

Research Article

Bandwidth Selection for Recursive Kernel Density Estimators Defined by Stochastic Approximation Method

Yousri Slaoui

Université de Poitiers, Laboratoire de Mathématiques et Application, 86962 Futuroscope, Chasseneuil, France

Correspondence should be addressed to Yousri Slaoui; yousri.slaoui@math.univ-poitiers.fr

Received 18 January 2014; Revised 18 May 2014; Accepted 18 May 2014; Published 2 June 2014

Academic Editor: Shein-chung Chow

Copyright © 2014 Yousri Slaoui. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose an automatic selection of the bandwidth of the recursive kernel estimators of a probability density function defined by the stochastic approximation algorithm introduced by Mokkadem et al. (2009a). We showed that, using the selected bandwidth and the stepsize which minimize the MISE (mean integrated squared error) of the class of the recursive estimators defined in Mokkadem et al. (2009a), the recursive estimator will be better than the nonrecursive one for small sample setting in terms of estimation error and computational costs. We corroborated these theoretical results through simulation study.

1. Introduction

The problem of automatic choice of smoothing parameters has been widely studied. There are many reasons to use an automatic choice of smoothing. One is in many situations the smoothing which is used by nonexperts. In this paper we focus only on one-dimensional kernel density estimation. The main ideas are useful in all types of nonparametric curve estimation, including regression, distribution, and time series. The bandwidth selection methods studied in the literature can be divided into two broad classes: the cross-validation techniques and the plug-in ideas.

There are many varieties of the technique cross-validation: pseudolikelihood cross-validation [1], least squares cross-validation [2], and biased cross-validation [3]. Reviews of all these bandwidth selection methods can be found in Marron [4].

Plug-in methods [5], also called “second generation methods” [6], need to use a pilot bandwidth to estimate the unknown quantities. For a choice of pilot bandwidth, a number of approaches have been proposed; see Jones et al. [7] for details and references. An interesting approach to choose the pilot bandwidth is the smoothed bootstrap [8]. In this paper, we developed a specific second generation bandwidth selection method of the recursive kernel estimators

of a probability density function defined by the stochastic approximation algorithm introduced by Mokkadem et al. [9].

Let X_1, \dots, X_n be independent, identically distributed random variables and let f denote the probability density of X_1 . To construct a stochastic algorithm, which approximates the function f at a given point x , Mokkadem et al. [9] define an algorithm of search of the zero of the function $h : y \rightarrow f(x) - y$. Following Robbins-Monro's procedure, this algorithm is defined by setting $f_0(x) \in \mathbb{R}$, and, for all $n \geq 1$,

$$f_n(x) = f_{n-1}(x) + \gamma_n W_n, \quad (1)$$

where $W_n(x)$ is an “observation” of the function h at the point $f_{n-1}(x)$ and the stepsize (γ_n) is a sequence of positive real numbers that goes to zero. To define $W_n(x)$, Mokkadem et al. [9] follow the approach of Révész [10, 11] and of Tsybakov [12] and introduce a kernel K (i.e., a function satisfying $\int_{\mathbb{R}} K(x) dx = 1$), a bandwidth (h_n) (i.e., a sequence of positive real numbers that goes to zero), and set $W_n(x) = h_n^{-1} K(h_n^{-1}(x - X_n)) - f_{n-1}(x)$. Then, the estimator f_n to recursively estimate the density function f at the point x can be written as

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n h_n^{-1} K(h_n^{-1}[x - X_n]). \quad (2)$$

This estimator was introduced by Mokkadem et al. [9] whose large and moderate deviation principles were established by Slaoui [13].

Throughout this paper, we suppose that $f_0(x) = 0$ and we let $\Pi_n = \prod_{j=1}^n (1 - \gamma_j)$; then it follows from (2) that one can estimate f recursively at the point x by

$$f_n(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-1} K\left(\frac{x - X_k}{h_k}\right). \quad (3)$$

Moreover, it was shown in Mokkadem et al. [9] that the bandwidth which minimizes the MISE of f_n depends on the choice of the stepsize (γ_n) ; they show in particular that the sequence $(\gamma_n) = (n^{-1})$ belongs to this set, under some conditions of regularity of f , and they show that the bandwidth (h_n) must equal

$$\left(\left(\frac{3}{10}\right)^{1/5} \left\{ \frac{\int_{\mathbb{R}} K^2(z) dz}{\left(\int_{\mathbb{R}} z^2 K(z) dz\right)^2 \int_{\mathbb{R}} (f^{(2)}(x))^2 dx} \right\}^{1/5} n^{-1/5} \right). \quad (4)$$

The first aim of this paper is to propose an automatic selection of such bandwidth through a plug-in method, and the second aim is to give the conditions under which the recursive estimator f_n will be better than the nonrecursive kernel density estimator introduced by Rosenblatt [14] (see also Parzen [15]) and defined as

$$\tilde{f}_n(x) = \frac{1}{nh_n} \sum_{k=1}^n K\left(\frac{x - X_k}{h_n}\right). \quad (5)$$

The simulation results given in Section 3 are corroborating these theoretical results. The remainder of the paper is organized as follows. In Section 2, we state our main results. Section 3 is devoted to our simulation results. We conclude the paper in Section 4.

2. Assumptions and Main Results

We define the following class of regularly varying sequences.

Definition 1. Let $\gamma \in \mathbb{R}$ and let $(v_n)_{n \geq 1}$ be a nonrandom positive sequence. One says that $(v_n) \in \mathcal{RS}(\gamma)$ if

$$\lim_{n \rightarrow +\infty} n \left[1 - \frac{v_{n-1}}{v_n} \right] = \gamma. \quad (6)$$

Condition (6) was introduced by Galambos and Seneta [16] to define regularly varying sequences (see also Bojanic and Seneta [17]) and by Mokkadem and Pelletier [18] in the context of stochastic approximation algorithms. Note that the acronym $\mathcal{RS}(\gamma)$ stands for [16]. Typical sequences in $\mathcal{RS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^\gamma (\log n)^b$, $n^\gamma (\log \log n)^b$, and so on.

The assumptions to which we will refer are as follows.

(A1) $K : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}} K(z) dz = 1$, $\int_{\mathbb{R}} zK(z) = 0$, and $\int_{\mathbb{R}} z^2 K(z) < \infty$.

(A2) (i) $(\gamma_n) \in \mathcal{RS}(-\alpha)$ with $\alpha \in]1/2, 1]$.
(ii) $(h_n) \in \mathcal{RS}(-a)$ with $a \in]0, 1[$.
(iii) $\lim_{n \rightarrow \infty} (n\gamma_n) \in]\min\{2a, (\alpha - a)/2\}, \infty[$.

(A3) f is bounded and differentiable and $f^{(2)}$ is bounded.

