Universität
Stuttgart

Fachbereich
Mathematik

Optimal convergence rates for density estimation from grouped data

Alexander Meister

Preprint 2006/008

Universität Stuttgart
Fachbereich Mathematik

Optimal convergence rates for density estimation from grouped data

Alexander Meister

Preprint 2006/008

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik/uni-stuttgart.de/preprints

ISSN **1613-8309**

Optimal convergence rates for density estimation from grouped data

Alexander Meister[1]

Institut für Stochastik und Anwendungen,
Universität Stuttgart
D-70569 Stuttgart, Germany

Abstract: We derive the optimal convergence rates for density estimation based on aggregated observations under common smoothness conditions for symmetric densities. We study a procedure for data-driven bandwidth selection and give an extent to skew densities.

Key words: aggregated data; cross-validation; deconvolution; minimax convergence rates; nonparametric estimation.

MSC: 62G07

[1]Alexander Meister
Institut für Stochastik und Anwendungen,
Universität Stuttgart
Pfaffenwaldring 57
D-70569 Stuttgart, Germany

email: meistear@mathematik.uni-stuttgart.de

1. Introduction

The statistical problem of estimating a density function when only aggregated data are observed has received considerable attention, where research is mainly motivated by real applications in the field of econometrics. For a study on parameter identification from averaged data, see the recent note of Machado & Santos Silva (2006). A nonparametric approach to density estimation from aggregated observations is given by Linton & Whang (2002).

In the mathematical model, we define the i.i.d. random variables X_{ij}, integer i, j, having the density function f_X, which we aim to estimate. The only empirical access is given by the data $Y_1, \dots, Y_n$ where

$$Y_i \;=\; \sum_{j=1}^{m} X_{ij}\,, \tag{1}$$

where m is the fixed size of the data groups. Note that our problem may also be seen as a missing data problem in time series analysis. Suppose an autoregressive process AR(1) involving Z_k, integer k, where $Z_{k+1} = Z_k + X_k$ and we are interested in the density of the X_k while only the $Z_{m \cdot l}$ are observed for each integer l.

The current note intends to advance the understanding of this problem by deriving the optimal rates of convergence under common smoothness conditions on f_X with respect to the mean integrated squared error (MISE). In Section 2, we study the optimal rates for symmetric densities. In Section 3, we describe a procedure for data-driven bandwidth selection; in Section 4 we give an extent to skew densities. The proofs are deferred to Section 5.

To give a survey on related problems of indirect density estimation, we mention problems of reconstructing a density from measurement error; that topic has become famous as density deconvolution (Stefanski & Carroll (1990), Carroll & Hall (1988), Zhang (1990), Horowitz & Markatou (1996) among many others). Another contribution is given by Schick & Wefelmeyer (2004) where estimation of the density of the sums of independent random variables is studied; hence, somehow, our consideration can be seen as the corresponding inverse problem. The problem of estimating the density of independent components of a Poisson sum is considered in van Es et al. (2005).

2. Minimax rates

We focus on those f_X which are symmetric around zero and, hence, have a real-valued Fourier transform $f_X^{ft}(t)$. As the density f_Y of each observation Y_i turns out to be the m-fold self-convolution of f_X, it is convenient to apply a Fourier approach, similarly to Linton & Whang (2002). With respect to the characteristic functions f_X^{ft} and f_Y^{ft}, we have

$$f_Y^{ft}(t) \;=\; \big[f_X^{ft}(t)\big]^m\,.$$

We assume that $f_X(x) = f_X(-x)$ for almost all x; and, for even m, in addition,

$$f_X^{ft}(t) \neq 0\,, \quad \forall t\,. \tag{2}$$

The necessity of condition (2) to ensure identifiability of f_X will be shown in the following:

We define the density

$$f_0(x) \;=\; \big(1 - \cos(x)\big)/(\pi x^2)\,, \tag{3}$$

having the triangle-shaped Fourier transform $f_0^{ft}(t) \;=\; 1 - |t|$ on $t \in [-1, 1]$. Therefore, the densities $f_\pm(x) = f_0(x) \pm (1/2) f_0(x) \cos(2x)$ possess the Fourier transforms

$$f_\pm^{ft}(t) \;=\; f_0^{ft}(t) \pm \frac{1}{4}\big[f_0^{ft}(t+2) + f_0^{ft}(t-2)\big]\,,$$

hence we have $\left[f_+^{ft}(t)\right]^2 = \left[f_-^{ft}(t)\right]^2$ for all t. This proves that $f_X = f_+$ cannot be uniquely reconstructed from the observation density $f_Y = f_+ * f_+$ in this example; where $*$ denotes convolution.

