



Méthodes récursives en estimation et prévision non paramétriques

Aboubacar Amiri

► **To cite this version:**

Aboubacar Amiri. Méthodes récursives en estimation et prévision non paramétriques. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494792>

HAL Id: inria-00494792

<https://hal.inria.fr/inria-00494792>

Submitted on 24 Jun 2010

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.

MÉTHODES RÉCURSIVES EN ESTIMATION ET PRÉVISION NON PARAMÉTRIQUES

Aboubacar AMIRI

*Université d'Avignon et des Pays de Vaucluse, Laboratoire d'Analyse Non Linéaire et
Géométrie (EA 2151), F-84018 Avignon
e-mail : aboubacar.amiri@univ-avignon.fr*

Résumé. Nous introduisons d'abord une famille paramétrique d'estimateurs récursifs de la densité indexée par un paramètre $\ell \in [0, 1]$. Leur comportement asymptotique en fonction du paramètre ℓ va nous amener à introduire des critères de comparaison basés sur les biais, variance et erreur quadratique asymptotiques. Pour ces critères, nous comparons nos estimateurs entre eux et aussi comparons notre famille d'estimateurs à l'estimateur non récursif de la densité de Parzen-Rozenblatt. Ensuite, nous définissons à partir de notre famille d'estimateurs de la densité une famille d'estimateurs récursifs à noyau de la fonction de régression. Nous étudions également ses propriétés asymptotiques en fonction du paramètre ℓ . Ces résultats permettent ainsi de construire une famille de prédicteurs non paramétriques qui permettent de réduire le temps calcul.

1 Cadre général et motivations

Nous nous intéressons dans cet exposé aux méthodes d'estimation non paramétriques à partir d'estimateurs récursifs à noyau ainsi qu'à leurs applications à la prévision. Considérons un processus à temps discret $\{X_k : k \geq 1\}$ et on souhaite prédire la valeur de X_{n+h} (prévision à l'horizon $h \in \mathbb{N}^*$). Plusieurs méthodes existent dans la littérature, cependant il n'existe pas de meilleure méthode que les autres dans toutes les situations. La méthode de prévision par k -moyennes mobiles, par exemple, consiste à prendre comme prévision la moyenne des observations des k périodes précédentes. La prévision est renouvelée de période en période. Cette méthode est simple d'utilisation, avec pour avantage d'atténuer suffisamment les fluctuations de la série tout en préservant son allure générale, cependant elle s'avère moins performante lorsque l'on prend un nombre élevé de données. La méthode du lissage exponentiel, quant à elle, prend en compte la prévision de la période antérieure. A cette prévision, on augmente l'écart subi, pondéré d'un coefficient a compris entre 0 et 1. Le prédicteur est donc défini, pour $h = 1$, par :

$$\widehat{X}_{n+1} = (1 - a) \sum_{j=0}^{n-1} a^j X_{n-j}, 0 < a < 1,$$

et peut se calculer de manière récursive par la relation :

$$\widehat{X}_{n+1} = a\widehat{X}_n + (1 - a)X_{n+1} \quad (0 < a < 1). \quad (1)$$

On constate à partir de (1) que la valeur prédite à l'instant $(n + 1)$ est une moyenne pondérée entre la valeur estimée faite en n et la dernière observation de la série. L'avantage de cette récursivité est que l'on n'a pas à relisser de nouveau le processus lorsqu'une nouvelle observation s'ajoute à la série. Ce qui n'est pas négligeable car cela permet de réduire considérablement le temps de calcul. Tout comme la méthode de prévision par moyenne mobile, le lissage exponentiel est simple d'utilisation et facilement compréhensible mais son principal inconvénient réside sur le choix de la constante de lissage a . Elle est également moins efficace pour prédire des longues séries. Pour des prévisions à long terme on privilégiera les méthodes de Box-Jenkins.

Encore plus robustes que les méthodes de Box-Jenkins, et ayant l'avantage d'une mise en pratique très facile, les méthodes dites non paramétriques sont plus récemment apparues pour tenter d'apporter un nouveau regard sur le problème de la prévision. Le principe de la prévision non paramétrique repose sur le fait que le problème de la prévision peut être vu comme un cas particulier de l'estimation de la régression dans le sens où si l'on suppose que le processus est Markovien, strictement stationnaire et d'ordre 1, alors prédire X_{n+1} revient à estimer l'espérance conditionnelle $E(X_{n+1}/X_n)$ par $\widehat{X}_{n+1} = \widehat{r}_{n-1}(X_n)$, où $\widehat{r}_n(x)$ est l'estimateur de la régression à noyau de $E(X_1/X_0 = x)$ basé sur les observations (X_i, X_{i+1}) pour $i = 1, \dots, n - 1$.

