Estimation adaptative de la densité par projection dans un modèle semiparamétrique à rapport de densités

Jean-Baptiste Aubin

e-mail: jean-baptiste.aubin@utc.fr
and

Samuela Leoni-Aubin

e-mail: samuela.leoni.aubin@gmail.com

Abstract:
Nous disposons de m échantillons pour lesquels les rapports des $m-1$ premières densités de probabilité par rapport à la $m-$ième sont de forme paramétrique connue. Ce modèle semiparamétrique apparaît naturellement notamment dans le modèle de régression logistique multinomiale. Nous introduisons de nouveaux estimateurs par projection adaptatifs semiparamétriques des m densités à partir des observations combinées de tous les échantillons. Nous établissons des vitesses de convergence suroptimales pour l'erreur quadratique intégrée dans un sous-ensemble dense dans l'ensemble des densités possibles.

Keywords and phrases: données "case-control", vraisemblance semiparamétrique empirique, estimation de la densité par projection, vraisemblance semiparamétrique empirique, indice de troncature adaptatif.

1. Introduction

Nous considérons m échantillons aléatoires indépendants $X_{i1}, \ldots, X_{in_i}$, $i = 1, \ldots, m$ de densités de probabilité respectives $g_i(x) = \mathrm{d}G_i(x)$, $i = 1, \ldots, m$. Soit le modèle semiparamétrique à rapport de densités suivant:

$$g_i(x) = w(x, \theta_i) g_m(x), \qquad i = 1, \ldots, m-1, \tag{1}$$

où $w(x, \theta_i) := \exp\{\theta_{1,i} + s(x)\theta_{2,i}\}$ est une fonction connue, positive et bornée et $\theta_i := \left[\theta_{1,i}, \theta_{2,i}^t\right]^t$, $i = 1, \ldots, m-1$ est un vecteur de paramètres de dimension finie d. Le support commun aux lois G_i peut être connu ou inconnu, discret ou continu. Les m densités sont supposées inconnues mais sont reliées les unes aux autres par leur rapport qui est, lui, connu.
Le modèle à rapport de densités généralise certains modèles (régression logistique multinomiale, analyse de la variance "normal-based one-way" ou encore modèle "biased sampling"). Des applications de ce modèle existent dans des domaines variés tels que l'économétrie, la biostatistique, ou en météorologie. Pour plus de détails, on se reportera à [1] et à sa bibliographie.

Par ailleurs, des problèmes d'inférence pour les paramètres du modèle (1) dans le cas $m = 2$ ont été étudiés par [18], [13] ou encore [12].
Pour le problème de l'estimation des densités dans le modèle (1), deux démarches ont été envisagées.
La première méthode, introduite par [9], consiste à estimer g_m par un estimateur à noyau puis à en déduire $\widehat{\theta}$.
De récents travaux ([8], [10], [1] et [20]) ont adopté une approche radicalement différente en utilisant un modèle à rapport de densités pour m échantillons ([10] et [1]) ou à deux échantillons ([8] et [20]).
La première étape consiste en l'estimation des paramètres de dimension finie en maximisant une fonction de vraisemblance semiparamétrique, la seconde, en l'obtention d'un estimateur de la fonction de répartition inconnue par maximum de vraisemblance semiparamétrique en posant des poids sur toutes les observations. Un lissage (utilisant un noyau pour [8], [10] et [20] ou une méthode de projection pour [1]) mène alors à une nouvelle estimation de la densité.
Plus spécifiquement, [10] a montré que dans le modèle à rapport de densités (1), les observations combinées mènent à des estimateurs à noyau des distributions inconnues asymptotiquement plus efficaces, dans le sens où ces estimateurs, s'ils ont un biais identique à celui obtenu par les estimateurs classiques à noyau, ont une variance plus faible.
Des résultats analogues ont été démontrés par [1] dans le cas d'un lissage par projection. De plus, une amélioration par rapport aux estimateurs semi-paramétriques à noyau a été constatée dans le cas où la base de projection est convenablement choisie, c'est-à-dire dans le cas où les coefficients de Fourier décroîssent assez vite.
[8] et [20] ont étudié le problème de l'estimation de la densité à noyau sous le modèle (1) avec $m = 2$. Le critère de sélection de la fenêtre de [8] est basé sur la validation croisée. Les estimateurs en résultant ont été employés pour un test d'ajustement du modèle à rapport de densités pour deux échantillons en utilisant la distance L_2 entre les estimateurs à noyau semiparamétrique et non-paramétrique. Ils ont également démontré que leur estimateur semiparamétrique n'est pas seulement consistant, mais a aussi la plus petite variance asymptotique parmi les estimateurs à noyau classiques. [20] ont utilisé une approche itérative pour sélectionner le fenêtre et établissent la normalité asymptotique des estimateurs.

Le but de cette contribution est d'estimer des densités inconnues en deux étapes, en utilisant les données combinées de tous les échantillons, donc en tirant partie de l'information contenue dans tous les échantillons. Premièrement, en appliquant la méthode du maximum de vraisemblance empirique au modèle (1), deuxièmement, en procédant à une estimation par projection adaptative de la densité recherchée. La méthode par projection a été introduite par [7] et la méthode adaptative utilisée a été définie dans [6] et [5] pour les méthode développée ici est qu'elle permet d'atteindre dans un ensemble dense dans celui des densités “possibles” des vitesses suroptimales pour l'erreur quadratique intégrée, et ainsi de faire mieux que les estimateurs présentés précédemment. Dans la seconde partie nous rappelons la méthode d'estimation des paramètres de dimension finie dans le modèle (1) basée sur la vraisemblance empirique (voir [18] et ses références). La partie 3, en connexion avec la théorie de la partie 2, définit l'estimateur par projection adaptative de la densité inconnue. La partie 4 présente quelques propriétés asymptotiques de l'indice de troncature adaptatif introduit dans la partie 3. Enfin, le comportement asymptotique de l'estimateur considéré est étudié dans la partie 5. En particulier, nous démontrons que, lorsque la base de projection est choisie de telle sorte que les coefficients de Fourier s'annulent à partir d'un certain rang, l'estimateur proposé atteint des vitesses de convergence pour l'erreur quadratique intégrée suroptimales, faisant ainsi mieux que les estimateurs semiparamétriques présentés jusqu'alors.

