

# Estimation localement suroptimale et adaptative de la densité

Denis Bosq

74, rue Dunois, 75013 Paris, France

Reçu le 21 novembre 2001 ; accepté le 29 janvier 2002

Note présentée par Paul Deheuvels.

---

## Résumé

On étudie une version tronquée de l'estimateur par projection dans un cadre général. On montre que cet estimateur atteint une vitesse suroptimale sur un ensemble dense dans la classe des densités à estimer et une vitesse quasi-optimale ailleurs. Cet ensemble peut être choisi par le statisticien et la vitesse suroptimale est atteinte pour l'erreur quadratique intégrée et la convergence uniforme presque sûre. Une version adaptative de l'estimateur est également considérée. *Pour citer cet article : D. Bosq, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 591–595.* © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

## Locally superoptimal and adaptive projection density estimators

## Abstract

We study a data-driven version of the density projection estimator in a general framework. We show that this estimator reaches a superoptimal rate on a dense set in the density class, and a quasi-optimal rate elsewhere. This set can be chosen by the statistician, and the superoptimal speed is reached for integrated quadratic error and almost sure uniform convergence. An adaptive version of the estimator is also considered. *To cite this article: D. Bosq, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 591–595.* © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

---

## Abridged English version

We study a data-driven version of the density projection estimator in a general framework. More precisely let  $(X_n, n \geq 0)$  be a sequence of i.i.d. r.v.'s with values in a probability space  $(E, \mathcal{B}, \mu)$ . Let  $1 = e_0, e_1, e_2, \dots$  be a bounded orthonormal basis of  $L^2(\mu)$ . If  $X_0$  has a density  $f$  that belongs to  $L^2(\mu)$  one may define the estimator

$$\hat{f}_n = \sum_{j=0}^{\hat{k}_n} \hat{a}_{j_n} e_j,$$

where  $\hat{a}_{j_n} = \frac{1}{n} \sum_{i=1}^n e_j(X_i)$  and  $\hat{k}_n = \max\{j : 0 \leq j \leq k_n, |\hat{a}_{j_n}| \geq \gamma_n\}$ .  $k_n$  and  $\gamma_n \in [0, 1]$  are chosen by the statistician.  $\hat{f}_n$  is not a threshold estimator in the sense of [3] and [4]. We show that the index  $\hat{k}_n$

---

Adresse e-mail : bosq@ccr.jussieu.fr (D. Bosq).

has a precise asymptotic behaviour in various situations. In particular, if  $f = \sum_{j=0}^K a_j(f)e_j$  ( $a_K(f) \neq 0$ ) then  $\hat{k}_n \rightarrow K$  a.s. This property implies  $E_f \|\hat{f}_n - f\|^2 = O(1/n)$  and  $\|\hat{f}_n - f\|_\infty = O((\ln \ln n/n)^{1/2})$  a.s. (superoptimal rates).

Now consider  $\mathcal{F}_\varphi = \{f : |a_j(f)| \leq \varphi(j)\}$  where  $\varphi(j) \downarrow 0$  and  $\sum_j \varphi(j) < \infty$ . Then there exists  $f^* \in \mathcal{F}_\varphi$  such that  $|a_j(f^*)| \downarrow 0$  and  $|a_j(f^*)| \geq \delta\varphi(j) > 0, j \geq 0$ , and, if  $\gamma_n = c(\frac{\ln n}{n})^{1/2}$ , one has  $\sup_{f \in \mathcal{F}_\varphi} E_f \|\hat{f}_n - f\|^2 \leq 8c^2 v_n \ln n$  where  $v_n = k_n^*/n + \sum_{j > k_n^*} \varphi^2(j)$  and  $k_n^*$  is characterized by  $a_{k_n^*}^2(f^*) > \alpha n^{-1}, a_{k_n^*+1}^2(f^*) \leq \alpha n^{-1}$  with  $0 < \alpha < 1$ . Similar quasi-optimal rates are obtained for a.s. uniform convergence.

Finally an adaptive version of  $\hat{f}_n$  is considered. It is noteworthy that the statistician may choose  $(e_j)$  in order to reach the suroptimal rates over an enough large family of densities.

