Journal of the Indian Statistical Association
Vol. 54, No. 1 & 2, 2016, 193-220.

Smoothing parameters for deconvolution recursive kernel density estimators defined by stochastic approximation method with Laplace errors

Yousri Slaoui

Université de Poitiers, France

Abstract

In this paper, we propose a bandwidth selection in deconvolution recursive kernel estimators of a probability density function defined by the stochastic approximation algorithm for Laplace errors. We show that, using the proposed bandwidth selection and the stepsize which minimize the $MISE$ (Mean Integrated Squared Error), the recursive estimator will be better than the nonrecursive one for small sample setting and when the error variance is controlled by the noise to signal ratio. We corroborate these theoretical results through simulations and a real dataset.

Key Words : *Bandwidth selection, deconvolution, density estimation, plug-in methods, stochastic approximation algorithm*

1 Introduction

Suppose we observe contamined data $Y_1, \ldots, Y_n$ which are independent, identically distributed random variables, and let f_Y denote the probability density of Y_1, where

$$Y_i = X_i + \varepsilon_i, \qquad i = 1, \ldots, n$$

Received : March 2016

and where $X_1, \ldots, X_n$ are independent, identically distributed random variables, and f_X denote the probability density of X_1. We assume that X and ε are mutually independent. The density function of ε is denoted by f_ε, assumed known. Throught out this paper we suppose that ε is a centred double exponentielly distributed, also called Laplace distribution, and denoted by $\varepsilon \sim \mathcal{E}d(\sigma)$, with σ is the scale parameter. Following Robbins-Monro's procedure, we construct a stochastic algorithm, which approximates the function f_X at a given point x, by defining an algorithm of search of the zero of the function $h : y \rightarrow f_X(x) - y$. This algorithm can be defined by setting $f_{0,X}(x) \in \mathbb{R}$, and, for all $n \geq 1$,

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

where $W_n(x)$ is an observation of the function h at the point $f_{n-1,X}(x)$, and the stepsize (γ_n) is a sequence of positive real numbers that goes to zero. To define $W_n(x)$, we follow the approach of Révész (1973, 1977), Tsybakov (1990), Mokkadem et al. (2009a, b) and of Slaoui (2014b, 2015a), and we introduce a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and a kernel K (that is, a function satisfying $\int_{\mathbb{R}} K(x)\,dx = 1$), and a deconvoluting kernel K^ε defined as follows:

$$K^\varepsilon(u) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itu}\frac{\phi_K(t)}{\phi_\varepsilon\left(\frac{t}{h_n}\right)}dt,$$

with ϕ_L the Fourier transform of a function or a random variable L, and sets $W_n(x) = h_n^{-1}K^\varepsilon\left(h_n^{-1}(x - Y_n)\right) - f_{n-1,X}(x)$. Then, the estimator $f_{n,X}$ to recursively estimate the density function f_X at the point x can be written as

$$f_{n,X}(x) = (1-\gamma_n) f_{n-1,X}(x) + \gamma_n h_n^{-1} K^\varepsilon\left(h_n^{-1}(x - Y_n)\right). \tag{1}$$

This estimator was introduced by Mokkadem et al. (2009a) in the error-free context (i.e. when the data are observed without measurement errors) and whose large and moderate deviation principles were established by Slaoui (2013).

In this paper we suppose that $f_{0,X}(x) = 0$, and we let $\Pi_n = \prod_{j=1}^{n}(1-\gamma_j)$. Then the proposed estimator (1) can be rewritten as:

$$f_{n,X}(x) \quad = \quad \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^{-1} K^{\varepsilon}\left(\frac{x-Y_k}{h_k}\right). \tag{2}$$

The aim of this paper is to study the properties of the recursive deconvolution kernel density estimator defined by the stochastic approximation algorithm (1) for Laplace errors, and its comparison with the nonrecursive deconvolution kernel density estimator introduced by Carroll and Hall (1988); Stefanski and Carroll (1990), and defined as

$$\widetilde{f}_{n,X}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K^{\varepsilon}\left(\frac{x-Y_i}{h_n}\right). \tag{3}$$

This estimator have been investigated by Carroll and Hall (1988); Stefanski and Carroll (1990); Fan (1991a,b,c, 1992); among many others.

We first compute the bias and the variance of the recursive estimator $f_{n,X}$ defined by (1). It turns out that they heavily depend on the choice of the stepsize (γ_n). Moreover, we proposed a plug-in estimate which minimize an estimate of the Mean Weighted Integrated Squared Error ($MWISE$), using the density function as weight function to implement the bandwith selection of the proposed estimator.

The remainder of the paper is organized as follows. In Section 2, we state our main results. Section 3 is devoted to our application results, first by simulations (Subsection 3.1) and second using real dataset through a plug-in method (Subsection 3.2), we give our conclusion in Section 4, whereas the technical details are deferred to Section 5.

2 Assumptions and main results

We define the following class of regularly varying sequences.

Definition 1. *Let $\gamma \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if*

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{4}$$

Condition (4) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1973), and by Mokkadem and Pelletier (2007) in the context of stochastic approximation algorithms. Noting that the acronym $\mathcal{GS}$ stand for (Galambos & Seneta). Typical sequences in $\mathcal{GS}(\gamma)$ 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 shall refer are the following

(A1) $\varepsilon \sim \mathcal{E}d(\sigma)$, i.e. $f_{\varepsilon}(x) = \exp(-|x|/\sigma)/(2\sigma)$.

(A2) The function K equal to $K(x) = (2\pi)^{-1/2}\exp\left(-x^2/2\right)$.

(A3) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in (1/2, 1]$.
ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in (0,1)$.
iii) $\lim_{n\to\infty}(n\gamma_n) \in (\min\{2a, (\alpha - 5a)/2\}, \infty]$.

(A4) f_X is bounded, differentiable, and $f_X^{(2)}$ is bounded.

Throughout this paper we shall use the following notations:

$$\xi = \lim_{n\to\infty}(n\gamma_n)^{-1}, \tag{5}$$

$$\begin{aligned} I_1 &= \int_{\mathbb{R}} f_Y^2(x)\,dx, \qquad I_2 = \int_{\mathbb{R}}\left(f_X^{(2)}(x)\right)^2 f_Y(x)\,dx, \\ R(K) &= \int_{\mathbb{R}} K^2(z)\,dz, \quad \mu_j(K) = \int_{\mathbb{R}} z^j K(z)\,dz. \end{aligned}$$

Our first result is the following proposition, which gives the bias and the variance of $f_{n,X}$.

Proposition 1 (Bias and variance of $f_{n,X}$). *Let Assumptions* $(A1)-(A4)$ *hold, and assume that* $f_X^{(2)}$ *is continuous at* x*, then we have*

1. *If* $a \in (0, \alpha/9]$*, then*

$$\mathbb{E}\left[f_{n,X}\left(x\right)\right]-f_X\left(x\right)=\frac{1}{2\left(1-2a\xi\right)}h_n^2 f_X^{(2)}\left(x\right)+o\left(h_n^2\right). \tag{6}$$

If $a \in (\alpha/9, 1)$*, then*

$$\mathbb{E}\left[f_{n,X}\left(x\right)\right]-f_X\left(x\right)=o\left(\sqrt{\gamma_n h_n^{-5}}\right).$$

2. *If* $a \in [\alpha/9, 1)$*, then*

$$Var\left[f_{n,X}\left(x\right)\right] = \frac{3\sigma^4}{8\sqrt{\pi}}\frac{1}{2-\left(\alpha-5a\right)\xi}\frac{\gamma_n}{h_n^5}f_Y\left(x\right)+o\left(\frac{\gamma_n}{h_n^5}\right). \tag{7}$$

If $a \in (0, \alpha/9)$*, then*

$$Var\left[f_{n,X}\left(x\right)\right]=o\left(h_n^4\right).$$

3. *If* $\lim_{n\to\infty}\left(n\gamma_n\right) > \max\left\{2a, \left(\alpha-5a\right)/2\right\}$*, then (6) and (7) hold simultaneously.*

The bias and the variance of the estimator $f_{n,X}$ defined by the stochastic approximation algorithm (2) then heavily depend on the choice of the stepsize (γ_n). Let us now state the following theorem, which gives the weak convergence rate of the estimator $f_{n,X}$ defined in (2).

