



Erreur quadratique asymptotique optimale pour les estimateurs à noyau de dérivées de Radon-Nikodym générales

Nadia Bensaïd, Pierre Jacob

► To cite this version:

Nadia Bensaïd, Pierre Jacob. Erreur quadratique asymptotique optimale pour les estimateurs à noyau de dérivées de Radon-Nikodym générales. Annales de l'ISUP, 1998, XXXXII (2-3), pp.3-14. <hal-03658855>

HAL Id: hal-03658855

<https://hal.science/hal-03658855v1>

Submitted on 4 May 2022

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

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



HAL Authorization

Erreur quadratique asymptotique optimale pour les estimateurs à noyau de dérivées de Radon-Nikodym générales

Nadia Bensaïd, Pierre Jacob

Laboratoire de Probabilités et Statistique,
 Université Montpellier II,

Place Eugène Bataillon, 34095 Montpellier Cedex 5

1 Introduction

Nous poursuivons dans cet article l'étude d'un problème d'inférence statistique relatif à une classe de processus ponctuels très généraux que nous appelons processus ponctuels composés (p.p.c.), dans un sens beaucoup moins restrictif que le sens habituel de "compound point processes", que l'on trouve dans les ouvrages de référence. Il s'agit pour nous de la classe des mesures aléatoires de Radon sur \mathbb{R}^p , discrètes et positives. On sait que (Kallenberg 1976 et Jacob 1979) l'on peut écrire une telle mesure aléatoire η sous la forme

$$\sum_{i=1}^N \alpha_i \delta_{X_i}, \text{ où } N \text{ est une variable aléatoire à valeurs dans } \mathbb{N} \cup \{+\infty\}, (\alpha_i) \text{ une suite de variables aléatoires positives, et } (X_i) \text{ une suite de variables aléatoires à valeurs dans } \mathbb{R}^p. \text{ A}$$

ce niveau de généralité, rien ne peut être dit de la loi conjointe du triplet $(N, (\alpha_i), (X_i))$, et d'ailleurs il n'y a pas d'unicité pour ce type de représentation. Evidemment, dans les applications, la situation est très différente. Les modèles de p.p.c. qu'on y rencontre

sont fournis directement sous la forme $\eta = \sum_{i=1}^N \alpha_i \delta_{X_i}$, à l'aide de choix bien précis du

triplet $(N, (\alpha_i), (X_i))$, selon les nécessités de la modélisation. Dans les constructions que nous présentons au paragraphe 2, on verra que les processus les plus élaborés proviennent de processus plus simples, de telle manière que les mesures moyennes de ceux-ci dominent les mesures moyennes de ceux-là. Les dérivées de Radon Nikodym des unes par rapport aux autres ont alors une signification spécifique à la modélisation envisagée, et on peut avoir envie de les estimer sur la base des réalisations des p.p.c.. Or, c'est justement la force du traitement statistique que nous faisons que de se passer de toute hypothèse relative à la loi conjointe des triplets tels que $(N, (\alpha_i), (X_i))$. Les propriétés ainsi obtenues, lorsqu'elles sont traduites dans le contexte souhaité, se révèlent être aussi performantes que celles qui ont pu être développées spécifiquement. On trouve déjà les prémisses de ces idées dans Mendes-Lopes (1984), Karr (1985), Saleh (1985). Cet article est plus spécialement consacré à la vitesse de convergence en moyenne quadratique de l'estimateur par la méthode du noyau, et complète essentiellement l'article de Jacob et Oliveira (1997). On verra qu'en dépit du cadre très général de notre étude, nous obtenons la même vitesse optimale qu'en estimation de la densité.

2 Quelques modèles

Rappelons qu'une mesure aléatoire est une variable aléatoire η à valeurs dans l'espace mesurable $(\mathcal{M}, B_{\mathcal{M}})$ des mesures de Radon positives sur \mathbb{R}^p (assimilées aux mesures boréliennes localement finies) muni de la tribu borélienne pour la topologie vague. Si les réalisations de η sont discrètes, on peut toujours représenter η sous la forme $\sum_{i=1}^N \alpha_i \delta_{X_i}$, indiquée dans l'introduction. De telles mesures aléatoires seront appelées processus ponctuels composés. Nous supposons toujours remplies, sous autre forme de procès, les conditions adéquates pour que les p.p.c. que nous manipulons possèdent des mesures moyennes $\mu = E(\eta)$, qui soient de Radon, c'est à dire telles que $\mu(A) = E\{\eta(A)\}$ soit fini pour tout borélien borné A de \mathbb{R}^p .

L'ingrédient de base pour élaborer des p.p.c. est simplement la mesure de dirac δ_X d'une variable aléatoire X à valeurs dans \mathbb{R}^p .

Si α est une variable aléatoire réelle (que nous supposons positive pour rester dans le cadre habituel des mesures positives), on construit un p.p.c. très élémentaire en posant $\eta = \alpha \delta_X$. La mesure moyenne μ de η est absolument continue par rapport à la mesure moyenne ν de δ_X (qui est en fait la loi de probabilité de X). Il est facile de vérifier que la dérivée de Radon-Nikodym $\frac{d\mu}{d\nu}$ n'est rien d'autre que l'espérance conditionnelle $E(\alpha/X = \cdot)$. Ainsi, l'estimation de cette densité, dans un cas aussi rudimentaire, pose déjà tout le problème de l'estimation de la régression.

