

Copyright © 1976, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

SUR UN ESTIMATEUR NON PARAMÉTRIQUE  
DE LA DENSITÉ DE PROBALILITÉ

by

G. Banon

Memorandum No. ERL-M606

27 October 1976

ELECTRONICS RESEARCH LABORATORY

College of Engineering  
University of California, Berkeley  
94720

Abstract

The usual Kernel estimates of the probability density function, those of Rosenblatt-Parzen, Yamato and Deheuvels are presented with their quadratic mean convergence assumptions.

A criterium based on the asymptotic behavior of the variance is given to compare them.

Using additional assumptions, it is shown that such criterium orders these three estimates.

Finally, a new recursive estimate is introduced for which the asymptotic behavior of the bias and the quadratic mean convergence are studied. Using the above criterium, it is shown that such estimate is better than the previous ones.

### Résumé

On rappelle dans cette note les estimateurs connus de la densité de probabilité, ceux de Rosenblatt-Parzen, Yamato et Deheuvels, ainsi que les conditions respectives de convergence en moyenne quadratique.

On introduit un critère de comparaison fondé sur le comportement asymptotique de la variance lorsque le nombre d'observations devient infini. Moyennant l'ajonction d'hypothèses supplémentaires on montre que le critère choisi permet d'ordonner les trois estimateurs.

Enfin on introduit un nouvel estimateur récurrent de la densité pour lequel on étudie le comportement du biais, on démontre la convergence en moyenne quadratique et on montre que sur la base du critère choisi ce dernier estimateur est meilleur que les précédents.

### Mots clés

Estimation, non paramétrique, récurrente, densité de probabilité.

Sommaire

|  |       |
|--|-------|
| 1) Nature du problème                      | p. 1  |
| 2) Rappel historique                       | p. 2  |
| 3) Définition d'un nouvel estimateur       | p. 6  |
| 4) Propriétés de l'estimateur $f_{4,n}(x)$ | p. 9  |
| Références                                 | p. 17 |
| Remerciements                              | p. 19 |
| Figures                                    | p. 20 |

1) Nature du problème

Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires réelles, indépendantes et identiquement distribuées suivant une fonction de répartition  $F(x)$  possédant une densité de probabilité  $f(x)$  bornée et continue sur  $\mathbb{R}$ :

$$F(x) = \int_{-\infty}^x f(y) dy. \quad (1-1)$$

Le problème auquel on s'intéresse ici est l'étude d'estimateurs  $f_n(x)$  de la densité de probabilité  $f(x)$ , fonction des variables aléatoires  $X_1, X_2, \dots, X_n$ , qui convergent en moyenne quadratique:

$$E(f_n(x) - f(x))^2 \xrightarrow{n \rightarrow \infty} 0, \quad (1-2)$$

ou encore, puisque:

$$E(f_n(x) - f(x))^2 = \text{Var } f_n(x) + (E f_n(x) - f(x))^2, \quad (1-3)$$

d'estimateurs tels que l'on ait simultanément:

$$E f_n(x) \xrightarrow{n \rightarrow \infty} f(x) \quad (1-4)$$

et

$$\text{Var } f_n(x) \xrightarrow{n \rightarrow \infty} 0. \quad (1-5)$$

Les estimateurs vérifiant la propriété (1-4) sont dits asymptotiquement sans biais.

Dans cette note, on se limitera à l'étude de la convergence en moyenne quadratique.

2) Rappel historique

Plusieurs estimateurs de la densité de probabilité ont été proposés par le passé. Chronologiquement on peut rappeler les travaux de M. Rosenblatt [8], puis ceux de E. Parzen [7] sur l'étude d'un estimateur de la forme:

$$f_{1,n}(x) = \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right), \quad (2-1)$$

où  $K(y)$  est une fonction mesurable sur  $\mathbb{R}$ , éventuellement non négative (condition nécessaire si l'on désire que  $f_n(x)$  soit une densité), bornée et vérifiant:

$$\int_{-\infty}^{\infty} K(y) dy = 1, \quad (2-2)$$

et où  $h_n$  est le terme général d'une suite numérique  $\{h_n\}$  positive telle que:

$$h_n \xrightarrow[n \rightarrow \infty]{} 0 \quad (2-3)$$

et

$$nh_n \xrightarrow[n \rightarrow \infty]{} \infty. \quad (2-4)$$

Sous ces conditions, on montre en particulier (voir par exemple [9] p. 1816) que  $f_{1,n}(x)$  converge en moyenne quadratique vers  $f(x)$ , et que (voir [9] p. 1818):

$$\text{Var } f_{1,n}(x) \underset{n \rightarrow \infty}{\sim} \frac{1}{nh_n} f(x) \int_{-\infty}^{\infty} K^2(y) dy. \quad (2-5)$$

On notera par  $C_1$  l'ensemble des couples  $(K(y), \{h_n\})$  tels que  $K(y)$  soit une fonction mesurable non négative, bornée et satisfaisant (2-2), et tels que  $\{h_n\}$  soit une suite numérique positive vérifiant (2-3) et (2-4).

