u) du.$$

Ce qui permet d'identifier $f \cdot \bar{G}$ puis $\lambda = f/\bar{F}$ qui est la fonction de hasard.

2.3.2 Estimateur de Kaplan-Meier

Si l'on ne peut pas supposer à priori que la loi de la durée de survie obéit à un modèle paramétrique, on peut estimer la fonction de survie S grâce à plusieurs méthodes non paramétriques dont la plus intéressante est celle de Kaplan-Meier [20] (aussi appelée méthode du produit-limite).

On définit, au temps t , l'estimateur standard de Kaplan-Meier comme suit :

$$\hat{S}_{KM}(t) = \prod_{Z_i \leq t} \left(1 - \frac{M_i}{R_i}\right) \quad \forall t < Z_{(1)}, \quad \hat{S}_{KM}(t) = 1.$$

Les instants de discrétisation sont les instants, aléatoires, d'observations :

$$Z_{(0)} = 0 \leq Z_{(1)} \leq \dots \leq Z_{(n)}$$

M_i = nombres de morts à $Z_{(i)}$

R_i = nombres de sujets à risque à $Z_{(i)}$ (ni morts ni censuré juste avant $Z_{(i)}$)

Remarquons que :

$$M_i = M(Z_{(i)}) = \sum_{j=1}^n \mathbb{I}_{\{X_j = Z_{(i)}\}};$$

$$R_i = R(Z_{(i)}) = \sum_{j=1}^n \mathbb{I}_{\{Z_j > Z_{(i)}\}} = n - \sum_{j=1}^n \mathbb{I}_{\{Z_j \leq Z_{(i-1)}\}} = \sum_{j=1}^n \mathbb{I}_{\{Z_j \geq Z_{(i)}\}}.$$

Expression en absence d'ex-aequos

Toujours vrai si les lois des pannes et censures sont continues :

$$\begin{aligned}\widehat{S}_{KM}(t) &= \prod_{i=1}^n \left(1 - \frac{\delta_{(i)}}{n-i+1}\right)^{\mathbb{I}_{\{Z_i \leq t\}}} \\ &= \prod_{\substack{i=1 \\ \{Z_i \leq t\}}}^n \left(1 - \frac{\delta_{(i)}}{n-i+1}\right) \\ &= \prod_{\substack{i=1 \\ \{Z_i \leq t\}}}^n \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}\end{aligned}$$

L'estimateur de Kaplan-Meier $\widehat{S}_{KM}(t)$ est une fonction en escalier décroissante, constante par morceaux et continue à droite, qui ne saute qu'aux instants de temps morts réelles. En l'absence de censure, elle est égale à la fonction de survie empirique.

L'estimateur de KM est l'estimateur du maximum de vraisemblance non-paramétrique de la survie dans l'espace des fonctions des répartitions (de dimension infinie).

La méthode de Kaplan-Meier repose sur les principes suivant :

- Elle ne s'appuie sur aucun regroupement de données.
- les intervalles de temps sont déterminés à posteriori par les moments de décès observés : les probabilités conditionnelles de survie entre deux dates de décès sont estimées

Une importante propriété de cet estimateur de la fonction de survie est qu'il est asymptotiquement distribué comme une loi normale de moyenne $S(t)$.

Par ailleurs, la variance de l'estimateur de Kaplan-Meier pour chaque temps t est estimée par la formule de Greenwood,

$$\widehat{\sigma}_{KM}^2(t) = \widehat{S}_{KM}^2(t) \sum_{Z_{(i)} \leq t} \frac{M_i}{R_i(R_i - M_i)}$$

En l'absence d'ex-Aequos :

$$\widehat{\sigma}_{KM}^2(t) = \widehat{S}_{KM}^2(t) \sum_{Z_{(i)} \leq t} \frac{\delta_i}{(n-i)(n-i+1)}$$

L'estimateur de Greenwood est un estimateur consistant de la variance.

2.4 Propriétés de l'estimateur de Kaplan-Meier

2.4.1 Cohérence

L'estimateur de Kaplan-Meier $\widehat{S}_{KM}(t)$ de la fonction de survie S est dit "*cohérent*" si, pour tout t il vérifie

$$\widehat{S}_{KM}(t) = \frac{1}{n} \left[\sum_{i=1}^n 1_{\{Z_i > t\}} + \sum_{i=1}^n 1_{\{Z_i \leq t, \delta_i = 0\}} \cdot \frac{\widehat{S}_{KM}(t)}{\widehat{S}_{KM}(Z_i)} \right]$$

La probabilité de décéder au-delà de la date t est la somme des probabilités de n'être ni décédé, ni censuré à cette date et celle d'avoir été censuré avant la date t , et être toujours en vie en t , d'où le terme $\frac{\widehat{S}_{KM}(t)}{\widehat{S}_{KM}(Z_i)}$.

L'estimateur de Kaplan-Meier est **l'unique** estimateur *cohérent* de la survie. En particulier, lorsqu'il n'y a pas de censure, $\widehat{S}_{KM}(t)$ se réduit à l'estimateur non paramétrique usuel de S : la fonction de répartition empirique.

2.4.2 Loi fortes

En analyse des données de survie, $\widehat{S}_{KM}(t)$ joue pour les données incomplètes le même rôle que la fonction de répartition empirique pour les données classiques. C'est ainsi qu'il possède beaucoup de propriétés analogues à celles de la fonction de répartition empirique.

Rappelons qu'on se place toujours dans le cadre de la censure aléatoire droite, et que H , la survie des observations Z , vérifie $H = S.G$ (G survie de la censure droite C). Soit A l'ensemble des atomes de Z , i.e. l'ensemble des points de discontinuités de H . Soit $\tau_H = \inf \{x \geq 0 / H(x) = 0\} \in \overline{\mathbb{R}} = \mathbb{R} \cup +\infty$ la borne supérieure du support de H , et on notera $\Delta S(a) = S(a) - S(a-)$ le saut éventuel de S en a . Alors,

Théorème 2.1 (Stute [41]). *Si S et G n'ont pas de discontinuité en commun, pour toute fonction mesurable ϕ telle que $\int |\phi| dS < \infty$, on a presque sûrement et dans $L^1(\mathbb{R})$*

$$\begin{aligned} \lim_{n \rightarrow \infty} \int \phi d\widehat{S}_{KM} &= \int_{x \notin A, x < \tau_H} \phi(x) S(dx) + \sum_{a \in A} \phi(a) \Delta S(a) \\ &= \int_{x < \tau_H} \phi(x) S(dx) + \mathbb{I}_{\{\tau_H \in A\}} \phi(\tau_H) \Delta S(\tau_H), \end{aligned}$$

et le dernier terme est nul si $\tau_H = +\infty$.

Notons que lorsqu'il existe des discontinuités communes à S et G , on ne peut pas estimer ces distributions, sans hypothèses supplémentaires.

Remarque 2.1. *Si l'on n'y a pas de censure (ou plus exactement quand la censure est portée "à l'infini", telle une masse de Dirac en un point t_0 qui tend vers $+\infty$), on retrouve le résultat classique*

$$\int \phi dS_n \xrightarrow[n \rightarrow \infty]{} \int \phi dS,$$

presque sûrement et dans L^1 . Mais il est remarquable que cette formule reste "quasiment vraie" sur données censurées; en particulier, Si S est continue en τ_H ,

$$\int \phi \widehat{S}_{KM}(dt) \xrightarrow[n \rightarrow \infty]{} \int_{x < \tau_H} \phi S(dt) = \int_{x \leq \tau_H} \phi S(dt).$$

Précisons la notation $\int \phi d\widehat{S}_{KM}$, avec les notations précédentes,

$$\int \phi d\widehat{S}_{KM} = \sum_{i=1}^k \phi(Z_{(i)}) \Delta \widehat{S}_{KM}(Z_{(i)}),$$

$$\Delta \widehat{S}_{KM}(Z_{(i)}) = \widehat{S}_{KM}(Z_{(i)}) - \widehat{S}_{KM}(Z_{(i)}^-) = -\frac{M_i}{R_i} \widehat{S}_{KM}(Z_{(i)}^-).$$

On en déduit un estimateur de la durée de vie moyenne lorsque les données sont censurées.

Corollaire 2.1 (Fermanian [12]). *Si S et G n'ont pas de discontinuité commune, et si $\int |x| S(dx) < \infty$, on a alors presque sûrement et dans L^1*

$$E_{\widehat{S}_{KM}} [X - t/X > t] = \widehat{S}_{KM}^{-1}(t) \int_t^\infty (u - t) \widehat{S}_{KM}(du)$$

$$\xrightarrow[n \rightarrow \infty]{} \begin{cases} E_S [1_{\{X < \tau_H\}}(X - t)/X > t] & \text{si } \Delta S(\tau_H) < 0 \text{ (mais alors } G(\tau_H^-) = 0) \\ E_S [1_{\{X \leq \tau_H\}}(X - t)/X > t] & \text{sinon} \end{cases}$$

On a également l'équivalent du théorème de Glivenko-Cantelli lorsque les données sont censurées.

Théorème 2.2 (Fermanian[12]). *Si S et G n'ont pas de discontinuité en commun, on a presque sûrement*

$$\sup_{t \leq \tau_H} \left| \widehat{S}_{KM}(t) - \widetilde{S}(t) \right| \xrightarrow[n \rightarrow \infty]{} 0.$$

où $\widetilde{S}(t) = S(t)$ si $t < \tau_H$, et $\widetilde{S}(t) = S(\tau_H^-) + \mathbb{I}_{\{\tau_H \in A\}} \Delta S(\tau_H)$ si $t \geq \tau_H$.

Remarquons que ce résultat nous dit que le théorème (2.1) est vrai uniformément sur la classe de fonctions $F = \{\phi_t = \mathbb{I}_{]-\infty, t]}, t \in \mathbb{R}\}$.

En pratique, on cherche à estimer la survie sur des intervalles $[0, \tau]$, $\tau < \tau_H$, pour éviter les problèmes de biais aux bornes, et travailler ainsi avec une limite "classique" proche du cas standard *i.i.d.* Les limites obtenues dans les théorèmes (2.1) et (2.2) s'expriment alors très simplement. Lorsque τ_H est inconnue, un principe de prudence consiste à ne pas choisir $\tau_H \gg Z_{(n)}$ (car on sait que $\tau_H \geq Z_{(n)}$).

De plus, on a le résultat suivant plus précis que le précédent sur l'intervalle $[0, \tau]$, $\tau < \tau_H$.

Théorème 2.3 (Lo, Marck et Wang [24]). *Si S est continue sur $[0, \tau]$, $\tau < \tau_H$, alors il existe une constante C_0 telle que presque sûrement*

$$\left(\frac{n}{\ln(\ln n)}\right)^{1/2} \sup_{x \in [0, \tau]} \left| \widehat{S}_{KM}(t) - S(t) \right| < C_0.$$

2.4.3 Convergence en loi

Nous sommes toujours dans le cadre de la censure à droite. Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable telle que

$$\int \phi^2 dF < +\infty.$$

On peut sous des hypothèses assez faibles énoncer un théorème central limite pour toute quantité $\int \phi d\widehat{S}_{KM}$, appelée intégrale de Kaplan-Meier.

Théorème 2.4 (Akritas [1]). *Si $\max(Z_1, \dots, Z_n) < \tau_H$ ou $\phi(\tau_H) = 0$, et si*

$$\int_{-\infty}^{\tau_H} \frac{\phi^2(s)}{1 - G(s-)} dF(s) < \infty,$$

alors

$$n^{1/2} \int_{-\infty}^{\tau_H} \phi(s) d\left(\widehat{S}_{KM} - S\right)(s) \xrightarrow[n \rightarrow \infty]{L} N(0, \sigma^2),$$

avec

$$\sigma^2 = \int_{-\infty}^{\tau_H} \frac{S(s)}{1 - H(s-)} [\phi(s) - \bar{\phi}(s)]^2 dF(s),$$

$$\bar{\phi}(s) = S^{-1}(s) \int_s^{\tau_H} \phi(u) dF(u).$$

Ce théorème a été prouvé en utilisant des techniques de marginales, et améliore celui établi par Stute [41] par des techniques plus classiques de processus empiriques et de représentations en sommes de termes *i.i.d.*

2.5 Estimation de la fonction de densité de probabilité dans un modèle censuré

Le but de cette section est l'estimation de la densité des durées de survie X dans le cadre des données censurées.

Soient X_1, X_2, \dots, X_n une suite de *v.a* positives *i.i.d.* de fonction de répartition F représentant les durées de survie et C_1, C_2, \dots, C_n est la suite de *v.a i.i.d.* représentant les censures de fonction de répartition G .

On suppose que la suite (X_i) est indépendante de la suite (C_i) et on observe $(Z_i, \delta_i), i = 1, \dots, n$ pour une taille d'échantillon $n \geq 1$, où pour $i = 1, \dots, n$,

$$\begin{cases} Z_i = X_i \wedge C_i \\ \delta_i = \mathbb{I}_{\{X_i \leq C_i\}} \end{cases} \quad (2.2)$$

L'estimateur de Kaplan-Meier $\widehat{S}_{KM}(t)$ de la fonction de survie S est donné par :

$$\widehat{S}_{KM}(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{\delta_{(i)}}{n-i+1}\right) & \text{si } t \leq Z_{(n)} \\ \{Z_{(i)} \leq t\} & \\ 1 - \widehat{F}_n(Z_{(n)}) & \text{si } t > Z_{(n)} \end{cases} \quad (2.3)$$

où $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$ est la statistique d'ordre de l'échantillon Z_1, Z_2, \dots, Z_n et $\delta_{(i)}$ est la valeur de δ_i (l'indicateur de censure à droite) associée à $Z_{(i)}$.

(Pour $t > Z_{(n)}$ on a posé $1 - \widehat{F}_n(t) = 1 - \widehat{F}_n(Z_{(n)})$ puisque il n'y a plus d'observation après t).

On suppose que F admet une densité f par rapport à la mesure de Lebesgue qu'on se propose d'estimer en utilisant les observations $(Z_i, \delta_i), i = 1, \dots, n$.

En se basant sur l'estimateur de Kaplan Meier, Blum et Susarla [5] ont proposé un estimateur de la densité par la méthode du noyau donné par

$$\widehat{f}_n(x) = \frac{1}{h_n} \int_0^\infty K\left(\frac{x-s}{h_n}\right) d\widehat{F}_n(s) \quad (2.4)$$

où \widehat{F}_n est une fonction empirique, $(h_n)_{n \geq 1}$ est la fenêtre avec h_n tend vers 0 lorsque n tend vers l'infini, et K un noyau de support $[-1, 1]$.

Cet estimateur f_n a été étudié en particulier par Blum et Susarla [5] et Mielniczuk [25].

On va étudier les propriétés asymptotiques de f_n en utilisant la théorie des martingales pour les processus continus.

Nous utilisons les notations suivantes :

$$S(t) = 1 - F(t) = \bar{F}(t), \quad C(t) = 1 - G(t) = \bar{G}(t),$$

$$\pi(t) = P(Z_1) = S(t)C(t^-) = \bar{F}(t)\bar{G}(t^-),$$

$\lambda(t) = \frac{f(t)}{S(t)}$, $\Lambda(t) = \int_0^t \lambda(s)ds$ où λ et Λ sont respectivement le taux de hasard et le taux de hasard cumulé.

On considère les processus ponctuels suivants définis sur $[0, \infty[$:

Le processus $N_i(t)$ indicateur de survenue de l'évènement.

$$N_i(t) = \mathbb{I}_{\{Z_i \leq t, \delta_i = 1\}}.$$

Le processus $N_n(t) = \sum_{i=1}^n N_i(t)$ nombre d'évènement observé dans $]0, t]$, ou nombre d'observations non censurées inférieur ou égale à t .

Le processus risque $Y_i(t) = \mathbb{I}_{\{Z_i \geq t\}}$.

Le processus $Y_n(t) = \sum_{i=1}^n Y_i(t)$ nombre d'individus à risque à l'instant t , ou nombre de sujets présents jusqu'à t .

On note aussi les processus suivants :

$A(t) = \int_0^t Y_n(s)d\Lambda(s)$ le processus intensité cumulée.

$$M_n(t) = N_n(t) - A(t) = N_n(t) - \int_0^t Y_n(s)\lambda(s)ds.$$

$$\Lambda_n = \int_0^t \frac{dN_n(u)}{Y_n(u)}.$$

$$\Lambda_n^* = \int_0^t J_n(u)\lambda(u)du \quad \text{où} \quad J_n(t) = 1_{\{Y_n(t) > 0\}}.$$

$$f_n^*(t) = \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) \widehat{S}_{KM}(s^-) d\Lambda_n^* \quad (2.5)$$

$$\tilde{f}_n(t) = \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) f(s) ds$$

Le processus $(M(t), t \geq 0)$ est une F_t martingale où

$$F_t = \sigma(\mathbb{I}_{Z_i \leq s, \delta_i=1}, \mathbb{I}_{Z_i \leq s, \delta_i=0}; 0 \leq s \leq t, i = 1, \dots, n)$$

.

$(M(t), t \geq 0)$ est aussi une martingale locale de carré intégrable.

On suppose que :

(C1) : le noyau K est borné sur $[-1, 1]$ et satisfait les conditions suivantes

$$\int_{-1}^1 K(t) dt = 1, \quad \int_{-1}^1 tK(t) dt = 0, \quad \int_{-1}^1 t^2 K(t) dt = k_2 \neq 0.$$

2.5.1 Etude du biais de l'estimateur \hat{f}_n

Dans le cas général où F est une *f.d.r.* quelconque, la fonction de survie $S(t)$ est l'unique solution de l'équation intégrale suivante :

$$S(t) = 1 - F(t) = 1 - \int_0^t S(s^-) d\Lambda(s)$$

et l'estimateur de Kaplan-Meier $\hat{S}_{KM}(t)$ noté dans la suite par $S_n(t)$, vérifie

$$\hat{S}_{KM}(t) = S_n(t) = 1 - \int_0^t S_n(s^-) d\Lambda_n(s) =: 1 - \hat{F}_n(t)$$

où encore

$$dS_n(t) = -S_n(t^-) d\Lambda_n(s)$$

En effet

$$\begin{aligned} S_n(t) - S_n(t^-) &= \Delta S_n(t) \\ &= -S_n(t^-) [\Lambda_n(t) - \Lambda_n(t^-)] \\ &=: S_n(t^-) \Delta \Lambda_n(t) \\ &= -S_n(t^-) \frac{\Delta N_n(t)}{Y_n(t)} \quad \text{par construction de } \Lambda_n. \end{aligned}$$

D'où

$$S_n(t) = S_n(t^-) \left(1 - \frac{\Delta N_n(t)}{Y_n(t)} \right)$$

où $\Delta N_n(t) = N_n(t) - \Delta N_n(t^-)$.

Comme Λ_n présente des sauts seulement aux points t_i , on obtient :

$$S_n(t) = \prod_{s \leq t} \left(1 - \frac{\Delta N_n(s)}{Y_n(s)} \right) = \prod_{t_i \leq t} (1 - \Delta \Lambda_n(s)). \quad (2.6)$$

Ainsi de l'équation $S_n(t) = 1 - \widehat{F}_n(t)$, l'estimateur à noyau (2.4) peut s'écrire comme suit :

$$\widehat{f}_n(t) = \frac{1}{h_n} \int_0^t K \left(\frac{t-s}{h_n} \right) S_n(s^-) d\Lambda_n(s) \quad (2.7)$$

car $d\widehat{F}_n(t) = -dS_n(t)$ où $S_n(t) = \widehat{S}_{KM}$ par notation.

Nous avons

$$\begin{aligned} \Lambda_n(t) - \Lambda_n^*(t) &= \int_0^t \frac{J_n(u)}{Y_n(u)} dN_n(u) - \int_0^t J_n(u) \lambda(u) du \\ &= \int_0^t \frac{J_n(u)}{Y_n(u)} (dN_n(u) - Y_n(u) \lambda(u) du) \\ &= \int_0^t \frac{J_n(u)}{Y_n(u)} dM_n(u) \end{aligned}$$

où $M_n(u) = \sum_{i=1}^n M_i(u)$, et la martingale $M_i(u) = N_i(u) \int_0^u Y_i(s) \lambda(s) ds$.

Par suite de (2.5) et (2.7), nous avons :

$$\begin{aligned} \widehat{f}_n(t) - f_n^*(t) &= \frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) S_n(s^-) d(\Lambda_n - \Lambda_n^*)(s) \\ &= \frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) S_n(s^-) \frac{J_n(s)}{Y_n(s)} dM_n(s) \\ &= \sum_{i=1}^n \frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) S_n(s^-) \frac{J_n(s)}{Y_n(s)} dM_i(s) \\ &= \sum_{i=1}^n \int_0^\infty L_i(s) dM_i(s). \end{aligned} \quad (2.8)$$

Comme $L_i(s)$ sont des processus localement bornés sur $([0, t])$ et F_s^- -prévisible, et $M_i(s)$ est une martingale locale de carré intégrable, alors par le théorème 2.4.5 de Fleming et Harrington [10] $(\widehat{f}_n(t) - f_n^*(t), t \geq 0)$ est une martingale locale de carré intégrable.

