



Optimal convergence rates for density estimation from grouped data

Alexander Meister

► To cite this version:

Alexander Meister. Optimal convergence rates for density estimation from grouped data. *Statistics and Probability Letters*, 2009, 77 (11), pp.1091. 10.1016/j.spl.2007.01.013 . hal-00565445

HAL Id: hal-00565445

<https://hal.science/hal-00565445>

Submitted on 13 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Optimal convergence rates for density estimation
from grouped data

Alexander Meister

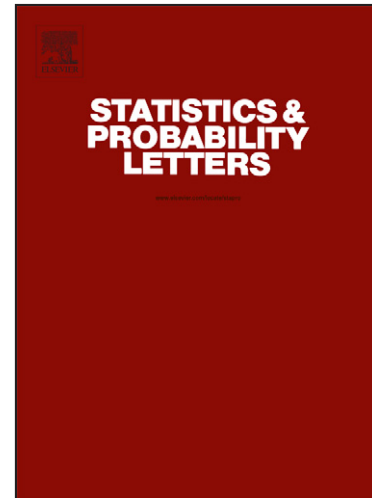
PII: S0167-7152(07)00047-8
DOI: doi:10.1016/j.spl.2007.01.013
Reference: STAPRO 4580

To appear in: *Statistics & Probability Letters*

Received date: 7 September 2006
Revised date: 1 December 2006
Accepted date: 17 January 2007

Cite this article as: Alexander Meister, Optimal convergence rates for density estimation from grouped data, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.01.013

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

Optimal convergence rates for density estimation from grouped data

Alexander Meister¹

*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 extension to skew densities.

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

MSC: 62G07

¹Alexander Meister

Institut für Stochastik und Anwendungen,
Universität Stuttgart
Pfaffenwaldring 57
D-70569 Stuttgart, Germany

email: meister@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.

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

$$Y_i = \sum_{j=1}^m X_{ij}, \quad (1)$$

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. Consider the moving average process $(Z_k)_k$ with $Z_k = \sum_{j=k}^{m+k-1} X_j$ with i.i.d. X_1, X_2, \dots having the density f_X . Then, assume that only the data Z_{1+mk} are observed for integer $k \geq 0$.

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 extension 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), Fan (1991) among many others). See the recent note of Machado & Santos Silva (2006) for a study on parameter identification from averaged data. A nonparametric approach to density estimation from aggregated observations is given by Linton & Whang (2002), where the authors consider a more complicated model with an additional independent error component for each data group. Pointwise asymptotic normality of the authors' estimator is studied at certain convergence rates. 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. Belomestny (2003) studies a procedure for estimating the density of a component of a moving average process (MA(1)) where the

other component is scaled with some factor ϑ with $|\vartheta| < 1$. 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. With respect to the characteristic functions f_X^{ft} and f_Y^{ft} , we have

$$f_Y^{ft}(t) = [f_X^{ft}(t)]^m.$$

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

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

The necessity of a condition as (2) to ensure identifiability of f_X is shown in the following:

We define the density

$$f_0(x) = (1 - \cos(x))/(\pi x^2), \quad (3)$$

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

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

hence we have $[f_+^{ft}(t)]^2 = [f_-^{ft}(t)]^2$ 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,

$$\hat{f}_X(x) = \frac{1}{2\pi} \int \exp(-itx) K^{ft}(th) |\hat{f}_Y^{ft}(t)|^{1/m} dt, \quad (4)$$

where K denotes a square-integrable kernel function with $K^{ft}(0) = 1$ 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 $\mathcal{F}_{\beta C}$ of even densities which satisfy (2) and

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

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

$$\int |f_X^{ft}(t)|^2 \exp(C_0|t|^\gamma) dt \leq C_1.$$

Those densities are collected into the class $\mathcal{G}_{C_0 C_1 \gamma}$. 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*

$$\begin{aligned} \sup_{f_X \in \mathcal{F}_{\beta C}} E\|\hat{f}_X - f_X\|^2 &= O(n^{-2\beta/\lceil m(2\beta+1) \rceil}), \\ \sup_{f_X \in \mathcal{G}_{C_0 C_1 \gamma}} E\|\hat{f}_X - f_X\|^2 &= O((\ln n)^{1/\gamma} n^{-1/m}), \end{aligned}$$

when selecting $h = c_n n^{-1/\lceil m(2\beta+1) \rceil}$ where $c_n > 0$ is bounded away from both ∞ and 0; and $h = d(\ln n)^{-1/\gamma}$ with a constant $d \leq C_0^{1/\gamma}$, respectively.

