doi: 10.1111/j.1467-9469.2006.00517.x

Integrated Square Error Asymptotics for Supersmooth Deconvolution

HAJO HOLZMANN and LEIF BOYSEN

Institut für Mathematische Stochastik, University of Göttingen

ABSTRACT. We derive the asymptotic distribution of the integrated square error of a deconvolution kernel density estimator in supersmooth deconvolution problems. Surprisingly, in contrast to direct density estimation as well as ordinary smooth deconvolution density estimation, the asymptotic distribution is no longer a normal distribution but is given by a normalized chi-squared distribution with 2 d.f. A simulation study shows that the speed of convergence to the asymptotic law is reasonably fast.

Key words: deconvolution, degenerate U-statistic, integrated square error, inverse problems, non-parametric density estimation

1. Introduction

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) real-valued observations with density g. The kernel estimator $\hat{g}_n$ of g with kernel K and bandwidth $h > 0$, introduced by Rosenblatt (1956) and Parzen (1962), is given by

$$\hat{g}_n(x) = \frac{1}{n} \sum_{k=1}^{n} K_h(x - X_k),$$

where $K_h(x) = K(x/h)/h$ and K is a function integrating to one. Properties of $\hat{g}_n$ are very well developed. In particular, for $n \to \infty$ and $h \to 0$, under some additional assumptions one has asymptotic normality both for the pointwise error

$$\sqrt{nh}\left(\hat{g}_n(x) - K_h * g(x)\right) \xrightarrow{\mathcal{L}} N\left(0, g(x) \int_{\mathbb{R}} K^2(t)\,\mathrm{d}t\right), \tag{1}$$

as well as for the integrated square error (cf. Bickel & Rosenblatt, 1973; Hall, 1984a)

$$n\sqrt{h} \int_{\mathbb{R}} \left(\hat{g}_n(x) - K_h * g(x)\right)^2 \mathrm{d}x - \int K^2(t)\,\mathrm{d}t/h^{1/2} \xrightarrow{\mathcal{L}} N\left(0, \tau^2\right), \tag{2}$$

where

$$\tau^2 = \int_{\mathbb{R}} (K * K)^2(x)\,\mathrm{d}x \int_{\mathbb{R}} g^2(x)\,\mathrm{d}x.$$

In (1) and (2) we use $K_h * g(x) = E\hat{g}_n(x)$ instead of $g(x)$ so as to avoid consideration of terms involving the bias $K_h * g - g$. However (1) and (2) continue to hold true with $K_h * g$ replaced by g, if for example, the bias is corrected by undersmoothing (cf. Bickel & Rosenblatt, 1973).

Often, the observations X_i are only noisy versions of the random variables Z_i of interest, i.e. $X_i = Z_i + \epsilon_i$, where ϵ_i and Z_i are independent, the errors ϵ_i have known density ψ and the Z_i have density f. Note that $g = f * \psi$. Estimating f from the observations X_i is therefore called the deconvolution problem.

To fix the notation, the Fourier transform of f is given by $\Phi_f(t) = \int_{\mathbb{R}} f(x)\,\mathrm{e}^{itx}\,\mathrm{d}x$. Under the assumption $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}$, a standard estimator of f is the kernel deconvolution density estimator

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-itx} \Phi_K(ht) \frac{\hat{\Phi}_n(t)}{\Phi_\psi(t)} \, \mathrm{d}t.$$

Here K is a kernel function such that Φ_K has compact support, $h > 0$ is a smoothing parameter called bandwidth and $\hat{\Phi}_n(t) = 1/n \sum_k \mathrm{e}^{itX_k}$ is the empirical characteristic function of $X_1, \ldots, X_n$. For $\hat{f}_n(x)$ we have that

$$E\hat{f}_n = K_h * f,$$

i.e. the bias is equal to that of an ordinary kernel estimator and hence does not depend on the error structure ψ. Kernel deconvolution type estimators have been studied by many authors, we mention Carroll & Hall (1988), Stefanski & Carroll (1990), Zhang (1990), Fan (1991a, b) and van Es & Uh (2004, 2005).

It turns out that the deconvolution problem depends sensitively on the Fourier transform Φ_ψ of the error density ψ. If

$$\Phi_\psi(t) \sim C_\psi t^{-\beta}, \quad t \to \infty,$$

for some $\beta > 0$ and $C_\psi \in \mathbb{C}$, the error density is called ordinary smooth. In this case the optimal rate of convergence for estimating f is of polynomial order, and is achieved by $\hat{f}_n$. Further, for $\hat{f}_n$ one has asymptotic normality both of the pointwise error (cf. Fan, 1991b) as well as of the integrated square error (Piterbarg & Penskaya, 1993; Holzmann *et al.*, 2006), in analogy to the statements (1) and (2).

In this paper, we suppose that the error distribution is supersmooth. More precisely, we assume that

$$\Phi_\psi(t) \sim C_\psi |t|^{\lambda_0} \, \mathrm{e}^{-|t|^\lambda/\mu}, \quad |t| \to \infty, \tag{3}$$

for $\lambda > 1$, $\mu > 0$ and $\lambda_0, C_\psi \in \mathbb{R}$, and that $\Phi_\psi(t) \neq 0$ for all t. Note that (3) includes the particularly important case of a normal error distribution, which is very popular among practitioners and is often used in parametric deconvolution and errors in variables models (see, e.g. Bickel & Ritov, 1987). However, it excludes the Cauchy and all other distributions for which the tail of the characteristic function decreases more slowly than $|t|^{\lambda_0} \, \mathrm{e}^{-|t|/\mu}$, $\lambda_0 \in \mathbb{R}$, $\mu > 0$, in particular the t-distribution (cf. Kotz & Nadarajah, 2004).

