Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 347 (2009) 309–314
COMPTES RENDUS MATHEMATIQUE

Statistique

Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortement mélangeant

Aboubacar Amiri

Université d'Avignon et des Pays de Vaucluse, Laboratoire d'analyse non linéaire et géométrie (EA 2151), 84018 Avignon, France

Reçu le 25 novembre 2008 ; accepté après révision le 26 janvier 2009

Disponible sur Internet le 24 février 2009

Présenté par Paul Deheuvels

Résumé

Soit $(X_t, t \in \mathbb{N})$ un processus α-mélangeant, où les X_t sont des vecteurs de $\mathbb{R}^d$ de même loi, de densité de probabilité inconnue f. Nous nous proposons d'estimer f de manière récursive à l'aide des observations $X_1, \ldots, X_n$. Pour cela, nous considérons une sous-famille des estimateurs récursifs généraux initiés par Deheuvels (1974), incluant les estimateurs récursifs les plus utilisés. Pour cette sous-famille, nous obtenons l'erreur quadratique asymptotique exacte, ensuite, nous introduisons des critères de comparaison qui nous permettent de classifier et comparer nos estimateurs. ***Pour citer cet article : A. Amiri, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Abstract

On a parametric family of sequential estimators of the density for a strong mixing process. Let $(X_t, t \in \mathbb{N})$ be a $\mathbb{R}^d$-valued α-mixing process, where the X_t's have the same unknown density f. We suggest to estimate f, recursively, from the data $X_1, \ldots, X_n$. So, we introduce a subfamily of the general recursive kernel estimators initiated by Deheuvels (1974), including the most popular recursive estimators. For this subfamily, we establish the exact asymptotic square error and then we introduce criteria for comparison that allow us to make a choice among our estimators. ***To cite this article: A. Amiri, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Abridged English version

Let $(X_t, t \in \mathbb{N})$ be a $\mathbb{R}^d$-valued α-mixing process, where the X_t's have the same unknown density f. To estimate f, from the observations $X_1, \ldots, X_n$, we introduce the following parametric family of recursive estimators defined as:

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

Recursive relationship. For all $l \in [0,1]$, one has the following recursive relationship:

Adresse e-mail : aboubacar.amiri@univ-avignon.fr.

1631-073X/$ – see front matter © 2009 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.
doi:10.1016/j.crma.2009.01.026

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

The family $(f_n^l(x))$ is a subfamily of the general recursive kernel estimators:

$$f_n^H(x) := \left(\sum_{i=1}^n h_i H(h_i)\right)^{-1} \sum_{i=1}^n H(h_i) K\left(\frac{x - X_i}{h_i}\right),$$

initiated by Deheuvels [6] for $d = 1$. Our family includes the most famous recursive estimators: Wagner–Wolvertone [11] for $l = 1$, Deheuvels [6] for $l = 0$ and Wegman–Davis [10] for $l = \frac{1}{2}$. Our first thought was to generalize the Wegman–Davis's idea, consisting in splitting the bandwidth in two powers: here, we choose the powers l and $1 - l$, $l \in [0, 1]$. Moreover, contrary to them, we consider here asymptotically unbiased estimators that conduct us to the family $(f_n^l(x))$. In this Note, we firstly establish the exact asymptotic values of bias, variance and mean square error of $f_n^l(x)$ according to l. So we have the following:

Theorem. *Under the assumptions $\mathcal{K}$, $\mathcal{H}$, $\mathcal{P}$, and $\mathcal{Q}$ (assumptions and notations are exposed in the French version):*

(a)

$$h_n^{-4}\left(\mathrm{E} f_n^l(x) - f(x)\right)^2 \underset{n\to\infty}{\longrightarrow} \left(\frac{\beta_{d(1-l)+2}}{\beta_{d(1-l)}}\right)^2 b_2^2(x),$$

with

$$b_2^2(x) := \frac{1}{4}\left(\sum_{1\leqslant i,j\leqslant d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v)\,\mathrm{d}v\right)^2.$$