Etant donné une suite (X_i) de variables aléatoires indépendantes et de même loi à valeurs dans \mathbb{R}^p , on définit pour tout n le processus empirique $\sum_{i=1}^n \delta_{X_i}$. Si N est une variable aléatoire de loi de Poisson, indépendante de (X_i) , on définit un processus ponctuel de Poisson fini en posant $\xi = \sum_{i=1}^N \delta_{X_i}$. Si (α_i) est une suite de variables aléatoires positives indépendantes et de même loi, indépendantes de ξ , on définit un processus ponctuel de Poisson composé en posant $\eta = \sum_{i=1}^N \alpha_i \delta_{X_i}$. L'idée de composition est prise ici dans un sens très restrictif. Alors $E(\eta) = \mu \ll \nu = E(\xi)$, et $\frac{d\mu}{d\nu}$ est constante et égale à $E(\alpha_1)$.

Un cas beaucoup plus intéressant est celui où, étant donné un processus $\xi = \sum_{i=1}^N \delta_{X_i}$, et une fonction mesurable p de \mathbb{R}^p dans $[0, 1]$, on considère une suite (α_i) de variables aléatoires de Bernoulli, indépendantes conditionnellement à ξ , et telles que, pour tout i , le paramètre de la loi de α_i soit $p(X_i)$. Le p.p.c. $\eta = \sum_{i=1}^N \alpha_i \delta_{X_i}$ est appelé "aminci de ξ " ou "thinned point process". Si ξ est un processus de Poisson, η est aussi un processus de Poisson. Dans tous les cas, $E(\eta) = \mu \ll \nu = E(\xi)$ et $\frac{d\mu}{d\nu}$ est tout simplement la fonction d'amincissement p .

Le cas des processus ponctuels marqués, ou "marked point processes", est susceptible de nombreuses interprétations concrètes, parmi lesquelles celle de points colorés, est

très évocatrice. Un tel processus peut s'écrire $\sum_{i=1}^N \delta_{(X_i, Y_i)}$ où les X_i sont des variables aléatoires à valeurs dans \mathbb{R}^p et les Y_i sont des variables aléatoires à valeurs dans un espace mesurable (S, \mathcal{S}) appelé espace des marques. Par exemple $E \in \mathcal{S}$ pourra représenter la couleur rouge. Si N est presque sûrement fini, $\xi = \sum_{i=1}^N \delta_{X_i}$ est aussi un processus ponctuel. On pose $\eta = \sum_{i=1}^N \alpha_i \delta_{X_i}$, où pour tout i , $\alpha_i = 1_{\{Y_i \in E\}}$, le processus ponctuel des points rouges. A nouveau, $E(\eta) = \mu \ll \nu = E(\xi)$ et $\frac{d\mu}{d\nu}$ représente la densité moyenne du rouge.

Les deux cas spéciaux de processus marqués qui suivent méritent un traitement à part: les marques y sont elles mêmes des répartitions ponctuelles.

Un processus ponctuel par grappe ou "cluster point process" est un processus ponctuel sur $\mathbb{R}^p \times \mathbb{R}^p$ de la forme $\zeta = \sum_{i=1}^N \sum_{j=1}^{N_i} \delta_{(X_i, Y_j^i)}$. On suppose que la projection $\gamma = \sum_{i=1}^N \sum_{j=1}^{N_i} \delta_{Y_j^i}$, est aussi un processus ponctuel, ce qui est vérifié par exemple si N et les (N_i) sont des variables aléatoires presque sûrement finies. Le processus ponctuel $\xi = \sum_{i=1}^N \delta_{X_i}$,

représente les "centres" de grappes et les processus ponctuels $\zeta_{X_i} = \sum_{j=1}^{N_i} \delta_{Y_j^i}$ représentent les grappes elles-mêmes. Pour préciser la loi de ζ , on se donne un noyau markovien $(\pi_x, x \in \mathbb{R}^p)$ de lois de processus ponctuels, de moyennes $(\mu_x, x \in \mathbb{R}^p)$, et on suppose que, conditionnellement à $\xi = \sum_{i=1}^n \delta_{x_i}$, $(\zeta_{x_1}, \dots, \zeta_{x_n})$ a pour loi $\pi_{x_1} \otimes \dots \otimes \pi_{x_n}$. Ayant fixé un borélien borné B de \mathbb{R}^p , on pose $\eta(A) = \zeta(A \times B)$ pour tout borélien borné A de \mathbb{R}^p . Alors $E(\eta) = \mu \ll \nu = E(\xi)$ et $\frac{d\mu}{d\nu}(x) = \mu_x(B)$, ν presque partout.

Les translations de processus ponctuels ou "Markov shifts", sont présentées dans (Daley et Vere-Jones 1988), comme un cas très spécial du précédent, où la grappe est réduite à un seul point, qui est interprétée comme une translation du point central. On considère donc un processus ponctuel ζ sur $\mathbb{R}^p \times \mathbb{R}^p$ de la forme $\zeta = \sum_{i=1}^N \delta_{(X_i, Y_i)}$, et, pour reprendre les notations du processus ponctuel par grappe, $\gamma = \sum_{i=1}^N \delta_{Y_i}$ et $\zeta_{X_i} = \delta_{Y_i}$. Le noyau markovien $(\pi_x, x \in \mathbb{R}^p)$ est ici réduit à un système de lois conditionnelles régulières sur \mathbb{R}^p , $(\mu_x, x \in \mathbb{R}^p)$, où, pour tout borélien borné B , $\mu_x(B)$ représente la probabilité pour que le point x soit envoyé dans B . On suppose que, conditionnellement à $\xi = \sum_{i=1}^n \delta_{x_i}$, $(\zeta_{x_1}, \dots, \zeta_{x_n})$ a pour loi $\pi_{x_1} \otimes \dots \otimes \pi_{x_n}$. Autrement dit, (Y_1, \dots, Y_n) a pour loi $\mu_{x_1} \otimes \dots \otimes \mu_{x_n}$. Soit B un borélien borné de \mathbb{R}^p , on pose $\eta(A) = \zeta(A \times B)$ pour tout borélien borné A de \mathbb{R}^p . Alors $E(\eta) = \mu \ll \nu = E(\xi)$ et $\frac{d\mu}{d\nu}(x) = \mu_x(B) = P(Y \in B/X = x)$.

