IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 153–168

DOI: 10.1214/10-IMSCOLL716

Integral functionals of the density

David M. Mason*,[1] Elizbar Nadaraya[2] and Grigol Sokhadze[3]

University of Delaware and Tbilisi State University

Abstract: We show how a simple argument based on an inequality of McDiarmid yields strong consistency and central limit results for plug-in estimators of integral functionals of the density.

1. Introduction

Let X be a random variable with cumulative distribution function F having density f. Let us consider a general class of integral functionals of the form

$$T(f) = \int_{\mathbb{R}} \Phi\left(f^{(0)}(x), f^{(1)}(x), \ldots, f^{(k)}(x)\right) \mathrm{d}x, \tag{1.1}$$

with $k \geq 0$, where $f^{(0)} = f$ and $f^{(j)}$ denotes the jth derivative of f, for $j = 1, \ldots, k$, if $k \geq 1$, and Φ is a smooth function defined on $\mathbb{R}^{k+1}$. Under suitable regular conditions, which will be specified below, $T(f)$ is finite. Some special cases of (1.1) are

$$\text{(i)} \int_{\mathbb{R}} \phi\left(f(x)\right) f\left(x\right) \mathrm{d}x, \text{ (ii)} \int_{\mathbb{R}} \Phi\left(f(x)\right) \mathrm{d}x \text{ and (iii)} \int_{\mathbb{R}} \left(f^{(k)}(x)\right)^2 \mathrm{d}x. \tag{1.2}$$

The estimation of integral functionals of the density and its derivatives has been studied by a large number statisticians over many decades. Such integral functionals frequently arise in nonparametric procedures such as bandwidth selection in density estimation and in location and regression estimation using rank statistics. For good sources of references to current and past research literature in this area along with statistical applications consult Nadaraya [9], Levit [7], and Giné and Mason [5].

We shall be studying *plug-in estimators* of $T(f)$. These estimators are obtained by replacing $f^{(j)}$, for $j = 0, \ldots, k$, by kernel estimators based on a random sample of $X_1, \ldots, X_n$, $n \geq 1$, i.i.d. X, defined as follows. Let $K(\cdot)$ be a kernel defined on $\mathbb{R}$ with properties soon to be stated. For $h > 0$ and each $x \in \mathbb{R}$ define the function on $\mathbb{R}$

$$K_h(x - \cdot) = h^{-1} K\left((x - \cdot)/h\right).$$

[1]Statistics Program, University of Delaware, 206 Townsend Hall, Newark, DE 19716, USA, E-mail: davidm@udel.edu
[2] Department of Mathematical Statistics, Tbilisi State University 2, Universitet str., Tbilisi, 0143, Republic of Georgia, E-mail: elizbar.nadaraya@tsu.ge
[3] Department of Probability Theory and Mathematical Statistics, Tbilisi State University 2, Universitet str., Tbilisi, 0143, Republic of Georgia, E-mail: giasokhi1@i.ua
*Research partially an NSF Grant.
AMS 2000 subject classifications: Primary 62E05, 62E20; secondary 62F07.
Keywords and phrases: Integral functionals, kernel density estimators, inequalities, consistency and central limit theorem.

The kernel estimator of f based on $X_1, \ldots, X_n$, $n \geq 1$, and a sequence of positive constants $h = h_n$ converging to zero, is

$$\widehat{f}_{h_n}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - X_i), \text{ for } x \in \mathbb{R},$$

and the kernel estimator of $f^{(j)}$, for $j = 1, \ldots, k$, is

$$\widehat{f}_{h_n}^{(j)}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h_n}^{(j)}(x - X_i), \text{ for } x \in \mathbb{R},$$

where $K_{h_n}^{(j)}$ is the jth derivative of K_{h_n}. Note that $K_{h_n}^{(j)} = h_n^{-j-1} K^{(j)}$, where $K^{(j)}$ is the jth derivative of K. We shall often write $\widehat{f}_h^{(0)}(x) = \widehat{f}_h(x)$ and $K_h^{(0)}(x) = K_h(x)$. Also we denote the expectation of these estimators as

$$f_h^{(j)}(x) = E\widehat{f}_h^{(j)}(x), \text{ for } j = 0, \ldots, k. \tag{1.3}$$

