SORT 28 (1) January-June 2004, 9-26

Asymptotic normality of the integrated square error of a density estimator in the convolution model

C. Butucea*

Université de Paris

Abstract

In this paper we consider a kernel estimator of a density in a convolution model and give a central limit theorem for its integrated square error (ISE). The kernel estimator is rather classical in minimax theory when the underlying density is recovered from noisy observations. The kernel is fixed and depends heavily on the distribution of the noise, supposed entirely known. The bandwidth is not fixed, the results hold for any sequence of bandwidths decreasing to 0. In particular the central limit theorem holds for the bandwidth minimizing the mean integrated square error (MISE). Rates of convergence are sensibly different in the case of regular noise and of super-regular noise. The smoothness of the underlying unknown density is relevant for the evaluation of the MISE.

MSC: 62G05, 62G20

Keywords: convolution density estimation, nonparametric density estimation, central limit theorem, integrated squared error, noisy observations

1 Introduction

In this paper we consider the following convolution model:

$$Z_i = X_i + e_i,$$

where $X_i,\ i = 1, \dots, n$ are i.i.d. random variables of unknown density f which we need to recover from noisy observations $Y_i,\ i = 1, \dots, n$. The noise variables e_i are supposed i.i.d. of known fixed distribution, having a density function η in L_1 and L_2 and a characteristic function (c. f.) Φ^η.

* *Address for correspondence:* Modal'X, Université Paris X, Bât. G, F92001 Nanterre Cedex. LPMA, Université Paris VI, 175, rue du Chevaleret, F75013 Paris. butucea@ccr.jussieu.fr.

Received: October 2003
Accepted: March 2004

We suggest here an estimator f_n of f from noisy observations and study the asymptotic normality of its integrated square error (ISE)

$$ISE(f_n, f) = \int (f_n(x) - f(x))^2 dx. \tag{1}$$

Let us suppose for the beginning that f belongs to a Sobolev class $W(r, L)$ of densities, i.e.

$$W(r, L) = \left\{ f \text{ density} : f \in L_2, \quad \int |\Phi(u)|^2 |u|^{2r} du \leq 2\pi L \right\}$$

where $\Phi(u) = \int \exp(iux) f(x) dx$ denotes its Fourier transform, for some fixed $r > 1/2$ and a constant $L > 0$. This roughly means these densities are continuously derivable up to order r and their r-th derivative has bounded L_2 norm.

It is known from estimation theory in the convolution model, that the rates and behaviours of estimators are sensibly different if the characteristic function of the noise decreases polynomially or exponentially asymptotically. We suppose in a first part that the noise is “polynomial”, i.e.

$$|\Phi^\eta(u)| \sim |u|^{-s}, \text{ as } |u| \to \infty,$$

where $\sim$ means that the functions behave similarly and $s > 0$ such that $r > s$.

Let us denote $g = f \star \eta$ the common density of $Y_i, i = 1, \ldots, n$ and $\Phi^g = \Phi \cdot \Phi^\eta$ its Fourier transform.

In Section 3, we state our results for different setups. In Section 3.1 we consider classes of supersmooth densities in association with polynomial noise. We say that f is a supersmooth density if f belongs to the class

$$S(\alpha, r, L) = \left\{ f \text{ density} : f \in L_2, \quad \int |\Phi(u)|^2 \exp(2\alpha |u|^r) du \leq 2\pi L \right\},$$

for some $\alpha, r, L > 0$. In Section 3.2 we consider Sobolev densities in association with exponentially decreasing noise. Exponential noise means

$$|\Phi^\eta(u)| \sim \exp(-\gamma |u|^s), \text{ as } |u| \to \infty,$$

where $\gamma, s > 0$. We work here with a kernel estimator of the deconvolution density

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_h^n (x - Y_i), \tag{2}$$

where $h > 0$ is small, K_h^n denotes $K^n(\cdot/h)/h$ and the kernel K^n is defined via its Fourier transform

$$\Phi^{K^n}(u) = \frac{\Phi^K(u)}{\Phi^\eta(u/h)}, \text{ where } \Phi^K(u) = I[|u| \leq 1]. \tag{3}$$

Since pioneering work by Carroll and Hall (1988), the deconvolution density was already estimated in many setups. We shall cite here only works very much related to our framework and problems. Such kernel estimates were used on classes similar to the Sobolev class by Fan (1991*a*), who computed the rates of convergence of the minimax L_2 risk. Recently wavelet estimators were proven to attain the same rates on Besov bodies and these rates are known to be optimal in the minimax approach, see Fan and Koo (2002).

In the setup of Sobolev densities, Goldenshluger (1999) generalized the minimax rate for estimating f with pointwise risk to adaptive (to the Sobolev smoothness) rates when the noise is either polynomial or exponential (without loss of rate in this last case). Efromovich (1997) computed exact asymptotic risks (pointwise and in L_2 norm) for estimating Sobolev densities in the presence of exponentially decreasing noise.

The kernel estimator in (2) (with adequate bandwidth) was proven to be minimax for estimating supersmooth densities with polynomial noise in Butucea (2004) and with exponential noise in Butucea and Tsybakov (2003). The same kernel estimator was proven asymptotically normal when the noise is either polynomial or exponential in Fan (1991*b*) and Fan and Liu (1997).