En ce qui concerne le comportement asymptotique du biais  $E f_{1,n}(x) - f(x)$ , on démontre (voir par exemple [9] p. 1818) que si  $K(y)$  est de plus une fonction paire telle que

$$\int_{-\infty}^{\infty} y^2 K(y) dy < \infty, \quad (2-6)$$

et si  $f'(x)$  et  $f''(x)$  sont continues et bornées ( $f''(x) \neq 0$ ) alors:

$$E f_{1,n}(x) - f(x) \underset{n \rightarrow \infty}{\sim} \frac{1}{2} h_n^2 f''(x) \int_{-\infty}^{\infty} y^2 K(y) dy. \quad (2-7)$$

Des propriétés asymptotiques de la variance et du biais on peut déduire, sous les conditions qui ont conduit à (2-5) et (2-7), le comportement asymptotique de l'erreur quadratique moyenne de l'estimateur de Rosenblatt-Parzen pour  $n \rightarrow \infty$  :

$$E(f_{1,n}(x) - f(x))^2 = \frac{1}{nh_n} f(x) \int_{-\infty}^{\infty} K^2(y) dy + \frac{1}{4} h_n^4 f''^2(x) \left( \int_{-\infty}^{\infty} y^2 K(y) dy \right)^2 + o\left(\frac{1}{nh_n} + h_n^4\right). \quad (2-8)$$

#### Remarque 2-1

Si l'on prend  $h_n = n^{-s}$  avec  $s \in ]0, 1[$ , on déduit de (2-7) que le biais converge vers zéro comme  $n^{-2s}$ , et de (2-8) on conclut que la vitesse de convergence vers zéro de l'erreur quadratique moyenne est



maximum pour  $s = \frac{1}{5}$ . A partir de (2-8) on constate aussi que pour  $s > \frac{1}{5}$  le biais converge vers zéro plus rapidement que la variance et que par conséquent l'erreur quadratique moyenne se comporte alors asymptotiquement comme la variance. #

On peut rappeler aussi les travaux de H. Yamato [10] qui a introduit un estimateur récurrent de la densité satisfaisant l'équation:

$$f_n(x) = \frac{n-1}{n} f_{n-1}(x) + \frac{1}{nh_n} K\left(\frac{x-X_n}{h_n}\right), \quad (2-9)$$

pour  $n = 1, 2, \dots$ .

L'estimateur de Yamato solution de (2-9) s'écrit:

$$f_{2,n}(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right). \quad (2-10)$$

On démontre en particulier (voir [10] théorème 2 p. 4) que pour tout couple  $(K(y), \{h_n\})$  appartenant à  $C_2$  sous ensemble de  $C_1$  pour lequel les suites  $\{h_n\}$  sont décroissantes,  $f_{2,n}(x)$  converge en moyenne quadratique vers  $f(x)$ .

#### Remarque 2-2

L'hypothèse de suite décroissante pour  $\{h_n\}$  n'est pas en général une condition nécessaire pour assurer les propriétés de convergence, mais puisque c'est une hypothèse simplificatrice dans les démonstrations et que ce n'est pas une limitation très forte en pratique, on la conservera dans toute la suite de cette note. #

Si de plus  $(K(y), \{h_n\})$  appartient à  $C'_2$  sous ensemble de  $C_2$  pour lequel les suites  $\{h_n\}$  vérifient:

$$\frac{h_n}{n} \sum_{j=1}^n \frac{1}{h_j} \xrightarrow{n \rightarrow \infty} \alpha \quad (2-11)$$

avec  $\alpha \in ]0,1[$ , alors on montre (voir [10] théorème 3 p. 6) que:

$$\lim_{n \rightarrow \infty} n h_n \text{Var } f_{2,n}(x) = \alpha f(x) \int_{-\infty}^{\infty} K^2(y) dy. \quad (2-12)$$

Enfin, P. Deheuvels [3] a proposé un autre estimateur récurrent de la densité de la forme:

$$f_{3,n}(x) = \frac{1}{b_n} \sum_{j=1}^n K\left(x, \frac{x-X_j}{h_j}\right) \quad (2-13)$$

$$\text{avec } b_n = \sum_{j=1}^n h_j,$$

où  $K(x,y)$  est une fonction non négative, bornée, mesurable sur  $\mathbb{R}^2$  et vérifiant:

$$\int_{-\infty}^{\infty} K(x,y) dy = 1 \quad \forall x \in \mathbb{R} \quad (2-14)$$

et

$$\sup_{x \in \mathbb{R}} \int_{|y| > t} |y| K(x,y) dy \xrightarrow{t \rightarrow \infty} 0, \quad (2-15)$$

et où  $h_j$  est le terme général d'une suite numérique  $\{h_j\}$  positive, décroissante vers zéro et vérifiant:

$$b_n \xrightarrow{n \rightarrow \infty} \infty. \quad (2-16)$$

Parmi de nombreux autres résultats P. Deheuvels a démontré que sous les conditions précédentes  $f_{3,n}(x)$  converge en moyenne quadratique

vers  $f(x)$ , et qu'en particulier (voir [3] théorème 1 p. 1119):

$$\text{Var } \hat{f}_{3,n}(x) \underset{n \rightarrow \infty}{\sim} \frac{1}{b_n} f(x) \int_{-\infty}^{\infty} K^2(x,y) dy . \quad (2-17)$$

Remarque 2-3

Puisque l'on se limite ici au cas de la convergence en moyenne quadratique, on se restreindra dans la suite de ce travail au cas où la fonction  $K(x,y)$  ne dépend que de  $y$ . D'autre part, on pourrait vérifier qu'alors la condition (2-15) est superflue.