Our plug-in estimator of $T(f)$ is

$$T(\widehat{f}_h) = \int_{\mathbb{R}} \Phi\left(\widehat{f}_h(x), \widehat{f}_h^{(1)}(x), \ldots, \widehat{f}_h^{(k)}(x)\right) \mathrm{d}x. \tag{1.4}$$

The goal of this paper is to show how a simple argument based on an inequality of McDiarmid yields a useful representation for $T(\widehat{f}_h)$. This means that it can be written as a sum of i.i.d. random variables plus a remainder term that converges to zero at a good stochastic rate. This will permit us to establish a nice strong consistency result and central limit theorem for $T(\widehat{f}_h)$. In the process we shall generalize and extend the results and methods of Mason [8] to multivariate integral functionals and estimators of the form (1.1) and (1.4). The [8] paper dealt solely with the special case in example (i).

In a paper closely related to this one, [5] investigated the Levit [7] estimator of integral functionals of the density:

$$\Theta(F) = \int_{\mathbb{R}} \varphi(x, F(x), f(x), \ldots, f^{(k)}(x))\, \mathrm{d}F(x), \tag{1.5}$$

which is formed by replacing in (1.5) the cumulative distribution function F by the empirical distribution function F_n and the $f^{(j)}$ by modified kernel estimators. They used very powerful U–statistics inequalities to obtain uniform in bandwidth type consistency and central limit results for the Levit estimator. These are results that hold uniformly in $a_n \leq h \leq b_n$, where a_n and b_n are suitable sequences of positive constants converging to zero.

With a lot more effort, we could derive analog results here for $T\left(\widehat{f}_h\right)$ using the methods in [5], as well as the modern empirical process tools developed in Einmahl and Mason [4] and Dony, Einmahl and Mason [2] in their work on uniform in bandwidth consistency of kernel type estimators. However, such an endeavour is well beyond the scope of the present paper. We should point out that one cannot extend our approach to handle the addition of x, F and F_n into $T(f)$ and $T\left(\widehat{f}_h\right)$ without imposing moment conditions on F and Φ. The reason is that one has to integrate with respect to dx instead of $dF(x)$.

Our representation theorem is stated and proved in Section 2. In Section 3 we use it to derive a strong consistency result and central limit theorem for $T\left(\widehat{f}_h\right)$. We conclude by applying our central limit theorem to the three examples in (1.2).

2. A representation theorem

Before we state our representation theorem, we shall gather together our basic assumptions along with some of their implications that will be used throughout this paper.

Assumptions on the density f.

(F.i) The density function f is continuously differentiable up to order $k \geq 1$, if $k \geq 1$.

(F.ii) For some constant $M > 0$, $\sup_{x \in \mathbb{R}} \left| f^{(j)}(x) \right| \leq M$ for $j = 0, \ldots, k$.

(F.iii) For each $j = 0, \ldots, k$, $f^{(j)} \in L_1(\mathbb{R})$.

Assumptions on the kernel K.

(K.i) $\int_{\mathbb{R}} |K|(x)\,\mathrm{d}x = \kappa < \infty$.

(K.ii) $\int_{\mathbb{R}} K(x)\,\mathrm{d}x = 1$.

(K.iii) The kernel K is $k+1$-times continuously differentiable.

(K.iv) For some $D > 0$, $\sup_{x \in \mathbb{R}} |K^{(j)}(x)| \leq D < \infty$, $j = 0, \ldots, k+1$.

(K.v) For each $j = 0, \ldots, k$, $\lim_{|x| \to \infty} K^{(j)}(x) = 0$ and $K^{(j)} \in L_1(\mathbb{R})$.

