
Volume 137
Issue 3
1 March 2007
ISSN 0378-3758
ELSEVIER
journal of statistical planning and inference
Special Issue on Nonparametric Statistics and Related Topics: In honor of M.L. Puri
Edited by: S. Ralescu
G. Haiman
F. Ruymgaart
For Contents see page iii
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 712–719
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Estimation of a quadratic regression functional using the sinc kernel

Nicolai Bissantz*, Hajo Holzmann

Institute for Mathematical Stochastics, Georg-August-University Göttingen, Maschmühlenweg 8-10, D-37073 Göttingen, Germany

Available online 18 July 2006

Abstract

We use the sinc kernel to construct an estimator for the integrated squared regression function. Asymptotic normality of the estimator at different rates is established, depending on whether the regression function vanishes or not.

MSC: Primary 62G07; 94A12; secondary 41A25; 41A45

Keywords: Asymptotic normality; Nonparametric regression; Quadratic regression functional; Sinc kernel

1. Introduction

Suppose that f is a function with compact support, say $\operatorname{supp}(f) \subset [-1, 1]$, and consider the regression model with fixed, equidistant design

$$Y_k = f(t_k) + \varepsilon_k, \quad k = -n, \ldots, n.$$

Here $t_k = k/n$, $k = -n, \ldots, n$, and ε_i is an i.i.d. additive noise with $E\varepsilon_k = 0$, $\operatorname{Var} \varepsilon_k = \sigma^2$ and $E\varepsilon_k^4 < \infty$.

In this paper we focus on estimation of the quadratic regression functional

$$\|f\|^2 = \int_{-1}^{1} f^2(t)\, dt.$$

To this end we use an estimator based on the so-called sinc kernel $\operatorname{sinc}(t) = \sin(t)/(\pi t)$. The sinc kernel has been widely applied in the context of density estimation, see e.g. Davis (1975, 1977), Devroye (1992) or Glad et al. (2003), because it has certain optimality properties in terms of mean square error and mean integrated square error for sufficiently smooth densities. It also arises naturally from a spectral cut-off procedure which is frequently applied in inverse statistical estimation (cf. Rooij et al., 1999).

As an estimator for the regression function itself we use

$$\hat{f}_{n,m}(t) = n^{-1} m \sum_{k=-n}^{n} \operatorname{sinc}(m(t - t_k))\, Y_k, \quad m > 0, \tag{1}$$

* Corresponding author. Fax: +49 551 39 13505.
E-mail address: nicolai.bissantz@gmx.de (N. Bissantz)
URL: http://www.stochastik.math.uni-goettingen.de .

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.06.004

which we call the *Fourier kernel estimator* due to its origin as inverse Fourier transform of the indicator function, and for the quadratic regression functional

$$\hat{N}_n^2 := \int_{-1}^{1} \hat{f}_{n,m}^2(t)\,\mathrm{d}t.$$

In Section 2, we show uniform convergence of $\hat{f}_{n,m}$ to the regression f with optimal rates under certain smoothness assumptions on f, given in terms of the tail behaviour of its Fourier transform. However, it is well known that estimators based on the sinc kernel have a certain wiggliness, and often do not accurately represent the target function. Therefore, in order to estimate the regression function itself, flat-top or supersmooth kernels can be a better choice (cf. McMurry and Politis, 2004). Supersmooth kernels also simplify bandwidth choice, particularly since they are absolutely integrable (cf. Devroye and Lugosi, 2001, Chapter 17, and Delaigle and Gijbels, 2004, for a simulation study in the context of density estimation).

The main goal of this paper is to examine the asymptotic distribution of the estimator $\hat{N}_n^2$ of the quadratic regression functional $\|f\|^2$. In Section 3 we show that $\hat{N}_n^2$ is asymptotically normally distributed at a $\sqrt{n}$-rate. However, for $f=0$, the limiting distribution is degenerate, and a non-degenerate normal limit law appears with rate $n/\sqrt{m}$. Similar results have been obtained by Huang and Fan (1999) for a regression function and its derivatives using local polynomial estimators based on compactly supported kernels, and by Bickel and Ritov (1988) for density estimators. However, the phenomenon of different rates for $f \neq 0$ and $f=0$ appears to be new in the context of quadratic regression functionals. Let us stress that the above-mentioned wiggliness of the Fourier kernel estimator is inessential when using it to estimate quadratic regression functionals. Also the choice of bandwidth seems to be less important for the estimation of such functionals than for estimation of the function itself. Recommendations could be based on a simulation study; cf. Dette (1999) for an ad hoc choice in a related context.