(b) *For all $l \in [(\frac{d-2}{2d})^+, 1]$, $nh_n^d \operatorname{Var} f_n^l(x) \to \frac{\beta_{d(1-2l)}}{\beta_{d(1-l)}^2} f(x) \int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u$, $n \to \infty$, if $\rho > 2$.*

(c) *If $d \geqslant 3$ and $l \in [0, \frac{d-2}{2d}[$, the conclusion of* (b) *is true if $\rho > \frac{d+2}{2}$.*

(d) *Under conditions of* (b) (*for $\rho > 2$*) *or of* (c) (*for $\rho > \frac{d+2}{2}$*) *with $f(x) > 0$, the choice $h_n = C_n n^{-\frac{1}{d+4}}$, $C_n \to c > 0$, implies that:*

$$n^{\frac{4}{d+4}} \mathrm{E}\left(f_n^l(x) - f(x)\right)^2 \longrightarrow c^4 \left(\frac{4 + dl}{2 + dl}\right)^2 b_2^2(x) + \frac{(4 + dl)^2 f(x) \|K\|_2^2}{2c^d (4 + d)(2 + dl)}, \quad n \to \infty,$$

for the respective values of l for any x such that $f(x) > 0$, where $x^+ := \max(x, 0)$, and the coefficient ρ is defined by the geometrically decrease of the strong mixing coefficient.

Note that if we specify the form of h_n, the (c) of the previous theorem is rewritten under the form:

(c′) *If $d \geqslant 3$ and $l \in [(1 - \frac{1}{2\nu d})^+, \frac{d-2}{2d}[$, the choice $h_n = C_n n^{-\nu}$, $C_n \to c > 0$, with $0 < \nu < \frac{1}{d+2}$, implies:*

$$nh_n^d \operatorname{Var} f_n^l(x) \to \frac{(1 - \nu d(1 - l))^2}{1 - \nu d(1 - 2l)} f(x) \int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u, \quad n \to \infty, \text{ if } \rho > 2.$$

Secondly, we classify $(f_n^l(x))$ according to the values of l with respect to various criteria. So we have:

Theorem. *Under the assumptions $\mathcal{K} - \mathcal{Q}$, $\rho > \max(2, \frac{d+2}{2})$, let $h_n = C_n n^{-\nu}$, $C_n \to c > 0$, $0 < \nu < \frac{1}{d+2}$. Then:*

(a) *the efficiency of $(f_n^l(x))$ is decreasing (resp. increasing) according to l in terms of the variance criterion (resp. bias criterion).*

(b) *If $f(x) > 0$, and $\nu = \frac{1}{d+4}$, the efficiency of $(f_n^l(x))$ is increasing according to l in terms of MSE criterion.*

Finally, we compare our family of estimators with the standard Parzen–Rozenblatt's estimator $(f_n^{\mathrm{PR}}(x))$ using the three criteria:

Theorem. *Under the assumptions $\mathcal{K} - \mathcal{Q}$, $\rho > \max(2, \frac{d+2}{2})$, let $h_n = C_n n^{-\nu}$, $C_n \to c > 0$, $0 < \nu < \frac{1}{d+2}$. Then:*

(a) *all estimators $f_n^l(x)$, $l \in [0, 1]$ are preferable than $f_n^{\mathrm{PR}}(x)$ in terms of the variance criterion.*
(b) *no estimator in the family $f_n^l(x)$, $l \in [0, 1]$ is preferable than $f_n^{\mathrm{PR}}(x)$ in terms of the bias criterion.*
(c) *for $d = 1$, if $f(x) > 0$, and $\nu = \frac{1}{d+4}$, $f_n^{\mathrm{PR}}(x)$ is preferable than all $f_n^l(x)$, $l \in [0, 1]$ in term of the MSE criterion.*

