AUSTRIAN JOURNAL OF STATISTICS
Volume 32 (2003), Number 1&2, 131–142

Representations for Integral Functionals of Kernel Density Estimators

David M. Mason
University of Delaware, Newark

Abstract: We establish a representation as a sum of independent random variables, plus a remainder term, for estimators of integral functionals of the density function, which have a certain simple structure. From this representation we derive a central limit theorem, a law of large numbers and a law of the iterated logarithm.

Keywords: Kernel Density Estimators, Integral Functionals, Law of Large Numbers, Central Limit Theorem, Law of the Iterated Logarithm.

1 Introduction

For several decades there has been considerable interest in the problem of estimating integral functionals of the density function of the form

$$T(f) = \int_{\mathbb{R}} \varphi\left(f(x), f'(x), \ldots, f^{(k)}(x), x\right)\, dx, \tag{1}$$

where f is a Lebesgue density. Such integral functionals appear in the asymptotic variance of certain nonparametric statistics. They also arise in a number of bandwidth selection procedures for density estimators. (Refer, especially, to Bickel and Ritov, 1988; Birgé and Massart, 1995; Hall and Marron, 1987; Laurent, 1996, 1997; Levit, 1978; Nadaraya, 1989, along with the references therein.). Some motivating special cases are

$$T_1(f) := \int_{\mathbb{R}} \left(f'(x)\,\right)^2 dx,\ T_2(f) := \int_{\mathbb{R}} f^2(x)dx \text{ and } T_3(f) =: \int_{\mathbb{R}} f(x)\log(f(x))\, dx.$$

Let $\widehat{f}_h(x)$ denote the *kernel density estimator*

$$\widehat{f}_h(x) = n^{-1}\sum_{i=1}^{n} h^{-1}K\left(\frac{x - X_i}{h}\right) := n^{-1}\sum_{i=1}^{n} K_h\left(x - X_i\right), \tag{2}$$

where K is a *kernel* satisfying conditions (K.i) and (K.ii) below and $h_n = h$ is a sequence of positive constants converging to zero. Set

$$f_h(x) = E\widehat{f}_h(x) = f * K_h(x). \tag{3}$$

We will assume that K satisfies

(K.i) K is non-negative and bounded by some $0 < \kappa < \infty$;

(K.ii) $\int_{-\infty}^{\infty} K(u)du = 1$.

Sometimes we will also impose the condition

(K.iii) $\int_R \Psi(x)dx < \infty$, where $\Psi(x) = \sup_{|y|\geq|x|} |K(y)|$.

Here are some examples of kernels that satisfy these conditions.

1. $K(u) = 1\{u \in [-1/2, 1/2]\}$ (Uniform Kernel);
2. $K(u) = \frac{3}{4}(1 - x^2)_+$ (Epanechnikov Kernel);
3. $K(u) = \frac{1}{2}\exp(-|u|)$ (Double exponential Kernel);
4. $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$ (Standard Normal Kernel).

Assuming a smooth enough density f, Eggermont and LaRiccia (1999) (also see their monograph Eggermont and LaRiccia, 2001) established an almost sure representation for the kernel density function negative entropy estimator

$$T(\widehat{f}_h) = \int_{\mathbb{R}} \widehat{f}_h(x)\log(\widehat{f}_h(x))\,dx, \tag{4}$$

where $K(u) = 2^{-1}\exp(-|u|)$, $u \in \mathbb{R}$, is the double exponential kernel. Namely they proved under suitable regularity conditions on $h = h_n$ and f that the following representation as a sum of independent random variables, plus a remainder term, holds for $T(\widehat{f}_h)$:

$$T(\widehat{f}_h) - T(f) = n^{-1}\sum_{i=1}^{n}\{\log f(X_i) - E\log f(X_i)\} + R_n^*,$$

where $R_n^* = o\left(n^{-1/2}\right)$, $a.s.$ From this representation they inferred a law of the iterated logarithm [LIL] for $T(\widehat{f}_h)$, along with best asymptotic normality by applying a result of Levit (1978). Their approach was based upon a special submartingale structure of the estimator, which follows from a Green's function property possessed by the double exponential kernel. For other approaches towards proving consistency and best asymptotic normality of estimators of integral functionals of f consult Laurent (1996), who concentrates on the special cases $T_2(f)$ and $T_3(f)$, and the references therein.