2. Inférence dans le modèle à rapport de densités

Soient m échantillons dont les m densités correspondantes satisfont l'équation (1), soit de plus $n := \sum_{i=1}^{m} n_i$ le nombre total d'observations . Considérons par ailleurs la vraisemblance empirique (voir [15] et [16]) basée sur toutes les observations des m échantillons $\{X_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, m\}$

$$L(\theta, G_m) = \left\{ \prod_{j=1}^{n_1} p_{1j} w(X_{1j}, \theta_1) \right\} \left\{ \prod_{j=1}^{n_2} p_{2j} w(X_{2j}, \theta_2) \right\} \ldots \prod_{j=1}^{n_m} p_{mj},$$

où $p_{ij} := \mathrm{d}G_m(X_{ij})$ et $\theta = (\theta_1^t, \ldots, \theta_{m-1}^t)^t$ est un vecteur de dimension $(m-1)d$. On écrit la log-vraisemblance comme suit:

$$l(\theta, p) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log(p_{ij}) + \sum_{i=1}^{m-1} \sum_{j=1}^{n_i} \log\{w(X_{ij}, \theta_i)\}, \tag{2}$$

où $p := \{p_{ij}, j = 1, \ldots, n_i,\ i = 1, \ldots, m\}$.
La maximisation de l'équation (2) est obtenue en employant l'approche "profile" en deux étapes décrite dans [19]. Cette procédure se base tout d'abord sur la maximisation de la partie nonparamétrique dans la fonction de vraisemblance globale avec θ fixé, puis sur la maximisation de la fonction de log-vraisemblance "profile" par rapport à θ. Celle-ci vaut $l(\theta) = \sup_{p \in \mathcal{C}_\theta} l(\theta, p)$, où p est restreint à l'ensemble

$$\mathcal{C}_\theta := \left\{ p \in \mathbb{R}_+^n : \sum_{i=1}^{m} \sum_{j=1}^{n_i} p_{ij} = 1, \sum_{i=1}^{m} \sum_{j=1}^{n_i} p_{ij} \{w(X_{ij}, \theta_k) - 1\} = 0,\ k = 1, \ldots, m-1 \right\}.$$

La maximisation utilise la méthode des mutiplicateurs de Lagrange, et il vient que si l'ensemble $\mathcal{C}_\theta$ n'est pas vide,

$$p_{ij}(\theta) = \frac{1}{n} \frac{1}{1 + \sum_{k=1}^{m-1} \frac{n_k}{n} \{w(X_{ij}, \theta_k) - 1\}}, \tag{3}$$

et la log-vraisemblance profile devient, à une constante près,

$$l(\theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left[1+\sum_{k=1}^{m-1}\frac{n_k}{n}\{w(X_{ij},\theta_k)-1\}\right]+\sum_{i=1}^{m-1}\sum_{j=1}^{n_i}\log[w(X_{ij},\theta_i)].$$

On remarque que l'on peut voir $\widehat{\theta}$ comme un M-estimateur qui maximise la fonction critère $l(\theta)$, donc aussi

$$M_n(\theta) := \sum_{i=1}^{m}\mathbb{G}_{i,n_i}m_{\theta,i,\rho_{n_i}}$$

où $\mathbb{G}_{i,n_i}m_{\theta,i,\rho_{n_i}} := \frac{1}{n_i}\sum_{j=1}^{n_i}m_{\theta,i,\rho_{n_i}}(X_{ij}),\ i=1,\ldots,m,$

$$m_{\theta,i,\rho_{n_i}}(x) := \rho_{n_i}\left\{-\log\left[1+\sum_{k=1}^{m-1}\rho_{n_k}\{w(x,\theta_k)-1\}\right]+\log[w(x,\theta_i)]\right\}$$

pour $i=1,\ldots,m-1$ et

$$m_{\theta,m,\rho_{n_m}}(x) := -\rho_{n_m}\log\left[1+\sum_{k=1}^{m-1}\rho_{n_k}\{w(x,\theta_k)-1\}\right],$$

où $\rho_{n_i} := \frac{n_i}{n}$, $i=1,\ldots,m$.

Dans la suite, on suppose que $\rho_{n_i}\to\rho_i>0$, pour $n\to\infty$, $i=1,\ldots,m$, et l'on note de façon analogue $M(\theta) := \sum_{i=1}^{m}G_i m_{\theta,i,\rho_i}$, où G_i est la loi du $i-$ième échantillon, $i=1,\ldots,m$, et

$$m_{\theta,i,\rho_i}(x) := \rho_i\left\{-\log\left[1+\sum_{k=1}^{m-1}\rho_k\{w(x,\theta_k)-1\}\right]+\log[w(x,\theta_i)]\right\}$$

pour $i=1,\ldots,m-1$ et

$$m_{\theta,m,\rho_m}(x) := -\rho_m\log\left[1+\sum_{k=1}^{m-1}\rho_k\{w(x,\theta_k)-1\}\right].$$

On a utilisé la notation $G_i f = \int f dG_i$. Avec ces conventions, on peut appliquer une variation du Théorème 5.7. de [21] et l'on obtient le Lemme suivant:

Lemme 1 *Si* $\forall\ \varepsilon>0$,

$$\sup_{\theta\ \in\ \Theta}|M_n(\theta)-M(\theta)|\to 0\ p.s., \tag{4}$$

et

$$\sup_{\theta:d(\theta,\theta_0)\geq\varepsilon} M(\theta) < M(\theta_0), \tag{5}$$

alors, si $\widehat{\theta}$ *satisfait*

$$M_n(\widehat{\theta}) \geq M_n(\theta_0) - o(1), \tag{6}$$

alors

$$\widehat{\theta} \to \theta_0 \ p.s.$$

Nota Bene: dans le Lemme précédent, nous indiquons la vraie valeur du paramètre θ_0 dans un souci de simplicité. De plus, la convergence presque sûre l'est par rapport à G_m.