1. Introduction

Soit $(X_t, t \in \mathbb{N})$ un processus à valeurs dans $\mathbb{R}^d$, α-mélangeant. Les X_t sont supposés équidistribués de densité f inconnue. Parmi les estimateurs les plus utilisés pour estimer f à partir des observations $X_1, \ldots, X_n$, il y a les histogrammes et les polygones de fréquences. L'histogramme mobile est un cas particulier du célèbre estimateur à noyau introduit par Rosenblatt [9] et Parzen [8] défini par : $f_n^{\mathrm{PR}}(x) := \frac{1}{nh_n^d} \sum_{i=1}^n K(\frac{x-X_i}{h_n})$. L'étude de cet estimateur a donné lieu à une vaste littérature statistique. Les vitesses de convergence de son erreur quadratique ont été étudiées dans le cas ponctuel par Bosq [2]. Nous nous intéressons aux estimateurs récursifs introduits pour la première fois par Wolverton Wagner [11] et Yamato [12] sous la forme : $f_n^{\mathrm{WW}}(x) := \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^d} K(\frac{x-X_i}{h_i})$. Cet estimateur possède en effet, les mêmes propriétés asymptotiques que $f_n^{\mathrm{PR}}(x)$ et il peut également s'écrire sous la forme : $f_n^{\mathrm{WW}}(x) = R_n(x, f_{n-1}^{\mathrm{WW}}(x), X_n)$, avec $R_n(x, a, b) = \frac{n-1}{n} a + \frac{1}{h_n^d} K(\frac{x-b}{h_n})$. L'intérêt d'une telle relation est que l'ajout d'une nouvelle observation ne conduit pas à recalculer intégralement l'estimateur : il suffit d'utiliser sa valeur à l'instant $n-1$ et la n^{ieme} observation pour déduire son expression à l'instant n. C'est donc une relation très intéressante surtout si on traite un nombre important de données et qui évolue au fur et à mesure. De nombreuses variantes récursives ont également été proposées et étudiées depuis. En particulier, Deheuvels [4–6] s'est intéressé à la famille suivante :

$$f_n^H(x) := \left(\sum_{i=1}^n h_i H(h_i)\right)^{-1} \sum_{i=1}^n H(h_i) K\left(\frac{x - X_i}{h_i}\right).$$

L'estimateur $f_n^{\mathrm{DHV}}(x)$ obtenu dans le cas où $H(u) = 1$ est connu dans la littérature sous le nom d'estimateur de Deheuvels. Aussi, Wegman–Davis [10] étudient l'estimateur récursif suivant :

$$f_n^{\mathrm{DW}}(x) := \frac{1}{n\sqrt{h_n^d}} \sum_{i=1}^n \frac{1}{h_i^{d/2}} K\left(\frac{x - X_i}{h_i}\right).$$

Leur idée consiste à partager le paramètre de lissage de l'estimateur à noyau habituel en deux puissances $\frac{1}{2}$. Cet estimateur est asymptotiquement biaisé et son intégrale n'est pas égale à 1, mais il est très intéressant car une fois ces problèmes corrigés, sa variance asymptotique est plus petite que celle de $f_n^{\mathrm{WW}}(x)$.

2. Présentation de l'estimateur

Nous proposons la sous-famille paramétrique d'estimateurs récursifs à noyau définie par :

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

qui correspond pour $d = 1$ au cas $H(u) = u^{-l}$. Remarquons que notre famille contient les estimateurs $f_n^{\mathrm{DHV}}(x)$ (avec $l = 0$), $f_n^{\mathrm{WW}}(x)$ (avec $l = 1$) et la version asymptotiquement débiaisée de $f_n^{\mathrm{DW}}(x)$, (avec $l = \frac{1}{2}$).

Nous donnons ici, les biais, variance et erreur quadratique asymptotiques exacts de $f_n^l(x)$, en fonction de l, ensuite nous introduisons trois critères de comparaison qui nous permettent de préférer ou non notre sous-famille à l'estimateur à noyau habituel et aussi de classer nos estimateurs en fonction de la valeur de l.