The optimal rate of convergence for estimating f in supersmooth deconvolution problems, which is once more achieved by $\hat{f}_n$, is only logarithmic (cf. Fan, 1991a). Recently, van Es & Uh (2004, 2005) obtained the asymptotic distribution of the pointwise error with an explicit rate for the variance. They showed that as $n \to \infty$ and $h \to 0$, under assumption 2 of section 2 (for $\alpha = 0$, $A = 1$) and $EX_1^2 < \infty$,

$$\frac{\sqrt{n}}{h^{\lambda + \lambda_0 - 1} \, \mathrm{e}^{1/(\mu h^\lambda)}} (\hat{f}_n(x) - K_h * f(x)) \xrightarrow{\mathcal{L}} N\left(0, \frac{\mu^2}{2\pi^2 \lambda^2 C_\psi^2}\right). \tag{4}$$

Although the asymptotic distribution remains normal, when comparing (4) with (1), van Es & Uh (2005) noticed two differences. First, as was to be expected, the normalization now reflects the super smoothness of the deconvolution problem. Secondly, in contrast to (1) and also to the ordinary smooth deconvolution problem, the asymptotic variance in (4) does no longer depend on f and x, but only depends on the error density ψ through λ, λ_0, μ and C_ψ.

In this paper, we study the asymptotic distribution of the integrated square error

$$T_n = \int_{\mathbb{R}} (\hat{f}_n - K_h * f)^2(x) \, \mathrm{d}x$$

in supersmooth deconvolution problems (3). This has applications to constructing goodness-of-fit tests for the density f (cf. Bickel & Rosenblatt, 1973) or to assessing the variability of cross-validation when used as an automatic selector for the bandwidth (cf. Hall, 1984b).

As was already indicated by the asymptotics of the pointwise error, the situation will be substantially different from the none-noisy case as well as from the ordinary smooth case. In fact, it turns out that the limit distribution of T_n is no longer normal, but equal to a normalized chi-squared distribution with 2 d.f. This result is derived in section 2. It disproves a claim by Butucea (2004) who asserted asymptotic normality for T_n for general $\lambda > 0$. Our proof relies on the limit theory for degenerate U-statistics with fixed kernel as well as on an approximation technique for triangular arrays of degenerate U-statistics, which might be of some independent interest.

In section 3, we conduct a small simulation experiment, in which we examine the speed of convergence to the asymptotic law. It turns out that the asymptotic approximation is quite reasonable even for small sample sizes.

Finally, let us remark that the deconvolution problem can be studied within the general framework of statistical inverse problems (cf. Mair & Ruymgaart, 1996). Therefore, it is tempting to conjecture that the integrated square error of regularized inverse estimators in other severely ill-posed inverse problems, e.g. the heat equation, may also not be asymptotically normally distributed.

2. Integrated square error asymptotics

To derive the asymptotic distribution of T_n, we need the following assumptions.

Assumption 1

The density f is square-integrable. Further, $EX_1^2 < \infty$.

Assumption 2

The Fourier transform Φ_K of the kernel K is real-valued, symmetric and supported on $[-1, 1]$. Moreover $\Phi_K(0) = 1$ and there exist $A > 0, \alpha \geq 0$ such that

$$\Phi_K(1-t) = At^{\alpha} + o(t^{\alpha}), \quad t \searrow 0. \tag{5}$$

Examples for kernel functions satisfying assumption 2 are the sinc kernel $K(x) = \sin(x)/(\pi x)$, for which $A = 1$ and $\alpha = 0$, and the kernel with Fourier transform $\Phi_K(t) = (1-t^2)^3$ used in Fan (1992), for which $\alpha = 3$ and $A = 8$.

For simplicity, let us first consider the case of normal deconvolution, where the characteristic function of the error variable is given by

$$\Phi_{\psi}(t) = e^{-t^2/2}.$$

Theorem 1

Under the assumptions 1 and 2, for $n \to \infty$ and $h \to 0$ we have for the integrated square error in the normal deconvolution problem that

$$\frac{n 2^{1+2\alpha}\pi}{h^{1+4\alpha}\exp(1/h^2)A^2\Gamma(2\alpha+1)} T_n \xrightarrow{\mathcal{L}} \frac{(Y_1^2 + Y_2^2)}{2}, \tag{6}$$

where Y_1 and Y_2 are independent standard normal random variables and $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution.

There are several striking differences when comparing (6) with (2), the case of direct density estimation. The rates once more reflect the supersmoothness of the normal deconvolution problem. Further, the rate of the standard deviation of T_n is equal to that of its expectation, and in fact from the proof it follows that their quotient tends to 1. Therefore, when standardizing T_n, we do not get a negative drift term as in (2) but asymptotically just subtract one. This directly implies that the asymptotic distribution of the standardized version of the non-negative random variables T_n, being asymptotically bounded from below by -1, cannot be normal. This being established, it is quite intuitive that the asymptotic law should be chi-squared. Note that in (6) we do not centre T_n and therefore get an asymptotic expectation of 1.

Before giving the proof of theorem 1, it might be helpful to present an outline of its main steps. In the first step, we compute the asymptotic expectation and variance of T_n and show that it can be written asymptotically in form of degenerate U-statistics based on a triangular array of random variables. Here, we extend arguments used in van Es & Uh (2005). In the second step we show that these degenerate U-statistics are asymptotically equivalent to U-statistics based on a sequence of i.i.d. observations. Finally, an application of the limit theory for degenerate U-statistics finishes the proof.

