Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Acad. Sci. Paris, Ser. I 342 (2006) 873–876
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Comportement asymptotique d'estimateurs de la densité par projection tronqués

Jean-Baptiste Aubin

Université Paris 6, LSTA, boîte 158, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 15 avril 2005 ; accepté après révision le 6 mars 2006

Disponible sur Internet le 2 mai 2006

Présenté par Paul Deheuvels

Résumé

Nous étudions deux versions tronquées de l'estimateur de la densité (par rapport à une mesure σ-finie μ) par projection. Ces versions se basent sur des indices de troncature dépendants des données et elles seront étudiées dans divers contextes. Nous décrivons d'abord le comportement asymptotique des indices de troncature. Nous montrons alors que les estimateurs correspondants atteignent une vitesse suroptimale au sens de l'erreur quadratique intégrée sur un sous-ensemble $\mathcal{F}_0$ dense de $L_2(\mu)$. De plus, nous déterminons les cas dans lesquels les estimateurs atteignent une vitesse quasi-optimale pour cette même erreur sur le complémentaire de $\mathcal{F}_0$. ***Pour citer cet article : J.-B. Aubin, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

Abstract

Asymptotic behavior of truncated projection density estimators. We present different versions of the truncated projection estimator of a density with respect to a σ-finite measure μ, where the traditional truncation index k_n is replaced by $\hat{k}_n$ (or $\tilde{k}_n$) under various conditions. First, we describe the asymptotic behaviour of $\hat{k}_n$ (or $\tilde{k}_n$). Next, we show that these estimators reach a superoptimal rate for the mean square error on a dense subset $\mathcal{F}_0$ of $L^2(\mu)$. We finally state under which conditions these estimators reach quasioptimal rate of convergence when the unknown density f belongs to $\mathcal{F}_0^c$. ***To cite this article: J.-B. Aubin, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

1. Introduction

On dispose d'une suite de n variables aléatoires $X_1, \ldots, X_n$ définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$ et à valeurs dans un espace mesuré $(E, \mathcal{B}, \mu)$ où μ est finie. On suppose qu'elles ont toutes la même loi qu'une v.a. X_0, et qu'elles admettent une densité f par rapport à μ telle que $f \in \mathcal{F}$ où $\mathcal{F} \subseteq L^2(\mu)$. L'espace de Hilbert $(L^2(\mu), \|\cdot\|)$ est supposé séparable, de dimension infinie et muni d'une base $(e_j)_{j\in\mathbb{N}}$ orthonormale où $e_0 = c_0$ constante dépendant de la masse totale de μ. L'estimateur de $f = \sum_{j=0}^{\infty} a_j e_j$ par projection (cf. Cencov [6]) et par projection tronquée (cf. Bosq [4]) sont définis respectivement par les relations :

Adresse e-mail : jbaubin@ccr.jussieu.fr (J.-B. Aubin).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2006.03.012

$$f_{k_n} = \sum_{j=0}^{k_n} \hat{a}_{j,n} e_j, \quad k_n \geqslant 0 \quad \text{et} \quad \hat{f}_n = \sum_{j=0}^{\hat{k}_n} \hat{a}_{j,n} e_j, \quad k_n \geqslant \hat{k}_n \geqslant 0$$

où $\hat{a}_{j,n} = \frac{1}{n}\sum_{i=1}^{n} e_j(X_i)$, où la suite $(k_n)_{n\in\mathbb{N}}$ est à choisir tendant vers l'infini et où l'indice de troncature $\hat{k}_n$ dépend des données.

Bosq [4] a étudié l'estimateur par projection tronquée dans le cas **A** où

- la base de projection est uniformément bornée i.e. $\exists M < \infty\ \forall j \in \mathbb{N}, \|e_j\|_\infty < M$;
- $\hat{k}_n = \max\{j\colon\ 0 \leqslant j \leqslant k_n\colon\ |\hat{a}_{j,n}| \geqslant \gamma_n\}$ où $(k_n)_{n\in\mathbb{N}}$ est une suite d'entiers telle que $(k_n) \uparrow \infty$ et $k_n/n \to 0$ et où $\gamma_n = \max(c\sqrt{\frac{\log n}{n}}, c_0)$;
- les X_i sont indépendants.

