J. Japan Statist. Soc.
Vol. 46 No. 1 2016 1–26

BANDWIDTH SELECTION IN DECONVOLUTION KERNEL DISTRIBUTION ESTIMATORS DEFINED BY STOCHASTIC APPROXIMATION METHOD WITH LAPLACE ERRORS

Yousri Slaoui*

In this paper we consider the kernel estimators of a distribution function defined by the stochastic approximation algorithm when the observation are contamined by measurement errors. It is well known that this estimators depends heavily on the choice of a smoothing parameter called the bandwidth. We propose a specific second generation plug-in method of the deconvolution kernel distribution estimators defined by the stochastic approximation algorithm. We show that, using the proposed bandwidth selection and the stepsize which minimize the *MISE* (Mean Integrated Squared Error), the proposed estimator will be better than the classical one for small sample setting 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 and phrases: Bandwidth selection, deconvolution, distribution estimation, plug-in methods, stochastic approximation algorithm.

1. Introduction

We suppose that we observe the contamined data $Y_1, \dots, Y_n$ instead of the uncontamined data $X_1, \dots, X_n$, where $Y_1, \dots, Y_n$ are generated from an additive measurement error model

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

and where $X_1, \dots, X_n$ are independent, identically distributed random variables, and let f_X and F_X denote respectively the probability density and the distribution function of X_1, the errors $\varepsilon_1, \dots, \varepsilon_n$ are identically distributed random variables. We assume that X and ε are mutually independent. The distribution function of ε is denoted by F_ε, assumed known. This problem is motivated by a wide set of practical applications in different fields such as, for example, astronomy, public health, and econometrics. In the classical deconvolution literature, the error distributions are classified into two classes: Ordinary smooth distribution and supersmooth distribution Fan (1991). Examples of ordinary smooth distributions include Laplacian, gamma, and symmetric gamma; examples of supersmooth distributions are normal, mixture normal and Cauchy. From a theoretical point of view, the rate of convergence cannot be faster than logarithmic for supersmooth errors, whereas for ordinary smooth errors the rate of

Received November 22, 2015. Revised March 17, 2016. Accepted April 21, 2016.

*Université de Poitiers, Laboratoire de Mathématiques et Application, Futuroscope Chasseneuil, France. Email: Yousri.Slaoui@math.univ-poitiers.fr

convergence of F_X is of a much better polynomial rate. For a practical point of view, Delaigle and Gijbels (2004) noted that the deconvolution estimators that assume Laplace error always gives better results than the Gaussian case, and as an application, they consider data from the second National Health and Nutrition Examination Survey (NHANES), which is a cohort study consisting of thousands of women who were investigated about their nutrition habits and then evaluated for evidence of cancer. The primary variable of interest in the study of the long-term log daily saturated fat intake which was known to be imprecisely measured, for more details, see Stefanski and Caroll (1990) and Carroll *et al.* (1995). 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. To construct a stochastic algorithm, which approximates the function F_X at a given point x, we define an algorithm of search of the zero of the function $h : y \to F_X(x) - y$. Following Robbins-Monro's procedure, this algorithm is 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 Slaoui (2013, 2014a, b) 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 $\int_{\mathbb{R}} K(x)dx = 1$), a function $\mathcal{K}$ (that is, a function defined by $\mathcal{K}(z) = \int_{-\infty}^{z} K(u)du$), and a deconvoluting kernel K^{ε} defined as follows:

$$K^{\varepsilon}(u) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itu}\frac{\phi_K(t)}{\phi_{\varepsilon}\left(\dfrac{t}{h_n}\right)}dt, \tag{1.1}$$

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

$$F_{n,X}(x) = (1-\gamma_n)F_{n-1,X}(x) + \gamma_n\mathcal{K}^{\varepsilon}(h_n^{-1}(x - Y_n)). \tag{1.2}$$

This estimator was introduced by Slaoui (2014b) in the error-free data. Now, we suppose that $F_0(x) = 0$, and we let $\Pi_n = \prod_{j=1}^{n}(1-\gamma_j)$. Then in this paper we propose to study the following estimator of F at the point x:

$$F_{n,X}(x) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\mathcal{K}^{\varepsilon}\left(\frac{x-Y_k}{h_k}\right). \tag{1.3}$$

The aim of this paper is to study the properties of the proposed deconvolution kernel distribution estimator defined by the stochastic approximation algorithm (1.2), and its comparison with the deconvolution Nadaraya's kernel distribution estimator defined as

$$\widetilde{F}_{n,X}(x) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}^{\varepsilon}\left(\frac{x-Y_i}{h_n}\right). \tag{1.4}$$

This estimator was introduced by Nadaraya (1964) in the error-free data and whose large and moderate deviation principles were established by Slaoui (2014c) in the context of error-free data. We first compute the bias and the variance of the proposed estimator $F_{n,X}$ defined by (1.2). It turns out that they heavily depend on the choice of the stepsize (γ_n), and on the distribution of ε and on the kernel K. Moreover, we proposed a plug-in estimate which minimize an estimate of the mean weighted integrated squared error, 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*

$$\lim_{n\to+\infty} n\left[1-\frac{v_{n-1}}{v_n}\right] = \gamma. \tag{2.1}$$

Condition (2.1) 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 and 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 $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$.

(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) $\lim_{n\to\infty}(n\gamma_n) \in (\min\{2a, (\alpha-3a)/2\}, \infty]$.

(A4) f_X is bounded, differentiable, and f_X' is bounded.

Remark 1. Assumption (A3)(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 shall use the following notations:

$$\begin{aligned}
\xi &= \lim_{n\to\infty} (n\gamma_n)^{-1}, \\
\Pi_n &= \prod_{j=1}^{n}(1-\gamma_j), \\
I_1 &= \int_{\mathbb{R}} f_Y^2(x)dx, \qquad I_2 = \int_{\mathbb{R}} (f_X'(x))^2 f_Y(x)dx.
\end{aligned} \tag{2.2}$$

Our first result is the following Proposition, which gives the bias and the variance of the proposed recursive deconvolution kernel distribution function.

PROPOSITION 1 (Bias and variance of $F_{n,X}$). *Let Assumptions* (A1)–(A4) *hold, and assume that f_X' is continuous at x, then we have*

1. *If $a \in (0, \alpha/7]$, then*

$$\mathbb{E}[F_{n,X}(x)] - F_X(x) = \frac{1}{2(1-2a\xi)} h_n^2 f_X'(x) + o(h_n^2). \tag{2.3}$$

   If $a \in (\alpha/7, 1)$, then

$$\mathbb{E}[F_{n,X}(x)] - F_X(x) = o\left(\sqrt{\gamma_n h_n^{-3}}\right). \tag{2.4}$$

2. *If $a \in [\alpha/7, 1)$, then*

$$\operatorname{Var}[F_{n,X}(x)] = \frac{\sigma^4}{4\sqrt{\pi}} \frac{1}{(2-(\alpha-3a)\xi)} \frac{\gamma_n}{h_n^3} f_Y(x) + o\left(\frac{\gamma_n}{h_n^3}\right). \tag{2.5}$$

   If $a \in (0, \alpha/7)$, then

$$\operatorname{Var}[F_{n,X}(x)] = o(h_n^4). \tag{2.6}$$

3. *If* $\lim_{n\to\infty}(n\gamma_n) > \max\{2a, (\alpha-3a)/2\}$, *then* (2.3) *and* (2.5) *hold simultaneously.*

The bias and the variance of the estimator $F_{n,X}$ defined by the stochastic approximation algorithm (1.3) 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 (1.3).