Démonstration La condition (4) implique que $M_n(\theta_0) \to M(\theta_0)$ p.s. . On en déduit à l'aide de la condition (6) que $M_n(\widehat{\theta}) \geq M(\theta_0) - o(1)$, d'où

$$M(\theta_0)-M(\widehat{\theta}) \leq M_n(\widehat{\theta})+o(1)-M(\widehat{\theta}) \leq \sup_{\theta} |M_n - M|(\theta)+o(1) \to 0 \quad p.s.. \tag{7}$$

Il vient de la condition (5) que

$$\forall\varepsilon > 0,\ \exists\eta > 0 \text{ tel que } \{d(\widehat{\theta},\theta_0) \geq \varepsilon\} \subseteq \{M(\widehat{\theta}) < M(\theta_0) - \eta\}.$$

Le résultat (7) permet de conclure la démonstration. ♢

En conclusion, une procédure “profile” permet d'aboutir à une estimation des paramètres de dimension finie.
La seconde étape consiste en l'estimation semiparamétrique des densités à partir des estimateurs obtenus ci-dessus.

3. Nouveaux estimateurs de la densité semiparamétriques par projection adaptative

Bien que le problème de l'estimation de la densité puisse être résolu avec un simple histogramme (voir [11]), une estimation plus lisse sera en général préférable. [8], [10] et [20] proposent des estimateurs semiparamétriques en modifiant les estimateurs classiques à noyau (essentiellement en lissant les sauts de $\widehat{G}_i$, $i = 1,\ldots,m$, où $\widehat{G}_i$ est l'estimateur du maximum de vraisemblance de G_i).
[1] introduisent des estimateurs semiparamétriques par projection. La méthode de projection consiste en la projection de la densité à estimer sur un espace de dimension finie (par exemple celui généré par les premières composantes d'une base de l'espace des densités possibles). On estime ensuite cette projection par une méthode des moments (voir [7] dans le cadre nonparamétrique).
Nous désignerons les observations des m échantillons $(X_{11},\ldots,X_{1n_1},\ldots,X_{mn_m})$ par $(T_1,...,T_n)$. Nous supposons que pour $l = 1,\ldots,m$, le $l-$ième échantillon admet la densité g_l par rapport à μ telle que $g_l \in L^2(\mu)$, où μ est une mesure

finie de mesure totale σ_μ. L'espace de Hilbert $(L^2(\mu), \|.\|)$ est supposé séparable, de dimension infinie et muni d'une base $e_1, e_2, \dots$ orthonormale.
On rappelle que pour tout $g \in L^2(\mu)$, la suite de ses coefficients de Fourier $(a_j)_{j\in\mathbb{N}^*}$ est l'unique suite de paramètres définissant g,

$$g = \sum_{j=1}^{\infty} a_j e_j,$$

où $a_j = \langle g, e_j \rangle_{L^2(\mu)} = \int g e_j \, \mathrm{d}\mu$, $j \geq 1$.
Dans ce travail, nous noterons donc a_j le $j-$ième coefficient de Fourier de la densité g_m.
L'estimateur par projection classique (voir [7]) pour $g_m = \sum_{j=1}^{\infty} a_j e_j$ est

$$\overline{g}_{m_{n_m}} = \sum_{j=1}^{k_{n_m}} \overline{a}_{j,n_m} e_j,$$

où (k_{n_m}) est une suite d'indices de troncature telle que $k_{n_m} \leq n_m$, $(n_m/k_{n_m}) \uparrow \infty$ et $(k_{n_m}) \uparrow \infty$ lorsque $n_m \uparrow \infty$. $\overline{a}_{j,n_m} = \frac{1}{n_m} \sum_{i=1}^{n_m} e_j(X_{mi})$ est un estimateur sans biais du $j-$ième coefficient de Fourier a_j. Nous supposons de plus que la base $(e_j)_{j\in\mathbb{N}^*}$ est uniformément bornée (telle que $\exists M < \infty : \sup_j \|e_j\|_\infty < M$).
Comme dans [1], il est possible de tirer profit de l'information de tous les échantillons (au lieu de ne considérer que le dernier) pour estimer plus efficacement g_m . Plus précisément, les auteurs ont utilisé un estimateur par projection semiparamétrique à indice de troncature déterministe:

$$\widehat{g}_{m_n} = \sum_{j=1}^{k_n} \widehat{a}_{j,n} e_j, \qquad \text{avec} \qquad \widehat{a}_{j,n} := \sum_{i=1}^{n} \widehat{p}_i e_j(T_i),$$

où la suite (k_n) est tel que $k_n \leq n$, $n/k_n \uparrow \infty$ et $(k_n) \uparrow \infty$ lorsque $n \uparrow \infty$.
Chaque T_i est associé à un p_i, $i = 1, \dots, n$ (après un réarrangement des (3)), qui est estimé par la méthode de vraisemblance empirique. Nous rappelons que $\sum_{i=1}^{n} p_i = 1$.
Ici, nous proposons d'utiliser un indice de troncature aléatoire dans l'estimateur adaptatif semiparamétrique par projection suivant:

$$\widehat{g}_{m_n} = \sum_{j=1}^{\widehat{k}_n} \widehat{a}_{j,n} e_j, \qquad \text{avec} \qquad \widehat{a}_{j,n} := \sum_{i=1}^{n} \widehat{p}_i e_j(T_i), \tag{8}$$

où

$$\widehat{k}_n := \max\{j : 1 \leq j \leq k_n : |\widehat{a}_{j,n}| \geq \gamma_n\}$$

et (k_n) vérifie les mêmes conditions que précédemment. Par ailleurs, la suite de seuils (γ_n) est donnée par

$$\gamma_n = \min\left(c\sqrt{\frac{\log n}{n}}, 1\right),\ c > 0.$$

Dans un souci de simplicité, dans la suite nous écrirons plutôt $\gamma_n = c\sqrt{\frac{\log n}{n}}$, ce qui est vrai pour n assez grand. Notons que le choix de c est laissé à l'utilisateur et qu'un tel choix influe sur $\widehat{k}_n$. Dans le cadre non-paramétrique, [3] a étudié le comportement de l'estimateur pour différentes valeurs de c.
Par ailleurs, afin d'assurer l'existence de $\widehat{k}_n$, nous prenons la première composante de la base de projection e_1 constante et égale à $\frac{1}{\sqrt{\sigma_\mu}}$.
Notons que l'équation (8) définit un estimateur de la densité semiparamétrique puisqu'il dépend à la fois de la fonction de répartition inconnue et des paramètres du modèle (1).
L'estimateur étudié ici est une modification de celui introduit dans le cas non-paramétrique par [14] pour une base de projection d'ondelettes, et par [6] et [5] pour d'autres choix de bases. D'autres travaux sur cet estimateur dans le cadre nonparamétrique ont été menés notamment dans le cas d'une base d'ondelettes (voir [4]) ou afin d'affaiblir certaines conditions portant par exemple sur le fait que la base de projection doit être uniformément bornée (voir [2]).