The empirical Fourier transform is denoted by $\hat f_Y^{ft}(t) = \frac{1}{n}\sum_{k=1}^n \exp(itY_k)$. We define the estimator of f_X by Fourier inversion,

$$\hat f_X(x) = \frac{1}{2\pi}\int \exp(-itx)K^{ft}(th)\big|\hat f_Y^{ft}(t)\big|^{1/m}dt\,, \tag{4}$$

where K denotes a square-integrable kernel function where $K^{ft}(0) = 1$ and K^{ft} is compactly supported; parameter h denotes the bandwidth.

In order to establish rates of convergence, we propose common smoothness conditions on f_X by assuming a uniform upper bound on the Sobolev norm of f_X. We introduce the class $\mathcal{F}_{\beta C}$; whose elements f are even densities; they satisfy (2) and

$$\int |f_X^{ft}(t)|^2(1+t^2)^\beta dt \le C\,,$$

where β describes the smoothness degree. Further, we consider so-called supersmooth densities whose Fourier transforms satisfy

$$\int |f_X^{ft}(t)|^2 \exp\big(C_0|t|^\gamma\big)dt \;\le\; C_1\,.$$

Those densities are collected into the class $\mathcal{G}_{C_0C_1\gamma}$. We give the following theorem

Theorem 1 *Let $\|\cdot\|$ denote the $L_2(\mathbb{R})$-norm. As the kernel function K, we choose the sinc kernel with $K^{ft}(t) = \chi_{[-1,1]}(t)$, i.e. the indicator function of the interval $[-1,1]$.*
(a) Take estimator f_X as in (4). Then, we have

$$\sup_{f_X\in\mathcal{F}_{\beta C}} E\|\hat f_X - f_X\|^2 \;=\; O\big(n^{-2\beta/\left[m(2\beta+1)\right]}\big)\,,$$
$$\sup_{f_X\in\mathcal{G}_{C_0C_1\gamma}} E\|\hat f_X - f_X\|^2 \;=\; O\big((\ln n)^{1/\gamma}n^{-1/m}\big)\,,$$

when selecting $h = c_n n^{-1/\left[m(2\beta+1)\right]}$ where $c_n > 0$ is bounded away from both ∞ and 0; and $h = d(\ln n)^{-1/\gamma}$ with a constant $d \le C_0^{1/\gamma}$, respectively.
(b) Assume an arbitrary estimator $\hat f$ of f_X based on the data $Y_1,\ldots,Y_n$. Then, for $\gamma\in(0,1)$ and $\beta > 1/2$, there is a constant $c > 0$ so that

$$\sup_{f_X\in\mathcal{F}_{\beta C}} E\|\hat f - f_X\|^2 \;\ge\; c\cdot n^{-2\beta/\left[m(2\beta+1)\right]}\,,$$
$$\sup_{f_X\in\mathcal{G}_{C_0C_1\gamma}} E\|\hat f - f_X\|^2 \;\ge\; c\cdot(\ln n)^{1/\gamma}n^{-1/m}\,.$$

Hence, we have established rate optimality of our estimation procedure. We notice deterioration of the convergence rate compared to density estimation based on direct data where the well-known rates $n^{-2\beta/(2\beta+1)}$ and $(\ln n)^{1/\gamma}n^{-1}$, resp., occur; they are included into our framework for $m = 1$. The rates become worse when m increases. In fact we have algebraic rates for supersmooth densities contrarily to the expectation stated in Linton & Whang (2002), p. 433. We mention that Theorem 1(a) can be extended to more general kernels K as long as $K^{ft}(t) = 1$ on an open interval around $t = 0$ and K^{ft} is compactly supported.

3. Adaptive estimation

The choice of the bandwidth h as given in Theorem 1 leads to optimal rates; however it requires knowledge of the parameters β, γ, C_0. Therefore our goal is to find a fully data-driven bandwidth selector. In classical density estimation, cross-validation (CV) is a famous procedure for adaptive bandwidth choice. As mentioned in Linton & Whang (2002) there is no straight-forward extent of the underlying theory to aggregated data problems as the estimators are non-linear.

Nevertheless, we can apply CV to estimate the observation density f_Y based on the direct data $Y_1, \ldots, Y_n$. The outcome bandwidth is denoted by $\hat{h}_C$, see Hall & Marron (1987) for the methodology and theory for that problem. In this section we restrict our consideration on densities whose Fourier transforms satisfy

$$C_2|t|^{-\beta-1/2} \leq |f_X^{ft}(t)| \leq C_3|t|^{-\beta-1/2} \quad , \qquad \forall |t| \geq T \tag{5}$$

for some T; and $|f_X^{ft}(t)| \geq |f_X^{ft}(T)|$ for all $|t| \leq T$. Those smoothness assumption is closely related to $f_X \in \mathcal{F}_{\beta C}$ with appropriate constants; indeed, the optimal convergence rates are the same under the corresponding constraints. Therefore, the mean integrated squared error for the estimation of f_Y is minimised by $h = h_0 \sim n^{-1/(2\beta m+m)}$. Surprisingly, those selection rule also minimises the MISE in our aggregated data problem when estimating f_X according to Theorem 1. That inspires us to employ $\hat{h} = \hat{h}_C$ as the bandwidth selector.