THEOREM 1 (Weak pointwise convergence rate). *Let Assumptions* (A1)–(A4) *hold, and assume that f_X' is continuous at x.*

1. *If there exists $c \geq 0$ such that $\gamma_n^{-1} h_n^7 \to c$, then*

$$\begin{aligned}
&\sqrt{\gamma_n^{-1} h_n^3}\,(F_{n,X}(x) - F_X(x)) \\
&\quad \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{2(1-2a\xi)} f_X'(x), \frac{\sigma^4}{4\sqrt{\pi}} \frac{1}{2-(\alpha-3a)\xi} f_Y(x)\right).
\end{aligned}$$

2. *If* $\gamma_n^{-1}h_n^7 \to \infty$, *then*

$$\frac{1}{h_n^2}(F_{n,X}(x) - F_X(x)) \stackrel{\mathbb{P}}{\to} \frac{1}{2(1-2a\xi)} f'_X(x),$$

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

The convergence rate of the proposed estimator (1.3) is smaller than the ordinary kernel distribution estimator Slaoui (2014b). This is the price paid for not measuring $\{\varepsilon_i\}_{i=1}^n$ precisely. In order to measure the quality of our proposed estimator (1.3), we use the following quantity,

$$\begin{aligned} MISE^*[F_{n,X}] &= \mathbb{E}\int_{\mathbb{R}} [F_{n,X}(x) - F_X(x)]^2 f_Y(x)dx \\ &= \int_{\mathbb{R}} (\mathbb{E}(F_{n,X}(x)) - F_X(x))^2 f_Y(x)dx + \int_{\mathbb{R}} \mathrm{Var}(F_{n,X}(x)) f_Y(x)dx. \end{aligned}$$

Moreover, in the case $a = \alpha/7$, it follows from the Proposition 1 that

$$AMISE^*[F_{n,X}] = \frac{\sigma^4}{4\sqrt{\pi}(2-(\alpha-3a)\xi)}\gamma_n h_n^{-3} I_1 + \frac{1}{4(1-2a\xi)^2} h_n^4 I_2. \tag{2.7}$$

Let us underline that first term in (2.7) can be larger than the variance component of the integrated mean squared error of the proposed kernel distribution estimator with error free data Slaoui (2014b). Corollary 1 gives the $AMISE^*$ of the proposed deconvolution kernel estimators (1.2) using the centred double exponentialle error distribution $f_\varepsilon(x) = \exp(-|x|/\sigma)/(2\sigma)$. Throughout this paper, we used the standard normal kernel. The following corollary gives the bandwidth which minimize the $AMISE^*$ and the corresponding $AMISE^*$.

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

$$\left(\left(\frac{3\sigma^4}{4\sqrt{\pi}}\right)^{1/7} \frac{(1-2a\xi)^{2/7}}{(2-(\alpha-3a)\xi)^{1/7}} \left\{\frac{I_1}{I_2}\right\}^{1/7} \gamma_n^{1/7}\right).$$

Then, the asymptotic dominating term of the $MISE^*$ *is*

$$AMISE^*[F_{n,X}] = \frac{7}{12}\left(\frac{3\sigma^4}{4\sqrt{\pi}}\right)^{4/7} (1-2a\xi)^{-6/7}(2-(\alpha-3a)\xi)^{-4/7} I_1^{4/7} I_2^{3/7} \gamma_n^{4/7}.$$

The following corollary shows that, for a special choice of the stepsize $(\gamma_n) = (\gamma_0 n^{-1})$, 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 $AMISE^*$ depend on γ_0.

Corollary 2. *Let Assumptions* (A1)–(A4) *hold. To minimize the* $AMISE^*$ *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*

$$\left(\left(\frac{3\sigma^4}{8\sqrt{\pi}}\right)^{1/7}(\gamma_0-2/7)^{1/7}\left\{\frac{I_1}{I_2}\right\}^{1/7}n^{-1/7}\right). \tag{2.8}$$

Then, the asymptotic dominating term of the $MISE^*$ *is*

$$AMISE^*[F_{n,X}] = \frac{7}{12}\left(\frac{3\sigma^4}{8\sqrt{\pi}}\right)^{4/7}\frac{\gamma_0^2}{(\gamma_0-2/7)^{10/7}}I_1^{4/7}I_2^{3/7}n^{-4/7}.$$

Moreover, the minimum of $\gamma_0^2(\gamma_0-2/7)^{-10/7}$ is reached at $\gamma_0=1$; then the bandwidth (h_n) must equal

$$\left(0.7634\sigma^{4/7}\left\{\frac{I_1}{I_2}\right\}^{1/7}n^{-1/7}\right). \tag{2.9}$$

Then, the asymptotic dominating term of the $MISE^*$ is

$$AMISE^*[F_{n,X}] = 0.3883\sigma^{16/7}I_1^{4/7}I_2^{3/7}n^{-4/7}. \tag{2.10}$$

In order to estimate the optimal bandwidth (2.9), 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:

$$\widehat{I}_1 = \frac{\Pi_n}{n}\sum_{i,k=1}^{n}\Pi_k^{-1}\gamma_k b_k^{-1}K_b^{\varepsilon}\left(\frac{Y_i-Y_k}{b_k}\right), \tag{2.11}$$

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

$$b_n = n^{-\beta}\min\left\{\widehat{s},\frac{Q_3-Q_1}{1.349}\right\},\qquad \beta\in(0,1) \tag{2.12}$$

(see Silverman (1986)) with $\widehat{s}$ the sample standard deviation, and Q_1, Q_3 denoting the first and third quartiles, respectively. We followed simlar steps as in the previous works (Slaoui (2014a, 2015a)), we prove 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 stepsize (γ_n) should be equal to $(1.93n^{-1})$. Then to estimate I_1, we use $\widehat{I}_1$, with

b_n equal to (2.12), and $\beta = 2/9$. Furthermore, to estimate I_2, we followed the approach of Slaoui (2014a) and we introduced the following kernel estimator:

$$\widehat{I}_2 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k=1 \\ j \neq k}}^{n} \Pi_j^{-1} \Pi_k^{-1} \gamma_j \gamma_k b_j'^{-2} b_k'^{-2} K_{b'}^{\varepsilon(1)} \left(\frac{Y_i - Y_j}{b_j'} \right) K_{b'}^{\varepsilon(1)} \left(\frac{Y_i - Y_k}{b_k'} \right), \tag{2.13}$$

where $K_{b'}^{\varepsilon(1)}$ is the first order derivative of a deconvoluting kernel $K_{b'}$, and b' the associated bandwidth. Following similar steps as in the previous works (Slaoui (2014a, 2015a)), we prove that in order to minimize the $MISE$ of $\widehat{I}_2$, the pilot bandwidth (b_n) should belong to $\mathcal{GS}(-1/6)$, and the stepsize (γ_n) should be equal to $(1.736 n^{-1})$. Then to estimate I_2, we use $\widehat{I}_2$, with b_n equal to (2.12), and $\beta = 1/6$.

Finally, the plug-in estimator of the bandwidth (h_n) using the proposed algorithm (1.3) must be equal to

$$\left(0.7634 \sigma^{4/7} \left\{ \frac{\widehat{I}_1}{\widehat{I}_2} \right\}^{1/7} n^{-1/7} \right). \tag{2.14}$$

Then, it follows from (2.10) that the asymptotic dominating term of the $MISE^*$ can be estimated by

$$\widehat{AMISE}^*[F_{n,X}] = 0.3883 \sigma^{16/7} \widehat{I}_1^{4/7} \widehat{I}_2^{3/7} n^{-4/7}.$$

Now, let us recall that under the assumptions (A1), (A2), (A3)ii) and (A4), the asymptotic dominating term of the $MISE^*$ of the deconvolution Nadaraya's kernel distribution estimator $\widetilde{F}_{n,X}$ is given by

$$AMISE^*[\widetilde{F}_{n,X}] = \frac{\sigma^4}{4\sqrt{\pi}} \frac{1}{n h_n^3} I_1 + \frac{1}{4} h_n^4 I_2.$$

Lemma 1 gives the $AMISE^*$ of the deconvolution Nadaraya's kernel distibution (1.4) estimator using the centred double exponentialle error distribution.

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

$$\left(0.884 \sigma^{4/7} \left\{ \frac{I_1}{I_2} \right\}^{1/7} n^{-1/7} \right). \tag{2.15}$$

Then, the asymptotic dominating term of the $MISE^*$ *is*

$$AMISE^*[\widetilde{F}_{n,X}] = 0.357 \sigma^{16/7} I_1^{4/7} I_2^{3/7} n^{-4/7}. \tag{2.16}$$

To estimate the optimal bandwidth (2.15), we must estimate I_1 and I_2. As suggested by Hall and Maron (1987), we use the following kernel estimator of I_1:

$$\widetilde{I}_1 = \frac{1}{n(n-1)b_n} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} K_b^{\varepsilon} \left(\frac{Y_i - Y_j}{b_n} \right). \tag{2.17}$$

where (b_n) equal to (2.12), with $\beta = 2/9$. and to estimate I_2, we use the following kernel estimator:

$$\widetilde{I}_2 = \frac{1}{n^3 b_n^4} \sum_{\substack{i,j,k=1 \\ j \neq k}}^{n} K_{b'}^{\varepsilon(1)} \left(\frac{Y_i - Y_j}{b'_n} \right) K_{b'}^{\varepsilon(1)} \left(\frac{Y_i - Y_k}{b'_n} \right), \tag{2.18}$$

where (b'_n) equal to (2.12), with $\beta = 1/6$. Finally, the plug-in estimator of the bandwidth (h_n) using the deconvolution Nadaraya's kernel distribution estimator (1.4) must be equal to

$$\left(0.884 \sigma^{4/7} \left\{ \frac{\widetilde{I}_1}{\widetilde{I}_2} \right\}^{1/7} n^{-1/7} \right). \tag{2.19}$$