Relation récursive. Pour tout $l \in [0,1]$, on a la relation récursive suivante :

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

3. Hypothèses et résultats

Hypothèse $\mathcal{K}$.

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

Hypothèse $\mathcal{H}$.

(i) h_n est une suite réelle qui décroît vers 0 et $nh_n^{d+2} \to \infty$ lorsque $n \to \infty$;
(ii) $B_{n,r} := \frac{1}{n}\sum_{i=1}^n (\frac{h_i}{h_n})^r \to \beta_r < \infty$, $n \to \infty$ $\forall r \in]-\infty, d+2]$.

Hypothèse $\mathcal{P}$. $f \in C_d^2(b)$, où $C_d^2(b)$ désigne l'ensemble des fonctions $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ telles que $\psi^{(2)}$ existe pour toute dérivée partielle d'ordre 2, continue et bornée.

Hypothèse $\mathcal{Q}$.

(i) Le processus (X_t) est $2-\alpha$-mélangeant avec : $\alpha^{(2)}(k) \leqslant \gamma k^{-\rho}$, $k \geqslant 1$ 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

$$\sup_{|s-t|\geqslant 1} \|g_{s,t}\|_\infty < \infty$$

où $g_{s,t} := f_{(X_s,X_t)} - f \otimes f$.

Les hypothèses $\mathcal{P}$ et $\mathcal{Q}$ sont classiques dans ce domaine : en particulier, il n'y a pas d'hypothèse de stationnarité de second ordre statuée sur le processus. On les rencontre par exemple dans Bosq–Blanke [3]. L'hypothèse $\mathcal{H}$(ii) est très utile dans nos calculs, et est propre à la récursivité.

Nous pouvons maintenant déterminer les biais, variance et erreur quadratique asymptotiques de notre famille d'estimateurs :

Théorème 3.1. *Sous les hypothèse $\mathcal{K}$, $\mathcal{H}$, $\mathcal{P}$, et $\mathcal{Q}$:*

(a) $h_n^{-4}(\mathrm{E} f_n^l(x) - f(x))^2 \underset{n\to\infty}{\longrightarrow} (\frac{\beta_{d(1-l)+2}}{\beta_{d(1-l)}})^2 b_2^2(x)$, *avec* $b_2^2(x) := \frac{1}{4}(\sum_{1\leqslant i,j\leqslant d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v)\,\mathrm{d}v)^2$.
(b) *Pour tout* $l \in [(\frac{d-2}{2d})^+, 1]$, $nh_n^d \operatorname{Var} f_n^l(x) \to \frac{\beta_{d(1-2l)}}{\beta_{d(1-l)}^2} f(x) \int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u$, $n \to \infty$, *si* $\rho > 2$, *où :* $x^+ = \max(x, 0)$.
(c) *Si* $d \geqslant 3$ *et* $l \in [0, \frac{d-2}{2d}[$, *la conclusion du* (b) *reste encore vraie si* $\rho > \frac{d+2}{2}$.
(d) *Sous les conditions du* (b) *(avec* $\rho > 2$*) ou du* (c) *(avec* $\rho > \frac{d+2}{2}$*), le choix* $h_n = C_n n^{-\frac{1}{d+4}}$, $C_n \to c > 0$, *entraîne que*

$$n^{\frac{4}{d+4}} \mathrm{E}\big(f_n^l(x) - f(x)\big)^2 \longrightarrow c^4\left(\frac{4+dl}{2+dl}\right)^2 b_2^2(x) + \frac{(4+dl)^2 f(x)\|K\|_2^2}{2c^d(4+d)(2+dl)}, \quad n \to \infty,$$

pour les valeurs respectives de l *en tout point où* $f(x) > 0$.