The resulting estimator is denoted by $\hat{f}_{X,\hat{h}}$. With respect to the convergence rates, we give a weak individual version. We write const. for a generic positive constant.

Proposition 1 *Assume f_X satisfies (5) where $\beta > 7/2$; and f_X', f_X'' are integrable. We apply the sinc kernel in $\hat{f}_{X,\hat{h}}$. Then, for all $c > 0$, we have*

$$\limsup_{n\to\infty} P\Big(n^{2\beta/\left[m(2\beta+1)\right]} \|\hat{f}_{X,\hat{h}} - f_X\|^2 > c\Big) \leq \textit{const.} \cdot c^{-1}$$

Therefore, the adaptive estimator $\hat{f}_{X,\hat{h}}$ keeps the optimal rates from Theorem 1 under certain circumstances. The case of supersmooth f_X is more difficult to address.

4. Skew densities

When f_X is no longer assumed to be symmetric around zero, its Fourier transform f_X^{ft} is not real-valued. Therefore the inversion procedure becomes more difficult as we have m different complex roots of $f_Y^{ft}(t)$ and its empirical version $\hat{f}_Y^{ft}(t)$. Let $R(t), \varphi(t)$ denote the absolute value and the angle of $f_Y^{ft}(t)$ in the polar representation of complex numbers. We face the problem that the angle is not uniquely defined. Our intention is to specify $\varphi(t)$ so that those functions are continuous for all t and $\varphi(0) = 0$.

We introduce the intervals $I_j = ((j/2-1)\pi, (j/2+1)\pi]$ for $j = 0, \ldots, 3$. Considering that $R(t) \neq 0$, the angle $\varphi(t)$ is uniquely determined by $f_Y^{ft}(t)$ if their images are restricted to I_j for any j. Therefore, we denote the angles within I_j by $\varphi_j(t)$. Setting $\varphi(t) = \varphi_{j(t)}(t)$, we start with $t_0 = 0, j(t_0) = 0$. Then, given t_k, we denote by t_{k+1} the smallest $t > t_k$ where $\varphi_{j(t_k)}(t)$ crosses either $(j(t_k)/2-1/2)\pi$ or $(j(t_k)/2+1/2)\pi$; in the first case, we put $j(t_{k+1}) = \left[j(t_k)-1\right] \mod 4$; in the latter case, we define $j(t_{k+1}) = \left[j(t_k)+1\right] \mod 4$. It follows from there that $\varphi(t_k) = j(t_k)\pi/2$ holds for any k. The sequence $(t_k)_{k>0}$ tends to infinity as, otherwise, the continuity of $f_Y^{ft}(t)$ is violated at the limit of $(t_k)_{k>0}$. Then we define $\varphi(t) = \varphi_{j(\tau(t))}(t)$ with $\tau(t) = \max\{t_k : t_k \leq t\}$ for $t \geq 0$; for $t < 0$ we set $\varphi(t) = -\varphi(-t)$. Then φ is a continuous function on the whole real line with $\varphi(0) = 0$.

When determining an empirical version $\hat{\varphi}(t)$ for $\varphi(t)$, we must consider that $\hat{f}_Y^{ft}(t)$ may have some isolated zeros. Therefore we introduce a parameter $\rho_n > 0$. We realise that the set $\hat{N} = \{t : |\hat{f}_Y^{ft}(t)| \leq \rho_n\}$ may be written as the disjoint union of countably many intervals, $[\tau_j, \tau_{j+1}]$ say. We introduce a function $\tilde{f}_Y^{ft}(t)$ which is equal to $\hat{f}_Y^{ft}(t)$ outside the set $\hat{N}$; while on any interval $[\tau_j, \tau_{j+1}]$ we put $\tilde{f}_Y^{ft}$ equal to the shortest connection between $\big(\tau_j, \hat{f}_Y^{ft}(\tau_j)\big)$ and $\big(\tau_{j+1}, \hat{f}_Y^{ft}(\tau_{j+1})\big)$ under the constraint $|\tilde{f}_Y^{ft}(t)| = \rho_n$ for all $t \in [\tau_j, \tau_{j+1}]$.