Then, it follows from (2.16) that the asymptotic dominating term of the $MISE^*$ can be estimated by

$$AMISE^*[\widetilde{F}_{n,X}] = 0.357 \sigma^{16/7} \widetilde{I}_1^{4/7} \widetilde{I}_2^{3/7} n^{-4/7}.$$

The following Theorem gives the conditions under which the expected $AMISE^*$ of the proposed estimator $F_{n,X}$ will be smaller than the expected $AMISE^*$ of the deconvolution Nadaraya's kernel distribution estimator $\widetilde{F}_{n,X}$. Following similar steps as in Slaoui (2014a) and Slaoui (2015a), we prove the following Theorem:

THEOREM 2. *Let the assumptions* (A1)–(A4) *hold, and the stepsize* $(\gamma_n) = (n^{-1})$. *We have*

$$\frac{\mathbb{E}[AMISE^*[F_{n,X}]]}{\mathbb{E}[AMISE^*[\widetilde{F}_{n,X}]]} < 1 \quad \textit{for small sample setting} \tag{2.20}$$

Then, the expected $AMISE^*$ *of the proposed estimator defined by* (1.3) *is smaller than the expected* $AMISE^*$ *of the deconvolution Nadaraya's kernel distribution estimator defined by* (1.4) *for small sample setting.*

3. Applications

The aim of our applications is to compare the performance of the deconvolution Nadaraya's kernel estimator defined in (1.4) with that of the proposed deconvolution distribution kernel estimators defined in (1.2).

3.1. Simulations

The aim of our simulation study is to compare the performance of the deconvolution Nadaraya's kernel estimator defined in (1.4) with that of the proposed deconvolution distribution kernel estimators defined in (1.3).

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

- The function K, we choose the standard normal kernel.
- The stepsize $(\gamma_n) = ([2/3 + c]n^{-1})$, with $c \in [0, 1]$.
- The bandwidth (h_n) is chosen to be equal to (2.8). To estimate I_1, we use the estimator $\widehat{I}_1$ given in (2.11), with K_b^ε is the standard normal kernel, the pilot bandwidth (b_n) is chosen to be equal to (2.12), with $\beta = 2/9$, and $(\gamma_n) = (1.93n^{-1})$. Moreover, to estimate I_2, we use the estimator $\widehat{I}_2$ given in (2.13), with $K_{b'}^\varepsilon$ is the standard normal kernel, the pilot bandwidth (b'_n) is chosen to be equal to (2.12), with $\beta = 1/6$, and $(\gamma_n) = (1.736n^{-1})$.

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

- The function K, we use the normal kernel.
- The bandwidth (h_n) is chosen to be equal to (2.15). To estimate I_1, we used the estimator $\widetilde{I}_1$ given in (2.17), with K_b^ε is the standard normal kernel, the pilot bandwidth (b_n) is chosen to be equal to (2.12), with $\beta = 2/9$. Moreover, to estimate I_2, we used the estimator $\widetilde{I}_2$ given in (2.18), with $K_{b'}^\varepsilon$ is the standard normal kernel, the pilot bandwidth (b'_n) is chosen to be equal to (2.12), with $\beta = 1/6$.

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 $\texttt{NSR} = \operatorname{Var}(\varepsilon)/\operatorname{Var}(X)$. We consider three sample sizes: $n = 25$, $n = 50$ and 150, and five distribution functions: normal $\mathcal{N}(0, 1/2)$ (see Table 1), standard normal $\mathcal{N}(0, 1)$ (see Table 2), normal $\mathcal{N}(0, 2)$ distribution (see Table 3), the normal mixture $\frac{1}{2}\mathcal{N}(1/2, 1) + \frac{1}{2}\mathcal{N}(-1/2, 1)$ (see Table 4), the exponential distribution of parameter 1/2 $\mathcal{E}(1/2)$ (see Table 5). For each of these five cases, 500 samples of sizes $n = 25$, $n = 50$ and 150 were generated. For each fixed $\texttt{NSR} \in [5\%, 30\%]$, the number of simulations is 500. We denote by F_i^* the reference distribution, and by F_i the test distribution, and then we compute the following measures: Robust Mean Relative Error ($RMRE = n^{-1} \sum_{i,|F_i|>\varepsilon} |\frac{F_i}{F_i^*} - 1|$), (which simply is the mean relative error obtained by removing the observations close to zero) and the linear Correlation ($\mathrm{Cor} = \mathbb{C}\mathrm{ov}(F_i, F_i^*)\sigma(F_i)^{-1}\sigma(F_i^*)^{-1}$).

From Tables 1, 2, 3, 4 and 5, we conclude that

(i) in all the cases, the $RMRE$ of the proposed distribution estimator (1.2), with the choice of the stepsize $(\gamma_n) = (n^{-1})$ is smaller than the deconvolution Nadaraya's kernel distribution estimator (1.4).
(ii) the $RMRE$ decrease as the sample size increase.
(iii) the $RMRE$ increase as the value of `NSR` increase.
(iv) the `CPU` time are approximately two times faster using the proposed distribution estimator (1.2) compared to the deconvolution Nadaraya's kernel

Table 1. Quantitative comparison between the deconvolution Nadaraya's estimator (1.4) and four proposed estimators; estimator 1 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([2/3]n^{-1})$, estimator 2 correspond to the estimator (1.2) with the choice of $(\gamma_n) = (n^{-1})$, estimator 3 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([4/3]n^{-1})$ and estimator 4 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([5/3]n^{-1})$. Here we consider the normal distribution $X \sim \mathcal{N}(0, 1/2)$ 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 $n = 25$, $n = 50$ and $n = 150$, the number of simulations is 500, and we compute the robust mean relative error ($RMRE$), the linear correlation (Cor) and the `CPU` time in seconds.

	Nadaraya	estimator 1	estimator 2	estimator 3	estimator 4
$n = 25$			$\texttt{NSR} = 5\%$		
$RMRE$	0.1109	0.1148	0.1089	0.1085	0.1094
Cor	0.993	0.993	0.993	0.993	0.993
`CPU`	13	7	7	7	7
$n = 50$					
$RMRE$	0.0764	0.0791	0.0756	0.0759	0.0766
Cor	0.996	0.996	0.996	0.996	0.996
`CPU`	41	23	24	23	22
$n = 150$					
$RMRE$	0.0395	0.0422	0.0394	0.0395	0.0399
Cor	0.999	0.999	0.999	0.999	0.999
`CPU`	395	216	212	213	215
$n = 25$			$\texttt{NSR} = 10\%$		
$RMRE$	0.1170	0.1209	0.1151	0.1150	0.1163
Cor	0.993	0.993	0.993	0.993	0.993
`CPU`	11	6	6	6	6
$n = 50$					
$RMRE$	0.0801	0.0835	0.0792	0.0794	0.0803
Cor	0.996	0.996	0.996	0.996	0.996
`CPU`	41	24	23	24	23
$n = 150$					
$RMRE$	0.0413	0.0430	0.0411	0.0414	0.0417
Cor	0.999	0.999	0.999	0.999	0.999
`CPU`	394	222	224	221	218
$n = 25$			$\texttt{NSR} = 20\%$		
$RMRE$	0.1225	0.1269	0.1207	0.1203	0.1215
Cor	0.992	0.992	0.992	0.992	0.993
`CPU`	9	5	5	5	5
$n = 50$					
$RMRE$	0.0838	0.0873	0.0835	0.0836	0.0842
Cor	0.996	0.996	0.996	0.996	0.996
`CPU`	39	21	21	22	23
$n = 150$					
$RMRE$	0.0421	0.0452	0.0422	0.0426	0.0431
Cor	0.998	0.998	0.998	0.998	0.998
`CPU`	388	209	207	205	209

Table 2. Quantitative comparison between the deconvolution Nadaraya's estimator (1.4) and four estimators; estimator 1 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([2/3]n^{-1})$, estimator 2 correspond to the estimator (1.2) with the choice of $(\gamma_n) = (n^{-1})$, estimator 3 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([4/3]n^{-1})$ and estimator 4 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([5/3]n^{-1})$. Here we consider the standard normal distribution $X \sim \mathcal{N}(0,1)$ 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 $n = 25$, $n = 50$ and $n = 150$, the number of simulations is 500, and we compute the robust mean relative error ($RMRE$), the linear correlation (Cor) and the CPU time in seconds.