Finally, in the Appendix we collect some technical lemmas.

2. Uniform convergence

In this section we show uniform pointwise convergence of the Fourier kernel estimator (1). The following regularity condition on the tail behaviour of the Fourier transform χ_f of the regression function f will be essential.

Assumption 1. The Fourier transform χ_f of f satisfies

$$|\chi_f(\omega)| \leqslant d|\omega|^{-(\alpha+1/2)}, \quad |\omega| \geqslant 1, \text{ for some } d \geqslant 0 \text{ and some } \alpha > \tfrac{1}{2}. \tag{2}$$

Theorem 2. *Suppose that* $m = \mathrm{o}(n^{2/3})$ *for* $n, m \to \infty$, $\operatorname{supp}(f) \subseteq [-1,1]$, *f is Lipschitz continuous, and that the Fourier transform* χ_f *of f satisfies Assumption* 1. *Then* $\hat{f}_{n,m}$ *is uniformly consistent for all* $t \in [-1,1]$ *with*

$$\sup_{t\in[-1,1]} \operatorname{Bias} \hat{f}_{n,m}(t) = \sup_{t\in[-1,1]} |E\hat{f}_{n,m}(t) - f(t)| = \mathrm{O}(m^{-(\alpha-1/2)}) + \mathrm{O}(m^{3/2}/n) \tag{3}$$

and variance

$$\sup_{t\in[-1,1]} \operatorname{Var} \hat{f}_{n,m}(t) = \mathrm{O}(m/n).$$

The mean-squared error converges uniformly to 0 *with rate*

$$\sup_{t\in[-1,1]} E((\hat{f}_{n,m}(t) - f(t))^2) = \mathrm{O}(m^3/n^2) + \mathrm{O}(m^{-(2\alpha-1)}) + \mathrm{O}(m/n).$$

Remark 1. Lipschitz continuity of f follows from Assumption 1 if $\alpha > \frac{3}{2}$.

Remark 2. Assumption 1 on the tails of the Fourier transform of χ_f implies continuity of f on the whole real line. Since $\operatorname{supp}(f) \subseteq [-1,1]$ it follows in particular that $f(1) = f(-1) = 0$. This property of f allows to show uniform convergence of the estimator on $[-1,1]$. Without such a condition kernel regression estimators without boundary correction converge to $f(x)/2$ at the boundary points, and not to the regression function (cf. Efromovich, 1999, p. 330).

Remark 3. The sinc-kernel estimator achieves asymptotic (rate) optimality. This can be seen as follows. Let, say, f be an L_1-function with Fourier transform χ_F, and $|\chi_f(\omega)| \leqslant d|\omega|^{-1-m}$, $m \in \mathbb{N}$, $m \geqslant 2$. This is (1) with $\alpha = m + \frac{1}{2}$. Then, according to Theorem 2, the pointwise MSE of the sinc-kernel estimator is $\mathrm{O}(n^{-2m/(2m+1)})$. The class of functions for which Assumption 1 holds with $\alpha = m + 1/2$ (see e.g. Chandrasekharan, 1989, p. 20) contains the class of m-times continuously differentiable functions. This class is basically similar to the class $\mathbf{C}_m$ defined in Fan and Gijbels (1996) if some additional regularity assumptions on the mth derivative of f are made. For the class $\mathbf{C}_m$ the rate of convergence of the linear minimax risk is $n^{-2m/(2m+1)}$. Therefore, the sinc-kernel estimator achieves the minimax rate of convergence in such function classes.

Proof of Theorem 2. We proceed mostly along the lines of the proof of Theorem 1 in Pawlak and Stadtmüller (1997), which is concerned with bandlimited functions and a regression model on the whole real line. Write

$$
\begin{aligned}
E(\hat{f}_n(t)) &= m \int_{-(n+1/2)/n}^{(n+1/2)/n} f(u)\,\mathrm{sinc}(m(t-u))\,\mathrm{d}u \\
&\quad + m \sum_{|k| \leqslant n} f(t_k) \left(n^{-1}\,\mathrm{sinc}(m(t-t_k)) - \int_{(k-1/2)/n}^{(k+1/2)/n} \mathrm{sinc}(m(t-u))\,\mathrm{d}u \right) \\
&\quad + m \sum_{|k| \leqslant n} \int_{(k-1/2)/n}^{(k+1/2)/n} \mathrm{sinc}(m(t-u))(f(t_k) - f(u))\,\mathrm{d}u \\
&= \mathbf{A}(t) + \mathbf{B}(t) + \mathbf{C}(t).
\end{aligned}
$$