4. Comportement asymptotique de $\widehat{k}_n$

Par la suite, nous nous intéressons au comportement asymptotique de l'indice de troncature. Il apparaît que celui-ci se fixe presque sûrement sur le dernier terme non nul du développement de g_m sur la base orthonormale $(e_j)_{j\in\mathbb{N}^*}$ de l'espace des densités “possibles” considérée. De façon analogue, il tend presque sûrement vers l'infini si le nombre de termes dans le développement est infini. Un encadrement plus précis de l'indice de troncature est fourni par la suite. Nous distinguons par la suite les fonctions admettant un développement fini par rapport à la base $(e_j)_{j\in\mathbb{N}^*}$ (définissant le sous-espace $\mathcal{G}_0$) des autres (définissant le sous-espace $\mathcal{G}_1$).
Des propriétés suroptimales de l'estimateur sont ensuite données. Celles-ci concernent l'espérance quadratique intégrée pour $\mathcal{G}_0$ dense dans l'espace $L^2(\mu)$ des densités “possibles”.
Enfin, nous précisons le comportement de l'estimateur sur le complémentaire de $\mathcal{G}_0$ dans $L^2(\mu)$, $\mathcal{G}_1$.

Dans la suite, nous distinguons donc:

$$\begin{aligned}
\mathcal{G}_0(K) &:= \left\{g_m \in L^2(\mu) : a_K \neq 0, a_j = 0, j > K\right\}, \\
\mathcal{G}_0 &:= \bigcup_{K=1}^{\infty} \mathcal{G}_0(K) \\
\mathcal{G}_1 &:= L^2(\mu) - \mathcal{G}_0.
\end{aligned}$$

Sur $\mathcal{G}_0$, la vitesse pour l'espérance quadratique intégrée est meilleure que celle obtenue par la méthode d'estimation à noyau, à la fois dans les cas classique (nonparamétrique) et semiparamétrique (voir [8], [10] et [20]).
Nous déduisons enfin les estimateurs pour les autres densités g_l, $l = 1, \ldots, m-1$ comme suit:

$$\widehat{g}_{l_n} = \sum_{j=1}^{\widehat{k}_n} \widehat{a}_{j,n} e_j, \qquad \text{avec} \qquad \widehat{a}_{j,n} := \sum_{i=1}^{n} \widehat{p}_i w(T_i, \widehat{\theta}_l) e_j(T_i), \qquad l = 1, \ldots, m-1.$$

Evidemment, ces estimateurs jouissent des mêmes propriétés asymptotiques que (8).
Le comportement asymptotique de $\widehat{k}_n$ a une importance capitale pour celui de notre estimateur. Pour l'étudier nous utiliserons notmment l'inégalité de Bernstein (voir [17]). On notera dans la suite $M_p = \max_i(np_i)$. Par construction, on sait que $M_p < \infty$ (voir (3)).

Nous commençons par étudier $\widehat{k}_n$ pour $g_m \in \mathcal{G}_0(K)$. Nous voyons dans la propriété suivante de $\widehat{k}_n$ est alors un estimateur de K.

Théorème 4.1 *Si $g_m \in \mathcal{G}_0(K)$ et sous les hypothèses du Lemme 1, alors, pour n assez grand,*

$$\mathbb{P}\left(\widehat{k}_n > K\right) \leq 2n^{1-\frac{c^2}{4M^2M_p^2}} \quad \textit{et}$$

$$\mathbb{P}\left(\widehat{k}_n < K\right) \leq 2\exp\left\{\frac{-n|a_K|^2}{8\left(M_p^2M^2 + M_pM\frac{|a_K|}{6}\right)}\right\}.$$

Si, de plus, $c > 2\sqrt{2}M_pM$, alors on a pour n assez grand

$$\widehat{k}_n = K \quad p.s.$$

Démonstration Pour n assez grand on a $k_n > K$ donc

$$\left\{\widehat{k}_n > K\right\} \Rightarrow \bigcup_{j>K} \left\{|\widehat{a}_{j,n}| > \gamma_n\right\},$$

$$\mathbb{P}\left(\widehat{k}_n > K\right) \leq (k_n - K) \max_{k_n \geq j > K} \mathbb{P}\left(|\widehat{a}_{j,n}| > \gamma_n\right).$$

Nous ne pouvons pas appliquer l'inégalité de Bernstein pour la probabilité précédente car les $\widehat{p}_i$ dépendent de toutes les observations, et nous n'avons donc pas indépendance entre les $Y_i := n\widehat{p}_i e_j(T_i)$. Nous introduisons donc les quantités

$$\tilde{a}_{j,n} := \sum_{i=1}^{n} p_i e_j(T_i). \tag{9}$$

Comme vu dans [1], nous rappelons que

Lemme 2 *$\tilde{a}_{j,n}$ est un estimateur sans biais de a_j.*

Nous savons que $\widehat{p}_i = p_i$ presque sûrement puisque p_i et $\widehat{p}_i$ sont des fonctions continues différentiables respectivement des θ_i et des $\widehat{\theta}_i$ et que $\widehat{\theta}_i = \theta_i$ presque sûrement (voir Lemme 1). Nous en déduisons que $\widehat{a}_{j,n} = \tilde{a}_{j,n}$ presque sûrement comme intersection dénombrable d'événements presque sûrs. Il vient alors que

$$\mathbb{P}\left(\widehat{k}_n > K\right) \leq (k_n - K) \max_{k_n \geq j > K} \mathbb{P}\left(|\tilde{a}_{j,n}| > \gamma_n\right).$$