Comme dans le cas de l'estimateur de Rosenblatt-Parzen avec  $h_n = n^{-s}$ , on pourrait vérifier aussi (voir les résultats sur la convergence des espérances mathématiques p. 29 th. 6 dans [5]) que pour  $s > \frac{1}{5}$  les biais des estimateurs de Yamato et de Deheuvels convergent vers zéro plus rapidement que leurs variances. #

3) Définition d'un nouvel estimateur

Afin de dégager l'intérêt relatif des estimateurs de la densité présentés au paragraphe précédent on introduit maintenant un critère de comparaison fondé sur le comportement asymptotique de la variance.

Bien d'autres critères pourraient être choisis (convergence de l'erreur quadratique moyenne ou de son intégrale sur  $R$ , convergence presque sûre par exemple), les remarques 2-1 et 2-3 (deuxième paragraphe) font cependant le lien (et sont ainsi un motif de justification) entre la convergence de la variance et celle de l'erreur quadratique moyenne dans le cas pratique important où  $h_n = n^{-s}$  et où  $s$  doit être pris supérieur à  $\frac{1}{5}$ .

Définition 3-1 (critère de comparaison)

C désignant un sous ensemble produit de fonction  $K(y)$  mesurable sur  $R$  et de suite numérique  $\{h_n\}$ , on dira que relativement à C l'estimateur  $f$  est meilleur que l'estimateur  $g$  au sens de la variance et on notera  $f \prec g$  si et seulement si à tout couple  $(K(y), \{h_n\})$  de C égal il existe un réel non négatif  $a$  inférieur à l'unité tel que pour tout  $\epsilon$  positif il existe un entier positif  $N(\epsilon)$  tel que  $n > N$  implique  $\frac{\text{Var } f_n(x)}{\text{Var } g_n(x)} < a + \epsilon$ , en résumé, relativement à C

$$f \prec g \Leftrightarrow \exists a \in [0,1[; \forall \epsilon > 0, \exists N(\epsilon);$$

$$n > N \Rightarrow \frac{\text{Var } f_n(x)}{\text{Var } g_n(x)} < a + \epsilon. \quad \# \quad (3-1)$$

Sur la base de la définition 3-1 on établit les deux lemmes suivants:

Lemme 3-1 (transitivité)

Si relativement à C  $f \prec g$  et relativement à C'  $g \prec h$  alors relativement à  $C \cap C'$   $f \prec h$ . #

Lemme 3-2 (condition suffisante)

Si pour tout  $(K(y), \{h_n\})$  appartenant à C,  $\frac{\text{Var } f_n(x)}{\text{Var } g_n(x)} \leq u_n$  et  $\lim_{n \rightarrow \infty} u_n = b$  avec  $b \in [0,1[$  alors relativement à C  $f \prec g$ . #

Remarque 3-1

Avec la définition 3-1, pour comparer deux estimateurs, il n'est pas nécessaire que la limite de  $\frac{\text{Var } f_n(x)}{\text{Var } g_n(x)}$  existe, il suffit par exemple que le lemme 3-2 soit vérifié. Dans ce cas on vérifie que  $\lim_{n \rightarrow \infty} u_n$  est inférieure à l'unité plutôt que  $u_n$  inférieur à l'unité (ce qui conduirait

à une autre définition du critère de comparaison) en effet on peut se trouver parfois dans la situation où  $u_n < 1$  et  $\lim_{n \rightarrow \infty} u_n = 1$  comme le montre l'exemple suivant:

$$u_n = \frac{1}{n \log n} \sum_{j=2}^n \text{Log } j \quad \text{pour } n = 2, 3, \dots \quad (\text{Voir [10] p. 9}). \quad \#$$

En appliquant le critère de comparaison défini plus haut aux trois estimateurs déjà rencontrés on obtient le résultat suivant:

Proposition 3-1

$f_1$ ,  $f_2$  et  $f_3$  désignant respectivement les estimateurs de Rosenblatt-Parzen, Yamato et Deheuvels et  $C'_3$  étant le sous ensemble de  $C'_2$  pour lequel  $\{h_n\}$  vérifie

$$\frac{nh_n}{b_n} \xrightarrow[n \rightarrow \infty]{} \beta \quad \text{avec } \beta \in [0, \alpha[, \text{ alors} \quad (3-2)$$

$$\text{relativement à } C'_2 \text{ on a } f_2 \prec f_1, \quad (3-3)$$

$$\text{relativement à } C'_3 \text{ on a } f_3 \prec f_2, \quad (3-4)$$

$$\text{et } f_3 \prec f_1. \quad \# \quad (3-5)$$

Démonstration

De (2-5) et (2-12) on déduit que pour tout élément de  $C'_2$   $\lim_{n \rightarrow \infty} \frac{\text{Var } f_{2,n}(x)}{\text{Var } f_{1,n}(x)} = \alpha$ , en appliquant le lemme 3-2 avec  $b = \alpha$  on obtient (3-3). En rapprochant (2-12) et (2-17) on obtient que pour tout élément de  $C'_3$   $\lim_{n \rightarrow \infty} \frac{\text{Var } f_{3,n}(x)}{\text{Var } f_{2,n}(x)} = \frac{\beta}{\alpha}$ , or  $\frac{\beta}{\alpha} < 1$  aussi en appliquant le lemme 3-2 avec  $b = \frac{\beta}{\alpha}$  on obtient (3-4). (3-5) découle du lemme 3-1 puisque  $C'_3 \subset C'_2$ . #