The following estimates, which hold uniformly for $t \in [-1, 1]$, are obtained by straightforward calculations using the Cauchy–Schwarz inequality and Lipschitz continuity of f:

$$
|\mathbf{B}(t)| = \mathrm{O}(m^{3/2}/n), \quad |\mathbf{C}(t)| = \mathrm{O}(m^{1/2}/n). \tag{4}
$$

To deal with the term $\mathbf{A}$ we write $\mathbf{A}(t) = f(t) - \int_{|\omega| \geqslant m} \chi_f(\omega) \mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega$, and estimate the remainder term as follows:

$$
\left| \int_{|\omega| \geqslant m} \chi_f(\omega) \mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega \right| \leqslant \int_{|\omega| \geqslant m} \mathrm{d}|\omega|^{-(\alpha+1/2)}\,\mathrm{d}\omega \leqslant 2\,\mathrm{d}m^{-(\alpha-1/2)}, \tag{5}
$$

which again holds uniformly in t. Now, (3) follows from (4) and (5). As for the variance of $\hat{f}_{n,m}(t)$, we estimate

$$
\mathrm{Var}\,\hat{f}_{n,m}(t) = \sigma^2 m\, n^{-1} \left(\int_{-m}^{m} \mathrm{sinc}^2(mt - x)\,\mathrm{d}x + \mathrm{O}(m/n) \right) = \mathrm{O}(m/n). \qquad \square
$$

3. Asymptotic normality

In this section we analyse the asymptotic distribution of the quadratic regression functional

$$
\hat{N}_n^2 := \int_{-1}^{1} \hat{f}_{n,m}^2(t)\,\mathrm{d}t = Y^{\mathrm{T}} A Y,
$$

where

$$
A = (a_{j,k})_{|j|,|k| \leqslant n}, \quad a_{j,k} = (m/n)^2 \int_{-1}^{1} \mathrm{sinc}(m(t-t_j))\,\mathrm{sinc}(m(t-t_k))\,\mathrm{d}t.
$$

In the following theorem, we show asymptotic normality of $\hat{N}_n^2$ in two different cases.

Theorem 3. *If* $f=0$ *and* $\log(n)/\sqrt{m} \to 0$, $m^{3/2}/n \to 0$ *as* $n, m \to \infty$, *we have*

$$n\, m^{-1/2}\, \hat{N}_n^2 - 2m^{1/2}\sigma^2/\pi \xrightarrow{\mathbf{D}} \mathbf{N}(0, 4\sigma^4/\pi). \tag{6}$$

If $f \not\equiv 0$ *is Lipschitz-continuous, has compact support* $\operatorname{supp}(f) \subseteq [-1, 1]$, *and satisfies Assumption* 1, *and if* $m^2 = \mathrm{o}(n)$, *and* $m^{-2\alpha}\sqrt{n} = \mathrm{o}(1)$ *as* $n, m \to \infty$, *then*

$$\sqrt{n}(\hat{N}_n^2 - \|f\|^2) \xrightarrow{\mathbf{D}} \mathbf{N}(0, 4\sigma^2\|f\|^2).$$

Remark 4. It is remarkable that different rates appear in the two cases $f=0$ and $f \not\equiv 0$. Such a phenomenon was also observed by Dette (1999). He considered a statistic based on the difference of a parametric and a nonparametric variance estimator to test the validity of a linear regression model, and obtained different rates under the hypothesis and under fixed alternatives. Our rates correspond to those of Dette, if the smoothing parameter m is replaced by the multiplicative inverse of a bandwidth.