Notons que si l'on précise la forme de h_n, le résultat (c) du Théorème 3.1 se réécrit sous la forme :

(c′) *Si* $d \geqslant 3$ *et* $l \in [(1-\frac{1}{2\nu d})^+, \frac{d-2}{2d}[$*, le choix* $h_n = C_n n^{-\nu}$*,* $C_n \to c > 0$*, avec* $0 < \nu < \frac{1}{d+2}$*, entraîne :*

$$nh_n^d \operatorname{Var} f_n^l(x) \to \frac{(1-\nu d(1-l))^2}{1-\nu d(1-2l)} f(x) \int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u, \quad n \to \infty, \text{ si } \rho > 2.$$

4. Comparaison d'estimateurs

Définition 4.1 *(Critères de comparaison).* Soient $f_n(x)$ et $g_n(x)$ deux estimateurs à noyau de f.

(i) On dira que $f_n(x)$ est préférable à $g_n(x)$ au sens de la variance (et on notera $f_n(x) \prec g_n(x)$) si :

$$0 \leqslant \lim_{n\to\infty} \frac{\operatorname{Var}(f_n(x))}{\operatorname{Var}(g_n(x))} < 1.$$

(ii) On dira que $f_n(x)$ est préférable à $g_n(x)$ au sens du biais (et on notera $f_n(x) \ltimes g_n(x)$) si :

$$0 \leqslant \lim_{n\to\infty} \frac{(Ef_n(x)-f(x))^2}{(Eg_n(x)-f(x))^2} < 1.$$

(iii) On suppose que $f(x) > 0$, on choisit : $h_n = C_n n^{-\frac{1}{d+4}}$, $C_n \to c > 0$, avec $c = c_{\min}(f_n(x))$ (resp. $c = c_{\min}(g_n(x))$) pour l'estimateur $f_n(x)$ (resp. $g_n(x)$). $c_{\min}(\diamond)$ désigne la constante qui minimise l'erreur quadratique asymptotique de l'estimateur $\diamond$. Sous ces conditions, on dira que $f_n(x)$ est préférable à $g_n(x)$ au sens de la moyenne quadratique (ou du *MSE* et on notera $f_n(x) \propto g_n(x)$) si : $0 \leqslant \lim_{n\to\infty} \frac{E(f_n(x)-f(x))^2}{E(g_n(x)-f(x))^2} < 1$.

Le critère (i) a été introduit par Banon [1]. Notre premier résultat de cette partie permet de classifier nos estimateurs selon les valeurs de l par les critères précédents :

Théorème 4.2. *On suppose que les hypothèses* $\mathcal{K}-\mathcal{Q}$ *sont vérifiées avec* $\rho > \max(2, \frac{d+2}{2})$*. On choisit* $h_n = C_n n^{-\nu}$*,* $C_n \to c > 0$*,* $0 < \nu < \frac{1}{d+2}$*. Alors :*

(a) *pour tous* $l_1, l_2 \in [0,1]$*,* $f_n^{l_1}(x) \prec f_n^{l_2}(x)$ *(resp.* $f_n^{l_2}(x) \ltimes f_n^{l_1(x)}$*) si* $l_1 < l_2$*.*
(b) *Si* $f(x) > 0$*, et* $\nu = \frac{1}{d+4}$*, pour tous* $l_1, l_2 \in [0,1]$*,* $f_n^{l_2}(x) \propto f_n^{l_1(x)}$ *si* $l_1 < l_2$*.*

Notre deuxième résultat compare notre famille d'estimateurs à l'estimateur à noyau usuel $f_n^{\mathrm{PR}}(x)$:

Théorème 4.3. *On se place sous les hypothèses du Théorème* 4.2*. Alors* :