La question que l'on se pose maintenant est de savoir s'il existe au moins un estimateur meilleur que l'estimateur de Deheuvels au sens de

la définition 3-1 pour au moins un ensemble non vide de couples  $(K(y), \{h_n\})$ .

Au prix d'une complexité accrue, la réponse est affirmative, on est dans cette situation dans le cas suivant:

$$f_{4,n}(x) = \frac{1}{b_n} \sum_{j=1}^n \frac{h_j}{b_j} \sum_{k=1}^j K\left(\frac{x-X_j}{h_k}\right), \quad (3-6)$$

avec  $b_n = \sum_{j=1}^n h_j$  comme pour l'estimateur  $f_{3,n}(x)$ .

On constate que l'estimateur donné par l'expression (3-6) est récurrent et vérifie l'équation:

$$f_n(x) = \frac{b_{n-1}}{b_n} f_{n-1}(x) + \frac{h_n}{b_n^2} \sum_{k=1}^n K\left(\frac{x-X_n}{h_k}\right), \quad (3-7)$$

pour  $n = 1, 2, \dots$ , avec  $b_0 = 0$  par convention.

#### Remarque 3-2

Dans le cas du noyau unité  $(K(y) = \frac{1}{2} 1_{A} ]-1,1])$ , où  $1_A$  est la fonction indicatrice de l'ensemble A) et de la suite  $\{h_n\}$  de terme général  $h_n = \frac{1}{\sqrt{n}}$ , la contribution (additive) de la  $n^{\text{ième}}$  variable aléatoire  $X_n$  dans l'expression (2-13) se calcule à partir de la fonction représentée sur la figure 3-1, cette contribution dans l'expression (3-6) se calcule, par contre, à partir de la fonction représentée sur la figure 3-2. #

#### 4) Propriétés de l'estimateur $f_{4,n}(x)$

Dans la suite de ce travail on va définir des conditions suffisantes sur la fonction  $K(y)$  et la suite  $\{h_n\}$  pour que  $f_{4,n}(x)$  converge en moyenne quadratique vers  $f(x)$  et ait la propriété énoncée ci dessus. On vérifiera que ces conditions sont satisfaites pour au moins une classe importante de cas pratiques. En outre on étudiera le comportement asymptotique du biais.

Théorème 4-1 (convergence asymptotique sans biais).

Si  $(K(y), \{h_n\}) \in C_2$  alors

$$E f_{4,n}(x) \xrightarrow[n \rightarrow \infty]{} f(x). \quad \# \quad (4-1)$$

Démonstration

D'après la définition de  $f_{4,n}(x)$  on a:

$$E f_{4,n}(x) = \frac{1}{b_n} \sum_{j=1}^n \frac{h_j}{b_j} \sum_{k=1}^j EK\left(\frac{x-X_{j-1}}{h_k}\right). \quad (4-2)$$

D'autre part en remarquant que  $\frac{1}{h_n} EK\left(\frac{x-X_1}{h_n}\right) = E f_{1,n}(x)$  on peut appliquer un résultat partiel établi par M. Rosenblatt (voir [9] p. 1816 expression (7)) et obtenir que pour tout élément de  $C_2$ :

$$\frac{1}{h_n} EK\left(\frac{x-X_1}{h_n}\right) \xrightarrow[n \rightarrow \infty]{} f(x). \quad (4-3)$$

En utilisant (4-3) et en appliquant deux fois le lemme de Toeplitz rappelé ci-dessous on trouve que la limite de l'expression (4-2) est bien  $f(x)$ . #

Lemme 4-1 (lemme de Toeplitz - énoncé partiel)

Si  $b_n = \sum_{k=1}^n a_k \xrightarrow[n \rightarrow \infty]{} \infty$ , alors  $x_n \xrightarrow[n \rightarrow \infty]{} x$  implique que  $\frac{1}{b_n} \sum_{k=1}^n a_k x_k \xrightarrow[n \rightarrow \infty]{} x$ . #

(Pour l'énoncé complet et la démonstration du lemme voir [6] p. 238).