Remark 5. Theorem 3 gives rise to a consistent test for the hypothesis $f=0$. The potential power of this test is indicated by the consideration of local alternatives. Similar to Dette (1999) we obtain for the limiting variance under local alternatives the value $4\sigma^4/\pi$ as in (6). Dette's (1999) result closely resembles (6), if the smoothing length h of the nonparametric estimator in Dette's test is replaced by the multiplicative inverse of our smoothing parameter Ω. However, Dette's regression model is $y_{j,n} = y(t_{j,n}) = m(t_{j,n}) + \varepsilon_{j,n}$, $j = 1, \ldots, n$ for design points $t_{1,n}, \ldots, t_{n,n} \in [0, 1]$. This differs slightly from our setting, both in the number of design points (n instead $2n+1$) and the size of the support of the design density ([0, 1] instead of $[-1, 1]$). A close inspection of our proofs shows that if our regression model is changed into n equally spaced observations on [0, 1], the variance of (6) becomes $\mu_0^2 := 2\sigma^4/\pi \approx 0.64\sigma^4$. The asymptotic variance μ_0^2 of Dette's test (his equation (2.13)) depends on the kernel used for the nonparametric variance estimator. In his numerical simulations he used the Epanechnikov kernel. For this kernel $\mu_0^2 \approx 1.70\sigma^4$. Furthermore, for the Gaussian kernel $\mu_0^2 \approx 0.81\sigma^4$, and for the Fourier estimate kernel as discussed in this paper $\mu_0^2 = 2\sigma^4/\pi$, thus the variance for our test with this kernel is recovered by Dette's equation (2.13). However, note that for the Gaussian kernel and the Fourier estimate kernel the assumption of compact kernel support does not hold, so Dette's theorems cannot be applied to these kernels. Hence our result extends Dette's result to the Fourier estimate kernel which is more powerful than tests based on the kernels mentioned above.

Remark 6. An inspection of the proof indicates that a similar result could be obtained for a fixed, but nonequidistant design with differentiable, nonzero design density. However, in that case the design density appears in certain integrals, which makes application of Fourier-based methods almost impossible, see e.g. Lemma 4. This could make the proof much more tedious.

Proof of Theorem 3. The proof goes along the lines of proof of Theorems 1 and 2 in Dette (1999). However, the actual computations, based on arguments involving the Fourier transform, are completely different. The expectation of our statistic is given by

$$E\hat{N}_n^2 = \sigma^2 \operatorname{tr}(A) + \sum_{|j|,|k| \leqslant n} a_{j,k} f(t_j) f(t_k), \tag{7}$$

and thus

$$\hat{N}_n^2 - E[\hat{N}_n^2] = \sum_{\substack{|j|,|k| \leqslant n,\\ j \neq k}} a_{j,k}(\varepsilon_j\varepsilon_k - f(t_j)f(t_k)) + \sum_{|j| \leqslant n} a_{j,j}(\varepsilon_j - \sigma^2) = T_1 + T_2.$$

Firstly let us consider the case $f=0$. Then

$$E\,T_1 = E\,T_2 = E(T_1T_2) = \operatorname{Cov}(T_1, T_2) = 0$$

and

$$\operatorname{Var} T_1 = 2\sigma^4 \sum_{\substack{|j|,|k|\leqslant n,\\ j\neq k}} a_{j,k}^2, \quad \operatorname{Var} T_2 = (\mu_4 - \sigma^4) \sum_{|j|\leqslant n} a_{j,j}^2.$$

Next we estimate the asymptotic behaviour of expectation and variance. For the expectation tedious but straightforward computations show that

$$\operatorname{tr}(A) = (m/n)^2 \sum_{k=-n}^{n} m^{-1} \int_{m(-1-t_k)}^{m(1-t_k)} \operatorname{sinc}^2(t)\,\mathrm{d}t = 2m/(\pi n) + \mathrm{O}(\log(n)/n).$$

As for the variance, we start by estimating $\operatorname{Var} T_2$

$$\sum_{|j|\leqslant n}^{n} a_{j,j}^2 = m^2 n^{-4} \sum_{|j|\leqslant n} \left(\int_{-\infty}^{\infty} \operatorname{sinc}^2(t)\,\mathrm{d}t - \int_{|t|\geqslant m(1-t_j)} \operatorname{sinc}^2(t)\,\mathrm{d}t \right)^2 = \mathrm{O}(m^2/n^3).$$