A présent, nous pouvons appliquer l'inégalité de Bernstein pour la probabilité précédente car, i_0 fixé, p_{i_0} ne dépend que de T_{i_0} (nous avons ainsi indépendance entre les $Y_i := np_i e_j(T_i)$). De plus, nous déduisons du Lemme 2 que $\mathbb{E}(Y_i) = a_j$. Par ailleurs, nous précisons que les Y_i ne sont pas identiquement distribuées puisque les T_i ne le sont pas. En effet, $T_1, \ldots, T_{n_1}$ sont des réalisations d'une variable aléatoire de densité g_1, $T_{n_1+1}, \ldots, T_{n_1+n_2}$ sont des réalisations d'une variable aléatoire de densité g_2, etc ...
Grâce à l'inégalité de Hölder, on a $\max_i(\mathbb{E}((np_i e_j(T_i))^2)) < M_p^2 M^2 < \infty$, où M est la borne uniforme de la base. Ceci implique que

$$\mathbb{P}\left(\widehat{k}_n > K\right) \leq 2(k_n - K) \exp \frac{-n\gamma_n^2}{2\left(M^2 M_p^2 + M_p M \frac{\gamma_n}{3}\right)}.$$

Or, puisque $\gamma_n \to 0$, à partir d'un certain rang, $M^2 M_p^2 > M_p M \frac{\gamma_n}{3}$. On en déduit que pour n assez grand

$$\mathbb{P}\left(\widehat{k}_n > K\right) \leq 2(k_n - K) \exp \frac{-n\gamma_n^2}{4M^2 M_p^2},$$

donc

$$\exists N : \forall n > N, \quad \mathbb{P}\left(\widehat{k}_n > K\right) \leq 2n^{1 - \frac{c^2}{4M^2 M_p^2}}.$$

On peut prendre $c > 2\sqrt{2} M M_p$. Pour un tel c, on conclut par le Lemme de Borel-Cantelli que $\widehat{k}_n \leq K$ presque sûrement.
Reste à démontrer que $\widehat{k}_n \geq K$ presque sûrement. Pour cela, on considère l'événement $\left\{\widehat{k}_n < K\right\}$.
A partir d'un certain rang, puisque $\gamma_n \to 0$ et $|a_K| \neq 0$, on a

$$\left\{\widehat{k}_n < K\right\} \Rightarrow \left\{|\widehat{a}_{K,n}| < \frac{|a_K|}{2}\right\},$$

d'où il vient que

$$\mathbb{P}\left(\widehat{k}_n < K\right) \leq \mathbb{P}\left(|\widehat{a}_{K,n} - a_K| > \frac{|a_K|}{2}\right).$$

De nouveau, on remarque que

$$\mathbb{P}\left(\widehat{k}_n < K\right) \leq \mathbb{P}\left(|\tilde{a}_{K,n} - a_K| > \frac{|a_K|}{2}\right).$$

Grâce à l'inégalité de Bernstein, on obtient alors

$$\mathbb{P}\left(\widehat{k}_n < K\right) \leq 2\exp\frac{-n|a_K|^2}{8\left(M_p^2 M^2 + M_p M\frac{|a_K|}{6}\right)}.$$

On conclut alors de manière analogue que $\widehat{k}_n \geq K$ p.s. et donc que $\widehat{k}_n = K$ p.s. $\diamondsuit$

Maintenant, si $g_m \in \mathcal{G}_1$, g_m n'admet pas de développement fini par rapport à la base $(e_j)_{j\in\mathbb{N}^*}$; nous obtenons alors le résultat suivant:

Théorème 4.2 *Si $g_m \in \mathcal{G}_1$ et sous les hypothèses du Lemme 1, alors on a pour n assez grand*

$$\widehat{k}_n \to \infty \quad p.s.$$

Démonstration Il suffit d'utiliser le fait que

$$\widehat{k}_n > j \quad p.s.$$

avec $a_j \neq 0$ (preuve pour j=K dans la démonstration précédente). On a alors que

$$\liminf_{n\to\infty} \widehat{k}_n > j \quad p.s.$$

On conclut en remarquant que sur $\mathcal{G}_1$, on a une infinité de tels j. $\diamondsuit$

Dans la suite, nous allons préciser le comportement asymptotique de $\widehat{k}_n$ à l'aide d'un nouveau paramètre dépendant de la densité g_m à estimer:

$$q(\eta) = min\left\{q \in \mathbb{N} : |a_j| \leq \eta, \quad \forall j > q\right\}, \quad \eta > 0.$$

Nous noterons par la suite $q_n(\varepsilon)$ (respectivement $q'_n(\varepsilon')$) l'entier $q((1+\varepsilon)\gamma_n)$ (respectivement $q((1-\varepsilon')\gamma_n)$) pour $\varepsilon > 0$ et $\varepsilon' \in\ ]0;1[$.

$\forall g_m \in L^2(\mu)$, $\sum_{i=0}^{\infty} a_j^2 < \infty$, ce qui implique la convergence de la suite $(a_j)_{j\in\mathbb{N}^*}$ vers 0 et donc que q est toujours défini sur $L^2(\mu)$.
La propriété suivante donne un résultat presque sûr sur le comportement asymptotique de $\widehat{k}_n$.

Théorème 4.3 *Si $g_m \in \mathcal{G}_1$ et sous les hypothèses du Lemme 1,*
$\forall \varepsilon > 0, \delta > 0\ \exists N_{\delta,\varepsilon} :\ \forall n > N_{\delta,\varepsilon}$,

$$\mathbb{P}\left(\widehat{k}_n < q_n(\varepsilon)\right) \leq 2n^{\frac{-c^2\varepsilon^2}{8(1+\delta)M_p^2M^2}}.$$

$\forall \varepsilon' \in]0;1[, \delta' > 0\ \exists N_{\delta',\varepsilon'} :\ \forall n > N_{\delta',\varepsilon'}$,

$$\mathbb{P}\left(\widehat{k}_n > q'_n(\varepsilon') \wedge k_n\right) \leq 2k_n n^{\frac{-\varepsilon'^2c^2}{2(1+\delta')M_p^2M^2}}.$$

On en déduit que si $c > max(D_{\delta,\varepsilon}, 2D_{\delta',\varepsilon'})$ avec $D_{\delta,\varepsilon} := \frac{2(1+\delta)M_p^2M^2}{\varepsilon}$, alors, si $n > max(N_{\delta,\varepsilon}, N_{\delta',\varepsilon'})$,

$$\widehat{k}_n \in [q_n(\varepsilon), q'_n(\varepsilon') \wedge k_n] \quad p.s.$$