It would be very useful to develop a general approach to establish the asymptotic distribution and exact rate of consistency under minimal assumptions of estimators of $T(f)$ of the form

$$\int_{\mathbb{R}} \varphi\left(\widehat{f}_h(x), \widehat{f}_h^{(1)}(x), \ldots, \widehat{f}_h^{(k)}(x), x\right) dx, \tag{5}$$

where for each $1 \le i \le k$, $\widehat{f}_h^{(i)}$ is the ith derivative of $\widehat{f}_h$, when K is i times continuously differentiable. However, at present this goal appears to be much too ambitious. In this note we obtain a representation in terms of a sum of independent random variables, plus a remainder term, for integral functionals of density estimators having a certain simple structure. From our representation we will derive a central limit theorem, a law of large numbers and an LIL for $T(\widehat{f}_h)$. In the last section we will apply our results to the special case of the kernel density estimator $T_2(\widehat{f}_h)$ of the integrated squared density $T_2(f)$.

2 Representations of Integral Functionals of Density Estimators

In this note we shall restrict ourselves to integral functionals of the following form:

$$T(f) = \int_{\mathbb{R}} f(x)\phi(f(x))dx, \tag{6}$$

where ϕ is a twice continuously differentiable function defined on an open set containing the closure of the image of f, $\{f(x) : x \in R\}$. The motivating special cases are $T_2(f)$ and $T_3(f)$. Set $\Phi(u) = u\phi(u)$,

$$\Phi_1(u) = \frac{d}{dx}\Phi(x)|_{x=u} \text{ and } \Phi_2(u) = \frac{d}{dx}\Phi_1(x)|_{x=u}. \tag{7}$$

By Taylor's formula we have

$$T(\widehat{f}_h) - T(f_h) =: \int_{\mathbb{R}} \Phi_1(f_h(x)) \left(\widehat{f}_h(x) - f_h(x)\right) dx + R_n, \tag{8}$$

where

$$R_n = \int_{\mathbb{R}} \left\{ \int_{f_h(x)}^{\widehat{f}_h(x)} \left(\widehat{f}_h(x) - u\right) \Phi_2(u)\, du \right\} dx. \tag{9}$$

We shall confine ourselves to determining the rate at which the remainder term R_n in (9) converges to zero, when there exists a positive nondecreasing function φ with derivative φ' such that

(φ.i) $|\Phi_2(u)| \le \varphi'(u)$;

(φ.ii) φ satisfies a Hölder condition with exponent $0 < \alpha \le 1$.

Notice that the class of integral functionals satisfying these conditions includes the $T_2(f)$ functional, but not the negative entropy functional $T_3(f)$. To establish a useful representation for $T_3(f)$ under minimal assumptions requires too much space and therefore will be presented elsewhere.

Theorem 1. *Assume that K satisfies (K.i) and (K.ii) and that (φ.i) and (φ.ii) hold. Then for R_n in (9) and any sequence of positive $h = h_n \le 1$, we have*

$$R_n = O\left(\left(\frac{(\log n)^{(1+\alpha)/2}}{n^{(1-\alpha)/2}} + 1\right)\frac{1}{(nh)^{\alpha}}\right), \text{ a.s.} \tag{10}$$

and

$$R_n = O_p\left(\frac{1}{(nh)^{\alpha}}\right). \tag{11}$$

Corollary 1. I*n addition to assumptions (K.i), (K.ii), (φ.i) and (φ.ii), assume that there exists a $C_\alpha > 0$ such that*

$$\sup_{0<h\le 1} \int_{\mathbb{R}} \{f_h(x)\}^{(1+\alpha)/2} dx \le C_\alpha. \tag{12}$$