From Lemma 4 it follows that the asymptotic variance of T_1 is given by

$$\operatorname{Var} T_1 = 4\sigma^4 m/(\pi n^2) + \mathrm{o}(m/n^2).$$

Therefore, in case $f = 0$ we have

$$\hat{N}_n^2 = 2\sigma^2 m/(\pi n) + \mathrm{O}(\log(n)/n) + T_1 + O_P(m/n^{3/2}),$$

and it suffices to show asymptotic normality of T_1. To this end we apply Theorem 5.2 in de Jong (1987). Assumptions (1) and (2) of de Jong's theorem hold with $K(n) = m^{1/4}$. Moreover, the maximal eigenvalue of A is estimated as

$$\sigma(n)^{-1} \max_{i=-n}^{n} |\mu_i| = \mathrm{O}(m^{-1/2}(\log m)^2),$$

by applying Gerschgorin's theorem. In fact,

$$\begin{aligned}\sigma(n)^{-1} \max_{|j|\leqslant n} |\mu_j| &\leqslant \sigma(n)^{-1} \max_{|j|\leqslant n} \sum_{|k|\leqslant n} |a_{j,k}| \\ &= c m^{-1/2} \max_{|j|\leqslant n} n^{-1} \sum_{|k|\leqslant n} \left| \int_{-m}^{m} \operatorname{sinc}(t - mt_j) m \operatorname{sinc}(t - mt_k)\,\mathrm{d}t \right| = \mathrm{O}(m^{-1/2}\log^2(m)),\end{aligned}$$

where we have used that

$$\sup_{x\in[0,m]} \int_0^m |\operatorname{sinc}(z-x)|\,\mathrm{d}z = \sup_{x\in[0,m]} \int_{-x}^{m-x} |\operatorname{sinc}(z)|\,\mathrm{d}z = \mathrm{O}(\log(m)),$$

and c is some generic constant. This proves (6).

Now let us consider the case $f \neq 0$. In this case the second term on the right-hand side of (7) is $\|f\|^2 + \mathrm{o}(n^{-1/2})$ by Lemma 5. In the variance $\operatorname{Var}(\hat{N}_n^2 - E\hat{N}_n^2)$ there appears an additional term

$$4\sigma^2 \sum_{i,j,l=-n}^{n} a_{ij} a_{il} f(t_j) f(t_l) = 4\sigma^2 m^2/n^2 \sum_{|k|\leqslant n} r_k^2,$$

where

$$r_k = \int_{-1}^{1} \left(mn^{-1} \sum_{|j|\leqslant n} f(t_j) \operatorname{sinc}(m(t-t_j)) \right) \operatorname{sinc}(m(t-t_k))\,\mathrm{d}t.$$

By Lemma 6 this equals

$$4\sigma^2 m^2 n^{-2} \sum_{|k|\leqslant n} \left(\int_{-1}^{1} f(t)\,\mathrm{sinc}(m(t-t_k))\,\mathrm{d}t + \mathrm{O}(m^{1/2}n^{-1/4})\right)^2$$
$$= 4\sigma^2 m^2 n^{-2} \sum_{|k|\leqslant n} \left(n^{-1} \sum_{|j|\leqslant n}^{n} f(t_j)\,\mathrm{sinc}(m(t_j-t_k)) + \mathrm{O}(m^{1/2}n^{-1/4}) + \mathrm{O}(m/n)\right)^2.$$

Using Theorem 2 we can further simplify the last expression as

$$(m/n)^2(2\sigma/m)^2 \sum_{|k|\leqslant n} ((f(t_k)+\mathrm{o}(1)) + \mathrm{O}(m^{1/2}n^{-1/4}))^2$$
$$= 4\sigma^2 n^{-2} \sum_{|k|\leqslant n} (f^2(t_k)+\mathrm{o}(1)) = 4\sigma^2 n^{-1}\|f\|^2 + \mathrm{o}(1/n),$$

and $\mathrm{Var}\,\hat{N}_n^2$ is asymptotically equal to $4\sigma^2\|f\|^2/n$. In summary, if $f \neq 0$ we have

$$\hat{N}_n^2 = \|f\|^2 + 2\sum_{|j|,|k|\leqslant n} a_{j,k}\varepsilon_j f(t_k) + \mathrm{o}_P(n^{-1/2}),$$

and the conclusion follows by an application of Lyapounov's theorem to the variance dominating term

$$\sum_{|j|,|k|\leqslant n} a_{j,k}\varepsilon_j f(t_k). \qquad \square$$

Acknowledgements

We would like to thank A. Munk, M. Pawlak, F. Ruymgaart and U. Stadtmüller for helpful comments. Furthermore, we would like to thank an anonymous referee for helpful suggestions and for pointing out several important references.