### 1. Estimation de la densité par projection tronquée

Soit  $(X_n, n \in \mathbb{N})$  une suite de v.a. i.i.d. définies sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans un espace probabilisé  $(E, \mathcal{B}, \mu)$ .  $X_0$  admet une densité  $f$  par rapport à  $\mu$  telle que  $f \in \mathcal{F}$  avec  $\mathcal{F} \subseteq L^2(\mu)$ . L'espace de Hilbert  $(L^2(\mu), \|\cdot\|)$  est supposé séparable, de dimension infinie, et muni d'une base orthonormale  $1 = e_0, e_1, e_2, \dots$ . Pour simplifier les énoncés on se limite au cas où  $M = \sup_{j \geq 0} \|e_j\|_\infty < \infty$ .

L'estimateur par projection de  $f = \sum_{j=0}^\infty a_j(f)e_j$  s'écrit  $f_n = \sum_{j=0}^{k_n} \hat{a}_{j_n} e_j$  avec  $\hat{a}_{j_n} = \frac{1}{n} \sum_{i=1}^n e_j(X_i)$  (cf. [2]) alors que l'estimateur par projection tronquée est défini par

$$\hat{f}_n = \sum_{j=0}^{\hat{k}_n} \hat{a}_{j_n} e_j,$$

où  $\hat{k}_n = \max\{j : 0 \leq j \leq k_n, |\hat{a}_{j_n}| \geq \gamma_n\}$ . L'entier  $k_n$  et  $\gamma_n \in [0, 1]$  sont à choisir.  $\hat{f}_n$  est différent de l'estimateur à seuil étudié dans [3] et [4].

Dans la suite on pose  $\mathcal{F}_0 = \bigcup_{K=0}^\infty \mathcal{F}_0(K)$  où  $\mathcal{F}_0(K) = \{f \in L^2(\mu) : a_K(f) \neq 0; a_j(f) = 0, j > K\}$  et  $\mathcal{F}_1 = L^2(\mu) \setminus \mathcal{F}_0$ . Par ailleurs on suppose que  $k_n \rightarrow \infty, k_n/n \rightarrow 0, \gamma_n = c(\frac{\ln n}{n})^{1/2}$  où  $c > 2M$ . Enfin on pose  $m(f) = \inf_{j \geq 1} (\int e_j^2 f - a_j^2(f))$ .

### 2. Comportement asymptotique de l'indice de troncature

PROPOSITION 1. – Si  $f \in \mathcal{F}_0(K), \hat{k}_n \rightarrow K$  p.s. ( $K \geq 0$ ); sinon,  $\hat{k}_n \rightarrow \infty$  p.s.

Pour préciser le comportement asymptotique de  $\hat{k}_n$  quand  $f \in \mathcal{F}_1$  on pose

$$q(\eta) = \min\{q \in \mathbb{N} : |a_j| \leq \eta \text{ pour tout } j > q\}, \quad \eta > 0.$$

PROPOSITION 2. – Pour  $\varepsilon > M\sqrt{2}/c, 2M/c < \varepsilon' < 1$  et  $k_n \geq q((1 - \varepsilon')\gamma_n)$  on a, presque sûrement pour  $n$  assez grand,

$$q_n := q((1 + \varepsilon)\gamma_n) \leq \hat{k}_n \leq q((1 - \varepsilon')\gamma_n) =: q'_n.$$

En particulier, si  $|a_j| = \alpha \rho^j$  ( $\alpha > 0, 0 < \rho < 1$ )  $\hat{k}_n$  se stabilise :

$$\frac{\hat{k}_n}{\ln n} \rightarrow \left(2 \ln \frac{1}{\rho}\right)^{-1} \text{ p.s.}$$

ce qui permet de construire un estimateur convergent de  $\rho$  en posant  $\hat{\rho}_n = n^{-1/2}(\hat{k}_n+1)$ .

PROPOSITION 3. – Si  $X_0$  a pour loi  $\mu$ , on a la majoration

$$d_V(P_{\hat{k}_n}, \nu_n) \leq 2c_0 M^3 \frac{k_n^5}{\sqrt{n}},$$

où  $c_0$  est une constante universelle,  $d_V$  la distance en variation et  $\nu_n$  la loi discrète définie par

$$\nu_n(k) = p_n^{\mathbb{1}_{\{k>0\}}} (1 - p_n)^{k_n - k}, \quad 0 \leq k \leq k_n,$$

avec  $p_n = P(|N| > c \ln n)$  ( $N \sim \mathcal{N}(0, 1)$ ).

### 3. Vitesses suroptimales

On choisit  $c > M\sqrt{6}$ .