Then for R_n in (9) *and any sequence of positive $h = h_n \le 1$, we have*

$$R_n = O\left(\frac{(\log n)^{(1+\alpha)/2}}{n^{(1+\alpha)/2}h^{\alpha}} + \frac{1}{(nh)^{(1+\alpha)/2}}\right), \text{ a.s.} \tag{13}$$

and

$$R_n = O_p\left(\frac{1}{(nh)^{(1+\alpha)/2}}\right). \tag{14}$$

Remark 1. Note that for $0 < \alpha < 1$, the R_n in (10) is $= O\left(1/\left(nh\right)^{\alpha}\right)$, a.s., and when $\alpha = 1$, this $R_n = O\left(\log n/\left(nh\right)\right)$, a.s.
Remark 2. Assume $h = h_n = n^{-\gamma}$ for some $\gamma > 0$. Notice that under the assumptions of Theorem 1, that if $1/2 < \alpha \leq 1$ and $1 - \frac{1}{2\alpha} > \gamma > 0$, then $R_n = o(n^{-1/2})$, a.s., and under the assumptions of Corollary 1, if we choose $0 < \gamma < \alpha/(1+\alpha)$ then $R_n = o(n^{-1/2})$, a.s.
Remark 3. Condition (12) is satisfied whenever for some $\lambda > (1-\alpha)/(1+\alpha)$,

$$\int_{\mathbb{R}} |x|^{\lambda} f(x)dx + \int_{\mathbb{R}} |y|^{\lambda} K(y)dy < \infty.$$

Consult Devroye (1987) for a proof of this fact.

2.1 Proof of Theorem 1

Notice that by (φ.i), the remainder term R_n defined in (9) satisfies

$$|R_n| \leq \int_{\mathbb{R}} \left(\widehat{f}_h(x) - f_h(x)\right)\left(\varphi(\widehat{f}_h(x)) - \varphi(f_h(x))\right) dx,$$

which by (φ.ii) is, for some $C > 0$,

$$\leq C \int_{\mathbb{R}} \left(\left|\widehat{f}_h(x) - f_h(x)\right|^{1+\alpha}\right) dx =: C\Delta_n(\alpha). \tag{15}$$

Write for $i = 1, 2, \ldots, n$,

$$Y_i(x) = n^{-1}\left\{K_h\left(x - X_i\right) - f_h(x)\right\},$$

so that

$$\left(\Delta_n(\alpha)\right)^{1/(1+\alpha)} = \left\|\sum_{i=1}^{n} Y_i\right\|_{1+\alpha}. \tag{16}$$

Furthermore, for each $i = 1, 2, \ldots, n$, we get

$$\left\|n^{-1}K_h\left(\cdot - X_i\right)\right\|_{1+\alpha} = \frac{1}{nh^{\alpha/(1+\alpha)}}\left(\int_{\mathbb{R}} h^{-1}K\left(\frac{x - X_i}{h}\right)^{1+\alpha} dx\right)^{1/(1+\alpha)} = \frac{||K||_{1+\alpha}}{nh^{\alpha/(1+\alpha)}}.$$

Thus by applying the McDiarmid inequality, exactly as on page 39 of Devroye (1991) (also refer to Pinelis (1990)), we get for any $t > 0$,

$$P\left\{\left|\left\|\sum_{i=1}^{n} Y_i\right\|_{1+\alpha} - E\left\|\sum_{i=1}^{n} Y_i\right\|_{1+\alpha}\right| > t\right\} \leq 2\exp\left(-\frac{t^2 nh^{2\alpha/(1+\alpha)}}{2||K||_{1+\alpha}^2}\right). \tag{17}$$

From this inequality we readily conclude via the Borel-Cantelli lemma that

$$\left\|\sum_{i=1}^{n} Y_i\right\|_{1+\alpha} = E\left\|\sum_{i=1}^{n} Y_i\right\|_{1+\alpha} + O\left(\frac{\sqrt{\log n}}{\sqrt{n}h^{\alpha/(1+\alpha)}}\right), \text{ a.s.} \tag{18}$$