	Nadaraya	estimator 1	estimator 2	estimator 3	estimator 4
$n = 25$			$\texttt{NSR} = 5\%$		
$RMRE$	0.0975	0.1024	0.0934	0.0942	0.0968
Cor	0.993	0.993	0.993	0.993	0.993
CPU	7	4	4	4	4
$n = 50$					
$RMRE$	0.0779	0.0811	0.0745	0.0745	0.0755
Cor	0.997	0.997	0.997	0.997	0.997
CPU	38	20	21	20	21
$n = 150$					
$RMRE$	0.0357	0.0379	0.0349	0.0345	0.0346
Cor	0.999	0.999	0.999	0.999	0.999
CPU	374	194	195	193	196
$n = 25$			$\texttt{NSR} = 10\%$		
$RMRE$	0.1150	0.1180	0.1133	0.1130	0.1139
Cor	0.993	0.993	0.993	0.993	0.993
CPU	8	4	4	4	4
$n = 50$					
$RMRE$	0.0797	0.0805	0.0772	0.0773	0.0778
Cor	0.997	0.997	0.997	0.997	0.997
CPU	35	18	17	16	20
$n = 150$					
$RMRE$	0.0374	0.0394	0.0363	0.0366	0.0370
Cor	0.999	0.999	0.999	0.999	0.999
CPU	369	189	187	186	191
$n = 25$			$\texttt{NSR} = 20\%$		
$RMRE$	0.1137	0.1176	0.1127	0.1139	0.1158
Cor	0.993	0.993	0.993	0.993	0.993
CPU	8	4	4	4	4
$n = 50$					
$RMRE$	0.0834	0.0864	0.0832	0.0842	0.0851
Cor	0.996	0.996	0.996	0.996	0.996
CPU	37	19	18	20	21
$n = 150$					
$RMRE$	0.0397	0.0419	0.0393	0.0395	0.0397
Cor	0.999	0.999	0.999	0.999	0.999
CPU	379	203	202	203	205

Table 3. Quantitative comparison between the deconvolution Nadaraya's estimator (1.4) and four estimators; estimator 1 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([2/3]n^{-1})$, estimator 2 correspond to the estimator (1.2) with the choice of $(\gamma_n) = (n^{-1})$, estimator 3 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([4/3]n^{-1})$ and estimator 4 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([5/3]n^{-1})$. Here we consider the normal distribution $X \sim \mathcal{N}(0, 2)$ 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 $n = 25$, $n = 50$ and $n = 150$, the number of simulations is 500, and we compute the robust mean relative error ($RMRE$) and the linear correlation (Cor), and the `CPU` time in seconds.

	Nadaraya	estimator 1	estimator 2	estimator 3	estimator 4
$n = 25$			$\texttt{NSR} = 5\%$		
$RMRE$	0.0948	0.0982	0.0903	0.0915	0.0940
Cor	0.995	0.995	0.995	0.995	0.995
`CPU`	10	5	5	5	5
$n = 50$					
$RMRE$	0.0768	0.0789	0.0751	0.0753	0.0760
Cor	0.997	0.997	0.997	0.997	0.997
`CPU`	43	25	24	23	24
$n = 150$					
$RMRE$	0.0357	0.0373	0.0344	0.0340	0.0341
Cor	0.998	0.998	0.998	0.998	0.998
`CPU`	403	215	213	212	216
$n = 25$			$\texttt{NSR} = 10\%$		
$RMRE$	0.0946	0.1030	0.0927	0.0916	0.0931
Cor	0.994	0.994	0.994	0.994	0.994
`CPU`	11	6	6	6	6
$n = 50$					
$RMRE$	0.0803	0.0808	0.0749	0.0748	0.0755
Cor	0.997	0.997	0.997	0.997	0.997
`CPU`	42	23	23	24	25
$n = 150$					
$RMRE$	0.0497	0.0490	0.0467	0.0469	0.0469
Cor	0.998	0.998	0.998	0.998	0.998
`CPU`	401	206	205	208	205
$n = 25$			$\texttt{NSR} = 20\%$		
$RMRE$	0.0993	0.1049	0.0940	0.0921	0.0932
Cor	0.993	0.993	0.993	0.993	0.993
`CPU`	10	5	5	5	5
$n = 50$					
$RMRE$	0.0812	0.0837	0.0805	0.0805	0.0809
Cor	0.997	0.996	0.996	0.996	0.996
`CPU`	43	25	24	23	23
$n = 150$					
$RMRE$	0.0762	0.0709	0.0685	0.0675	0.0666
Cor	0.998	0.998	0.998	0.998	0.998
`CPU`	394	202	204	203	201

Table 4. Quantitative comparison between the deconvolution Nadaraya's estimator (1.4) and four estimators; estimator 1 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([2/3]n^{-1})$, estimator 2 correspond to the estimator (1.2) with the choice of $(\gamma_n) = (n^{-1})$, estimator 3 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([4/3]n^{-1})$ and estimator 4 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([5/3]n^{-1})$. Here we consider the normal mixture distribution $X \sim 1/2\mathcal{N}(1/2, 1) + 1/2\mathcal{N}(-1/2, 1)$ with $\mathtt{NSR} = 5\%$ in the first block, $\mathtt{NSR} = 10\%$ in the second block and $\mathtt{NSR} = 20\%$ in the last block, we consider three sample sizes $n = 25$, $n = 50$ and $n = 150$, the number of simulations is 500, and we compute the robust mean relative error ($RMRE$), the linear correlation (Cor) and the `CPU` time in seconds.

	Nadaraya	estimator 1	estimator 2	estimator 3	estimator 4
$n = 25$			$\mathtt{NSR} = 5\%$		
$RMRE$	0.0831	0.0888	0.0790	0.0795	0.0817
Cor	0.994	0.994	0.994	0.994	0.994
CPU	13	7	7	7	7
$n = 50$					
$RMRE$	0.0494	0.0522	0.0486	0.0489	0.0497
Cor	0.996	0.996	0.996	0.996	0.996
CPU	45	25	24	23	24
$n = 150$					
$RMRE$	0.0163	0.0180	0.0149	0.0143	0.0141
Cor	0.999	0.999	0.999	0.999	0.999
CPU	423	228	224	225	226
$n = 25$			$\mathtt{NSR} = 10\%$		
$RMRE$	0.0841	0.0895	0.0807	0.0818	0.0844
Cor	0.992	0.992	0.992	0.992	0.992
CPU	12	7	7	7	7
$n = 50$					
$RMRE$	0.0547	0.0579	0.0540	0.0539	0.0544
Cor	0.994	0.994	0.994	0.994	0.994
CPU	44	24	23	23	24
$n = 150$					
$RMRE$	0.0213	0.0246	0.0211	0.0218	0.0225
Cor	0.997	0.997	0.997	0.997	0.997
CPU	425	226	225	225	229
$n = 25$			$\mathtt{NSR} = 20\%$		
$RMRE$	0.0846	0.0907	0.0808	0.0806	0.0828
Cor	0.991	0.991	0.991	0.991	0.991
CPU	13	7	7	7	7
$n = 50$					
$RMRE$	0.0582	0.0616	0.0580	0.0581	0.0587
Cor	0.993	0.993	0.993	0.993	0.993
CPU	45	23	24	24	23
$n = 150$					
$RMRE$	0.0219	0.0249	0.0213	0.0222	0.0228
Cor	0.995	0.995	0.995	0.995	0.995
CPU	435	232	228	229	230

Table 5. Quantitative comparison between the deconvolution Nadaraya's estimator (1.4) and four estimators; estimator 1 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([2/3]n^{-1})$, estimator 2 correspond to the estimator (1.2) with the choice of $(\gamma_n) = (n^{-1})$, estimator 3 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([4/3]n^{-1})$ and estimator 4 correspond to the estimator (1.2) with the choice of $(\gamma_n) = ([5/3]n^{-1})$. Here we consider the exponetial distribution $X \sim \mathcal{E}(1/2)$ with $\mathtt{NSR} = 5\%$ in the first block, $\mathtt{NSR} = 10\%$ in the second block and $\mathtt{NSR} = 20\%$ in the last block, we consider three sample sizes $n = 25$, $n = 50$ and $n = 150$, the number of simulations is 500, and we compute the robust mean relative error ($RMRE$), the linear correlation (Cor) and the CPU time in seconds.