Assumption (A2)(iii) on the limit of $(n\gamma_n)$ as n goes to infinity is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $([n\gamma_n]^{-1})$ is finite. Throughout this paper we will use the following notations:

$$\begin{aligned} \xi &= \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}, \\ R(K) &= \int_{\mathbb{R}} K^2(z) dz, \\ \mu_j(K) &= \int_{\mathbb{R}} z^j K(z) dz, \\ \Theta(K) &= R(K)^{4/5} \mu_2(K)^{2/5}, \\ \Gamma(K) &= R(K^{(2)})^{8/7} \mu_2(K)^{10/7}, \\ \zeta(K) &= R(K^{(2)})^{8/7} \mu_2(K)^{-4/7} \mu_4(K), \\ I_1 &= \int_{\mathbb{R}} f^2(x) dx, \\ I_2 &= \int_{\mathbb{R}} (f^{(2)}(x))^2 f(x) dx. \end{aligned} \quad (7)$$

In order to measure the quality of our recursive estimator (3), we use the following quantity:

$$\begin{aligned} \text{MISE}^*[f_n] &= \mathbb{E} \int_{\mathbb{R}} [f_n(x) - f(x)]^2 f(x) dx \\ &= \int_{\mathbb{R}} (\mathbb{E}(f_n(x)) - f(x))^2 f(x) dx \\ &\quad + \int_{\mathbb{R}} \text{Var}(f_n(x)) f(x) dx. \end{aligned} \quad (8)$$

Moreover, in the case $a = \alpha/5$, Proposition 1 in Mokkadem et al. [9] shows that

$$\mathbb{E}[f_n(x)] - f(x) = \frac{1}{2(1 - 2a\xi)} h_n^2 f^{(2)}(x) \mu_2(K) + o(h_n^2) \quad (9)$$

and that

$$\text{Var}[f_n(x)] = \frac{1}{2 - (\alpha - a)\xi} \gamma_n h_n^{-1} f(x) R(K) + o(\gamma_n h_n^{-1}). \quad (10)$$

Then

$$\begin{aligned} \text{MISE}^*[f_n] &= \frac{1}{2 - (\alpha - a)\xi} \gamma_n h_n^{-1} R(K) I_1 \\ &\quad + \frac{1}{4(1 - 2a\xi)^2} h_n^4 \mu_2^2(K) I_2 + o(\gamma_n h_n^{-1} + h_n^4). \end{aligned} \quad (11)$$

The following corollary ensures that the bandwidth which minimizes the MISE* depends on the stepsize (γ_n) and then the corresponding MISE* depends also on the stepsize (γ_n).

Corollary 2. *Let assumptions (A1)–(A3) hold. To minimize the MISE* of f_n , the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and the bandwidth (h_n) must equal*

$$\left(\frac{(1 - 2a\xi)^{2/5}}{(2 - (\alpha - a)\xi)^{1/5}} \left\{ \frac{R(K) I_1}{\mu_2^2(K) I_2} \right\}^{1/5} \gamma_n^{1/5} \right). \quad (12)$$

Then, one has

$$\begin{aligned} \text{MISE}^* [f_n] &= \frac{5}{4} (1 - 2a\xi)^{-2/5} (2 - (\alpha - a)\xi)^{-4/5} \\ &\times \Theta(K) I_1^{4/5} I_2^{1/5} \gamma_n^{4/5} + o(\gamma_n^{4/5}). \end{aligned} \quad (13)$$

The following corollary shows that, for a special choice of the stepsize ($\gamma_n = (\gamma_0 n^{-1})$), which fulfilled that $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$ and that ($\gamma_n \in \mathcal{GS}(-1)$), the optimal value for h_n depends on γ_0 and then the corresponding MISE* depends on γ_0 .

Corollary 3. *Let assumptions (A1)–(A3) hold. To minimize the MISE* of f_n , the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$, $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$, and the bandwidth (h_n) must equal*

$$\left(2^{-1/5} \left(\gamma_0 - \frac{2}{5} \right)^{1/5} \left\{ \frac{R(K) I_1}{\mu_2^2(K) I_2} \right\}^{1/5} n^{-1/5} \right), \quad (14)$$

and one then has

$$\begin{aligned} \text{MISE}^* [f_n] &= \frac{5}{4} \frac{1}{2^{4/5}} \frac{\gamma_0^2}{(\gamma_0 - 2/5)^{6/5}} \Theta(K) I_1^{4/5} I_2^{1/5} n^{-4/5} \\ &+ o(n^{-4/5}). \end{aligned} \quad (15)$$

Moreover, the minimum of $\gamma_0^2(\gamma_0 - 2/5)^{-6/5}$ is reached at $\gamma_0 = 1$; then the bandwidth (h_n) must equal

$$\left(\left(\frac{3}{10} \right)^{1/5} \left\{ \frac{R(K) I_1}{\mu_2^2(K) I_2} \right\}^{1/5} n^{-1/5} \right), \quad (16)$$

and one then has

$$\begin{aligned} \text{MISE}^* [f_n] &= \frac{5}{4} 2^{2/5} \left(\frac{5}{6} \right)^{6/5} \Theta(K) I_1^{4/5} I_2^{1/5} n^{-4/5} \\ &+ o(n^{-4/5}). \end{aligned} \quad (17)$$

In order to estimate the optimal bandwidth (16), we must estimate I_1 and I_2 . We followed the approach of Altman and Léger [19], which is called the plug-in estimate, and we use the following kernel estimator of I_1 :

$$\hat{I}_1 = \frac{\Pi_n}{n} \sum_{i,k=1}^n \Pi_k^{-1} \gamma_k b_k^{-1} K_b \left(\frac{X_i - X_k}{b_k} \right), \quad (18)$$

where K_b is a kernel and b is the associated bandwidth.

In practice, we take

$$b_n = n^{-\beta} \min \left\{ \hat{s}, \frac{Q_3 - Q_1}{1.349} \right\}, \quad \beta \in]0, 1[\quad (19)$$

(see Silverman [20]) with \hat{s} the sample standard deviation and Q_1, Q_3 denoting the first and third quartiles, respectively.

The following theorem gives the bias and variance of \hat{I}_1 .

Theorem 4. *Let assumptions (A2)–(A3) hold and suppose that the kernel K_b satisfies assumption (A1) and ($b_n \in \mathcal{GS}(-\beta)$), with $\beta \in]0, 1[$; one has*

$$\begin{aligned} \text{Bias} [\hat{I}_1] &= \frac{1}{2(1 - 2a\xi)} b_n^2 \mu_2(K_b) \\ &\times \int_{\mathbb{R}} f^{(2)}(x) f(x) dx + o(b_n^2), \\ \text{Var} [\hat{I}_1] &= \frac{1}{2 - (\alpha - a)\xi} \frac{\gamma_n}{n b_n} R(K_b) I_1 \\ &+ \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) + o\left(\frac{1}{n} + \frac{\gamma_n}{n b_n} \right). \end{aligned} \quad (20)$$

The following corollary shows that the bandwidth which minimizes the MISE of \hat{I}_1 depends on the stepsize (γ_n) and then the corresponding MISE depends also on the stepsize (γ_n).