Proof of theorem 1

Using Parseval's equation we compute

$$\begin{aligned}
T_n &= \frac{1}{2\pi}\int_{\mathbb{R}} |\Phi_K(ht)|^2\, e^{t^2}\, |\hat{\Phi}_n(t) - \Phi_g(t)|^2\, dt\\
&= \frac{1}{2\pi n^2}\sum_{k=1}^{n}\int_{\mathbb{R}} |\Phi_K(ht)|^2\, e^{t^2}\, |e^{iX_k t} - \Phi_g(t)|^2\, dt\\
&\quad + \frac{1}{\pi n^2}\sum_{1\le j<k\le n} \Re\left(\int_{\mathbb{R}} |\Phi_K(ht)|^2\, e^{t^2}\left(e^{iX_k t} - \Phi_g(t)\right)\overline{\left(e^{iX_j t} - \Phi_g(t)\right)}\, dt\right)\\
&= S_1 + S_2.
\end{aligned} \tag{7}$$

First consider S_1. We have

$$\begin{aligned}
ES_1 &= \frac{1}{2\pi n}\int_{\mathbb{R}} |\Phi_K(ht)|^2\, e^{t^2}(1 - |\Phi_g(t)|^2)\, dt\\
&= \frac{1}{2\pi nh}\int_{-1}^{1} |\Phi_K(u)|^2\, e^{u^2/h^2}\, du - \frac{1}{2\pi n}\int_{\mathbb{R}} |\Phi_K(ht)|^2 |\Phi_f(t)|^2\, dt\\
&= \frac{1}{\pi nh}\int_0^1 |\Phi_K(u)|^2\, e^{u^2/h^2}\, du + O\left(\frac{1}{n}\right),
\end{aligned}$$

as $\Phi_f \in L_2$ by assumption 1. Note that from assumption 2, the kernel function satisfies $\Phi_K^2(1-t) = A^2 t^{2\alpha} + o(t^{2\alpha})$ as $t \searrow 0$. From lemma 5 in van Es & Uh (2005), it follows that

$$\int_0^1 |\Phi_K(u)|^2\, e^{u^2/h^2}\, du \sim A^2 h^{2+4\alpha} \exp(1/h^2)\frac{\Gamma(2\alpha+1)}{2^{1+2\alpha}} \tag{8}$$

and

$$\int_0^1 (1-u)|\Phi_K(u)|^2\, e^{u^2/h^2}\, du \sim A^2 h^{4+4\alpha} \exp(1/h^2)\frac{\Gamma(2+2\alpha)}{2^{2+2\alpha}}, \tag{9}$$

where $\Gamma(s)$ denotes the Gamma function. From (8), we get that

$$ES_1 \sim \frac{A^2 h^{1+4\alpha}\exp(1/h^2)\Gamma(2\alpha+1)}{2^{1+2\alpha}\pi n}. \tag{10}$$

The variance of S_1 can be bounded by

$$\operatorname{var} S_1 \leq \frac{1}{4\pi^2 n^4}\sum_{k=1}^{n} E\left(\int_{\mathbb{R}} |\Phi_K(ht)|^2\, \mathrm{e}^{t^2}\, |\mathrm{e}^{iX_k t} - \Phi_g(t)|^2\, \mathrm{d}t\right)^2 \leq C\frac{h^{2+8\alpha}\,\mathrm{e}^{2/h^2}}{n^3}$$

for some $C>0$, by bounding $|\mathrm{e}^{iX_k t} - \Phi_g(t)|^2 \leq 4$ and using (8). Therefore

$$S_1 = ES_1 + O_P\left(h^{1+4\alpha}\mathrm{e}^{1/h^2} n^{-3/2}\right). \tag{11}$$

Next consider a term in the second sum S_2. Using $\Re(\Phi_g)(-t) = \Re(\Phi_g)(t)$ and $\Im(\Phi_g)(-t) = -\Im(\Phi_g)(t)$, we compute

$$\begin{aligned}
&\Re\left(\int_{\mathbb{R}} |\Phi_K(ht)|^2\, \mathrm{e}^{t^2} (\mathrm{e}^{iX_k t} - \Phi_g(t))\overline{(\mathrm{e}^{iX_j t} - \Phi_g(t))}\, \mathrm{d}t\right)\\
&= \int_{\mathbb{R}} |\Phi_K(ht)|^2\, \mathrm{e}^{t^2} \big((\cos(tX_k) - \Re(\Phi_g)(t))(\cos(tX_j) - \Re(\Phi_g)(t))\\
&\quad + (\sin(tX_k) - \Im(\Phi_g)(t))(\sin(tX_j) - \Im(\Phi_g)(t))\big)\, \mathrm{d}t\\
&= \frac{2}{h}\int_0^1 |\Phi_K(u)|^2\, \mathrm{e}^{u^2/h^2}\left(\left(\cos\left(\frac{uX_k}{h}\right) - \Re(\Phi_g)\left(\frac{u}{h}\right)\right)\left(\cos\left(\frac{uX_j}{h}\right) - \Re(\Phi_g)\left(\frac{u}{h}\right)\right)\right.\\
&\quad \left. + \left(\sin\left(\frac{uX_k}{h}\right) - \Im(\Phi_g)\left(\frac{u}{h}\right)\right)\left(\sin\left(\frac{uX_j}{h}\right) - \Im(\Phi_g)\left(\frac{u}{h}\right)\right)\right)\mathrm{d}u.
\end{aligned}$$