3 Position du problème d'estimation

La situation suivante est constante dans les modèles que nous avons évoqués: on dispose de deux p.p.c. η et ξ , de mesures moyennes μ et ν , respectivement, et μ a une densité relativement à ν . On souhaite estimer cette densité.

Lorsqu'on a des processus ponctuels qui ont presque sûrement un nombre fini de points, on ne peut pas espérer obtenir des résultats asymptotiques en regardant une seule réalisation et en faisant grossir le domaine d'observation. Or, le fait d'avoir presque sûrement un nombre fini de points est dans beaucoup d'exemples une condition inévitable, pour l'existence même des processus que nous voulons étudier. De plus, même dans le cas où une infinité de points seraient disponibles, on n'a pas en général des propriétés de stationnarité et d'ergodicité. D'ailleurs, le cas où les mesures moyennes sont, à une constante multiplicative près la mesure de Lebesgue, perd beaucoup d'intérêt par rapport au cas général.

On doit donc se résoudre à supposer que l'on dispose d'un n -uple d'observations $(\eta_1, \xi_1), \dots, (\eta_n, \xi_n)$ de processus ponctuels de même loi que (η, ξ) . Le cas où ces couples sont indépendants est peu réaliste, mais il faut bien commencer par là. On trouvera dans Jacob et Oliveira (1995 et 1997) les premiers traitements généraux de ces questions. Des traitements plus partiels ont précédé (Mendes-Lopes 1984, Karr 1985, Saleh 1985, Jacob et Niéré 1990, Bensaïd 1997). D'autres auteurs ont depuis, envisagés des modèles beaucoup plus réalistes, où les suites de processus ponctuels sont mélangées (Bensaïd et Fabre 1998) ou associées (Ferrieux 1996). Pour en revenir au cas i.i.d., l'étude de la vitesse de convergence en moyenne quadratique de l'estimateur à noyau n'avait pas été faite par Jacob et Oliveira. Cet article y pourvoit. Il est aussi l'occasion de revisiter quelques méthodes classiques en estimation de la régression, et permet de séparer dans celles-ci le particulier du général.

Dans tout ce qui suit, (η, ξ) désigne un couple de p.p.c. sur \mathbb{R}^p , tel que $\mu = E(\eta)$ soit de densité φ par rapport à $\nu = E(\xi)$. La notion de mesure moyenne a été rappelée au paragraphe 2. Pour étudier la convergence en moyenne quadratique d'estimateurs de φ , nous aurons évidemment à supposer l'existence de moments d'ordre supérieur. Soit (ζ^1, ζ^2) l'un quelconque des couples de p.p.c. (ξ, ξ) , (ξ, η) , (η, ξ) , (η, η) . La mesure moment d'ordre deux $E(\zeta^1 \otimes \zeta^2)$ est définie sur la tribu borélienne de $\mathbb{R}^p \times \mathbb{R}^p$, dès que pour tout pavé mesurable borné $A \times B$:

$$E(\zeta^1(A)\zeta^2(B)) < +\infty$$

On pose alors $E(\zeta^1 \otimes \zeta^2)(A \times B) = E(\zeta^1(A)\zeta^2(B))$, et les théorèmes de prolongement habituels font de $E(\zeta^1 \otimes \zeta^2)$ une mesure localement finie sur $\mathbb{R}^p \times \mathbb{R}^p$. Il est important de noter qu'une telle mesure n'est généralement pas à densité par rapport à la mesure de Lebesgue $\lambda \otimes \lambda$ de $\mathbb{R}^p \times \mathbb{R}^p$. Par exemple, si ξ est un processus ponctuel obtenu par effacement de points du processus ponctuel η , le produit $\eta \otimes \xi$ comporte sur la diagonale de $\mathbb{R}^p \times \mathbb{R}^p$ tous les couples de points de ξ . Cette remarque aidera à comprendre l'hypothèse H4 que nous ferons plus loin. En attendant, nous notons $\Delta = \{(x, x), x \in \mathbb{R}^p\}$ la diagonale de $\mathbb{R}^p \times \mathbb{R}^p$ et λ^* la mesure de Lebesgue sur Δ .

Le noyau utilisé pour construire un estimateur de φ est une densité de probabilité K ,

bornée sur \mathbb{R}^p . Etant donné une suite de nombres positifs (h_n) qui décroît vers zéro quand n tend vers l'infini, on pose pour tout x de \mathbb{R}^p :

$$K_n(x) = h_n^{-p} K(h_n^{-1}x).$$

Etant donné une suite d'observations $(\eta_i, \xi_i)_{i \geq 1}$ de même loi que (η, ξ) , on pose pour tout entier $n \geq 1$ et tout $x \in \mathbb{R}^p$:

$$g_n(x) = \sum_{i=1}^n \eta_i \star K_n(x)/n \text{ et } f_n(x) = \sum_{i=1}^n \xi_i \star K_n(x)/n,$$

où \star désigne l'opérateur de convolution.