We shall repeatedly use the fact following by integration by parts that under our assumptions on f and K, that for $j = 0, \ldots, k$,

(2.1)
$$f_h^{(j)}(x) = h^{-j-1} \int_{\mathbb{R}} K^{(j)}\left(\frac{x-y}{h}\right) f(y)\,\mathrm{d}y = h^{-1} \int_{\mathbb{R}} K\left(\frac{x-y}{h}\right) f^{(j)}(y)\,\mathrm{d}y.$$

For $j = 0, 1, \ldots, k$, set

$$g_{n,h}^{(j)}(x) = h^j \widehat{f}_h^{(j)}(x) = \frac{1}{nh} \sum_{i=1}^{n} K^{(j)}((x - X_i)/h).$$

Our assumptions on f and K permit us to apply Theorem 2 of [2] to get for some $h_0 > 0$, every $c > 0$ and each $j = 0, 1, \ldots, k$, with probability 1,

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \leq h \leq h_0} \sup_{x \in \mathbb{R}} \frac{\sqrt{nh}\left| g_{n,h}^{(j)}(x) - E g_{n,h}^{(j)}(x) \right|}{\sqrt{|\log h| \vee \log\log n}} =: G_j(c) < \infty. \tag{2.2}$$

This implies that as long as h_n converges to zero at a rate such that $h_n \geq (c \log n)/n$ for some $c > 0$, for each $j = 0, 1, \ldots, k$, with probability 1,

$$\sup_{x \in \mathbb{R}} \left| \widehat{f}_{h_n}^{(j)}(x) - f_{h_n}^{(j)}(x) \right| = O\left(\frac{\sqrt{|\log h_n| \vee \log\log n}}{\sqrt{n} h_n^{1/2+j}} \right). \tag{2.3}$$

To see this, notice that

$$h^{-j} E g_{n,h}^{(j)}(x) = f_h^{(j)}(x) = \int_{\mathbb{R}} h^{-j-1} K^{(j)}\left(\frac{x-y}{h}\right) f(y)\,\mathrm{d}y,$$

where $f_h^{(j)}(x)$ is as in (1.3). Now by applying the formula (2.1) we get

$$f_h^{(j)}(x) = h^{-1} \int_{\mathbb{R}} K\left(\frac{x-y}{h}\right) f^{(j)}(y) \, \mathrm{d}y,$$

which, in turn, by the change of variables $v = \frac{x-y}{h}$ or $y = x - hv$

$$= \int_{\mathbb{R}} K(v) f^{(j)}(x - hv) \, \mathrm{d}v. \tag{2.4}$$

From (2.4) we get via (K.i) and (F.ii) that

$$\sup_{x \in \mathbb{R}} \left| f_h^{(j)}(x) \right| \le \kappa M, \; 0 \le j \le k. \tag{2.5}$$

Therefore as long as

$$\sqrt{|\log h_n| \vee \log\log n}/(\sqrt{n} h_n^{1/2+k}) \to 0, \text{ as } n \to \infty, \tag{2.6}$$

we can infer from (2.3) that with probability 1 for all large enough n

$$\left\{ \left(\widehat{f}_h(x), \widehat{f}_h^{(1)}(x), \ldots, \widehat{f}_h^{(k)}(x) \right) : x \in \mathbb{R} \right\} \subset C, \tag{2.7}$$

where C is any open convex set such that

$$[-\kappa M, \kappa M]^{k+1} \subset C. \tag{2.8}$$

Assumptions on Φ

(Φ.i) $\Phi(0, \ldots, 0) = 0$.

(Φ.ii) The function Φ possesses all derivatives up to second order on an open convex set C containing $[-\kappa M, \kappa M]^{k+1}$.

(Φ.iii) The second order derivatives of Φ are uniformly bounded on C by a constant $B_\Phi > 0$.

For $j = 0, \ldots, k$, let

$$\Phi_j(y_0, y_1, \ldots, y_k) = \frac{\partial \Phi(y_0, y_1, \ldots, y_k)}{\partial y_j}; \tag{2.9}$$

and for $0 \le i, j \le k$ set

$$\Phi_{i,j}(y_0, y_1, \ldots, y_k) = \frac{\partial^2 \Phi(y_0, y_1, \ldots, y_k)}{\partial y_i \partial y_j}.$$

Our assumptions on Φ say that for all $0 \le i, j \le k$,

$$\sup \{ |\Phi_{i,j}| (y_0, y_1, \ldots, y_k) : (y_0, y_1, \ldots, y_k) \in C \} \le B_\Phi. \tag{2.10}$$

We shall first verify that $T(f)$, $T(f_h)$ *and* $T\left(\widehat{f}_h\right)$ *are finite.*