Here we study the asymptotic normality of the ISE in (1) and will discuss several important applications of results issued from these computations. Such computations can be found in Hall (1984) for a nonparametric density estimator with direct observations. His study is a direct application of a Central Limit Theorem of degenerate U-statistics of second order. He motivates this by the practical use in simulations of ISE as a measure of the performance of a density estimator. The main goal is to evaluate c_n and σ_n such that

$$\sigma_n^{-1}(ISE(f_n, f) - c_n) \to N(0, 1),$$

when $h \to 0$ and $n \to \infty$. This subject is strongly related to estimating the L_2 norm of the density f from noisy observations. Indeed, a natural estimator d_n^2 of $\|f\|_2^2$ can be decomposed such that one of the terms is the degenerate second order U-statistic S_2 defined later in (8). For not too smooth densities S_2 is the dominating term and this gives the rate of estimating $\|f\|_2^2$. Estimating the L_2 norm of a density is furthermore useful for nonparametric testing in the convolution model. These problems will be soon the subject of scientific communications.

Another related problem can be further investigated starting with these calculations, namely that of bandwidth selection for the kernel deconvolution density estimator f_n in (2), via cross-validation.

2 Results

As a first step it is natural to replace c_n by $E_f[ISE(f_n, f)]$ also denoted by $MISE(f_n, f)$ for mean integrated square error. From now on P_f, E_f, and V_f denote the probability, the expectation and the variance when the true underlying density of the model is f. We may use constants $c, C, C', \dots$ which are different throughout the whole proof.

Note that the density of our observations is $g = f \star \eta$. We note next that

$$\begin{aligned} ISE(f_n, f) &= \int (f_n(x) - E_f[f_n(x)] + E_f[f_n(x)] - f(x))^2 dx \\ &= \int (f_n(x) - E_f[f_n(x)])^2 dx + \int (E_f[f_n(x)] - f(x))^2 dx. \end{aligned}$$

Indeed, the cross product term is null, see Lemma 2. We replace from now on $E_f[f_n(x)]$ by its value $K_h \star f$. Then

$$MISE(f_n, f) = E_f[ISE(f_n, f))] = E_f \left[\int (f_n(x) - E_f[f_n(x)])^2 dx \right] + \int (E_f[f_n(x)] - f(x))^2 dx$$

and we write

$$ISE(f_n, f) - E_f[ISE(f_n, f))] = I_n - E_f[I_n],$$

where $I_n = \int (f_n(x) - E_f[f_n(x)])^2 dx$. Computation of $E_f[I_n]$ and of the bias $B^2(f_n) = \int (E_f[f_n(x)] - f(x))^2 dx$ is rather classical in minimax theory.

Lemma 1 *Let $f_n(\cdot, Y_1, \dots, Y_n)$ be the kernel density estimator defined in (2) based on the noisy observations in our convolution model with a bandwidth $h \to 0$ when $n \to \infty$. Then*

$$E_f[I_n] = \frac{1 + o(1)}{\pi(2s + 1)nh^{2s+1}}.$$

If the underlying density belongs to a Sobolev smoothness class $W(r, L)$ with $r > 1/2$, then

$$\sup_{f \in W(r,L)} B^2(f_n) = \sup_{f \in W(r,L)} \int \left(E_f[f_n(x)] - f(x)\right)^2 dx = Lh^{2r} = o(1).$$

In conclusion, $MISE(f_n, f)$ converges to 0, if and only if $nh^{2s+1} \to \infty$ when $n \to \infty$ and the bandwidth minimizing $\sup_{f \in W(r,L)} MISE(f_n, f)$ is

$$h_{MISE} = (L\pi(2s + 1)n)^{-\frac{1}{2(r+s)+1}}.$$

Proof. We present here only exact calculation of $E_f[I_n]$, since the remaining results are obvious or not entirely new. We have

$$\begin{aligned} E_f[I_n] &= \frac{1}{n} \int \left(\int \left(K_h^n(x - y) - K_h \star f \right)^2 (x) dx \right) g(y) dy \\ &= \frac{1}{n} \left(\int \left(\int \left(K_h^n(x - y) \right)^2 dx \right) g(y) dy - \|K_h \star f\|_2^2 \right). \end{aligned}$$

We know that $\|K_h \star f\|_2^2$ is equal to $\|f\|_2^2$ plus some estimation bias which tends to 0 when $h \to 0$ on a smoothness class like the Sobolev class, $W(r, L)$. So, the main term is $\int \left(\int \left(K_h^n(x-y) \right)^2 dx \right) g(y) dy$. Use Lemma 2:

$$\begin{aligned}
\int \left(\int \left(K_h^n(x-y) \right)^2 dx \right) g(y) dy &= \frac{1}{h} \int (K^n)_h^2 \star g(x) dx = \frac{1}{2\pi h} \Phi^{(K^n)_h^2 \star g}(0) \\
&= \frac{1}{2\pi h} \Phi^g(0) \Phi^{(K^n)^2}(0) = \frac{1}{2\pi h} \int \Phi^{K^n}(-u) \Phi^{K^n}(u) du \\
&= \frac{1 + o(1)}{\pi (2s+1) h^{2s+1}}.
\end{aligned}$$

□

Remark that in previous equations and in the following proofs, we compute integrals like $\int (\Phi^{K^n})^2$ by actually replacing the c. f. of the noise by $|u|^{-s}$, its asymptotic expression. We do this for simplicity, since calculation would actually need splitting integration domain into $|u| \leq M$ and $M < |u| < 1/h$, for some large enough, but fixed $M > 0$. If M is large enough, Φ^η is almost $|u|^{-s}$ and the second integral is always dominating over the first and gives the order of the whole expression. For a complete and explicit computation of $\|K_n\|_2^2$ see Butucea (2004).