Then, define $\hat{R}(t) = |\tilde{f}_Y^{ft}(t)|$ and $\hat{\varphi}(t)$ by applying the procedure for deriving $\varphi(t)$ to $\tilde{f}_Y^{ft}(t)$ instead of $f_Y^{ft}(t)$.

Further, we define the empirical version of $f_X^{ft}(t)$ by

$$\hat{f}_X^{ft}(t) \,=\, \hat{R}^{1/m}(t) \exp\big(it\hat{\varphi}(t)/m\big)\,, \qquad \forall t \in (-T, T)\,, \tag{6}$$

and, accordingly,

$$\hat{f}_X(x) \,=\, \frac{1}{2\pi} \int \exp(-itx) K^{ft}(th) \hat{f}_X^{ft}(t)\, dt\,, \tag{7}$$

while stipulating that K^{ft} is supported on $[-1, 1]$ and $h > 1/T$.

In order to give convergence rates we need more restrictive conditions compared to symmetric densities, namely $\int x^2 f_X(x) dx \leq C_7$ and (5). As an analogue for (5) for supersmooth densities we use

$$C_5 |t|^{(\gamma-1)/2} \exp(-C_4|t|^\gamma) \leq |f_X^{ft}(t)| \leq C_6 |t|^{(\gamma-1)/2} \exp(-C_4|t|^\gamma) \quad , \qquad \forall |t| \geq T\,. \tag{8}$$

Densities satisfying those conditions are collected into the class $\mathcal{F}'_{\beta C_7 C_2 C_3}$, which corresponds to $\mathcal{F}_{\beta C}$. When assuming (8) instead of (5), we call the density class $\mathcal{G}'_{\gamma C_7 C_4 C_5 C_6}$, as the analogue of $\mathcal{G}_{C_0 C_1 \gamma}$.

Proposition 2 *Take estimator $\hat{f}_X$ as defined in (7) and K as in Theorem 1. Choose $\rho_n = \exp(-n)$. Under the constraint $f_X \in \mathcal{F}'_{\beta C_7 C_2 C_3}$, select $h = h_n = [C_h (\ln n)/n]^{1/[m(2\beta+1)]}$ with a constant $C_h > (24m)/C_2^{2m}$. Then, for $\beta > 1/2$, we have*

$$\sup_{f_X \in \mathcal{F}'_{\beta C_7 C_2 C_3}} E\|\hat{f}_X - f_X\|^2 \,=\, O\Big((\ln n/n)^{2\beta/[m(2\beta+1)]}\Big)\,.$$

If $f_X \in \mathcal{G}'_{\gamma C_7 C_4 C_5 C_6}$ and $\max\{1, \gamma\} < m$, select $h = h_n = \Big\{2C_4 m/\big[1 - \nu(\ln\ln n)/\ln n\big]\Big\}^{1/\gamma} \cdot (\ln n)^{-1/\gamma}$ with $\nu \in (1 - m(\gamma-1)/\gamma, m/\gamma]$ to obtain

$$\sup_{f_X \in \mathcal{G}'_{\gamma C_7 C_4 C_5 C_6}} E\|\hat{f}_X - f_X\|^2 \,=\, O\Big((\ln n)^{1/\gamma} n^{-1/m}\Big)\,.$$

Therefore, the rates are kept in the case of $f_X \in \mathcal{F}'_{\beta C_7 C_2 C_3}$ from the symmetric constraint $f_X \in \mathcal{F}_{\beta C}$ (see Theorem 1) up to a logarithmic factor; while, for $f_X \in \mathcal{G}'_{\gamma C_7 C_4 C_5 C_6}$, they are exactly the same as for symmetric $f_X \in \mathcal{G}_{C_0 C_1 \gamma}$ in Theorem 1.

5. Proofs

Proof of Theorem 1: (a) By Parseval's identity and Fubini's theorem, we obtain, for $f_X \in \mathcal{F}_{\beta C}$,

$$E\|\hat{f}_X - f_X\|^2 \,=\, \int |K^{ft}(t/h)|^2 E\big||\hat{f}_Y^{ft}(t)|^{1/m} - |f_X^{ft}(t)|\big|^2 dt \,+\, O(h^{2\beta})\,. \tag{9}$$

In the case of $f_X \in \mathcal{G}_{C_0 C_1 \gamma}$, the bias term in (9) changes from $O(h^{2\beta})$ to $O\big(\exp(-C_0 h^{-\gamma})\big)$.