N.B : Lepskii et al. [8] et Aubin et Massiani [2] ont utilisé un estimateur analogue dans le cas des ondelettes et Bosq et Blanke [5] dans le cas continu alors que celui d'Efromovich [7] par exemple est voisin mais différent. Dans cette étude, nous nous intéressons aux trois cas où (toutes les autres hypothèses étant égales à celles du cas **A**) :

- cas **B** : la **B**ase de projection n'est plus uniformément bornée mais simplement bornée :

  i.e. $\exists (M_{k_n})_{n\in\mathbb{N}}\colon\ \forall n \in \mathbb{N}, \quad \sup_{j\leqslant k_n} \|e_j\|_\infty < M_{k_n}$.

  Ici, on suppose que $M_{k_n} = \mathrm{o}(\sqrt{\frac{n}{\log n}})$ et que $\mathcal{F} \subseteq \{f \in L^2(\mu)\colon\ \exists M' < \infty, \|f\|_\infty < M'\}$.
- cas **C** : le **C**hoix de l'indice de troncature devient :

  $$\tilde{k}_n = \min\bigl\{j\colon\ 0 \leqslant j \leqslant k_n\colon\ \|f_{k_n} - f_j\|^2 \leqslant \delta_n\bigr\}$$

  avec $\delta_n = c k_n \sqrt{\frac{\log n}{n}}$, $c > 0$. De plus, $(k_n) \uparrow \infty$ et $k_n = \mathrm{o}(\sqrt{\frac{n}{\log n}})$.
- cas **D** : Les X_i sont **D**épendants. Les v.a. $e_j(X_i) - a_j$ sont α-mélangeantes (voir par exemple Bosq [3] p. 7 pour définition) avec $\alpha(m) \leqslant a\exp(-bm)$, $a > 0$ et $b > 0$. On prend alors $\gamma_n = c\log^{\Gamma} n/\sqrt{n}$ avec $\Gamma > 1$, $c > 0$.

Les cas **B** et **D** sont plus généraux que le cas **A** (dès que l'on admet que f est uniformément bornée pour **B** ce qui ne devrait pas poser de problème en pratique). Le cas **C** fait, lui, intervenir plusieurs $\hat{a}_{j,n}$ dans le critère de sélection de l'indice de troncature ce qui devrait rendre ce dernier moins sensible aux variations d'un seul $\hat{a}_{j,n}$. Dans la suite, on pose $\mathcal{F}_0 = \bigcup_{K\geqslant 0} \mathcal{F}_0(K)$ où

$$\mathcal{F}_0(K) = \left\{ f \in L^2(\mu)\colon\ f = \sum_{j=0}^{K} a_j e_j \right\},$$

$\mathcal{F}_1 = L^2(\mu)\backslash\mathcal{F}_0$ ainsi que $m_{k_n} = \inf_{j\leqslant k_n} \mathrm{var}(e_j(X_0))$.

2. Comportement asymptotique de l'indice de troncature

Proposition 1. *Ici, $\breve{k}_n$ désigne selon le cas $\hat{k}_n$ ou $\tilde{k}_n$. Pour c assez grand, on a*

(i) *Si $f \in \mathcal{F}_0(K)$, alors, presque sûrement pour n assez grand, $\breve{k}_n = K$.*
(ii) *Si $f \in \mathcal{F}_1$, alors, presque sûrement, $\breve{k}_n \to \infty$ pour n assez grand.*

Remarque 1. Respectivement dans les cas **B**, **C** et **D**, c doit être supérieur à $4M'$, $2\sqrt{6}M^2$ et $8\sqrt{2}M$.