Theorem 1 (Weak pointwise convergence rate). *Let Assumptions* $(A1)-(A4)$ *hold, and assume that* $f_X^{(2)}$ *is continuous at* x*.*

1. *If there exists* $c \geq 0$ *such that* $\gamma_n^{-1}h_n^9 \to c$*, then*

$$\sqrt{\gamma_n^{-1}h_n^5}\left(f_{n,X}\left(x\right)-f_X\left(x\right)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\tfrac{\sqrt{c}}{2(1-2a\xi)}f_X^{(2)}\left(x\right), \tfrac{3\sigma^4}{8\sqrt{\pi}}\tfrac{1}{2-(\alpha-5a)\xi}f_Y\left(x\right)\right).$$

2. If $\gamma_n^{-1} h_n^9 \to \infty$, then

$$\frac{1}{h_n^2}\left(f_{n,X}(x) - f_X(x)\right) \xrightarrow{\mathbb{P}} \frac{1}{2(1-2a\xi)} f_X^{(2)}(x),$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution, $\mathcal{N}$ the Gaussian-distribution and $\xrightarrow{\mathbb{P}}$ the convergence in probability.

In order to measure the quality of our recursive estimator (2), we use the following quantity,

$$MWISE\left[f_{n,X}\right] = \int_{\mathbb{R}} \left(\mathbb{E}\left(f_{n,X}(x)\right) - f_X(x)\right)^2 f_Y(x)\,dx + \int_{\mathbb{R}} Var\left(f_{n,X}(x)\right) f_Y(x)\,dx.$$

Moreover, in the case $a = \alpha/9$, it follows from the proposition 1 that

$$MWISE\left[f_{n,X}\right] \simeq \frac{3\sigma^4}{8\sqrt{\pi}\left(2-(\alpha-5a)\xi\right)} \gamma_n h_n^{-5} I_1 + \frac{1}{4(1-2a\xi)^2} h_n^4 I_2. \quad (8)$$

The first term in (8) can be much larger than the variance component of the integrated mean squared error of an ordinary recursive kernel density estimator Mokkadem et al. (2009a). This is the price paid for not measuring $\{\varepsilon_i\}_{i=1}^n$ precisely. Corollary 1 gives the $MWISE$ of the deconvolution recursive kernel estimators (1) using the centred double exponentialle error distribution $f_\varepsilon(x) = \exp\left(-|x|/\sigma\right)/(2\sigma)$. Throughout this paper, we used standard normal kernel. The following corollary gives the bandwidth which minimize the $MWISE$.

Corollary 1. *Let Assumptions $(A1)-(A4)$ hold. To minimize the $MWISE$ of $f_{n,X}$, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$, the bandwidth (h_n) must equal*

$$\left(\left(\frac{15\sigma^4}{8\sqrt{\pi}}\right)^{1/9} \frac{(1-2a\xi)^{2/9}}{\left(2-(\alpha-5a)\xi\right)^{1/9}} \left\{\frac{I_1}{I_2}\right\}^{1/9} \gamma_n^{1/9}\right).$$

Then, the $MWISE\left[f_{n,X}\right]$

$$\simeq \frac{5}{4}\left(\frac{15\sigma^4}{8\sqrt{\pi}}\right)^{4/9} (1-2a\xi)^{-10/9} \left(2-(\alpha-5a)\xi\right)^{-4/9} I_1^{4/9} I_2^{5/9} \gamma_n^{4/9}.$$

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

Corollary 2. *Let Assumptions* $(A1)-(A4)$ *hold. To minimize the* $MWISE$ *of* $f_{n,X}$*, the stepsize* (γ_n) *must be chosen in* $\mathcal{GS}(-1)$, $\lim_{n\to\infty} n\gamma_n = \gamma_0$, *and the bandwidth* (h_n) *must equal*

$$\left(\left(\frac{15\sigma^4}{16\sqrt{\pi}}\right)^{1/9} (\gamma_0 - 2/9)^{1/9} \left\{\frac{I_1}{I_2}\right\}^{1/9} n^{-1/9}\right).$$

Then, the $MWISE$ *is*

$$MWISE\left[f_{n,X}\right] \simeq \frac{9}{20}\left(\frac{15\sigma^4}{16\sqrt{\pi}}\right)^{4/9} \frac{\gamma_0^2}{(\gamma_0 - 2/9)^{16/9}} I_1^{4/9} I_2^{5/9} n^{-4/9}.$$

Moreover, the minimum of $\gamma_0^2 (\gamma_0 - 2/9)^{-16/9}$ is reached at $\gamma_0 = 1$; then the bandwidth (h_n) must equal

$$\left(0.906\sigma^{4/9} \left\{\frac{I_1}{I_2}\right\}^{1/9} n^{-1/9}\right). \tag{9}$$

Then, the $MWISE$ is

$$MWISE\left[f_{n,X}\right] \simeq 0.530\,\sigma^{16/9} I_1^{4/9} I_2^{5/9} n^{-4/9}. \tag{10}$$

In order to estimate the optimal bandwidth (9), we must estimate I_1 and I_2. We followed the approach of Altman and Leger (1995), which is called the plug-in estimate, and we use the following kernel estimator of I_1 introduced in Slaoui (2014a) to implement the bandwidth selection in recursive kernel estimator of probability density function in the error-free context and in Slaoui (2014b) to implement the bandwidth selection in recursive kernel estimator of distribution function also in the error-free data context:

$$\widehat{I_1} = \frac{\Pi_n}{n} \sum_{i,k=1}^{n} \Pi_k^{-1} \alpha_k b_k^{-1} K_b^{\varepsilon}\left(\frac{Y_i - Y_k}{b_k}\right), \tag{11}$$

where K_b^{ε} is a deconvoluting kernel and b is the associated bandwidth, called the pilot bandwidth and α the pilot stepsize. In practice, we take

$$b_n = n^{-\beta} \min\left\{\widehat{s}, \frac{Q_3 - Q_1}{1.349}\right\}, \quad \beta \in (0,1) \tag{12}$$