Corollary 5. *Let the assumptions of Theorem 4 hold. To minimize the MISE of \hat{I}_1 , the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and the bandwidth (b_n) must equal*

$$\begin{aligned} &\left(\frac{(1 - 2a\xi)^{2/5}}{(2 - (1 - a)\xi)^{1/5}} \right. \\ &\times \left. \left\{ \frac{R(K_b) I_1}{\mu_2^2(K_b) \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/5} \left(\frac{\gamma_n}{n} \right)^{1/5} \right). \end{aligned} \quad (21)$$

Then, one has

$$\begin{aligned} \text{MISE} [\hat{I}_1] &= \frac{5}{4} (1 - 2\xi)^{-2/5} (2 - (1 - a)\xi)^{-4/5} \Theta(K_b) I_1^{4/5} \\ &\times \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{2/5} \left(\frac{\gamma_n}{n} \right)^{4/5} \\ &+ \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) + o\left(\left(\frac{\gamma_n}{n} \right)^{4/5} \right). \end{aligned} \quad (22)$$

The following corollary shows that, for a special choice of the stepsize ($\gamma_n = (\gamma_0 n^{-1})$), which fulfilled that $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$ and that ($\gamma_n \in \mathcal{GS}(-1)$), the optimal value for b_n depends on γ_0 and the corresponding MISE depends on γ_0 .

Corollary 6. Let the assumptions of Theorem 4 hold. To minimize the MISE of \hat{I}_1 , the stepsize (γ_n) must be chosen in $\mathcal{ES}(-1)$, $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$, and the bandwidth (b_n) must equal

$$\left(\frac{(\gamma_0 - 4/5)^{2/5}}{(2\gamma_0 - 3/5)^{1/5}} \times \left\{ \frac{R(K_b)}{\mu_2^2(K_b)} \frac{I_1}{\left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/5} n^{-2/5} \right). \quad (23)$$

Then, one has

$$\begin{aligned} \text{MISE}[\hat{I}_1] &= \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) + \frac{5}{4} \gamma_0^2 \\ &\times \left(\gamma_0 - \frac{4}{5} \right)^{-2/5} \left(2\gamma_0 - \frac{3}{5} \right)^{-4/5} \Theta(K_b) I_1^{4/5} \\ &\times \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{2/5} n^{-8/5} + o(n^{-8/5}). \end{aligned} \quad (24)$$

Moreover, the minimum of $\gamma_0^2(\gamma_0 - 4/5)^{-2/5}(2\gamma_0 - 3/5)^{-4/5}$ is reached at $\gamma_0 = (9 + \sqrt{21})/10$; then the bandwidth (b_n) must equal

$$\left(0.6817256 \times \left\{ \frac{R(K_b)}{\mu_2^2(K_b)} \frac{I_1}{\left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/5} n^{-2/5} \right). \quad (25)$$

Then

$$\begin{aligned} \text{Bias}[\hat{I}_1] &\approx 0.5653751 \times \Theta(K_b)^{1/2} I_1^{2/5} \\ &\times \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{1/5} n^{-4/5} + o(n^{-4/5}), \\ \text{Var}[\hat{I}_1] &\approx \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) + 1.278596 \\ &\times \Theta(K_b) I_1^{4/5} \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{2/5} \\ &\times n^{-8/5} + o(n^{-8/5}). \end{aligned} \quad (26)$$

It follows that

$$\begin{aligned} \text{MISE}[\hat{I}_1] &\approx \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) + 1.598245 \\ &\times \Theta(K_b) I_1^{4/5} \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{2/5} n^{-8/5} \\ &+ o(n^{-8/5}). \end{aligned} \quad (27)$$

Furthermore, to estimate I_2 , we introduce the following kernel estimator:

$$\begin{aligned} \hat{I}_2 &= \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k=1 \\ j \neq k}}^n \Pi_j^{-1} \Pi_k^{-1} \gamma_j \gamma_k b_j^{l-3} b_k^{l-3} \\ &\times K_{b'}^{(2)} \left(\frac{X_i - X_j}{b'_j} \right) K_{b'}^{(2)} \left(\frac{X_i - X_k}{b'_k} \right), \end{aligned} \quad (28)$$

where $K_{b'}^{(2)}$ is the second order derivative of a kernel $K_{b'}$. The bias and variance of \hat{I}_2 are computed in the following theorem.

Theorem 7. Let assumptions (A2)-(A3) hold, and suppose that the kernel K_b^l satisfies assumption (A1) and $(b'_n) \in \mathcal{ES}(-\beta')$, with $\beta' \in]0, 1[$; one has

$$\begin{aligned} \text{Bias}[\hat{I}_2] &= \frac{1}{(1 - 2a\xi)} b_n^{l/2} \mu_2(K_{b'}) \\ &\times \int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \\ &+ \frac{1}{4(1 - 2a\xi)^2} b_n^{l/4} \mu_2^2(K_{b'}) \\ &\times \int_{\mathbb{R}} (f^{(4)}(x))^2 f(x) dx \\ &+ \frac{1}{12(1 - 4a\xi)} b_n^{l/4} \mu_4(K_{b'}) \\ &\times \int_{\mathbb{R}} f^{(6)}(x) f^{(4)}(x) f(x) dx \\ &- \frac{1}{2 - \xi} I_2 \gamma_n + o(b_n^{l/4} + \gamma_n), \\ \text{Var}[\hat{I}_2] &= \left(\frac{1}{2 - (\alpha - 5a)\xi} \right)^2 \frac{\gamma_n^2}{n b_n^{l/10}} R(K_{b'}^{(2)})^4 \\ &\times \int_{\mathbb{R}} f^3(x) dx \\ &+ \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) \\ &+ o\left(\frac{1}{n} + \frac{\gamma_n^2}{n b_n^{l/10}} \right). \end{aligned} \quad (29)$$

The following corollary ensures that the bandwidth which minimizes the MISE of \hat{I}_2 depends on the stepsize (γ_n) and then the corresponding MISE depends also on the stepsize (γ_n) .