On note :

$$\begin{aligned}\tilde{f}_n(t) &= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) f(s) ds \\ &= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) S(s) d\Lambda(s).\end{aligned}$$

Nous avons le résultat suivant :

Théorème 2.5 (Fleming et Harrington [10]). *Supposons que f est continue*

(i) : *Si $h_n \xrightarrow[n \rightarrow \infty]{} 0$, alors*

$$\tilde{f}_n(t) - f(t) = \int_{-1}^1 K(u) [f(t - h_n u) - f(t)] du \xrightarrow[n \rightarrow \infty]{} 0$$

(ii) : *Pour tout $t \geq 0$, nous avons :*

$$E\hat{f}_n(t) = Ef_n^*(t) = \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-)J_n(s)) d\Lambda(s)$$

(iii) : *Si la fonction $\pi(\cdot) = \mathbb{P}(Z_1 \geq \cdot)$ est strictement positive dans un voisinage d'un point t , alors nous avons :*

$$\begin{aligned}E(\hat{f}_n(t) - \tilde{f}_n(t)) &= E(f_n^*(t) - \tilde{f}_n(t)) \\ &= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E\left(\mathbb{I}_{\{\tau_Y < s\}} \frac{S_n(\tau_Y)S(s)}{S(\tau_Y)}\right) d\Lambda(s)\end{aligned}$$

où $\tau_Y = \inf \{s : Y(s) = 0\}$

(iv) : *Sous les mêmes conditions de (iii), nous avons pour n assez grand*

$$\left|E(\hat{f}_n(t) - \tilde{f}_n(t))\right| \leq \exp^{-n\pi(t+h_n)} \int_{-1}^1 |K(u)| h(t - h_n u) du$$

C'est à dire

$$\left|E(\hat{f}_n(t) - \tilde{f}_n(t))\right| \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{avec une vitesse exponentielle.}$$

(v) : *Si les conditions de (iii) sont vérifiées, f de classe C^2 et $h_n \rightarrow 0$, alors*

$$E\left(\hat{f}_n(t) - f(t)\right) = \frac{1}{2} h_n^2 f''(t) k_2 + o(h_n^2) + o\left(\frac{1}{n}\right)$$

où k_2 est donné par C1.

Preuve. (i) : Nous avons

$$\begin{aligned}\tilde{f}_n(t) - f(t) &= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) f(s) ds - f(t) \\ &= \int_{-1}^1 K(u) f(t - h_n u) du - \int_{-1}^1 K(u) f(t) du \\ &= \int_{-1}^1 K(u) [f(t - h_n u) - f(t)] du\end{aligned}$$

Par continuité de f , nous avons

$$|f(t - h_n u) - f(t)| \xrightarrow[n \rightarrow \infty]{} 0$$

et comme K est borné, nous avons par le théorème de convergence dominée

$$\begin{aligned}|\tilde{f}_n(t) - f(t)| &\leq \int_{-1}^1 |K(u)| |f(t - h_n u) - f(t)| du \\ &\leq \sup_{u \in [-1, 1]} |K(u)| \int_{-1}^1 |f(t - h_n u) - f(t)| du \xrightarrow[n \rightarrow \infty]{} 0\end{aligned}$$

et par suite $\tilde{f}_n(t) \xrightarrow[n \rightarrow \infty]{} f(t)$.

Nous avons

$$\tilde{f}_n(t) = (K_{h_n} * f)(t)$$

Comme f est continue et $h_n \rightarrow 0$, alors :

$$(K_{h_n} * f)(t) \xrightarrow[n \rightarrow \infty]{} f(t)$$

donc

$$\tilde{f}_n(t) - f(t) \xrightarrow[n \rightarrow \infty]{} 0.$$

(ii) : Nous avons par (2.8)

$$\hat{f}_n(t) - f_n^*(t) = \sum_{i=1}^n \int_0^\infty L_i(s) dM_i(s)$$

et en notant que

$$\langle M_n \rangle_t = \int_0^t Y_n(s) \lambda(s) ds,$$

nous avons

$$\begin{aligned}E \left[\hat{f}_n(t) - f_n^*(t) \right]^2 &= \frac{1}{h_n^2} E \left[\int_0^t K^2\left(\frac{t-s}{h_n}\right) S_n^2(s^-) \frac{J_n(s)}{Y_n^2(s)} d \langle M_n \rangle_s \right] \\ &= \frac{1}{h_n} \int_0^{t/h_n} K^2(u) E \left[S_n^2[(t - h_n u)^-] \frac{J_n(t - h_n u)}{Y_n(t - h_n u)} \right] \lambda(t - h_n u) du \\ &\leq \frac{1}{h_n} \int_0^1 K^2(u) \lambda(t - h_n u) du < \infty.\end{aligned}$$

où nous avons utilisé que :

$$E \left[S_n^2 [(t - h_n u)^-] \frac{J_n(t - h_n u)}{Y_n(t - h_n u)} \right] \leq 1$$

car $S_n \leq 1$, $J_n \leq 1$, $Y_n \leq 1$ par le théorème (2.4.5) de Fleming et Harrington [10], pp. 73.

Nous avons pour n assez grand

$$E \left(\widehat{f}_n(t) - f_n^*(t) \right) = 0$$

ce qui donne $E(\widehat{f}_n(t)) = E(f_n^*(t))$ et nous avons

$$\begin{aligned} E(f_n^*(t)) &= E \left(\frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) S_n d\Lambda_n^*(s) \right) \\ &= E \left(\frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) S_n(s^-) J_n(s) \lambda(s) ds \right) \\ &= E \left(\frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) S_n(s^-) J_n(s) d\Lambda(s) \right) \\ &= \frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) E(S_n(s^-) J_n(s)) d\Lambda(s) \end{aligned}$$

(iii) et (iv) : En utilisant (ii) et le lemme (3.2.1) de Fleming et Harrington [10], pp. 99, on obtient du fait que

$$J_n(s) = 1 - \mathbb{I}_{\{Y_n(s)=0\}}$$

les égalités suivantes :

$$E \left[\widehat{f}_n(t) - \widetilde{f}_n(t) \right] = E \left[f_n^*(t) - \widetilde{f}_n(t) \right]$$

$$\begin{aligned}
&= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-) J_n(s)) d\Lambda(s) \\
&\quad - \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) S(s) d\Lambda(s) \\
&= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-)) d\Lambda(s) \\
&\quad - \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-) \mathbb{I}_{\{Y_n(s)=0\}}) d\Lambda(s) \\
&\quad - \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) S(s) d\Lambda(s) \\
&= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-) - S(s^-)) d\Lambda(s) && \text{où } (S(s) = S(s^-)) \\
&\quad - \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-) \mathbb{I}_{\{Y_n(s)=0\}}) && \text{car } S \text{ est continue} \\
&= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E\left(\mathbb{I}_{\{\tau_Y < s^-\}} \frac{S_n(\tau_Y) [S(\tau_Y) - S(s^-)]}{S(\tau_Y)}\right) d\Lambda(s) \\
&\quad - \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(S_n(s^-) \mathbb{I}_{\{\tau_Y < s\}}) d\Lambda(s)
\end{aligned}$$

D'où :

$$\begin{aligned}
&E\left[\widehat{f}_n(t) - \widetilde{f}_n(t)\right] \\
&= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E\left(\mathbb{I}_{\{\tau_Y < s\}} \frac{S_n(\tau_Y) [S(\tau_Y) - S(s^-)]}{S(\tau_Y)} - S_n(s^-) \mathbb{I}_{\{\tau_Y < s\}}\right) d\Lambda(s) \\
&= \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E(\mathbb{I}_{\{\tau_Y < s\}} (S_n(\tau_Y) - S_n(s^-))) d\Lambda(s) \\
&\quad - \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E\left(\mathbb{I}_{\{\tau_Y < s\}} \frac{S_n(\tau_Y) S(s)}{S(\tau_Y)}\right) d\Lambda(s) \\
&= -\frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) E\left(\mathbb{I}_{\{\tau_Y < s\}} \frac{S_n(\tau_Y) S(s)}{S(\tau_Y)}\right) d\Lambda(s)
\end{aligned}$$

$$\mathbb{I}_{\{\tau_Y < s\}} (S_n(\tau_Y) - S_n(s^-)) = 0 \text{ car } S_n(\tau_Y) = S_n(S^-) \forall \tau_Y < s.$$

Par suite pour n assez grand, nous avons

$$\begin{aligned}
\left|E\left[\widehat{f}_n(t) - \widetilde{f}_n(t)\right]\right| &\leq \frac{1}{h_n} \int_0^\infty \left|K\left(\frac{t-s}{h_n}\right)\right| E\left(\mathbb{I}_{\{\tau_Y < s\}} \frac{S_n(\tau_Y) S(s)}{S(\tau_Y)}\right) d\Lambda(s) \\
&\leq \frac{1}{h_n} \int_0^\infty \left|K\left(\frac{t-s}{h_n}\right)\right| E(\mathbb{I}_{\{\tau_Y < s\}}) d\Lambda(s) \quad \text{où } \left(\frac{S(s)}{S(\tau_Y)} \leq 1, S_n \leq 1\right)
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{h_n} \int_0^\infty \left| K\left(\frac{t-s}{h_n}\right) \right| E(\mathbb{I}_{\{Y(s)=0\}}) \lambda(s) ds \\
&= \int_{-1}^1 |K(u)| [1 - \pi(t - h_n u)]^n \lambda(t - h_n u) du \\
&\leq [1 - \pi(t + h_n u)]^n \int_{-1}^1 |K(u)| \lambda(t - h_n u) du \quad (\text{car } \pi \searrow) \\
&\leq e^{-n\pi(t+h_n)} \int_{-1}^1 |K(u)| \lambda(t - h_n u) du \quad (\text{car } 1 - x \leq e^{-x})
\end{aligned}$$

car nous avons

$$E(\mathbb{I}_{\{Y_n(s)=0\}}) = \mathbb{P}(Y_n(s) = 0) = [(1 - (1 - F(s))(1 - G(s^-)))]^n = [1 - \pi(s)]^n.$$

(v) : Nous savons que $E\widehat{f}_n(t) = (K_{h_n} * f)(t)$ et aussi $\widetilde{f}_n(t) = (K_{h_n} * f)(t)$ où $K_{h_n}(\cdot) = \frac{1}{h_n} K\left(\frac{\cdot}{h_n}\right)$.

Nous avons par (i), $\widetilde{f}_n(t) \xrightarrow[n \rightarrow \infty]{} f(t)$ et par (iv) nous avons

$$n \left| E \left[\widehat{f}_n(t) - \widetilde{f}_n(t) \right] \right| \leq C n e^{-n\pi(t+h_n)} \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{où } C > 0$$

Nous avons

$$\begin{aligned}
E\widehat{f}_n(t) - f(t) &= E\left(\widehat{f}_n(t) - \widetilde{f}_n(t)\right) + \left(\widetilde{f}_n(t) - f(t)\right) \\
&= o\left(\frac{1}{h_n}\right) + \int_{-1}^1 K(u) [f(t - h_n u) - f(t)] du
\end{aligned}$$

Un développement de Taylor

$$f(t + h_n u) - f(t) = -h_n u f'(t) + \frac{h_n^2}{2} u^2 f''(t) + o(h_n^2)$$

et alors par C1

$$\begin{aligned}
\int_{-1}^1 K(u) [f(t - h_n u) - f(t)] du &= -h_n f'(t) \int_{-1}^1 u K(u) du \\
&\quad + \frac{h_n^2}{2} f''(t) \int_{-1}^1 u^2 K(u) du + o(h_n^2) \\
&= \frac{h_n^2}{2} f''(t) k_2 + o(h_n^2)
\end{aligned}$$

et par suite

$$E\widehat{f}_n(t) - f(t) = \frac{1}{2} h_n^2 f''(t) k_2 + o(h_n^2) + o\left(\frac{1}{n}\right),$$

ce qui achève la démonstration.

Remarque 2.2. le point (v) montre que l'estimateur $\widehat{f}_n(t)$ est asymptotiquement sans biais.

2.5.2 consistance de $\widehat{f}_n(t)$

Le théorème suivant donne la convergence en probabilité de l'estimateur $\widehat{f}_n(t)$.

Théorème 2.6 (Fleming et Harrington [10]). *Si f est continue en t et π est positive au voisinage de t , $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ alors*

$$\widehat{f}_n(t) \xrightarrow[n \rightarrow \infty]{} f(t)$$

Preuve. Puisque $\pi > 0$ sur un voisinage V_t d'un point $t > 0$, f est continue et bornée, S continue et $S > 0$ sur V_t , alors $\lambda = \frac{f}{S}$ est aussi continue et bornée dans V_t .

Nous avons

$$\left| \widehat{f}_n(t) - f(t) \right| \leq \left| \widehat{f}_n(t) - f^*(t) \right| + \left| f_n^*(t) - \widetilde{f}(t) \right| + \left| \widetilde{f}_n(t) - f(t) \right|$$

Par le théorème précédent (i) nous avons

$$\left| \widetilde{f}_n(t) - f(t) \right| \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{quand } h_n \rightarrow \infty$$

Reste à montrer que

$$\left| \widehat{f}_n(t) - f^*(t) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad \text{et} \quad \left| f_n^*(t) - \widetilde{f}(t) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (2.9)$$

Par (2.8) et le corollaire 3.4.1 de Fleming et Harrington [10] et l'inégalité de Lengart, nous avons pour $\epsilon, \eta > 0$ et n assez grand

$$\begin{aligned}
& \mathbb{P} \left(\left| \widehat{f}_n(t) - f_n^*(t) \right| > \epsilon \right) \\
&= \mathbb{P} \left(\left| \frac{1}{h_n} \int_0^\infty \left| K \left(\frac{t-u}{h_n} \right) S_n(u^-) \frac{J_n(u)}{Y_n(u)} dM_n(u) \right| > \epsilon \right) \\
&\leq \mathbb{P} \left(\sup_{t-h_n \leq s \leq t+h_n} \left| \frac{1}{h_n} \int_0^\infty \left| K \left(\frac{t-u}{h_n} \right) S_n(u^-) \frac{J_n(u)}{Y_n(u)} dM_n(u) \right| > \epsilon \right) \\
&\leq \frac{\eta}{\epsilon^2} + \mathbb{P} \left(\frac{1}{h_n^2} \int_{t-h_n}^{t+h_n} K^2 \left(\frac{t-u}{h_n} \right) S_n^2(u^-) \frac{J_n(u)}{Y_n^2(u)} d\langle M_n \rangle_u \geq \eta \right) \\
&= \frac{\eta}{\epsilon^2} + \mathbb{P} \left(\frac{1}{h_n} \int_{-1}^1 K^2(v) S_n^2[(t-h_nv)^-] \frac{J_n(t-h_nv)}{Y_n(t-h_nv)} \lambda(t-h_nv) dv \geq \eta \right) \\
&\leq \frac{\eta}{\epsilon^2} + \mathbb{P} \left(\frac{1}{nh_n n^{-1} Y_n(t+h_n)} \sup_{s \in [t-h_n, t+h_n]} \lambda(s) \int_{-1}^1 K^2(v) dv \geq \eta \right) \\
&\leq \frac{\eta}{\epsilon^2} + \mathbb{P} \left(\frac{1}{n^{-1} Y_n(t+h_n)} \geq C \cdot \eta n h_n \right)
\end{aligned}$$

Comme $n^{-1}Y(t+h_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (1-F(t))(1-G(t^-))$ et $nh_n \rightarrow \infty$, la probabilité ci-dessous tend vers 0. or ϵ et η sont arbitraires et donc nous avons

$$\left| \widehat{f}_n(t) - f_n^*(t) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$$

D'autre part, nous avons

$$\begin{aligned}
\left| f_n^*(t) - \widetilde{f}_n(t) \right| &= \left| \frac{1}{h_n} \int_0^\infty K \left(\frac{t-s}{h_n} \right) [S_n(s^-)J(s) - S(s)] d\Lambda(s) \right| \\
&\leq \int_{-1}^1 |K(u)| \mathbb{I}_{\{Y_n(t-uh_n) > 0\}} |S_n((t-uh_n)^-) - S(t-uh_n)| \lambda(t-uh_n) du \\
&\quad + \int_{-1}^1 |K(u)| \mathbb{I}_{\{Y_n(t-uh_n) = 0\}} f(t-uh_n) du \\
&\leq \left[\sup_{s \in [t-h_n, t+h_n]} \lambda(s) \sup_{s \in [t-h_n, t+h_n]} |S_n(s) - S(s)| + \mathbb{I}_{\{Y_n(t-uh_n) = 0\}} \sup_{s \in [t-h_n, t+h_n]} f(s) \right] \int_{-1}^1 |K(u)| du \\
&\hspace{20em} (2.10)
\end{aligned}$$

Comme $\mathbb{I}_{\{Y_n(t-uh_n) = 0\}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$ et par le théorème 3.4.2 de Fleming et Harrington [10], nous avons

$$\sup_{s \in [t-h_n, t+h_n]} |S_n(s) - S(s)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$$

On en déduit que

$$\left| f_n^*(t) - \widetilde{f}_n(t) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$$

D'où le résultat du théorème.

¶

Chapitre 3

Vitesse de Convergence de l'Estimateur des Dérivées de la Densité dans un Modèle Complet

3.1 Introduction

X_1, \dots, X_n sont des variables aléatoires continues indépendantes, de fonction de densité de probabilité f commune et de *f.d.r* F commune. Soit F_n la *f.d.r* empirique basée sur X_1, \dots, X_n ; i.e., $nF_n(x)$ est le nombre des X_i tel que $X_i \leq x$ où $1 \leq i \leq n$.

Les fonctions de la forme

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K\left(\frac{x - X_i}{h_n}\right)$$

où K est une fonction de densité de probabilité et h_n est une suite des nombres positifs convergeant vers 0, ont été suggérées comme des estimations pour f par Rosenblatt [29] et les diverses propriétés de ces estimations comprenant la convergence uniforme ont été étudiées par Parzen [28].

Définissons l'estimateur à noyau de la r -ième dérivée $f^{(r)}$ de f par

$$f_n^{(r)}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n^{r+1}} K^{(r)}\left(\frac{x - X_i}{h_n}\right) \quad (3.1)$$

où $K^{(r)}$ est la r -ième dérivée du noyau K .

Quand $r = 0$, $f_n^{(0)} = f_n$ représente l'estimateur à noyau de la fonction de densité de probabilité de f , et évidemment $K^{(0)} = K$.

Dans ce chapitre nous examinerons la convergence des variables aléatoires

$$\xi_n^{(r)} = \sup_{|x| \leq a_n} |f_n^{(r)}(x) - f^{(r)}(x)|$$

pour $r = 0, 1, 2, \dots$ où pour toute fonction f , $f^{(0)} = f$, et $\lim_{n \rightarrow \infty} a_n = \infty$.

Nous notons ici que, pour $r = 0$, l'assertion selon laquelle $\xi_n^{(0)}$ converge vers 0 en probabilité est plus faible que la consistance uniforme de f_n prouvée par Parzen [28], et les hypothèses que nous allons faire au sujet de f sont plus fortes que celles établies par Parzen.

3.2 Convergence uniforme des $f_n^{(r)}$

Supposons que f et ses premières $(r + 1)$ dérivées sont bornées.

Nous donnons d'abord les hypothèses nécessaires pour établir les résultats :

Soit K un noyau sur \mathbb{R} , c'est-à-dire une fonction numérique mesurable vérifiant les conditions suivantes :

(K1) : $K > 0$ et $\int K(u)du = 1$.

(K2) : K est symétrique autour de 0 tel que $\int_{-\infty}^{+\infty} |u| K(u)du < \infty$.

(K3) : $K^{(s)}$ est une fonction continue à variations bornées sur \mathbb{R} , pour $s = 0, 1, \dots, r + 1$.

(K4) : $K^{(s)}(x)$, $s = 0, 1, \dots, r - 1$ tend vers 0 quand x tend vers l'infini.

Et soit h_n une suite de nombres réels positifs tel que $h_n =: h$ et vérifiant :

(H1) : h_n décroît vers zéro lorsque n tend vers l'infini

(H2) : nh_n^{2r+2} tend vers l'infini lorsque n tend vers l'infini.

Lemme 3.1. (Bhattacharaya [3]) : Sous les hypothèses (K1)-(K4) et (H1)-(H2) :

Si ψ une fonction bornée sur \mathbb{R} ainsi que ses $(r + 1)$ dérivées bornées, alors il existe une constante C qui ne dépend pas de h telle que :

$$\sup_x \left| \frac{1}{h^{r+1}} \int_{-\infty}^{+\infty} K^{(r)}\left(\frac{x-y}{h}\right) \psi(y)dy - \psi^{(r)}(x) \right| \leq Ch.$$

Preuve. Comme ψ est bornée, l'intégrale qui nous intéresse existe.