Pour préciser le comportement asymptotique de $\hat{k}_n$ (respectivement $\tilde{k}_n$) lorsque $f \in \mathcal{F}_1$, on pose :

$$q(\eta) = \min\bigl\{q \in \mathbb{N}\colon\ |a_j| \leqslant \eta,\ \forall j > q\bigr\}, \quad \eta > 0$$

(respectivement $r(\eta) = \min\{0 \leqslant q \leqslant k_n\colon\ \sum_{j=q}^{k_n} a_j^2 < \eta\}$ et $r(\eta) = k_n$ si $a_{k_n}^2 > \eta$).

Proposition 2. *On a*

- *cas* $\boldsymbol{B}$: $\forall\delta>0$, *si* $c>2\sqrt{2M'}$, $\varepsilon>\frac{\sqrt{(2+2\delta)M'}}{c}$, $1>\varepsilon'>\frac{2\sqrt{2M'}}{c}$ *et* $k_n\geqslant q((1+\varepsilon)\gamma_n)$, *alors* $q((1+\varepsilon)\gamma_n)\leqslant\hat{k}_n\leqslant q((1-\varepsilon')\gamma_n)$ *p.s. pour n assez grand* ;
- *cas* $\boldsymbol{C}$: *si* $c>\frac{4\sqrt{2}M^2}{\varepsilon\wedge\varepsilon'}$, *alors* $r((1+\varepsilon)\delta_n)\leqslant\tilde{k}_n\leqslant r((1-\varepsilon')\delta_n)$ *p.s. pour n assez grand* ;
- *cas* $\boldsymbol{D}$: *si* $c>8\sqrt{2}M$ *et* $k_n\geqslant q((1+\varepsilon)\gamma_n)$, *alors* $q((1+\varepsilon)\gamma_n)\leqslant\hat{k}_n\leqslant q((1-\varepsilon')\gamma_n)$ *p.s. pour n assez grand.*

3. Suroptimalité de l'estimateur sur $\mathcal{F}_0$

Ici, on suppose que les densités étudiées f sont éléments de $\mathcal{F}_0(K)$ où $K\in\mathbb{N}$.

Proposition 3. $\check{f}_n$ *désigne selon le cas* $\hat{f}_n$ *ou* $\tilde{f}_n$. *Pour c assez grand, on a*

- *cas* $\boldsymbol{B}$ *et* $\boldsymbol{C}$: $n\mathbb{E}_f\|\check{f}_n-f\|^2\to\sum_{j=1}^K\{\int e_j^2 f\,\mathrm{d}\mu-a_j^2\}$;
- *cas* $\boldsymbol{D}$: $n\mathbb{E}_f\|\hat{f}_n-f\|^2=\mathcal{O}(1)$

Remarque 2. Respectivement dans les cas **B**, **C** et **D**, c doit être supérieur à $16M'$, $2\sqrt{10}M^2$ et $8\sqrt{3}M$.

4. Comportement de l'estimateur sur $\mathcal{F}_1$

Proposition 4. *Sous les conditions de la Proposition 2, on a pour n assez grand*

- *cas* $\boldsymbol{B}$:

$$\frac{m_{k_n}q((1+\varepsilon)\gamma_n)}{n}+\sum_{j>q((1-\varepsilon')\gamma_n)\wedge k_n}a_j^2\leqslant\mathbb{E}_f\|\hat{f}_n-f\|^2\leqslant\frac{4M_{k_n}^2k_n}{n}+\sum_{j>q((1+\varepsilon)\gamma_n)}a_j^2+\mathrm{o}\left(\frac{1}{n}\right),$$

- *cas* $\boldsymbol{C}$:

$$\frac{m_{k_n}r((1+\varepsilon)\delta_n)}{n}+\sum_{j>r((1-\varepsilon')\delta_n)}a_j^2\leqslant\mathbb{E}_f\|\tilde{f}_n-f\|^2\leqslant\frac{4M^2r((1-\varepsilon')\delta_n)}{n}+\sum_{j>r((1+\varepsilon)\delta_n)}a_j^2+\mathrm{o}\left(\frac{1}{n}\right),$$