(see Silverman (1986)) with $\widehat{s}$ the sample standard deviation, and Q_1, Q_3 denoting the first and third quartiles, respectively. In order to estimate I_1, we need to estimate the optimal pilot bandwidth and the optimal pilot stepsize. For this purpose, we should calculate the bias and variance of $\widehat{I}_1$, we followed the same steps as in Slaoui (2014a) and we showed that in order to minimize the $MISE$ of $\widehat{I}_1$, the pilot bandwidth (b_n) should belong to $\mathcal{GS}(-2/9)$, and the pilot stepsize (α_n) should be equal to $\left(1.93\, n^{-1}\right)$. Then to estimate I_1, we use $\widehat{I}_1$, with b_n equal to (12), and $\beta = 2/9$ and $(\alpha_n) = \left(1.93\, n^{-1}\right)$.
Furthermore, to estimate I_2, we followed the approach of Slaoui (2014a) and we introduced the following kernel estimator:

$$\begin{aligned}\widehat{I}_2 \quad = \quad & \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k=1 \\ j\neq k}}^{n} \Pi_j^{-1}\Pi_k^{-1}\alpha_j'\alpha_k' b_j'^{-3} b_k'^{-3} K_{b'}^{\varepsilon(2)}\left(\frac{Y_i - Y_j}{b_j'}\right) \\ & \times K_{b'}^{\varepsilon(2)}\left(\frac{Y_i - Y_k}{b_k'}\right),\end{aligned} \tag{13}$$

where $K_{b'}^{\varepsilon(2)}$ is the second order derivative of a deconvoluting kernel $K_{b'}$, and b' the associated bandwidth and α' the pilot stepsize. In order to estimate I_2, we need to estimate the optimal pilot bandwidth b_n' and the optimal pilot stepsize α_n'. For this purpose, we should calculate the bias and variance of $\widehat{I}_2$,we followed the same steps as in Slaoui (2014a) and we showed that in order to minimize the $MISE$ of $\widehat{I}_2$, the pilot bandwidth (b_n') should belong to $\mathcal{GS}(-3/22)$, and the pilot stepsize (α_n') should be equal to $\left(1.65\, n^{-1}\right)$. Then to estimate I_2, we use $\widehat{I}_2$, with b_n' equal to (12), and $\beta = 3/22$ and $(\alpha_n') = \left(1.65\, n^{-1}\right)$.
Finally, the plug-in estimator of the bandwidth (h_n) using the recursive

algorithm (2) must equal to

$$\left(0.906\sigma^{4/9}\left\{\frac{\widehat{I}_1}{\widehat{I}_2}\right\}^{1/9}n^{-1/9}\right). \tag{14}$$

Then, it follows from (10) that the $MWISE$ can be estimated by

$$\widehat{MWISE}\left[f_{n,X}\right] \simeq 0.530\,\sigma^{16/9}\widehat{I}_1^{4/9}\widehat{I}_2^{5/9}n^{-4/9}.$$

Now, let us recall that under the assumptions $(A1)$, $(A2)$, $(A3)\,ii)$ and $(A4)$, the $MWISE$ of the nonrecursive deconvolution kernel density estimator $\widetilde{f}_{n,X}$ (see Stefanski and Carroll (1990)) is given by

$$MWISE\left[\widetilde{f}_{n,X}\right] \simeq \frac{3\sigma^4}{8\sqrt{\pi}}\frac{1}{nh_n^5}I_1+\frac{1}{4}h_n^4I_2.$$

Lemma 1 gives the $MWISE$ of the deconvolution nonrecursive kernel density (3) estimator using the centred double exponentialle error distribution.

Lemma 1. *Let Assumptions* $(A1)$, $(A2)$, $(A3)\,ii)$ *and* $(A4)$ *hold. To minimize the MWISE of* $\widetilde{f}_{n,X}$, *the bandwidth* (h_n) *must equal*

$$\left(1.006\,\sigma^{4/9}\left\{\frac{I_1}{I_2}\right\}^{1/9}n^{-1/9}\right). \tag{15}$$

Then, the MWISE is

$$MWISE\left[\widetilde{f}_{n,X}\right] \simeq 0.461\,\sigma^{16/9}I_1^{4/9}I_2^{5/9}n^{-4/9}. \tag{16}$$

To estimate the optimal bandwidth (15), we must estimate I_1 and I_2. As suggested by Hall and Maron (1987), we we use the following two kernel estimators to estimate respectively I_1 and I_2:

$$\widetilde{I}_1 = \frac{1}{n\,(n-1)\,b_n}\sum_{\substack{i,j=1\\ i\neq j}}^{n}K_b^{\varepsilon}\left(\frac{Y_i-Y_j}{b_n}\right), \tag{17}$$

$$\widetilde{I}_2 = \frac{1}{n^3b_n'^6}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n}K_{b'}^{\varepsilon(2)}\left(\frac{Y_i-Y_j}{b_n'}\right)K_{b'}^{\varepsilon(2)}\left(\frac{Y_i-Y_k}{b_n'}\right). \tag{18}$$

The following Lemma gives the bias and variance of $\widetilde{I}_2$. We showed that in order to minimize the $MISE$ of $\widetilde{I}_1$ respectively of $\widetilde{I}_2$, the pilot bandwidth (b_n) respectively (b'_n) must belong to $\mathcal{GS}\left(-2/9\right)$, respectively to $\mathcal{GS}\left(-3/22\right)$. Then, the plug-in estimator of the bandwidth (h_n) using the nonrecursive algorithm (3) must equal to

$$\left(1.006\,\sigma^{4/9}\left\{\frac{\widetilde{I}_1}{\widetilde{I}_2}\right\}^{1/9} n^{-1/9}\right). \tag{19}$$

Then, it follows from (16) that the $MWISE$ can be estimated by

$$\widetilde{MWISE}\left[\widetilde{f}_{n,X}\right] \simeq 0.461\,\sigma^{16/9}\widetilde{I}_1^{4/9}\widetilde{I}_2^{5/9}n^{-4/9}.$$

The following Theorem gives the conditions under which the expected $MWISE$ of the recursive estimator $f_{n,X}$ will be smaller than the expected $MWISE$ of the nonrecursive estimator $\widetilde{f}_{n,X}$.

Theorem 2. *Let the assumptions* $(A1)$*-*$(A2)$ *hold, and the bandwidth* (h_n) *equal to (19) and the stepsize* $(\gamma_n) = \left(n^{-1}\right)$*. We have*

$$\frac{\mathbb{E}\left[\widehat{MWISE}\left[f_{n,X}\right]\right]}{\mathbb{E}\left[\widetilde{MWISE}\left[\widetilde{f}_{n,X}\right]\right]} < 1 \quad \textit{for small sample setting and small error variance}$$

Then, the expected $MWISE$ *of the recursive estimator defined by (2) is smaller than the expected* $MWISE$ *of the nonrecursive estimator defined by (3) for small sample setting and when the error variance is controlled by the noise to signal ratio.*

Following similar step as in Slaoui (2014a), we can proof Theorem 2.

3 Applications

The aim of our applications is to compare the performance of the nonrecursive deconvolution density kernel estimator defined in (3) with that of the recursive deconvolution density kernel estimators defined in (1).

3.1 Simulations

The aim of our simulation study is to compare the performance of the nonrecursive estimator defined in (3) with that of the recursive estimators defined in (2).

When applying $f_{n,X}$ one need to choose three quantities:

- The function K, we choose the standard normal kernel.
- The stepsize $(\gamma_n) = \left([5/9 + c]\, n^{-1}\right)$, with $c \in [0, 7/9]$.
- The bandwidth (h_n) is chosen to be equal to (14). To estimate I_1, we use the estimator $\widehat{I}_1$ given in (11), with K_b^{ε} is the standard normal kernel, the pilot bandwidth (b_n) is chosen to be equal to (12), with $\beta = 2/9$, and the pilot stepsize $(\alpha_n) = \left(1.93\, n^{-1}\right)$. Moreover, to estimate I_2, we use the estimator $\widehat{I}_2$ given in (13), with $K_{b'}^{\varepsilon}$ is the standard normal kernel, the pilot bandwidth (b'_n) is chosen to be equal to (12), with $\beta = 3/22$, and the pilot stepsize $(\alpha'_n) = \left(1.65\, n^{-1}\right)$.