Let

$$\begin{aligned}
S_{2,1} &= \frac{2}{\pi n^2 h}\sum_{1\leq j<k\leq n}\int_0^1 |\Phi_K(u)|^2\, \mathrm{e}^{u^2/h^2}\\
&\quad \times\left[\left(\cos\left(\frac{uX_k}{h}\right) - \Re\left(\Phi_g\right)\left(\frac{u}{h}\right)\right)\left(\cos\left(\frac{uX_j}{h}\right) - \Re\left(\Phi_g\right)\left(\frac{u}{h}\right)\right)\right]\mathrm{d}u,\\
\tilde{S}_{2,1} &= \frac{2}{\pi n^2 h}\int_0^1 |\Phi_K(u)|^2\, \mathrm{e}^{u^2/h^2}\, \mathrm{d}u\\
&\quad \times\sum_{1\leq j<k\leq n}\left[\left(\cos\left(\frac{X_k}{h}\right) - \Re\left(\Phi_g\right)\left(\frac{1}{h}\right)\right)\left(\cos\left(\frac{X_j}{h}\right) - \Re\left(\Phi_g\right)\left(\frac{1}{h}\right)\right)\right],
\end{aligned} \tag{12}$$

and

$$\begin{aligned}
S_{2,2} &= \frac{2}{\pi n^2 h}\sum_{1\leq j<k\leq n}\int_0^1 |\Phi_K(u)|^2\, \mathrm{e}^{u^2/h^2}\\
&\quad \times\left[\left(\sin\left(\frac{uX_k}{h}\right) - \Im\left(\Phi_g\right)\left(\frac{u}{h}\right)\right)\left(\sin\left(\frac{uX_j}{h}\right) - \Im\left(\Phi_g\right)\left(\frac{u}{h}\right)\right)\right]\mathrm{d}u,\\
\tilde{S}_{2,2} &= \frac{2}{\pi n^2 h}\int_0^1 |\Phi_K(u)|^2\, \mathrm{e}^{u^2/h^2}\, \mathrm{d}u\\
&\quad \times\sum_{1\leq j<k\leq n}\left[\left(\sin\left(\frac{X_k}{h}\right) - \Im\left(\Phi_g\right)\left(\frac{1}{h}\right)\right)\left(\sin\left(\frac{X_j}{h}\right) - \Im\left(\Phi_g\right)\left(\frac{1}{h}\right)\right)\right].
\end{aligned} \tag{13}$$

We will show that

$$S_{2,i} - \tilde{S}_{2,i} = O_P(h^{2+4\alpha}\,\mathrm{e}^{1/h^2} n^{-1}), \quad i = 1, 2. \tag{14}$$

Consider (14) for $i=2$, the case $i=1$ is similar. We have

$$\sin\left(\frac{uX_k}{h}\right)-\sin\left(\frac{X_k}{h}\right)=2\cos\big((u+1)X_k/(2h)\big)\sin\left(\frac{(u-1)X_k}{(2h)}\right),$$

therefore for $u\in[0,1]$, using $|\sin x|\le|x|$,

$$\left|\sin\left(\frac{uX_k}{h}\right)-\sin\left(\frac{X_k}{h}\right)\right|\le(1-u)|X_k|/h,$$

and

$$\left|\sin\left(\frac{uX_k}{h}\right)\sin\left(\frac{uX_j}{h}\right)-\sin\left(\frac{X_k}{h}\right)\sin\left(\frac{X_j}{h}\right)\right|\le(1-u)\frac{(|X_k|+|X_j|)}{h}. \tag{15}$$

Using $|\Phi_f(t)|\le 1$, we estimate for $0\le u\le 1$,

$$\begin{aligned}&\left|\Im(\Phi_g)^2\left(\frac{1}{h}\right)-\Im(\Phi_g)^2\left(\frac{u}{h}\right)-\Im(\Phi_g)\left(\frac{1}{h}\right)\left(\sin\left(\frac{X_k}{h}\right)+\sin\left(\frac{X_j}{h}\right)\right)\right.\\&\qquad\left.+\Im(\Phi_g)\left(\frac{u}{h}\right)\left(\sin\left(\frac{uX_k}{h}\right)+\sin\left(\frac{uX_j}{h}\right)\right)\right|\\&\quad\le 3\left|\Phi_\psi\left(\frac{u}{h}\right)\right|+3\left|\Phi_\psi\left(\frac{1}{h}\right)\right|\le 6\exp\left(\frac{-u^2}{(2h^2)}\right).\end{aligned}\tag{16}$$

Now $E(S_{2,2}-\tilde S_{2,2})=0$ and using (15) and (16),

$$\begin{aligned}\operatorname{var}(S_{2,2}-\tilde S_{2,2})&\le\frac{C}{n^4h^2}\sum_{1\le j\ne k\le n}\frac{E(|X_k|+|X_j|)^2}{h^2}\left(\int_0^1(1-u)|\Phi_K(u)|^2,\mathrm{e}^{u^2/h^2}\,\mathrm{d}u\right)^2\\&\quad+\frac{C'}{n^2h^2}\left(\int_0^1|\Phi_K(u)|^2\,\mathrm{e}^{u^2/(2h^2)}\,\mathrm{d}u\right)^2\\&=O\left(\frac{h^{4+8\alpha}}{n^2}\mathrm{e}^{2/h^2}\right)+O\left(\frac{h^{2+8\alpha}}{n^2}\mathrm{e}^{1/h^2}\right)=O\left(\frac{h^{4+8\alpha}}{n^2}\mathrm{e}^{2/h^2}\right)\end{aligned}$$

for some $C,C'>0$, using (8) and (9). This proves (14).

From (11) and (14),

$$T_n=ES_1+O_P\left(h^{1+4\alpha}\,\mathrm{e}^{1/h^2}n^{-3/2}\right)+\tilde S_{2,1}+\tilde S_{2,2}+O_P\left(h^{2+4\alpha}\,\mathrm{e}^{1/h^2}n^{-1}\right),\tag{17}$$

where ES_1 satisfies (10).