Corollary 8. Let the assumptions of Theorem 7 hold. To minimize the MISE of \widehat{I}_2 , the stepsize (γ_n) must be chosen in $\mathcal{ES}(-1)$ and the bandwidth (b'_n) must equal

$$\left(\left(\frac{5}{2} \right)^{1/14} \frac{(1 - 2a\xi)^{1/7}}{(2 - (1 - 5a)\xi)^{1/7}} \times \left\{ \frac{R(K_{b'}^{(2)})^4 \int_{\mathbb{R}} f^3(x) dx}{\mu_2^2(K_{b'}) \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/14} \times \left(\frac{\gamma_n^2}{n} \right)^{1/14} \right); \quad (30)$$

then one has

$$\begin{aligned} \text{MISE} [\widehat{I}_2] &= \frac{7}{5} \left(\frac{5}{2} \right)^{2/7} (1 - 2a\xi)^{-10/7} \\ &\times (2 - (1 - 5a)\xi)^{-4/7} \Gamma(K_{b'}) \\ &\times \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{10/7} \left(\frac{\gamma_n^2}{n} \right)^{2/7} \\ &+ \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) + o(n^{-1}). \end{aligned} \quad (31)$$

The following corollary shows that, for a special choice of the stepsize $(\gamma_n) = (\gamma_0 n^{-1})$, which fulfilled that $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$ and that $(\gamma_n) \in \mathcal{ES}(-1)$, the optimal value for b'_n depends on γ_0 and the corresponding MISE depends on γ_0 .

Corollary 9. Let the assumptions of Theorem 7 hold. To minimize the MISE of \widehat{I}_2 , the stepsize (γ_n) must be chosen in $\mathcal{ES}(-1)$, $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$, and the bandwidth (b'_n) must equal

$$\left(\left(\frac{5}{2} \right)^{1/14} \frac{\gamma_0^{1/7} (\gamma_0 - 3/7)^{1/7}}{(2\gamma_0 + 1/14)^{1/7}} \times \left\{ \frac{R(K_{b'}^{(2)})^4 \int_{\mathbb{R}} f^3(x) dx}{\mu_2^2(K_{b'}) \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/14} \times n^{-3/14} \right); \quad (32)$$

then one has

$$\begin{aligned} \text{MISE} [\widehat{I}_2] &= \frac{7}{5} \left(\frac{5}{2} \right)^{2/7} \frac{\gamma_0^{18/7}}{(\gamma_0 - 3/7)^{10/7} (2\gamma_0 + 1/14)^{4/7}} \\ &\times \Gamma(K_{b'}) \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{10/7} n^{-6/7} \\ &+ \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) + o(n^{-1}). \end{aligned} \quad (33)$$

Moreover, the minimum of $\gamma_0^{18/7} (\gamma_0 - 3/7)^{-10/7} (2\gamma_0 + 1/14)^{-4/7}$ is reached at $\gamma_0 = (10 + \sqrt{227/2})/14$; then the bandwidth (b'_n) must equal

$$\left(0.9699894 \times \left\{ \frac{R(K_{b'}^{(2)})^4 \int_{\mathbb{R}} f^3(x) dx}{\mu_2^2(K_{b'}) \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/14} \times n^{-3/14} \right); \quad (34)$$

then we have

$$\begin{aligned} \text{Bias} [\widehat{I}_2] &\simeq 1.326127 \times \Gamma(K_{b'})^{1/2} \left(\int_{\mathbb{R}} f^3(x) dx \right)^{1/7} \\ &\times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{5/7} n^{-3/7} \\ &+ \left\{ 0.4396531 \times \Gamma(K_{b'}) \int_{\mathbb{R}} (f^{(4)}(x))^2 f(x) dx \right. \\ &\quad \left. + 0.1760697 \times \zeta(K_{b'}) \right. \\ &\quad \left. \times \int_{\mathbb{R}} f^{(6)}(x) f^{(2)}(x) f(x) dx \right\} \\ &\times \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{-4/7} n^{-6/7} \\ &- 1.115801 \times I_2 n^{-1} + o(n^{-1}), \end{aligned}$$

$$\begin{aligned} \text{Var} [\hat{I}_2] &\approx 0.703445 \times \Gamma(K_{b'}) \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{10/7} n^{-6/7} \\ &\quad + \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) + o(n^{-1}). \end{aligned} \quad (35)$$

It follows that

$$\begin{aligned} \text{MISE} [\hat{I}_2] &\approx 2.462058 \times \Gamma(K_{b'}) \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{10/7} n^{-6/7} \\ &\quad + \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) + o(n^{-1}). \end{aligned} \quad (36)$$

Finally, the plug-in estimator of the bandwidth (h_n) using the recursive algorithm (3) must equal

$$\left(\left(\frac{3}{10} \right)^{1/5} \left(\frac{R(K)}{\mu_2^2(K)} \right)^{1/5} \left(\frac{\hat{I}_1}{\hat{I}_2} \right)^{1/5} n^{-1/5} \right), \quad (37)$$

$$\begin{aligned} \text{MISE}^* [f_n] &= \frac{5}{4} 2^{2/5} \left(\frac{5}{6} \right)^{6/5} \Theta(K) \hat{I}_1^{4/5} \hat{I}_2^{1/5} n^{-4/5} \\ &\quad + o(n^{-4/5}). \end{aligned} \quad (38)$$

Now, let us recall that the bias and variance of Rosenblatt's estimator \tilde{f}_n are given by

$$\begin{aligned} \mathbb{E} [\tilde{f}_n(x)] - f(x) &= \frac{1}{2} h_n^2 f^{(2)}(x) \mu_2(K) + o(h_n^2), \\ \text{Var} [\tilde{f}_n(x)] &= \frac{1}{nh_n} f(x) R(K) + o\left(\frac{1}{nh_n}\right). \end{aligned} \quad (39)$$

It follows that

$$\begin{aligned} \text{MISE}^* [\tilde{f}_n] &= \frac{1}{nh_n} I_1 R(K) + \frac{1}{4} h_n^4 I_2 \mu_2^2(K) \\ &\quad + o\left(h_n^4 + \frac{1}{nh_n}\right). \end{aligned} \quad (40)$$

To minimize the MISE* of \tilde{f}_n , the bandwidth (h_n) must equal

$$\left(\left\{ \frac{R(K) I_1}{\mu_2^2(K) I_2} \right\}^{1/5} n^{-1/5} \right), \quad (41)$$

and then we have

$$\text{MISE}^* [\tilde{f}_n] = \frac{5}{4} \Theta(K) I_1^{4/5} I_2^{1/5} n^{-4/5} + o(n^{-4/5}). \quad (42)$$

To estimate the optimal bandwidth (41), we must estimate I_1 and I_2 . As suggested by Hall and Marron [21], we use the following kernel estimator of I_1 :

$$\tilde{I}_1 = \frac{1}{n(n-1)b_n} \sum_{\substack{i,j=1 \\ i \neq j}}^n K_b \left(\frac{X_i - X_j}{b_n} \right). \quad (43)$$

The following lemma gives the bias and variance of \tilde{I}_1 .