	Nadaraya	estimator 1	estimator 2	estimator 3	estimator 4
$n = 25$			$\mathtt{NSR} = 5\%$		
$RMRE$	0.1298	0.1336	0.1239	0.1242	0.1265
Cor	0.955	0.954	0.953	0.952	0.952
CPU	9	5	5	5	5
$n = 50$					
$RMRE$	0.1263	0.1274	0.1217	0.1217	0.1224
Cor	0.965	0.964	0.964	0.963	0.962
CPU	38	20	21	20	21
$n = 150$					
$RMRE$	0.0808	0.0790	0.0759	0.0751	0.0748
Cor	0.984	0.984	0.985	0.983	0.982
CPU	384	198	199	197	198
$n = 25$			$\mathtt{NSR} = 10\%$		
$RMRE$	0.1350	0.1403	0.1300	0.1297	0.1317
Cor	0.939	0.938	0.938	0.939	0.939
CPU	9	5	5	5	5
$n = 50$					
$RMRE$	0.1284	0.1311	0.1250	0.1249	0.1257
Cor	0.950	0.949	0.949	0.948	0.948
CPU	39	20	19	20	21
$n = 150$					
$RMRE$	0.1190	0.1092	0.1073	0.1064	0.1054
Cor	0.942	0.944	0.954	0.954	0.953
CPU	392	204	203	202	204
$n = 25$			$\mathtt{NSR} = 20\%$		
$RMRE$	0.1669	0.1525	0.1509	0.1494	0.1479
Cor	0.934	0.934	0.944	0.943	0.943
CPU	9	5	5	5	5
$n = 50$					
$RMRE$	0.1363	0.1382	0.1289	0.1289	0.1305
Cor	0.944	0.944	0.948	0.949	0.949
CPU	37	19	21	21	20
$n = 150$					
$RMRE$	0.1258	0.1213	0.1160	0.1151	0.11508
Cor	0.933	0.938	0.937	0.937	0.938
CPU	378	195	197	194	194

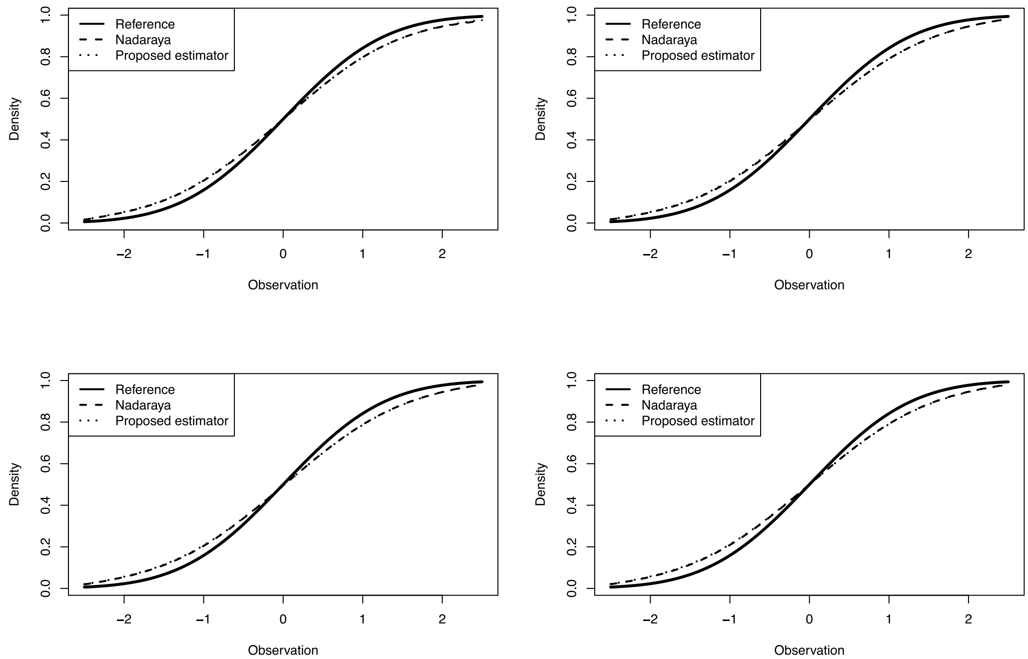

Figure 1. Qualitative comparison between the deconvolution Nadaraya's estimator (1.4) and the proposed estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of size 200, with NSR equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the normal distribution $X \sim \mathcal{N}(0, 1/2)$.

distribution estimator (1.4).

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

(vi) the $RMRE$ decrease as the value of NSR increase.

From Figs. 1, 2, 3, 4 and 5, we conclude that, our proposed kernel distribution estimator (1.2), with the choice of the stepsize $(\gamma_n) = (n^{-1})$ can be closer to the true distribution function as compared to the deconvolution Nadaraya's kernel distribution estimator (1.4), especially for small NSR. For our last choice of distribution function (see Fig. 5), even when the value of NSR is equal to 30% our proposed estimator is closer to the true distribution function.

3.2. Real dataset

Salmon Dataset: This data is from Simonoff (1996). It concerns the size of the annual spawning stock and its production of new catchable-sized fish for 1940 through 1967 for the Skeena river sockeye salmon stock (in thousands of fish). The dataset was available in the R package idr and contained 28 observations on the following three variables; year, spawness and recruits, for more details see Simonoff (1996). In order to investigate the comparison between the two estimators, we consider the annual recruits: 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 $AMISE^*$. The plug-in estimators (2.14),

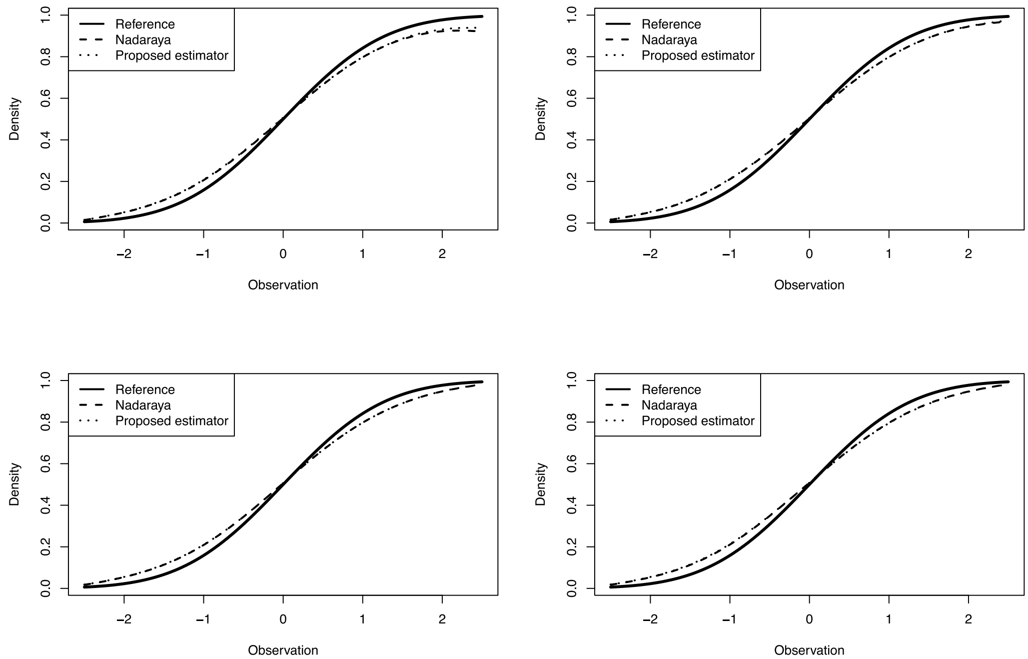

Figure 2. Qualitative comparison between the deconvolution Nadaraya's estimator (1.4) and the proposed estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of size 200, with NSR equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the standard normal distribution $X \sim \mathcal{N}(0,1)$.

(2.19) 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 (2.12), with β equal respectively to 2/9 and 1/6.

From the Table 6, we conclude that, the $\widehat{AMISE}^*$ of proposed estimator is quite better than the $\widetilde{AMISE}^*$ of the deconvolution Nadaraya's kernel distribution estimator. From the Fig. 6, 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 distribution function in the case of deconvolution kernel estimators with Laplace measurement errors. The estimators are compared to the deconvolution distribution estimator (1.4). We showed that using the selected bandwidth and the stepsizes $(\gamma_n) = (n^{-1})$, the proposed estimator will be better than the estimator (1.4) 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 estimator was more computing efficiency than the estimator (1.4).