In the second part we study the asymptotics of $\tilde S_{2,i}$, $i=1,2$. First note that the factor $\int_0^1|\Phi_K(u)|^2\,\mathrm{e}^{u^2/h^2}\,\mathrm{d}u$ satisfies (8). Further, from van Es & Uh (2005), proof of lemma 6, we have that X_k/h mod $2\pi\xrightarrow{\mathcal{L}}U_k$, $h\to 0$ for each fixed $k\ge 1$, where U_k is uniform on $(0,2\pi)$. Let us show that in $\tilde S_{2,i}$, $i=1,2$, we can asymptotically replace X_k/h by U_k.

Lemma 1

Let $(U_{k,n})_{k\ge 1}$ be i.i.d. random variables (r.v.s) for each $n\ge 1$, and let $(U_k)_{k\ge 1}$ be i.i.d. r.v.s such that $U_{1,n}\xrightarrow{\mathcal{L}}U_1$ as $n\to\infty$. Then on a joint probability space Ω there exist r.v.s $(V_{k,n}),(V_k)$, such that $(V_{k,n})_{k\ge 1}$ are i.i.d. for each $n\ge 1$, $(V_k)_{k\ge 1}$ are i.i.d., $U_{k,n}\stackrel{\mathcal{L}}{=}V_{k,n}$, $U_k\stackrel{\mathcal{L}}{=}V_k$, and $V_{k,n}\to V_k$ a.s. as $n\to\infty$ for each k.

Proof. From Skorohod's theorem (cf. Billingsley, 1995, theorem 25.6, p. 333), there exists a probability space Ω_0 and random variables $V_{1,n}$, V_1, such that $V_{1,n}\stackrel{\mathcal{L}}{=}U_{1,n}$, $V_1\stackrel{\mathcal{L}}{=}U_1$, and $V_{1,n}\to V_1$ a.s. Now take the product $\Omega=\Omega_0^{\mathbb{N}}$ and construct $V_{k,n}$ and V_k as $V_{1,n}$, V_1, but only depending on the kth coordinate in Ω, to obtain the result.

Lemma 2

Let $(U_{k,n}), (V_{k,n}), (U_k), (V_k)$, be as in lemma 1. Suppose that $f_{n,l}$, f_l, $l=1,\dots,p$, $n\geq 1$ are continuous functions for some $p\geq 1$, such that $|f_{n,l}|$, $|f_l|\leq C$ for some $C>0$ and that $f_{n,l}\to f_l$ uniformly as $n\to\infty$. Further, assume that for $W_{k,n,l}=f_{n,l}(U_{k,n})$, $W_{k,l}=f_l(U_k)$ we have that $EW_{k,n,l}=EW_{k,l}=0$. Set $W'_{k,n,l}=f_{n,l}(V_{k,n})$, $W'_{k,l}=f_l(V_k)$. Then

$$\sum_{1\leq j<k\leq n}\sum_{l=1}^{p} W_{k,n,l}W_{j,n,l} \stackrel{\mathcal{L}}{=} \sum_{1\leq j<k\leq n}\sum_{l=1}^{p} W'_{k,n,l}W'_{j,n,l},\quad n\geq 1,$$

and

$$E\left(\frac{2}{n}\sum_{1\leq j<k\leq n}\sum_{l=1}^{p}(W'_{k,n,l}W'_{j,n,l}-W'_{k,l}W'_{j,l})\right)^2\to 0,\quad n\to\infty.$$

In consequence,

$$\frac{2}{n}\sum_{1\leq j<k\leq n}\sum_{l=1}^{p} W_{k,n,l}W_{j,n,l}\quad \text{and}\quad \frac{2}{n}\sum_{1\leq j<k\leq n}\sum_{l=1}^{p} W_{k,l}W_{j,l}$$

will have the same limiting distribution.

Proof. The first statement follows directly from the properties of the $(V_{k,n})$, as given in lemma 1. For the second statement, for simplicity, we will only consider the case $p=1$ and drop the l-index from the notation. However, the proof of the general case is completely analogous. Since the functions f_n, f are continuous and as $f_n\to f$ uniformly, we have for $W'_{k,n}$ and W'_k that $W'_{k,n}\to W'_k$ a.s., $n\to\infty$. As f_n and f are uniformly bounded, we also get that

$$E(W'_{k,n}-W'_k)^2\to 0,\quad n\to\infty,$$

for each $k\geq 1$. Furthermore, from the properties of the $V_{k,n}$ and V_k and from the assumptions, we have that $EW'_{k,n}=EW'_k=0$ and that

$$E(W'_{k,n}-W'_k)^2=E(W'_{1,n}-W'_1)^2.$$

Now

$$W'_{k,n}W'_{j,n}-W'_kW'_j=(W'_{j,n}-W'_j)W'_{k,n}+(W'_{k,n}-W'_k)W'_j$$

and

$$\left(\sum_{1\leq j<k\leq n}(W'_{k,n}W'_{j,n}-W'_kW'_j)\right)^2 \leq 2\left(\sum_{1\leq j<k\leq n}(W'_{j,n}-W'_j)W'_{k,n}\right)^2+2\left(\sum_{1\leq j<k\leq n}(W'_{k,n}-W'_k)W'_j\right)^2. \tag{18}$$

Further

$$E\left(\frac{2}{n}\sum_{1\leq j<k\leq n}(W'_{j,n}-W'_j)W'_{k,n}\right)^2 =\frac{4}{n^2}\sum_{1\leq j<k\leq n}E(W'_{j,n}-W'_j)^2EW'^2_{k,n} \leq\frac{4}{n^2}\frac{n(n-1)}{2}E(W'_{1,n}-W'_1)^2EW'^2_{1,n}\to 0,\quad n\to\infty.$$

The second term in (18) is dealt with similarly. This proves the lemma.