PROPOSITION 4. –

$$\left\| n\mathbb{C}(\hat{f}_n - f) - \mathbb{C}\left(\sum_{j=0}^K e_j(X_0)e_j - f\right) \right\|_{\mathcal{N}} \rightarrow 0, \quad f \in \mathcal{F}_0(K) \quad (K \geq 0),$$

où  $\|\cdot\|_{\mathcal{N}}$  désigne la norme nucléaire et  $\mathbb{C}(Z)$  l'opérateur d'ordre 2 de la v.a. Hilbertienne  $Z$ . Par conséquent

$$nE_f \|\hat{f}_n - f\|^2 \rightarrow \sum_{j=0}^K \left( \int e_j^2 f \, d\mu - a_j^2(f) \right).$$

De plus

$$\limsup_{n \rightarrow \infty} \sup_{f \in \mathcal{F}_0(K, 2\gamma_n)} nE_f \|\hat{f}_n - f\|^2 \leq KM^2,$$

où  $\mathcal{F}_0(K, 2\gamma_n) = \{f : f \in \mathcal{F}_0(K), |a_K(f)| \geq 2\gamma_n\}$ .

Un raisonnement classique (« méthode des 2 fonctions ») montre que la vitesse  $1/n$  est minimax pour  $K > 0$ . Pour  $K = 0$  la vitesse est un  $o(1/n)$ .

PROPOSITION 5. – Pour tout entier  $K \geq 0$

$$\overline{\lim}_{n \rightarrow \infty} \left( \frac{n}{\ln \ln n} \right)^{1/2} \|\hat{f}_n - f\|_{\infty} \leq \sqrt{2}KM^{3/2} \quad p.s., \quad f \in \mathcal{F}_0(K),$$

en particulier

$$\|\hat{f}_n - 1\|_{\infty} = o\left(\left(\frac{\ln \ln n}{n}\right)^{1/2}\right) \quad p.s. \text{ si } P_{X_0} = \mu.$$

PROPOSITION 6. – Si  $f \in \mathcal{F}_0(K)$ ,  $\sqrt{n}(\hat{f}_n - f)$  vérifie le TCL dans  $L^2(\mu)$  et  $\left(\frac{n}{\ln \ln n}\right)^{1/2}(\hat{f}_n - f)$  vérifie la LLI dans  $L^2(\mu)$ . La limite et la limite supérieure sont dégénérées si  $f = 1$ .

**4. Vitesses quasi-optimales**

PROPOSITION 7. – Si  $f \in \mathcal{F}_1$ ,  $E_f \|\widehat{f}_n - f\|^2 \rightarrow 0$ .

PROPOSITION 8. – Sous les conditions de la Proposition 2

$$E_f \|\widehat{f}_n - f\|^2 \leq M \frac{q'_n}{n} + \sum_{j>q_n} a_j^2(f) + o\left(\frac{1}{n}\right)$$

et

$$E_f \|\widehat{f}_n - f\|^2 \geq m(f) \frac{q_n}{n} + \sum_{j>q'_n} a_j^2(f) + o\left(\frac{1}{n}\right)$$

et par conséquent

$$E_f \|\widehat{f}_n - f\|^2 = O\left(\frac{k_n}{n} \ln n\right) + \sum_{j>k_n} a_j^2(f).$$

Notons que  $f_n$  vérifie

$$E_f \|f_n - f\|^2 = O\left(\frac{k_n}{n}\right) + \sum_{j>k_n} a_j^2(f).$$

Posons maintenant  $\mathcal{F}_\varphi = \{f : |a_j(f)| \leq \varphi(j)\}$  où  $\varphi(j) \downarrow 0$  et  $\sum \varphi(j) < \infty$ . Alors il existe  $f^* \in \mathcal{F}_\varphi$  telle que  $|a_j(f^*)| \downarrow 0$  et  $|a_j(f^*)| \geq \delta \varphi(j) > 0$ . On caractérise  $k_n^*$  par  $a_{k_n^*}^2(f^*) > \alpha n^{-1}$ ,  $a_{k_n^*+1}^2(f^*) \leq \alpha n^{-1}$  ( $0 < \alpha < 1$ ) et on pose  $v_n = k_n^*/n + \sum_{j>k_n^*} \varphi^2(j)$ .