L'estimateur de φ , par la méthode du noyau est alors défini par:

$$\varphi_n(x) = \frac{g_n(x)}{f_n(x)}, \quad x \in \mathbb{R}^p.$$

Bien entendu, dans le cas particulier évoqué au paragraphe 2, $\xi = \delta_X$ et $\eta = Y\delta_X$, on reconnaît en φ_n l'estimateur de Nadaraya-Watson de la régression $E(Y/X = \cdot)$.

4 Erreur quadratique asymptotique optimale

Dans tout ce qui suit on fait les hypothèses suivantes:

H1: K est une densité de probabilité de \mathbb{R}^p dans \mathbb{R} symétrique, bornée et à support compact S_K .

H2: $h_n \rightarrow 0$ et $nh_n^p \rightarrow +\infty$ quand $n \rightarrow +\infty$.

H3: les mesures moyennes μ et ν ont des densités par rapport à la mesure de Lebesgue λ de \mathbb{R}^p . Ces densités ont des versions respectives g et f , strictement positives, 2 fois différentiables; et leurs dérivées partielles du second ordre sont continues et bornées.

Les trois premières hypothèses sont familières en estimation fonctionnelle. Les deux hypothèses suivantes relèvent plus spécifiquement de la théorie des mesures aléatoires.

H4: on suppose que, pour tout choix de (ζ^1, ζ^2) , $E(\zeta^1 \otimes \zeta^2)$ existe sur $\mathbb{R}^p \times \mathbb{R}^p$ et se décompose en la somme de deux mesures $\gamma_{\zeta^1, \zeta^2}^1 + \gamma_{\zeta^1, \zeta^2}^2$, la première ayant une densité bornée g_{ζ^1, ζ^2}^1 par rapport à la mesure de Lebesgue $\lambda \otimes \lambda$ de $\mathbb{R}^p \times \mathbb{R}^p$, la seconde ayant une densité continue et bornée g_{ζ^1, ζ^2}^2 par rapport à la mesure de Lebesgue λ^* de Δ .

H5: pour tout $i = 3, 4$, il existe une fonction positive C_i définie sur \mathbb{R}^p telle que pour toute suite $I_n \searrow \{s\}$,

$$E(\zeta^i(I_n)) / E(\zeta^{i-1}(I_n)) \longrightarrow C_i(s).$$

Le théorème suivant généralise le théorème V.2 de Bosq et Lecoutre (1987).

Théorème 4.1 Si H1, H2, H3 et H4 sont vérifiées alors:

$$E(f_n(x) - f(x))^2 = (h_n^4/4) \chi_{\nu, p}^2(x) + (R/nh_n^p)(g_{\xi, \xi}^2(x, x)) + o(h_n^4 + (1/nh_n^p)),$$

$$\begin{aligned}
E(g_n(x) - g(x))^2 &= (h_n^4/4) \chi_{\mu,p}^2(x) + (R/nh_n^p)(g_{\eta,\eta}^2(x, x)) + o(h_n^4 + (1/nh_n^p)), \\
E((f_n(x) - f(x))(g_n(x) - g(x))) &= (h_n^4/4) \chi_{\mu,p}(x) \chi_{\nu,p}(x) + (R/nh_n^p)(g_{\eta,\xi}^2(x, x)) \\
&\quad + o(h_n^4 + (1/nh_n^p)).
\end{aligned}$$

$$\begin{aligned}
Où \chi_{\nu,p}(x) &= \sum_{1 \leq i, j \leq p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int t_i t_j K(t) dt, \\
\chi_{\mu,p}(x) &= \sum_{1 \leq i, j \leq p} \frac{\partial^2 g}{\partial x_i \partial x_j}(x) \int t_i t_j K(t) dt \quad \text{et} \quad R = (\int K^2(s) \lambda^*(ds))^{-1}.
\end{aligned}$$

Démonstration: $E(f_n(x) - f(x))^2 = E(f_n(x) - Ef_n(x))^2 + (Ef_n(x) - f(x))^2$.

$$\begin{aligned}
Ef_n(x) &= E(\xi \star K_n(x)) = (1/h_n^p) \int K\left(\frac{x-u}{h_n}\right) \frac{d\nu}{d\lambda}(u) \lambda(du) \\
&= (1/h_n^p) \int K\left(\frac{x-u}{h_n}\right) f(u) \lambda(du).
\end{aligned}$$

$$Ef_n(x) - f(x) = \int K(v)(f(x - h_n v) - f(x)) dv.$$

En développant f en série de Taylor on obtient:

$$\begin{aligned}
Ef_n(x) - f(x) &= -h_n \sum_{i=1}^p \frac{\partial f}{\partial x_i}(x) \int v_i K(v) dv \\
&\quad + (h_n^2/2) \sum_{1 \leq i, j \leq p} \int \frac{\partial^2 f}{\partial x_i \partial x_j}(x - \theta_n h_n v) v_i v_j K(v) dv,
\end{aligned}$$

où $\theta_n \in]0, 1[$.