When applying $\widetilde{f}_n$ one need to choose two quantities:

- The function K, as in the recursive framework, we use the normal kernel.
- The bandwidth (h_n) is chosen to be equal to (19). To estimate I_1, we used the estimator $\widetilde{I}_1$ given in (17), with K_b^{ε} is the standard normal kernel, the pilot bandwidth (b_n) is chosen to be equal to (12), with $\beta = 2/9$. Moreover, to estimate I_2, we used the estimator $\widetilde{I}_2$ given in (18), with $K_{b'}^{\varepsilon}$ is the standard normal kernel, the pilot bandwidth (b'_n) is chosen to be equal to (12), with $\beta = 3/22$.

In order to investigate the comparison between the two estimators, we consider $\varepsilon \sim \mathcal{E}d\,(\sigma)$ (i.e. centred double exponentielle with the scale parameter

σ). The error variance was controlled by the noise to signal ratio, denoted by NSR and defined by NSR $= Var(\varepsilon)/Var(X)$. We consider three sample sizes: $n = 50$, $n = 100$ and 500, and four density functions: the normal $\mathcal{N}(0,2)$ distribution (see Table 1), the chi-squared distribution with four degrees of freedom distribution $\chi^2(4)$ (see Table 2), the gamma mixture distribution $0.4\mathcal{G}(5) + 0.6\mathcal{G}(13)$ (see Table 3). For each of these four cases, 500 samples of sizes $n = 50$, $n = 100$ and 500 were generated. For each fixed NSR $\in [5\%, 30\%]$, the number of simulations is 500. We denote by f_i^* the reference density, and by f_i the test density, and then we compute the following measures : Mean squared Error ($MSE = n^{-1}\sum_i (f_i - f_i^*)^2$) and the linear Correlation ($Cor = \mathbb{C}ov(f_i, f_i^*)\,\sigma(f_i)^{-1}\,\sigma(f_i^*)^{-1}$).

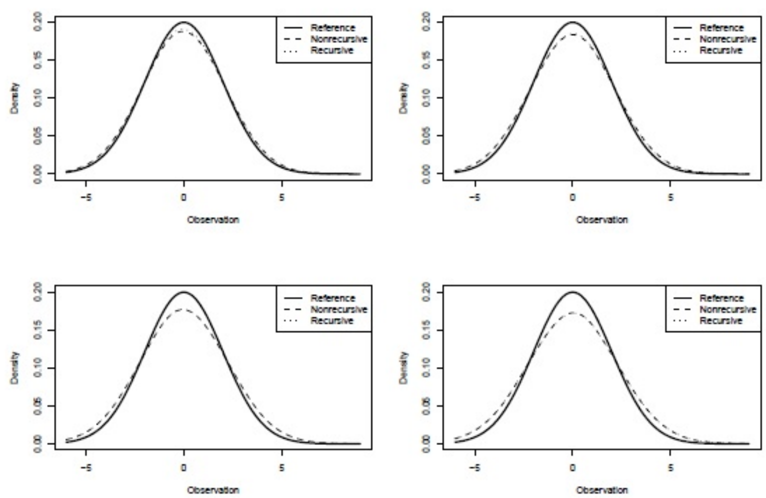

Figure 1: Qualitative comparison between the nonrecursive estimator (3) and the proposed estimator (1) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of size 200, with NSR equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the normal distribution $X \sim \mathcal{N}(0,2)$.

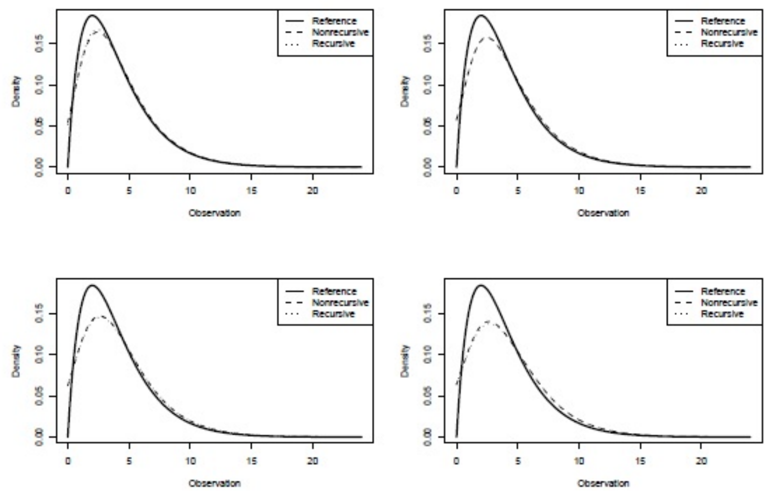

Figure 2: Qualitative comparison between the nonrecursive estimator (3) and the proposed estimator (1) with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$, for 500 samples of size 200, with `NSR` equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the chi-squared distribution with four degrees of freedom distribution $X \sim \chi(4)$.

From Tables 1, 2 and 3, we conclude that

(i) in all the cases of the Table 1, the MSE of the proposed density estimator (1), with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$ is smaller than the nonrecursive deconvolution density estimator (3).

(ii) in all the cases of the Table 2, the MSE of the proposed density estimator (1), with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$ is smaller than the nonrecursive deconvolution density estimator (3), when the `NSR` equal to 5% or 10%.

(iii) in all the cases of the Table 3, the MSE of the proposed density estimator (1), with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$ is smaller than the nonrecursive deconvolution density estimator (3), when the

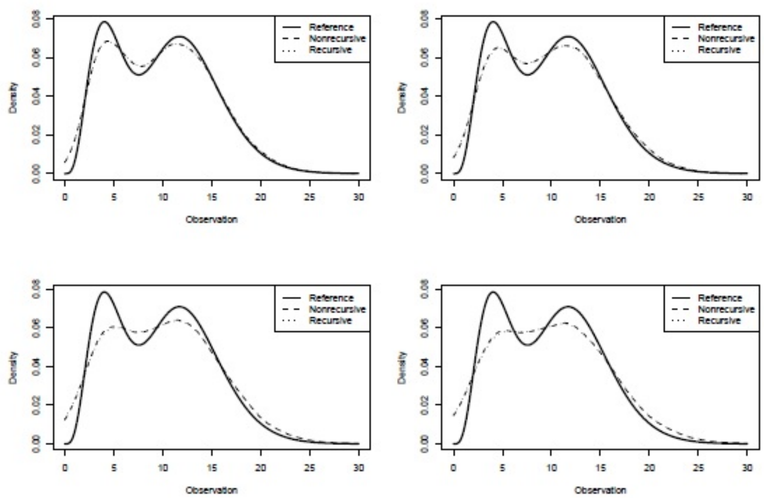

Figure 3: Qualitative comparison between the nonrecursive estimator (3) and the proposed estimator (1) with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$, for 500 samples of size 200, with NSR equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the gamma mixture distribution $X \sim 0.4\mathcal{G}(5) + 0.6\mathcal{G}(13)$.

NSR equal to 5% and $n = 50$ or $n = 100$.

(iv) the MSE decrease as the sample size increase.

(v) the MSE increase as the value of NSR increase.

(vi) the Cor increase as the sample size increase.

(vii) the Cor decrease as the value of NSR increase.