L'objectif principal de cet exposé est d'améliorer les performances du prédicteur non paramétrique par noyau, en terme de temps de calcul en utilisant des estimateurs récursifs à noyau de la régression pour la construction du prédicteur. Pour cela nous avons besoin d'estimer la régression de manière récursive et donc nous devons d'abord estimer la densité de manière recursive.

2 Présentation de nos estimateurs

Considérons un processus stochastique bivarié $\{\xi_t = (X_t, Y_t), t \in \mathbb{N}\}$ défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$, à valeurs dans $\mathbb{R}^d \times \mathbb{R}^{d'}$. A partir d'une suite de variables équidistribuées (X_n, Y_n) , on cherche à estimer la densité f des X_i et une version la fonction de régression définie par :

$$r(x) := \begin{cases} E(m(Y_0) | X_0 = x) = \frac{\int_{\mathbb{R}^{d'}} m(y) f^*(x, y) dy}{f(x)} := \frac{\varphi(x)}{f(x)}, & \text{si } f(x) > 0 \\ Em(Y_0), & \text{si } f(x) = 0, \end{cases}$$

m est une fonction Borélienne de $\mathbb{R}^{d'}$ à valeurs dans \mathbb{R} telle que $\omega \mapsto m^2(Y_t(\omega))$ soit P -intégrable et f^* est la densité de probabilité de ξ_n .

Pour estimer la densité f , nous proposons la famille paramétrique d'estimateurs récursifs

à noyau définie par :

$$f_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n \frac{1}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R}^d \quad (\ell \in [0, 1]), \quad (2)$$

vérifiant la relation recursive suivante :

$$f_{n+1}^\ell(x) = \frac{\sum_{i=1}^n h_i^{d(1-\ell)}}{\sum_{i=1}^{n+1} h_i^{d(1-\ell)}} f_n^\ell(x) + K_{n+1}^\ell(x - X_{n+1}), \quad K_i^\ell(\cdot) := \frac{1}{h_i^{d\ell} \sum_{j=1}^i h_j^{d(1-\ell)}} K\left(\frac{\cdot}{h_i}\right). \quad (3)$$

Quant à la fonction de régression elle est estimée par la famille d'estimateurs récursifs définie par :

$$r_n^\ell(x) := \frac{\varphi_n^\ell(x)}{f_n^\ell(x)} \quad \text{où} \quad \varphi_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{(1-\ell)d}} \sum_{i=1}^n \frac{m(Y_i)}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right). \quad (4)$$

Ces estimateurs peuvent se calculer de manière récursive par :

$$r_{n+1}^\ell(x) = \frac{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right) \varphi_n^\ell(x) + \left(\sum_{i=1}^{n+1} h_i^{d(1-\ell)}\right) m(Y_{n+1}) K_{n+1}^\ell(x - X_{n+1})}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right) f_n^\ell(x) + \left(\sum_{i=1}^{n+1} h_i^{d(1-\ell)}\right) K_{n+1}^\ell(x - X_{n+1})},$$

où $K_i^\ell(\cdot)$ est défini dans (3).

3 Hypothèses et quelques résultats théoriques

L'étude théorique des estimateurs (2) et (4) nécessite les hypothèses suivantes :

Hypothèses H.1.

- (i) $K : \mathbb{R}^d \mapsto \mathbb{R}$ est une densité de probabilité, strictement positive, symétrique et bornée ;
- (ii) $\lim_{\|x\| \rightarrow \pm\infty} \|x\|^d K(x) = 0, \quad \forall x \in \mathbb{R}^d;$
- (iii) $\int_{\mathbb{R}^d} |v_i v_j| K(v) dv < \infty, \quad i, j = 1, \dots, d.$

Les hypothèses **H.1** sont classiques en estimation non paramétrique, vérifiées en particulier par les noyaux d'Epanechnikov, Gaussien etc.