Sous les hypothèses H1 et H3 on déduit, en utilisant le théorème de la convergence dominée, que:

$$Ef_n(x) - f(x) = (h_n^2/2) X_{\nu,p}(x) + o(h_n^2). \quad (1)$$

Maintenant,

$$V(f_n(x)) = (1/n)E(\xi \star K_n(x))^2 - (1/n)E^2(\xi \star K_n(x)).$$

$$\begin{aligned}
(1/n)E(\xi \star K_n(x))^2 &= (1/nh_n^{2p}) \int_{\mathbb{R}^{2p}} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) g_{\xi,\xi}^1(u, v) \lambda \otimes \lambda(du, dv) \\
&\quad + (1/nh_n^{2p}) \int_{\mathbb{R}^p} K^2\left(\frac{x-u}{h_n}\right) g_{\xi,\xi}^2(u, u) \lambda^*(du) \\
&= (1/n) \int_{\mathbb{R}^{2p}} K(t) K(z) g_{\xi,\xi}^1(x - h_n t, x - h_n z) \lambda \otimes \lambda(dt, dz) \\
&\quad + (1/nh_n^p) \int_{\mathbb{R}^p} K^2(t) g_{\xi,\xi}^2(x - h_n t, x - h_n t) \lambda^*(dt).
\end{aligned}$$

En utilisant le théorème de la convergence dominée, on obtient sous H4:

$$(1/n)E(\xi \star K_n(x))^2 = (1/nh_n^p) g_{\xi,\xi}^2(x, x) \int_{\mathbb{R}^p} K^2(t) \lambda^*(dt) + o(1/(nh_n^p)), \quad (2)$$

par conséquent

$$V f_n(x) = (1/n h_n^p) g_{\xi, \xi}^2(x, x) \int_{\mathbb{R}^p} K^2(t) \lambda^*(dt) + o(1/(n h_n^p)) + O(1/n), \quad (3)$$

ce qui donne le résultat.

On traite $E(g_n(x) - g(x))^2$ et $E((g_n(x) - g(x))(f_n(x) - f(x)))$ exactement de la même manière que $E(f_n(x) - f(x))^2$.

Corollaire 4.1 Si H_4 est vérifiée et si $h_n \rightarrow 0$, alors pour tout compact S_K de \mathbb{R}^p , $(1/h_n^p)(E\xi(x - h_n S_K)^2)$ est une suite convergente.

Démonstration:

Il suffit d'appliquer la démonstration de (2) à un noyau K constant sur S_K .

Remarque 4.1 On retrouve les résultats du théorème 4.1 si on remplace l'hypothèse H_1 par:

H'1: K est une densité de probabilité de \mathbb{R}^p dans \mathbb{R} symétrique, bornée, et telle que $\lim_{\|y\| \rightarrow \infty} \|y\|^p K(y) = 0$ et $\int \|y\|^2 K(y) dy < +\infty$.

Théorème 4.2 a) Si H_1, H_2, H_3, H_4 et H_5 sont vérifiées alors

$$E(f_n(x) - f(x))^4 = O(h_n^8 + \frac{h_n^4}{n h_n^p} + \frac{1}{(n h_n^p)^2}).$$

b) Si, en particulier $h_n = c n^{\frac{-1}{p+4}}$, $c > 0$, alors:

$$E(f_n(x) - f(x))^4 = O(n^{\frac{-8}{p+4}}).$$

Démonstration:

$$\begin{aligned} E(f_n(x) - f(x))^4 &= (E f_n(x) - f(x))^4 + 6E(f_n(x) - E f_n(x))^2 (E f_n(x) - f(x))^2 \\ &\quad + 4E(f_n(x) - E f_n(x))^3 (E f_n(x) - f(x)) + E(f_n(x) - E f_n(x))^4. \end{aligned}$$

Par (1) et (3), $(E f_n(x) - f(x))^4$ est un $O(h_n^8)$ et $E(f_n(x) - E f_n(x))^2 (E f_n(x) - f(x))^2$ est un $O(\frac{h_n^4}{n h_n^p})$.

Maintenant,

$$\begin{aligned} E(f_n(x) - E f_n(x))^3 &= (1/n^2) E(\xi \star K_n(x) - \nu \star K_n(x))^3 \\ &= (1/n^2) \{E(\xi \star K_n(x))^3 - 3E(\xi \star K_n(x))^2 (\nu \star K_n(x)) + 2(\nu \star K_n(x))^3\}. \end{aligned}$$

En posant $M = \sup_x K(x)$ et $K^*(x) = M^{-1} K(x)$,

$$\begin{aligned} E(\xi \star K_n(x))^3 &= E(1/(h_n^p) \int K(\frac{x-y}{h_n}) \xi(dy))^3 \\ &= E(M/(h_n^p) \int K^*(\frac{x-y}{h_n}) \xi(dy))^3 \\ &\leq M^3 / (h_n^{3p}) E(\xi(x - h_n S_K))^3. \end{aligned}$$

Comme pour n assez grand, $x - h_n S_K \searrow \{x\}$, $E(\xi(x - h_n S_K))^3 / E(\xi(x - h_n S_K))^2$ converge, sous H_5 , vers $C_3(x)$ et par conséquent on peut déduire par le corollaire 4.1 que

$$E(\xi \star K_n(x))^3 = O(h_n^{-2p}).$$

Finalement par (1) et (2) et on obtient:

$$E(f_n(x) - Ef_n(x))^3(Ef_n(x) - f(x)) = O(h_n^2/(nh_n^p)^2).$$

Il reste à examiner $E(f_n(x) - Ef_n(x))^4$. Un calcul de routine montre que:

$$\begin{aligned} E(f_n(x) - Ef_n(x))^4 &= (1/n^3)\{E(\xi \star K_n(x))^4 - 4(\nu \star K_n(x))E(\xi \star K_n(x))^3 \\ &\quad + 6(\nu \star K_n(x))^2E(\xi \star K_n(x))^2 - 4(\nu \star K_n(x))^4\} \\ &\quad + (3(n-1)/n^3)E^2(\xi \star K_n(x) - \nu \star K_n(x))^2. \end{aligned}$$