3.2 Real dataset

We use `Salvister` data which appears in the R package `kerdiest` (Quintela-del-Río and Estévez-Pérez (2012)). These data contains the annual peak instantaneous flow levels of the Salt River near Roosevelt, AZ, USA, for the period 1924-2009, obtained from the National Water Information System. In order to investigate the comparison between the two estimators, we consider the annual peak : for 500 samples of Laplacian errors $\varepsilon \sim \mathcal{E}d\,(\sigma)$, with NSR $\in [5\%, 30\%]$. For each fixed NSR, we computed the mean (over the 500 samples) of I_1, I_2, h_n and $MWISE$. The plug-in estimators (14), (19) requires two kernels to estimate I_1 and I_2. In both cases we use the normal kernel with b_n and b_n' are given in (12), with β equal respectively to 2/9 and 3/22.

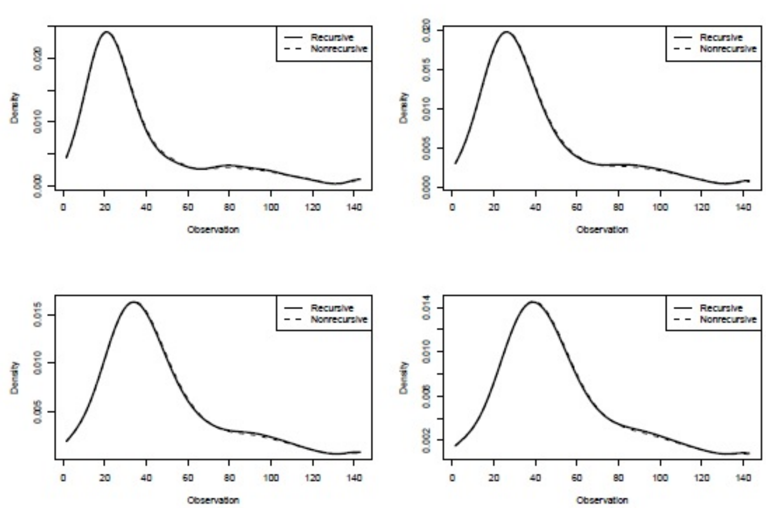

Figure 4: Qualitative comparison between the nonrecursive estimator (3) and the proposed estimator (1) with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$, for 500 samples of Laplacian errors with NSR equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the `Salvister` data of the package `kerdiest` and through a plug-in method.

From the table 4, we conclude that, the $\widehat{MWISE}$ of proposed estimator is quite similar to the $\widetilde{MWISE}$ of the nonrecursive estimator. From the Figures 1, 2, 3 and 4, we conclude that the two estimators present a quite similar behavior for all the fixed NSR.

4 Conclusion

This paper propose an automatic selection of the bandwidth of a probability density function in the case of deconvolution recursive kernel estimators. The estimators are compared to the nonrecursive deconvolution density estimator (3). We showed that using the selected bandwidth and the stepsizes $(\gamma_n) = \left(n^{-1}\right)$, the recursive estimator will be better than the nonrecursive one for small sample setting and when the error variance is controlled by the noise to signal ratio. The simulation study corroborated these theoretical results. Moreover, the simulation results indicate that the proposed recursive estimator was more computing efficiency than the nonrecursive estimator.

In conclusion, the proposed estimators allowed us to obtain quite better results then the nonrecursive estimator proposed by Carroll and Hall (1988); Stefanski and Carroll (1990). Moreover, we plan to make an extensions of our method in future and to consider the case of a regression function (see Mokkadem et al. (2009b) and Slaoui (2015a,b,c, 2016)) in the error-free context.

5 Technical proofs

Throughout this section we use the following notation:

$$Z_n\left(x\right) = h_n^{-1}K\left(\frac{x-X_n}{h_n}\right),$$

and

$$Z_n^{\varepsilon}\left(x\right) = h_n^{-1}K^{\varepsilon}\left(\frac{x-Y_n}{h_n}\right). \tag{20}$$

Let us first state the following technical lemma.

Lemma 2. *Let* $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, *and* $m > 0$ *such that* $m - v^*\xi > 0$ *where* ξ *is defined in* (5). *We have*

$$\lim_{n\to+\infty} v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\xi}.$$

Moreover, for all positive sequence (α_n) *such that* $\lim_{n\to+\infty} \alpha_n = 0$, *and all* $\delta \in \mathbb{R}$,

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + \delta \right] = 0.$$

Lemma 2 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption $(A3)(iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity.

Our proofs are organized as follows. Proposition 1 in Section 5.1, Theorem 1 in Section 5.2.

5.1 Proof of Proposition 1

Proof. In view of (2) and (20), we have

$$\begin{aligned} f_{n,X}(x) - f_X(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left(Z_k^{\varepsilon}(x) - f_X(x) \right) \\ &\quad + \Pi_n \left(f_{0,X}(x) - f_X(x) \right). \end{aligned} \tag{21}$$

Then, it follows that

$$\mathbb{E}\left(f_{n,X}(x)\right) - f_X(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left(\mathbb{E}\left(Z_k^{\varepsilon}(x)\right) - f_X(x) \right) + \Pi_n \left(f_{0,X}(x) - f_X(x) \right).$$

Moreover, an interchange of expectation and integration, justified by Fubini's Theorem and assumptions $(A1)$ and $(A2)$, shows that $\mathbb{E}\left\{Z_k^{\varepsilon}(x) | X_k\right\} = Z_k(x)$, which ensure that $\mathbb{E}\left[Z_k^{\varepsilon}(x)\right] = \mathbb{E}\left[Z_k(x)\right]$. Taylor's expansion with integral remainder implies that

$$\begin{aligned}\mathbb{E}\left[Z_k(x)\right]-f_X(x) &= \int_{\mathbb{R}} K(z)\left[f_X(x-zh_k)-f_X(x)\right]dz \\ &= \frac{h_k^2}{2}\mu_2(K)\, f_X^{(2)}(x)+h_k^2\delta_k(x)\end{aligned}$$

with

$$\delta_k(x)=h_k^{-2}\int_{\mathbb{R}} K(z)\left[f_X(x-zh_k)-f_X(x)-z^2\frac{h_k^2}{2}f_X^{(2)}(x)\right]dz,$$

and, since $f_X^{(2)}$ is bounded and continuous at x, we have $\lim_{k\to\infty}\delta_k(x)=0$. In the case $a\leq \alpha/9$, we have $\lim_{n\to\infty}(n\gamma_n)>2a$; the application of Lemma 2 then gives

$\mathbb{E}\left[f_{n,X}(x)\right]-f_X(x)$

$= \frac{1}{2}\mu_2(K)\, f_X^{(2)}(x)\,\Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^2[1+o(1)]+\Pi_n\left(f_{0,X}(x)-f_X(x)\right)$

$= \frac{1}{2(1-2a\xi)}\mu_2(K)\, f_X^{(2)}(x)\left[h_n^2+o(1)\right],$

(6) follows. In the case $a>\alpha/9$, we have $h_n^2=o\left(\sqrt{\gamma_n h_n^{-5}}\right)$; since $\lim_{n\to\infty}(n\gamma_n)>(\alpha-5a)/2$, Lemma 2 then ensures that

$$\begin{aligned}\mathbb{E}\left[f_{n,X}(x)\right]-f_X(x) &= \Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k o\left(\sqrt{\gamma_k h_k^{-5}}\right)+O(\Pi_n) \\ &= o\left(\sqrt{\gamma_n h_n^{-5}}\right),\end{aligned}$$

which gives (7). Now, we have

$$\begin{aligned}Var\left[f_{n,X}(x)\right] &= \Pi_n^2\sum_{k=1}^n \Pi_k^{-2}\gamma_k^2 Var\left[Z_k^{\varepsilon}(x)\right] \\ &= \Pi_n^2\sum_{k=1}^n \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[f_Y(x)\,R(K^{\varepsilon})+\nu_k(x)-h_k\tilde{\nu}_k(x)\right]\end{aligned}$$

with $\nu_k(x)=\int_{\mathbb{R}}(K^{\varepsilon}(z))^2\left[f_Y(x-zh_k)-f_Y(x)\right]dz$, and $\tilde{\nu}_k(x)=\left(\int_{\mathbb{R}}K(z)\, f_X(x-zh_k)\,dz\right)^2$. In view of $(A2)$ and $(A4)\,i)$, we have