Démonstration Si $q_n(\varepsilon) = 0$, la première inéquation est trivialement vérifiée. Sinon, par définition de $q_n(\varepsilon)$, on a que $|a_{q_n(\varepsilon)}| > (1+\varepsilon)\gamma_n$. Donc, puisque $k_n > q_n(\varepsilon)$, on a

$$\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \Rightarrow \left\{|\widehat{a}_{q_n(\varepsilon),n} - a_{q_n(\varepsilon)}| > \varepsilon\gamma_n\right\}.$$

Or, après le désormais classique passage à $\tilde{a}_{q_n(\varepsilon),n}$ et d'après l'inégalité de Bernstein avec $Y_i = np_{q_n(\varepsilon)}e_{q_n(\varepsilon)}(T_i)$,

$$\mathbb{P}\left(|\tilde{a}_{q_n(\varepsilon),n} - a_{q_n(\varepsilon)}| > \varepsilon\gamma_n\right) \leq 2\exp\left(\frac{-n\varepsilon^2\gamma_n^2}{2(\sigma^2 + \|Y_i\|_\infty \frac{\varepsilon\gamma_n}{3})}\right).$$

D'où il vient, grâce aux majorations analogues à celles effectuées dans la démonstration précédente sur σ^2 et $\|Y_i\|_\infty$, que

$$\mathbb{P}\left(\widehat{k}_n < q_n(\varepsilon)\right) \leq 2\exp\left(\frac{-c^2\varepsilon^2\log n}{2(M_p^2M^2 + M_pM\frac{\varepsilon c\sqrt{\log n}}{3\sqrt{n}})}\right).$$

Mais $\forall \delta > 0, \quad \exists N_\delta : \forall n > N_\delta, \quad M_pM\frac{\varepsilon c\sqrt{\log n}}{3\sqrt{n}} \leq \delta M_p^2M^2$. Donc, pour tout δ, à partir d'un certain rang,

$$\mathbb{P}\left(\widehat{k}_n < q_n(\varepsilon)\right) \leq 2n^{\frac{-c^2\varepsilon^2}{2(1+\delta)M_p^2M^2}},$$

ce qui est le premier résultat du Théorème. La seconde partie s'obtient en remarquant que, si $k_n > q'_n(\varepsilon')$ (sinon, elle est triviale), on doit démontrer que:

$$\mathbb{P}\left(\widehat{k}_n > q'_n(\varepsilon')\right) \leq 2k_n \exp \frac{-\varepsilon'^2 c^2 \log n}{2(1+\delta')M_p^2 M^2}.$$

On considère pour cela l'événement $\left\{\widehat{k}_n > q'_n(\varepsilon')\right\}$:

$$\left\{\widehat{k}_n > q'_n(\varepsilon')\right\} \Rightarrow \bigcup_{k_n \geq j > q'_n(\varepsilon')} \left\{|\widehat{a}_{j,n} - a_j| > \varepsilon' \gamma_n\right\}.$$

Ici encore, on applique l'inégalité de Bernstein aux composants de $\tilde{a}_{j,n}$ pour obtenir:

$$\mathbb{P}\left(\widehat{k}_n > q'_n(\varepsilon')\right) \leq 2\left(k_n - q'_n(\varepsilon')\right) \exp\left(\frac{-n\varepsilon'^2 \gamma_n^2}{2\left(M_p^2 M^2 + M_p M \frac{\varepsilon' c \sqrt{\log n}}{3\sqrt{n}}\right)}\right),$$

d'où il vient que

$$\forall \delta' > 0, \quad \exists N_{\delta'} : \forall n > N_{\delta'}, \quad \mathbb{P}\left(\widehat{k}_n > q'_n(\varepsilon')\right) \leq 2k_n \exp \frac{-\varepsilon'^2 c^2 \log n}{2(1+\delta')M_p^2 M^2},$$

ce qui conduit à l'inégalité annoncée.
Pour terminer, l'appartenance presque sûre à $[q_n(\varepsilon), q'_n(\varepsilon') \wedge k_n]$ est obtenue en remarquant que si $c > max(D_{\delta,\varepsilon}, 2D_{\delta',\varepsilon'})$ avec $D_{\delta,\varepsilon} := \frac{2(1+\delta)M_p^2 M^2}{\varepsilon}$, alors, si $n > \max(N_{\delta,\varepsilon}, N_{\delta',\varepsilon'})$,

$$\exists \eta < -1, \ \mathbb{P}\left(\widehat{k}_n < q_n(\varepsilon)\right) = \mathcal{O}(n^{\eta})$$

et

$$\exists \eta' < -1, \ \mathbb{P}\left(\widehat{k}_n > q'_n(\varepsilon') \wedge k_n\right) = \mathcal{O}(n^{\eta'}).$$

On conclut alors de nouveau par le Lemme de Borel-Cantelli. $\diamondsuit$

Les résultats de cette partie font donc de $\widehat{k}_n$ un estimateur de l'ordre de développement de g_m par rapport à la base de projection.

5. Comportement asymptotique de $\widehat{g}_{m_n}$

Dans cette partie, nous considérons l'erreur quadratique intégrée asymptotique de l'estimateur de la densité semiparamétrique $\widehat{g}_{m_n}$ (défini par (8)) comme mesure de sa qualité.
Pour étudier les propriétés statistiques de $\widehat{g}_{m_n}$, nous serons amenés à considérer

$$\tilde{g}_{m_n} = \sum_{j=1}^{\widehat{k_n}} \tilde{a}_{j,n} e_j,$$

avec $\tilde{a}_{j,n}$ défini en (9).

Théorème 5.1 *Sous les hypothèses du Lemme 1 et si $g_m \in \mathcal{G}_0(K)$, alors, pour n assez grand,*

$$\mathbb{E}\|\widehat{g}_{m_n} - g_m\|^2 = \mathcal{O}\left(\frac{1}{n}\right).$$

Notons que la vitesse optimale “classique”dans le cas semiparamétrique (voir [1]) est en $\mathcal{O}\left(\frac{k_n}{n}\right) + \sum_{j>k_n} a_j^2$.