(b) *Assume an arbitrary estimator \hat{f} of f_X based on the data Y_1, \dots, Y_n . Then, for $\gamma \in (0, 1)$ and $\beta > 1/2$, C, C_1 sufficiently large, there is a constant $c > 0$ so that*

$$\begin{aligned} \sup_{f_X \in \mathcal{F}_{\beta C}} E\|\hat{f} - f_X\|^2 &\geq c \cdot n^{-2\beta/\lceil m(2\beta+1) \rceil}, \\ \sup_{f_X \in \mathcal{G}_{C_0 C_1 \gamma}} E\|\hat{f} - f_X\|^2 &\geq c \cdot (\ln n)^{1/\gamma} n^{-1/m}. \end{aligned}$$

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. 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 extension 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, \dots, 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 to densities whose Fourier transforms satisfy

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

for some T ; and $|f_X^{ft}(t)| \geq |f_X^{ft}(T)|$ for all $|t| \leq T$. This 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 $h = h_0 \sim n^{-1/(2\beta+m)}$. Surprisingly, this 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*

$$\limsup_{n \rightarrow \infty} P\left(n^{2\beta/\lceil m(2\beta+1) \rceil} \|\hat{f}_{X,\hat{h}} - f_X\|^2 > c\right) \leq \text{const.} \cdot c^{-1}$$

Therefore, the adaptive estimator $\hat{f}_{X,h}$ 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 integer j . Assuming that $R(t) \neq 0$, for all t , the angle $\varphi(t)$ is uniquely determined by $f_Y^{ft}(t)$ if its image is restricted to I_j for any j . Therefore, we denote the angle 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 $j(t_{k+1}) = j(t_k) - 1$; in the latter case, we define $j(t_{k+1}) = j(t_k) + 1$; otherwise, put $t_{k+1} = \infty$. 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 $\rho_n \in (0, 1)$. 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 $(\tau_j, \hat{f}_Y^{ft}(\tau_j))$ and $(\tau_{j+1}, \hat{f}_Y^{ft}(\tau_{j+1}))$ 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

$$\hat{f}_X^{ft}(t) = \hat{R}^{1/m}(t) \exp(it\hat{\varphi}(t)/m), \quad (6)$$

and, accordingly,

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

while stipulating that K^{ft} is supported on $[-1, 1]$.

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

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

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 $C_h > (576m)/C_2^{2m}$. Then, for $\beta > 1/2$, we have*

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

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

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

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 $f_X \in \mathcal{F}_{\beta C}$,

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

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 $O(\exp(-C_0 h^{-\gamma}))$.

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

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

independently of t . Therefore, the mean integrated squared error is bounded above by $O(h^{-1} n^{-1/m}, h^{2\beta})$ for $f_X \in \mathcal{F}_{\beta C}$ and $O(h^{-1} n^{-1/m}, \exp(-C_0 h^{-\gamma}))$ 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

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

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

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

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 $\hat{f}(x) = \hat{f}(x; Y_1, \dots, Y_n)$. In the sequel, we write $\theta_{j,b} = (\theta_{k_n}, \dots, \theta_{j-1}, b, \theta_{j+1}, \dots, \theta_{2k_n})$ 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,

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

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

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

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. That method was influenced by Donoho & Liu (1991a,b). 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

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

where

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

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 $k_n = \lfloor (1/4) \cdot [\ln n / (mC')]^{1/\gamma} \rfloor$ with some constant $C' > C_0$ in (10). Then the proof follows analogously. ■

Proof of Proposition 1: We consider

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

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

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 and the third term in (14) converge to zero as $n \rightarrow \infty$ so that the proposition follows. \blacksquare

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

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

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

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

Parseval's identity leads to

$$E\|\hat{f}_X - f_X\|^2 = O\left(h^{-1}n^{-1/m}, h^{2\beta}, \sup_{|t| \leq 1/h} \left\{P(|\hat{\varphi}(t) - \varphi(t)| > \pi)\right\}^{1/m}\right), \quad (15)$$

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

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

$$\begin{aligned} P(|\hat{\varphi}(t) - \varphi(t)| > \pi) &\leq P\left(\exists j = 1, \dots, M-1 : |\hat{f}_Y^{ft}(s_j) - f_Y^{ft}(s_j)| \geq (C_2^m/6) \cdot h^{m(\beta+1/2)}\right) \\ &\quad + P\left(\frac{1}{Mnh} \sum_{k=1}^n |Y_k| \geq \text{const.} \cdot h^{m(\beta+1/2)}\right) \\ &\quad + \chi_{\{h^{-1}M^{-1} \geq \text{const.} \cdot h^{m(\beta+1/2)}\}}. \end{aligned} \quad (16)$$

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

$$O(M) \cdot \exp\left(-\frac{1}{288} C_2^{2m} h^{m(2\beta+1)} n\right).$$

By Markov's inequality, the second addend is bounded above by

$$\chi_{\{h^{-1}M^{-1} \geq \text{const.} \cdot h^{m(\beta+1/2)}\}} + O(1/n)$$

The term above as well as the third addend in (16) are negligible when selecting $M = C_M h^{-m(\beta+1/2)-1}$ 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

- Belomestny, D. (2003). Rates of convergence for constrained deconvolution problem. Preprint *arXiv math.ST/0306237 v1*.

- Carroll, R.J. & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186.
- Donoho, D.L. & Liu, R.C. (1991a). Geometrizing rate of convergence II. *Ann. Statist.* **19**, 633–667.
- Donoho, D.L. & Liu, R.C. (1991b). Geometrizing rate of convergence III. *Ann. Statist.* **19**, 668–701.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
- 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. Preprint *arXiv math.ST/0505355 v2*
- Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806–830.