In conclusion, the proposed estimators allowed us to obtain quite better

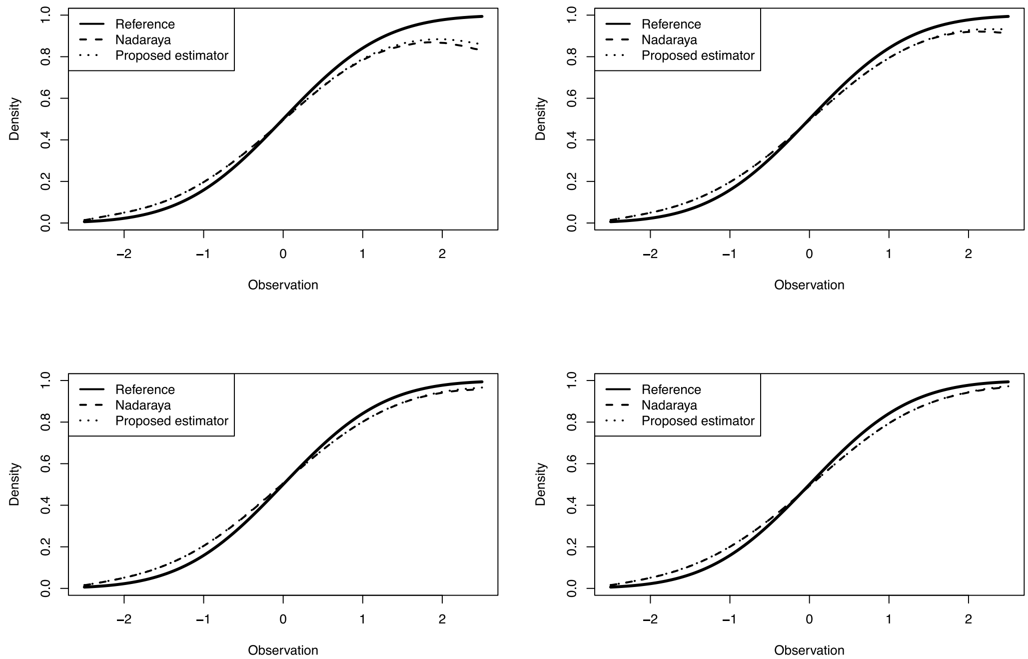

Figure 3. Qualitative comparison between the deconvolution Nadaraya's estimator (1.4) and the proposed estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of size 200, with NSR equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the normal distribution $X \sim \mathcal{N}(0, 2)$.

results then the deconvolution Nadaraya's estimator. 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, and to consider the case of supersmooth measurements error distribution (e.g. normal distribution).

5. Technical proofs

Throughout this section we use the following notations:

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

$$Z_n^{\varepsilon}(x) = \mathcal{K}^{\varepsilon}\left(\frac{x - Y_n}{h_n}\right) \tag{5.1}$$

$$\mu_2(K) = \int_{\mathbb{R}} z^2 K(z) dz$$

$$\psi(K) = \int_{\mathbb{R}} z K(z) \mathcal{K}(z) dz \tag{5.2}$$

Let us first state the following technical lemma.

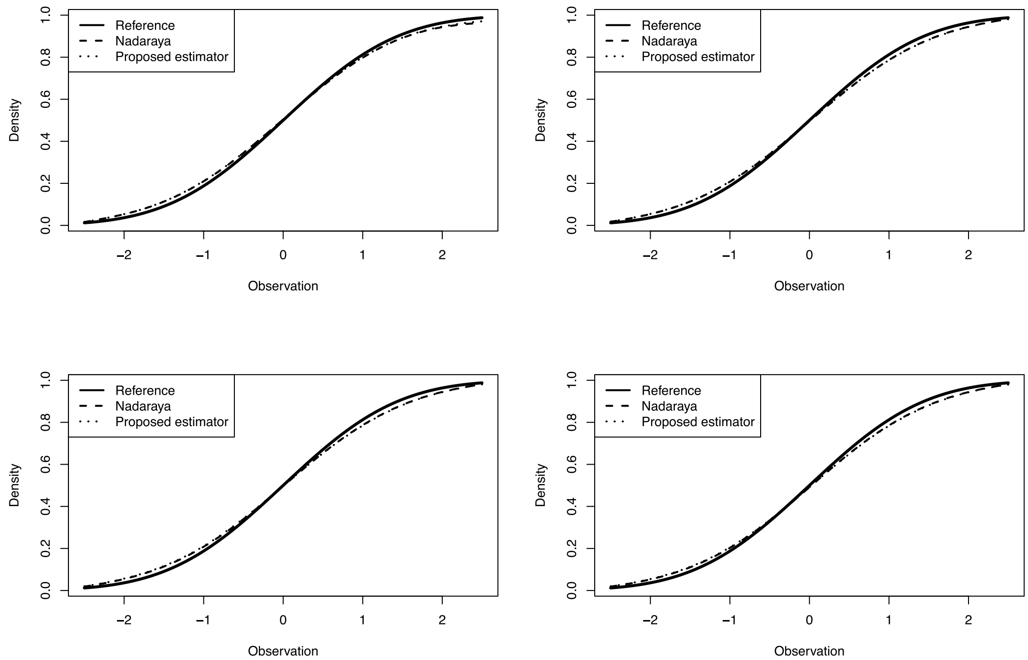

Figure 4. Qualitative comparison between the deconvolution Nadaraya' estimator (1.4) and the proposed estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of size 200, with `NSR` equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the normal mixture distribution $X \sim 1/2\mathcal{N}(1/2, 1) + 1/2\mathcal{N}(-1/2, 1)$.

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* (2.2). *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 Subsection 5.1, Theorem 1 in Subsection 5.2.

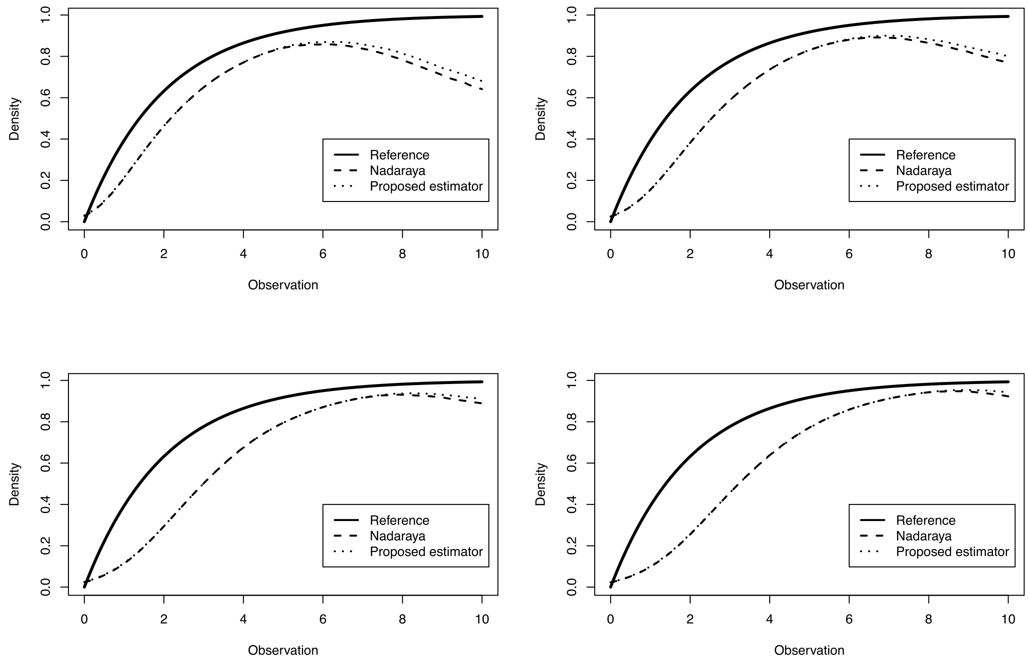

Figure 5. Qualitative comparison between the deconvolution Nadaraya's estimator (1.4) and the proposed estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of size 200, with `NSR` equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the exponetial distribution $X \sim \mathcal{E}(1/2)$.

5.1. Proof of Proposition 1

Proof. In view of (1.3) and (5.1), we have

$$\begin{aligned}
F_{n,X}(x) - F_X(x) \quad & \\
&= (1-\gamma_n)(F_{n-1,X}(x) - F_X(x)) + \gamma_n(Z_n^{\varepsilon}(x) - F_X(x)) \\
&= \sum_{k=1}^{n-1}\left[\prod_{j=k+1}^{n}(1-\gamma_j)\right]\gamma_k(Z_k^{\varepsilon}(x) - F_X(x)) + \gamma_n(Z_n^{\varepsilon}(x) - F_X(x)) \\
&\quad + \left[\prod_{j=1}^{n}(1-\gamma_j)\right](F_{0,X}(x) - F_X(x)) \\
&= \Pi_n \sum_{k=1}^{n}\Pi_k^{-1}\gamma_k(Z_k^{\varepsilon}(x) - F_X(x)) + \Pi_n(F_{0,X}(x) - F_X(x)).
\end{aligned} \tag{5.3}$$

It follows that

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

Table 6. The comparison between the $AMISE^*$ of the deconvolution Nadaraya's distribution estimator (1.4) and the $AMISE^*$ of the proposed distribution estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$ 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.