En procédant de la même manière que pour $E(\xi \star K_n(x))^3$, à l'aide de l'hypothèse H5, on montre que:

$$E(\xi \star K_n(x))^4 = O(h_n^{-3p}).$$

Et en utilisant les résultats intermédiaires (1) et (2) on déduit que:

$$E(f_n(x) - Ef_n(x))^4 = O(1/(nh_n^p)^2).$$

Ainsi,

$$\begin{aligned} E(f_n(x) - f(x))^4 &= O(h_n^8 + \frac{h_n^4}{nh_n^p} + \frac{h_n^2}{(nh_n^p)^2} + \frac{1}{(nh_n^p)^2}) \\ &= O(h_n^8 + \frac{h_n^4}{nh_n^p} + \frac{1}{(nh_n^p)^2}), \end{aligned}$$

d'où le résultat.

Remarque 4.2 On obtient les mêmes résultats que dans le théorème 4.2 si on remplace l'hypothèse H1 par H'1 et H5 par des hypothèses H'4 et H''4 du même type que H4 sur l'existence et la décomposition des moments $E(\zeta^1 \otimes \zeta^2 \otimes \zeta^3)$ et $E(\zeta^1 \otimes \zeta^2 \otimes \zeta^3 \otimes \zeta^4)$. Pour le voir, il suffit de reprendre les étapes de la démonstration de (2), où on a montré que $E(\xi \star K_n(x))^2 = O(h_n^{-p})$, pour déduire de la même manière $E(\xi \star K_n(x))^3 = O(h_n^{-2p})$ et $E(\xi \star K_n(x))^4 = O(h_n^{-3p})$.

Remarque 4.3 On peut facilement vérifier qu'un processus de Poisson vérifie l'hypothèse H5 et que pour $i = 3, 4$, $C_i(s) = 1$.

Remarque 4.4 L'hypothèse H5 peut être remplacée par une hypothèse du type Cramer à l'ordre 4:

H'5 : Il existe une constante C telle que pour tout $I \subset B$, B compact de \mathbb{R}^p ,

$$E(\zeta^4(I)) \leq CE(\zeta^2(I)),$$

dans ce cas les résultats des théorèmes 4.2 et 4.3 restent valables pour tout $x \in \overset{\circ}{B}$, l'intérieur du compact B .

Pour démontrer le théorème donnant l'erreur quadratique asymptotique optimale nous aurons besoin d'une variante d'un lemme de Jacob et Niéré (1990):

Lemme 4.1 Soient X et Y des variables positives. Pour tout ϵ tel que $0 < \epsilon E(Y)/E(X) < 1$, on a:

$$\left\{ \left| \frac{X}{Y} - \frac{E(X)}{E(Y)} \right| > \epsilon \right\} \subset \left\{ \left| \frac{X}{E(X)} - 1 \right| > \frac{\epsilon E(Y)}{4 E(X)} \right\} \cup \left\{ \left| \frac{Y}{E(Y)} - 1 \right| > \frac{\epsilon E(Y)}{4 E(X)} \right\}.$$

Théorème 4.3 Si $H1, H3, H4$, et $H5$ sont vérifiées, si $h_n = c n^{\frac{-1}{p+4}}$, $c > 0$, alors pour tout x vérifiant $E(\varphi_n^4(x) 1_{\{\varphi_n(x) > n^{1/2(p+4)}\}}) = o(1)$ on a :

$$E(\varphi_n(x) - \varphi(x))^2 = n^{\frac{-4}{p+4}} \{D_{\mu, \nu, p}(x) + o(1)\},$$

$$\text{où } D_{\mu, \nu, p}(x) = \frac{c^4}{4f^2(x)} \{\varphi(x) \chi_{\nu, p}(x) - \chi_{\mu, p}(x)\}^2 \\ + \frac{R}{c^p f^2(x)} \{\varphi^2(x) g_{\xi, \xi}^2(x, x) - 2\varphi(x) g_{\xi, \eta}^2(x, x) + g_{\eta, \eta}^2(x, x)\}.$$

Démonstration:

Comme dans Bosq et Cheze (1993), on décompose $E(\varphi_n(x) - \varphi(x))^2$ de la manière suivante:

$$E(\varphi_n(x) - \varphi(x))^2 = \frac{\varphi^2(x)}{f^2(x)} E(f_n(x) - f(x))^2 + \frac{1}{f^2(x)} E(g_n(x) - g(x))^2 - \frac{2\varphi(x)}{f^2(x)} E(g_n(x) - g(x))(f_n(x) - f(x)) + \frac{1}{f^2(x)} E\{(\varphi_n^2(x) - \varphi^2(x))(f_n(x) - f(x))^2\} - \frac{2}{f^2(x)} E\{(\varphi_n(x) - \varphi(x))(f_n(x) - f(x))(g_n(x) - g(x))\}. \quad (4)$$

Les trois premiers termes de droite de (4) sont traités dans le théorème 4.1, il reste à montrer que les deux derniers sont des $o(n^{\frac{-4}{p+4}})$.