Notice that by Taylor's theorem for each $(y_0, y_1, \ldots, y_k) \in C$ for some $\widetilde{y}_k \in C$

$$|\Phi|(y_0, y_1, \ldots, y_k) = \left| \sum_{j=0}^{k} \Phi_j(0, 0, \ldots, 0) y_j + \frac{1}{2} \sum_{i,j=0}^{k} \int_{\mathbb{R}} \Phi_{i,j}(\widetilde{y}_k) y_i y_j \right|,$$

which for some constant A_Φ is

$$\leq A_\Phi \left(\sum_{j=0}^{k} |y_j| + \sum_{j=0}^{k} |y_j|^2 \right).$$

This implies using (2.10) that for any $k+1$ bounded measurable functions $\varphi_0, \ldots, \varphi_k$ in $L_1(\mathbb{R})$ taking values in C,

$$\int_{\mathbb{R}} |\Phi| (\varphi_0(x), \varphi_1(x), \ldots, \varphi_k(x)) \, \mathrm{d}x < \infty. \tag{2.11}$$

From the assumptions on f and K we can easily infer that the functions $f^{(j)}$ and $f_h^{(j)}$, $j = 0, \ldots, k$ are bounded and in $L_1(\mathbb{R})$. This when combined with (2.5) and (2.11) implies that both $T(f)$ and $T(f_{h_n})$ are finite. Similarly, the assumptions on K imply that each $\widehat{f}_h^{(j)}$ is bounded and in $L_1(\mathbb{R})$, which in combination with (2.7) and (2.11) gives, with probability 1, that the estimator $T(\widehat{f}_{h_n})$ is finite for all n sufficiently large.

Next we shall represent the difference $T(\widehat{f}_{h_n}) - T(f_{h_n})$ as a sum of i.i.d. random variables $S_n(h_n)$ plus a remainder term R_n. By Taylor's formula we can write

$$T(\widehat{f}_{h_n}) - T(f_{h_n}) = S_n(h_n) + R_n, \tag{2.12}$$

where for any $h > 0$, $S_n(h)$ is the sum of i.i.d. random variables

$$S_n(h) = \sum_{j=0}^{k} \int_{\mathbb{R}} \Phi_j(f_h(x), f_h^{(1)}(x), \ldots, f_h^{(k)}(x)) \left(\widehat{f}_h^{(j)}(x) - f_h^{(j)}(x) \right) \mathrm{d}x; \tag{2.13}$$

and R_n is the remainder term

$$R_n = \frac{1}{2} \sum_{i,j=0}^{k} \int_{\mathbb{R}} \Phi_{i,j}(\widetilde{y}_k(x)) \left(\widehat{f}_{h_n}^{(i)}(x) - f_{h_n}^{(i)}(x) \right) \left(\widehat{f}_{h_n}^{(j)}(x) - f_{h_n}^{(j)}(x) \right) \mathrm{d}x, \tag{2.14}$$

with $\widetilde{y}_k(x)$ on the line joining

$$(f_h(x), f_h^{(1)}(x), \ldots, f_h^{(k)}(x)) \text{ and } (\widehat{f}_h(x), \widehat{f}_h^{(1)}(x), \ldots, \widehat{f}_h^{(k)}(x)).$$

Here is our representation theorem. It determines the size of the stochastic remainder term R_n in the representation (2.12). Our consistency result and central limit theorem for $T\left(\widehat{f}_h\right)$ will follow from it.

Theorem 2.1. *Assume the above conditions on the density f, the kernel K and the function Φ. Then for any positive sequence $h = h_n \leq 1$ converging to zero at the rate (2.6) the remainder term in the representation (2.12) satisfies, with probability* 1,

$$R_n = O\left(\log n / (n h_n^{2k+1})\right). \tag{2.15}$$

Moreover,

$$R_n = O_p\left(1/(n h_n^{2k+1})\right). \tag{2.16}$$

Remark 2.2. *We call (2.15) a strong representation and (2.16) a weak representation.*

Proof of Theorem 2.1. Applying standard inequalities, we get from (2.10), (2.5) and (2.7) that for some $C_\Phi > 0$, with probability 1 for all large n,

$$|R_n| \leq C_\Phi \int_{\mathbb{R}} \sum_{j=0}^{k} \left(\widehat{f}_{h_n}^{(j)}(x) - f_{h_n}^{(j)}(x) \right)^2 \mathrm{d}x. \tag{2.17}$$