Let us look closer at I_n:

$$\begin{aligned}
I_n &= \frac{1}{n^2} \int \left(\sum_{i=1}^n (K_h^n(x - Y_i) - K_h \star f(x)) \right)^2 dx \\
&= \frac{1}{n^2} \sum_{i=1}^n \|K_h^n(\cdot - Y_i) - K_h \star f\|_2^2 + \frac{1}{n^2} \sum_{i \neq j = 1}^n \langle K_h^n(\cdot - Y_i) - K_h \star f, K_h^n(\cdot - Y_j) - K_h \star f \rangle,
\end{aligned}$$

where $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the L_2 norm and the scalar product in L_2, respectively. If we denote by

$$U_i = U_i(x, h, Y_i) = K_h^n(x - Y_i) - K_h \star f(x), \tag{4}$$

these variables are centred and independent. We get

$$\begin{aligned}
I_n - E_f[I_n] &= \frac{1}{n^2} \sum_{i=1}^n \left(\|U_i\|_2^2 - E_f[\|U_i\|_2^2] \right) + \frac{1}{n^2} \sum_{i \neq j = 1}^n \langle U_i, U_j \rangle \\
&= S_1 + S_2, \text{ say.}
\end{aligned}$$

It is easy to see that variables in S_1 and in S_2 are uncorrelated:

$$E_f[(\|U_k\|_2^2 - E_f[\|U_k\|_2^2])(\langle U_i, U_j \rangle)] = 0,$$

for all $k, i, j = 1, \ldots, n$ and $i \neq j$. It is necessary now to compute the variance of each sum and compare. What we prove in the following is that S_2 has a larger variance (in order) than S_1, for any $h \to 0$ and $n \to \infty$. Then we prove its asymptotic normality and deduce the asymptotic normality of $ISE(f_n, f) - E_f[ISE(f_n, f)]$. The main difficulty comes from the fact that S_2 is an U-statistic of order 2 and degenerate. Indeed,

$$\begin{aligned} E_f[\langle U_i, U_j\rangle / Y_j = y_j] &= E_f[\langle K_h^n(\cdot - Y_i) - K_h \star f, K_h^n(\cdot - y_j) - K_h \star f\rangle] \\ &= \langle E_f[K_h^n(\cdot - Y_i)] - K_h \star f, K_h^n(\cdot - y_j) - K_h \star f\rangle = 0. \end{aligned}$$

Nevertheless, each term of the sum depends on n and we apply a central limit theorem for degenerate U-statistics by Hall (1984), which he already applied in his paper for the ISE of a nonparametric estimator with direct observations. Here, we have noisy observations and a particular choice of the kernel (motivated by the minimax theory in this field) giving sensibly different asymptotic behaviours and rates.

Theorem 1 *Let $f_n(\cdot, Y_1, \dots, Y_n)$ be the kernel density estimator defined in (2) based on the noisy observations in our convolution model and a bandwidth $h \to 0$ such that $nh^{2s+1} \to \infty$, when $n \to \infty$. Then*

$$\sqrt{\frac{\pi(4s+1)n^2h^{4s+1}}{2\|g\|_2^2}}\Big(ISE(f_n, f) - E_f[ISE(f_n, f)]\Big) \to N(0,1)$$

where the convergence is in law when $n \to \infty$.

Corollary 2 *Let $f_n(\cdot, Y_1, \dots, Y_n)$ be the kernel density estimator in (2) based on the noisy observations with noise having polynomially decreasing Fourier transform and a bandwidth $h \to 0$ such that $nh^{2s+1} \to \infty$, when $n \to \infty$. Then I_n is asymptotically normally distributed with*

$$E_f[I_n] = \frac{1+o(1)}{\pi(2s+1)nh^{2s+1}} \text{ and } V_f[I_n] = \frac{2\|g\|_2^2(1+o(1))}{\pi(4s+1)n^2h^{4s+1}}; \tag{5}$$

if f belongs to the Sobolev class $W(r, L)$, the integrated square error $ISE(f_n, f)$ is asymptotically normally distributed with

$$MISE(f_n, f) \le Lh^{2r} + \frac{1+o(1)}{\pi(2s+1)nh^{2s+1}} \text{ and } V_f[ISE(f_n, f)] = \frac{2\|g\|_2^2(1+o(1))}{\pi(4s+1)n^2h^{4s+1}}$$

and the $MISE$ becomes minimal (and of the order of the minimax L_2 risk) for $h_* = (L\pi(2s+1)n)^{1/(2(r+s)+1)}$

$$\inf_{h>0} \sup_{f \in W(r,L)} MISE(f_n, f) = L^{\frac{1}{2(r+s)+1}}(\pi(2s+1)n)^{-\frac{2r}{2(r+s)+1}}.$$

Notice that for constructing a confidence interval of $ISE(f_n, f)$ using its asymptotic normality, both $MISE(f_n, f)$ and $V_f[ISE(f_n, f)]$ still depend on unknown quantities. This was already noted by Hall (1984). The mean of $ISE(f_n, f)$ depends on unknown f

via the bias of f_n: $B^2(f_n) = ||E_f[f_n] - f||_2^2$ that we can bound from above by Lh^{2r}. The variance of $ISE(f_n, f)$ depends on unknown $||g||_2^2$. Nevertheless, g is the density of our observations and can be directly evaluated at a faster rate than f (the same holds for the other frameworks). Indeed, not only we have direct observations, moreover, g is more regular than f due to the convolution (which adds smoothness). The estimation of the L_2 norm of a regular enough density, having a smoothness $> 1/4$, can be done efficiently at rate $1/\sqrt{n}$, see e.g. Laurent (1996).

Note also that if we use another bandwidth h satisfying $nh^{2r+2s+1} \to \infty$, when $n \to \infty$, the associated $MISE$ is $(1 + o(1))/(\pi(2s + 1)nh^{2s+1})$. Indeed, whatever the bias of the estimator f_n is, it is smaller than $Lh^{2r} = o(1/(nh^{2s+1}))$. In this case, the confidence interval $IC_{1-\delta}$ of risk $\delta > 0$, writes

$$IC_{1-\delta} = \left[\frac{1}{\pi(2s+1)nh^{2s+1}} \pm z_{1-\delta/2} \frac{||g||_2}{nh^{2s+1/2}} \sqrt{\frac{2}{\pi(4s+1)}} \right], \tag{6}$$

where z_δ is the δ-quantile of $N(0, 1)$, a gaussian law.