Now

$$\left(E\left\|\sum_{i=1}^{n}Y_i\right\|_{1+\alpha}\right)^{1+\alpha}\le E\left\|\sum_{i=1}^{n}Y_i\right\|_{1+\alpha}^{1+\alpha}=\int_{\mathbb{R}}E\left|\sum_{i=1}^{n}Y_i(x)\right|^{1+\alpha}dx,$$

which by the moment inequality of von Bahr and Esseen (1965) is

$$\le 2\sum_{i=1}^{n}\int_{\mathbb{R}}E|Y_i(x)|^{1+\alpha}dx.$$

Next observe that by Jensen's inequality for each $i=1,\dots,n$,

$$\sum_{i=1}^{n}\int_{\mathbb{R}}E|Y_i(x)|^{1+\alpha}dx\le$$

$$\sum_{i=1}^{n}\frac{2^{\alpha}}{n^{1+\alpha}}\left[\int_{\mathbb{R}}E\left(K_h\left(x-X_i\right)\right)^{1+\alpha}dx+\int_{\mathbb{R}}\left(\int_{\mathbb{R}}K_h\left(x-y\right)f\left(y\right)dy\right)^{1+\alpha}dx\right],$$

which, in turn, again by Jensen's inequality and $|K|\le\kappa$, is

$$\le\frac{2^{\alpha+1}\kappa^{\alpha}}{n^{\alpha}h^{\alpha}}\int_{\mathbb{R}^2}K_h\left(x-y\right)f(y)dydx=\frac{2^{\alpha+1}\kappa^{\alpha}}{n^{\alpha}h^{\alpha}}. \tag{19}$$

Thus we have by (15), (16), (18) and (19),

$$\Delta_n(\alpha)=O\left(\frac{(\log n)^{(1+\alpha)/2}}{n^{(1+\alpha)/2}h^{\alpha}}+\frac{1}{(nh)^{\alpha}}\right),\text{ a.s.}$$

Assertion (11) follows in the same way from (17). □

2.2 Proof of Corollary 1

We will only prove (13). Assertion (14) is proved similarly. Notice that

$$E\left\|\sum_{i=1}^{n}Y_i\right\|_{1+\alpha}^{1+\alpha}=\int_{\mathbb{R}}E\left|\sum_{i=1}^{n}Y_i(x)\right|^{1+\alpha}dx\le\int_{\mathbb{R}}\left[E\left(\sum_{i=1}^{n}Y_i(x)\right)^2\right]^{(1+\alpha)/2}dx$$

$$\le\int_{\mathbb{R}}\left[\frac{1}{nh^2}\int_{\mathbb{R}}\left(K\left(\frac{x-y}{h}\right)\right)^2f(y)dy\right]^{(1+\alpha)/2}dx,$$

which by using $|K|\le\kappa$, is

$$\le\frac{\kappa^{(1+\alpha)/2}}{(nh)^{(1+\alpha)/2}}\int_{\mathbb{R}}\left[\int_{\mathbb{R}}K_h\left(x-y\right)f(y)dy\right]^{(1+\alpha)/2}dx.$$

By assumption (12) this last bound is $\le\kappa^{(1+\alpha)/2}C_{\alpha}/\left(nh\right)^{(1+\alpha)/2}$. Hence from (18) we conclude that whenever (12) holds we have almost surely that

$$\Delta_n(\alpha)=O\left(\frac{(\log n)^{(1+\alpha)/2}}{n^{(1+\alpha)/2}h^{\alpha}}+\frac{1}{(nh)^{(1+\alpha)/2}}\right).$$

□

3 Central Limit Theorem and Law of Large Numbers

Our next theorem, when combined with Theorem 1 and Corollary 1, provides conditions under which the central limit theorem and law of large numbers hold for $T(\widehat{f}_h)$. Introduce the assumptions

(Φ.i) f is bounded by some constant $M > 0$;

(Φ.ii) Φ_1 is continuous and bounded by a constant L on $[0, M]$.