Appendix

The Fourier transform is given by $\mathscr{F}(f)(\omega) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega x} f(x)\,\mathrm{d}x$, so that $\mathscr{F}(\mathrm{sinc})(\omega) = \mathbf{1}_{[-1,1]}(\omega)$.

Lemma 4. *Let* $m^{3/2} = \mathrm{o}(n)$ *and* $\log(n) = \mathrm{o}(m^{1/2})$ *as* $n, m \to \infty$, *then*

$$\sum_{\substack{|j|,|k|\leqslant n,\\ j\neq k}} \left[\int_{-1}^{1} \mathrm{sinc}(m(t-t_j))\,\mathrm{sinc}(m(t-t_k))\,\mathrm{d}t\right]^2 = 2n^2/(\pi m^3) + \mathrm{o}(n^2/m^3).$$

Proof. First observe that

$$\int_{-1}^{1} \mathrm{sinc}(m(t-t_j))\,\mathrm{sinc}(m(t-t_k))\,\mathrm{d}t$$
$$= m^{-1}\int_{-m}^{m} \mathrm{sinc}(mt_j - t)\,\mathrm{sinc}(mt_j + (m(t_k-t_j)-t))\,\mathrm{d}t = m^{-1}(\mathrm{sinc}(m(t_j+t_k)) - \tilde{c}_{j,k}),$$

where

$$\tilde{c}_{j,k} := \int_{|t|\geqslant m} \mathrm{sinc}(t-mt_j)\,\mathrm{sinc}(t-mt_k)\,\mathrm{d}t.$$

Next we estimate the sum over the squared sinc functions. In fact, tedious but straightforward calculations show that

$$
\begin{aligned}
&\sum_{\substack{|j|,|k|\leqslant n,\\ j\neq k}} \operatorname{sinc}^2(m(t_j+t_k))\\
&= \sum_{|j|,|k|\leqslant n} \operatorname{sinc}^2(m(t_j+t_k)) - \sum_{|j|\leqslant n} \operatorname{sinc}^2(2mt_j)\\
&= 2\cdot\left((2n+1)\sum_{k=1}^{2n}\operatorname{sinc}(mt_k)^2 - \sum_{k=1}^{2n} k\operatorname{sinc}(mt_k)^2 + \mathrm{O}(n)\right)\\
&= 2\cdot\left((2n+1)\left(\int_{1/2}^{2n+1/2}\operatorname{sinc}(mt/n)^2\,\mathrm{d}t + \mathrm{O}(m/n)\right)\right.\\
&\quad\left. + \int_{1/2}^{2n+1/2} t\operatorname{sinc}(mt/n)^2 + \mathrm{O}(n^2\log(n)/m^2) + \mathrm{O}(n)\right)\\
&= 2n^2/(\pi m) + \mathrm{o}(n^2/m).
\end{aligned}
\tag{8}
$$

Finally, for the sum over the squared error terms $\tilde{c}^2_{j,k}$ is tedious, but straightforward computations yield

$$
\sum_{\substack{|j|,|k|\leqslant n,\\ j\neq k}} \tilde{c}^2_{j,k} = \mathrm{o}(n^2/m),
$$

which shows its asymptotic negligibility as compared to the sum over the squared sinc-functions. This completes the proof of the lemma. □

Lemma 5. *Suppose* $\operatorname{supp}(f)\subseteq[-1,1]$, *that f is Lipschitz-continuous, and that the Fourier transform* χ_f *of f satisfies Assumption* 1. *If* $m^{-2\alpha}\sqrt{n}=\mathrm{o}(1)$ *and* $m^2=\mathrm{o}(n)$, *then*

$$
\|E\hat{f}_{n,m} - f\|^2_{L_2[-1,1]} = \mathrm{o}(n^{-1/2}),
$$

and in particular

$$
\|E\hat{f}_{n,m}\|^2 = \|f\|^2_{L_2[-1,1]} + \mathrm{o}(n^{-1/2}).
$$

Proof. We let χ_n denote the Fourier transform of $E\hat{f}_{n,m}$. Using Parseval's equality, we compute

$$
\|f_n - f\|^2_{L_2[-1,1]} \leqslant \|f_n - f\|^2_{L_2(-\infty,\infty)} = (2\pi)^{-1}\|\chi_n - \chi_f\|^2_{L_2(-\infty,\infty)} \tag{9}
$$

$$
= (2\pi)^{-1}\left(\int_{-m}^{m}\left|\chi_n(\omega)-\chi_f(\omega)\right|^2\mathrm{d}\omega + \int_{|\omega|\geqslant m}|\chi_f|^2(\omega)\,\mathrm{d}\omega\right) = \mathbf{E}/(2\pi) + \mathrm{O}(m^{-2\alpha}). \tag{10}
$$