Let W_k be the Sobolev space of functions g having continuous derivatives of order up to $k \geq 1$, each in $L_2(\mathbb{R})$, with the Sobolev norm

$$\|g\|_k = \sqrt{\sum_{j=0}^{k} \int_{\mathbb{R}} |g^{(j)}(x)|^2 \, \mathrm{d}x.}$$

The space W_k has the inner product

$$\langle g_1, g_2 \rangle_k = \sum_{j=0}^{k} \int_{\mathbb{R}} g_1^{(j)}(x) g_2^{(j)}(x) \, \mathrm{d}x.$$

Set $r_n(k) = \|\widehat{f}_{h_n} - f_{h_n}\|_k^2$. We see that with this notation, $|R_n| \leq C_\Phi r_n(k)$. Next set

$$Y_i = Y_i(x) = \frac{1}{n} \left\{ K_{h_n}(x - X_i) - f_{h_n}(x) \right\},$$

where $f_{h_n}(x) = E\widehat{f}_{h_n}(x)$. Then

$$\sum_{i=1}^{n} Y_i(x) = \frac{1}{n} \sum_{i=1}^{n} \{K_{h_n}(x - X_i) - f_{h_n}(x)\} = \widehat{f}_{h_n}(x) - f_{h_n}(x).$$

Therefore

$$r_n(k) = \left\| \sum_{i=1}^{n} Y_i \right\|_k^2. \tag{2.18}$$

Let us now estimate the $\|\cdot\|_k$ norm of the function $g_i = g_i(x) = \frac{1}{n} K_{h_n}(x - X_i)$ for each $i = 1, \ldots, n$. We have

$$\|g_i\|_k = \left(\sum_{j=0}^{k} \frac{1}{n^2} \int_{\mathbb{R}} \left(K_{h_n}^{(j)}(x - X_i) \right)^2 \mathrm{d}x \right)^{1/2}$$

$$= \left(\frac{1}{n^2} \sum_{j=0}^{k} \int_{\mathbb{R}} \left(\frac{1}{h_n^{j+1}} K^{(j)} \left(\frac{x - X_i}{h_n} \right) \right)^2 \mathrm{d}x \right)^{1/2}$$

$$= \frac{1}{n} \left(\sum_{j=0}^{k} \frac{1}{h_n^{2j+1}} \int_{\mathbb{R}} \left(K^{(j)} \left(\frac{x - X_i}{h_n} \right) \right)^2 \mathrm{d}\, \frac{x - X_i}{h_n} \right)^{1/2}$$

$$\leq \left(\sum_{j=0}^{k} \int_{\mathbb{R}} \left(K^{(j)}(u) \right)^2 \mathrm{d}u \right)^{1/2} \Big/ \left(n \sqrt{h_n^{2k+1}} \right).$$

Therefore

$$\|g_i\|_k \le \|K\|_k \Big/ \left(n\sqrt{h_n^{2k+1}}\right) =: D_n/2. \tag{2.19}$$

Note that (K.iv) and (K.v) imply that $\|K\|_k^2$ is finite. Observe that (2.19) yields the bound,

$$\|Y_i\|_k \le \|g_i\|_k + E\|g_i\|_k \le D_n. \tag{2.20}$$

We shall control the size of $r_n(k)$ using McDiarmid's inequality, which for convenience we state here.

McDiarmid's inequality (See Devroye [1]) *Let $Y_1, \ldots, Y_n$ be independent random variables taking values in a set A and assume that the function $H : A^n \to \mathbb{R}$, satisfies for each $i = 1, \ldots, n$ and some c_i,*

$$\sup_{y_1,\ldots,y_n,y,\in A} |H(y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_n) - H(y_1, \ldots, y_{i-1}, y, y_{i+1}, \ldots, y_n)| \le c_i.$$

then for every $t > 0$,

$$P\left\{|H(Y_1, \ldots, Y_n) - EH(Y_1, \ldots, Y_n)| \ge t\right\} \le 2\exp\left(-2t^2 \Big/ \sum_{i=1}^n c_i^2\right).$$