A propos du comportement asymptotique du biais on peut établir le résultat particulier suivant:

Théorème 4-2

Si  $(K(y), \{h_n\}) \in C_2$ , si de plus  $K(y)$  est une fonction paire satisfaisant

$$\text{Var } f_{4,n}(x) \leq \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^j \frac{1}{h_j} \frac{1}{h_k} (2j-2k+1) E K^2 \left( \frac{h^k}{x-X_1} \right) \quad (4-7) \quad \#$$

sur les intervalles y positif et y négatif, alors:  
 Si  $(K(y), \{h_n\}) \in C_4$  sous ensemble de  $C_2$  tel que  $K(y)$  soit monotone  
 Lemme 4-2 (majoration de la variance)

démontrer le lemme suivant:

Avant d'étudier la convergence en moyenne quadratique, on va  
 le biais converge vers zéro comme  $n^{-2}$  (voir remarque 2-1).  
 on constate que comme dans le cas de l'estimateur de Rosenblatt-Parzen

$$\sum_{j=1}^n \frac{1-s}{n} \sim \frac{1-s}{n} \quad \text{As} > 1, \quad (4-6)$$

Si on prend  $h_n = n^{-s}$  avec  $s \in ]0,1[$  et si on utilise la propriété:

Remarque 4-1

en rapprochant (4-5) et (4-2) on établit alors (4-4).  
 #

$$E K \left( \frac{h^n}{x-X_1} \right) = h^n f(x) + \frac{1}{2} h^3 f''(x) \int_{-\infty}^{\infty} y^2 K(y) dy + o(h^n), \quad (4-5)$$

x on obtient pour  $n \rightarrow \infty$ :

Sous les conditions du théorème 4-2, en développant  $f(x)$  au voisinage de

Démonstration

(4-4)

$$E f_{4,n}(x) - f(x) \sim \frac{1}{2} f''(x) \int_{-\infty}^{\infty} y^2 K(y) dy \left( \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^j \frac{1}{h_j} \frac{1}{h_k} \right) \quad \#$$

(2-6) et si  $f'(x)$  et  $f''(x)$  sont continues et bornées ( $f''(x) \neq 0$ ) alors:



Démonstration

D'après la définition de  $f_{4,n}(x)$  et à cause de l'indépendance des  $X_j$  on a :

$$\text{Var } f_{4,n}(x) \leq \frac{1}{b_n^2} \sum_{j=1}^n \frac{h_j^2}{b_j^2} \sum_{k,\ell=1}^j \text{EK}^2\left(\frac{x-X_1}{h_k}\right) \text{K}\left(\frac{x-X_1}{h_\ell}\right). \quad (4-8)$$

La monotonie de  $K(y)$  et  $\{h_n\}$  implique l'inégalité :

$$\text{EK}\left(\frac{x-X_1}{h_k}\right) \text{K}\left(\frac{x-X_1}{h_\ell}\right) \leq \text{EK}^2\left(\frac{x-X_1}{h_{\min(k,\ell)}}\right). \quad (4-9)$$

En considérant (4-8) et (4-9) simultanément on obtient :

$$\text{Var } f_{4,n}(x) \leq \frac{1}{b_n^2} \sum_{j=1}^n \frac{h_j^2}{b_j^2} \sum_{k,\ell=1}^j \text{EK}^2\left(\frac{x-X_1}{h_{\min(k,\ell)}}\right), \quad (4-10)$$

ce qui implique (4-7). #

Théorème 4-3 (convergence en moyenne quadratique)

Si  $(K(y), \{h_n\}) \in C_4$  alors :

$$E(f_{4,n}(x) - f(x))^2 \xrightarrow[n \rightarrow \infty]{} 0. \quad \# \quad (4-11)$$

Démonstration

Puisque d'après le théorème 4-1 le biais converge vers zéro relativement à  $C_1 \supset C_4$ , il suffit de démontrer maintenant que la variance converge vers zéro. A partir de l'expression (4-7) du lemme précédent on obtient les majorants suivant pour la variance :

$$\text{Var } f_{4,n}(x) \leq \frac{2}{b_n^2} \sum_{j=1}^n \frac{j h_j^2}{b_j^2} \sum_{k=1}^j \text{EK}^2\left(\frac{x-X_1}{h_k}\right) \quad (4-12)$$

$$\leq \frac{2}{b_n^2} \sum_{j=1}^n \frac{h_j}{b_j} \sum_{k=1}^j EK^2\left(\frac{x-X_1}{h_k}\right), \quad (4-13)$$

(4-13) découlant de la monotonie de  $\{h_n\}$ . Enfin, soit  $M$  tel que

$\sup_{x \in \mathbb{R}} f(x) \leq M$ , on a alors:

$$EK^2\left(\frac{x-X_1}{h_k}\right) \leq h_k M \int_{-\infty}^{\infty} K^2(y) dy. \quad (4-14)$$

Finalement en rapprochant (4-13) et (4-14) on trouve:

$$\text{Var } f_{4,n}(x) \leq \frac{2M}{b_n} \int_{-\infty}^{\infty} K^2(y) dy, \quad (4-15)$$

ce qui implique la convergence vers zéro de la variance lorsque  $n$  devient infini. #

Afin de comparer  $f_{4,n}(x)$  avec les estimateurs présentés au paragraphe 2, on va démontrer le lemme suivant:

#### Lemme 4-3

Si  $(K(y), \{h_n\}) \in C'_3$  et si la suite  $\{h_n\}$  vérifie:

$$\frac{h_n}{b_n^2} \sum_{j=1}^n j h_j \xrightarrow{n \rightarrow \infty} \gamma, \text{ alors:} \quad (4-16)$$