For the variance term we use the inequalities $|x^{1/m} - y^{1/m}|^m \leq |x-y|$ for $x, y > 0$ and even m as well as $|x^{1/m} - y^{1/m}|^m \leq 2^m|x-y|$ for all x, y in the case of odd m; combined with Jensen's inequality, we obtain

$$E\big||\hat{f}_Y^{ft}(t)|^{1/m} - |f_X^{ft}(t)|\big|^2 \leq \text{const.} \cdot \Big(E\big|\hat{f}_Y^{ft}(t) - f_Y^{ft}(t)\big|^2\Big)^{1/m} = O\big(n^{-1/m}\big),$$

independently of t. Therefore, the mean integrated squared error is bounded above by $O\big(h^{-1}n^{-1/m}, h^{2\beta}\big)$ for $f_X \in \mathcal{F}_{\beta C}$ and $O\big(h^{-1}n^{-1/m}, \exp(-C_0 h^{-\gamma})\big)$ for $f_X \in \mathcal{G}_{C_0 C_1 \gamma}$. Choosing h as stated in the theorem leads to the given rates.

(b) First we consider $f_X \in \mathcal{F}_{\beta C}$. Take f_0 as in (3). Furthermore, we introduce the supersmooth Cauchy density $f_1(x) = \pi^{-1}(1+x^2)^{-1}$ with $f_1^{ft}(t) = \exp(-|t|)$. From there, we construct the following subclass of densities

$$f_{n,\theta}(x) = \frac{1}{2}\big\{f_1(x) + f_0(x)\big\} + \text{const.} \cdot \sum_{2k_n \geq j \geq k_n} \theta_j j^{-\beta-(1/2)} \cos(2jx) f_0(x), \tag{10}$$

where k_n denotes a positive integer still to be determined and const. is sufficiently small; and all $\theta_j \in \{0,1\}$ are i.i.d. random variables with $P(\theta_j = 0) = 1/2$. The corresponding Fourier transforms are given by

$$f_{n,\theta}^{ft}(t) = \frac{1}{2}\big\{f_1^{ft}(t) + f_0^{ft}(t)\big\} + \text{const.} \cdot \sum_{2k_n \geq j \geq k_n} \theta_j j^{-\beta-(1/2)} \big\{f_0^{ft}(t-2j) + f_0^{ft}(t+2j)\big\}.$$

By that we may verify that all $f_{n,\theta}^{ft}$ are non-vanishing and $f_{n,\theta} \in \mathcal{F}_{\beta C}$.

Now assume an arbitrary estimator $\hat{f}(x) = \hat{f}(x; Y_1, \ldots, Y_n)$. In the sequel, we write $\theta_{j,b} = \big(\theta_{k_n}, \ldots, \theta_{j-1}, b, \theta_{j+1}, \ldots, \theta_{2k_n}\big)$ where $b \in \{0,1\}$ and $f_{Y,\theta_{j,b}} = f_{n,\theta_{j,b}}^{*,n}$, i.e. the n-fold self-convolution of $f_{n,\theta_{j,b}}$, which, therefore, denotes the density of each observation Y_j. We consider its mean integrated squared error, using Parseval's identity and Fubini's theorem,

$$\begin{aligned}
\sup_{f_x \in \mathcal{F}_{\beta C}} E\|\hat{f} - f_X\|^2 &\geq \text{const.} \sum_{2k_n \geq j \geq k_n} E_{\theta_j} E_{f_{n,\theta_j}} \int_{2j-1}^{2j+1} \big|\hat{f}^{ft}(t) - f_{n,\theta}^{ft}(t)\big|^2 dt \\
&\geq \text{const.} \sum_{2k_n \geq j \geq k_n} \int_{2j-1}^{2j+1} \big|f_{n,\theta_{j,0}}^{ft}(t) - f_{n,\theta_{j,1}}^{ft}(t)\big|^2 dt \\
&\quad \cdot \int \cdots \int \min\{f_{Y,\theta_{j,0}}(y_1) \cdots f_{Y,\theta_{j,0}}(y_n),\, f_{Y,\theta_{j,1}}(y_1) \cdots f_{Y,\theta_{j,1}}(y_n)\} dy_1 \cdots dy_n \\
&\geq \text{const.} \sum_{2k_n \geq j \geq k_n} j^{-2\beta-1} \geq \text{const.} k_n^{-2\beta},
\end{aligned} \tag{11}$$

if the χ^2-distance between $f_{Y,\theta_{j,0}}$ and $f_{Y,\theta_{j,1}}$ satisfies

$$\chi^2(f_{Y,\theta_{j,0}}, f_{Y,\theta_{j,1}}) = \int \big|f_{Y,\theta_{j,0}}(x) - f_{Y,\theta_{j,1}}(x)\big|^2 \big[f_{Y,\theta_{j,0}}(x)\big]^{-1} dx = O(1/n) \tag{12}$$