	I_1	I_2	h_n	$AMISE^*$
		`NSR` $= 5\%$		
Nadaraya	$1.14e^{-01}$	$5.47e^{-04}$	0.661	$6.09e^{-04}$
Proposed estimator	$1.15e^{-01}$	$1.24e^{-02}$	0.368	$2.49e^{-04}$
		`NSR` $= 10\%$		
Nadaraya	$1.11e^{-01}$	$4.84e^{-04}$	0.825	$5.66e^{-04}$
Proposed estimator	$1.12e^{-01}$	$4.05e^{-04}$	0.819	$1.52e^{-04}$
		`NSR` $= 20\%$		
Nadaraya	$1.07e^{-01}$	$4.31e^{-04}$	1.025	$5.17e^{-04}$
Proposed estimator	$1.08e^{-01}$	$3.67e^{-04}$	1.020	$3.13e^{-04}$
		`NSR` $= 30\%$		
Nadaraya	$1.03e^{-01}$	$4.16e^{-04}$	1.167	$4.95e^{-04}$
Proposed estimator	$1.05e^{-01}$	$3.83e^{-04}$	1.150	$4.86e^{-04}$

Moreover, an interchange of expectation and integration, justified by Fubini's Theorem and assumptions (A1) and (A2), shows that

$$\mathbb{E}\{Z_k^{\varepsilon}(x) \mid X_k\} = Z_k(x),$$

which ensure that

$$\mathbb{E}[Z_k^{\varepsilon}(x)] = \mathbb{E}[Z_k(x)].$$

Moreover, by integration by parts, we have

$$\begin{aligned} \mathbb{E}[Z_k(x)] &= \int_{\mathbb{R}} \mathcal{K}\left(\frac{x-y}{h_k}\right) f_X(y)dy \\ &= \int_{\mathbb{R}} K(z) F_X(x+zh_k)dz. \end{aligned} \tag{5.4}$$

It follows that

$$\begin{aligned} \mathbb{E}[Z_k(x)] - F(x) &= \int_{\mathbb{R}} K(z)[F_X(x+zh_k) - F_X(x)]dz \\ &= \frac{h_k^2}{2} f_X'(x)\mu_2(K) + \beta_k(x), \end{aligned} \tag{5.5}$$

with

$$\beta_k(x) = \int_{\mathbb{R}} K(z)\left[F_X(x+zh_k) - F_X(x) - zh_k f_X(x) - \frac{1}{2}z^2h_k^2 f_X'(x)\right]dz,$$

and, since F_X is bounded and continuous at x, we have $\lim_{k\to\infty}\beta_k(x) = 0$. In the case $a \leq \alpha/7$, we have $\lim_{n\to\infty}(n\gamma_n) > 2a$; the application of Lemma 2 then

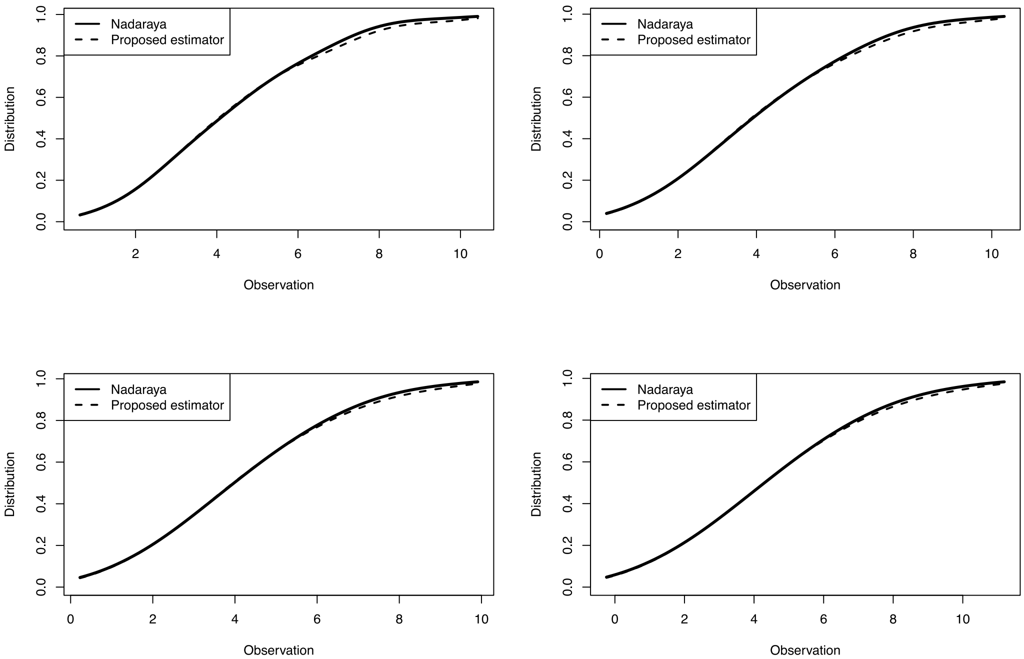

Figure 6. Qualitative comparison between the deconvolution Nadaraya's kernel estimator (1.4) and the proposed estimator (1.2) with the choice of the stepsize $(\gamma_n) = (n^{-1})$, for 500 samples of Laplacian errors with `NSR` equal respectively to 5% (in the top left panel), equal to 10% (in the top right panel), equal to 20% (in the down left panel) and equal to 30% (in the down right panel) for the `salmon` data of the package `idr` and through a plug-in method.

gives

$$\begin{aligned}
\mathbb{E}[F_{n,X}(x)] - F_X(x) &= \frac{1}{2} f'_X(x) \int_{\mathbb{R}} z^2 K(z) dz \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^2 [1 + o(1)] \\
&\quad + \Pi_n (F_{0,X}(x) - F_X(x)) \\
&= \frac{1}{2(1-2a\xi)} f'_X(x) \mu_2(K) [h_n^2 + o(1)],
\end{aligned}$$

and (2.3) follows. In the case $a > \alpha/7$, we have $h_n^2 = o(\sqrt{\gamma_n h_n^{-3}})$, and $\lim_{n\to\infty}(n\gamma_n) > (\alpha - 3a)/2$, then Lemma 2 ensures that

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

which gives (2.4). Now, we have

$$\begin{aligned}
\operatorname{Var}[F_{n,X}(x)] &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \operatorname{Var}[Z_k^{\varepsilon}(x)] \\
&= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 (\mathbb{E}((Z_k^{\varepsilon}(x))^2) - (\mathbb{E}(Z_k(x)))^2).
\end{aligned} \tag{5.6}$$

Moreover, by integration by parts, we have

$$\begin{aligned}
\mathbb{E}((Z_k^{\varepsilon}(x))^2) &= \int_{\mathbb{R}} \left(\mathcal{K}^{\varepsilon}\left(\frac{x-y}{h_k}\right)\right)^2 f_Y(y)dy \\
&= 2\int_{\mathbb{R}} K^{\varepsilon}(z)\mathcal{K}^{\varepsilon}(-z)F_Y(x+zh_k)dz \\
&= F_Y(x) - h_k f_Y(x)\psi(K^{\varepsilon}) + \nu_k(x),
\end{aligned} \tag{5.7}$$

with

$$\nu_k(x) = 2\int_{\mathbb{R}} K^{\varepsilon}(z)\mathcal{K}^{\varepsilon}(-z)[F_Y(x+zh_k) - F_Y(x) - zh_k f_Y(x)]dz.$$

Let us now state the following lemma:

Lemma 3. *Let Assumptions* (A1)–(A2) *hold, then we have*

$$\psi(K^{\varepsilon}) = -\frac{1}{4\sqrt{\pi}}\left(\left(\frac{\sigma}{h_k}\right)^4 + o(1)\right).$$

Proof. First, under the assumptions (A1) and (A2), we have $\phi_{\varepsilon}(t) = (1+\sigma^2t^2)^{-1}$ and $\phi_K(t) = \exp(-t^2/2)$, then, it follows from (1.1), that

$$\begin{aligned}
K^{\varepsilon}(u) &= \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-itu)\exp(-t^2/2)\left(1+t^2\frac{\sigma^2}{h_n^2}\right)dt \\
&= \frac{1}{2\pi}\left\{\int_{\mathbb{R}} \exp(-(itu+t^2/2))dt + \frac{\sigma^2}{h_n^2}\int_{\mathbb{R}} t^2\exp(-(itu+t^2/2))dt\right\}.
\end{aligned}$$