On pose

$$\tilde{\varphi}_n(x) = \frac{E g_n(x)}{E f_n(x)}.$$

$$E\{(\varphi_n^2(x) - \varphi^2(x))(f_n(x) - f(x))^2\} = E\{(\varphi_n^2(x) - \tilde{\varphi}_n^2(x))(f_n(x) - f(x))^2\} \\ + (\tilde{\varphi}_n^2(x) - \varphi^2(x)) E(f_n(x) - f(x))^2.$$

Par (1) et (3) on a :

$$(\tilde{\varphi}_n^2(x) - \varphi^2(x)) E(f_n(x) - f(x))^2 = o(n^{\frac{-4}{p+4}}).$$

Il nous reste à examiner $E\{(\varphi_n^2(x) - \tilde{\varphi}_n^2(x))(f_n(x) - f(x))^2\}$.

$$E\{(\varphi_n^2(x) - \tilde{\varphi}_n^2(x))(f_n(x) - f(x))^2\} \\ \leq (\alpha_n + \tilde{\varphi}_n(x)) E\{|\varphi_n(x) - \tilde{\varphi}_n(x)| (f_n(x) - f(x))^2 1_{\{\varphi_n(x) \leq \alpha_n\}} 1_{\{|\varphi_n(x) - \tilde{\varphi}_n(x)| \leq \epsilon_n\}}\} \\ + (\alpha_n + \tilde{\varphi}_n(x)) E\{|\varphi_n(x) - \tilde{\varphi}_n(x)| (f_n(x) - f(x))^2 1_{\{\varphi_n(x) \leq \alpha_n, |\varphi_n(x) - \tilde{\varphi}_n(x)| > \epsilon_n\}}\} \\ + E\{(\varphi_n^2(x) - \tilde{\varphi}_n^2(x))(f_n(x) - f(x))^2 1_{\{\varphi_n(x) > \alpha_n\}}\}. \\ = A_n + B_n + C_n.$$

Si on suppose que $\epsilon_n \rightarrow 0$, $\alpha_n \rightarrow +\infty$ et $\alpha_n \epsilon_n = o(1)$ alors comme $A_n \leq (\alpha_n + \tilde{\varphi}_n(x)) \epsilon_n E(f_n(x) - f(x))^2$ on peut déduire par (1) et (3) que $A_n = o(n^{\frac{-4}{p+4}})$.

Maintenant,

$$B_n \leq (\alpha_n + \tilde{\varphi}_n(x))^2 E\{(f_n(x) - f(x))^2 1_{\{|\varphi_n(x) - \tilde{\varphi}_n(x)| > \epsilon_n\}}\} \\ \leq (\alpha_n + \tilde{\varphi}_n(x))^2 E^{\frac{1}{2}}(f_n(x) - f(x))^4 P^{\frac{1}{2}}\{|\varphi_n(x) - \tilde{\varphi}_n(x)| > \epsilon_n\}.$$

En utilisant le lemme 4.1, on obtient pour n assez grand,

$$P\{|\varphi_n(x) - \tilde{\varphi}_n(x)| > \epsilon_n\} \leq P\{|g_n(x) - Eg_n(x)| > \frac{\epsilon_n}{4} Ef_n(x)\} \\ + P\{|f_n(x) - Ef_n(x)| > \frac{\epsilon_n}{4} \frac{E^2 f_n(x)}{Ef_n(x)}\}.$$

Et grâce au théorème 4.1 on peut déduire que:

$$B_n = O((\alpha_n)^2 E^{\frac{1}{2}}(f_n(x) - f(x))^4 P^{\frac{1}{2}}\{|f_n(x) - Ef_n(x)| > \epsilon_n\}).$$

Il suffit donc de montrer que $\alpha_n^2 P^{\frac{1}{2}}\{|f_n(x) - Ef_n(x)| > \epsilon_n\} = o(1)$ pour déduire, par le théorème 4.2, que $B_n = o(n^{\frac{-4}{p+4}})$.

Comme,

$$\alpha_n^2 P^{\frac{1}{2}}\{|f_n(x) - Ef_n(x)| > \epsilon_n\} \leq \frac{\alpha_n^2}{\epsilon_n^2} E^{\frac{1}{2}}(f_n(x) - Ef_n(x))^4,$$

il suffit de prendre $\epsilon_n = n^{\frac{-1}{p+4}}$ et $\alpha_n = n^{\frac{1}{2(p+4)}}$ pour avoir par le théorème 4.2 le résultat souhaité.

Il reste enfin à montrer que $C_n = o(n^{\frac{-4}{p+4}})$.

$$C_n = E\{(\varphi_n^2(x) - \tilde{\varphi}_n^2(x))(f_n(x) - f(x))^2 1_{\{\varphi_n(x) > n^{\frac{1}{2(p+4)}}\}}\} \\ \leq E^{\frac{1}{2}}(f_n(x) - f(x))^4 E^{\frac{1}{2}}(\varphi_n^4(x) 1_{\{\varphi_n(x) > n^{\frac{1}{2(p+4)}}\}}) \\ + E^{\frac{1}{2}}(f_n(x) - f(x))^4 \tilde{\varphi}_n^2(x) P^{\frac{1}{2}}(1_{\{\varphi_n(x) > n^{\frac{1}{2(p+4)}}\}}).$$

Comme $E(\varphi_n^4(x) 1_{\{\varphi_n(x) > n^{\frac{1}{2(p+4)}}\}}) = o(1)$, on conclut grâce au (1) et le théorème 4.2.