$\lim_{k\to\infty}\nu_k(x)=0$ and $\lim_{k\to\infty}h_k\tilde{\nu}_k(x)=0$. Then, we have

$$Var\left[f_{n,X}(x)\right] = \frac{1}{2\pi}\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2h_k^{-1}R\left(K^{\varepsilon}\right)\left[f_Y(x)+o(1)\right].$$

Let us now state the following Lemma:

Lemma 3.

$$R\left(K^{\varepsilon}\right) = \frac{1}{2\sqrt{\pi}}\left(1+\left(\frac{\sigma}{h_k}\right)^2+\frac{3}{4}\left(\frac{\sigma}{h_k}\right)^4\right).$$

Proof of Lemma 3

Proof. First, we have

$$R\left(K^{\varepsilon}\right) = \frac{1}{2\pi}\int_R\phi_K^2(t)\left|\phi_{\varepsilon}(t/h_n)\right|^{-2}dt. \tag{22}$$

Moreover, since ε follows Laplace errors $\mathcal{L}(0,\sigma)$, we have $\phi_{\varepsilon}(t)=\frac{1}{(1+\sigma^2t^2)}$, and since $K\sim\mathcal{N}(0,1)$, we have $\phi_K(t)=\exp\left(-t^2/2\right)$. Then, it follows that from (22), that

$$R\left(K^{\varepsilon}\right)=\frac{1}{2\pi}\left\{\int_{\mathbb{R}}\exp\left(-t^2\right)dt+2h_n^{-2}\sigma^2\int_{\mathbb{R}}t^2\exp\left(-t^2\right)dt\right.$$
$$\left.+h_n^{-4}\sigma^4\int_{\mathbb{R}}t^4\exp\left(-t^2\right)dt\right\}$$

Moreover, we have $\int_{\mathbb{R}}\exp\left(-t^2\right)dt=\sqrt{\pi}$, $\int_{\mathbb{R}}t^2\exp\left(-t^2\right)dt=\sqrt{\pi}/4$ and $\int_{\mathbb{R}}t^4\exp\left(-t^2\right)dt=3\sqrt{\pi}/8$, which concludes the proof of Lemma 3. □

Now, when $a\geq\alpha/9$, we have $\lim_{n\to\infty}(n\gamma_n)>(\alpha-5a)/2$, and the application of Lemma 2 gives

$$Var\left[f_{n,X}(x)\right] = \frac{3\sigma^4}{8\sqrt{\pi}}\frac{1}{2\pi\left(2-(\alpha-5a)\xi\right)}\frac{\gamma_n}{h_n^5}\left[f_Y(x)+o(1)\right]$$

which proves (7). In the case $a<\alpha/9$, we have $\gamma_nh_n^{-5}=o\left(h_n^4\right)$; since $\lim_{n\to\infty}(n\gamma_n)>2a$, Lemma 2 then ensures that $Var\left[f_{n,X}(x)\right]=o\left(h_n^4\right)$, which gives (8). □

5.2 Proof of Theorem 1

Proof. Let us at first assume that, if $a \geq \alpha/9$, then

$$\sqrt{\gamma_n^{-1} h_n^5} \left(f_{n,X}(x) - \mathbb{E}\left[f_{n,X}(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{3\sigma^4}{8\sqrt{\pi}} \frac{1}{2\sqrt{\pi}\left(2 - (\alpha - 5a)\xi\right)} f_Y(x)\right), \quad (23)$$

In the case when $a > \alpha/9$, Part 1 of Theorem 1 follows from the combination of (7) and (23). In the case when $a = \alpha/9$, Parts 1 and 2 of Theorem 1 follow from the combination of (6) and (23). In the case $a < \alpha/5$, we have

$$h_n^{-2}\left(f_{n,X}(x) - f_X(x)\right) = \frac{1}{\sqrt{\gamma_n^{-1} h_n^9}} \sqrt{\gamma_n^{-1} h_n^5}\left(f_{n,X}(x) - \mathbb{E}\left[f_{n,X}(x)\right]\right) + h_n^{-2}\left(\mathbb{E}\left[f_{n,X}(x)\right] - f_X(x)\right)$$

and the application of (6) gives Part 2 of Theorem 1.

We now prove (23). In view of (2), we have

$$f_{n,X}(x) - \mathbb{E}\left[f_{n,X}(x)\right]$$

$$= (1 - \gamma_n)\left(f_{n-1,X}(x) - \mathbb{E}\left[f_{n-1,X}(x)\right]\right) + \gamma_n\left(Z_n^{\varepsilon}(x) - \mathbb{E}\left[Z_n^{\varepsilon}(x)\right]\right)$$

$$= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left(Z_k^{\varepsilon}(x) - \mathbb{E}\left[Z_k^{\varepsilon}(x)\right]\right).$$

Set

$$Y_k(x) = \Pi_k^{-1} \gamma_k \left(Z_k^{\varepsilon}(x) - \mathbb{E}\left(Z_k^{\varepsilon}(x)\right)\right).$$

The application of Lemma 2 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^{n} Var\left(Y_k\left(x\right)\right) \\ &= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var\left(Z_k^{\varepsilon}\left(x\right)\right) \\ &= \sum_{k=1}^{n} \frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[f_Y\left(x\right)R\left(K^{\varepsilon}\right)+o\left(1\right)\right] \\ &= \frac{3\sigma^4}{8\sqrt{\pi}}\frac{1}{2\sqrt{\pi}\left(2-\left(\alpha-5a\right)\xi\right)}\frac{1}{\Pi_n^2}\frac{\gamma_n}{h_n^5}\left[f_Y\left(x\right)+o\left(1\right)\right]. \end{aligned}$$

On the other hand, we have, for all $p>0$,

$$\mathbb{E}\left[\left|Z_k^{\varepsilon}\left(x\right)\right|^{2+p}\right] = O\left(\frac{1}{h_k^{(1+p)}}\right),$$

and, since $\lim_{n\to\infty}\left(n\gamma_n\right) > \left(\alpha-5a\right)/2$, there exists $p>0$ such that $\lim_{n\to\infty}\left(n\gamma_n\right) > \frac{1+p}{2+p}\left(\alpha-5a\right)$. Applying Lemma 2, we get

$$\begin{aligned} \sum_{k=1}^{n}\mathbb{E}\left[\left|Y_k\left(x\right)\right|^{2+p}\right] &= O\left(\sum_{k=1}^{n}\Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left[\left|Z_k^{\varepsilon}\left(x\right)\right|^{2+p}\right]\right) \\ &= O\left(\sum_{k=1}^{n}\frac{\Pi_k^{-2-p}\gamma_k^{2+p}}{h_k^{(1+p)}}\right) \\ &= O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}h_n^{(1+p)}}\right), \end{aligned}$$

and we thus obtain

$$\frac{1}{v_n^{2+p}}\sum_{k=1}^{n}\mathbb{E}\left[\left|Y_k\left(x\right)\right|^{2+p}\right] = O\left(\left[\gamma_n^{p/2}h_n^{4+3/2p}\right]\right) = o\left(1\right).$$

The convergence in (23) then follows from the application of Lyapounov's Theorem. □

References

[1] **Altman, N.** and **Leger, C.** (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46(2)**, 195-214.