(a) *tous les estimateurs* $f_n^l(x)$*,* $l \in [0,1]$ *sont préférables à* $f_n^{\mathrm{PR}}(x)$ *au sens de la variance*;
(b) *aucun estimateur* $f_n^l(x)$*,* $l \in [0,1]$ *n'est préférable à* $f_n^{\mathrm{PR}}(x)$ *au sens du biais*;
(c) *pour* $d = 1$*, si* $f(x) > 0$*, et* $\nu = \frac{1}{d+4}$*,* $f_n^{\mathrm{PR}}(x)$ *est préférable à tous les estimateurs* $f_n^l(x)$*,* $l \in [0,1]$ *au sens du* $M.S.E$*, pour les choix « optimaux » respectifs de* c*.*

On retiendra que $f_n^0(x)$ est optimal dans la famille $f_n^l(x)$, pour le critère de la variance. Ceci est sans surprise car Deheuvels [6] avait déjà montré que cet estimateur était optimal pour la variance sur la famille des $f_n^H(x)$. En revanche, cet même estimateur est le « moins préférable » de la famille $f_n^l(x)$, au sens du biais. Pour ce critère, c'est $f_n^1(x)$ qui est optimal dans la famille $f_n^l(x)$.

5. Indications pour les preuves

Théorème 3.1. (a) : Le biais se traite par un développement de Taylor (hyp. $\mathcal{P}$), le résultat suit grâce à l'hypothèse $\mathcal{H}$ par application du théorème de convergence dominée suivi du lemme de Toeplitz.

(b) et (c) : On décompose la variance de l'estimateur en un terme principal et un terme de covariance. L'hypothèse $\mathcal{H}$, le lemme de Bochner et l'application du lemme de Toeplitz (cf. Masry [7]) entraînent le résultat annoncé (les termes de covariance étant négligeables et contrôlés par l'hypothèse $\mathcal{Q}$).

Le (d) découle de la décomposition de l'erreur quadratique en biais au carré plus variance.

Théorème 4.2. (a) et (b) : Minimisation en l des variance, biais et MSE asymptotiques de $f_n^l(x)$.

Théorème 3.3. (a) (b) et (c) : Etude du rapport des biais, variances et MSE des $f_n^l(x)$ par ceux de $f_n^{\mathrm{PR}}(x)$.

Références

[1] G. Banon, Sur un estimateur non paramétrique de la densité de probabilité, Revue de Statistique Appliquée 24 (4) (1976) 61–73.
[2] D. Bosq, Nonparametric Statistics for Stochastic Processes. Estimation and Prediction, 2nd edn., Lecture Notes in Statistics, vol. 110, Springer-Verlag, New York, 1998.
[3] D. Bosq, D. Blanke, Inference and Prediction in Large Dimensions, Wiley Series in Probability and Statistics, ISBN 978-0-470-08147-1, 2007.
[4] P. Deheuvels, Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, C. R. Acad. Sci. Paris 278 (1973) 1217–1220.
[5] P. Deheuvels, Sur l'Estimation séquentielle de la densité, C. R. Acad. Sci. Paris, Ser. A 276 (1973) 1119–1121.
[6] P. Deheuvels, Sur une famille d'estimateurs de la densité d'une variable aléatoire, C. R. Acad. Sci. Paris, Ser. A 276 (1974) 1013–1015.
[7] E. Masry, Recursive Probability Density Estimation for Weakly Dependent Stationary Processes, IEEE, 1986.
[8] E. Parzen, On the estimation of a probability density function and the mode, Ann. Math. Stat. 33 (1962) 1065–1076.
[9] M. Rozenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Stat. 27 (1956) 832–837.
[10] E.J. Wegman, H.I. Davis, Remarks on some recursive estimators of a probability density, Ann. Statist. 7 (2) (1979) 316–327.
[11] C.T. Wolverton, T.J. Wagner, Recursive estimates of probability densities, IEEE Trans. on Systems Sciences and Cybernetics 5 (1969) 307.
[12] H. Yamato, Sequential estimation of a continuous probability density function and mode, Bull. Math. Statist. Jap. 14 (1972) 1–12.