holds for all $k_n \leq j \leq 2k_n$. There we have used a result in Fan (1993) which has been developed for the classical deconvolution problem. From the definition of the density subclass, we derive $f_{Y,\theta_{j,0}}(x) \geq 2^{-m} f_1^{*,m}(x)$. That implies

$$f_{Y,\theta_{j,0}}(x) \geq \text{const.}(1+x^2)^{-1},$$

and, from there, we see that (12) may be represented equivalently in the Fourier domain by

$$\int_{2j-1}^{2j+1} \big[\big|g_j(t)\big|^2 + \big|g_j'(t)\big|^2\big] dt = O(1/n), \tag{13}$$

where

$$\begin{aligned}|g_j(t)| &= \left|\left(f_{Y,\theta_{j,0}} - f_{Y,\theta_{j,0}}\right)^{ft}(t)\right| \\ &= \left|\left(\frac{1}{2}f_1^{ft}(t) + j^{-\beta-1/2}f_0^{ft}(t+2j)\right)^m - \left(j^{-\beta-1/2}f_0^{ft}(t+2j)\right)^m\right| \\ &\le \text{ const.} \cdot j^{m(-\beta-1/2)},\end{aligned}$$

while $j \geq k_n$ and $|t| \geq 2j-1$. We can derive the same bound for $|g_j'(t)|$. Then, (13) and, hence, (11) follow when selecting $k_n = \lfloor n^{1/[m(2\beta+1)]} \rfloor$.

In the case $f_X \in \mathcal{G}_{C_0 C_1 \gamma}$, we replace $j^{-\beta-(1/2)}$ by $n^{-1/(2m)}$ and set $k_n = \lfloor (1/2)\cdot\left[\ln n/(mC')\right]^{1/\gamma}\rfloor$ with some constant $C' > C_0$ in (10). Then the proof follows analogously. ■

Proof of Proposition 1: We consider

$$\begin{aligned}&P\left(n^{2\beta/\left[m(2\beta+1)\right]} \cdot \|\hat{f}_{X,\hat{h}} - f_X\|^2 > c\right) \\ &\qquad \le P\left(n^{2\beta/\left[m(2\beta+1)\right]} \int_{|t|\le(dh_0)^{-1}} \left|\hat{f}_{X,\hat{h}}^{ft}(t) - f_X^{ft}(t)\right|^2 dt > c/2\right) \\ &\qquad\quad + P(\hat{h} < dh_0) + P\left(n^{2\beta/\left[m(2\beta+1)\right]}\hat{h}^{2\beta} > c/2\right) \qquad (14)\end{aligned}$$

for some $d \in (0,1)$. The first addend is seen to be bounded above by const. $\cdot\, c^{-1}$ when using Markov's inequality. The second and the third terms are bounded above by $P\left(|\hat{h} - h_0| \geq \text{const.}\cdot h_0\right)$.

Both $n^{3/10}(\hat{h}_0 - h_0)$ and $n^{3/10}(\hat{h} - \hat{h}_0)$ are asymptotically normal distributed, due to Theorem 2.1 in Hall & Marron (1987); where $\hat{h}_0$ is the bandwidth which minimises the integrated squared error; the conditions are satisfied for $\beta > 7/2$, f_X', f_X'' integrable and an appropriate kernel; also note that the kernel used for deriving $\hat{h}_C$ need not coincide with K in estimator $\hat{f}_{X,\hat{h}}$. It follows from there that the second term in (14) converges to zero as $n \to \infty$ so that the proposition follows. ■

Proof of Proposition 2: Note that, for $|t| \leq 1/h$, we have

$$|\hat{f}_Y^{ft}(t) - f_Y^{ft}(t)| \geq |\tilde{f}_Y^{ft}(t) - f_Y^{ft}(t)| - 2\rho_n$$

and hence $|\hat{f}_Y^{ft}(t) - f_Y^{ft}(t)| \geq (1/2)|\tilde{f}_Y^{ft}(t) - f_Y^{ft}(t)|$.

With $\hat{f}_X^{ft}$ as in (6), we derive

$$\begin{aligned}E\left|\hat{f}_X^{ft}(t) - f_X^{ft}(t)\right|^2 &= E\left|\hat{R}^{1/m}(t)\exp\left[i\hat{\varphi}(t)/m\right] - R^{1/m}(t)\exp\left[i\varphi(t)/m\right]\right|^2 \\ &\le E\left|\hat{R}^{1/m}(t) - R^{1/m}(t)\right|^2 + E\hat{R}^{1/m}(t)R^{1/m}(t)\left[1 - \cos\left((\hat{\varphi}(t) - \varphi(t))/m\right)\right] \\ &= O\left(n^{-1/m}\right) + E\hat{R}^{1/m}(t)R^{1/m}(t)\left[1 - \cos\left((\hat{\varphi}(t) - \varphi(t))/m\right)\right]\cdot\chi_{\{|\hat{\varphi}(t)-\varphi(t)|>\pi\}} \\ &\quad + E\hat{R}^{1/m}(t)R^{1/m}(t)\left[1 - \cos\left((\hat{\varphi}(t) - \varphi(t))/m\right)\right]\cdot\chi_{\{|\hat{\varphi}(t)-\varphi(t)|\le\pi\}}\,.\end{aligned}$$