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{j=1}^n \frac{h_j^2}{b_j^2} \sum_{k=1}^j (2j-2k+1) EK^2\left(\frac{x-X_1}{h_k}\right) = 2(\beta-\gamma) f(x) \int_{-\infty}^{\infty} K^2(y) dy. \quad \# \quad (4-17)$$

#### Démonstration

On écrit le terme de gauche de l'expression (4-17) sous la forme:

$$u_n = \frac{1}{b_n} \sum_{j=1}^n h_j v_j \quad (4-18)$$

avec

$$v_j = \frac{h_j}{b_j^2} \sum_{k=1}^j (2j-2k+1) EK^2\left(\frac{x-X_1}{h_k}\right) . \quad (4-19)$$

En remarquant que (4-3) peut s'étendre au cas des moments supérieurs à l'unité, on a en particulier:

$$\frac{1}{h_n} EK^2\left(\frac{x-X_1}{h_n}\right) \xrightarrow{n \rightarrow \infty} f(x) \int_{-\infty}^{\infty} K^2(y) dy \quad (4-20)$$

et en appliquant trois fois le lemme de Toeplitz dans l'expression (4-19) on établit que:

$$v_j \underset{j \rightarrow \infty}{\sim} 2 \left( \frac{j h_j}{b_j} - \frac{h_j}{b_j^2} \sum_{k=1}^j k h_k \right) f(x) \int_{-\infty}^{\infty} K^2(y) dy. \quad (4-21)$$

En tenant compte de (3-2) et (4-16) on déduit de (4-21) que:

$$\lim_{j \rightarrow \infty} v_j = 2(\beta - \gamma) f(x) \int_{-\infty}^{\infty} K^2(y) dy. \quad (4-22)$$

Enfin en rapprochant (4-18) et (4-22) et en appliquant une dernière fois le lemme de Toeplitz on trouve (4-17). #

Proposition 4-1

$f_4$  désignant l'estimateur donné par (3-6) et  $C'_4$  étant le sous ensemble de  $C'_3 \cap C_4$  pour lequel  $\{h_n\}$  vérifie (4-16) avec  $\gamma \in ]\beta - \frac{1}{2}, \beta]$ , alors relativement à  $C'_4$  on a:

$$f_4 < f_3, \quad (4-23)$$

$$f_4 < f_2, \quad (4-24)$$

et

$$f_4 < f_1. \quad \# \quad (4-25)$$

### Démonstration

Le lemme 4-2 permet d'établir que:

$$\frac{\text{Var } f_{4,n}(x)}{\text{Var } f_{3,n}(x)} \leq u_n$$

$$\text{avec } u_n = \frac{1}{\text{Var } f_{3,n}(x)} \frac{1}{b_n^2} \sum_{j=1}^n \frac{h_j^2}{b_j^2} \sum_{k=1}^j (2j-2k+1) \text{EK}^2\left(\frac{x-X_1}{h_k}\right). \quad (4-26)$$

Le lemme 4-3 et (2-17) permettent de conclure que:

$$\lim_{n \rightarrow \infty} u_n = 2(\beta - \gamma). \quad (4-27)$$

Avec l'hypothèse sur  $\gamma$  on a  $2(\beta - \gamma) \in [0, 1[$ , aussi en appliquant le lemme 3-2 avec  $b = 2(\beta - \gamma)$  on obtient (4-23), (4-24) et (4-25) découlent du lemme 3-1 sur la transitivité. #

A l'aide d'exemples on vérifie maintenant que  $C'_4$  n'est pas l'ensemble vide.

Les conditions de monotonie sur  $K(y)$  sont vérifiées pour une classe importante de fonction densité. Une telle classe contient par exemple les fonctions densité correspondant au cas des distributions uniforme, triangulaire, gaussienne, exponentielle et de Cauchy.

Quant aux conditions sur  $\{h_n\}$ , elles sont vérifiées par exemple dans le cas pratique important où  $h_n = n^{-s}$  avec  $s \in ]0, 1[$ . En effet, en utilisant une fois de plus la propriété:

$$\frac{1}{n^{1-s}} \sum_{j=1}^n j^{-s} \xrightarrow{n \rightarrow \infty} \frac{1}{1-s} \quad \forall s < 1 \quad (4-28)$$

on trouve que:

$$\alpha = \frac{1}{s+1}, \quad (4-29)$$

$$\beta = 1-s, \quad (4-30)$$

et

$$\gamma = \frac{(1-s)^2}{2-s}. \quad (4-31)$$

On peut vérifier, que pour tout  $s \in ]0,1[$ ,  $\alpha$ ,  $\beta$  et  $\gamma$  appartiennent bien aux intervalles  $]0,1[$ ,  $]0,\alpha[$  et  $]\beta - \frac{1}{2}, \beta[$  respectivement.

Remarque 4-2

A partir de (4-30) et (4-31) on trouve que  $2(\beta-\gamma) = 2 \frac{1-s}{2-s}$ , ce qui permet d'apprécier quantitativement le gain obtenu en utilisant  $f_{4,n}(x)$  au lieu de  $f_{3,n}(x)$  pour les suites  $\{h_n\}$  du type  $h_n = n^{-s}$  (Voir (4-27)).