Moreover, it is easy to check that $\int_{\mathbb{R}} \exp(-(itu+t^2/2))dt = \sqrt{2\pi}$ and $\int_{\mathbb{R}} t^2\exp(-(itu+t^2/2))dt = \sqrt{2\pi}\exp(-u^2/2)(1-u^2)$, then, it follows that

$$K^{\varepsilon}(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)\left(1+\frac{\sigma^2}{h_n^2}(1-u^2)\right). \tag{5.8}$$

Now, we let $\phi(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}\exp(-t^2/2)dt$, then, we can check that

$$\mathcal{K}^{\varepsilon}(u) = \phi(u) + \frac{1}{\sqrt{2\pi}}\frac{\sigma^2}{h_n^2}u\exp(-u^2/2). \tag{5.9}$$

The combinations of equations (5.2), (5.8) and (5.9) leads to

$$\begin{aligned}\psi(K^{\varepsilon}) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} u\exp(-u^2/2)\phi(u)du + \frac{1}{\sqrt{2\pi}}\frac{\sigma^2}{h_n^2}\int_{\mathbb{R}}(u-u^3)\exp(-u^2/2)\phi(u)du \\ &\quad + \frac{1}{2\pi}\frac{\sigma^2}{h_n^2}\int_{\mathbb{R}} u^2\exp(-u^2)du + \frac{1}{2\pi}\frac{\sigma^4}{h_n^4}\int_{\mathbb{R}}(u^2-u^4)\exp(-u^2)du \\ &= \frac{1}{2\pi}\frac{\sigma^4}{h_n^4}\int_{\mathbb{R}}(u^2-u^4)\exp(-u^2)du + o\left(\frac{\sigma^4}{h_n^4}\right).\end{aligned}$$

Moreover, since $\int_{\mathbb{R}} u^2\exp(-u^2)du = \sqrt{\pi}/2$ and $\int_{\mathbb{R}} u^4\exp(-u^2)du = \frac{3}{4}\sqrt{\pi}$, we conclude the proof of Lemma 3. □

Moreover, it follows from (5.4), that

$$\begin{aligned}\mathbb{E}[Z_k(x)] &= F_X(x) + \int_{\mathbb{R}} K(z)[F_X(x+zh_k) - F_X(x)]dz \\ &= F_X(x) + \widetilde{\nu}_k(x),\end{aligned} \tag{5.10}$$

with

$$\widetilde{\nu}_k(x) = \int_{\mathbb{R}} K(z)[F_X(x+zh_k) - F_X(x)]dz.$$

Then, it follows from (5.6), (5.7) and (5.10), that

$$\begin{aligned}\mathrm{Var}[F_{n,X}(x)] &= (F_Y(x) - F_X^2(x))\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 \\ &\quad - f_Y(x)\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2 h_k\psi(K^{\varepsilon}) \\ &\quad + (\nu_k(x) - 2F(x)\widetilde{\nu}_k(x) - \widetilde{\nu}_k^2(x))\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2.\end{aligned} \tag{5.11}$$

Since F_X and F_Y is bounded continuous, we have $\lim_{k\to\infty}\nu_k(x) = 0$ and $\lim_{k\to\infty}\widetilde{\nu}_k(x) = 0$. In the case $a \geq \alpha/7$, we have $\lim_{n\to\infty}(n\gamma_n) > (\alpha - 3a)/2$, and the application of Lemma 2 gives

$$\mathrm{Var}[F_{n,X}(x)] = \frac{\gamma_n}{2-\alpha\xi}(F_Y(x) - F_X^2(x)) + \frac{\sigma^4}{\sqrt{\pi}}\frac{\gamma_n h_n^{-3}}{2-(\alpha-3a)\xi}f_Y(x) + o(\gamma_n h_n^{-3}),$$

which proves (2.5). Now, in the case $a < \alpha/7$, we have $\gamma_n h_n^{-3} = o(h_n^4)$, and $\lim_{n\to\infty}(n\gamma_n) > 2a$, then the application of Lemma 2 gives

$$\begin{aligned}\mathrm{Var}[F_{n,X}(x)] &= \Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k o(h_k^4) \\ &= o(h_n^4),\end{aligned}$$

which proves (2.6). □

5.2. Proof of Theorem 1

PROOF. Let us at first assume that, if $a \geq \alpha/7$ then

$$\sqrt{\gamma_n^{-1} h_n^3}\left(F_{n,X}(x) - \mathbb{E}[F_{n,X}(x)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^4}{4\sqrt{\pi}(2-(\alpha-3a)\xi)} f_Y(x)\right). \tag{5.12}$$

In the case when $a > \alpha/7$, Part 1 of Theorem 1 follows from the combination of (2.4) and (5.12). In the case when $a = \alpha/7$, Parts 1 and 2 of Theorem 1 follow from the combination of (2.3) and (5.12). In the case $a < \alpha/7$, (2.6) implies that

$$h_n^{-2}\left(F_{n,X}(x) - \mathbb{E}(F_{n,X}(x))\right) \xrightarrow{\mathbb{P}} 0,$$

and the application of (2.3) gives Part 2 of Theorem 1.

We now prove (5.12). In view of (1.3), we have

$$\begin{aligned} F_{n,X}(x) - \mathbb{E}[F_{n,X}(x)] &= (1-\gamma_n)\left(F_{n-1,X}(x) - \mathbb{E}[F_{n-1,X}(x)]\right) \\ &\quad + \gamma_n\left(Z_n^{\varepsilon}(x) - \mathbb{E}[Z_n(x)]\right) \\ &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\left(Z_k^{\varepsilon}(x) - \mathbb{E}[Z_k(x)]\right). \end{aligned}$$

Set

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

The application of Lemma 2 ensures that

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

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

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

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

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

and we thus obtain

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

The convergence in (5.12) then follows from the application of Lyapounov's Theorem. □

Acknowledgements

We are grateful to the associate editor and reviewer for their helpful comments, which have led to this substantially improved version of the paper.

References

Altman, N. and Leger, C. (1995). Bandwidth selection for kernel distribution function estimation, *J. Statist. Plann. Inference.*, **46**, 195–214.

Bojanic, R. and Seneta, E. (1973). A unified theory of regularly varying sequences, *Math. Z.*, **134**, 91–106.

Carroll, R. J., Ruppert, D. and Stefanski, L. (1995). *Measurement Error in Nonlinear Models*, Chapman & Hall, London.

Delaigle, A. and Gijbels, I. (2004). Practical bandwidth selection in deconvolution Kernel density estimation, *Comput. Statist. Data Anal.*, **45**, 249–267.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems, *Ann. Statist.*, **19**, 1257–1272.

Galambos, J. and Seneta, E. (1973). Regularly varying sequences, *Amer. Math. Soc.*, **41**, 110–116.

Hall, P. and Maron, J. S. (1987). Estimation of integrated squared density derivatives, *Statist. Probab. Lett.*, **6**, 109–115.

Mokkadem, A. and Pelletier, M. (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *Ann. Statist.*, **35**, 1749–1772.

Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009a). The stochastic approximation method for the estimation of a multivariate probability density, *J. Statist. Plann. Inference.*, **139**, 2459–2478.

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.

Nadaraya, E. A. (1964). Some new estimates for distribution functions, *Theory Probab. Appl.*, **9**, 497–500.

Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I, *Studia Sci. Math. Hung.*, **8**, 391–398.

Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function, *Math. Operationsforsch. Statist.*, *Ser. Statistics.*, **8**, 119–126.

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

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Spinger-Verlag, New York.

Slaoui, Y. (2013). Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method, *Serdica Math. J.*, **39**, 53–82.

Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method, *J. Probab. Stat.*, **2014**, ID 739640, doi:10.1155/2014/739640.

Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function, *Math. Methods Statist.*, **23**, 306–325.

Slaoui, Y. (2014c). Large and moderate deviation principles for kernel distribution estimator, *Int. Math. Forum.*, **18**, 871–890.

Slaoui, Y. (2015a). Plug-In Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method, *Stat. Neerl.*, **69**, 483–509.

Slaoui, Y. (2015b). Large and moderate deviation principles for averaged stochastic approximation method for the estimation of a regression function, *Serdica Math. J.*, **41**, 307–328.

Slaoui, Y. (2015c). Moderate deviation principles for recursive regression estimators defined by stochastic approximation method, *Int. J. Math. Stat.*, **16**, 51–60.

Slaoui, Y. (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators, *Stat. Interface*, **9**, 375–388.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators, *Statistics*, **2**, 169–184.

Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution, *Probl. Inf. Transm.*, **8**, 119–126.