Lemma 10. Suppose that the kernel K_b satisfies assumption (A1), (b_n), and assumption (A2)(ii) and f satisfies assumption (A3)

$$\begin{aligned} \text{Bias} [\tilde{I}_1] &= \frac{1}{2} b_n^2 \mu_2(K_b) \int_{\mathbb{R}} f^{(2)}(x) f(x) dx + o(b_n^2), \\ \text{Var} [\tilde{I}_1] &= \frac{1}{n(n-1)b_n} R(K_b) I_1 \\ &\quad + \frac{1}{n-1} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) \\ &\quad + o\left(\frac{1}{n} + \frac{1}{n^2 b_n}\right). \end{aligned} \quad (44)$$

To minimize the MISE of \tilde{I}_1 , the bandwidth (b_n) must equal

$$\left(\left\{ \frac{R(K_b) I_1}{\mu_2^2(K_b) \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/5} n^{-2/5} \right); \quad (45)$$

then one has

$$\begin{aligned} \text{Bias} [\tilde{I}_1] &= \frac{1}{2} \Theta(K_b)^{1/2} I_1^{2/5} \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{1/5} \\ &\quad \times n^{-4/5} + o(n^{-4/5}), \\ \text{Var} [\tilde{I}_1] &= \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) \\ &\quad + \Theta(K_b) I_1^{4/5} \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{2/5} \\ &\quad \times n^{-8/5} + o(n^{-8/5}), \\ \text{MISE} [\tilde{I}_1] &= \frac{1}{n} \left(\int_{\mathbb{R}} f^3(x) dx - I_1^2 \right) \\ &\quad + \frac{5}{4} \Theta(K_b) I_1^{4/5} \left(\int_{\mathbb{R}} f^{(2)}(x) f(x) dx \right)^{2/5} \\ &\quad \times n^{-8/5} + o(n^{-8/5}). \end{aligned} \quad (46)$$

Furthermore, to estimate I_2 , we use the following kernel estimator introduced in Hall and Marron [21]:

$$\tilde{I}_2 = \frac{1}{n^3 b_n'^6} \sum_{\substack{i,j,k=1 \\ j \neq k}}^n K_{b'}^{(2)} \left(\frac{X_i - X_j}{b'_n} \right) K_{b'}^{(2)} \left(\frac{X_i - X_k}{b'_n} \right). \quad (47)$$

The following lemma gives the bias and variance of \tilde{I}_2 .

Lemma 11. Suppose that the kernel $K_{b'}$ satisfies assumption (A1), (b'_n) , and assumption (A2)(ii) and f satisfies assumption (A3)

$$\begin{aligned} \text{Bias} [\tilde{I}_2] &= b_n'^2 \mu_2(K_{b'}) \int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \\ &\quad + \frac{1}{4} b_n'^4 \mu_2^2(K_{b'}) \int_{\mathbb{R}} (f^{(4)}(x))^2 f(x) dx \\ &\quad + \frac{1}{12} b_n'^4 \mu_4(K_{b'}) \int_{\mathbb{R}} f^{(6)}(x) f^{(2)}(x) f(x) dx \\ &\quad + o(b_n'^4), \\ \text{Var} [\tilde{I}_2] &= \frac{1}{n^3 b_n'^{10}} R(K_{b'}^{(2)})^4 \int_{\mathbb{R}} f^3(x) dx \\ &\quad + \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) \\ &\quad + o\left(\frac{1}{n} + \frac{1}{n^3 b_n'^{10}}\right). \end{aligned} \quad (48)$$

To minimize the MISE of \tilde{I}_2 , the bandwidth (b'_n) must equal

$$\left(\left\{ \frac{5}{2} \frac{R(K_{b'}^{(2)})^4 \int_{\mathbb{R}} f^3(x) dx}{\mu_2^2(K_{b'}) \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^2} \right\}^{1/14} \times n^{-3/14} \right); \quad (49)$$

then one has

$$\begin{aligned} \text{Bias} [\tilde{I}_2] &\approx 1.139852 \times \Gamma(K_{b'})^{1/2} \left(\int_{\mathbb{R}} f^3(x) dx \right)^{1/7} \\ &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{5/7} n^{-3/7} \\ &\quad + \left\{ 0.3248158 \times \Gamma(K_{b'}) \int_{\mathbb{R}} (f^{(4)}(x))^2 f(x) dx \right. \\ &\quad \left. + 0.1082719 \times \zeta(K_{b'}) \right. \\ &\quad \left. \times \int_{\mathbb{R}} f^{(6)}(x) f^{(2)}(x) f(x) dx \right\} \\ &\quad \times \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{-4/7} n^{-6/7} \\ &\quad + o(n^{-6/7}), \end{aligned}$$

$$\begin{aligned} \text{Var} [\tilde{I}_2] &\approx 0.5197053 \times \Gamma(K_{b'}) \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{10/7} n^{-6/7} \\ &\quad + \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) + o(n^{-1}). \end{aligned} \quad (50)$$

It follows then that

$$\begin{aligned} \text{MISE} [\tilde{I}_2] &\approx 1.81189 \times \Gamma(K_{b'}) \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\ &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{10/7} n^{-6/7} \\ &\quad + \frac{1}{n} \left(\int_{\mathbb{R}} (f^{(2)}(x))^4 f(x) dx - I_2^2 \right) + o(n^{-1}). \end{aligned} \quad (51)$$

Then the plug-in estimator of the bandwidth (h_n) using the nonrecursive estimator (5) must equal

$$\left(\left(\frac{R(K)}{\mu_2^2(K)} \right)^{1/5} \left(\frac{\tilde{I}_1}{\tilde{I}_2} \right)^{1/5} n^{-1/5} \right), \quad (52)$$

$$\text{MISE}^* [\tilde{f}_n] = \frac{5}{4} \Theta(K) \tilde{I}_1^{4/5} \tilde{I}_2^{1/5} n^{-4/5} + o(n^{-4/5}). \quad (53)$$

The following corollary gives the expected MISE* of the recursive estimator f_n and the nonrecursive estimator \tilde{f}_n .

Corollary 12. Let the assumptions of Theorem 7 hold. Then

$$\begin{aligned} \mathbb{E} [\text{MISE}^* [f_n]] &= \frac{5}{4} 2^{2/5} \left(\frac{5}{6} \right)^{6/5} \Theta(K) I_1^{4/5} I_2^{1/5} \\ &\quad \times \left[\left(1 + 0.26522 \times C_1 I_2^{-1} n^{-3/7} \right. \right. \\ &\quad \left. \left. + (0.08793 \times C_2 + 0.03521 \times C_3) \right. \right. \\ &\quad \left. \left. \times I_2^{-1} n^{-6/7} - 0.22316 \times n^{-1} \right) + o(n^{-1}) \right] \\ &\quad \times n^{-4/5} (1 + o(1)), \end{aligned} \quad (54)$$

$$\begin{aligned} \mathbb{E} [\text{MISE}^* [\tilde{f}_n]] &= \frac{5}{4} \Theta(K) I_1^{4/5} I_2^{1/5} \\ &\quad \times \left[\left(1 + 0.22797 \times C_1 I_2^{-1} n^{-3/7} \right. \right. \\ &\quad \left. \left. + (0.06496 \times C_2 + 0.02165 \times C_3) \right. \right. \\ &\quad \left. \left. \times I_2^{-1} n^{-6/7} \right) \right] n^{-4/5} (1 + o(1)), \end{aligned}$$

TABLE 1: $X \rightsquigarrow \mathcal{N}(0, 1)$.