[2] **Bojanic, R.** and **Seneta, E.** (1973). A unified theory of regularly varying sequences. *Mathematische Zeitschrift*, **134**, 91-106.

[3] **Carroll, R.J.** and **Hall, P.** (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association.* **83(404)**, 1184-1186.

[4] **Fan, J.** (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhya*, Series A, **53(1)**, 97-110.

[5] **Fan, J.** (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, **19(3)**, 1257-1272.

[6] **Fan, J.** (1991c). Global behaviour of deconvolution kernel estimates. *Statistica Sinica*, **1(2)**, 541-551.

[7] **Fan, J.** (1992). Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics*, **20(2)**, 155-169.

[8] **Galambos, J.** and **Seneta, E.** (1973). Regularly varying sequences. *Proceedings of the American Mathematical Society*, **41(1)**, 110-116.

[9] **Hall, P.** and **Maron, J. S.** (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters.* **6(2)**, 109-115.

[10] **Mokkadem, A.** and **Pelletier, M.** (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *The Annals of Statistics*, **35(4)**, 1749-1772.

[11] **Mokkadem, A. Pelletier, M.** and **Slaoui, Y.** (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *Journal of Statistical Planning and Inference*, **139(7)**, 2459-2478.

[12] **Mokkadem, A. Pelletier, M.** and **Slaoui, Y.** (2009b). Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA. Latin American Journal of Probability and Mathematical Statistics*, **6**, 63-114.

[13] **Quintela-del-Río, A.** and **Estévez-Pérez, G.** (2012). Nonparametric kernel distribution function estimation with kerdist: An R package for bandwidth choice and applications, *Journal of Statistical Software.* **50(8)**, 1-21.

[14] **Révész, P.** (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Scientiarum Mathematicarum Hungarica*, **8**, 391-398.

[15] **Révész, P.** (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Mathematical Operationsforsch Statistics, Series of Statistics*, **8(1)**, 119-126.

[16] **Silverman, B. W.** (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

[17] **Slaoui, Y.** (2013). Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method. *Serdica Mathematical Journal*, **39(1)**, 53-82.

[18] **Slaoui, Y.** (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *Journal of Probability & Statistics*, 2014, ID 739640, doi:10.1155/2014/739640.

[19] **Slaoui, Y.** (2014b). The stochastic approximation method for the estimation of a distribution function. *Mathematical Methods of Statistics*, **23(4)**, 306-325.

[20] **Slaoui, Y.** (2015a). Plug-In Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Statistica Neerlandica*, **69(4)**, 483-509.

[21] **Slaoui, Y.** (2015b). Moderate deviation principles for recursive regression estimators defined by stochastic approximation method. *International Journal of Mathematics and Statistics*, **16(2)**, 51-60.

[22] **Slaoui, Y.** (2015c). Large and moderate deviation principles for averaged stochastic approximation method for the estimation of a regression function. *Serdica Mathematical Journal*, **41(2-3)**, 307-328.

[23] **Y. Slaoui** (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators, *Statistics and its Interface*, **9(3)**, 375-388.

[24] **Stefanski, L. A.** and **Carroll, R. J.** (1990). Deconvoluting kernel density estimators. *Statistics*, 21, 169-184.

[25] **Tsybakov, A. B.** (1990). Recurrent estimation of the mode of a multidimensional distribution. *Problems of Information Transmission*, **26(1)**, 31-37.

Yousri Slaoui
Associate professor at the University of Poitiers
Laboratoire de Mathmatiques et Applications
UMR 7348 du CNRS, Téléport 2- BP 30179
11 Boulevard Marie et Pierre Curie
86962 Futuroscope Chasseneuil, France
E-mail: *Yousri.Slaoui@math.univ-poitiers.fr*

	nonrecursive	estimator 1	estimator 2	estimator 3	estimator 4
$n = 50$			NSR $= 5\%$		
MSE	$4.18e^{-05}$	$1.20e^{-04}$	$3.68e^{-05}$	$2.36e^{-05}$	$2.11e^{-05}$
Cor	0.99833	0.99887	0.99899	0.99904	0.99911
$n = 100$					
MSE	$3.12e^{-05}$	$6.90e^{-05}$	$2.38e^{-05}$	$1.80e^{-05}$	$1.79e^{-05}$
Cor	0.99864	0.99913	0.99918	0.99916	0.99912
$n = 500$					
MSE	$2.13e^{-05}$	$2.93e^{-05}$	$1.51e^{-05}$	$1.32e^{-05}$	$1.29e^{-05}$
Cor	0.99913	0.99937	0.99944	0.99946	0.99946
$n = 50$			NSR $= 10\%$		
MSE	$6.89e^{-05}$	$1.70e^{-04}$	$7.10e^{-05}$	$5.36e^{-05}$	$5.13e^{-05}$
Cor	0.99727	0.99780	0.99790	0.99789	0.99784
$n = 100$					
MSE	$6.61e^{-05}$	$1.26e^{-04}$	$6.45e^{-05}$	$5.35e^{-05}$	$5.11e^{-05}$
Cor	0.99725	0.99753	0.99771	0.99777	0.99781
$n = 500$					
MSE	$4.06e^{-05}$	$5.80e^{-05}$	$3.71e^{-05}$	$3.35e^{-05}$	$3.25e^{-05}$
Cor	0.99844	0.99847	0.99865	0.99873	0.99877
$n = 50$			NSR $= 20\%$		
MSE	$1.14e^{-03}$	$2.44e^{-03}$	$1.31e^{-03}$	$1.10e^{-03}$	$1.09e^{-03}$
Cor	0.99586	0.99615	0.99616	0.99601	0.99581
$n = 100$					
MSE	$1.05e^{-04}$	$1.93e^{-04}$	$1.17e^{-04}$	$1.02e^{-04}$	$9.98e^{-05}$
Cor	0.99591	0.99564	0.99595	0.99604	0.99607
$n = 500$					
MSE	$8.48e^{-05}$	$1.17e^{-04}$	$8.82e^{-05}$	$8.33e^{-05}$	$8.27e^{-05}$
Cor	0.99692	0.99665	0.99691	0.99698	0.99700

Table 1: Quantitative comparison between the nonrecursive estimator (3) and four estimators; estimator 1 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([5/9]\, n^{-1}\right)$, estimator 2 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([7/9]\, n^{-1}\right)$, estimator 3 correspond to the estimator (1) with the choice of $(\gamma_n) = \left(n^{-1}\right)$ and estimator 4 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([4/3]\, n^{-1}\right)$. Here we consider the normal distribution $X \sim \mathcal{N}(0, 2)$ with NSR $= 5\%$ in the first block, NSR $= 10\%$ in the second block and NSR $= 20\%$ in the last block, we consider three sample sizes $n = 50$, $n = 100$ and $n = 500$, the number of simulations is 500, and we compute the Mean squared error (MSE) and the linear correlation (Cor).