On prouve le lemme d'abord pour $r = 0$:

Sous l'hypothèse (K_2) , nous avons pour x arbitraire :

$$\begin{aligned} \left| \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{x-y}{h}\right) \psi(y) dy - \psi(x) \right| &= \left| \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right) \psi(y) dy - \psi(x) \right| \\ &= \left| \int_{-\infty}^{+\infty} K(u) \{ \psi(x+hu) - \psi(x) \} du \right| \\ &\leq \int_{-\infty}^{+\infty} K(u) |\psi(x+hu) - \psi(x)| \\ &\leq \sup_x |\psi'(x)| \cdot h \cdot \int_{-\infty}^{+\infty} |u| K(u) du, \end{aligned}$$

et le lemme est vérifié pour $r = 0$ avec :

$$C = \sup_x |\psi'(x)| \int_{-\infty}^{+\infty} |u| K(u) du.$$

Pour $r \geq 1$ il est facile de vérifier que :

$$\frac{1}{h^{r+1}} K^{(r)}\left(\frac{x-y}{h}\right) \psi(y) = \frac{1}{h} K\left(\frac{x-y}{h}\right) \psi^{(r)}(y) - \frac{d}{dy} \sum_{s=0}^{r-1} \frac{1}{h^{s+1}} K^{(s)}\left(\frac{x-y}{h}\right) \psi^{(r-s-1)}(y),$$

d'où

$$\begin{aligned} \frac{1}{h^{r+1}} \int_{-\infty}^{+\infty} K^{(r)}\left(\frac{x-y}{h}\right) \psi(y) dy &= \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{x-y}{h}\right) \psi^{(r)}(y) dy \\ &\quad - \lim_{y \rightarrow +\infty} \sum_{s=0}^{r-1} \frac{1}{h^{s+1}} K^{(s)}\left(\frac{x-y}{h}\right) \psi^{(r-s-1)}(y) \\ &\quad + \lim_{y \rightarrow -\infty} \sum_{s=0}^{r-1} \frac{1}{h^{s+1}} K^{(s)}\left(\frac{x-y}{h}\right) \psi^{(r-s-1)}(y). \end{aligned}$$

Comme ψ et toutes ses dérivées impliquées dans l'expression ci-dessus sont bornées et comme $K^{(s)}\left(\frac{x-y}{h}\right)$ tend vers 0, $s = 0, 1, \dots, r-1$ quand $|x| \rightarrow +\infty$,

$$\frac{1}{h^{r+1}} \int_{-\infty}^{+\infty} K^{(r)}\left(\frac{x-y}{h}\right) \psi(y) dy = \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{x-y}{h}\right) \psi^{(r)}(y) dy.$$

La même démarche que pour $r = 0$, conduit au résultat et complète la preuve pour $r \geq 1$.

Lemme 3.2. (Shuster [31]) :

Il existe une constante universelle C telle que pour tout $n > 0$, $\varepsilon_n > 0$ et une f.d.r F

$$\mathbb{P}_F \left\{ \sup_x |F_n(x) - F(x)| > \varepsilon_n \right\} \leq C \exp \left\{ -2n \varepsilon_n^2 \right\}.$$

Preuve. Pour le cas où F est continue (voir Dvoretzky, Kiefer and Wolfowitz [8]). Si F est discontinue en quelques points, alors il existe une f.d.r continue \bar{F} pour laquelle

$$\mathbb{P}_F \left\{ \sup_x |F_n(x) - F(x)| > \varepsilon_n \right\} \leq \mathbb{P}_F \left\{ \sup_x |F_n(x) - \bar{F}(x)| > \varepsilon_n \right\}$$

(voir Kiefer and Wolfowitz [18]). Ainsi, le lemme est vrai pour toute F univariée.

Notons la variation bornée de $K^{(s)}$ par μ_s .

L'hypothèse de continuité sur $K^{(s)}$ a été faite uniquement pour s'assurer que $\sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)|$ est une v.a.

Lemme 3.3. (Shuster [31]) : Sous les hypothèses $(K1), (K3), (K4)$ et $(H1)-(H2)$:

Si f et ses premières $(r + 1)$ dérivées sont bornées, alors il existe une constante C telle que pour tout $n > 0$, $\varepsilon_n > 0$ et une f.d.r F

$$\mathbb{P}_F \left\{ \sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)| > \varepsilon_n \right\} \leq C \exp \left\{ -2n \varepsilon_n^2 h_n^{2r+2} / \mu_r^2 \right\}.$$

Preuve. Sous l'hypothèse (K_3) , nous avons $K^{(r)}$ est borné et que $\lim_{x \rightarrow +\infty} K^{(r)}(x)$ et $\lim_{x \rightarrow -\infty} K^{(r)}(x)$ existent tous les deux.

Pour $r = 0$, sous l'hypothèse (K_1) , $\lim_{x \rightarrow +\infty} K^{(r)}(x)$ et $\lim_{x \rightarrow -\infty} K^{(r)}(x)$ doivent être nulles.

Si $r \geq 1$, alors la fonction $K^{(r-1)}$ a une dérivé bornée sur $[-a, a]$ pour tout a , et (voir Natansan [27], p133) donc K est Lebesgue-intégrable sur $[-a, a]$. Ainsi, (voir Natansan [27], p259)

$$V_{-a}^a [K^{(r-1)}] = \int_{-a}^a |K^{(r)}(u)| du.$$

Maintenant

$$V_{+\infty}^{-\infty} [K^{(r-1)}] = \lim_{a \rightarrow \infty} V_{-a}^a [K^{(r-1)}] = \lim_{a \rightarrow \infty} \int_{-a}^a |K^{(r)}(u)| du =: \int_{-\infty}^{+\infty} |K^{(r)}(u)| du$$

pour que $\int |K^{(r)}(u)| du$ soit fini.

Ce fait et l'existence de $\lim_{x \rightarrow +\infty} K^{(r)}(x)$ et $\lim_{x \rightarrow -\infty} K^{(r)}(x)$ impliquent que ces limites doivent être nulles.

Par une intégration par parties et en rappelant que $\lim_{x \rightarrow +\infty} K^{(r)}(x) = 0$ et $\lim_{x \rightarrow -\infty} K^{(r)}(x) = 0$, on trouve que

$$\begin{aligned} \sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)| &= \sup_x h_n^{-(r+1)} \left| \int K^{(r)} \left(\frac{x-y}{h_n} \right) dF_n(y) - \int K^{(r)} \left(\frac{x-y}{h_n} \right) dF(y) \right| \\ &= h_n^{-(r+1)} \sup_x \left| \left[\{F_n(y) - F(y)\} K^{(r)} \left(\frac{x-y}{h_n} \right) \right]_{-\infty}^{+\infty} - \int \{F_n(y) - F(y)\} dK^{(r)} \left(\frac{x-y}{h_n} \right) \right| \\ &= h_n^{-(r+1)} \sup_x \left| \int \{F_n(y) - F(y)\} dK^{(r)} \left(\frac{x-y}{h_n} \right) \right| \\ &\leq h_n^{-(r+1)} \sup_x |F_n(x) - F(x)| \mu_r. \end{aligned}$$

Par conséquent, en appliquant le lemme (3.2), nous avons

$$\begin{aligned} \mathbb{P}_F \left\{ \sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)| > \varepsilon_n \right\} &\leq \mathbb{P}_F \left\{ \sup_x |F_n(x) - F(x)| > \varepsilon_n h_n^{r+1} / \mu_r \right\} \\ &\leq C \exp \left\{ -2n \varepsilon_n^2 h_n^{2r+2} / \mu_r^2 \right\}. \end{aligned}$$

et la preuve est finie.

Lemme 3.4. (Shuster [31]) : Sous les hypothèses (K1), (K3)-(K4) et (H1)-(H2) :

Soit X une v.a absolument continue de fonction de densité de probabilité f et h un certain nombre réel positif. Si f et ses premières $(r+1)$ dérivées sont bornées, alors il existe une constante C , ne dépend pas de h , telle que :

$$\sup_x \left| E \left[\frac{1}{h^{r+1}} K^{(r)} \left(\frac{x-X}{h} \right) \right] - f^{(r)}(x) \right| \leq Ch.$$

Lemme 3.5. (Bhattacharaya [3]) : Sous les hypothèses (K1)-(K2), (K4) et (H1)-(H2) :

Si f et ses premières $(r+1)$ dérivées sont bornées, et si ε_n est une suite de nombres positifs telle que $h_n = o(\varepsilon_n)$ quand $n \rightarrow \infty$, alors ils existent des constantes positives C_1 et C_2 telles que :

$$\mathbb{P} \left[\sup_{|x| \leq a_n} |f_n^{(r)}(x) - f^{(r)}(x)| \geq \varepsilon_n \right] \leq \frac{C_1 a_n}{h_n^{r+2} \varepsilon_n} \exp \left\{ -n h_n^{2r+2} \varepsilon_n^2 / C_2 \right\}$$

pour n suffisamment grand.

Preuve. Soit $\delta_n = h_n^{r+2} \varepsilon_n / 4M_1$ est une borne supérieure pour $\{f^{(r+1)}(x)\}$ ainsi que $\{K^{(r+1)}(x)\}$. Aussi,

$$|f^{(r)}(x_1) - f^{(r)}(x_2)| \leq \varepsilon_n/4 \quad \text{et} \quad |f_n^{(r)}(x_1) - f_n^{(r)}(x_2)| \leq \varepsilon_n/4$$

quand $|x_1 - x_2| \leq \delta_n$. Divisons l'intervalle $[-a_n, a_n]$ en intervalles consécutifs de longueur δ_n chacun. Il y aura

$$b_n = \frac{2a_n}{\delta_n} = \frac{8M_1 a_n}{h_n^{r+2} \varepsilon_n}$$

intervalles. Notons ces intervalles par J_{n1}, \dots, J_{nb_n} et soit x_{nj} un point arbitraire de J_{nj} . Comme $h_n = o(n)$ quand $n \rightarrow \infty$, il résulte du lemme 3.4 au lemme 3.1 et à partir des propriétés des intervalles J_{nj} que pour n assez grand,

$$\begin{aligned} \sup_{|x| \leq a_n} |f_n^{(r)}(x) - f^{(r)}(x)| &= \max_{j=1, \dots, b_n} \sup_{x \in J_{nj}} | \{f_n^{(r)}(x) - f_n^{(r)}(x_{nj})\} + \{f_n^{(r)}(x_{nj}) - E[f_n^{(r)}(x_{nj})]\} \\ &\quad + \{E[f_n^{(r)}(x_{nj})] - f^{(r)}(x_{nj})\} - \{f^{(r)}(x) - f^{(r)}(x_{nj})\} | \\ &\leq \frac{\varepsilon_n}{2} + Ch_n + \max_{j=1, \dots, b_n} |f_n^{(r)}(x_{nj}) - E[f_n^{(r)}(x_{nj})]| \\ &\leq \frac{3\varepsilon_n}{4} + \max_{j=1, \dots, b_n} |f_n^{(r)}(x_{nj}) - E[f_n^{(r)}(x_{nj})]|. \end{aligned}$$

Ainsi, pour n grand

$$\begin{aligned} \mathbb{P} \left[\sup_{|x| \leq a_n} |f_n^{(r)}(x) - f^{(r)}(x)| \geq \varepsilon_n \right] &\leq \mathbb{P} \left[\max_{j=1, \dots, b_n} |f_n^{(r)}(x_{nj}) - E[f_n^{(r)}(x_{nj})]| \geq \varepsilon_n/4 \right] \\ &\leq \sum_{j=1}^{b_n} \mathbb{P} \left[\left| \sum_{i=1}^n K^{(r)} \left(\frac{x_{nj} - X_i}{h_n} \right) - E \left[\sum_{i=1}^n K^{(r)} \left(\frac{x_{nj} - X_i}{h_n} \right) \right] \right| \geq n h_n^{r+1} \varepsilon_n/4 \right]. \end{aligned}$$

Soit M_2 une borne supérieure pour $K^{(r)}(x)$. Il en résulte alors du théorème de Hoeffding [16], concernant les probabilités de queue des sommes de v.a indépendantes et bornées, que pour x arbitraire

$$\begin{aligned} P \left[\left| \sum_{i=1}^n K^{(r)} \left(\frac{x - X_i}{h_n} \right) - E \left[\sum_{i=1}^n K^{(r)} \left(\frac{x - X_i}{h_n} \right) \right] \right| \geq n h_n^{r+1} \varepsilon_n/4 \right] \\ \leq 2 \exp \left\{ -n h_n^{2r+2} \varepsilon_n^2 / 32M_2^2 \right\}, \end{aligned}$$

ce qui achève la preuve du lemme avec $C_1 = 16M_1$ et $C_2 = 32M_2^2$.

Lemme 3.6. (Shuster [31]) : Sous les hypothèses (K1),(K3)-(K4) et (H1)-(H2) :

Si f et ses premières $(r + 1)$ dérivées sont bornées, et si ε_n est une suite de nombres positifs telle que $h_n = o(\varepsilon_n)$, alors ils existent des constantes positives C_1 et C_2 telles que :

$$\mathbb{P} \left[\sup_x |f_n^{(r)}(x) - f^{(r)}(x)| \geq \varepsilon_n \right] \leq C_1 \exp \left\{ -C_2 n \varepsilon_n^2 h_n^{2r+2} \right\}$$

pour n suffisamment grand.

Preuve. Nous avons à l'aide du lemme (3.4)

$$\begin{aligned} \sup_x |f_n^{(r)}(x) - f^{(r)}(x)| &\leq \sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)| + \sup_x |E f_n^{(r)}(x) - f^{(r)}(x)| \\ &\leq \sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)| + C h_n. \end{aligned}$$

Comme $h_n = o(\varepsilon_n)$ donc pour n suffisamment grand

$$\mathbb{P} \left\{ \sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \varepsilon_n \right\} \leq \mathbb{P} \left\{ \sup_x |f_n^{(r)}(x) - E f_n^{(r)}(x)| > \varepsilon_n/2 \right\}$$

Une application du lemme (3.3) donne le résultat souhaité.

Le théorème qui suit indique que, pour certaines suites h_n , $\sup_x |f_n^{(r)}(x) - f^{(r)}(x)|$ converge vers 0 avec probabilité 1.

Une suite b_n avec $\lim_{n \rightarrow \infty} b_n = \infty$ est introduite pour indiquer la vitesse à laquelle la convergence ci-dessus a lieu.

Théorème 3.1. (Shuster [31]) : Si f et ses premières $(r + 1)$ dérivées sont bornées, et si les suites $\{h_n\}$ et $\{b_n\}$ sont telles que $h_n b_n = o(1)$ et $\sum_{n=1}^{\infty} \exp \{-c n h_n^{2r+2}/b_n^2\}$ est fini pour tout c positif, alors

$$\lim_{n \rightarrow \infty} \sup_x b_n |f_n^{(r)}(x) - f^{(r)}(x)| = 0$$

avec probabilité 1.

Preuve. $\forall \varepsilon > 0$, on obtient par le lemme 3.6

$$\mathbb{P}_F \left\{ \sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \varepsilon/b_n \right\} \leq C_1 \exp \left(-C_2 \varepsilon^2 n h_n^{2r+2}/b_n^2 \right)$$

pour n suffisamment grand. Comme $\sum_{n=1}^{\infty} \exp \{-c n h_n^{2r+2}/b_n^2\}$ est fini pour tout c positif, alors

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \varepsilon/b_n \right\}$$

est fini pour tout ε positif. Par conséquent, et à l'aide du lemme de Borel–Cantelli, on a $\lim_{n \rightarrow \infty} \sup_x b_n \left| f_n^{(r)}(x) - f^{(r)}(x) \right| = 0$ avec probabilité 1.

Le fait que $f^{(r+1)}$ soit bornée, est utilisé pour assurer que $\sup_x \left| E f_n^{(r)}(x) - f^{(r)}(x) \right| = O(h_n)$. Afin d'établir $\lim_{n \rightarrow \infty} \sup_x \left| f_n^{(r)}(x) - f^{(r)}(x) \right| = 0$ avec probabilité 1, nous avons besoin que $\sup_x \left| E f_n^{(r)}(x) - f^{(r)}(x) \right| = o(1)$ qui est le cas, par exemple, si $f^{(r)}$ était uniformément continue.

Le corollaire suivant indique la vitesse de convergence pour un choix particulier de h_n .

Corollaire 3.1. (Shuster [31] :)

Si f et ses premières $(r+1)$ dérivées sont bornées, $h_n = n^{-1/(2r+4)}$ et $0 < c < \frac{1}{2r+4}$, alors

$$\lim_{n \rightarrow \infty} \sup_x n^c \left| f_n^{(r)}(x) - f^{(r)}(x) \right| = 0$$

avec probabilité 1.

3.3 Expressions du biais et de la variance pour l'estimateur à noyau des dérivées de la densité :

Notons que $MISE = IMSE$.

$$\begin{aligned} MISE(\widehat{f}^{(r)}, f^{(r)}) &= E \left[\int \left(\widehat{f}^{(r)}(x) - f^{(r)}(x) \right)^2 dx \right] \\ &= \int E \left(\widehat{f}^{(r)}(x) - f^{(r)}(x) \right)^2 dx \\ &= IMSE(\widehat{f}^{(r)}, f^{(r)}) \end{aligned} \tag{3.2}$$

L'erreur quadratique moyenne (MSE) peut être décomposée comme suit :

$$\begin{aligned} MSE(\widehat{f}^{(r)}, f^{(r)}, x) &= E \left(\widehat{f}^{(r)}(x) - f^{(r)}(x) \right)^2 \\ &= Var(\widehat{f}^{(r)}(x)) + \left(E(\widehat{f}^{(r)}(x)) - f^{(r)}(x) \right)^2 \end{aligned} \tag{3.3}$$

Si le noyau K est r fois différentiable, alors l'estimateur de la r ième dérivée de la densité $\widehat{f}^{(r)}(x)$ peut être écrite comme suit :

$$\widehat{f}^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{i=1}^n K^{(r)} \left(\frac{x - X_i}{h} \right)$$

$$= \frac{1}{n} \sum_{i=1}^n K_h^{(r)}(x - X_i) \quad (3.4)$$

où $K_h^{(r)}(x) = (1/h^{r+1})K^{(r)}(x/h)$.

Afin de tirer une approximation du *MISE*, il est nécessaire d'établir les conditions suivantes sur f , h et K .

1. La $(r + 2)$ ième dérivée $f^{(r+2)}$ est continue, de carré intégrable et monotone sur $(-\infty, -M)$ et (M, ∞) pour $M > 0$;
2. $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} nh^{2r+1} = \infty$;
3. Pour que $\hat{f}(x)$ soit une densité, on suppose que $K(u) \geq 0$ et $\int K(u)du = 1$. La fonction noyau est supposée être symétrique autour de zéro, *i.e.*, $\int uK(u)du = 0$, et possède un moment d'ordre 2 fini, *i.e.*, $\int u^2K(u)du < \infty$.

3.3.1 Biais

La moyenne de l'estimateur peut être écrite comme suit :

$$\begin{aligned} E(\hat{f}^{(r)}(x)) &= \frac{1}{n} \sum_{i=1}^n E\left(K_h^{(r)}(x - X_i)\right) \\ &= E\left(K_h^{(r)}(x - X)\right) \\ &= \int K_h^{(r)}(x - y)f(y)dy. \end{aligned} \quad (3.5)$$

En utilisant le produit de convolution $*$ nous avons

$$E(\hat{f}^{(r)}(x)) = (K_h^{(r)} * f)(x) = (K_h * f^{(r)})(x) \quad (3.6)$$

où nous avons utilisé la relation $K_h^{(r)} * f = K_h * f^{(r)}$.

Tirons maintenant une approximation de la moyenne. En utilisant un changement de variable, la moyenne peut être écrite comme suit :

$$\begin{aligned} E(\hat{f}^{(r)}(x)) &= (K_h * f^{(r)})(x) = \int K_h(x - y)f^{(r)}(y)dy \\ &= \int K(u)f^{(r)}(x - hu)du. \end{aligned} \quad (3.7)$$

En utilisant le développement de Taylor, $f^{(r)}(x - hu)$ est développé comme

$$f^{(r)}(x - hu) = f^{(r)}(x) - huf^{(r+1)}(x) + \frac{1}{2}h^2u^2f^{(r+2)}(x) + o(h^2). \quad (3.8)$$

D'où

$$\begin{aligned}
 E(\widehat{f}^{(r)}(x)) &= \widehat{f}^{(r)}(x) \left[\int K(u) du \right] - h f^{(r+1)}(x) \left[\int u K(u) du \right] \\
 &\quad + \frac{1}{2} h^2 f^{(r+2)}(x) \left[\int u^2 K(u) du \right] + o(h^2).
 \end{aligned} \tag{3.9}$$

De l'hypothèse (3) nous avons,

$$\begin{aligned}
 \int K(u) du &= 1 \\
 \int u K(u) du &= 0 \\
 \mu_2(K) &= \int u^2 K(u) du < \infty
 \end{aligned} \tag{3.10}$$

D'où le biais s'écrit comme

$$E(\widehat{f}^{(r)}(x)) - f^{(r)}(x) = \frac{1}{2} h^2 \mu_2(K) f^{(r+2)}(x) + o(h^2). \tag{3.11}$$

L'estimation est asymptotiquement sans biais. Le biais estimant la r ième dérivée est proportionnel à la valeur de la $(r + 2)$ ième dérivée de la fonction densité.