Nous avons également besoin des conditions suivantes imposées à la fenêtre h_n :

Hypothèses H.2.

- (i) $h_n \downarrow 0$ et $nh_n^{d+2} \rightarrow \infty$ lorsque $n \rightarrow \infty$;
- (ii) $B_{n,r} := \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n}\right)^r \rightarrow \beta_r < \infty, \quad n \rightarrow \infty \quad \forall r \in]-\infty, d+2].$

La décroissance de h_n n'est pas nécessaire en estimation non récursive. La condition **H.2(ii)** est très utile dans nos calculs, et est également propre à la récursivité. Elle est souvent utilisée dans la littérature.

Nous avons le résultat suivant qui établit la vitesse de convergence presque sûre de notre famille d'estimateurs de la densité dans le cas iid.

Théorème 1. *Sous les hypothèses **H.1** et **H.2**, et pour des observations iid, si*

$$\frac{nh_n^d}{(\ln \ln n)^{2(\alpha+1)} \ln n} \rightarrow \infty, \text{ lorsque } n \rightarrow \infty, \text{ pour tout } \alpha \geq 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\ln h_n}{\ln n} < \infty,$$

(a) *alors pour tout x tel que $f(x) > 0$,*

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{nh_n^d}{\ln \ln n}} \left(f_n^\ell(x) - E f_n^\ell(x) \right) = \sigma_\ell \sqrt{2} \text{ p.s. },$$

avec :

$$\sigma_\ell^2(x) := \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} f(x) \int_{\mathbb{R}^d} K^2(x) dx. \quad (5)$$

(b) *De plus le choix $h_n = C_n \left(\frac{\ln \ln n}{n} \right)^{\frac{1}{d+4}}$, avec $C_n \downarrow c > 0$, implique que :*

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln \ln n} \right)^{\frac{2}{d+4}} \left(f_n^\ell(x) - f(x) \right) = \sigma_\ell \sqrt{2c^d} + \frac{c^2 \beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_f(x) \text{ p.s. },$$

pour tout x tel que $f(x) > 0$, avec :

$$b_f(x) := \frac{1}{2} \sum_{1 \leq i, j \leq d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v) dv. \quad (6)$$

Ensuite dans le cadre dépendant on considère une hypothèse supplémentaire afin d'établir la convergence en moyenne quadratique et la normalité asymptotique de nos estimateurs de la densité.

Hypothèses H.3.

(i) Le processus (X_t) est $2 - \alpha$ -mélangeant avec : $\alpha^{(2)}(k) \leq \gamma k^{-\rho}$, $k \geq 1$ pour deux constantes strictement positives γ et ρ .

(ii) Pour chaque couple (s, t) , $s \neq t$, le vecteur aléatoire (X_s, X_t) admet une densité $f_{(X_s, X_t)}$ telle que : $\sup_{|s-t| \geq 1} \|g_{s,t}\|_\infty < \infty$, où $g_{s,t} := f_{(X_s, X_t)} - f \otimes f$.

Les hypothèses **H.3** sont classiques dans ce domaine, en particulier, il n'y a pas d'hypothèse de stationnarité de second ordre statuée sur le processus.

Cela permet d'énoncer les deux résultats suivants :

Théorème 2. *Sous les hypothèses **H.1** – **H.3** :*

(a) *Pour tout $\ell \in \left[\left(\frac{d-2}{2d} \right)^+, 1 \right]$,*

$$nh_n^d \text{Var} f_n^\ell(x) \rightarrow \sigma_\ell^2, \text{ lorsque } n \rightarrow \infty, \text{ si } \rho > 2, \text{ avec } \sigma_\ell \text{ défini dans (5).}$$

- (b) Si $d \geq 3$ et $\ell \in \left[0, \frac{d-2}{2d}\right[$, la conclusion du (a) reste encore vraie si $\rho > \frac{d+2}{2}$.
- (c) Sous les conditions du (a) (avec $\rho > 2$) ou du (b) (avec $\rho > \frac{d+2}{2}$), le choix $h_n = C_n n^{-\frac{1}{d+4}}$, $C_n \downarrow c > 0$, entraîne que :

$$n^{\frac{4}{d+4}} E \left(f_n^\ell(x) - f(x) \right)^2 \longrightarrow c^4 \left(\frac{4 + d\ell}{2 + d\ell} \right)^2 b_f^2(x) + \frac{(4 + d\ell)^2 f(x) \|K\|_2^2}{2c^d(4 + d)(2 + d\ell)},$$

lorsque $n \rightarrow \infty$, pour les valeurs respectives de ℓ en tout point où $f(x) > 0$, où $b_f(x)$ est défini en (6).