Proof. **Convergence of S_1**

$$S_1 = \frac{1}{n^2} \sum_{i=1}^{n} \left(||U_i||_2^2 - E_f[||U_i||_2^2] \right).$$

Let us compute an upper bound of the variance of S_1. We have

$$\begin{aligned} V_f[S_1] &= \frac{1}{n^4} \sum_{i=1}^{n} E_f \left[\left(||U_i||_2^2 - E_f[||U_i||_2^2] \right)^2 \right] \\ &= \frac{1}{n^3} \left(E_f[||U_1||_2^4] - \left(E_f[||U_1||_2^2] \right)^2 \right) \leq \frac{E_f[||U_1||_2^4]}{n^3}. \end{aligned}$$

In order to evaluate an upper bound of this, we develop the square of sums in $E_f[||U_1||_2^4]$ and conclude by saying that the dominant term is given by one of positive terms (this expectation being a positive real number):

$$\begin{aligned} E_f[||U_1||_2^4] &= \int \left(\int \left(K_h^n(x-y) - K_h \star f(x) \right)^2 dx \right)^2 g(y)dy \\ &= \int \left(\int \left(K_h^n(x-y) \right)^2 dx \right)^2 g(y)dy \\ &\quad +2||K_h \star f||_2^2 \int \left(\int \left(K_h^n(x-y) \right)^2 dx \right) g(y)dy \\ &\quad +4 \int \left(\int K_h^n(x-y) K_h \star f(x) dx \right)^2 g(y)dy. \end{aligned}$$

Note that, by Cauchy-Schwarz and previous evaluations:

$$\begin{aligned}
&\int\left(\int K_h^n(x-y)K_h \star f(x)dx\right)^2 g(y)dy \\
\leq\ &\int\left(\int \left(K_h^n(x-y)\right)^2 dx\right)^{1/2} \|K_h \star f\|_2\, g(y)dy \leq \frac{O(1)}{h^{2s+1}}.
\end{aligned}$$

It remains to compute an asymptotic upper bound of $\int\left(\int\left(K_h^n(x-y)\right)^2 dx\right)^2 g(y)dy$. As previously,

$$\int\left(\int\left(K_h^n(x-y)\right)^2 dx\right)^2 g(y)dy \leq \frac{C}{h^2}\|K^n\|_2^4 \leq \frac{c}{h^{4s+2}}.$$

Then, for all $h > 0$ small such that $nh^{2s+1} \to \infty$,

$$V_f\left[\sqrt{\frac{\pi(4s+1)n^2h^{4s+1}}{2\|g\|_2^2}}S_1\right] \leq \frac{C}{nh} = o(1), \text{ when } n \to \infty \tag{7}$$

and then

$$\sqrt{\frac{\pi(4s+1)n^2h^{4s+1}}{2\|g\|_2^2}}S_1 \to_P 0, \text{ when } n \to \infty.$$

Convergence of S_2:

$$S_2 = \frac{1}{n^2}\sum_{i\neq j=1}^{n}\langle U_i, U_j\rangle. \tag{8}$$

The variables in S_2 are centred and, moreover, $E_f[\langle U_i, U_j\rangle\langle U_k, U_l\rangle] = 0$ as soon as $(i, j) \neq (k, l)$ and $(i, j) \neq (l, k)$. Then

$$V_f[S_2] = \frac{1}{n^4}E_f\left[\left(\sum_{i\neq j=1}^{n}\langle U_i, U_j\rangle\right)^2\right] = \frac{2}{n^4}n(n-1)E_f[\langle U_1, U_2\rangle^2] = \frac{2+o(1)}{n^2}E_f[\langle U_1, U_2\rangle^2]$$

If we develop this, we get

$$E_f[\langle U_1, U_2\rangle^2] = E_f[\langle K_h^n(x-Y_1), K_h^n(x-Y_2)\rangle^2] - \|K_h \star f\|_2^4.$$

We use again the fact that $\|K_h \star f\|_2^2$ is equal to $\|f\|_2^2$ plus some estimation bias which tends to 0 when $h \to 0$ on the class $W(r, L)$. So, the main term is the first one. Indeed:

$$\begin{aligned}
E_f[\langle K_h^n(\cdot - Y_1), K_h^n(\cdot - Y_2)\rangle^2] &= \int\int\left(\int K_h^n(x-u)K_h^n(x-v)dx\right)^2 g(u)g(v)dudv \\
&= \frac{1}{h}\int\int (M^n)_h^2(v-u)g(u)g(v)dudv,
\end{aligned}$$

where we put $M^n(x) = \int K^n(z+x)K^n(z)dz$. Note that

$$\begin{aligned}\int (M^n(x))^2 dx &= \frac{1}{2\pi}\int \left|\Phi^{\langle K^n(x+\cdot),K^n(\cdot)\rangle}(u)\right|^2 du = \frac{1}{2\pi}\int \left|\Phi^{K^n}(u)\Phi^{K^n}(-u)\right|^2 du \\ &= \frac{1}{2\pi}\int_{|u|\le 1} \frac{du}{|\Phi^\eta(u/h)|^2\,|\Phi^\eta(-u/h)|^2} = \frac{1+o(1)}{\pi(4s+1)h^{4s}}.\end{aligned}$$