3.3.2 Variance

La variance de l'estimateur est donnée par

$$\begin{aligned}
 Var(\widehat{f}^{(r)}(x)) &= \frac{1}{n} Var \left(K_h^{(r)}(x - X) \right) \\
 &= \frac{1}{n} \left[E \left(K_h^{(r)}(x - X)^2 \right) - E \left(K_h^{(r)}(x - X) \right)^2 \right].
 \end{aligned} \tag{3.12}$$

En utilisant l'équation (3.5), nous obtenons l'expression de la variance suivante

$$Var(\widehat{f}^{(r)}(x)) = \frac{1}{n} \left[\left(K_h^{(r)} * p \right) (x)^2 - \left(K_h^{(r)} * p \right)^2 (x) \right]. \tag{3.13}$$

En utilisant un changement de variable nous avons

$$\begin{aligned}
 Var(\widehat{f}^{(r)}(x)) &= \frac{1}{n} \left[\int K_h^{(r)}(x - y)^2 f(y) dy \right] - \frac{1}{n} \left[\int K_h^{(r)}(x - y) f(y) dy \right]^2 \\
 &= \frac{1}{nh^{2r+1}} \left[\int K^{(r)}(u)^2 f(x - hu) du \right] - \frac{1}{nh^{2r}} \left[\int K^{(r)}(u) f(x - hu) du \right]^2.
 \end{aligned} \tag{3.14}$$

En utilisant le développement de Taylor, $f(x - hu)$ est développée comme suit :

$$f(x - hu) = f(x) + o(1). \quad (3.15)$$

Nous avons besoin uniquement du premier terme à cause du facteur $1/n$. D'où

$$\begin{aligned} \text{Var}(\widehat{f}^{(r)}(x)) &= \frac{1}{nh^{2r+1}} [f(x) + o(1)] \int K^{(r)}(u)^2 du - \frac{1}{nh^{2r}} [f(x) + o(1)]^2 \left[\int K^{(r)}(u) \right]^2 du \\ &= \frac{1}{nh^{2r+1}} f(x) \int K^{(r)}(u)^2 du + o(1/nh^{2r+1}). \end{aligned} \quad (3.16)$$

En se basant sur l'hypothèse (2), $\lim_{n \rightarrow \infty} nh^{2r+1} = \infty$, la variable converge asymptotiquement vers zéro.

En utilisant les équations (3.11) et (3.16), la MSE en un point x s'écrit comme

$$MSE(\widehat{f}^{(r)}, f^{(r)}, x) = \frac{1}{nh^{2r+1}} f(x) R(K^{(r)}) + \frac{1}{4} h^4 \mu_2(K)^2 f^{(r+2)}(x)^2 + o(h^4 + 1/nh^{2r+1}) \quad (3.17)$$

où $R(K^{(r)}) = \int K^{(r)}(u)^2 du$.

MSE optimal de la fenêtre de lissage pour estimer $\widehat{f}^{(r)}(x)$ est d'ordre $n^{-1/(2r+5)}$.

3.3.3 MISE

Comme $MISE = IMSE$ nous avons,

$$\begin{aligned} MISE(\widehat{f}^{(r)}, f^{(r)}) &= \frac{1}{nh^{2r+1}} R(K^{(r)}) \int f(x) dx + \frac{1}{4} h^4 \mu_2(K)^2 \int f^{(r+2)}(x)^2 dx + o(h^4 + 1/nh^{2r+1}). \\ &= AMISE(\widehat{f}^{(r)}, f^{(r)}) + o(h^4 + 1/nh^{2r+1}) \end{aligned} \quad (3.18)$$

où

$$AMISE(\widehat{f}^{(r)}, f^{(r)}) = \frac{1}{nh^{2r+1}} R(K^{(r)}) + \frac{1}{4} h^4 \mu_2(K)^2 R(f^{(r+2)}). \quad (3.19)$$

Le paramètre de lissage optimal $h_{AMISE} := h_r^*$ pour l'estimation de la r-ème dérivée de la densité est donné par

$$h_r^* = \left[\frac{(2r+1)R(K^{(r)})}{\mu_2(K)^2 R(f^{(r+2)})} \right]^{1/(2r+5)} n^{-1/(2r+5)}$$

D'où

$$AMISE^*(\widehat{f}^{(r)}) = \frac{2r+5}{4} R(K^{(r)})^{\frac{4}{2r+5}} [\mu_2(K)^2 R(f^{(r+2)}) / (2r+1)]^{\frac{2r+1}{2r+5}} n^{-\frac{4}{2r+5}}$$

3.4 Expressions asymptotique du biais et de la variance pour l'estimateur à noyau des dérivées de la densité :

Nous voulons estimer la fonction Φ_r .

$$\Phi_r = \int f^{(r)}(x)f(x)dx = E [f^{(r)}(X)]. \quad (3.20)$$

Un estimateur de Φ_r est

$$\widehat{\Phi}_r = \frac{1}{n} \sum_{i=1}^n \widehat{f}^{(r)}(X_i). \quad (3.21)$$

Où $\widehat{f}^{(r)}(X_i)$ est l'estimateur de la r ième dérivée de la densité $f(x)$ quand $x = X_i$.

En utilisant l'estimation à noyau de la dérivée de la densité $\widehat{f}^{(r)}(X_i)$ (équation (3.1)) nous obtenons

$$\begin{aligned} \widehat{\Phi}_r &= \frac{1}{n^2 h^{r+1}} \sum_{i=1}^n \sum_{j=1}^n K^{(r)}\left(\frac{X_i - X_j}{h}\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_h^{(r)}(X_i - X_j). \end{aligned} \quad (3.22)$$

L'erreur quadratique moyenne (MSE) peut être décomposée comme suit :

$$MSE(\widehat{\Phi}_r, \Phi_r) = E[\widehat{\Phi}_r - \Phi_r]^2 = Var[\widehat{\Phi}_r] + (E[\widehat{\Phi}_r] - \Phi_r)^2. \quad (3.23)$$

Afin de tirer une approximation du MSE , il est nécessaire d'établir les conditions suivantes sur f , h et K .

1. La fonction de densité f possède $k > 2$ dérivée continue et monotone sur $(-\infty, -M)$ et (M, ∞) pour $M > 0$. La $(r+2)$ ième dérivée $f^{(r+2)}$ est continue, de carré intégrable et monotone sur $(-\infty, -M)$ et (M, ∞) pour $M > 0$;
2. $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} nh^{2r+1} = \infty$;
3. Pour que $\widehat{f}(x)$ soit une densité, on suppose que $K(u) \geq 0$ et $\int K(u)du = 1$. La fonction noyau est supposée être symétrique autour de zéro, *i.e.*, $\int uK(u)du = 0$, et possède un moment d'ordre 2 fini, *i.e.*, $\int u^2K(u)du < \infty$.

Nous pouvons écrire l'estimateur à noyau comme suit :

$$\widehat{\Phi}_r = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_h^{(r)}(X_i - X_j)$$

$$= \frac{1}{n} K_h^{(r)}(0) + \frac{1}{n^2} \sum_{i \neq j} \sum K_h^{(r)}(X_i - X_j). \quad (3.24)$$

Le premier terme est une constante indépendante des données.

3.4.1 Biais

L'espérance de l'estimateur est donnée par

$$\widehat{\Phi}_r = \frac{1}{n} K_h^{(r)}(0) + \left(1 - \frac{1}{n}\right) E \left[K_h^{(r)}(X_1 - X_2) \right]. \quad (3.25)$$

Le terme $E \left[K_h^{(r)}(X_1 - X_2) \right]$ peut être simplifié comme suit :

$$E \left[K_h^{(r)}(X_1 - X_2) \right] = \int \int K_h^{(r)}(x - y) f(x) f(y) dx dy$$

En utilisant la relation $K_h^{(r)} * f = K_h * f^{(r)}$ nous obtenons

$$E \left[K_h^{(r)}(X_1 - X_2) \right] = \int \int K_h(x - y) f(x) f^{(r)}(y) dx dy \quad (3.26)$$

Par un changement de variable, nous obtenons

$$E \left[K_h^{(r)}(X_1 - X_2) \right] = \int \int K(u) f(y + hu) f^{(r)}(y) du dy \quad (3.27)$$

En utilisant le développement de Taylor, $f(y + hu)$ peut être développée comme suit :

$$f(y + hu) = f(y) + huf'(y) + \frac{1}{2}h^2u^2f''(y) + O(h^3). \quad (3.28)$$

D'où

$$\begin{aligned} E \left[K_h^{(r)}(X_1 - X_2) \right] &= \left(\int K(u) du \right) \left(\int f^{(r)}(y) f(y) dy \right) \\ &\quad + h \left(\int uK(u) du \right) \left(\int f^{(r)}(y) f'(y) dy \right) \\ &\quad + \frac{1}{2}h^2 \left(\int u^2K(u) du \right) \left(\int f^{(r)}(y) f''(y) dy \right) + O(h^3). \end{aligned}$$

De l'hypothèse (3) nous avons,

$$\int K(u) du = 1$$

$$\int uK(u)du = 0$$

$$\mu_2(K) = \int u^2K(u)du < \infty \quad (3.29)$$

En remplaçant, nous avons

$$E \left[K_h^{(r)}(X_1 - X_2) \right] = \Phi_r + \frac{1}{2}h^2\mu_2(K) \left(\int f^{(r)}(y)f''(y)dy \right) + O(h^3). \quad (3.30)$$

Ceci peut être simplifié, le fait que les dérivées de la densité soient monotones sur $(-\infty, -M)$ et (M, ∞) pour $M > 0$ et en utilisant une intégration par partie .

$$E \left[K_h^{(r)}(X_1 - X_2) \right] = \Phi_r + \frac{1}{2}h^2\mu_2(K)\Phi_{r+2} + O(h^3).$$

D'où en remplaçant dans l'équation (3.25), le biais de l'estimateur peut être écrit comme suit :

$$E[\widehat{\Phi}_r] - \Phi_r = \frac{1}{nh^{r+1}}K^{(r)}(0) + \frac{1}{2}h^2\mu_2(K)\Phi_{r+2} + O(h^3) - \frac{1}{n}\Phi_r - \frac{1}{2n}h^2\mu_2(K)\Phi_{r+2}. \quad (3.31)$$

Le biais après avoir ignoré les $1/n$ termes est donné par

$$E[\widehat{\Phi}_r] - \Phi_r = \frac{1}{nh^{r+1}}K^{(r)}(0) + \frac{1}{2}h^2\mu_2(K)\Phi_{r+2} + O(h^3).$$

3.4.2 Variance

Si r est pair, alors la variance peut être écrite comme suit :

$$Var[\widehat{\Phi}_r] = \frac{2(n-1)}{n^3}Var[K_h^{(r)}(X_1 - X_2)]$$

$$+ \frac{4(n-1)(n-2)}{n^3}Cov[K_h^{(r)}(X_1 - X_2), K_h^{(r)}(X_2 - X_3)]. \quad (3.32)$$

D'abord nous calculons

$$E[K_h^{(r)}(X_1 - X_2)^2] = \int \int K_h^{(r)}(x - y)^2 f(x)f(y)dx dy$$

En utilisant un changement de variable, nous obtenons

$$E[K_h^{(r)}(X_1 - X_2)^2] = \frac{1}{h^{2r+1}} \int \int K^{(r)}(u)^2 f(y + hu)f(y)dudy$$

En utilisant un développement de Taylor, $f(y + hu)$ peut être développée comme suit :

$$f(y + hu) = f(y) + o(1). \quad (3.33)$$

D'où

$$E[K_h^{(r)}(X_1 - X_2)^2] = \frac{1}{h^{2r+1}} \Phi_0 R(K^{(r)}) + o(1/h^{2r+1}).$$

Aussi, nous avons

$$E[K_h^{(r)}(X_1 - X_2)] = \Phi_r + o(1). \quad (3.34)$$

D'après les deux dernières équations ci-dessus, la variance peut être écrite comme

$$\begin{aligned} Var[K_h^{(r)}(X_1 - X_2)] &= E[K_h^{(r)}(X_1 - X_2)^2] - E[K_h^{(r)}(X_1 - X_2)]^2 \\ &= \frac{1}{h^{2r+1}} \Phi_0 R(K^{(r)}) - \Phi_r^2 + o(1/h^{2r+1}). \end{aligned}$$

La covariance est alors donnée par

$$\begin{aligned} Cov[K_h^{(r)}(X_1 - X_2), K_h^{(r)}(X_2 - X_3)] &= E[K_h^{(r)}(X_1 - X_2)K_h^{(r)}(X_2 - X_3)] \\ &\quad - E[K_h^{(r)}(X_1 - X_2)]E[K_h^{(r)}(X_2 - X_3)] \end{aligned}$$

Le premier terme peut être simplifié comme suit :

$$\begin{aligned} &E[K_h^{(r)}(X_1 - X_2)K_h^{(r)}(X_2 - X_3)] \\ &= \int \int \int K_h^{(r)}(x - y)K_h^{(r)}(y - z)f(x)f(y)f(z) dx dy dz \\ &= \int \int \int K_h(x - y)K_h(y - z)f^{(r)}(x)f(y)f^{(r)}(z) dx dy dz \\ &= \int \int \int K(u)K(v)f^{(r)}(y + hu)f(y)f^{(r)}(y - hv) du dv dy \\ &= \int f^{(r)}(y)^2 f(y) dy + o(1). \end{aligned}$$

D'où

$$Cov[K_h^{(r)}(X_1 - X_2), K_h^{(r)}(X_2 - X_3)] = \int f^{(r)}(y)^2 f(y) dy - \Phi_r^2 + o(1).$$

En utilisant ces approximations, la variance peut être écrite comme suit :

$$Var[\widehat{\Phi}_r] = \frac{2}{n^2 h^{2r+1}} \Phi_0 R(K^{(r)}) + \frac{4}{n} \left[\int f^{(r)}(y)^2 f(y) dy - \Phi_r^2 \right] + o(1/n^2 h^{2r+1} + 1/n). \quad (3.35)$$

3.4.3 MSE

L'erreur quadratique moyenne asymptotique est donnée par

$$\begin{aligned}
 MSE(\widehat{\Phi}_r, \Phi_r) &= E[\widehat{\Phi}_r - \Phi_r]^2 \\
 &= Var[\widehat{\Phi}_r] + \left(E[\widehat{\Phi}_r] - \Phi_r\right)^2 \\
 &= \left[\frac{1}{nh^{r+1}}K^{(r)}(0) + \frac{1}{2}h^2\mu_2(K)\Phi_{r+2}\right]^2 + \frac{2}{n^2h^{2r+1}}\Phi_0R(K^{(r)}) \\
 &\quad + \frac{4}{n} \left[\int f^{(r)}(y)^2 f(y) dy - \Phi_r^2\right] + O(h^6) + o(1/n^2h^{2r+1} + 1/n). \quad (3.36)
 \end{aligned}$$

Chapitre 4

Estimateur à noyau des dérivées de la densité pour le modèle censuré

En se basant sur la méthode du noyau, on définit l'estimateur des dérivées de la fonction de densité lorsque la variable d'intérêt est censurée aléatoirement à droite, puis, on établit la vitesse de convergence uniforme presque sûre de cet estimateur, ainsi que sa vitesse de convergence.

4.1 Introduction :

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Soit X_1, \dots, X_n une suite de variables aléatoires d'intérêt *i.i.d.*, de *f.d.r* commune F et de densité f . Soit C_1, \dots, C_n une suite de variables aléatoires de censure *i.i.d.*, de *f.d.r* commune G , admettant une densité g . Les C_i sont supposées indépendantes des X_i .

L'échantillon observé est alors (Z_i, δ_i) , $i = 1, \dots, n$ où

$$\begin{cases} Z_i = X_i \wedge C_i \\ \delta_i = \mathbb{I}_{\{X_i \leq C_i\}} \end{cases} \quad (4.1)$$

$\mathbb{I}_{(E)}$ désigne la fonction indicatrice de E .

L'indicatrice $\delta_i = 1$ indique l'absence de censure ($Z_i = X_i$), et $\delta_i = 0$ sinon.

Pour $t \in \mathbb{R}$, soient $F(t) := \mathbb{P}(X \leq t)$, $G(t) := \mathbb{P}(C \leq t)$ et $H(t) := \mathbb{P}(Z \leq t)$ les fonctions de répartition de X , C et Z respectivement.

Pour toute *f.d.r.* L , nous noterons par $\tau_L = \sup\{t : L(t) < 1\}$ la borne supérieure du support de L . Il a été montré dans Stute et Wang [40], la relation $\tau_H = \tau_F \wedge \tau_G$.

Nous supposons que $\tau_F \leq \tau_G$. Cette hypothèse est nécessaire pour des raisons de consistance de l'estimateur de Kaplan-Meier sur toute la ligne réelle. En effet, quand $\tau_F > \tau_G$ toutes les valeurs entre τ_F et τ_G sont des censures et ne sont donc pas observées.

Nous considérons un compact Θ tel que $\Theta \subset (-\infty, \tau_F]$

Les fonctions de la forme

$$f_n(x) = \frac{1}{n h_n} \sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Z_i)} K\left(\frac{x - Z_i}{h_n}\right),$$

où $h_n =: h$ et G_n est l'estimateur empirique de G , ont été suggérées comme des estimateurs pour f dans le cas censuré.

Si le noyau K admet une dérivée d'ordre r , on définit l'estimateur à noyau de la r -ième dérivée $f^{(r)}$ de f par

$$f_n^{(r)}(x) = \frac{1}{n h_n^{r+1}} \sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Z_i)} K^{(r)}\left(\frac{x - Z_i}{h_n}\right). \quad (4.2)$$

Quand $r = 0$, $f_n^{(0)} = f_n$ représente l'estimateur à noyau de la fonction de densité de probabilité de f , et évidemment $K^{(0)} = K$.

Dans ce chapitre, nous étudions la convergence des variables aléatoires

$$\xi_n^{(r)} = \sup_{x \in \Theta} |f_n^{(r)}(x) - f^{(r)}(x)|, \text{ pour } r \geq 1.$$

Pour ce faire, nous introduisons le pseudo-estimateur $\tilde{f}_n^{(r)}$ défini par :

$$\tilde{f}_n^{(r)}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n^{r+1}} \frac{\delta_i}{\bar{G}(Z_i)} K^{(r)}\left(\frac{x - Z_i}{h_n}\right) \quad (4.3)$$

Nous remarquons que ce pseudo-estimateur n'est pas utile en pratique, car G est inconnue. Cependant, il servira à établir nos résultats.

4.2 Convergence uniforme de $f_n^{(r)}$:

Nous donnons d'abord les hypothèses nécessaires pour établir nos résultats :

Supposons que f et ses $(r + 2)$ premières dérivées sont bornées.

Pour K , on peut prendre n'importe quelle fonction de densité de probabilité symétrique autour de 0 qui satisfait les conditions suivantes :

(K1) : $K > 0$, $\int K(u)du = 1$, $\int u |K(u)| du < \infty$ et $\int u^2 K(u)du < \infty$.

(K2) : $K^{(s)}$ est Lipchitzienne pour $s = 0, 1, \dots, r$. et à support compact.

La fenêtre h_n satisfait :

(H1) : $h_n \rightarrow 0$ lorsque $n \rightarrow \infty$.

(H2) : $n h_n^{2r+1} / \log n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

(H3) : $n h_n^{2r+2} \rightarrow \infty$ lorsque $n \rightarrow \infty$.

(H4) : $h_n \geq n^{-1/2r+1}$.

Théorème 4.1. *Sous les hypothèses (K1) – (K2) et (H1) – (H4), nous avons pour $r \geq 1$*

$$\sup_{x \in \Theta} |f_n^{(r)}(x) - f^{(r)}(x)| = O \left\{ \max \left(\sqrt{\frac{\log n}{n h_n^{2r+1}}}, h_n^2 \right) \right\} \text{ p.s quand } n \rightarrow \infty$$

Preuve. *En utilisant l'inégalité triangulaire, nous obtenons*

$$\begin{aligned} \sup_{x \in \Theta} |f_n^{(r)}(x) - f^{(r)}(x)| &\leq \sup_{x \in \Theta} |f_n^{(r)}(x) - \tilde{f}_n^{(r)}(x)| + \sup_{x \in \Theta} |\tilde{f}_n^{(r)}(x) - E \tilde{f}_n^{(r)}(x)| \\ &\quad + \sup_{x \in \Theta} |E \tilde{f}_n^{(r)}(x) - f^{(r)}(x)| \end{aligned}$$

Pour la preuve du théorème (4.1), nous avons besoin d'établir sous forme de lemmes, les résultats intermédiaires suivants :

Lemme 4.1. *Sous les hypothèses (K1) – (K2) et (H1) – (H2), nous avons*

$$\sup_{x \in \Theta} |f_n^{(r)}(x) - \tilde{f}_n^{(r)}(x)| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ p.s quand } n \rightarrow \infty$$

Preuve.

$$\begin{aligned}
 \left| f_n^{(r)}(x) - \tilde{f}_n^{(r)}(x) \right| &= \frac{1}{nh_n^{r+1}} \left| \sum_{i=1}^n \frac{\delta_i}{\overline{G}_n(Z_i)} K^{(r)} \left(\frac{x - Z_i}{h_n} \right) - \frac{\delta_i}{\overline{G}(Z_i)} K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right| \\
 &= \frac{1}{nh_n^{r+1}} \left| \sum_{i=1}^n \frac{\mathbb{I}_{\{X_i \leq C_i\}}}{\overline{G}_n(Z_i)} K^{(r)} \left(\frac{x - Z_i}{h_n} \right) - \frac{\mathbb{I}_{\{X_i \leq C_i\}}}{\overline{G}(Z_i)} K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right| \\
 &\leq \frac{1}{nh_n^{r+1}} \sum_{i=1}^n \left| K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right| \cdot \left| \frac{1}{\overline{G}_n(Z_i)} - \frac{1}{\overline{G}(Z_i)} \right| \\
 &\leq \frac{1}{nh_n^{r+1}} \sum_{i=1}^n \left| K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right| \cdot \left| \frac{\overline{G}_n(Z_i) - \overline{G}(Z_i)}{\overline{G}_n(Z_i)\overline{G}(Z_i)} \right| \\
 &\leq \frac{1}{\overline{G}_n(\tau_F)\overline{G}(\tau_F)} \cdot \sup_{t \in \Theta} |\overline{G}_n(t) - \overline{G}(t)| \cdot \frac{\tau_F}{nh_n^{r+1}} \sum_{i=1}^n \left| K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right|
 \end{aligned}$$

Comme $\overline{G}(\tau) > 0$ et en utilisant la loi du logarithme itéré (LLI) (voir Deheuvels et Einmahl [6]), nous obtenons

$$\sup_{x \in \Theta} |\overline{G}_n(t) - \overline{G}(t)| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \quad p.s$$

Par conséquent, la loi forte des grands nombres (LFGN) permet d'avoir

$$\sup_{x \in \Theta} \left| f_n^{(r)}(x) - \tilde{f}_n^{(r)}(x) \right| \leq \frac{\tau_F}{\overline{G}^2(\tau_F)} \cdot E \left(\left| \frac{1}{h_n^{r+1}} K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right| \right) \sqrt{\frac{\log \log n}{n}}$$

Les hypothèses sur K et h_n achèvent la preuve.