Démonstration

$$\begin{aligned}\mathbb{E}\|\widehat{g}_{m_n} - g_m\|^2 &= \mathbb{E}\|\widehat{g}_{m_n} - \tilde{g}_{m_n}\|^2 + \mathbb{E}\|\tilde{g}_{m_n} - g_m\|^2 \\ &\quad +2\mathbb{E}\left[\int (\widehat{g}_{m_n} - \tilde{g}_{m_n})(\tilde{g}_{m_n} - g_m)\mathrm{d}\mu\right] = \\ &= A + B + C\end{aligned}$$

Analysons les trois termes de cette décomposition séparément.

Terme A

$$\widehat{g}_{m_n} - \tilde{g}_{m_n} = \sum_{j=1}^{\widehat{k_n}} e_j(\cdot)(\widehat{a}_{j,n} - \tilde{a}_{j,n}) = \sum_{j=1}^{\widehat{k_n}} \sum_{i=1}^{n} e_j(\cdot) e_j(T_i)(\widehat{p_i} - p_i).$$

Puisque nous avons que

$$\widehat{p_i} = p_i \quad p.s.,$$

alors, comme intersection finie des événements $\{\widehat{p_i} - p_i = 0\}$ presque sûrs et puisque la base de projection est uniformément bornée, on déduit que

$$\widehat{g}_{m_n}(\cdot) - \tilde{g}_{m_n}(\cdot) = 0 \quad p.s.$$

et par suite $\mathbb{E}\|\widehat{g}_{m_n} - \tilde{g}_{m_n}\|^2 = 0$.

Terme B

$$\begin{aligned}
\mathbb{E}\|\tilde{g}_{m_n} - g_m\|^2 &= \mathbb{E}\int \left(\sum_{j=1}^{\widehat{k}_n}(\tilde{a}_{j,n} - a_j)e_j(\cdot) - \sum_{j>\widehat{k}_n} a_j e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot) = \\
&= \mathbb{E}\int \left(\sum_{j=1}^{\widehat{k}_n}(\tilde{a}_{j,n} - a_j)e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot) \qquad (10)\\
&\quad -2\mathbb{E}\int \left(\sum_{j=1}^{\widehat{k}_n}(\tilde{a}_{j,n} - a_j)e_j(\cdot) \sum_{l>\widehat{k}_n} a_l e_l(\cdot)\right) \mathrm{d}\mu(\cdot) \qquad (11)\\
&\quad +\mathbb{E}\int \left(\sum_{j>\widehat{k}_n} a_j e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot). \qquad (12)
\end{aligned}$$

En remarquant que $\int e_j e_k \mathrm{d}\mu = \delta_{jk}$, nous obtenons pour le terme (10)

$$\mathbb{E}\int \left(\sum_{j=1}^{\widehat{k}_n}(\tilde{a}_{j,n} - a_j)e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot) \quad \leq \quad \mathbb{E}\left(\widehat{k}_n \mathcal{O}(1/n)\right).$$

Considérons l'événement $A_n := \{\widehat{k}_n = K\}$. D'après le Théorème 4.1, pour c assez grand,

$$\exists\ \eta < -1\ :\ \mathbb{P}\left(\overline{A_n}\right) \leq \frac{1}{n^\eta}.$$

Afin de majorer le *terme B*, nous cherchons à contrôler $\mathbb{E}\left(\widehat{k}_n\right)$:

$$\mathbb{E}\left(\widehat{k}_n\right) = \mathbb{E}\left(\widehat{k}_n(\mathbb{1}_{A_n} + \mathbb{1}_{\overline{A_n}})\right),$$

donc

$$\mathbb{E}\left(\widehat{k}_n\right) \leq K\mathbb{P}\left(A_n\right) + k_n\mathbb{P}\left(\overline{A_n}\right),$$

ce qui conduit à

$$\exists\beta < 0\ :\mathbb{E}\left(\widehat{k}_n\right) \leq K + \mathcal{O}(n^\beta).$$

On en déduit que

$$\mathbb{E}\int \left(\sum_{j=1}^{\widehat{k}_n}(\tilde{a}_{j,n} - a_j)e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot) \leq \mathcal{O}(1/n).$$

Le terme (11) s'annule directement du fait de l'orthogonalité de la base de projection choisie. Quant à (12), on peut le majorer comme suit:

$$\begin{aligned}
\mathbb{E}\int \left(\sum_{j>\widehat{k}_n} a_j e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot) &= \mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\right) = \mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2(\mathbb{1}_{A_n} + \mathbb{1}_{\overline{A_n}})\right) \\
&\leq \sum_{j>K} a_j^2 + \|g_m\|_2^2 \mathbb{P}\left(\overline{A_n}\right) = \|g_m\|_2^2 \mathbb{P}\left(\overline{A_n}\right).
\end{aligned}$$

Il vient donc que

$$\mathbb{E}\int \left(\sum_{j>\widehat{k}_n} a_j e_j(\cdot)\right)^2 \mathrm{d}\mu(\cdot) \leq o(1/n),$$

c'est-à-dire que le *terme B* est en $\mathcal{O}(1/n)$.

Finalement, puisque le *terme A* est négligeable par rapport au *terme B*, l'inégalité de Cauchy-Schwarz implique que le terme dominant de $\mathbb{E}\|\widehat{g}_{m_n} - g_m\|^2$ est en $\mathcal{O}\left(\frac{1}{n}\right)$, ce qui conclut la démonstration.

$\diamondsuit$

Nous nous intéressons à présent au comportement asymptotique de l'estimateur lorsque g_m n'admet pas de développement fini par rapport à la base de projection. Voici un encadrement asymptotique de $\mathbb{E}\|\widehat{g}_{m_n} - g_m\|^2$ dans ce cas.

Théorème 5.2 *Sous les conditions du Théorème* 4.3, *sous les hypothèses du Lemme 1 et si $g_m \in \mathcal{G}_1$, alors, pour n et c assez grands, on a presque sûrement*

$$\mathcal{O}\left(\frac{q((1+\varepsilon)\gamma_n)}{n}\right) + \sum_{j>q((1-\varepsilon')\gamma_n)\wedge k_n} a_j^2 \leq \mathbb{E}\|\widehat{g}_{m_n} - g_m\|^2 \leq \mathcal{O}\left(\frac{k_n}{n}\right) + \sum_{j>q((1+\varepsilon)\gamma_n)} a_j^2.$$

Démonstration L'erreur quadratique intégrée s'écrit comme suit:

$$\mathbb{E}\|\widehat{g}_{m_n} - g_m\|^2 = \mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n} (\widehat{a}_{j,n} - a_j)^2\right) + \mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\right).$$

Etudions le comportement du premier terme du membre de droite.