Write for $h > 0$,

$$S_n(h) := \int_{\mathbb{R}} \Phi_1(f_h(x)) \left(\widehat{f}_h\,(x) - f_h\,(x)\right)\, dx.$$

Theorem 2. *Assume that K satisfies (K.i), (K.ii) and (K.iii) and (Φ.i) and (Φ.ii) hold. Then for any sequence of positive constants $h = h_n \leq 1$ converging to zero and satisfying*

$$T(\widehat{f}_h) - T(f_h) = S_n(h) + R_n, \tag{20}$$

with $R_n = o(1)$ almost surely, we have, with probability 1, as $n \to \infty$,

$$T(\widehat{f}_h) \to \int_{\mathbb{R}} \Phi_1(f(y))f(y)dy, \tag{21}$$

and whenever $R_n = o_p\left(n^{-1/2}\right)$ in (12), *we have, as $n \to \infty$,*

$$\sqrt{n}\left\{T(\widehat{f}_h) - T(f_h)\right\} \to_d N\left(0, \sigma_f^2(\Phi_1)\right), \tag{22}$$

where

$$\sigma_f^2(\Phi_1) := \int_{\mathbb{R}} \Phi_1^2(f(y))f(y)dy - \left(\int_{\mathbb{R}} \Phi_1(f(y))f(y)dy\right)^2. \tag{23}$$

Remark 4. Applying the results of Levit (1978) (also consult Laurent, 1996) we see that $\sigma_f^2(\Phi_1)$ is the variance of any best asymptotically normal estimator of $T\,(f)$.

3.1 Proof of Theorem 2

Our proof requires the following fact and lemma. First we need a special case of a result in Stein (1970) (also consult Devroye and Györfi, 1985).

Fact 1. (Stein) *Let φ be a measurable function on $\mathbb{R}$ such that $\int_{\mathbb{R}} \varphi(x)dx = 1$ and set $\varphi_\varepsilon(x) = \varepsilon^{-1}\varphi(\varepsilon^{-1}x)$, $\varepsilon > 0$. Further let $\Psi(x) = \sup_{|y|\geq|x|} |\varphi(y)|$ and assume that $\int_{\mathbb{R}} \Psi(x)dx < \infty$. Then for all functions g in $L_1\,(\mathbb{R})$ and almost all $x \in \mathbb{R}$,*

$$\lim_{\varepsilon\searrow 0} g * \varphi_\varepsilon(x) = g(x). \tag{24}$$

For for $h > 0$, define

$$Z(h) = \int_{\mathbb{R}} \Phi_1(f_h(x))K_h(x - X)dx, \tag{25}$$

where X has density f.

Lemma 1. *Under the assumptions of Theorem 2, as* $h \searrow 0$,

$$E\left(Z(h)\right) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \Phi_1(f_h(s)) K_h\left(s-y\right) ds\right] f(y)dy \to \int_{\mathbb{R}} \Phi_1(f(y))f(y)dy \tag{26}$$

and

$$\begin{aligned} E\left(Z(h)\right)^2 &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}^2} \Phi_1(f_h(s))\Phi_1(f_h(t))K_h\left(s-y\right)K_h\left(t-y\right)dsdt\right] f(y)dy \\ &\to \int_{\mathbb{R}} \Phi_1^2(f(y))f(y)dy < \infty. \end{aligned} \tag{27}$$

Proof. First consider (27). Notice that after a change of variables the left hand side of (27) is equal to

$$\int_{\mathbb{R}} \left[\int_{\mathbb{R}^2} \Phi_1(f_h(y+hu))\Phi_1(f_h(y+hv))K\left(u\right)K\left(v\right)dudv\right] f(y)dy.$$

Writing

$$f_h(y+hu) = \int_{\mathbb{R}} h^{-1} K\left(\frac{y-t}{h} + u\right) f(t)dt,$$

we see by the Stein Fact 1 that for almost every y and all u, as $h \searrow 0$,

$$f_h(y+hu) \to f(y).$$

Therefore, since $|f_h| \le M$, we have by (Φ.ii) that for almost every y and every u and v,

$$\Phi_1(f_h(y+hu))\Phi_1(f_h(y+hv)) \to \Phi_1^2(f(y)).$$

Thus by the bounded convergence theorem for almost every y,

$$\varphi_h(y) := \int_{\mathbb{R}^2} \Phi_1(f_h(y+hu))\Phi_1(f_h(y+hv))K\left(u\right)K\left(v\right)dudv \to \Phi_1^2(f(y)).$$

Moreover, since each φ_h is bounded, we have as $h \searrow 0$,

$$\int_{\mathbb{R}} \varphi_h(y)f(y)dy \to \int_{\mathbb{R}} \Phi_1^2(y)f(y)dy.$$