.

Lemme 4.2. Sous les hypothèses $(K_1) - (K_2)$, (H_1) , nous avons :

$$\sup_{x \in \Theta} \left| E \tilde{f}_n^{(r)}(x) - f^{(r)}(x) \right| = O(h_n^2) \quad p.s \text{ quand } n \rightarrow \infty$$

Preuve. De (4.3) et en utilisant

$$\mathbb{I}_{\{X_1 \leq C_1\}} \varphi(Z_1) = \mathbb{I}_{\{X_1 \leq C_1\}} \varphi(X_1) \tag{4.4}$$

pour toute fonction mesurable φ . En utilisant un changement de variables, puis r inté-

grations par parties, nous obtenons

$$\begin{aligned}
 \left| E\tilde{f}_n^{(r)}(x) - f^{(r)}(x) \right| &= \left| E \left\{ \frac{1}{nh_n^{r+1}} \sum_{i=1}^n \frac{\delta_i}{\overline{G}(Z_i)} K^{(r)} \left(\frac{x - Z_i}{h_n} \right) \right\} - f^{(r)}(x) \right| \\
 &= \left| E \left\{ \frac{1}{h_n^{r+1}} \frac{\delta_1}{\overline{G}(Z_1)} K^{(r)} \left(\frac{x - Z_1}{h_n} \right) \right\} - f^{(r)}(x) \right| \\
 &= \left| \frac{1}{h_n^{r+1}} E \left\{ \frac{\mathbb{I}_{\{X_1 \leq C_1\}}}{\overline{G}(Z_1)} K^{(r)} \left(\frac{x - Z_1}{h_n} \right) \right\} - f^{(r)}(x) \right| \\
 &= \left| \frac{1}{h_n^{r+1}} E \left\{ K^{(r)} \left(\frac{x - Z_1}{h_n} \right) \overline{G}^{-1}(Z_1) E [\mathbb{I}_{\{X_1 \leq C_1\}} / X_1] \right\} - f^{(r)}(x) \right| \\
 &= \left| \frac{1}{h_n^r} \int K^{(r)}(s) f(t - sh_n) ds - f^{(r)}(t) \right| \\
 &= \left| \int K(s) f^{(r)}(t - sh_n) ds - f^{(r)}(t) \right| \\
 &= \left| \int K(s) [f^{(r)}(t - sh_n) - f^{(r)}(t)] ds \right|.
 \end{aligned}$$

En utilisant un développement de Taylor, nous avons

$$f^{(r)}(t - sh_n) - f^{(r)}(t) = -h_n s f^{(r+1)}(t) + \frac{h_n^2}{2} s^2 f^{(r+2)}(t_0)$$

où t_0 est compris entre t et $t - sh_n$. Alors

$$\begin{aligned}
 \sup_{x \in \Theta} \left| E\tilde{f}_n^{(r)}(x) - f^{(r)}(x) \right| &= \sup_{x \in \Theta} \left| \int K(s) \left[-h_n s f^{(r+1)}(t) + \frac{h_n^2}{2} s^2 f^{(r+2)}(t_0) \right] ds \right| \\
 &\leq h_n \sup_{x \in \Theta} \left| \int s K(s) f^{(r+1)}(t) ds \right| + h_n^2 \sup_{x \in \Theta} \left| \int \frac{s^2}{2} K(s) f^{(r+2)}(t_0) ds \right| \\
 &=: I + II
 \end{aligned}$$

Comme K est symétrique autour de zéro et sous les hypothèses $(K_1) - (K_2)$, nous avons $I = o(1)$. Aussi, comme f et ses $(r + 2)$ premières dérivées sont bornées, nous avons $\sup_{x \in \Theta} |f^{(r+2)}(t)| < \infty$, et sous les hypothèses $(H_1) - (H_2)$, nous avons $II = O(h_n^2)$, ce qui donne le résultat.

Lemme 4.3. sous les hypothèses $(K_1) - (K_2)$ et $(H_1) - (H_4)$, nous avons :

$$\sup_{x \in \Theta} \left| \tilde{f}_n^{(r)}(x) - E\tilde{f}_n^{(r)}(x) \right| = O \left(\sqrt{\frac{\log n}{nh_n^{2r+1}}} \right) \quad p.s \text{ quand } n \rightarrow \infty$$

Preuve. Comme Θ est un compact, il peut être recouvert par un nombre fini q_n d'intervalles de longueur $w_n = \sqrt{\frac{h_n^{2r+1}}{n}}$, centrés en x_1, \dots, x_{q_n} pour $r \geq 1$. Comme Θ est borné, il existe une constante A telle que $q_n \leq A/w_n$.

Pour tout $x \in \Theta$, nous posons

$$\Delta_i(x) = \frac{1}{nh_n^{r+1}} \frac{\delta_i}{\overline{G}(Z_i)} K^{(r)}\left(\frac{x - Z_i}{h_n}\right) - E\left(\frac{1}{nh_n^{r+1}} \frac{\delta_1}{\overline{G}(Z_1)} K^{(r)}\left(\frac{x - Z_1}{h_n}\right)\right).$$

Il est évident que

$$\sum_{i=1}^n \Delta_i(x) = \tilde{f}_n^{(r)}(x) - E\left(\tilde{f}_n^{(r)}(x)\right),$$

donc

$$\begin{aligned} \sum_{i=1}^n \Delta_i(x) &= \left\{ \left[\tilde{f}_n^{(r)}(x) - \tilde{f}_n^{(r)}(x_k) \right] - \left[E\tilde{f}_n^{(r)}(x) - E\tilde{f}_n^{(r)}(x_k) \right] \right\} + \left\{ \tilde{f}_n^{(r)}(x_k) - E\tilde{f}_n^{(r)}(x_k) \right\} \\ &=: \sum_{i=1}^n \tilde{\Delta}_i(x) + \sum_{i=1}^n \Delta_i(x_k) \end{aligned}$$

Alors

$$\sup_{x \in \Theta} \left| \sum_{i=1}^n \Delta_i(x) \right| \leq \sup_{x \in \Theta} \max_{1 \leq k \leq q_n} \underbrace{\left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right|}_{\Sigma_1} + \max_{1 \leq k \leq q_n} \underbrace{\left| \sum_{i=1}^n \Delta_i(x_k) \right|}_{\Sigma_2}$$

Nous observons que

$$\begin{aligned} \Sigma_1 &\leq \frac{1}{nh_n^{r+1}} \sum_{i=1}^n \delta_i \overline{G}^{-1}(Z_i) \left| K^{(r)}\left(\frac{x - Z_i}{h_n}\right) - K^{(r)}\left(\frac{x_k - Z_i}{h_n}\right) \right| \\ &\quad + \frac{1}{h_n^{r+1}} E \left[\delta_1 \overline{G}^{-1}(Z_1) \left| K^{(r)}\left(\frac{x - Z_1}{h_n}\right) - K^{(r)}\left(\frac{x_k - Z_1}{h_n}\right) \right| \right] \\ &=: \Sigma_{1n} + \Sigma_{2n} \end{aligned}$$

Comme $K^{(r)}$ est lipchitzien, alors

$$\begin{aligned} \Sigma_{1n} &\leq \left(\frac{C}{\overline{G}(\tau_F)} \frac{1}{h_n^{r+1}} \left| \frac{x - x_k}{h_n} \right| \right) \\ &\leq \frac{Cw_n}{\overline{G}(\tau_F)h_n^{r+2}} = O\left(\frac{1}{\sqrt{nh_n^3}}\right) = o(1). \end{aligned}$$

Les mêmes arguments ci-dessus conduit à la même borne pour Σ_{2n} .

Par conséquent

$$\Sigma_1 = o(1) \quad p.s.$$

Nous étudions maintenant $\Sigma_2 := |\sum_{i=1}^n \Delta_i(x_k)|$:

Pour tout $\epsilon > 0$ nous avons :

$$\mathbb{P} \left\{ \max_{1 \leq k \leq q_n} \left| \sum_{i=1}^n \tilde{\Delta}_i(x_k) \right| > \epsilon \right\} \leq \sum_{i=1}^{q_n} \mathbb{P} \left\{ \left| \sum_{i=1}^n \tilde{\Delta}_i(x_k) \right| \right\}$$

d'où

$$\mathbb{P} \left\{ \max_{1 \leq k \leq q_n} \left| \tilde{f}_n^{(r)}(x_k) - E\tilde{f}_n^{(r)}(x_k) \right| > \epsilon \right\} \leq q_n \left\{ \mathbb{P} \left| \tilde{f}_n^{(r)}(x_k) - E\tilde{f}_n^{(r)}(x_k) \right| > \epsilon \right\} \quad (4.5)$$

Nous posons $U_i := nh_n^{r+1} \Delta_i(x_k)$. Nous avons $E(U_i) = 0$ et

$$|U_i| \leq \frac{2\bar{K}}{\bar{G}(\tau_F)} =: M$$

où \bar{K} est une borne supérieure pour $K^{(r)}(x)$.

Afin d'appliquer l'inégalité de Bernstein, nous calculons $\sigma_1^2 = \text{Var}(U_1)$ car les U_i sont i.i.d.

$$\begin{aligned} \sigma_1^2 &= E \left[\left(\frac{\delta_1}{\bar{G}(Z_1)} K^{(r)} \left(\frac{x - Z_1}{h_n} \right) \right)^2 \right] - E^2 \left[\frac{\delta_1}{\bar{G}(Z_1)} K^{(r)} \left(\frac{x - Z_1}{h_n} \right) \right] \\ &= \int \frac{1}{\bar{G}(v)^2} K^{2(r)} \left(\frac{t - v}{h_n} \right) f(v) dv - \left(\int \frac{1}{\bar{G}(v)} K^{(r)} \left(\frac{t - v}{h_n} \right) f(v) dv \right)^2. \end{aligned}$$

Par un changement de variables, nous avons

$$S_1^2 = h_n \int (K^{(r)}(u))^2 \frac{f(t - uh_n)}{\bar{G}(t - uh_n)} du - h_n^2 \left(\int K^{(r)}(u) f(t - uh_n) du \right)^2.$$

Sous les hypothèses établies sur le noyau K , nous obtenons $\sigma_1^2 = O(h_n)$.

En appliquant l'inégalité de Bernstein, nous avons

$$\begin{aligned}
 \mathbb{P} \left\{ \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\} &= P \left\{ \left| \sum_{i=1}^n U_i \right| > \epsilon n h_n^{r+1} \right\} \\
 &\leq 2 \exp \left\{ \frac{-\epsilon^2 n^2 h_n^{2r+2}}{2 \left[V + \frac{1}{3} M \epsilon n h_n^{r+1} \right]} \right\} \quad \text{avec } V \geq n \sigma_1^2, \\
 &= 2 \exp \left\{ \frac{-\epsilon^2 n^2 h_n^{2r+2}}{2 n h_n \left[c + \frac{1}{3} M \epsilon h_n^r \right]} \right\} \\
 &\approx 2 \exp \left\{ -c \epsilon^2 n h_n^{2r+1} \right\}
 \end{aligned}$$

où c est une constante universelle.

Par conséquent (4.5) devient

$$\begin{aligned}
 \mathbb{P} \left\{ \max_{1 \leq k \leq q_n} \left| \tilde{f}_n^{(r)}(x_k) - E \tilde{f}_n^{(r)}(x_k) \right| > \epsilon \right\} &\leq 2 q_n \exp \left\{ \frac{-c \epsilon^2 n h_n^{2r+1}}{\log n} \log n \right\} \\
 &= 2 q_n n^{-c \epsilon^2 n h_n^{2r+1} / \log n}.
 \end{aligned}$$

En remplaçant q_n par sa valeur, nous obtenons

$$\mathbb{P} \left\{ \max_{1 \leq k \leq q_n} \left| \tilde{f}_n^{(r)}(x_k) - E \tilde{f}_n^{(r)}(x_k) \right| > \epsilon \right\} \leq 2A h^{-r-\frac{1}{2}} n^{\frac{1}{2}-c \epsilon^2 n h_n^{2r+1} / \log n}$$

Sous l'hypothèse (H_4) , nous avons

$$\mathbb{P} \left\{ \max_{1 \leq k \leq q_n} \left| \tilde{f}_n^{(r)}(x_k) - E \tilde{f}_n^{(r)}(x_k) \right| > \epsilon \right\} \leq 2A n^{\frac{1}{2}-c \epsilon^2 n h_n^{2r+1} / \log n} \quad (4.6)$$

Sous l'hypothèse (H_2) , le dernier terme de (4.6) est un terme général d'une série convergente. En remplaçant ϵ par $\epsilon_0 \left(\sqrt{\frac{\log n}{n h_n^{2r+1}}} \right)$, pour $\epsilon_0 > 0$, nous avons

$$\mathbb{P} \left\{ \max_{1 \leq k \leq q_n} \left| \tilde{f}_n^{(r)}(x_k) - E \tilde{f}_n^{(r)}(x_k) \right| > \epsilon_0 \sqrt{\frac{\log n}{n h_n^{2r+1}}} \right\} \leq n^{\frac{1}{2}-c \epsilon_0^2}. \quad (4.7)$$

Par un choix convenable de ϵ_0 , le terme de (4.7) est un terme général d'une série convergente, d'où

$$\sqrt{\frac{n h_n^{2r+1}}{\log n}} \max_{1 \leq k \leq q_n} \left| \tilde{f}_n^{(r)}(x_k) - E \tilde{f}_n^{(r)}(x_k) \right| = o(1) \quad p.s$$

et donc $\Sigma_2 = O\left(\sqrt{\frac{\log n}{nh_n^{2r+1}}}\right)$, et comme $\frac{\log n}{nh_n^{2r+1}} > \frac{1}{nh_n^3}$ pour tout $n \geq 3$, nous avons

$$\sup_{x \in \Theta} \left| \tilde{f}_n^{(r)}(x) - E\tilde{f}_n^{(r)}(x) \right| = O\left(\sqrt{\frac{\log n}{nh_n^{2r+1}}}\right) \quad p.s \text{ quand } n \rightarrow \infty$$

et par conséquent

$$\sup_{x \in \Theta} \left| f_n^{(r)}(x) - f^{(r)}(x) \right| = O\left\{ \max\left(\sqrt{\frac{\log n}{nh_n^{2r+1}}}, h_n^2\right) \right\} \quad p.s \text{ quand } n \rightarrow \infty$$

ce qui achève la démonstration du théorème.

Chapitre 5

Simulations

5.1 Estimation du paramètre de lissage

Dans ce chapitre, nous comparons par simulation sur différentes densités quatre méthodes de sélection du paramètre de lissage qui interviennent dans l'estimation de la densité et ses dérivées par la méthode du noyau :

- La méthode de maximum de vraisemblance avec validation croisée
- La validation croisée biaisée
- La validation croisée non biaisée
- La méthode du *AMISE*.

5.1.1 Maximum de vraisemblance avec validation croisée

Nous faisons intervenir la méthode du maximum de vraisemblance avec validation croisée (Maximum Likelihood Cross-Validation "MLCV") pour l'estimation de la densité ($r = 0$). Or l'estimation des dérivées de la densité pour ($r \geq 1$) ne semble pas possible à cause des valeurs négatives que peuvent prendre les dérivées du noyau. Ce qui empêche la fonction du logarithme d'être défini.

L'estimation du paramètre de lissage optimal par la méthode de maximum de vraisemblance avec validation croisée est obtenue en maximisant l'expression suivante

$$\begin{aligned} MLCV(h) &= \frac{1}{n} \sum_{i=1}^n \log \hat{f}_{h,-i}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \log \left[\sum_{j \neq i} K \left(\frac{X_i - X_j}{h} \right) \right] - \log \left[\frac{(n-1)}{h} \right] \end{aligned}$$

où

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

Le paramètre de lissage qui maximise la fonction $MLCV(h)$ est $h_{MLCV} = \arg \max_h MLCV(h)$.

5.1.2 Validation croisée non biaisée

Cette méthode de validation croisée non biaisée (Unbiased Cross-Validation "UCV") proposée par Rudemo [30] et Bowman [4], appelée aussi méthode des moindres carrés avec validation croisée (Least-Squares Cross-Validation "LSCV") est probablement la méthode la plus connue et la plus utilisée pour l'estimation du paramètre de lissage par la méthode des noyaux.

Considérons une mesure de distance alternative entre \widehat{f}_h et f , l'erreur quadratique intégrée (Integrated Squared Error "ISE") de la fonction de densité f :

$$ISE\{\widehat{f}_h\} = \int [\widehat{f}_h(x) - f(x)]^2 dx = d(\widehat{f}_h, f).$$

En développant, nous obtenons

$$ISE\{\widehat{f}_h\} = \int \widehat{f}_h^2(x) dx - 2 \int \widehat{f}_h(x) f(x) dx + \int f^2(x) dx$$

Par la suite on omet le dernier terme puisqu'il ne dépend pas de h . Donc, le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$\begin{aligned} UCV(h) &= \int_{\mathbb{R}} [\widehat{f}_h(x) - f(x)]^2 dx - \int_{\mathbb{R}} f^2(x) dx \\ &= \int_{\mathbb{R}} \widehat{f}_h^2(x) dx - 2 \int_{\mathbb{R}} \widehat{f}_h(x) f(x) dx \end{aligned}$$

Puisque $\int_{\mathbb{R}} \widehat{f}_h^2(x) dx$ ne dépend pas du paramètre de lissage h , nous pouvons choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de

$$\int_{\mathbb{R}} \widehat{f}_h^2(x) dx - 2 \int_{\mathbb{R}} \widehat{f}_h(x) f(x) dx$$

Le terme croisé s'interprète comme une espérance

$$\int_{\mathbb{R}} \widehat{f}_h(x) f(x) dx \equiv E[\widehat{f}_h(x)]$$

Un estimateur sans biais (leave one out estimator) de l'espérance de $\widehat{f}_h(x)$ est donné par

$$E\left[\widehat{f}_h(x)\right] = \frac{1}{n} \sum_{i=1}^n \widehat{f}_{h,-i}(X_i)$$

car pour la fonction de densité, le concept de validation croisée consiste à supprimer l'observation i de l'échantillon pour l'estimation de la fonction de densité $f(X_i)$ à ce point. cet estimateur est dénoté par $\widehat{f}_{h,-i}(X_i)$ et s'exprime de la manière suivante :

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

A cet effet, le critère à optimiser est alors :

$$UCV(h) = \int_{\mathbb{R}} \widehat{f}_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{h,-i}(X_i)$$

Par ailleurs, nous montrons que

$$\int_{\mathbb{R}} \widehat{f}_h^2(x) dx = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K * K\left(\frac{X_i - X_j}{h}\right) = R(\widehat{f}_h)$$

où $K * K$ est la convolution du noyau par lui même.

Finalement, nous obtenons un paramètre h en minimisant le critère suivant :

$$UCV(h) = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K * K\left(\frac{X_i - X_j}{h}\right) - \frac{2}{nh(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K\left(\frac{X_i - X_j}{h}\right)$$

Le paramètre de lissage qui minimise cette fonction est $h_{UCV} = \arg \min_h UCV(h)$.

Considérons maintenant une mesure de distance alternative entre $\widehat{f}_h^{(r)}$ et $f^{(r)}$, l'erreur quadratique intégrée (Integrated Squared Error "ISE") de la r-ème dérivée de la fonction de densité $f^{(r)}$:

$$ISE\left\{\widehat{f}_h^{(r)}\right\} = \int \left[\widehat{f}_h^{(r)}(x) - f^{(r)}(x)\right]^2 dx = d(\widehat{f}_h^{(r)}, f^{(r)}).$$

En développant, nous obtenons

$$ISE\left\{\widehat{f}_h^{(r)}\right\} = \int \widehat{f}_h^{(r)2}(x) dx - 2 \int \widehat{f}_h^{(r)}(x) f^{(r)}(x) dx + \int f^{(r)2}(x) dx$$

Le critère à optimiser est alors :

$$UCV_r(h) = \int_{\mathbb{R}} \widehat{f}_h^{(r)2}(x)dx - \frac{2}{n}(-1)^r \sum_{i=1}^n \widehat{f}_{h,-i}^{(2r)}(X_i)$$

où

$$\widehat{f}_{h,-i}^{(2r)}(X_i) = \frac{1}{h^{2r+1}(n-1)} \sum_{j=1, j \neq i}^n K^{(2r)}\left(\frac{X_i - X_j}{h}\right)$$

et

$$\int_{\mathbb{R}} \widehat{f}_h^{(r)2}(x)dx = \frac{(-1)^{(r)}}{n^2 h^{2r+1}} \sum_{i=1}^n \sum_{j=1}^n (K * K)^{(2r)}\left(\frac{X_i - X_j}{h}\right) = R(\widehat{f}_h^{(r)})$$

Nous obtenons un paramètre h en minimisant le critère suivant :

$$UCV_r(h) = \frac{(-1)^{(r)}}{nh^{2r+1}} \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (K * K)^{(2r)}\left(\frac{X_i - X_j}{h}\right) - \frac{2}{(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K^{(2r)}\left(\frac{X_i - X_j}{h}\right) \right]$$

où $(K * K)^{(2r)} = K^{(r)} * K^{(r)}$. Le paramètre de lissage qui minimise cette fonction est $h_{UCV_r} = \arg \min_h UCV_r(h)$.