- *cas* $\boldsymbol{D}$:

$$\frac{m_{k_n}q((1+\varepsilon)\gamma_n)}{n}+\sum_{j>q((1-\varepsilon')\gamma_n)\wedge k_n}a_j^2\leqslant\mathbb{E}_f\|\hat{f}_n-f\|^2\leqslant\frac{5aM^2k_n}{(\mathrm{e}^m-1)n}+\sum_{j>q((1+\varepsilon)\gamma_n)}a_j^2+\mathrm{o}\left(\frac{1}{n}\right).$$

Cette propriété nous permet d'énoncer le corollaire suivant :

Corollaire 4.1. *Sous les hypothèses de la propriété précédente, on a*

- *cas* $\boldsymbol{B}$: *Si* $k_n\geqslant q((1-\varepsilon')\gamma_n)$, *alors* $\mathbb{E}_f\|\hat{f}_n-f\|^2=\mathcal{O}(\frac{k_n(\log n\vee M_{k_n}^2)}{n})+\sum_{j>k_n}a_j^2$,
- *cas* $\boldsymbol{C}$: $\mathbb{E}_f\|\tilde{f}_n-f\|^2=\mathcal{O}(k_n\sqrt{\frac{\log n}{n}})+\sum_{j>k_n}a_j^2$,
- *cas* $\boldsymbol{D}$: *Si* $k_n\geqslant q((1-\varepsilon')\gamma_n)$, *alors* $\mathbb{E}_f\|\hat{f}_n-f\|^2=\mathcal{O}(\frac{k_n\log^\Gamma n}{n})+\sum_{j>k_n}a_j^2$, $\Gamma>1$.

Dans le cas **B**, sous réserve que M_{k_n} croisse assez lentement, le taux de convergence est identique à celui obtenu par Bosq [4] dans le cas **A**. De même dans le cas **D**, on peut parler de quasioptimalité car on observe la perte d'un $\log^{\Gamma-1}n$, $\Gamma>1$ par rapport à la vitesse optimale. Enfin, le taux obtenu dans le cas **C** ne permet plus de parler de quasioptimalité car la perte est alors de l'ordre de $\sqrt{\frac{n}{\log n}}$.

5. Commentaires

Les preuves (voir Aubin [1]) des deux premières propositions utilisent notamment les inégalités de Bernstein dans les cas **B** et **C** (voir par exemple Pollard [9] p. 193) et le Théorème 1.3 p. 23 (voir Bosq [3]) dans le cas **D**. Enfin, le Corollaire 1.1 p. 13 dans Rio [10] est utile pour démontrer la Proposition 4.

Remerciements

Je remercie M. Bosq pour ses suggestions constructives et ses encouragements.

Références

[1] J.B. Aubin, Rapport Technique, LSTA, 2005.
[2] J.B. Aubin, A. Massiani, Comportement asymptotique d'un estimateur de la densité adaptatif par méthode d'ondelettes, C. R. Acad. Sci. Paris, Ser. I 337 (4) (2003) 293–296.
[3] D. Bosq, Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, Springer-Verlag, New York, 1996.
[4] D. Bosq, Estimation localement suroptimale et adaptative de la densité, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 591–595.
[5] D. Bosq, D. Blanke, Local superefficiency of data-driven projection density estimators in continuous time, SORT 28 1 (2004) 37–53.
[6] N.N. Cencov, Soviet Math. Dokl. 3 (1962) 1559–1562.
[7] S. Efromovich, Nonparametric Curve Estimation, Springer-Verlag, New York, 1999.
[8] O.V. Lepskii, E. Mammen, V.G. Spokoiny, Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwith selection, Technical report, 1994.
[9] D. Pollard, Convergence of Stochastic Processes, Springer-Verlag, New York, 1984.
[10] E. Rio, Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants, Springer-Verlag, New York, 2000.