Since densities g are continuous functions, even $(r+s-1/2)$ - Lipschitz continuous, see Lemma 3, they are uniformly bounded over f in the Sobolev class $W(r,L)$ with any noise density η under our assumptions. Then for any small $\epsilon > 0$, such that $\epsilon/h \to \infty$, when $n \to \infty$:

$$\begin{aligned}&\left|\int\int (M^n)^2_h(v-u)g(u)g(v)dudv - \int (M^n)^2\|g\|_2^2\right| \\ =\ &\left|\int\int \left((M^n)^2_h(v-u)g(u) - g(v)\int (M^n)^2\right)dug(v)dv\right| \\ \le\ &\int\left|\int (M^n)^2(x)(g(v+hx)-g(v))dx\right| g(v)dv \\ \le\ &\int_{|hx|\le\epsilon} (M^n)^2(x)|hx|^{r+s-1/2}dx + 2\sup_{f,\eta}\|g\|_\infty \int_{|hx|>\epsilon}(M^n)^2(x)dx \le o\left(\int (M^n)^2\right).\end{aligned}$$

This means

$$E_f[\langle K_h^n(x-Y_1), K_h^n(x-Y_2)\rangle^2] = \frac{1+o(1)}{\pi(4s+1)h^{4s+1}}\|g\|_2^2$$

which implies that

$$V_f[S_2] = \frac{(2+o(1))\|g\|_2^2}{\pi(4s+1)n^2h^{(4s+1)}}. \tag{9}$$

Asymptotic normality of S_2. We apply here the following Proposition by Hall (1984):

Proposition 1 (see Theorem 1, Hall (1984)) *Assume $H_n(x,y)$ is a symmetric function such that $E[H_n(X_1,X_2)/X_1] = 0$ almost surely and $E[H_n^2(X_1,X_2)] < \infty$ for each n. Denote by*

$$G_n(x,y) = E[H_n(X_1,x)H_n(X_1,y)].$$

If

$$\left(E[G_n^2(X_1,X_2)] + n^{-1}E[H_n^4(X_1,X_2)]\right)/\left(E[H_n^2(X_1,X_2)]\right)^2 \to 0, \tag{10}$$

as $n\to\infty$, then

$$W_n \equiv \sum_{i<j=1}^n H_n(X_i,X_j)$$

is asymptotically normally distributed with zero mean and variance $n^2E[H_n^2(X_1,X_2)]/2$.

We apply this result to

$$n^2 S_2/2 = \sum_{i<j=1}^{n} \langle U_i, U_j \rangle.$$

We have seen already that this U-statistic is degenerate and that

$$E_f[\langle U_1, U_2 \rangle^2] = \frac{\|g\|_2^2 + o(1)}{\pi(4s+1)h^{4s+1}} < \infty.$$

In order to check (10) we evaluate and bound from above $E_f[G_n^2(Y_1, Y_2)]$ and $E_f[\langle U_1, U_2\rangle^4]$. First, if we replace U_1 and U_2 and we keep the dominant term in the expectation:

$$\begin{aligned}
E_f[\langle U_1, U_2\rangle^4] &\le \int \left(\int \frac{1}{h^2} K^n \left(\frac{u - y_1}{h} \right) K^n \left(\frac{u - y_2}{h} \right) \right)^4 g(y_1) g(y_2) dy_1 dy_2 \\
&\le \frac{1}{h^3} \int \frac{1}{h} \left(K^n \left(z + \frac{y_2 - y_1}{h} \right) K^n(z) dz \right)^4 g(y_1) g(y_2) dy_1 dy_2 \\
&\le \frac{1}{h^3} \int R_h^n(y_2 - y_1) g(y_1) g(y_2) dy_1 dy_2,
\end{aligned}$$

where $R^n(z) = (\int K^n(z+u) K^n(u) du)^4 = (M^n(z))^4$. As in the previous part of this proof, we need to evaluate

$$\int R^n(z) dz = \int (M^n)^4(z) dz = \frac{1}{2\pi} \int \left| \Phi^{M^n} \star \Phi^{M^n}(u) \right|^2 du \le \left(\int \left| \Phi^{M^n}(u) \right|^2 du \right)^2 \le \frac{c}{h^{8s}}.$$

Thus,

$$E_f[\langle U_1, U_2\rangle^4] / \left(n(E_f[\langle U_1, U_2\rangle^2])^2 \right) \le \frac{c/h^{8s+3}}{n/h^{8s+2}} \le \frac{C'}{nh} = o(1) \tag{11}$$

and this proves the first part of (10).

Now, recall (4) and write

$$G_n(y_1, y_2) = \int \langle U_1(\cdot, h, y_1), U_3(\cdot, h, y_3)\rangle \langle U_2(\cdot, h, y_2), U_3(\cdot, h, y_3)\rangle g(y_3) dy_3.$$

We have

$$\begin{aligned}
\langle U_1(\cdot, h, y_1), U_3(\cdot, h, y_3)\rangle &= \int \frac{1}{h^2} K^n \left(\frac{u - y_1}{h} \right) K^n \left(\frac{u - y_3}{h} \right) du \\
&\quad - \frac{1}{h} \int K_h \star f(u) \left[K^n \left(\frac{u - y_1}{h} \right) + K^n \left(\frac{u - y_3}{h} \right) \right] du \\
&\quad + \|K_h \star f\|_2^2
\end{aligned}$$