Considering the fact that the function $\left[1-\cos(\cdot/m)\right]/\left[1-\cos(\cdot)\right]$ is bounded on the interval $[-\pi,\pi]$, some standard techniques lead to

$$E\left|\hat{f}_X^{ft}(t) - f_X^{ft}(t)\right|^2 = O\left(n^{-1/m}, \left\{R^2(t)P(|\hat{\varphi}(t) - \varphi(t)| > \pi)\right\}^{1/m}\right).$$

Parseval's identity leads to

$$E\|\hat{f}_X - f_X\|^2 \;=\; O\Big(h^{-1}n^{-1/m}\,,\, h^{2\beta}\,,\, \sup_{|t|\le 1/h}\Big\{P(|\hat{\varphi}(t)-\varphi(t)|>\pi)\Big\}^{1/m}\Big)\,, \tag{15}$$

for $f_X \in \mathcal{F}'_{\beta C_7 C_2 C_3}$; the latter term needs more careful consideration for any $t \in [-1/h, 1/h]$.

As $|\hat{\varphi}(t)-\varphi(t)| > \pi$ implies the existence of at least one $s \in [0,t]$ so that $|\hat{\varphi}(s)-\varphi(s)| = \pi/2$ and hence $\big|\hat{f}_Y^{ft}(s) - f_Y^{ft}(s)\big| \ge (1/2)\cdot|f_Y^{ft}(s)|$. Applying the sequence $(s_j)_j$, $j = 1,\ldots,M-1$ of equidistant points $s_j = jt/M$ where $M = M_n$, we derive that

$$\begin{aligned}P(|\hat{\varphi}(t)-\varphi(t)|>\pi) \;\le\; & P\Big(\exists j = 1,\ldots,M-1\,:\, \big|\hat{f}_Y^{ft}(s_j) - f_Y^{ft}(s_j)\big| \ge (C_2^m/6)\cdot h^{m(\beta+1/2)}\Big)\\ & +\, P\Big(\frac{1}{Mn}\sum_{k=1}^{n}|Y_k| \ge \text{const.}\cdot h^{m(\beta+1/2)}\Big)\\ & +\, \chi_{\{M^{-1}\ge \text{const.}\cdot h^{m(\beta+1/2)}\}}\,. \end{aligned} \tag{16}$$

By Hoeffding's inequality, the first addend in (16) has the upper bound

$$O(M)\cdot\exp\Big(-\frac{1}{12}C_2^{2m}h^{m(2\beta+1)}n\Big).$$

The second addend is bounded above by

$$\begin{aligned}& P\Big(\frac{1}{Mn}\sum_k\big(|Y_k| - E|Y_k|\big) \ge \text{const.}\cdot h^{m(\beta+1/2)}\Big) \,+\, \chi_{\{M^{-1}\ge \text{const.}\cdot h^{m(\beta+1/2)}\}}\\ & \le\; O\big(M^{-2}n^{-1}h^{-m(2\beta+1)}\big) \,+\, \chi_{\{M^{-1}\ge \text{const.}\cdot h^{m(\beta+1/2)}\}}\,.\end{aligned}$$

The latter term above as well as the third addend in (16) vanish when selecting $M = C_M h^{-m(\beta+1/2)}$ with an appropriate constant $C_M > 0$.

In the case $f_X \in \mathcal{G}'_{\gamma C_7 C_4 C_5 C_6}$, the proof follows by replacing C_2 by C_5, $h^{2\beta}$ by $\exp(-2C_4 h^{-\gamma})$ in (15) and $h^{m(\beta+1/2)}$ by $\exp(-C_4 m h^{-\gamma})\cdot h^{-m(\gamma-1)/2}$ in the first two lines of (16) and, accordingly, in the sequel.

Then the specific choice of h as stated in the theorem leads to the desired rates in the view of (15). ■

References

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

Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21**, 600–610.

Hall, P. & Marron, J.S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probab. Th. Rel. Fields* **74**, 567–581.

Horowitz, J.L. & Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Rev. Econom. Stud.* **63**, 145–168.

Linton, O. & Whang, Y.L. (2002). Nonparametric estimation with aggregated data. *Econometric Theory* **18**, 420–468.