In order to estimate $\mathbf{E}$, observe that $\chi_n(\omega) = \mathbf{1}_{[-m,m]}(\omega)\left(\chi_f(\omega)+\bar{c}(\omega)\right)$, where $\bar{c}(\omega)$ is estimated as follows:

$$
\begin{aligned}
|\bar{c}(\omega)| &= \left|\sum_{|j|\leqslant n}\left(n^{-1}f\left(t_j\right)\mathrm{e}^{\mathrm{i}\omega t_j} - \int_{(j-1/2)/n}^{(j+1/2)/n}\mathrm{e}^{\mathrm{i}\omega t}f(t)\,\mathrm{d}t\right)\right|\\
&\leqslant n^{-1}\sum_{|j|\leqslant n}\ \sup_{\xi\in[(j-1/2)/n,(j+1/2)/n]} |\mathrm{e}^{\mathrm{i}\omega t_j}f(t_j) - \mathrm{e}^{\mathrm{i}\omega\xi}f(\xi/n)|\\
&\leqslant n^{-1}L'\cdot(2n+1)/n = \mathrm{O}(1/n),
\end{aligned}
\tag{11}
$$

independently of m and $\omega\in[-m,m]$, where L' is a Lipschitz constant for $\mathrm{e}^{\mathrm{i}\omega}\cdot f(\cdot)$. The lemma follows from (9) and (11). □

Lemma 6. *Let the conditions of Lemma* 5 *hold*. *Then*

$$\left| \int_{-1}^{1} \left((m/n) \sum_{|k| \leqslant n} f(t_k) \operatorname{sinc}(m(t-t_k)) \right) \operatorname{sinc}(m(t-t_j))\, \mathrm{d}t - \int_{-1}^{1} f(t) \operatorname{sinc}(m(t-t_j))\, \mathrm{d}t \right| = \mathrm{o}((nm)^{-1/2}).$$

The proof is straightforward using the Cauchy–Schwarz inequality and Lemma 5 and is therefore omitted.

References

Bickel, P., Ritov, Y., 1988. Estimating integrated squared density derivatives: sharp best order of convergence estimates. Sankhya Ser. A 50, 381–393.

Chandrasekharan, K., 1989. Classical Fourier Transforms. Springer, Berlin.

Davis, K.B., 1975. Mean square error properties of density estimates. Ann. Statist. 3, 1025–1030.

Davis, K.B., 1977. Mean integrated square error properties of density estimates. Ann. Statist. 5, 530–535.

de Jong, P., 1987. A central limit theorem for generalized quadratic forms. Probab. Theory Related Fields 75, 261–277.

Delaigle, A., Gijbels, I., 2004. Practical bandwidth selection in deconvolution kernel density estimation. Comput. Statist. Data Anal. 45, 249–267.

Dette, H., 1999. A consistent test for the functional form of a regression based on a difference of variance estimators. Ann. Statist. 27, 1012–1040.

Devroye, L., 1992. A note on the usefulness of superkernels in density estimation. Ann. Statist. 20, 2037–2056.

Devroye, L., Lugosi, G., 2001. Combinatorial Methods in Density Estimation. Springer, New York.

Efromovich, S., 1999. Nonparametric Curve Estimation. Springer, New York.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

Glad, I.K., Hjort, N.L., Ushakov, N.G., 2003. Correction of density estimators that are not densities. Scand. J. Statist. 30, 415–427.

Huang, L.-S., Fan, J., 1999. Nonparametric estimation of quadratic regression functionals. Bernoulli 5, 927–949.

McMurry, T.L., Politis, D.N., 2004. Nonparametric regression with infinite order flat-top kernels. J. Nonparametric Statist. 16, 549–562.

Pawlak, M., Stadtmüller, U., 1997. Nonparametric estimation of a class of smooth functions. J. Nonparametric Statist. 8, 149–183.

Rooij, A.C.M., Ruymgaart, F.H., van Zweet, W.R., 1999. Asymptotic efficiency of inverse estimators. Theory Probab. Appl. 4, 722–738.