Proof of theorem 1 continued
We apply lemma 2 with $U_{k,n} = X_k/h \bmod 2\pi$, $f_{n,1}(x) = \cos x - \Re(\Phi_g(1/h))$, $f_{n,1}(x) = \sin x - \Im(\Phi_g(1/h))$, $f_1(x) = \cos x$, $f_2(x) = \sin x$ and uniformly distributed U_ks. Thus we end up by analysing the asymptotic distribution of

$$\frac{2}{n} \sum_{1 \le j < k \le n} (\cos U_j \cos U_k + \sin U_j \sin U_k).$$

As $E \cos U_j = E \sin U_j = 0$, this is a degenerate U-statistic with kernel $a(x, y) = \cos x \cos y + \sin x \sin y$. From the limit theory for degenerate U-statistics (cf. Denker, 1985, proposition 2.2.1, p. 74),

$$\frac{2}{n} \sum_{1 \le j < k \le n} (\cos U_j \cos U_k + \sin U_j \sin U_k) \xrightarrow{\mathcal{L}} \sum_{k \ge 1} \lambda_k (Y_k^2 - 1),$$

where the Y_k are i.i.d. standard normal, and the λ_k are the eigenvalues of the integral operator

$$(Af)(x) = \int_0^1 \big(\cos(2\pi x)\cos(2\pi y) + \sin(2\pi x)\sin(2\pi y)\big) f(y)\,\mathrm{d}y, \quad x \in [0, 1].$$

These are 1/2, corresponding to the eigenfunctions $\cos(2\pi y)$ and $\sin(2\pi y)$, and 0, corresponding to the eigenfunctions 1, $\cos(2\pi n y)$ and $\sin(2\pi n y)$, $n \ge 2$. Combining this result with (17), (8) and (10) yields the theorem.

Remark 1. Hall's (1984a) result (2) is based on a central limit theorem for degenerate U-statistics with variable kernels. Also for the integrated square error (ISE) in normal deconvolution density estimation, one starts with a variable-kernel U-statistic. However, the proof of theorem 1 shows that in this case the statistic asymptotically behaves like a U-statistic with fixed kernel. Therefore, in contrast to Hall (1984a), we get a non-normal limit law.

In fact, often one studies the statistic

$$\begin{aligned} \mathrm{ISE}(\hat{f}_n) &= \int_{\mathbb{R}} \left(\hat{f}_n(x) - f(x)\right)^2 \mathrm{d}x \\ &= T_n + 2\int_{\mathbb{R}} \left(\hat{f}_n(x) - K_h * f(x)\right) \left(K_h * f(x) - f(x)\right) \mathrm{d}x \\ &\quad + \int_{\mathbb{R}} \left(K_h * f(x) - f(x)\right)^2 \mathrm{d}x. \end{aligned} \tag{19}$$

If one corrects the bias terms $K_h * f - f$ by undersmoothing (cf. Bickel & Rosenblatt, 1973), one can ensure that T_n dominates the asymptotics in (19). Otherwise, the second term in the expansion (19) might not be negligible. For the case of direct density estimation, for certain choices of kernels Hall (1984a) gives a complete treatment for which choices of bandwidth the distinct terms in (19) dominate the asymptotics of $\mathrm{ISE}(\hat{g}_n)$.

However, in deconvolution density estimation one typically uses different kernels, namely kernels with a compactly supported Fourier transform. For example, for the sinc kernel $K(x) = \sin(x)/(\pi x)$, for which $\Phi_K(t) = 1_{[-1,1]}(t)$, one can easily show that the second term in the expansion (19) vanishes identically. Further, T_n and not $\mathrm{ISE}(\hat{f}_n)$ is of particular interest when constructing tests for the simple hypothesis $H: f = f_0$ for the density f, and theorem 1 can be used to this end. For a detailed discussion of such indirect testing procedures in the ordinary smooth case cf. Holzmann *et al.* (2006).

Theorem 1 can be generalized to cover the supersmooth error structures (3).

Theorem 2
Under the assumptions 1 and 2, for $n \to \infty$ and $h \to 0$ we have for the integrated square error in the supersmooth deconvolution problem (3) that

$$\frac{(2\lambda)^{1+2\alpha}\pi C_\psi^2 n}{A^2\mu^{1+2\alpha}h^{\lambda-1+2\lambda\alpha+2\lambda_0}\exp(2/\mu h^\lambda)\Gamma(2\alpha+1)} T_n \xrightarrow{\mathcal{L}} \frac{(Y_1^2+Y_2^2)}{2}, \tag{20}$$

where Y_1 and Y_2 are independent standard normal random variables.

Proof of theorem 2. The proof is similar to that of theorem 1, with some additional technical details. Therefore we only present the main steps, in order to clarify where the condition $\lambda > 1$ on the tail behaviour (3) of the characteristic function of the error density ψ is needed. A complete proof can be found in Holzmann & Boysen (2006).