This proves (27). Assertion (26) is proved similarly. □

Turning to the proof of Theorem 2, write for $i = 1, \ldots, n$,

$$Z_i\left(h\right) := \int_{\mathbb{R}} \Phi_1(f_h(x))K_h\left(x - X_i\right)dx, \tag{28}$$

where each $X_1, \ldots, X_n$ are i.i.d. with density f. Note that since each $|Z_i\left(h\right)| \le L$ and

$$S_n(h) = n^{-1} \sum_{i=1}^{n} \left\{Z_i\left(h\right) - EZ_i\left(h\right)\right\},$$

we can apply Hoeffding's inequality to get for all $t > 0$,

$$P\left\{|S_n(h)| > t/\sqrt{n}\right\} \leq 2\exp\left(-2t^2/L^2\right),$$

from which we readily check using the Borel–Cantelli lemma that almost surely

$$S_n(h) = O\left(\sqrt{\log n/n}\right). \tag{29}$$

Thus (21) is a consequence of (26).

Turning to the proof of (22), observe by (26) and (27), as $h \searrow 0$,

$$Var\left(Z\left(h\right)\right) \rightarrow \sigma_f^2(\Phi_1).$$

Thus (22) follows directly from $|Z_i\left(h\right)| \leq L$, combined with Liapounov's central limit theorem (e.g. Shorack, 2000). □

4 Laws of the Iterated Logarithm

In this section we introduce a general approach to obtain laws of the iterated logarithm [LIL] for $T(\widehat{f}_h)$. This will be based upon the following special case of a LIL stated as Theorem 4.2 in Kuelbs (1976). Let $\mathcal{C}[0, h_0]$, $h_0 > 0$, denote the space of continuous functions on $[0, h_0]$ equipped with the supremum norm $||\cdot||_\infty$.

Proposition 1. *Let* $Z, Z_1, Z_2, \ldots,$ *be a sequence of i.i.d.* $\mathcal{C}[0, h_0]$, $h_0 > 0$, *valued random variables satisfying for all* $h \in [0, h_0]$,

$$VarZ(h) = \sigma^2(h) < \infty. \tag{30}$$

Also assume there exists a nonnegative random variable W *satisfying* $EW^2 < \infty$ *and an* $0 < \gamma \leq 1$ *such that for all* $h_1, h_2 \in [0, h_0]$

$$|Z(h_1) - Z(h_2)| \leq W\,|h_1 - h_2|^\gamma. \tag{31}$$

Then

$$\limsup_{n\rightarrow\infty} \sup_{h\in[0,h_0]} \pm \frac{\sum_{i=1}^n \{[Z_i(h) - EZ_i(h)\}}{\sqrt{2n\log\log n}} = \tau,\ \text{a.s.} \tag{32}$$

where

$$\tau = \sup_{h\in[0,h_0]} \sigma(h). \tag{33}$$

We will soon be applying the following straightforward corollary to Proposition 1.

Corollary 2. *In addition to the assumptions of Proposition 1, assume that*

$$\lim_{h\searrow 0} Var(Z(h) - Z(0)) = 0. \tag{34}$$

Then for any sequence of positive constants $h_n \in [0, h_0]$ *converging to zero,*

$$\limsup_{n\rightarrow\infty} \pm \frac{\sum_{i=1}^n \{Z_i(h_n) - EZ_i(h_n)\}}{\sqrt{2n\log\log n}} = \sigma(0),\ \text{a.s.} \tag{35}$$

4.1 The Aim of All This

The aim of all this is to show, under suitable regularity conditions, that almost surely

$$\limsup_{n\to\infty} \pm \frac{\sqrt{n}\left\{T(\widehat{f}_{h_n}) - T(f_{h_n})\right\}}{\sqrt{2\log\log n}} = \limsup_{n\to\infty} \pm \frac{\sum_{i=1}^{n}\left\{Z_i(h_n) - EZ_i(h_n)\right\}}{\sqrt{2n\log\log n}} = \sigma(0), \tag{36}$$

where for $h > 0$, $Z_1(h), Z_2(h), \ldots,$ are i.i.d. $Z(h)$, with $Z(h)$ as in (25), and $\sigma^2(0) = \sigma_f^2(\Phi_1)$, whenever $h = h_n$ is such that $R_n = o\left(\sqrt{\log\log n/n}\right)$, a.s. Here are some sufficient conditions that will permit us to apply Corollary 2 to give (36).