	$n = 50$					$n = 100$			
	Ref.	Rosenblatt		Recursive		Rosenblatt		Recursive	
		Mean	sd	Mean	sd	Mean	sd	Mean	sd
I_1	0.2821	0.2780	0.0357	0.2864	0.0358	0.2788	0.0255	0.2866	0.0257
I_2	0.0612	0.0994	0.2744	0.0690	0.1697	0.0749	0.1908	0.0575	0.1078
h_n		0.4801	0.1435	0.4408	0.1399	0.4347	0.1169	0.3894	0.1197
MISE*		0.0801	0.0288	0.0746	0.0280	0.0649	0.0203	0.0621	0.0195

TABLE 2: $X \rightsquigarrow \mathcal{E}(1/2)$.

	$n = 50$					$n = 100$			
	Ref.	Rosenblatt		Recursive		Rosenblatt		Recursive	
		Mean	sd	Mean	sd	Mean	sd	Mean	sd
I_1	0.25	0.1930	0.0331	0.1768	0.0303	0.1957	0.0232	0.1839	0.0221
I_2	0.0078	0.0353	0.0598	0.0196	0.0327	0.0369	0.0336	0.0206	0.0199
h_n		0.5825	0.1203	0.5049	0.0989	0.4611	0.0598	0.4032	0.0525
MISE*		0.0435	0.0166	0.0381	0.0144	0.0407	0.0100	0.0363	0.0091

where

$$\begin{aligned}
 C_1 &= \Gamma(K_{b'})^{1/2} \left(\int_{\mathbb{R}} f^3(x) dx \right)^{1/7} \\
 &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{5/7} \\
 C_2 &= \Gamma(K_{b'}) \int_{\mathbb{R}} (f^{(4)}(x))^2 f(x) dx \\
 &\quad \times \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\
 &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{-4/7} \quad (55) \\
 C_3 &= \zeta(K_{b'}) \int_{\mathbb{R}} f^{(6)}(x) f^{(2)}(x) f(x) dx \\
 &\quad \times \left(\int_{\mathbb{R}} f^3(x) dx \right)^{2/7} \\
 &\quad \times \left(\int_{\mathbb{R}} f^{(4)}(x) f^{(2)}(x) f(x) dx \right)^{-4/7}.
 \end{aligned}$$

The following theorem gives the conditions under which the expected MISE* of the recursive estimator f_n will be smaller than the expected MISE* of the nonrecursive estimator \tilde{f}_n .

Theorem 13. *Let the assumptions of Theorem 7 hold and let the bandwidth (h_n) equal (37) and the stepsize $(\gamma_n) = (n^{-1})$. One has*

$$\frac{\mathbb{E} [\text{MISE}^* [f_n]]}{\mathbb{E} [\text{MISE}^* [\tilde{f}_n]]} < 1 \quad \text{for small sample setting.} \quad (56)$$

Then, the expected MISE* of the recursive estimator defined by (3) is smaller than the expected MISE* of the

nonrecursive estimator defined by (5) for small sample setting.

3. Simulation

The aim of our simulation study is to compare the performance of the nonrecursive Rosenblatt's estimator defined in (5) with that of the recursive estimators defined in (3).

When applying f_n one needs to choose three quantities.

- (i) For the function K , we choose the normal kernel.
- (ii) The stepsize $(\gamma_n) = (n^{-1})$.
- (iii) The bandwidth (h_n) is chosen to be equal to (37).

When applying \tilde{f}_n one needs to choose two quantities.

- (i) For the function K , as in the recursive framework, we use the normal kernel.
- (ii) The bandwidth (h_n) is chosen to be equal to (52).

In order to investigate the comparison between the two estimators, we consider two densities of f : the standard normal $\mathcal{N}(0, 1)$ distribution (see Table 1) and the exponential $\mathcal{E}(1/2)$ distribution (see Table 2). For each of these two cases, 500 samples of sizes $n = 50$ and 100 were generated. For each fixed bandwidth h , we computed the mean and the standard deviation (over the 500 samples) of I_1 , I_2 , h_n , and MISE*. The plug-in estimators (37) and (52) require two kernels to estimate I_1 and I_2 . In both cases we use the normal kernel with b_n and b'_n given in (19), with β equal, respectively, to $2/5$ and $3/14$. Both tables show that (1) the bias, respectively, and the standard deviation of I_1 using the recursive algorithm (3) are very similar to the bias, respectively, and the standard deviation of I_1 using the nonrecursive estimator (5), (2) the bias, respectively, and the standard deviation of I_2 using the recursive algorithm (3) are always smaller than the

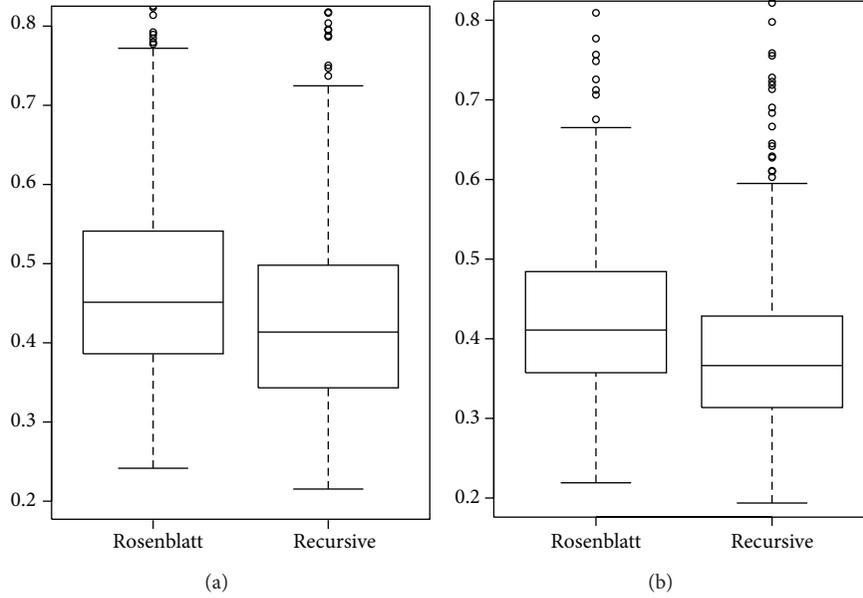


FIGURE 1: The bandwidths minimizing the $MISE^*$ of the nonrecursive estimator (5) and the recursive estimator (3) for 500 samples, respectively, of size $n = 50$ (a) and of size $n = 100$ (b) from a normal distribution.