By changing the variable, the first term on the right-hand side becomes

$$\frac{1}{h} \int K^n \left(u + \frac{y_3 - y_1}{h} \right) K(u) du = M_h^n(y_3 - y_1),$$

where again $M^n(z) = \int K^n(u+z)K^n(u)du$. Then, when we replace this into $E_f[G_n^2(Y_1,Y_2)]$, we keep only the dominant term:

$$
\begin{aligned}
E_f[G_n^2(Y_1,Y_2)] &\le \int\int\left(\int M_h^n(y_3-y_1)M_h^n(y_3-y_2)g(y_3)dy_3\right)^2 g(y_1)g(y_2)dy_1dy_2 \\
&\le \frac{1}{h}\int\int\frac{1}{h}\left(M^n\left(z+\frac{y_2-y_1}{h}\right)M^n(z)g(y_2+hz)dz\right)^2 g(y_1)g(y_2)dy_1dy_2 \\
&\le \frac{1}{h}\int\int\frac{1}{h}\int(M^n)^2\left(z+\frac{y_2-y_1}{h}\right)(M^n)^2(z)g(y_2+hz)dzg(y_1)g(y_2)dy_1dy_2 \\
&\le C\frac{1}{h}\int\int Q_h^n(y_2-y_1)g(y_1)g(y_2)dy_1dy_2,
\end{aligned}
$$

where we used Jensen inequality, the fact that densities g are uniformly bounded by a constant C depending only on r, s, L. We denoted by

$$Q^n(z) = (\int(M^n)^2(z+x)(M^n)^2(x)dx)^2.$$

Similarly to previous calculation of $E_f[\langle U_1,U_2\rangle^2]$

$$\int Q^n(z)dz \quad = \quad \int\int(M^n)^2(z+x)(M^n)^2(x)dxdz = \left(\int(M^n)^2(x)dx\right)^2 \le \frac{C''}{h^{8s}}.$$

Thus,

$$E_f[G_n^2(Y_1,Y_2)]/(E_f[\langle U_1,U_2\rangle^2])^2 \le C'''h = o(1). \tag{12}$$

Inequalities (11) and (12) imply verification of (10) and the proof of asymptotic normality. Thus, together with (9), we get the theorem: $ISE(f_n,f) - MISE(f_n,f)$ is asymptotically normally distributed with mean 0 and variance $2\|g\|_2^2/(\pi(4s+1)n^2h^{4s+1})$. If we take in consideration Lemma 1, plus simple computations, we get the Corollary.

□

3 Other frameworks

We study here the same problem in the framework of supersmooth densities observed with polynomial noise (Section 4.1) and that of Sobolev densities with exponential noise (Section 4.2). As it is known from deconvolution density estimation, the bandwidth minimizing $MISE$ provides much slower rates for smoother noise distribution. Smoother is the noise, harder is the deconvolution problem and slower is the convergence rate to the asymptotic gaussian law.

3.1 Supersmooth densities and polynomial noise

In the previous context, condition $nh^{2s+1} \to \infty$ was necessary to ensure consistency of the $MISE$, but we only need the more classical, less restrictive condition $nh \to \infty$ in order to have S_1 converging in probability to 0 (see (7)) and for the asymptotic normality of the ISE, see (11) necessary to get (10). The fact that f was in the Sobolev class allowed us to evaluate the bias term in $MISE$ and to minimize over $h > 0$ the $MISE$. If we consider instead of Sobolev smoothness classes, a class $S(\alpha, r, L)$ of supersmooth densities f as defined in the Introduction. We know (see Butucea (2004)) that

$$B^2(f_n) = \int \left(E_f[f_n(x)] - f(x)\right)^2 du \le L \exp\left(-\frac{2\alpha}{h^r}\right).$$

Theorem 3 *Let $f_n(\cdot, Y_1, \ldots, Y_n)$ be the kernel density estimator in (2) based on noisy observations with noise having polynomially decreasing Fourier transform and a bandwidth $h \to 0$ such that $nh^{2s+1} \to \infty$, when $n \to \infty$. Then Theorem 1 holds. Moreover, I_n is asymptotically normally distributed with mean and variance given by* (5) *in Corollary 2; if f belongs to the class $S(\alpha, r, L)$, the integrated square error $ISE(f_n, f)$ is asymptotically normally distributed with*

$$MISE(f_n, f) \le L \exp\left(-\frac{2\alpha}{h^r}\right) + \frac{1 + o(1)}{\pi(2s+1)nh^{2s+1}} \text{ and } V_f[ISE(f_n, f)] = \frac{2\|g\|_2^2(1 + o(1))}{\pi(4s+1)n^2h^{4s+1}}$$

and the $MISE$ becomes minimal for

$$h_* = \left(\frac{\log n}{2\alpha} - \frac{2s - r + 1}{2\alpha r} \log\log n\right)^{-1/r}$$

giving

$$\inf_{h>0} \sup_{f \in S(\alpha,r,L)} MISE(f_n, f) = \frac{1}{\pi(2s+1)n} \left(\frac{\log n}{2\alpha}\right)^{(2s+1)/r}.$$

The main density being here much smoother than the variance, we can at the same time choose a bandwidth h that minimizes the $MISE(f_n, f)$ and makes the bias term $\exp(-2\alpha/h^r)$ negligible. Indeed, consider,

$$h = \left(\frac{\log n}{2\alpha} - \sqrt{\log n}\right)^{-1/r}. \tag{13}$$

Then $h/h_* \to 1$, when $n \to \infty$,

$$\exp\left(-\frac{2\alpha}{h^r}\right) = \exp\left(-\frac{2\alpha}{h_*^r}\right) \exp\left(-2\alpha\sqrt{\log n} + \frac{2s+1}{r} \log\log n\right) = o\left(\exp\left(-\frac{2\alpha}{h_*^r}\right)\right)$$

and thus

$$MISE(f_n, f) = \frac{1 + o(1)}{\pi(2s+1)nh^{2s+1}} = \frac{1 + o(1)}{\pi(2s+1)nh_*^{2s+1}}$$

and the confidence interval can be written as in (6) for the bandwidth h in (13).

3.2 Sobolev densities and exponential noise

The situation changes completely if the noise is exponentially smooth. From Butucea and Tsybakov (2003) we know

$$E_f[I_n] = \frac{h^{s-1}(1+o(1))}{2\pi\gamma sn} \exp\left(\frac{2\gamma}{h^s}\right)$$

and this has to be $o(1)$ as a necessary condition for the $MISE$ to be consistent.