Example. Assume that Φ_1 is Hölder of order $0 < \alpha \le 1$ and f satisfies the condition that for some $0 < \beta \le 1$ and all $x, y \in R$

$$|f(x+y) - f(x)| \le g(x)|y|^{\beta}, \tag{37}$$

where $E\left(g^{\alpha}(X)\right)^2 < \infty$. Also assume that

$$\int_{-\infty}^{\infty} |u|^{\beta} K(u)du < \infty. \tag{38}$$

To verify that (31) holds, observe that by two changes of variables

$$Z(h) = \int_{\mathbb{R}} \Phi_1\left(\int_{\mathbb{R}} f\left(X + h\left(u - v\right)\right) K(v)dv\right) K(u)du,$$

Therefore for some constant $C > 0$

$$\begin{aligned} &|Z(h_1) - Z(h_2)| \\ &\le C\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\left\{f\left(X + h_1\left(u-v\right)\right) - f\left(X + h_2\left(u-v\right)\right)\right\}K(v)dv\right|^{\alpha} K(u)du, \end{aligned}$$

which by concavity ($0 < \alpha \le 1$) and (37) is for some constant $B > 0$

$$\le Bg^{\alpha}(X)\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\left(|u|^{\beta} + |v|^{\beta}\right)K(u)K(v)dudv\right|^{\alpha}|h_1 - h_2|^{\alpha\beta} =: Ag^{\alpha}(X)|h_1 - h_2|^{\alpha\beta}.$$

Therefore (31) holds with $W = Ag^{\alpha}(X)$ and $\gamma = \alpha\beta$.

Remark 5. Assumption (37) is a generalization of a condition of Bickel and Ritov (1988).

5 Application to the Estimation of the Integrated Squared Density

The functional $T_2(f)$ serves as a basic quantity that appears in expressions for the asymptotic efficiency of the Wilcoxon signed rank statistic for location hypotheses and the asymptotic variance of the Hodges-Lehmann location estimator. (Refer, for instance, to

Randles and Wolfe, 1991). In our final section we apply our results to the estimation of $T_2(f)$. First we employ Theorem 1 with $\phi(u) = u$ and $\alpha = 1$ to give for any density f,

$$T_2(\widehat{f}_h) - T_2(f_h) = 2\int_{\mathbb{R}} f_{h_n}(x)\left(\widehat{f}_h(x) - f_h(x)\right)dx + R_n$$

where

$$R_n = O\left(\frac{\log n}{nh}\right), \text{ a.s. and } R_n = O_p\left(\frac{1}{nh}\right).$$

Now assume that f is bounded. We see that whenever $h_n \to 0$ and $nh_n/\log n \to \infty$, we have $R_n = o(1)$ a.s., which by Theorem 2, implies

$$T_2(\widehat{f}_h) \to T_2(f), \text{ a.s.} \tag{39}$$

Further, whenever $h_n \to 0$ and $\sqrt{n}h_n \to \infty$, yielding $R_n = o_p(n^{-1/2})$, we get by Theorem 2 that

$$\sqrt{n}\left\{T_2(\widehat{f}_h) - T_2(f_h)\right\} \to_d N\left(0, \sigma_f^2(\Phi_1)\right), \tag{40}$$

where

$$\sigma_f^2(\Phi_1) = 4\left[\int_{\mathbb{R}} f^3(x)dx - \left(\int_{\mathbb{R}} f^2(x)dx\right)^2\right].$$

Moreover, whenever $h_n \to 0$ and $\frac{h_n\sqrt{n\log\log n}}{\log n} \to \infty$, giving $R_n = o\left(\sqrt{\log\log n/n}\right)$ a.s., and both (37) and (38) hold for some $0 < \beta \le 1$, then by Proposition 1 and the Example above,

$$\limsup_{n\to\infty} \pm\frac{\sqrt{n}\left\{T_2(\widehat{f}_h) - T_2(f_h)\right\}}{\sqrt{2\log\log n}} = \sigma_f(\Phi_1), \text{ a.s.} \tag{41}$$