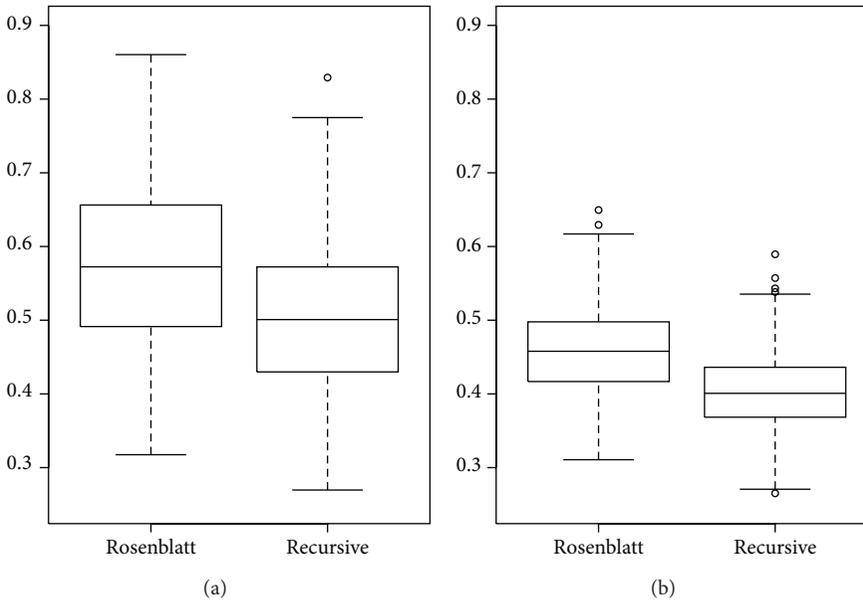


FIGURE 2: The bandwidths minimizing the $MISE^*$ of the nonrecursive estimator (5) and the recursive estimator (3) for 500 samples, respectively, of size $n = 50$ (a) and of size $n = 100$ (b) from an exponential distribution of rate $1/2$.

bias, respectively, and the standard deviation of I_2 using the nonrecursive estimator (5), (3) the mean, respectively, and the standard deviation of the bandwidths selected by the recursive estimator (3) are always smaller than the bias, respectively, and the standard deviation of the bandwidths selected by the nonrecursive estimator (5), and (4) the mean, respectively, and the standard deviation of the $MISE^*$ of the bandwidths selected by the recursive estimator (3) are always smaller than the bias, respectively, and the standard deviation of $MISE^*$ of the bandwidths selected by the nonrecursive

estimator (5). In Tables 1 and 2 the Ref. column could be used as a reference for the mean of I_1 and I_2 .

Figures 1 and 2 show boxplots of the selected bandwidths by the two algorithms (3) and (5), respectively. For samples of size 50 and 100, the bandwidths selected by the recursive estimator (3) are always smaller than the bandwidths selected by the nonrecursive estimator (5).

Figures 3 and 4 show boxplots of the expected $MISE^*$ by the two algorithms (3) and (5), respectively. For samples of size 50 and 100, the expected $MISE^*$ of the selected

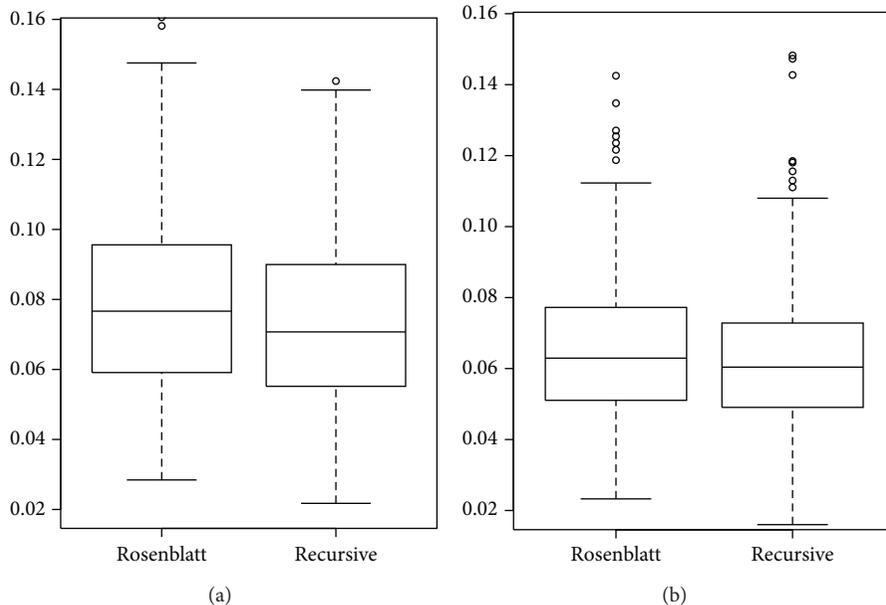


FIGURE 3: The $MISE^*$ of the optimal bandwidth of the nonrecursive estimator (5) and the recursive estimator (3) for 500 samples, respectively, of size $n = 50$ (a) and of size $n = 100$ (b) from a normal distribution.

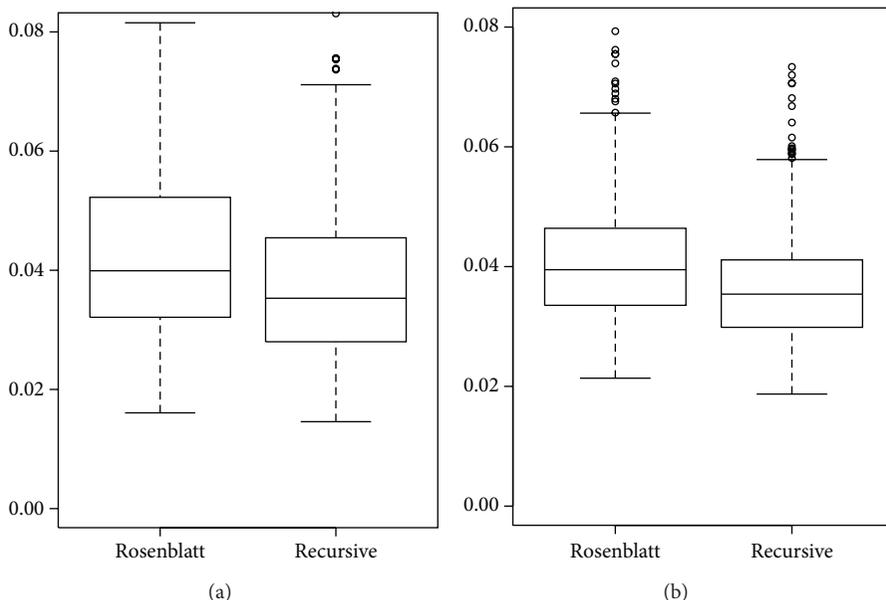


FIGURE 4: The $MISE^*$ of the optimal bandwidth of the nonrecursive estimator (5) and the recursive estimator (3) for 500 samples, respectively, of size $n = 50$ (a) and of size $n = 100$ (b) from an exponential distribution of rate $1/2$.

bandwidths by the recursive estimator (3) are always smaller than the expected $MISE^*$ of the selected bandwidths by the nonrecursive estimator (5).