5.1.3 Validation croisée biaisée

Le critère de validation croisée biaisée (Biased Cross-Validation "BCV") a été introduit par Scott et Terrell [32] pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance.

$$BCV(h) = \frac{1}{nh} R(K) + \frac{h^4}{4} \mu_2^2(K) \left[R(f^{(2)}) - \frac{1}{nh^5} R(K^{(2)}) \right]$$

où

$$R(f^{(2)}) = \int_{\mathbb{R}} [f^{(2)}(x)]^2 dx = \frac{1}{n^2 h^5} \sum_{i=1}^n \sum_{j=1}^n K^{(2)} * K^{(2)}\left(\frac{X_j - X_i}{h}\right)$$

et

$$R(K) = \int_{\mathbb{R}} K^2(u) du$$

Notons que

$$K^{(2)} * K^{(2)}(0) = \int K^{(2)}(0 - y) K^{(2)}(y) dy = \int (K^{(2)}(y))^2 dy = R(K^{(2)})$$

Nous obtenons

$$\begin{aligned}
 BCV(h) - \frac{1}{nh}R(K) &= \frac{h^4}{4}\mu_2^2(K) \left[\frac{1}{n^2h^5} \sum_{i=1}^n \sum_{j=1}^n K^{(2)} * K^{(2)} \left(\frac{X_j - X_i}{h} \right) - \frac{1}{nh^5}R(K^{(2)}) \right] \\
 &= \frac{\mu_2^2(K)}{4nh} \left[\frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n K^{(2)} * K^{(2)} \left(\frac{X_j - X_i}{h} \right) + K^{(2)} * K^{(2)}(0) - K^{(2)} * K^{(2)}(0) \right] \\
 &= \frac{\mu_2^2(K)}{2n^2h} \sum_{i=1}^{n-1} \sum_{j=i+1}^n K^{(2)} * K^{(2)} \left(\frac{X_j - X_i}{h} \right)
 \end{aligned}$$

d'où la quantité à minimiser est alors :

$$BCV(h) = \frac{1}{nh}R(K) + \frac{\mu_2^2(K)}{2n^2h} \sum_{i=1}^{n-1} \sum_{j=i+1}^n K^{(2)} * K^{(2)} \left(\frac{X_j - X_i}{h} \right)$$

Nous définissons $h_{BCV} = \arg \min_h BCV(h)$.

Le critère BCV pour les dérivées de la densité devient ainsi

$$BCV_r(h) = \frac{1}{nh^{2r+1}}R(K^{(r)}) + \frac{h^4}{4}\mu_2^2(K) \left[R(f^{(r+2)}) - \frac{1}{nh^{(2r+5)}}R(K^{(r+2)}) \right]$$

où

$$R(f^{(r+2)}) = \int_{\mathbb{R}} [f^{(r+2)}(x)]^2 dx = \frac{(-1)^{(r+2)}}{n^2h^{(2r+5)}} \sum_{i=1}^n \sum_{j=1}^n K^{(r+2)} * K^{(r+2)} \left(\frac{X_j - X_i}{h} \right)$$

et

$$R(K^{(r)}) = \int_{\mathbb{R}} K^{(r)2}(u) du$$

Notons que

$$K^{(r+2)} * K^{(r+2)}(0) = \int K^{(r+2)}(0 - y)K^{(r+2)}(y)dy = \int (K^{(r+2)}(y))^2 dy = R(K^{(r+2)})$$

La quantité à minimiser est alors :

$$BCV_r(h) = \frac{1}{nh^{2r+1}}R(K^{(r)}) + \frac{\mu_2^2(K)}{2n^2h^{2r+1}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n K^{(r+2)} * K^{(r+2)} \left(\frac{X_j - X_i}{h} \right)$$

Nous définissons $h_{BCV_r} = \arg \min_h BCV_r(h)$.

5.1.4 Critère du AMISE

Nous avons vu au chapitre (3) que

$$AMISE(\widehat{f}^{(r)}, f^{(r)}) = \frac{1}{nh^{2r+1}}R(K^{(r)}) + \frac{1}{4}h^4\mu_2(K)^2R(f^{(r+2)}).$$

Donc, le paramètre de lissage optimal h_{AMISE_r} pour l'estimation de la r ème dérivée de la densité est donné par

$$h_{AMISE_r} = \left[\frac{(2r+1)R(K^{(r)})}{\mu_2(K)^2R(f^{(r+2)})} \right]^{1/(2r+5)} n^{-1/(2r+5)}$$

D'où

$$AMISE_r(\widehat{f}^{(r)}) = \frac{2r+5}{4}R(K^{(r)})^{\frac{4}{2r+5}} [\mu_2(K)^2R(f^{(r+2)})/(2r+1)]^{\frac{2r+1}{2r+5}} n^{\frac{-4}{2r+5}}$$

5.2 Simulations et résultats

5.2.1 Simulations

Dans cette étude, nous considérons des simulations qui ont été conduites à partir de deux densités distinctes : une densité gaussienne $N(0; 1)$, et un mélange à 50% de deux densités gaussiennes $N(-1.5; \frac{1}{4})$ et $N(1.5; \frac{1}{4})$:

1. La loi bimodale (Bimodal separated) :

$$f(x) = \frac{1}{2}\phi(x, \mu = -1.5, \sigma^2 = \frac{1}{4}) + \frac{1}{2}\phi(x, \mu = 1.5, \sigma^2 = \frac{1}{4})$$

2. La loi normale centrée et réduite :

$$g(x) = \phi(x, \mu = 0, \sigma^2 = 1)$$

Nous simulons pour chacune de ces lois un échantillon de taille $n = 500$ et nous estimons le paramètre de lissage et l'estimateur à noyau de ces densités ainsi de leurs dérivées, et ce, pour différents noyaux. Les programmes de calculs sont réalisés par le langage R.3.0.0.

Afin de comparer les estimateurs à noyau des dérivées de la densité dans le cas complet iid et le cas censuré iid, nous simulons un échantillon $X = (X_1, X_2, \dots, X_n)$ de loi $N(0, 1)$ de taille $n = 500$ et un échantillon de la loi de censure $C = (C_1, C_2, \dots, C_n)$ de loi $N(0, 1)$

de taille $n = 500$, nous présentons l'estimateur à noyau des dérivées dans le cas complet et censuré dans une même figure.

Les r dérivées de la loi bimodale $f(x; r)$ pour $r = 0, 1, 2, 3, 4, 5, 6$ respectivement sont

$$\begin{aligned}
 f(x; 0) &= \frac{1}{2} \frac{\sqrt{2} e^{-2(x+\frac{3}{2})^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} e^{-2(x-\frac{3}{2})^2}}{\sqrt{\pi}} \\
 f(x; 1) &= \frac{1}{2} \frac{\sqrt{2} (-4x - 6) e^{-2(x+3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x + 6) e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 f(x; 2) &= -2 \frac{\sqrt{2} e^{-2(x+3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x - 6)^2 e^{-2(x+3/2)^2}}{\sqrt{\pi}} - 2 \frac{\sqrt{2} e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 &\quad + \frac{1}{2} \frac{\sqrt{2} (-4x + 6)^2 e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 f(x; 3) &= -6 \frac{\sqrt{2} (-4x - 6) e^{-2(x+3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x - 6)^3 e^{-2(x+3/2)^2}}{\sqrt{\pi}} \\
 &\quad - 6 \frac{\sqrt{2} (-4x + 6) e^{-2(x-3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x + 6)^3 e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 f(x; 4) &= 24 \frac{\sqrt{2} e^{-2(x+3/2)^2}}{\sqrt{\pi}} - 12 \frac{\sqrt{2} (-4x - 6)^2 e^{-2(x+3/2)^2}}{\sqrt{\pi}} \\
 &\quad + \frac{1}{2} \frac{\sqrt{2} (-4x - 6)^4 e^{-2(x+3/2)^2}}{\sqrt{\pi}} + 24 \frac{\sqrt{2} e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 &\quad - 12 \frac{\sqrt{2} (-4x + 6)^2 e^{-2(x-3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x + 6)^4 e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 f(x; 5) &= 120 \frac{\sqrt{2} (-4x - 6) e^{-2(x+3/2)^2}}{\sqrt{\pi}} - 20 \frac{\sqrt{2} (-4x - 6)^3 e^{-2(x+3/2)^2}}{\sqrt{\pi}} \\
 &\quad + \frac{1}{2} \frac{\sqrt{2} (-4x - 6)^5 e^{-2(x+3/2)^2}}{\sqrt{\pi}} + 120 \frac{\sqrt{2} (-4x + 6) e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 &\quad - 20 \frac{\sqrt{2} (-4x + 6)^3 e^{-2(x-3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x + 6)^5 e^{-2(x-3/2)^2}}{\sqrt{\pi}}
 \end{aligned}$$

$$\begin{aligned}
 f(x; 6) = & -480 \frac{\sqrt{2} e^{-2(x+3/2)^2}}{\sqrt{\pi}} + 360 \frac{\sqrt{2} (-4x - 6)^2 e^{-2(x+3/2)^2}}{\sqrt{\pi}} \\
 & - 30 \frac{\sqrt{2} (-4x - 6)^4 e^{-2(x+3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x - 6)^6 e^{-2(x+3/2)^2}}{\sqrt{\pi}} \\
 & - 480 \frac{\sqrt{2} e^{-2(x-3/2)^2}}{\sqrt{\pi}} + 360 \frac{\sqrt{2} (-4x + 6)^2 e^{-2(x-3/2)^2}}{\sqrt{\pi}} \\
 & - 30 \frac{\sqrt{2} (-4x + 6)^4 e^{-2(x-3/2)^2}}{\sqrt{\pi}} + \frac{1}{2} \frac{\sqrt{2} (-4x + 6)^6 e^{-2(x-3/2)^2}}{\sqrt{\pi}}
 \end{aligned}$$

Les r dérivées de la loi normale centrée réduite $g(x; r)$ pour $r = 0, 1, 2, 3, 4, 5, 6$ respectivement sont

$$g(x; 0) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

$$g(x; 1) = -\frac{1}{2} \frac{x e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}}$$

$$g(x; 2) = -\frac{1}{2} \frac{e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} + \frac{1}{2} \frac{x^2 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}}$$

$$g(x; 3) = \frac{3}{2} \frac{x e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} - \frac{1}{2} \frac{x^3 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}}$$

$$g(x; 4) = \frac{3}{2} \frac{e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} - 3 \frac{x^2 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} + \frac{1}{2} \frac{x^4 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}}$$

$$f(x; 5) = -\frac{15}{2} \frac{x e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} + 5 \frac{x^3 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} - \frac{1}{2} \frac{x^5 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}}$$

$$f(x; 6) = -\frac{15}{2} \frac{e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} + \frac{45}{2} \frac{x^2 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} - \frac{15}{2} \frac{x^4 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}} + \frac{1}{2} \frac{x^6 e^{-\frac{x^2}{2}} \sqrt{2}}{\sqrt{\pi}}$$

5.2.2 Résultats

Nous présentons les résultats (tables 5.1 et 5.2) des simulations obtenus pour l'estimation du paramètre de lissage des dérivées d'une densité dans le cas : d'une loi bimodale et d'une loi normale centrée réduite par les méthodes de sélection : *MLCV*, *UCV*, *BCV*

et *AMISE*.

Dans les figures 5.1 et 5.2, nous représentons l'estimateur de la fonction de densité de chacune des deux lois. L'estimateur représenté par la courbe noire est calculé à l'aide de la fenêtre optimale h_{MLCV} pour $r = 0$, pour un ensemble de 500 données simulées, l'autre courbe (rouge) correspond à la fonction de densité théorique.

Dans les figures 5.3 et 5.4, nous représentons l'estimateur des dérivées de la fonction de densité de la loi bimodale pour $r=0, 2$ respectivement, où l'estimateur est calculé à l'aide de la fenêtre h_{UCV} pour un ensemble de 500 données simulées.

Dans les figures 5.5 et 5.6, nous représentons l'estimateur des dérivées de la fonction de densité de la loi bimodale pour $r=1, 2$ respectivement, où l'estimateur est calculé à l'aide de la fenêtre h_{BCV} pour un ensemble de 500 données simulées.

Dans les figures 5.7 et 5.8, nous représentons l'estimateur des dérivées de la fonction de densité de la loi bimodale pour $r = 0, 3$ respectivement, où l'estimateur est calculé à l'aide de la fenêtre h_{AMISE} pour un ensemble de 500 données simulées.

Afin de voir l'impact du choix de la méthode d'estimation du paramètre de lissage sur la qualité d'estimation des dérivées de la fonction de densité, nous représentons dans les figures 5.9 et 5.10, l'estimateur des dérivées de la fonction de densité de chacune des deux lois pour $r = 1$, où l'estimateur est calculé à l'aide des fenêtres h_{UCV} , h_{BCV} et h_{AMISE} pour un ensemble de 500 données simulées avec un noyau gaussien.

Afin de pouvoir visualiser l'impact du choix du noyau sur la qualité d'estimation des deux lois pour différentes fenêtres, nous présentons sur les figures 5.11 et 5.12, l'estimateur des dérivées de la fonction de densité de la loi bimodale pour $r = 0, 1$ respectivement, calculé à l'aide de la fenêtre h_{UCV} , puis nous présentons sur la figure 5.13 l'estimateur de la dérivée de la fonction de densité de la loi $N(0; 1)$ pour $r = 1$ calculé à l'aide de la fenêtre h_{BCV} , et ce pour différents noyaux.

Les figure 5.14 et 5.15 présentent l'estimateur à noyau de la rème dérivée de la densité $N(0; 1)$ pour $r = 0$ et $r = 1, 3$ respectivement, pour le cas complet et censuré pour une fenêtre h donnée. Les valeurs de h sont celles obtenues par validation croisée non biaisée

"UCV" dans le cas complet (table 5.2) et donc ils sont choisis d'une manière arbitraire pour le cas censuré.

Kernel		Gaussian	Tricube	Triweight	Epanechnikov
Method	r	h	h	h	h
MLCV	0	0,1443	0,4818	0,5456	0,2872
UCV	0	0,2036	0,4633	0,6203	0,3565
	1	0,2845	0,4765	0,7862	0,1233
	2	0,3626	1,6537	1,7393	
	3	0,5072	0,1475	0,1685	
	4	0,5891	2,7798		
	5	0,6655			
	6	0,7392			
BCV	0	0,1956	0,4926	0,5874	1,1594
	1	0,2557	0,6787	0,6749	
	2	0,2671	1,3874	1,5849	
	3	0,1770	1,3848	1,5843	
	4	0,1798	1,3874	1,5839	
	5	0,1849	1,3840		
	6	0,1932	1,3870		
AMISE	0	0,2035	0,5111	0,6002	2,3236
	1	0,6568	1,6247	1,8304	
	2	1,0485	2,7774	3,1746	
	3	0,3180	0,1475	0,1685	
	4	1,0361	2,7808	3,1769	
	5	0,0550	0,1475		
	6	1,0367	2,7778		

TABLE 5.1 – Résultats des simulations de la loi Bimodale pour $n=500$

Compte tenu des résultats des simulations, on peut constater que l'estimateur de la dérivée d'une densité dépend du choix du noyau et de la fenêtre h .

Le choix du noyau influe peu sur les performances de ces estimateurs ce qui nous conduit à privilégier la solution classique du noyau gaussien pour $r = 0$.

Le choix de la fenêtre h est, en revanche crucial. Pour $r = 0$, comme fenêtre optimale, nous choisissons la valeur qui minimise le *MLCV*.

Pour $r \geq 1$ la meilleure méthode de sélection de la fenêtre change d'un noyau à un autre et d'une loi à une autre. Par exemple, pour $r = 1$, la valeur optimale qui minimise le *UCV* pour la loi $N(0; 1)$ est celle obtenue avec le noyau gaussien. Tandis que pour la loi bimodale, parmi les noyaux testés, nous remarquons que le noyau tricube améliore la qualité de l'estimateur de la dérivée.

Nous constatons, d'après les résultats obtenues, que la méthode de validation croisée non biaisée "UCV" donne de bons résultats pour les densités unimodales suffisamment lisses. En égard à la méthode de validation croisée biaisée "BCV", il serait intéressant – dans un travail futur – d'effectuer des simulations extensives et observer la robustesse de la méthode par rapport aux propriétés de la densité cible.

Cependant, un point important doit être impérativement souligné. La sélection de la largeur de fenêtre par AMISE n'est valable qu'asymptotiquement. A taille d'échantillon fixé, l'analyse s'avère délicate. Ensuite, les formules qui sont au coeur de la méthode imposent des restrictions sur la densité f souvent difficiles à vérifier dans la pratique. Il est instructif de souligner que dans le AMISE, le terme du biais et le terme de la variance varient en sens inverse par rapport à h : une largeur de fenêtre trop importante entrainera une augmentation du biais et une diminution de la variance (phénomène de sur-lissage), alors qu'une largeur de fenêtre trop petite provoquera une ination de la variance et une diminution du biais (phénomène de sous-lissage).

Aussi, en faisons les mêmes simulations pour différentes tailles d'échantillon $n = 100, 500$ et 1000 , nous constatons que la fenêtre h est une quantité déterministe (qui dépend du nombre d'observation n). Nous avons remarqué que les performances moyennes de l'estimateur h_n se dégradent lorsque la taille d'échantillon augmente.

Kernel		Gaussian	Cosine	Biweight	Triangular
Method	r	h	h	h	h
MLCV	0	0,3533	0,4545	0,5844	0,5615
UCV	0	0,3682	0,8381	0,9867	0,9145
	1	0,5262	0,0809	1,1217	
	2	0,6679	0,0809	0,0932	
	3	0,6761	0,0809		
	4	0,6768	0,0809		
	5	0,6769	0,0809		
	6	0,6769	0,0809		
BCV	0	0,3269	0,7605	0,8759	
	1	0,3379	0,7608	0,8751	
	2	0,1148	0,7598	0,8765	
	3	0,2067	0,7608		
	4	0,2128	0,7604		
	5	0,0878	0,7588		
	6	0,0890	0,7609		
AMISE	0	0,3404	1,5226	0,9321	
	1	0,6764	1,3605	0,6618	
	2	0,6772	1,5227	1,7551	
	3	0,0359	0,9451		
	4	0,6769	1,5245		
	5	0,0359	0,8786		
	6	0,6762	1,5228		

TABLE 5.2 – Résultats des simulations de la loi Normale centrée réduite pour n=500