Splitting T_n as in (7) (with $|\Phi_\psi(t)|^{-2}$ in place of $\exp(t^2)$), one shows that

$$ES_1 \sim C_{\alpha,A,\psi}\frac{h^{2\lambda\alpha+\lambda+2\lambda_0-1}}{n}\exp\left(\frac{2}{h^\lambda\mu}\right), \quad C_{\alpha,A,\psi} = \frac{\Gamma(2\alpha+1)A^2}{\pi C_\psi^2}\left(\frac{\mu}{2\lambda}\right)^{2\alpha+1}, \tag{21}$$

and that

$$S_1 = ES_1 + O_P\left(h^{2\lambda\alpha+\lambda+2\lambda_0-1}\exp\left(\frac{2}{h^\lambda\mu}\right)n^{-3/2}\right). \tag{22}$$

Define $S_{2,i}$, $i=1,2$, analogously as in (12) and (13), but replace the integrals in $\tilde{S}_{2,i}$, $i=1,2$, by $\int_\epsilon^1$ for some fixed $0<\epsilon<1$. Then one shows that

$$S_{2,i} - \tilde{S}_{2,i} = O_P\left(h^{2\lambda\alpha+2(\lambda-1)+2\lambda_0}\exp\left(\frac{2}{h^\lambda\mu}\right)n^{-1}\right), \quad i=1,2, \tag{23}$$

and that the norming factor of the $\tilde{S}_{2,i}$, $i=1,2$, satisfies

$$\int_\epsilon^1 |\Phi_K(u)|^2|\Phi_\psi(u/h)|^{-2}\,\mathrm{d}u \sim C_{\alpha,A,\psi}h^{2\lambda\alpha+\lambda+2\lambda_0}\exp\left(\frac{2}{h^\lambda\mu}\right). \tag{24}$$

From (21), (22) and (23),

$$\begin{aligned} &\frac{n}{C_{\alpha,A,\psi}h^{2\lambda\alpha+\lambda+2\lambda_0-1}}\exp\left(-\frac{2}{h^\lambda\mu}\right)T_n \\ &\quad = 1 + o(1) + O_P\left(n^{-1/2}\right) + O_P\left(h^{\lambda-1}\right) \\ &\qquad + \frac{n}{C_{\alpha,A,\psi}h^{2\lambda\alpha+\lambda+2\lambda_0-1}}\exp\left(-\frac{2}{h^\lambda\mu}\right)\left(\tilde{S}_{2,1}+\tilde{S}_{2,2}\right). \end{aligned} \tag{25}$$

Therefore, in order that the remainder terms in (25) vanish, it is essential that $\lambda > 1$. Using (24) the terms in (25) involving $\tilde{S}_{2,i}$, $i=1,2$, are now dealt with exactly as in the proof of theorem 1.

Remark 2. Van Es & Uh (2004) considered the asymptotic distribution of $\hat{f}_n(x)$ in a deconvolution problem where the error density follows a symmetric stable distribution, i.e. where

$$\Phi_\psi(t) = \mathrm{e}^{-|t|^\lambda/\mu}, \quad \mu > 0, \ 0 < \lambda \le 2.$$

They showed asymptotic normality for all $1/3 < \lambda \le 2$, but observed three different cases. For $1/3 < \lambda < 1$, the asymptotic variance still depends on the density of the observations $g(x)$ at x, as in the direct density estimation context and in the ordinary smooth case. For $\lambda = 1$, the asymptotic variance depends in a global way on g, and for $1 < \lambda \le 2$, it only depends on g through the error density ψ, but not on f or x. Thus, the asymptotics change drastically within the supersmooth error class, and not between ordinary smooth and supersmooth error distributions. We conjecture that a similar phenomenon occurs for the integrated square error as well, and that possibly T_n is asymptotically normal for $\lambda < 1$ (but not for $\lambda = 1$).

3. Simulations

To examine the speed of convergence and quality of approximation by the asymptotic law given in theorem 1 we conduct a small simulation experiment. We compute the statistic by its expression (7) in the Fourier domain, thus avoiding the inverse Fourier transform required to calculate $\hat{f}_n$. We perform 10^5 simulations for different values of n and h using the sinc kernel, and for f we choose the Laplace density and for ψ the normal density. Other choices of the kernel and of f lead to similar results.

For visualization we use $P-P$ plots, which show for each α in $[0, 1]$ the probability that $T_n \leq Q_\alpha$, where Q_α is the α-quantile of the asymptotic distribution. Figure 1 displays the result for fixed sample size $n = 50$ and different values of the bandwidth parameter h. In Fig. 2, for different sample sizes n we selected the bandwidth $h = h(n)$ such that the variance of T_n (as well as that of $\hat{f}_n(x)$) asymptotically disappears. Figure 3 visualizes the results by plotting a kernel estimate of the density of T_n as well as the density of the asymptotic distribution.

The plots show that the speed of convergence strongly depends on the choice of bandwidth h. In particular, taking a sample size of $n = 50$ and a small h already leads to very good results. This reflects the decomposition of the statistic T_n given in (17). In fact, we see that the asymptotic behaviour of the $a_{n,h}T_n$ is determined by $a_{n,h}(\tilde{S}_{2,1} + \tilde{S}_{2,2})$, where $a_{n,h}$ is the normalization factor from theorem 1, and the remainder terms are of the order $O_P(h)$ and $O_P(n^{-1/2})$. If one selects h in a way appropriate to estimate f, the approximation is less good

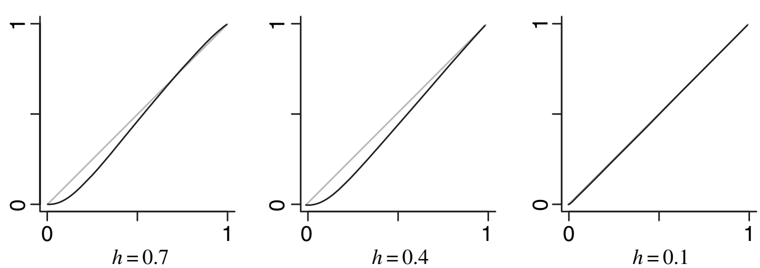

Fig. 1. $P-P$ plots comparing the empirical distribution of T_n to the asymptotic distribution for different bandwidths h, using the sample size $n = 50$, the sinc kernel, $f(x) = 1/2\exp(-|x|)$ and standard normal error.