Théorème 3. Supposons les hypothèses **H.1** – **H.3** satisfaites et que pour toutes suites d'entiers u_n et v_n , la suite h_n vérifie la condition $u_n \sim v_n \Rightarrow h_{u_n} \sim h_{v_n}$. S'il existe un réel positif ς_0 tel que $\frac{nh_n^d}{(\log n)^{\varsigma_0}} \rightarrow +\infty$ lorsque $n \rightarrow \infty$, alors pour tout x tel que $f(x) > 0$, on a :

$$\sqrt{nh_n^d} \left(f_n^\ell(x) - E f_n^\ell(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \sigma_\ell^2 \right),$$

lorsque $n \rightarrow \infty$ avec σ_ℓ défini dans (5).

Enfin, moyennant quelques hypothèses supplémentaires sur le processus associé (ξ_t) , on peut établir des résultats similaires pour les estimateurs de la régression (4).

Bibliographie

- [1] **Ahmad**, I.A. and **Lin**, P.E. (1976). Nonparametric sequential estimation of a multiple regression function. Bull. Math. Stat. 17, 63-75.
- [2] **Amiri**, A. (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortement mélangeant. C. R. Acad. Sci. Paris, Ser. I 347, 309-314.
- [3] **Bosq**, D. (1998). Nonparametric statistics for stochastic processes. Estimation and prediction, Lecture notes in statistics 110 2nd ed. Springer-Verlag, New York.
- [4] **Bosq**, D. and **Blanke**, D. (2007). Inference and prediction in large dimensions. Wiley Series in Prob. and Stat. ISBN 978-0-470-08147-1.
- [5] **Collomb**, G. (1981). Estimation non-paramétrique de la régression : revue bibliographique. Int. Stat. Rev., Vol. 49, 1, 75-93.
- [6] **Collomb**, G. (1985) Nonparametric regression : an up-to-date bibliography. Statistics 16, 2, 309-324.

- [7] **Davies**, H.L. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Stat.* 15, 49-54.
- [8] **Deheuvels**, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C. R. Acad. Sci. Paris Vol. 278*, 1217-1220.
- [9] **Devroye**, L., **Gyorfi**, L. (1985). Nonparametric density estimation : the L_1 view, Wiley, New-York, Russian translation : Mir, Moscow, 1988.
- [10] **Devroye**, L. and **Wagner**, T.J. (1980). On the L_1 convergence of kernel estimators of regression functions with application in discrimination. *Z. Wahrsch. verw. Gebiete* 51, 15-25.
- [11] **Gyorfi**, L., **Kholer**, M., **Krzyzak**, A. and **Walk**, H. (2002). A distribution-free theory of nonparametric regression, *S.-Ver. N. york*.
- [12] **Jain**, C., **Jogdeo**, K. and **Stout**, W. (1975). Upper and lower functions for martingales and mixing processes, *Ann. Prob.*, 3, 1, 119-145.
- [13] **Krzyzak**, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation, *IEEE Trans.*, 38, 4, 1323-1328.
- [14] **Prakasa Rao**, B.L.S. (1983), Nonparametric functional estimation, New-York : Academic Press.
- [15] **Roussas**, G.G. (1992). Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimation, *Nonparam. Stat.* 1, 171-195.
- [16] **Silverman**, B. W. (1986). Density estimation for statistics and data analysis. Monographs on Stat. and Appl. Prob. Chapman and Hall.
- [17] **Wegman**, E. and **Davies**, H.I. (1979). Remarks on some recursive estimators of a probability density, *Ann. Stat.*, 7, 2, 316-327.
- [18] **Wertz**, W. (1985). Sequential and recursive estimators of the probability density, *Statistics* 16, 277-295.
- [19] **Wolverton**, C. and **Wagner**, T.J. (1969). Recursive estimates of probability densities. *IEEE Trans.* 5, 307-308.
- [20] **Yamato**, H. (1972). Sequential estimation of a continuous propability density function and mode. *Bull. Math. Statist. Jap.*, Vol. 14, 1-12.