Theorem 4 *Let $f_n(\cdot, Y_1, \dots, Y_n)$ be the kernel density estimator defined in (2) based on noisy observations with noise having exponentially decreasing Fourier transform in our convolution model and a bandwidth $h \to 0$ such that $h^{s-1}\exp(2\gamma/h^s)/n \to 0$, when $n \to \infty$. Then*

$$\sqrt{\frac{2\pi\gamma sn^2}{h^{s-1}\exp(4\gamma/h^s)\|g\|_2^2}}\Big(ISE(f_n, f) - E_f[ISE(f_n, f)]\Big) \to N(0,1)$$

where the convergence is in law when $n \to \infty$. Moreover, I_n is asymptotically normally distributed with

$$E_f[I_n] = \frac{h^{s-1}}{2\pi\gamma sn} e^{2\gamma/h^s}(1+o(1)) \text{ and } V_f[I_n] = \frac{h^{s-1}\|g\|_2^2}{2\pi\gamma sn^2} e^{4\gamma/h^s}(1+o(1));$$

if f belongs to the class $W(r, L)$, the integrated square error $ISE(f_n, f)$ is asymptotically normally distributed with

$$MISE(f_n, f) \le Lh^{2r} + \frac{h^{s-1}}{2\pi\gamma sn} e^{2\gamma/h^s}(1+o(1)) \text{ and } V_f[ISE(f_n, f)] = \frac{h^{s-1}\|g\|_2^2}{2\pi\gamma sn^2} e^{4\gamma/h^s}(1+o(1))$$

and the $MISE$ becomes minimal (and of the order of the minimax L_2 risk, see Efromovich (1997)) for h_* of order $(\log n/(2\gamma))^{-1/s}$

$$\inf_{h>0} \sup_{f\in W(r,L)} MISE(f_n, f) = L\left(\frac{\log n}{2\gamma}\right)^{-2r/s}.$$

In this case the bias term, the bias term Lh_*^{2r} is dominating in the expression of $MISE(f_n, f)$.

Proof. Indeed, we can see that

$$V_f[S_1] \le C\frac{\|K^n\|_2^4}{h^2n^3} \le C\frac{h^{2s-2}}{(2\pi\gamma s)^2 n^3} \exp\left(\frac{4\gamma}{h^s}\right),$$

for some constant $C > 0$, and

$$V_f[S_2] = \frac{h^{s-1}\|g\|_2^2(1+o(1))}{2\pi\gamma sn^2} \exp\left(\frac{4\gamma}{h^s}\right).$$

We can see that $V_f[S_1]/V_f[S_2] \leq h^{s-1}/n = o(1)$ and thus S_2 is still the dominating term in the weak convergence to the normal law.

Moreover, $\int (M^n)^2 = (1+o(1))h^s \exp(4\gamma/h^s)/(4\pi\gamma s)$ and finally

$$\left(E[G_n^2(X_1,X_2)] + n^{-1}E[H_n^4(X_1,X_2)]\right) / \left(E[H_n^2(X_1,X_2)]\right)^2 \quad \leq \quad O(h^3) + \frac{O(1)}{nh^3} = o(1).$$

By Proposition 1 we deduce the asymptotic normality. □

4 Auxiliary results

Lemma 2 *Let f_n be the kernel estimator defined in (2) with the particular choice of the kernel and for arbitrary $h > 0$ small. Then*

$$E_f[f_n(x)] = K \star f(x).$$

Moreover, due to the choice of the kernel the cross term in $ISE(f_n, f)$ is null

$$\int \left(f_n(x) - E_f[f_n(x)]\right)\left(E_f[f_n(x)] - f(x)\right) dx = 0.$$

Proof. For the first part, we use the Fourier inversion formula, the expression of the Fourier transform of the kernel and the fact that $\Phi^g = \Phi \cdot \Phi^\eta$:

$$\begin{aligned} E_f[f_n(x)] &= \int \frac{1}{h} K^n\left(\frac{x-y}{h}\right) g(y)dy = \frac{1}{2\pi}\int e^{-ixu}\Phi^{K^n}(hu)\Phi^g(u)du \\ &= \frac{1}{2\pi}\int e^{-ixu}\Phi^K(hu)\Phi(u)du = \int \frac{1}{h}K\left(\frac{x-y}{h}\right) f(y)dy = K_h \star f(x). \end{aligned}$$

Next,

$$\begin{aligned} &\int \left(f_n(x) - E_f[f_n(x)]\right)\left(E_f[f_n(x)] - f(x)\right) dx \\ = &\int \left(f_n(x) - E_f[f_n(x)]\right) E_f[f_n(x)]dx - \int \left(f_n(x) - E_f[f_n(x)]\right) f(x)dx. \end{aligned} \tag{14}$$

Now, the first term of the difference, we use again Plancherel formula (saying that $\int p \cdot q = \int \Phi^p \cdot \overline{\Phi}^q / 2\pi$ for any functions p and q in L_1 and L_2):

$$\begin{aligned} &\int \left(f_n(x) - E_f[f_n(x)]\right) E_f[f_n(x)]dx \\ = &\ \frac{1}{n}\sum_{i=1}^n \int \left(K_h^n(x - Y_i) - K_h \star f\right) K_h \star f(x)dx \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi n}\sum_{i=1}^{n}\int e^{-ixu}\left(\frac{\Phi^K(hu)e^{iuY_i}}{\Phi^\eta(u)} - \Phi^K(hu)\Phi(u)\right)\overline{\Phi^K(hu)}\overline{\Phi(u)}du\\
&= \frac{1}{\pi n}\sum_{i=1}^{n}\int e^{-ixu}\left(\frac{\Phi^K(hu)e^{iuY_i}}{\Phi^\eta(u)} - \Phi^K(hu)\Phi(u)\right)\overline{\Phi(u)}du\\
&= \int \left(f_n(x) - E_f[f_n(x)]\right) f(x)dx,
\end{aligned}
$$

where $\overline{\Phi^K(u)}$ is the complex conjugate of $\Phi^K(u) = I[|u| \le 1]$ and we use the fact that $(\Phi^K)^2 = \Phi^K$. Then the difference in (14) is null. □