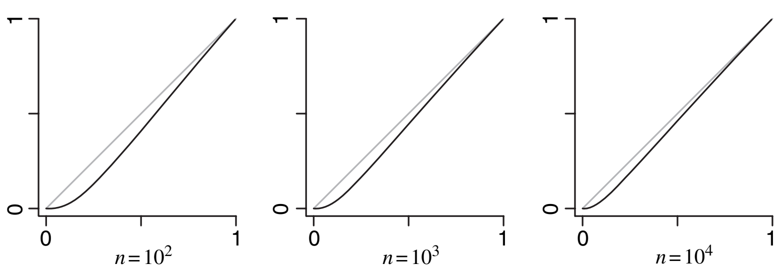

Fig. 2. $P-P$ plots comparing the empirical distribution of T_n to the asymptotic distribution for different samples sizes n, using $h = (\log(n))^{-1/2}$, the sinc kernel, $f(x) = 1/2\exp(-|x|)$ and standard normal error.

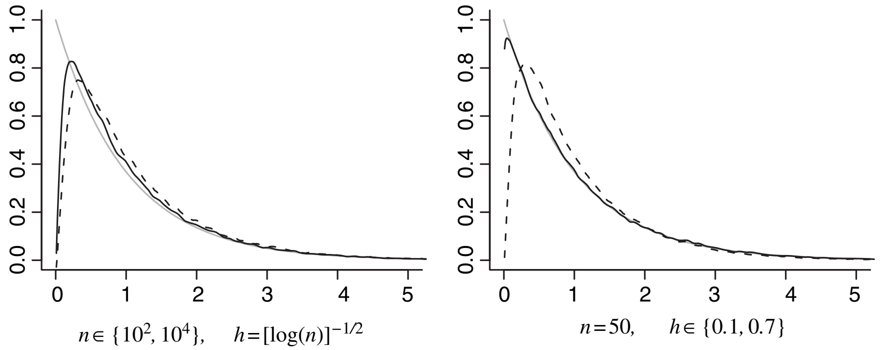

Fig. 3. Kernel density estimate for T_n. The grey line shows the asymptotic density. The solid line corresponds to $n = 10^4$ and $h = 0.1$, respectively, and the dashed line to $n = 10^2$ and $h = 0.7$ respectively. Close to zero a boundary kernel is used.

but still remains reasonable. This is especially true for $\alpha \in [0.9, 1]$, the region of main interest for testing.

We also conduct simulations for two cases to which theorem 2 does not apply directly, in order to see whether it could be further extended. First, we use as error distribution the symmetric stable distribution with characteristic function $\Phi_\psi(t) = \exp(-|t|^{3/2})$, for which the moment restriction $EX_1^2 < \infty$ of assumption 1 is not satisfied. The asymptotic distribution (20) seems to apply in this case as well. Further, we consider a t-distributed error with 5 d.f., for which

$$\Phi_\psi(t) = \left(1 + |\sqrt{5}t| + \frac{5}{3}t^2\right) \exp\left(-\sqrt{5}|t|\right).$$

For this error we do not have $\lambda > 1$ in (3). Simulations indicate that neither the asymptotic chi-squared-distribution nor an asymptotic normal law applies for T_n in this case.

Acknowledgements

Hajo Holzmann gratefully acknowledges financial support of the Deutsche Forschungsgemeinschaft through the DFG project 'Statistische Methoden der Modellwahl in der Regressionsanalyse'. Leif Boysen acknowledges support of the Georg Lichtenberg program 'Applied Statistics & Empirical Methods' and the DFG graduate program 'Identification in Mathematical Models'. We would like to thank two anonymous referees for helpful comments and for pointing out some relevant references.

References

Bickel, P. & Ritov, Y. (1987). Efficient estimation in the errors in variables model. *Ann. Statist.* **15**, 513–540.

Bickel, P. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095.

Billingsley, P. (1995). *Probability and measure*, 3rd edn. Wiley, New York.

Butucea, C. (2004). Asymptotic normality of the integrated square error of a density estimator in the convolution model. *SORT* **28**, 9–26.

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

Denker, M. (1985). *Asymptotic distribution theory in nonparametric statistics*. Vieweg, Braunschweig.

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

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

Fan, J. (1992). Deconvolution with supersmooth distributions. *Can. J. Statist.* **20**, 155–169.

Hall, P. (1984a). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14**, 1–16.

Hall, P. (1984b). Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Geb.* **67**, 175–196.

Holzmann, H. & Boysen, L. (2006). *Integrated square error asymptotics for supersmooth deconvolution: technical details*. Technical Report. Institute for Mathematical Stochastics, University of Göttingen, Göttingen, Germany.

Holzmann, H., Bissantz, N. & Munk, A. (2006). Density testing in a contaminated sample. *J. Multivariate Anal*. (in press).

Kotz, S. & Nadarajah, S. (2004). *Multivariate t distributions and their applications*. Cambridge University Press, Cambridge.

Mair, B. A. & Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56**, 1424–1444.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.

Piterbarg, V. I. & Penskaya, M. Y. (1993). On asymptotic distribution of integrated square error of an estimate of a component of a convolution. *Math. Methods Statist.* **2**, 30–41.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837.

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

van Es, A. J. & Uh, H.-W. (2004). Asymptotic normality for kernel type deconvolution density estimators: crossing the Cauchy boundary. *J. Nonparametr. Statist*. **16**, 261–277.

van Es, A. J. & Uh, H.-W. (2005). Asymptotic normality for kernel type deconvolution estimators. *Scand. J. Statist.* **32**, 467–483.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806–831.

Received August 2005, in final form March 2006

Hajo Holzmann, Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Maschmühlenweg 8–10, 37073 Göttingen, Germany.
E-mail: holzmann@math.uni-goettingen.de