In order to give some comparative elements with nonrecursive estimator (5), including computational costs, we consider 500 samples of size 50 generated from a standard normal distribution $\mathcal{N}(0, 1)$; moreover, we suppose that we receive an additional 500 samples of size 50 generated also from a standard normal distribution $\mathcal{N}(0, 1)$. Performing the two methods, the running time using the recursive estimator

defined by algorithm (2) with stepsize $(\gamma_n) = (n^{-1})$ and the bandwidth (h_n) given in (37) was roughly 6880 s on the author’s workstation, while the running time using the nonrecursive estimator defined by algorithm (5) with the bandwidth (h_n) given in (52) was roughly 14080 s on the author’s workstation.

This simulation study shows the good performance of the recursive estimator defined by algorithm (3) with stepsize $(\gamma_n) = (n^{-1})$ and the bandwidth (h_n) given in (37) for small sample setting.

4. Conclusion

In this paper we proposed an automatic selection of the bandwidth of the recursive kernel estimators of a probability density function defined by the stochastic approximation algorithm (2). We showed that, using the selected bandwidth and the stepsize $(\gamma_n) = (n^{-1})$ (the stepsize which minimizes the MISE of the class of the recursive estimators defined in Mokkadem et al. [9]), the recursive estimator will be better than the nonrecursive one for small sample setting. The simulation study corroborated these theoretical results. Moreover, the simulation results indicate that the proposed recursive estimator has more computing efficiency than the nonrecursive estimator.

In conclusion, the proposed method allowed us to obtain better results than the nonrecursive estimator proposed by Rosenblatt [14] for small sample setting. Moreover, we plan to make an extension of our method in the future and to consider the case of a regression function as in Härdle and Marron [22] in recursive way (see Mokkadem et al. [23]) and the case of time series as in Hart and Vieu [24] in recursive way.

Conflict of Interests

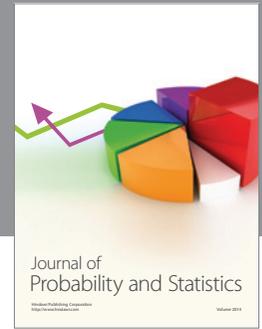
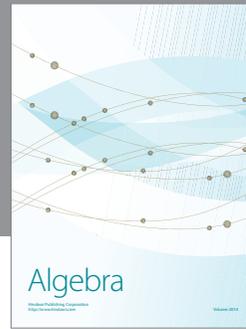
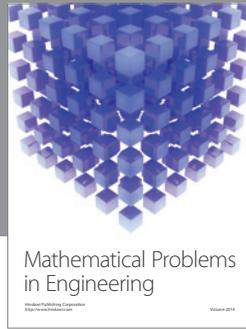
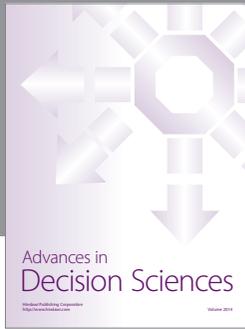
The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The author is grateful to the editor and to the referee for their helpful comments, which have led to this substantially improved version of the paper.

References

- [1] R. P. W. Duin, "On the choice of smoothing parameters for parzen estimators of probability density functions," *IEEE Transactions on Computers*, vol. C-25, no. 11, pp. 1175–1179, 1976.
- [2] M. Rudemo, "Empirical choice of histograms and kernel density estimators," *Scandinavian Journal of Statistics*, vol. 9, pp. 65–78, 1982.
- [3] D. W. Scott and G. R. Terrell, "Biased and unbiased cross-validation in density estimation," *Journal of the American Statistical Association*, vol. 82, pp. 1131–1146, 1987.
- [4] J. S. Marron, "Automatic smoothing parameter selection: a survey," *Empirical Economics*, vol. 13, no. 3-4, pp. 187–208, 1988.
- [5] M. Woodroffe, "On choosing a delta-sequence," *Annals of Statistics*, vol. 41, pp. 1665–1671, 1970.
- [6] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, UK, 1994.
- [7] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996.
- [8] J. S. Marron, "Bootstrap bandwidth selection," in *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, Eds., pp. 249–262, John Wiley, New York, NY, USA, 1992.
- [9] A. Mokkadem, M. Pelletier, and Y. Slaoui, "The stochastic approximation method for the estimation of a multivariate probability density," *Journal of Statistical Planning and Inference*, vol. 139, no. 7, pp. 2459–2478, 2009.
- [10] P. Révész, "Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I," *Studia Scientiarum Mathematicarum Hungarica*, vol. 8, pp. 391–398, 1973.
- [11] P. Révész, "How to apply the method of stochastic approximation in the nonparametric estimation of a regression function," *Mathematische Operationsforschung und Statistik, Series Statistics*, vol. 8, pp. 119–126, 1977.
- [12] A. B. Tsybakov, "Recurrent estimation of the mode of a multidimensional distribution," *Problems of Information Transmission*, vol. 8, pp. 119–126, 1990.
- [13] Y. Slaoui, "Large and moderate principles for recursive kernel density estimators defined by stochastic approximation method," *Serdica Mathematical Journal*, vol. 39, pp. 53–82, 2013.
- [14] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, vol. 27, pp. 832–837, 1956.
- [15] E. Parzen, "On estimation of a probability density and mode," *The Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.
- [16] J. Galambos and E. Seneta, "Regularly varying sequences," *American Mathematical Society*, vol. 41, pp. 110–116, 1973.
- [17] R. Bojanic and E. Seneta, "A unified theory of regularly varying sequences," *Mathematische Zeitschrift*, vol. 134, no. 2, pp. 91–106, 1973.
- [18] A. Mokkadem and M. Pelletier, "A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm," *Annals of Statistics*, vol. 35, no. 4, pp. 1749–1772, 2007.
- [19] N. Altman and C. Léger, "Bandwidth selection for kernel distribution function estimation," *Journal of Statistical Planning and Inference*, vol. 46, no. 2, pp. 195–214, 1995.
- [20] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, UK, 1986.
- [21] P. Hall and J. S. Marron, "Estimation of integrated squared density derivatives," *Statistics and Probability Letters*, vol. 6, no. 2, pp. 109–115, 1987.
- [22] W. Härdle and J. S. Marron, "Optimal bandwidth selection in nonparametric regression function estimation," *Annals of Statistics*, vol. 12, pp. 1465–1481, 1985.
- [23] A. Mokkadem, M. Pelletier, and Y. Slaoui, "Revisiting Révész's stochastic approximation method for the estimation of a regression function," *ALEA—Latin American Journal of Probability and Mathematical Statistics*, vol. 6, pp. 63–114, 2009.
- [24] J. D. Hart and P. Vieu, "Data-driven bandwidth choice for density estimation based on dependent data," *Annals of Statistics*, vol. 18, pp. 873–890, 1990.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