Lemma 3 *1) If f belongs to a Sobolev class $W(r, L)$ with $r > 1/2$, then $g = f \star \eta$, with η the density of a polynomial noise, is $(r + s - 1/2)$- Lipschitz continuous function. If f is a supersmooth density in $S(\alpha, r, L)$, then g is at least Lipschitz continuous.*
2) If f is either Sobolev or supersmooth density then f and $g = f \star \eta$ are uniformly bounded densities, whether the noise is polynomial or exponential. That means, there exists a constant $C > 0$, depending only on r, s, L, such that

$$\sup_f \|f\|_\infty \le C \text{ and } \sup_f \|g\|_\infty \le C.$$

3) If the noise is polynomial then the deconvolution kernel defined in (3) *has*

$$\|K^n\|_2^2 = \frac{1 + o(1)}{\pi(2s + 1)h^{2s}},$$

if the noise is exponential, then it has

$$\|K^n\|_2^2 = \frac{h^s(1 + o(1))}{2\pi\gamma s}\exp\left(\frac{2\gamma}{h^s}\right).$$

Proof. 1) If f is in the Sobolev class $W(r, L)$ and η is the density of a polynomial noise, we have:

$$
\begin{aligned}
|g(x + y) - g(x)| &= \frac{1}{2\pi}\left|\int (e^{-iu(x+y)} - e^{-iux})\Phi^g(u)du\right|\\
&\le \frac{1}{2\pi}\int \frac{|e^{-iuy} - 1|}{|u|^{r+s}}|\Phi(u)||u|^r|\Phi^\eta(u)||u|^s du\\
&\le \frac{1}{2\pi}\left(\int \frac{|e^{-iuy} - 1|^2}{|u|^{2(r+s)}}du \int |\Phi(u)|^2|u|^{2r}|\Phi^\eta(u)||u|^{2s}du\right)^{1/2}\\
&\le \frac{|y|^{r+s-1/2}}{2\pi}\left(\int \frac{|e^{-iv} - 1|^2}{|v|^{2(r+s)}}dv\right)^{1/2}\left(\int_{|u|\le M} |\Phi(u)|^2|\Phi^\eta(u)|^2|u|^{2(r+s)}du\right.\\
&\quad \left. + \int_{|u|>M} |\Phi(u)|^2|u|^{2r}du\right)^{1/2}
\end{aligned}
$$

and all the integrals are finite, for any $M > 0$ large enough but fixed. Then there exists a finite constant $C > 0$ that does not depend on x or y, such that

$$|g(x+y) - g(x)| \leq C|y|^{r+s-1/2}.$$

We omit the similar proofs in the cases where either the noise is exponential or the density f is supersmooth.

2) Probability density functions f in the Sobolev class are such that:

$$\begin{aligned} |f(x)| &= \frac{1}{2\pi}\left|\int e^{-ixu}\Phi(u)du\right| \\ &\leq \frac{1}{2\pi}\left(\int |\Phi(u)|^2(1+|u|^{2r})du \int (1+|u|^{2r})^{-1}du\right)^{1/2}, \end{aligned}$$

which is less than some constant C depending only on r and L. Similarly for g.

3) For this we refer to Butucea (2004) and Butucea and Tsybakov (2003). □

References

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise, *Canadian J. Statist.*, 32, to appear.

Butucea, C. and Tsybakov, A.B. (2003). Fast asymptotics in density deconvolution, Manuscript.

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *J. Amer. Statist. Assoc.*, 83, 1184-1186.

Fan, J. (1991*a*). Global behavior of deconvolution kernel estimates, *Statist. Sinica*, 1, 541-551.

Fan, J. (1991*b*). Asymptotic normality for deconvolution kernel density estimators, *Sankhyā Ser. A*, 53, 97-110.

Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution, *IEEE Trans. Inform. Theory*, 48, 734-747.

Fan, Y. and Liu, Y. (1997). A note on asymtptotic normality for deconvolution kernel estimators, *Sankhyā Ser. A*, 59, 138-141.

Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution, *Bernoulli*, 5, 907-926.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, 14, 1-16.

Laurent, B. (1996). Efficient estimation of integral functionals of a density, *Ann. Statist.*, 24, 659-681.

Resum

En aquest article considerem un estimador nucli de la densitat en un model de convolució i donem un teorema central del límit pel seu error quadràtic integrat. L'estimador nucli és força usual en teoria mínimax quan la densitat subjacent es recupera a partir d'observacions amb soroll. El nucli està fixat i depèn fortament de la distribució de l'error, la qual se suposa totalment coneguda. L'amplada de banda no està fixada, els resultats es verifiquen per qualsevol seqüència d'amplades decreixents cap a 0. En particular, es pot aplicar el teorema central del límit per l'amplada de banda que minimitza l'error quadràtic integrat mitjà. Les velocitats de convergència són força diferents en el cas de sorolls regulars i de sorolls super-regulars. La suavitat de la densitat subjacent és rellevant en l'avaluació del l'error quadràtic integrat mitjà.

MSC: 62G05, 62G20

Paraules clau: error quadràtic integrat, estimació no paramètrica de la densitat, model de convolució, observacions amb soroll, teorema central del límit