Il nous reste maintenant à étudier le cinquième terme de droite de (4) pour cela on peut remarquer que:

$$E\{(\varphi_n(x) - \varphi(x))(g_n(x) - g(x))(f_n(x) - f(x))\} \\ \leq E^{\frac{1}{2}}\{(\varphi_n(x) - \varphi(x))^2 (f_n(x) - f(x))^2\} E^{\frac{1}{2}}(g_n(x) - g(x))^2.$$

On montre que $E\{(\varphi_n(x) - \varphi(x))^2 (f_n(x) - f(x))^2\} = o(n^{\frac{-4}{p+4}})$ et ceci en reprenant la même technique que celle utilisée pour le terme $E\{(\varphi_n^2(x) - \varphi^2(x))(f_n(x) - f(x))^2\}$ puis on conclut par le théorème 4.1.

Remarque 4.5 La condition $E(\varphi_n^4(x) 1_{\{\varphi_n(x) > n^{1/2(p+4)}\}}) = o(1)$ peut paraître contraignante, alors qu'elle est vérifiée dès que par exemple $E(\varphi_n^5(x)) = O(1)$. D'autre part, si on se contente de la condition $E(\varphi_n^4(x)) = O(1)$, alors évidemment $E(\varphi_n^4(x) 1_{\{\varphi_n(x) > n^{1/2(p+4)}\}}) = O(1)$. Nous obtenons dans ce dernier cas le corollaire suivant, qui est légèrement plus faible que le théorème 4.3.

Corollaire 4.2 Si $H1, H2, H3, H4$, et $H5$ sont vérifiées, alors pour tout x vérifiant $E(\varphi_n^4(x) 1_{\{\varphi_n(x) > n^{1/2(p+4)}\}}) = O(1)$, on a:

$$\varphi_n(x) = O(n^{\frac{-4}{p+4}}).$$

Démonstration:

La démonstration est exactement la même que celle du théorème 4.3, sauf que la condition $E(\varphi_n^4(x)1_{\{\varphi_n(x) > n^{1/2(p+4)}\}}) = O(1)$, entraîne que les deux derniers termes de droite de (4) sont des $O(n^{\frac{-4}{p+4}})$.

Bibliographie

- Bensaïd, N. (1997). Non parametric inference for thinned point process. *Statist. Probab. Lett.*, 33, 253-258.
- Bensaïd, N., et Fabre, J.P. (1998). Convergence de l'estimateur à noyau de dérivées de Radon-Nikodym générale dans le cas mélangeant. *Revue Canadienne de Statistique*. (A paraître).
- Bosq, D., et Lecoutre, J.P. (1987). *Théorie de l'estimation fonctionnelle*. Economica, Paris.
- Bosq, D. et Cheze, N. (1993). Erreur quadratique asymptotique optimale de l'estimateur non paramétrique de la régression pour des observations discrétisées d'un processus stationnaire à temps continu. *C.R.Acad.Sci., Paris*, 317, Série I, 891-894.
- Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Lecture Notes in Statistics. Springer-Verlag, New York.
- Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrsch. Verw. Geb.*, 66, 441-460.
- Daley, D.J., Vere-Jones, D. (1988). *An introduction to the theory of point processes*. Springer Series in Statistics.
- Ferrieux, D. (1996). Estimation de densités de mesures moyennes de processus ponctuels associés. Thèse de l'Université de Montpellier II.
- Jacob, P. (1979). Sur diverses convergences stochastiques des processus ponctuels composés. *C.R.Acad.Sci., Paris*, 288, Série A, 355-357.
- Jacob, P., et Niéré, L. (1990). Contribution à l'estimation des lois de Palm d'une mesure aléatoire. *Pub. Inst. Stat. Univ. Paris*, XXXV, No.2. 39-49.
- Jacob, P., et Mendes-Lopes, N. (1991). Un théorème central limite fonctionnel pour un estimateur d'un noyau de transition. *Port. Math.*, 48, No.2.
- Jacob, P., et Oliveira, P.E. (1995). A general approach to non parametric histogram estimation. *Statistics*, 27, 73-92.
- Jacob, P., et Oliveira, P.E. (1997). Kernel estimators of general Radon-Nikodym derivatives. *Statistics*, 30, 25-46.
- Kallenberg, O. (1976). *Random Measures*. Academic Press.
- Karr, A. (1985). Inference for thinned point processes with applications to Cox process. *J. Multivariate Anal.*, 16, 368-392.
- Karr, A. (1986). *Point processes and their statistical inference*. Marcel Dekker, New York.

- Karr, A. (1987). Estimation of Palm measures of stationary point processes. *Prob. Theory Related Fields*, 74, No.1., 55-69.
- Karr, A. (1988). *Palm distributions of point processes and their applications to statistical inference, statistical inference from stochastic processes*. Contemp. Math. 87, Amer. Math. Soc., Providence.
- Mendes-Lopes, N. (1984). Convergence et optimisation d'un estimateur de la répartition locale des couleurs d'un processus ponctuel chromatique. *Pub. Inst. Stat. Univ. Paris*, XXIX, No.2. 49-68.
- Niére, L. (1987). Estimation des distributions de Palm. Mesure intensité conditionnelle. Thèse de 3^{ème} cycle, Lille.
- Reiss, R.-D. (1993). *A Course on Point Processes*. Springer-Verlag, New York.
- Saleh, S. (1985). Estimation de la distribution locale moyenne d'une mesure composite aléatoire. *C. R. Acad. Sci. Paris*, 301, Série 1, 10, 545-548.