On doit bien noter qu'on obtient (à cause de l'inégalité (4-9)) une évaluation pessimiste de ce gain.

Dans le cas où  $K(y)$  correspond à la "fenêtre rectangulaire" (noyau unité) où à la "fenêtre triangulaire" on peut faire le calcul exact de

$$\text{la limite de } \frac{\text{Var } f_{4,n}(x)}{\text{Var } f_{3,n}(x)}.$$

Tout calcul fait on aboutit aux résultats suivants:

$$\text{pour } K(y) = \frac{1}{2} \mathbb{1}_{]-1,1]}(y) \quad \lim_{n \rightarrow \infty} \frac{\text{Var } f_{4,n}(x)}{\text{Var } f_{3,n}(x)} = 2 \frac{(1-s)^2}{2-s}, \quad (4-32)$$

$$\text{pour } k(y) = (1-|y|) \mathbb{1}_{]-1,1]}(y) \quad \lim_{n \rightarrow \infty} \frac{\text{Var } f_{4,n}(x)}{\text{Var } f_{3,n}(x)} = \frac{(1-s)^2(3s+2)}{(2-s)(1+s)}. \quad (4-33)$$

Si on choisit  $s = \frac{1}{2}$ , la limite apparaissant dans l'expression (4-32) vaut  $\frac{1}{3}$ , celle apparaissant dans (4-33) vaut  $\frac{7}{18}$ . #

Références

- [1] DEHEUVELS P.  
Sur une application de la théorie des processus de renouvellement  
à l'estimation de la densité d'une variable aléatoire.  
C.R. Acad. Sc. Paris, (26 Mars 1973), t. 276, série A, p. 943-946.
- [2] DEHEUVELS P.  
Sur une famille d'estimateurs de la densité d'une variable aléatoire.  
C.R. Acad. Sc. Paris, (2 Avril 1973), t. 276, série A, p. 1013-1015.
- [3] DEHEUVELS P.  
Sur l'estimation séquentielle de la densité.  
C.R. Acad. Sc. Paris, (16 Avril 1973), t. 276, série A, p. 1119-1121.
- [4] DEHEUVELS P.  
Conditions nécessaires et suffisantes de convergence ponctuelle  
presque sûre et uniforme presque sûre des estimateurs de la densité.  
C.R. Acad. Sc. Paris, (29 Avril 1974), t. 278, série A, p. 1217-1220.
- [5] DEHEUVELS P.  
Estimation séquentielle de la densité  
Thèse présentée à l'Université de Paris VI, 1974, p. 1-121.
- [6] LOEVE M.  
Probability Theory.  
Third Edition, Van Nostrand Reinhold, 1963, p. 238.
- [7] PARZEN E.  
On the estimation of a probability density function and the mode.  
Ann. Math. Statist., 1962, vol. 33, p. 1065-1076.
- [8] ROSENBLATT M.  
Remarks on some non parametric estimates of a density function.  
Ann. Math. Statist., 1956, vol. 27, p. 832-837.

[9] ROSENBLATT M.

Curve Estimates

Ann. Math. Statist., 1971, vol. 42, p. 1815-1842.

[10] YAMATO H.

Sequential estimation of a continuous probability density function  
and mode.

Bull. Math. Statist. Jap., 1972, vol. 14, p. 1-12.

Remerciements

L'auteur tient ici à remercier Monsieur Chevalier pour ses remarques très précieuses qui ont permis d'améliorer substantiellement la première version de cette note.

Ce travail a été terminé à l'Université de Californie (Berkeley) grâce à une bourse (No. 01P75-04371) de la NSF (National Science Foundation) dans le cadre d'un programme d'échange entre l'Electronic Research Laboratory et le Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS.