	nonrecursive	estimator 1	estimator 2	estimator 3	estimator 4
$n = 50$			NSR $= 5\%$		
MSE	$1.07e^{-04}$	$1.71e^{-04}$	$1.01e^{-04}$	$8.58e^{-05}$	$8.23e^{-05}$
Cor	0.98876	0.99014	0.99075	0.99094	0.99106
$n = 100$					
MSE	$8.80e^{-05}$	$1.18e^{-04}$	$7.789e^{-05}$	$7.12e^{-05}$	$7.15e^{-05}$
Cor	0.99080	0.99218	0.99246	0.99243	0.99229
$n = 500$					
MSE	$6.16e^{-05}$	$6.96e^{-05}$	$5.32e^{-05}$	$5.03e^{-05}$	$4.97e^{-05}$
Cor	0.99384	0.99444	0.99483	0.99493	0.99497
$n = 50$			NSR $= 10\%$		
MSE	0.000141	0.000235	0.000154	0.000136	0.000131
Cor	0.98571	0.98546	0.98613	0.98629	0.98640
$n = 100$					
MSE	0.000128	0.000186	0.000135	0.000125	0.000124
Cor	0.98736	0.98706	0.98760	0.98769	0.98772
$n = 500$					
MSE	$1.00e^{-04}$	$1.24e^{-04}$	$1.01e^{-04}$	$9.78e^{-05}$	$9.70e^{-05}$
Cor	0.99048	0.98991	0.99053	0.99071	0.99079
$n = 50$			NSR $= 20\%$		
MSE	0.000235	0.000380	0.000280	0.000255	0.000249
Cor	0.97731	0.97398	0.97515	0.97547	0.97559
$n = 100$					
MSE	0.000190	0.000288	0.000222	0.000208	0.000206
Cor	0.98223	0.97943	0.98039	0.98062	0.98066
$n = 500$					
MSE	0.000169	0.000217	0.000188	0.000183	0.000183
Cor	0.98483	0.98288	0.98350	0.98358	0.98354

Table 2: Quantitative comparison between the nonrecursive estimator (3) and four estimators; estimator 1 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([5/9]\, n^{-1}\right)$, estimator 2 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([7/9]\, n^{-1}\right)$, estimator 3 correspond to the estimator (1) with the choice of $(\gamma_n) = \left(n^{-1}\right)$ and estimator 4 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([4/3]\, n^{-1}\right)$. Here we consider the chi-squared distribution with four degrees of freedom distribution $X \sim \chi(4)$ with NSR $= 5\%$ in the first block, NSR $= 10\%$ in the second block and NSR $= 20\%$ in the last block, we consider three sample sizes $n = 50$, $n = 100$ and $n = 500$, the number of simulations is 500, and we compute the Mean squared error (MSE) and the linear correlation (Cor).

	nonrecursive	estimator 1	estimator 2	estimator 3	estimator 4
$n = 50$			$\texttt{NSR} = 5\%$		
MSE	$1.85e^{-05}$	$4.04e^{-05}$	$2.24e^{-05}$	$1.92e^{-05}$	$1.83e^{-05}$
Cor	0.99067	0.98930	0.98998	0.99030	0.99063
$n = 100$					
MSE	$1.83e^{-05}$	$3.19e^{-05}$	$2.10e^{-05}$	$1.90e^{-05}$	$1.83e^{-05}$
Cor	0.99044	0.98909	0.98976	0.99005	0.99031
$n = 500$					
MSE	$1.41e^{-05}$	$1.88e^{-05}$	$1.53e^{-05}$	$1.48e^{-05}$	$1.47e^{-05}$
Cor	0.99278	0.99188	0.99232	0.99244	0.99247
$n = 50$			$\texttt{NSR} = 10\%$		
MSE	$3.06e^{-05}$	$5.66e^{-05}$	$3.74e^{-05}$	$3.42e^{-05}$	$3.37e^{-05}$
Cor	0.98376	0.98136	0.98174	0.98174	0.98176
$n = 100$					
MSE	$2.84e^{-05}$	$4.64e^{-05}$	$3.39e^{-05}$	$3.17e^{-05}$	$3.13e^{-05}$
Cor	0.98561	0.98323	0.98374	0.98379	0.98374
$n = 500$					
MSE	$2.74e^{-05}$	$3.50e^{-05}$	$3.05e^{-05}$	$2.98e^{-05}$	$2.98e^{-05}$
Cor	0.98609	0.98425	0.98470	0.98475	0.98471
$n = 50$			$\texttt{NSR} = 20\%$		
MSE	$5.39e^{-05}$	$9.02e^{-05}$	$6.62e^{-05}$	$6.08e^{-05}$	$5.91e^{-05}$
Cor	0.97311	0.96760	0.96878	0.96936	0.96996
$n = 100$					
MSE	$5.27e^{-05}$	$7.62e^{-05}$	$6.16e^{-05}$	$5.88e^{-05}$	$5.84e^{-05}$
Cor	0.97295	0.96818	0.96925	0.96959	0.96963
$n = 500$					
MSE	$4.82e^{-05}$	$6.00e^{-05}$	$5.40e^{-05}$	$5.29e^{-05}$	$5.27e^{-05}$
Cor	0.97603	0.97252	0.97326	0.97342	0.97345

Table 3: Quantitative comparison between the nonrecursive estimator (3) and four estimators; estimator 1 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([5/9]\, n^{-1}\right)$, estimator 2 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([7/9]\, n^{-1}\right)$, estimator 3 correspond to the estimator (1) with the choice of $(\gamma_n) = \left(n^{-1}\right)$ and estimator 4 correspond to the estimator (1) with the choice of $(\gamma_n) = \left([4/3]\, n^{-1}\right)$. Here we consider the gamma mixture distribution $X \sim 0.4\mathcal{G}\,(5) + 0.6\mathcal{G}\,(13)$ with $\texttt{NSR} = 5\%$ in the first block, $\texttt{NSR} = 10\%$ in the second block and $\texttt{NSR} = 20\%$ in the last block, we consider three sample sizes $n = 50$, $n = 100$ and $n = 500$, the number of simulations is 500, and we compute the Mean squared error (MSE) and the linear correlation (Cor).

	I_1	I_2	h_n	$MWISE$
		NSR $= 5\%$		
Nonrecursive	$1.9286e^{-2}$	$1.1327e^{-8}$	6.0943	$7.2029e^{-6}$
Recursive	$1.8140e^{-2}$	$8.2467e^{-9}$	6.0633	$7.2020e^{-6}$
		NSR $= 10\%$		
Nonrecursive	$1.9286e^{-2}$	$1.1339e^{-8}$	7.1083	$1.3346e^{-5}$
Recursive	$1.8143e^{-2}$	$8.2668e^{-9}$	7.0712	$1.3355e^{-5}$
		NSR $= 20\%$		
Nonrecursive	$1.9291e^{-2}$	$1.1338e^{-8}$	8.2924	$2.4714e^{-5}$
Recursive	$1.8147e^{-2}$	$8.2666e^{-9}$	8.2490	$2.4733e^{-5}$
		NSR $= 30\%$		
Nonrecursive	$1.9291e^{-2}$	$1.1342e^{-8}$	9.0739	$3.5446e^{-5}$
Recursive	$1.8145e^{-2}$	$8.2626e^{-9}$	9.0271	$3.5453e^{-5}$

Table 4: The comparison between the $MWISE$ of the Nadaraya's distribution estimator (3) and the $MWISE$ of the proposed distribution estimator (1) with the choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$ via the `Salvister` data of the package `kerdiest` and through a plug-in method, with `NSR` equal to 5% in the first block, 10% in the second block, 20% in the third block and 30% in the last block the number of simulations is 500.