Machado, J.A.F. & Santos Silva, J.M.C. (2006). A note on identification with averaged data. *Econometric Theory* **22**, 537–541.

Schick, A. & Wefelmeyer, W. (2004). Root n consistent density estimation for sums of independent random variables. *J. Nonparam. Statist.* **16**, 925–935.

Stefanski, L.A. & Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169–184.

van Es, B., Gugushvili, S. & Spreij, P. (2005). A kernel type nonparametric density estimator for decompounding. available online under *http://arxiv.org/pdf/math.ST/0505355 v2*

Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806–830.

Alexander Meister
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: meistear@mathematik.uni-stuttgart.de

Erschienene Preprints ab Nummer 2004/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2004/001 *Walk, H.:* Strong Laws of Large Numbers by Elementary Tauberian Arguments.

2004/002 *Hesse, C.H., Meister, A.:* Optimal Iterative Density Deconvolution: Upper and Lower Bounds.

2004/003 *Meister, A.:* On the effect of misspecifying the error density in a deconvolution problem.

2004/004 *Meister, A.:* Deconvolution Density Estimation with a Testing Procedure for the Error Distribution.

2004/005 *Efendiev, M.A., Wendland, W.L.:* On the degree of quasiruled Fredholm maps and nonlinear Riemann-Hilbert problems.

2004/006 *Dippon, J., Walk, H.:* An elementary analytical proof of Blackwell's renewal theorem.

2004/007 *Mielke, A., Zelik, S.:* Infinite-dimensional hyperbolic sets and spatio-temporal chaos in reaction-diffusion systems in $\mathbb{R}^n$.

2004/008 *Exner, P., Linde, H., Weidl T.:* Lieb-Thirring inequalities for geometrically induced bound states.

2004/009 *Ekholm, T., Kovarik, H.:* Stability of the magnetic Schrödinger operator in a waveguide.

2004/010 *Dillen, F., Kühnel, W.:* Total curvature of complete submanifolds of Euclidean space.

2004/011 *Afendikov, A.L., Mielke, A.:* Dynamical properties of spatially non-decaying 2D Navier-Stokes flows with Kolmogorov forcing in an infinite strip.

2004/012 *Pöschel, J.:* Hill's potentials in weighted Sobolev spaces and their spectral gaps.

2004/013 *Dippon, J., Walk, H.:* Simplified analytical proof of Blackwell's renewal theorem.

2004/014 *Kühnel, W.:* Tight embeddings of simply connected 4-manifolds.

2004/015 *Kühnel, W., Steller, M.:* On closed Weingarten surfaces.

2004/016 *Leitner, F.:* On pseudo-Hermitian Einstein spaces.

2004/017 *Förster, C., Östensson, J.:* Lieb-Thirring Inequalities for Higher Order Differential Operators.

2005/001 *Mielke, A.; Schmid, F.:* Vortex pinning in super-conductivity as a rate-independent model

2005/002 *Kimmerle, W.; Luca, F., Raggi-Cárdenas, A.G.:* Irreducible Components of the Burnside Ring

2005/003 *Höfert, C.; Kimmerle, W.:* On Torsion Units of Integral Group Rings of Groups of Small Order

2005/004 *Röhrl, N.:* A Least Squares Functional for Solving Inverse Sturm-Liouville Problems

2005/005 *Borisov, D.; Ekholm, T; Kovarik, H.:* Spectrum of the Magnetic Schrödinger Operator in a Waveguide with Combined Boundary Conditions

2005/006 *Zelik, S.:* Spatially nondecaying solutions of 2D Navier-Stokes equation in a strip

2005/007 *Meister, A.:* Deconvolving compactly supported densities

2005/008 *Förster, C., Weidl, T.:* Trapped modes for an elastic strip with perturbation of the material properties

2006/001 *Dippon, J., Schiemert, D.:* Stochastic differential equations driven by Gaussian processes with dependent increments

2006/002 *Lesky, P.A.:* Orthogonale Polynomlösungen von Differenzengleichungen vierter Ordnung

2006/003 *Dippon, J., Schiemert, D.:* Option Pricing in a Black-Scholes Market with Memory

2006/004 *Banchoff, T., Kühnel, W.:* Tight polyhedral models of isoparametric families, and PL-taut submanifolds

2006/005 *Walk, H.:* A universal strong law of large numbers for conditional expectations via nearest neighbors

2006/006 *Dippon, J., Winter, S.:* Smoothing spline regression estimates for randomly right censored data

2006/007 *Walk, H.:* Almost sure Cesàro and Euler summability of sequences of dependent random variables

2006/008 *Meister, A.:* Optimal convergence rates for density estimation from grouped data

2006/009 *Förster, C.:* Trapped modes for the elastic plate with a perturbation of Young's modulus