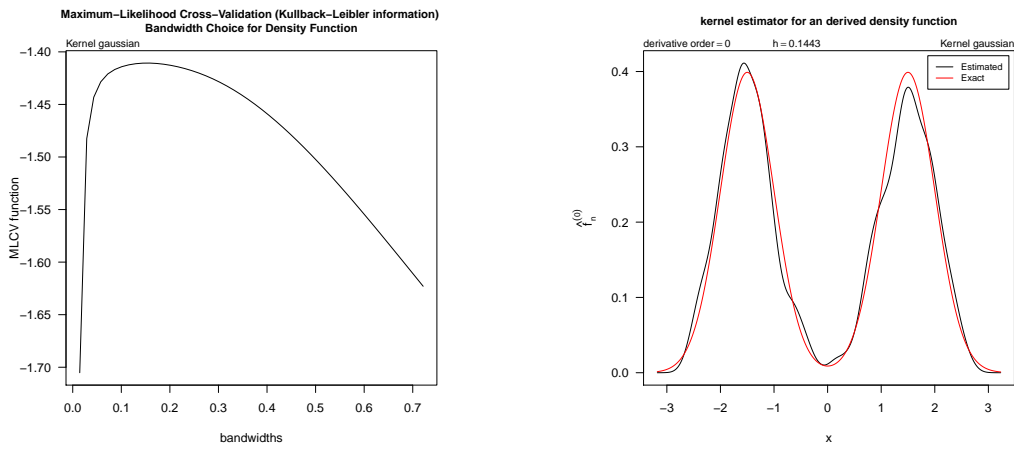


FIGURE 5.1 – MLCV pour la loi Bimodale, $n=500$

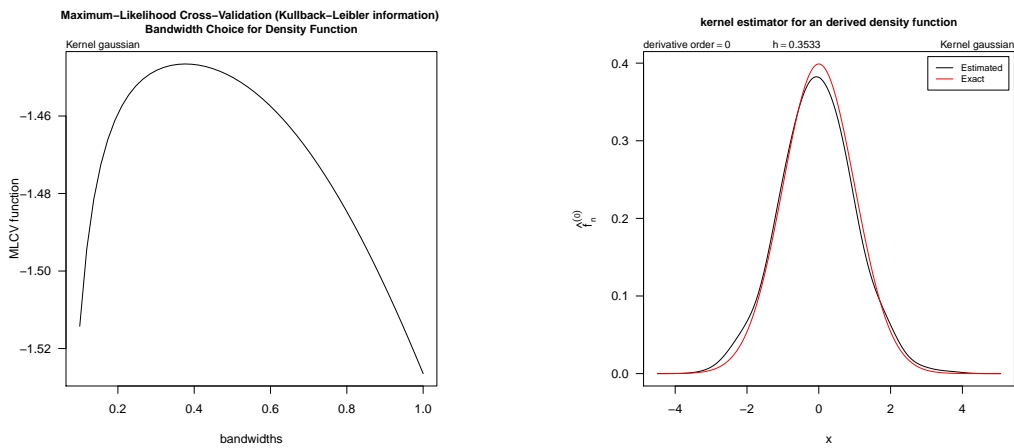


FIGURE 5.2 – MLCV pour la loi $N(0,1)$, $n=500$

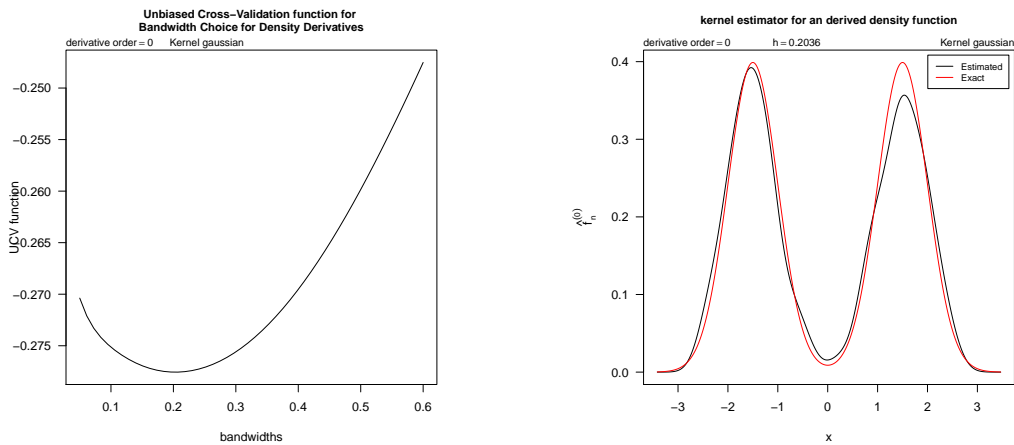


FIGURE 5.3 – UCV pour la loi Bimodale, $r=0$, $n=500$

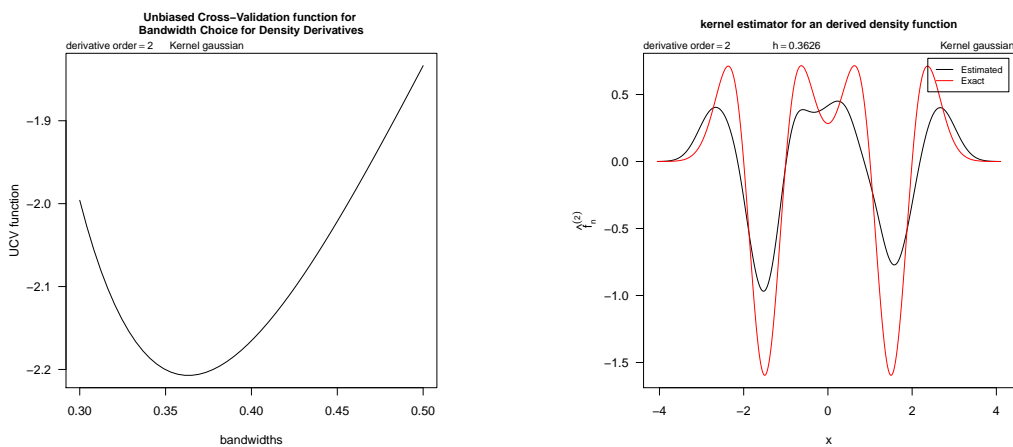


FIGURE 5.4 – UCV pour la loi Bimodale, $r=2$, $n=500$

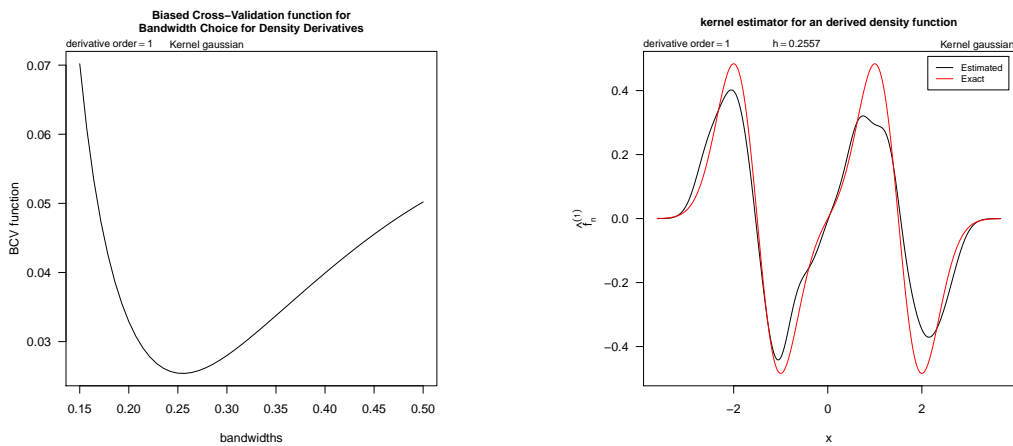


FIGURE 5.5 – BCV pour la loi Bimodale, $r=1$, $n=500$

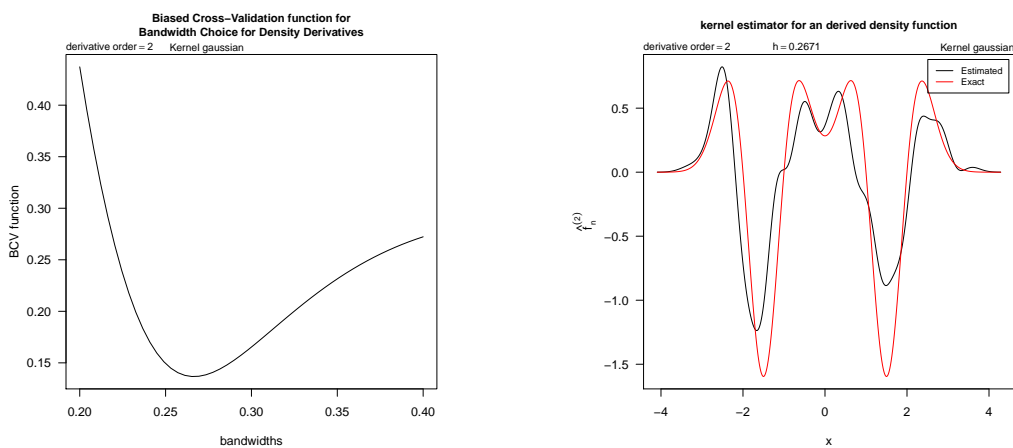


FIGURE 5.6 – BCV pour la loi Bimodale, $r=2$, $n=500$

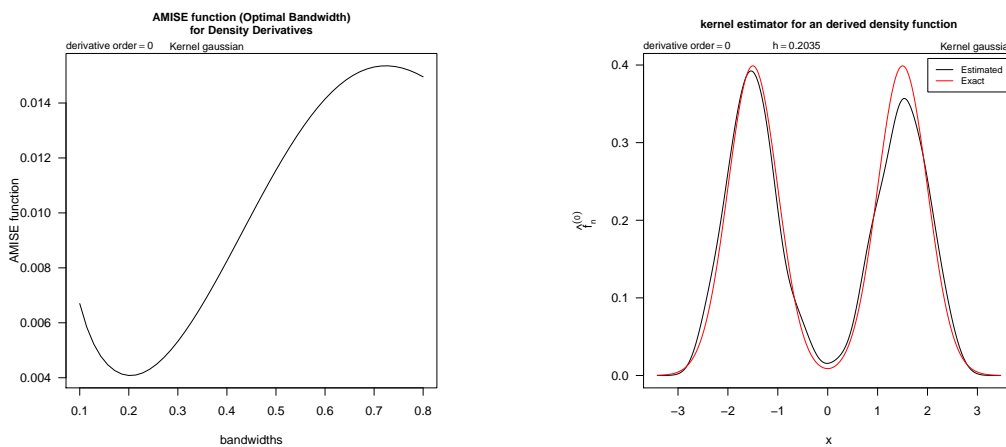


FIGURE 5.7 – AMISE pour la loi Bimodale, $r=0$, $n=500$

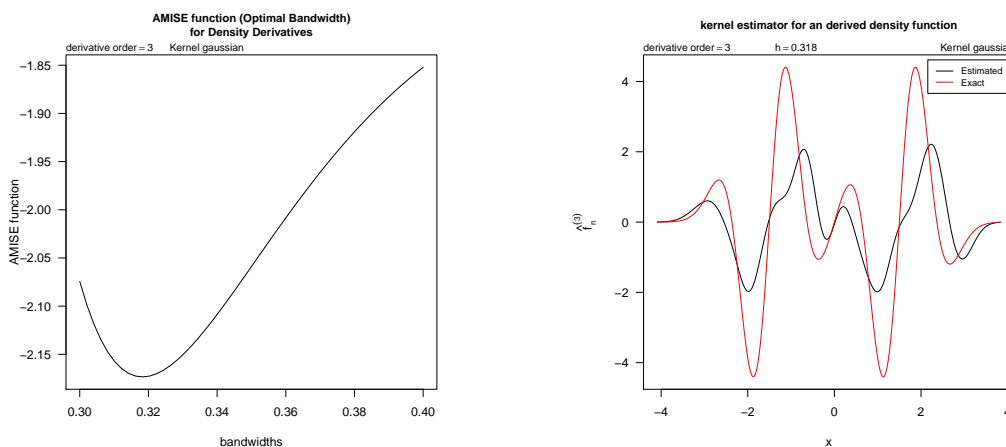


FIGURE 5.8 – AMISE pour la loi Bimodale, $r=3$, $n=500$

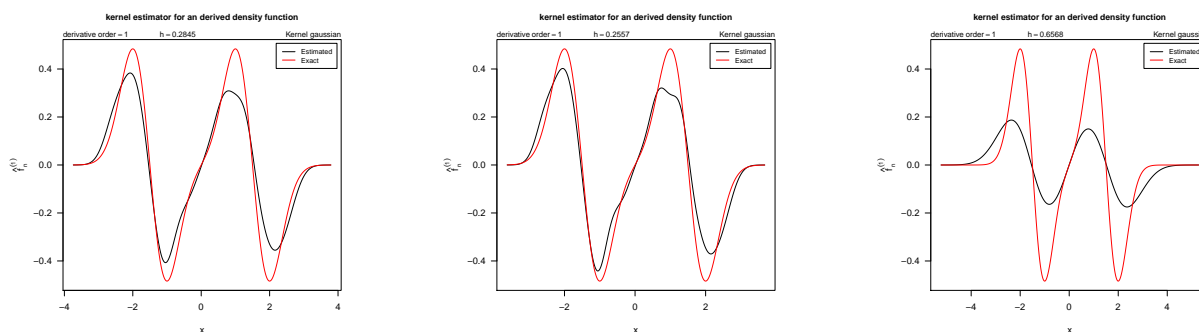


FIGURE 5.9 – Estimation de la loi Bimodale par les méthodes UCV, BCV, AMISE respectivement, pour noyau gaussien, $r=1$, $n=500$

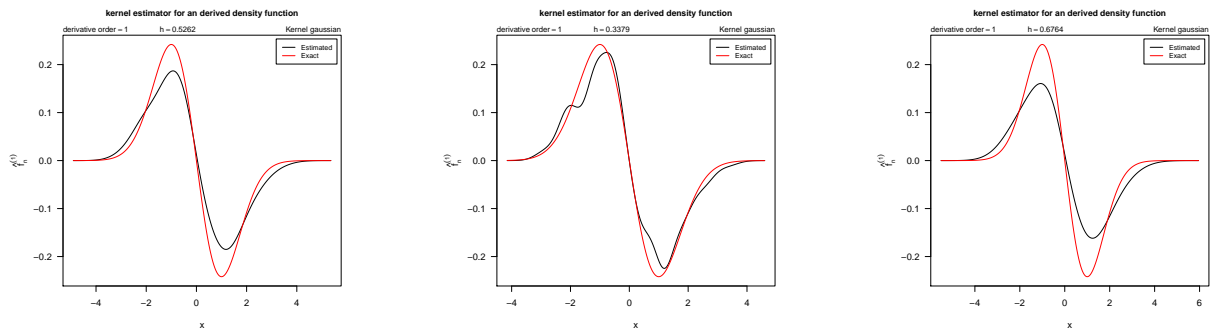


FIGURE 5.10 – Estimation de la loi $N(0, 1)$ par les méthodes UCV, BCV, AMISE respectivement, pour noyau gaussien, $r=1$, $n=500$

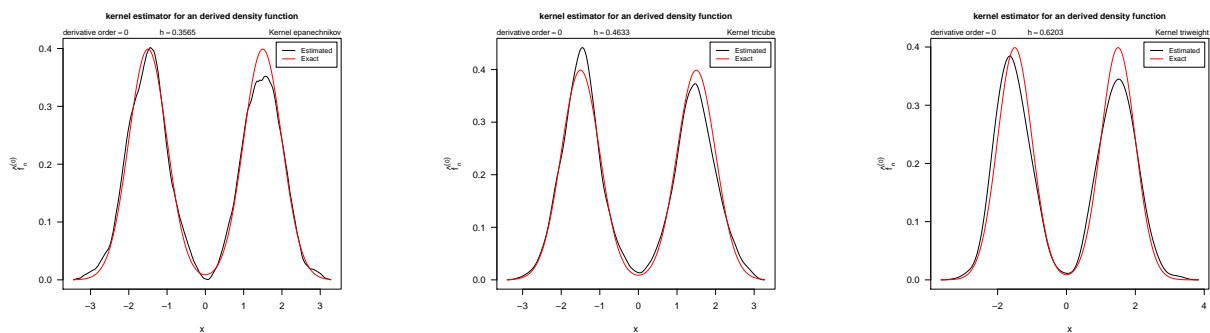


FIGURE 5.11 – Impact du choix du noyau, avec la méthode UCV, loi Bimodale, $r=0$, $n=500$

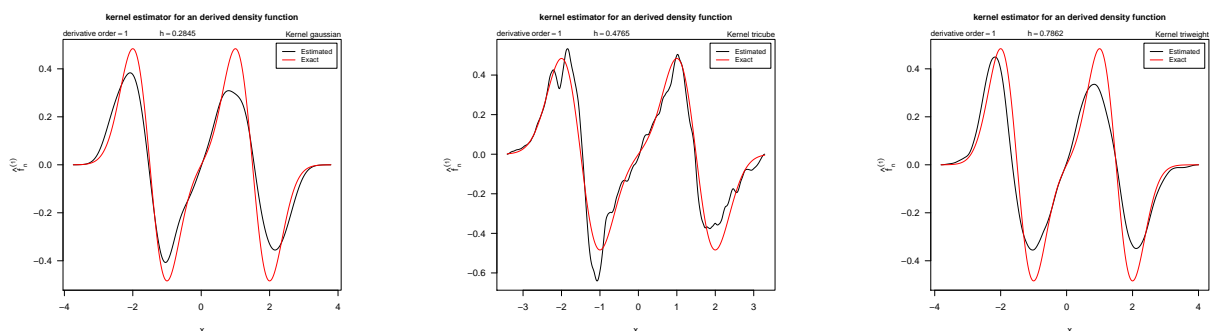


FIGURE 5.12 – Impact du choix du noyau, avec la méthode UCV, loi Bimodale, $r=1$, $n=500$

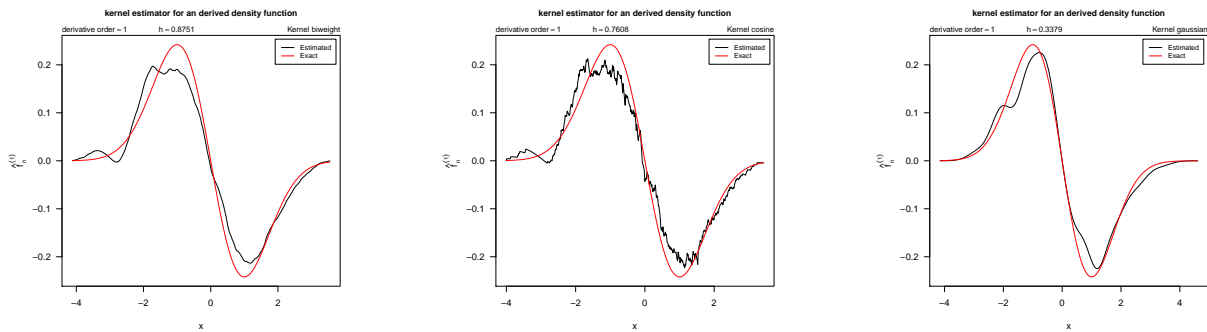


FIGURE 5.13 – Impact du choix du noyau, avec la méthode BCV, loi $N(0,1)$, $r=1$, $n=500$

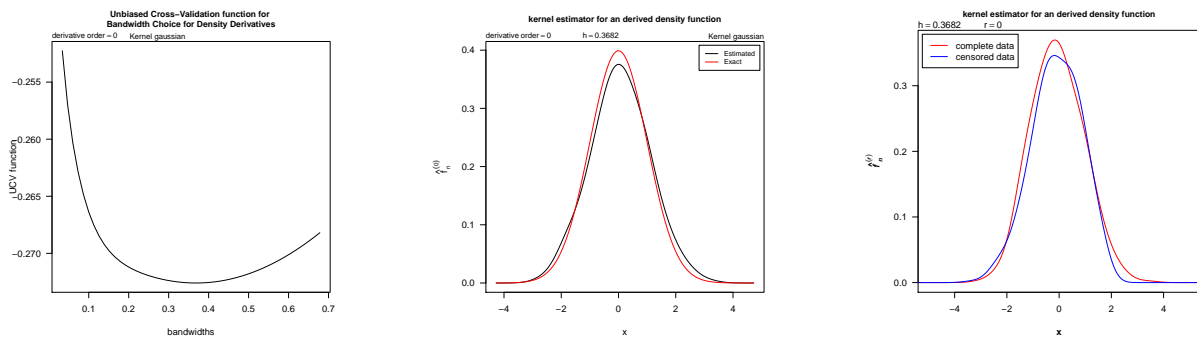


FIGURE 5.14 – $X : N(0,1)$, $C : N(0,1)$, $r=0$, $n=500$

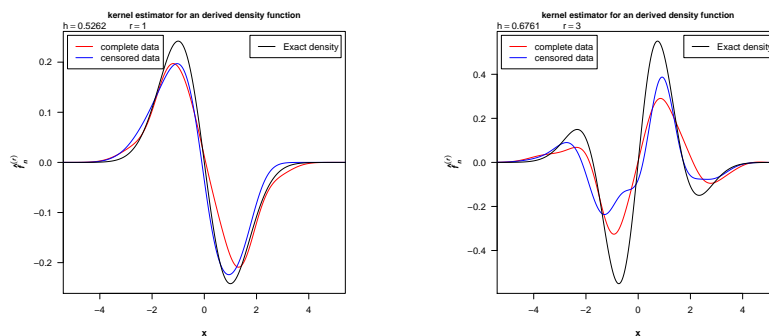


FIGURE 5.15 – $X : N(0,1)$, $C : N(0,1)$, $r=1, 3$ respectivement, $n=500$

Conclusion et Perspectives

Conclusion

La statistique non paramétrique est en plein développement mais reste encore une théorie inachevée. Les problèmes d'optimalité, notamment, ne sont pas résolus. Cependant les méthodes non paramétriques ont déjà fait leurs preuves : elles sont simples, robustes et souvent supérieures aux méthodes paramétriques.

L'estimation des dérivées d'une densité de probabilité à partir d'un échantillon de variables aléatoires indépendantes est un problème de la statistique non paramétrique qui a fait l'objet d'une littérature abondante, et qui a connu peu de développements théoriques et pratiques. Il nous a donc semblé intéressant de faire le point sur quelques propriétés de l'estimateur à noyau des dérivées de la densité pour des données censurées à droite avec l'hypothèse d'indépendance des durées de survie et les variables de censure.

L'estimateur à noyau des dérivées de la densité est flexible, dans la mesure où il laisse à l'utilisateur une grande latitude non seulement dans le choix du noyau K , mais encore dans le choix du paramètre de lissage h .

Lorsqu'on se limite aux noyaux K positifs, les vitesses de convergence varient peu en fonction de K et les critères essentiels du choix du noyau sont alors la simplicité, la vitesse de calcul d'une part et la régularité de la courbe à obtenir d'une autre part.

En revanche, le choix du paramètre de lissage étant crucial, aussi bien pour la précision locale que pour la précision globale de l'estimateur. A ce jour, de nombreuses méthodes (dites de sélection) ont été proposées afin de choisir la meilleure fenêtre possible.

Dans ce mémoire, nous avons effectué un point bibliographique sur quelques unes de ces

méthodes de sélection : le *UCV*, *BCV*, *MLCV* et le *AMISE*.