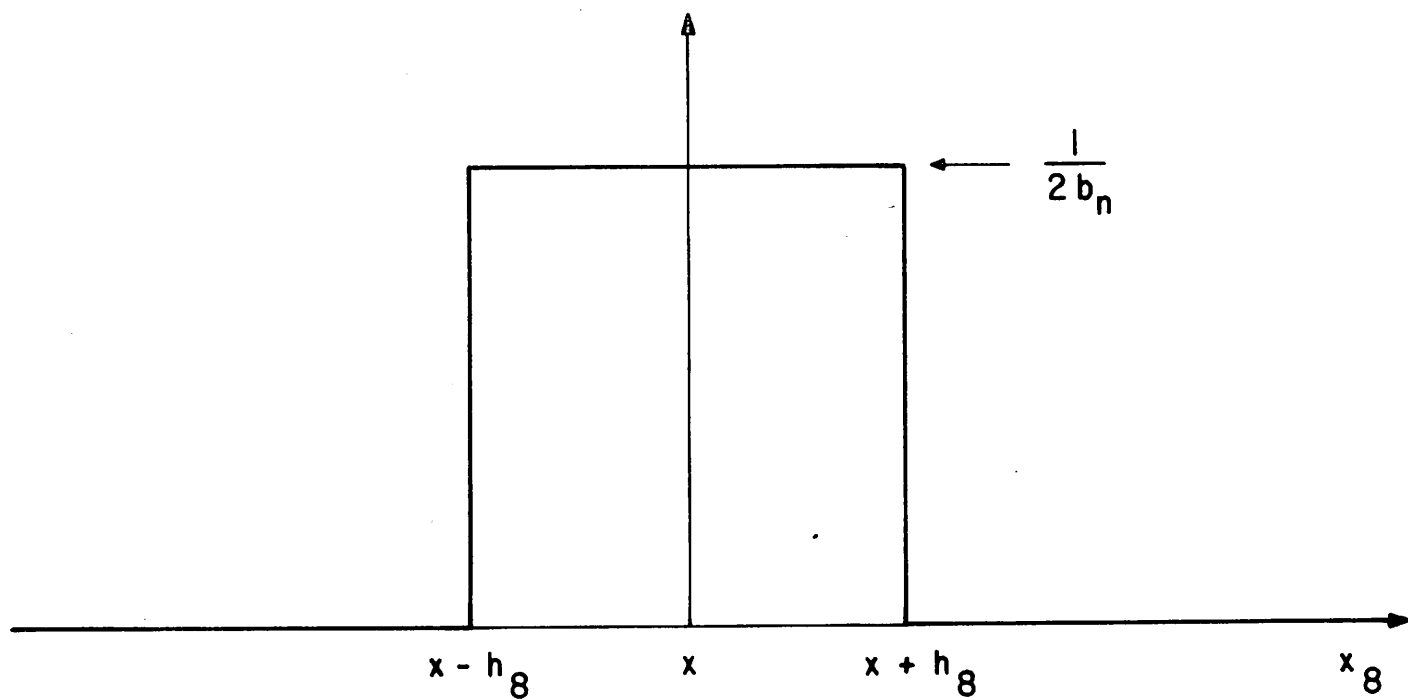


Figure 3-1

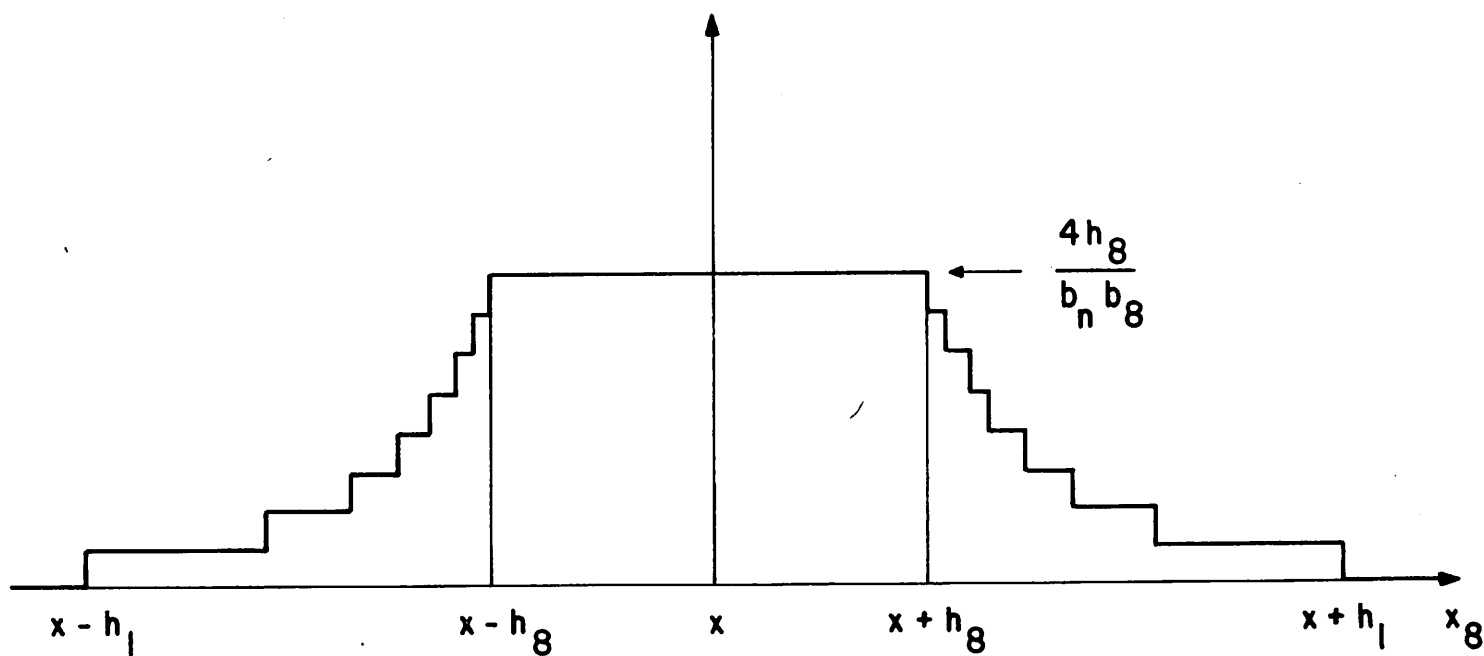


Figure 3-2