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n} (\widehat{a}_{j,n} - a_j)^2\right) \leq \sum_{j=0}^{k_n} \mathrm{Var}(\widehat{a}_{j,n}),$$

donc, puisque $\mathrm{Var}(\widehat{a}_{j,n}) = \mathcal{O}\left(\frac{1}{n}\right)$, on a

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{a}_{j,n}-a_j)^2\right) \leq \mathcal{O}\left(\frac{k_n}{n}\right).$$

Par ailleurs, considérons l'événement $A_n := \left\{q_n(\varepsilon) \leq \widehat{k}_n \leq q'_n(\varepsilon') \wedge k_n\right\}$. On a

$$\mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\right) = \mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2(\mathbb{1}_{A_n} + \mathbb{1}_{\overline{A_n}})\right),$$

d'où

$$\mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\right) \leq \sum_{j>q_n(\varepsilon)} a_j^2 + \|g_m\|_2^2\mathbb{P}(\overline{A_n}).$$

D'après le Théorème 4.3, on sait que $\mathbb{P}(\overline{A_n}) = o\left(\frac{1}{n}\right)$. La première partie de la propriété (concernant la majoration de l'erreur quadratique) est prouvée. Pour prouver la minoration, on utilise la même décomposition:

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{a}_{j,n}-a_j)^2\right) \geq \mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{a}_{j,n}-a_j)^2\mathbb{1}_{A_n}\right),$$

d'où

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{a}_{j,n}-a_j)^2\right) \geq \mathcal{O}\left(\frac{q_n(\varepsilon)}{n}\right).$$

De manière analogue, on a

$$\mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\right) \geq \mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\mathbb{1}_{A_n}\right),$$

donc

$$\mathbb{E}\left(\sum_{j>\widehat{k}_n} a_j^2\right) \geq \sum_{j>k_n\wedge q'_n(\varepsilon')} a_j^2,$$

ce qui conclut la démonstration.

$\diamondsuit$

6. Conclusions et Perspectives

Nous avons défini une nouvelle classe d'estimateurs semiparamétriques par projection de la densité. En plus de d'élargir les conditions usuelles de travail sur des problèmes à m échantillons, cette méthode a l'avantage de conduire, pour des bases de projection bien choisies par l'utilisateur, à des estimateurs pouvant atteindre des vitesses suroptimales (paramétriques) pour l'erreur quadratique intégrée asymptotique.

En outre, les logiciels statistiques usuellement utilisés sont bien adaptés pour calculer ces estimateurs. A ce propos, nous développons un "package" relatif à cette méthode pour le logiciel libre R.

Enfin, le choix aléatoire de l'indice de troncature $\widehat{k}_n$ s'appuie sur des choix préalables de constantes influant grandement sur la valeur de $\widehat{k}_n$ à distance finie. Il serait oppportun de procéder à de nombreuses simulations numériques afin de préciser le comportement de notre estimateur, d'un côté par rapport à différents choix de ces constantes, et d'un autre par rapport à ses concurrents naturels, les estimateurs semiparamétriques à noyau ou à projection non adaptatifs.

Acknowledgements

Nous remercions le professeur Bosq sans qui cet article n'aurait pas vu le jour ainsi que le referee qui a, par ses remarques, amélioré ce travail.

References

[1] Aubin, J.B. et Leoni-Aubin, S. (2008) Projection density estimation under a m-sample semiparametric model. *Computational Statistics and Data Analysis* **52**(5), 2451–2468.
[2] Aubin, J.B. (2006) Comportement asymptotique d'estimateurs de la densité par projection tronqués. *Comptes Rendus Mathematique* **342**(11), 873–876.
[3] Aubin, J.B. (2005) Estimation localement suroptimale de la densité par projection tronquée. *Annales de l'ISUP* **49**(2-3), 3–19.
[4] Aubin, J.B. et Massiani, A. (2003) Comportement asymptotique d'un estimateur de la densité adaptatif par méthode d'ondelettes. *Comptes Rendus Mathematique* **337**(4), 293–296.
[5] Bosq, D. (2005) *Inférence et prévision en grandes dimensions*. Economica, Paris.
[6] Bosq, D. (2002) Estimation localement suroptimale et adaptative de la densité. *Comptes Rendus Mathematique* **334**(7), 591–595.

[7] Cencov, N.N. (1962) Estimation of unknown distribution density from observations. *Soviet Math. Dokl.* **3**, 1559–1562.
[8] Cheng, K.F. and Chu, C.K. (2004) Semiparametric density estimation under a two-sample density ratio model. *Bernoulli* **10**(4), 583–604.
[9] Efron, B. et Tibshrani, R. (1996) Using specially designed exponential families for density estimation. *Ann. of Stat.* **24**(6), 2431–2461.
[10] Fokianos, K. (2004) Merging information for semiparametric density estimation. *J. R. Statist. Soc. B* **66**(4), 941–958.
[11] Fokianos, K., Kedem, B., Qin, J. Haferman, J. and Short, D.A. (1998) On combining instruments. *Journal of Applied Meteorology* **37**, 220–226.
[12] Keziou, A. and Leoni-Aubin, S. (2008) On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference* **138**(4), 915–928.
[13] Keziou, A. and Leoni-Aubin, S. (2005) Test of homogeneity in semiparametric two-sample density ratio models. *Comptes Rendus Mathematique* **340**(12), 905–910.
[14] Lepskii, O.V., Mammen, e. et Spokoiny v.g. (1994) Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwith selection. *Technical Report.*
[15] Owen, A.B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** (2), 237–249.
[16] Owen, A.B. (2001) *Empirical Likelihood*. Chapman and Hall, New York.
[17] pollard, d. (1984) Convergence of stochastic processes, *Springer-Verlag* New York.
[18] Qin, J. (1998) Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika* **85** (3), 619–630.
[19] Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.* **22** (1), 300–325.
[20] Qin, J. et Zhang, B. (2005) Density estimation under a two-sample semiparametric model. *Nonparametric Statistics* **17** (6), 665–683.
[21] Van der Vaart, A.W. (1998) Asymptotic Statistics. *Cambridge series in Statistical and Probabilistic Mathematics, Cambridge University Press.*