Remark 6. By taking advantage of the $U-$statistic structure of $T_2(\widehat{f}_h)$, Nadaraya (1989) establishes (39) in his Theorem 2.2 under the same assumptions that we impose on K, but with our condition $||f||_\infty < \infty$ replaced by the weaker assumption that $T_2(f) < \infty$. However, he requires a stronger condition on h_n, namely he substitutes our $nh_n/\log n \to \infty$ by the more restrictive condition that $nh_n^2/\log n \to \infty$. It is also interesting to compare his Theorem 2.3 with our assumptions leading to (40). He shows that with more smoothness assumptions $T_2(f_h)$ can be replaced by $T_2(f)$ in (40). Another approach to the LIL for $T_2(\widehat{f}_h)$, which would require fewer regularity conditions on the density function, can be based on a moderate deviation result for $T_2(\widehat{f}_h)$. For closely related details consult Giné and Mason (2002).

Acknowledgements

The author would like to thank his colleagues Paul Deheuvels, Paul Eggermont and William Wedin for useful comments and the referee for carefully reading the manuscript.

This work was completed with the support of NSA Grant MDA904–02–1–0034 and NSF Grant DMS–0203865.

References

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

L. Birgé and P. Massart. Estimation of integral functionals of a density. *Ann. Statist.*, 23: 11–29, 1995.

L. Devroye. *A Course in Density Estimation*. Birkhäuser, Boston, 1987.

L. Devroye. Exponential inequalities in nonparametric estimation. In G. Roussas, editor, *Nonparametric Functional Estimation and Related Topics*, pages 31–44. Kluwer Academic Publishers, 1991.

L. Devroye and L. Györfi. *Nonparametric Density Estimation: The L-1 View*. J. Wiley, New York, 1985.

P.P.B. Eggermont and V.N. LaRiccia. Best asymptotic normality of the kernel density entropy estimator for smooth densities. *IEEE Trans. Inform. Theory*, 45:1321–1326, 1999.

P.P.B. Eggermont and V.N. LaRiccia. *Maximum Penalized Likelihood: Volume I, Density Estimation*. Springer, New York, 2001.

E. Giné and D.M. Mason. The law of iterated logarithm for the integrated squared deviation of a kernel density estimator. *Preprint*, 2002.

P. Hall and J.S. Marron. Estimation for integrated squared density derivatives. *Statist. Probab. Lettres*, 6:109–155, 1987.

J. Kuelbs. The law of the iterated logarithm and related strong convergence theorems for Banach space valued random variables. In *Ecole d'Eté de Probabilités de Saint-Flour, V-1975, Lecture Notes in Math. 539*, pages 224–314. Springer, 1976.

B. Laurent. Efficient estimation of integral functionals of a density. *Ann. Statist.*, 24: 659–681, 1996.

B. Laurent. Estimation of integral functionals of a density and its derivatives. *Bernoulli*, 3:181–211, 1997.

B.Y. Levit. Asymptotically efficient estimation of nonlinear functionals. *Problems Inform. Transmission*, 14:204–209, 1978.

N.A. Nadaraya. *Nonparametric Estimation of Probability Densities and Regression Curves*. Mathematics and Its Applications. Kluwer Academic Publishers, 1989.

I.F. Pinelis. Inequalities for sums of independent random vectors and their application to estimating a density. *Theory Probab. Appl.*, 35:605–607, 1990.

R.H. Randles and D.A. Wolfe. *Introduction to the Theory of Nonparametric Statistics*. Krieger Publishing Company, Malabar, Florida, 2nd edition, 1991.

G.R. Shorack. *Probability for Statisticians*. Springer, New York, 2000.

E.M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton, New Jersey, 1970.

B. von Bahr and C. Esseen. Inequalities for the r-th absolute moment of a sum of random variables. *Ann. Math. Statist.*, 36:299–303, 1965.

Author's address:

David M. Mason
Statistics Program
University of Delaware
206 Townsend Hall
Newark, Delaware 19716
USA

Tel. 1-302-831-1188
E-mail: davidm@udel.edu