Il est essentiel de comprendre qu'il n'existe pas de méthode de sélection du paramètre de lissage qui soit intrinsèquement meilleure que toutes les autres. L'expérience montre que chaque méthode de sélection possède, en quelque sorte, ses densités de prédilection : une première méthode fonctionnera par exemple correctement pour des densités unimodales, telle que la loi $N(0, 1)$, alors qu'une seconde fournira de bien meilleurs résultats pour des densités à queues lourdes, telle que la loi bimodale.

Dès lors, toute comparaison objective et équitable entre les différentes méthodes de sélection devient difficile, voir impossible. Donc, il faut insister aujourd'hui sur la nécessité de s'orienter vers des méthodes de sélection plus universelles et effectuer des simulations extensives en observant la robustesse de la méthode par rapport aux propriétés de la densité cible.

Perspectives

En perspectives de recherche, nous présentons quelques projets théoriques et pratiques se situant directement en continuité de l'étude effectuée dans le cadre de ce mémoire.

En complément des résultats du chapitre 4, où nous avons établi la vitesse de convergence uniforme presque sûre d'un estimateur des dérivées de la fonction de densité censurée, divers travaux intéressants seront en cours au moyen terme. Ci dessus, nous citons certains de ces travaux :

- Etude du comportement asymptotiques de la fenêtre de lissage pour les dérivées de la densité pour des données censurées.
- Adapter cet estimateur au cadre conditionnel, i.e ; étudier les propriétés asymptotiques d'un estimateur à noyau (ou double noyau) pour les dérivées de la densité conditionnelles dans le cadre des données censurées.
- Affiner les critères de sélection du paramètre de lissage et utilisation du bootstrap pour le choix de la fenêtre optimale pour cet estimateur.

Annexe

Modes de convergences

Convergence en loi On dit que (X_n) converge en loi vers X ($X_n \xrightarrow{L} X$) si la suite de fonctions F_n converge, point par point, vers F :

$$\forall x, \lim_{n \rightarrow \infty} |F_n(x) - F(x)| = 0.$$

Convergence en probabilité On dit que (X_n) converge en probabilité vers X ($X_n \xrightarrow{\mathbb{P}} X$) si :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

La convergence en probabilité est utilisée pour exprimer la loi faible des grands nombres.

Convergence presque sûre On dit que (X_n) converge presque sûrement vers X ($X_n \xrightarrow{ps} X$) si et seulement si :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0, \forall n \geq 1.$$

La convergence presque sûre se distingue de la convergence en probabilité par le fait qu'elle est plus générale ($\forall n \geq 1$). Intuitivement, elle implique qu'une fois que la série X_n s'approche de X , elle ne s'en écarte pas.

On voit rapidement avec la définition (prenons $n=1$) que la convergence presque sûre implique la convergence en probabilité.

$$X_n \xrightarrow{ps} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$$

La convergence presque sûre est utilisée pour exprimer la loi forte des grands nombres.

Convergence en moyenne quadratique On dit que (X_n) converge en moyenne quadratique vers X ($X_n \xrightarrow{mq} X$) si :

$$\lim_{n \rightarrow \infty} E[|X_n - X|^2] \rightarrow 0.$$

Propriétés :

1. Convergence presque sûre \Rightarrow Convergence en probabilité.
2. Convergence en moyenne quadratique \Rightarrow Convergence en probabilité.
3. $X_n \rightarrow X$ en loi $\Rightarrow \varphi(X_n) \rightarrow \varphi(X)$ en loi pour toute fonction *continue* φ .
4. $E(X_n) \rightarrow c$ et $V(X_n) \rightarrow 0 \Rightarrow X_n \rightarrow c$ en moyenne quadratique.

Taux de convergence

A travers cette section, nous introduirons des notations très importantes, appelées **notations** O , o , qui se lisent "notation grand- O ", et "notation petit- o ". Ici O et o indiquent l'ordre et sont souvent considérés comme des symboles d'ordre. De manière approximative, nous dirons qu'une certaine quantité est, disons, $O(x)$, lorsqu'elle est du même ordre, asymptotiquement, que la quantité x , alors qu'elle sera $o(x)$ quand elle sera d'un ordre inférieur à la quantité x . La signification précise de tout cela sera indiquée par la suite.

Définition

Si $f(\cdot)$ et $g(\cdot)$ sont deux fonctions réelles de la variable positive entière n , alors la notation

$$f(n) = o(g(n)) \quad \text{quand } n \rightarrow \infty,$$

signifie que

$$\lim_{n \rightarrow \infty} \left(\frac{f(n)}{g(n)} \right) = 0.$$

Nous pouvons dire que $f(n)$ est d'un ordre plus faible que $g(n)$ asymptotiquement ou quand n tend vers l'infini.

Notons que $g(n)$ elle-même peut avoir toute sorte de comportement quand n tend vers l'infini. Elle peut posséder une limite ou pas, et si c'est le cas, cette limite peut être nulle, finie et non nulle, ou infinie. Ce qui est important est la comparaison opérée par le ratio. Le plus souvent $g(n)$ est une puissance de n , positive, négative, ou nulle. Dans le dernier cas, puisque $n^0 = 1$ pour tout n , nous devrions écrire $f(n) = o(1)$, et ceci signifierait d'après la définition que

$$\lim_{n \rightarrow \infty} \left(\frac{f(n)}{1} \right) = \lim_{n \rightarrow \infty} (f(n)) = 0,$$

autrement dit, que $f(n)$ tend vers zéro quand n tend vers l'infini. Mais si nous disons que $f(n) = o(n^{-1})$, par exemple, ou que $f(n)$ est $o(n^{-1})$, nous signifions que $f(n)$ tend vers zéro plus vite que n^{-1} . Nous pourrions également dire que $f(n)$ est $o(n)$, et nous ne saurions pas alors si $f(n)$ possède une limite quand $n \rightarrow \infty$. Mais nous savons que si $f(n)$ tend vers l'infini, c'est moins rapidement que n .

La notation grand- O , qui exprime la **relation d'ordre identique**, est plus précise que la notation petit- o , puisqu'elle nous indique le taux le plus fort auquel les quantités peuvent varier avec n . En voici la définition.

Définition

Si $f(\cdot)$ et $g(\cdot)$ sont deux fonctions réelles de la variable positive entière n , alors la notation

$$f(n) = O(g(n))$$

signifie qu'il existe une constante $K > 0$, indépendante de n , et un entier positif N tels que $|f(n)/g(n)| < K$ pour tout $n > N$.

Nous disons que $f(n)$ et $g(n)$ sont du même ordre asymptotiquement ou quand $n \rightarrow \infty$. Une fois encore c'est le ratio de $f(n)$ sur $g(n)$ qui est en cause. La définition n'exclut pas la possibilité que la limite du ratio soit nulle, de sorte que l'expression verbale "du même ordre" peut être trompeuse.

Une autre relation évite cette incertitude : nous l'appelons l'**égalité asymptotique**.

Définition : Si $f(\cdot)$ et $g(\cdot)$ sont des fonctions réelles de la variable positive entière n , alors elles sont asymptotiquement égales si

$$\lim_{n \rightarrow \infty} \left(\frac{f(n)}{g(n)} \right) = 1.$$

Nous notons $f(n) \stackrel{a}{=} g(n)$. La notation standard pour cette relation n'est pas $\stackrel{a}{=}$, mais \approx . Puisque le symbole \approx est utilisé pour désigner la distribution d'une variable aléatoire, cela justifiera l'usage de l'autre symbole dans ce mémoire.

L'égalité asymptotique évite la difficulté à laquelle nous avons fait allusion en connexion avec le grand- O ou relation du même ordre, aux dépens d'une condition plus forte. Contrairement à l'égalité asymptotique, la relation grand- O ne nécessite pas que le ratio $f(n)/g(n)$ ait une quelconque limite. Il peut en avoir une, mais il peut également varier perpétuellement entre des bornes.

Les relations définies plus tôt sont consacrées aux suites réelles non stochastiques. Les relations dites relations d'ordre stochastiques sont d'un intérêt encore plus grand. Elles sont parfaitement analogues aux relations définies plus haut, mais reposent par contre sur l'une ou l'autre forme de la convergence stochastique. Formellement :

Définition

Si a_n est une suite de variables aléatoires, et $g(n)$ une fonction réelle d'un argument positif entier n , alors la notation $a_n = o_p(g(n))$ signifie que

$$p \lim_{n \rightarrow \infty} \left(\frac{a_n}{g(n)} \right) = 0.$$

De façon similaire la notation $a_n = O_p(g(n))$ signifie qu'il existe une constante K telle que, pour tout $\epsilon > 0$, il existe un entier positif N tel que

$$Pr \left(\left| \frac{a_n}{g(n)} \right| > K \right) < \epsilon \quad \text{pour tout } n > N.$$

Si b_n est une autre suite de variables aléatoires, la notation $a_n \stackrel{a}{=} b_n$ signifie que

$$p \lim_{n \rightarrow \infty} \left(\frac{a_n}{b_n} \right) = 1.$$

Il sera toujours évident de savoir si les variables aléatoires en question sont aléatoires. Quand elle le sont, $O(\cdot)$ et $o(\cdot)$ devraient être interprétés comme $O_p(\cdot)$ et $o_p(\cdot)$.

Les symboles d'ordre sont très simples à manipuler, et nous présentons maintenant quelques règles de manipulations. Pour faire simple, nous nous restreignons aux fonctions $g(n)$ qui sont simplement des puissances de n . Les règles pour l'addition et la soustraction sont

$$O(n^p) \pm O(n^q) = O(n^{\max(p,q)});$$

$$o(n^p) \pm o(n^q) = o(n^{\max(p,q)});$$

$$O(n^p) \pm o(n^q) = O(n^p) \quad \text{si } p \geq q;$$

$$O(n^p) \pm o(n^q) = o(n^q) \quad \text{si } p < q.$$

Les règles pour la multiplication, et aussi pour la division, sont

$$O(n^p)O(n^q) = O(n^{p+q});$$

$$o(n^p)o(n^q) = o(n^{p+q});$$

$$O(n^p)o(n^q) = o(n^{p+q}).$$

Lemme de Borel-Cantelli

Soit (Ω, A, P) un espace de probabilité, et $(A_n)_{n \in \mathbb{N}}$ une famille d'évènement.

Si $\sum_{n \in \mathbb{N}} P(A_n) < +\infty$, alors $P(\overline{\lim_{n \in \mathbb{N}} A_n}) = 0$.

Si $\sum_{n \in \mathbb{N}} P(A_n) = +\infty$, et si les $(A_n)_{n \in \mathbb{N}}$ sont mutuellement indépendants, alors

$$P(\overline{\lim_{n \in \mathbb{N}} A_n}) = 1.$$

Fonction hölderienne

Soit f une fonction définie sur un intervalle I de \mathbb{R} . On dit que f est γ -hölderienne avec γ dans $]0, 1]$ s'il existe une constante C_γ telle que :

$$\forall x, y \in I, |f(x) - f(y)| \leq C_\gamma |x - y|^\gamma.$$

L'inégalité de Hoeffding

Soit X_1, X_2, \dots, X_n une suite de v.a indépendante avec $a_i \leq X_i \leq b_i$ pour $1 \leq i \leq n$. Alors $\forall \lambda > 0$

$$P(\sqrt{n}(\bar{X} - E\bar{X}) \geq \lambda) \leq \exp\left(-2\lambda^2 / \sum_{i=1}^n (b_i - a_i)^2\right).$$

L'inégalité de Bernstein

Soit U_1, U_2, \dots, U_n une suite de v.a indépendante et centrées avec $|U_i| \leq M, \forall M > 0$. Soit $S_i^2 = Var(U_i)$

$$P\left(\left|\sum_{i=1}^n U_i\right| > \eta\right) \leq 2 \exp\left\{\frac{-\eta^2}{2[V + \frac{1}{3}M \eta]}\right\},$$

où $V \geq \sum_{i=1}^n S_i^2$.

L'inégalité de Lengart

Nous rappelons ici un résultat bien connu dans l'étude des intégrales Kaplan-Meier. Il met en avant l'intérêt de l'approche martingale dans le contexte des données censurées. Au préalable, nous rappelons la définition de la variation quadratique $\langle M(t), M(t) \rangle$. Soit M une martingale par rapport à une filtration $\{F_t, t \geq 0\}$. Si M est continue à droite et si pour tout $t \geq 0$, $E[M^2(t)] < \infty$, alors il existe un unique processus croissant prévisible continu à droite noté $\langle M(\cdot), M(\cdot) \rangle$ tel que :

(i) $\langle M(0), M(0) \rangle = 0$ presque sûrement,

- (ii) pour tout $t \geq 0$, $E[\langle M(t), M(t) \rangle] < \infty$,
- (iii) $M(t)^2 - \langle M(t), M(t) \rangle$ est une martingale continue à droite par rapport à la filtration F_t .

Pour plus de détails sur les outils martingales dans le contexte des données censurées, voir Fleming et Harrington (1991).

Voici maintenant l'énoncé de l'inégalité de Langlart, sous une forme appropriée à notre contexte.

Théorème 5.1. *Soit M une martingale par rapport à la filtration $\{F_t, t \geq 0\}$ de carré intégrable. Soit f un processus prévisible, localement borné. Alors, pour tout temps d'arrêt τ tel que $\mathbb{P}(\tau < \infty) = 1$ et pour tous $\epsilon, \eta > 0$, nous avons :*

$$\mathbb{P} \left(\sup_{t \leq \tau} \left\{ \int_0^t f(s) dM(s) \right\}^2 \geq \epsilon \right) \leq \frac{\eta}{\epsilon} + \mathbb{P} \left(\int_0^\tau f(s)^2 d \langle M(t), M(t) \rangle (s) \geq \eta \right).$$

Nous pouvons retrouver l'énoncé ainsi que la démonstration de ce théorème dans Fleming and Harrington[10].

Bibliographie

- [1] Akritas, M.G.(2000) : The Central Limit Theorem Under Censoring. *Bernoulli*, Vol. 6, pp. 1109-1120.
- [2] Bartlett, M. S. (1963) : Statistical estimation of density functions. *Sankhya : The Indian Journal of Statistics, Series A*, Vol. 25, pp. 245-54.
- [3] Bhattacharya, P. K.(1967) : Estimation of a Probability Density Function and Its Derivatives. *Sankhya : The Indian Journal of Statistics, Series A*, Vol. 29, No. 4, pp.373-382.
- [4] Bowman, A. W. (1984) : An alternative method of cross-validation for the smoothing density estimates. *Biometrika*, Vol. 71, pp. 353-360.
- [5] Blum, J.R., and Susarla, V.(1980) : Maximal Deviation Theory of Density and Failure Rate Function Estimates Based on Censored Data. *In : Krishnaiah, P.K. (Ed.), Multivariate Analysis vol. 5. North-Holland, New York, pp. 213-222.*
- [6] Deheuvels, P. and Einmahl, John H. J. (2000) : Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. *Annals of Probability*, Vol. 28, No. 3, pp. 1301-1335.
- [7] Diehl, S. and Stute, W. (1988) : Kernel Density and Hazard Function Estimation in the Presence of Censoring. *Journal of Multivariate Analysis*, Vol. 25, pp. 299-310.
- [8] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) : Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator.

The Annals of Mathematical Statistics, Vol. 27, No. 3, pp. 642-669.

- [9] Földes, A., Rejtő, L. and Winter, B.B. (1981) : Strong Consistency Properties of Nonparametric Estimators for Randomly Censored Data,II : Estimation of density and failure rate. *Periodica Mathematica Hungarica, Vol. 12, No. 1, pp. 15-29.*
- [10] Fleming, T.R. and Harrington, D.P. (1991) : Counting processes and survival analysis. *New York : Wiley.*
- [11] Ferraty, F. and Vieu, P. (2006) : Nonparametric functional data analysis. *Theory and Practice. Springer , New York.*
- [12] Fermanian, J : Modèles de durées. *Cours ENSEE 3ème année.*
- [13] Hildenbrand, K . and Hildenbrand, W. (1986) : On the Mean Income Effect : a Data Analysis of the U.K. Family Expenditure Survey. In : Hildenbrand, W. and Mass-Collel, A. *Contributions to Matimatical Economics, in Honor of Gerard debreu. North-Holland, Amesterdam, pp. 247-268.*
- [14] Härdle, W., Marron, J.S. and Wand, M.P. (1990) : Bandwidth Choice for Density Derivatives. *Journal of the Royal Statistical Society. Series B (Methodological), Vol. 52, No. 1, pp. 223-232.*
- [15] Härdle, W. (1990) : Applied Nonparametric Regression. *Oxford University Press, Oxford.*
- [16] Hoeffding, W. (1963) : Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association , Vol. 58, pp. 13-30.*
- [17] Hubert, C : Modèles pour des durées de survie. *Cours de Statistiques, Université Paris 5.*
- [18] Kiefer, J. and Wolfowitz, J. (1958) : On The Deviations of the Empiric Distribution Function of Vector Chance Variables. *Transactions of the American Mathematical*

- Society, Vol.27, pp. 642-669.*
- [19] Karunamuni, R.J and Yang, S. (1991) : Weak and Strong Uniform Consistency Rates of Kernel Density Estimates for Randomly Censored Data. *The Canadian Journal of Statistics, Vol. 19, No. 4, pp. 349-359.*
- [20] Kaplan, E. L. and Meier, P. (1958) : Nonparametric Estimation From Incomplete Observation. *Journal of the American Statistical Association, Vol. 53, No. 282, pp. 457-481.*
- [21] Khardani, S., Lemdani, M. and Ould Saïd, E. (2012) : On the Strong Uniform Consistency of the Mode Estimator for Censored Time Series. *Metrika (2012), Vol. 75, pp. 229-241.*
- [22] Louani, D. (1998) : On the asymptotic normality of the kernel estimators of the density function and its derivatives under censoring. *Communications in Statistics -Theory and Methods, Vol. 27, No. 12, pp. 2909-2924.*
- [23] Lyndell-Bell, D. (1971) : A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomical Society, Vol. 155, pp. 95-118.*
- [24] Lo, S.H., Marck, Y.P. and Wang, J.L. (1989) : Density and Hazard Rate Estimation for Censored Data via Strong Representation of the Kaplan-Meier Estimator. *Probability Theory and Related Fields, Vol. 80, pp. 461-473.*
- [25] Mielniczuk, J. (1986) : Some Asymptotic Properties of Kernel Estimators of a Density Function in Case of Censored Data. *The Annals of Statistics, Vol. 14, No. 2, pp. 766-773.*
- [26] Müller, H.G (1988) : Nonparametric Analysis of Longitudinal Data. *Lecture Notes in Statistics, Vol. 46, Springer, New York.*

- [27] Natansan, I.P (1955) : Theory of Functions of a Real Variable 1 (Rev.Ed.). *Ungar, New York*.
- [28] Parzen, E. (1962) : On The Estimation of a Probability Density and Mode. *The Annals of Mathematical Statistics, Vol. 33, pp. 1065-1076*.
- [29] Rosenblatt, M. (1956) : Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics, Vol. 27, No. 3, pp. 832-837*.
- [30] Rudemo, M. (1982) : Empirical choice of histogram and kernel density estimators. *Scandinavian Journal of Statistics, Vol. 9, No. 2, pp. 65-78*.
- [31] Schuster, E. F. (1969) : Estimation of a Probability Density Functions and its Derivatives. *The Annals of Mathematical Statistics, Vol. 40, No. 4, pp. 1187-1195*.
- [32] Scott, D. W. and Terrell, G.R. (1987) : Biased and Unbiased Cross-Validation in Density Estimation. *Journal of the American Statistical Association, Vol. 82, No.400, pp. 1131-1146*.
- [33] Silverman, B. W. (1978) : Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives. *The Annals of Mathematical Statistics, Vol. 6, No. 1, pp. 177-184*.
- [34] Silverman, B. W. (1986) : Density Estimation for Statistics and Data Analysis. *Chapman and Hall, New York*.
- [35] Singh, R. S. (1977a) : Improvement on Some Known Nonparametric Uniformly Consistent Estimators of Derivatives of a Density. *The Annals of Statistics, Vol. 5, No. 2 , pp. 394-399*.
- [36] Singh, R. S. (1977b) : Applications of Estimators of a Density and its Derivatives to Certain Statistical Problems. *Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 3, pp. 357-363*.

- [37] Singh, R. S. (1979) : Mean Squared Errors of Estimates of a Density and Its Derivatives. *Biometrika*, Vol. 66, No. 1, pp. 177-180.

- [38] Shorack, G. and Wellner, J. (1986) : Empirical Processes With Applications to Statistics. *New York : Wiley*, PP.293-333.

- [39] Sheather, S. J and Jones, M.C.(1991) : A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B*, Vol.53, PP.683-690.

- [40] Stute, W. and Wang, J. L.(1993) : The strong Law Under random Censorship. *The Annals of Statistics*, Vol. 21, No. 3, pp. 1591-1607.

- [41] Stute, W.(1995) : The Central Limit Theorem Under Random Censorship. *The Annals of Statistics*, Vol. 23, No. 2, pp. 422-439.

- [42] Schwartz, S. C. (1967) : Estimation of a probability density of an orthogonal series. *The Annals of Mathematical Statistics*, Vol. 36, pp. 879-89.

- [43] Van Ryzin, J. (1969) : On strong consistency of density estimates. *The Annals of Mathematical Statistics*, Vol. 40, pp. 1765-1772.

- [44] Woodroffe, M. (1985) : Estimating a distribution function with truncated data. *The Annals of Statistics*, Vol. 13, pp. 163-177.