Enfin on désigne par  $\mathcal{T}_\ell$  la classe des estimateurs de la forme  $g_n(\cdot) = \frac{1}{n} \sum_{i=1}^n H_n(X_i, \cdot)$  où  $H_n = \sum_{j=0}^\infty \lambda_{j_n} e_j \otimes e_j$  ( $\sum \lambda_{j_n}^2 < \infty$ ). Alors

PROPOSITION 9. –

$$\inf_{g_n \in \mathcal{T}_\ell} \sup_{f \in \mathcal{F}_\varphi} E_f \|g_n - f\|^2 \geq \frac{(\alpha \wedge \delta^2) m(f^*)}{\alpha + M^2} v_n > 0$$

et pour  $k_n = k_n^*$

$$\sup_{f \in \mathcal{F}_\varphi} E_f \|f_n - f\|^2 \leq M^2 v_n,$$

alors que pour  $c > M\sqrt{6}$  et  $n$  assez grand

$$\sup_{f \in \mathcal{F}_\varphi} E_f \|\widehat{f}_n - f\|^2 \leq 8c^2 v_n \ln n.$$

On a des résultats analogues pour  $\mathcal{F}_\psi = \{f : \sum_{j>k} a_j^2(f) \leq \psi(k), k > 0\}$  où  $\psi(k) \downarrow 0$ , mais sous des conditions plus contraignantes.

PROPOSITION 10. – Si  $\sum_j |a_j(f)| < \infty$  alors  $\|\widehat{f}_n - f\|_\infty \rightarrow 0$  p.s.

PROPOSITION 11. – Si  $f \in \mathcal{F}_\varphi$  et  $c > M\sqrt{6}$  alors

$$\|\widehat{f}_n - f\|_\infty = O(u_n) \text{ p.s.,}$$

où  $u_n = k_n^* \left(\frac{\ln n}{n}\right)^{1/2} + \sum_{j>k_n^*} \varphi(j)$ .

## 5. Estimateur adaptatif

On considère le choix « omnibus »  $k_n = \left[ M^2 \frac{n}{\ln n} \right] + 1$  et on impose à  $a_j(f)$  la condition peu contraignante  $|a_j(f)| \leq M/j$ ,  $j \geq j_0$ . Enfin on choisit  $c = 3M + 1$ . Alors la plupart des résultats précédents subsistent :

PROPOSITION 12. –

$$\sup_{f \in \mathcal{F}_\varphi} E_f \|\widehat{f}_n - f\|^2 \leq M^2 \frac{q_{n,\varphi}}{n} + (6M + 1)^2 \frac{\ln n}{n} q'_{n,\varphi} + \sum_{j > q'_{n,\varphi}} \varphi^2(j) + o\left(\frac{1}{n}\right),$$

et

$$\begin{aligned} \|\widehat{f}_n - f\|_\infty &\leq M \sum_{j=0}^{q_{n,\varphi}} |\widehat{a}_{j_n} - a_j| + M(6M + 1) \left(\frac{\ln n}{n}\right)^{1/2} q_{n,\varphi} \\ &\quad + M \sum_{j > q_{n,\varphi}} \varphi(j) + M \left(2k_n + \sum_{j=0}^{\infty} \varphi(j)\right) \mathbb{1}_{\widehat{k}_n > q_{n,\varphi}}, \end{aligned}$$

où  $q_\varphi$  désigne  $q$  lorsque  $|a_j| = \varphi(j)$ .

On obtient alors des vitesses analogues aux précédentes.

## 6. Commentaires

Le statisticien peut choisir la base  $(e_j)$  en orthonormalisant une suite  $(h_j)$  de densités privilégiées, de façon à atteindre la vitesse suroptimale sur le convexe engendré par les  $h_j$ . Notons qu'inversement toute base orthonormale  $(e_j)$  constituée de fonctions bornées est engendrée par une suite de densités. D'autre part «  $f \in \mathcal{F}_\varphi$  » (ou «  $f \in \mathcal{F}_\psi$  ») n'est pas nécessairement une condition de régularité : c'est une condition d'approximation par des densités appartenant à  $\mathcal{F}_0$ , celles-ci pouvant éventuellement être irrégulières.

Les preuves utilisent notamment [4,5] et [7], elles apparaissent dans [1]. Enfin un estimateur analogue pour la régression figure dans [6].

## Références bibliographiques

- [1] D. Bosq, Rapport technique, LSTA, 2002.
- [2] N.N. Cencov, Soviet Math. Dokl. 3 (1962) 1559–1562.
- [3] D. Donoho, I. Johnstone, G. Kerkycharian, D. Picard, Ann. Statist. 24 (2) (1996) 508–539.
- [4] S. Efromovich, Nonparametric Curve Estimation, Springer, New York, 1999.
- [5] W. Hoeffding, J. Amer. Statist. Assoc. 58 (1963) 13–30.
- [6] G. Lee, J. Hart, J. Nonparametr. Statist. 12 (2000) 683–707.
- [7] V.V. Sazonov, Sankhya A 30 (1968) 191–204.