Applying McDiarmid's inequality, in our situation, with

$$H(Y_1, \ldots, Y_n) = \left\|\sum_{i=1}^n Y_i\right\|_k$$

and $c_i = 2D_n$, for $i = 1, \ldots, n$, which comes from (2.20), we obtain for every $t > 0$,

$$P\left\{\left|\left\|\sum_{i=1}^n Y_i\right\|_k - E\left\|\sum_{i=1}^n Y_i\right\|_k\right| \ge t\right\} \le 2\exp\left(-\frac{t^2 n h_n^{2k+1}}{2\|K\|_k^2}\right). \tag{2.21}$$

Setting $t = 2\sqrt{\log n}/\sqrt{nh_n^{2k+1}}$ into the probability bound in (2.21), we get via the Borel–Cantelli lemma that with probability 1,

$$\left\|\sum_{i=1}^n Y_i\right\|_k = E\left\|\sum_{i=1}^n Y_i\right\|_k + O\left(\frac{\sqrt{\log n}}{\sqrt{nh_n^{2k+1}}}\right). \tag{2.22}$$

Furthermore, by Jensen's inequality,

$$\left(E\left\|\sum_{i=1}^n Y_i\right\|_k\right)^2 \le E\left\|\sum_{i=1}^n Y_i\right\|_k^2 = \sum_{i=1}^n \sum_{j=0}^k E\int_{\mathbb{R}} \left(Y_i^{(j)}(x)\right)^2 \mathrm{d}x,$$

that is,

$$\left(E\left\|\sum_{i=1}^n Y_i\right\|_k\right)^2 \le \frac{1}{n^2}\sum_{i=1}^n \sum_{j=0}^k \int_{\mathbb{R}} E\left\{K_h^{(j)}(x - X_i) - f_h^{(j)}(x)\right\}^2 \mathrm{d}x$$

$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{k}\int_{\mathbb{R}} E\left(\frac{1}{h_n^{j+1}}K^{(j)}\left(\frac{x-X_i}{h_n}\right)\right)^2 \mathrm{d}x$$

$$\leq \frac{1}{n^2 h_n^{2k+2}}\sum_{i=1}^{n}\sum_{j=0}^{k}\int_{\mathbb{R}} E\left(K^{(j)}\left(\frac{x-X_i}{h_n}\right)\right)^2 \mathrm{d}x$$

$$= \frac{1}{n^2 h_n^{2k+2}}\sum_{i=1}^{n}\sum_{j=0}^{k}\int_{\mathbb{R}^2} \left(K^{(j)}\right)^2\left(\frac{x-y}{h_n}\right) f(y)\,\mathrm{d}y\mathrm{d}x,$$

which by using Fubini's theorem is seen to

$$= \|K\|_k^2/\left(nh_n^{2k+1}\right). \tag{2.23}$$

From (2.18), (2.22) and (2.23) we conclude for any positive sequence $h = h_n$ converging to zero at the rate (2.6) that $R_n = O\left(\log n/\left(nh_n^{2k+1}\right)\right)$, a.s.

The proof of (2.18) follows similar lines. Therefore we have proved our main result. □

3. Applications of the representation theorem

3.1. Consistency

As our first application of Theorem 2.1 we shall establish a strong consistency result for $T\left(\widehat{f}_h\right)$.

Theorem 3.1. *Assume the conditions of Theorem 2.1. If a positive sequence $h = h_n \leq 1$ is chosen so that*

$$\log n/\left(nh_n^{2k+1}\right) \to 0, \tag{3.1}$$

then with probability 1, we have, as $n \to \infty$,

$$T(\widehat{f}_{h_n}) \to T(f). \tag{3.2}$$

Proof of Theorem 3.1. First, by Theorem 2.1 and (3.1),

$$T(\widehat{f}_{h_n}) - T(f_{h_n}) = S_n(h_n) + R_n \ \text{ with } \ R_n = o(1), \text{ a.s.} \tag{3.3}$$

Let $X_1, \ldots, X_n$ be i.i.d. with density f. Recall the definition of Φ_j in (2.9) and set for $i = 1, \ldots, n$,

$$Z_i\left(h_n\right) := \sum_{j=0}^{k}\int_{\mathbb{R}} \Phi_j(f_{h_n}(x), f_{h_n}^{(1)}(x), \ldots, f_{h_n}^{(k)}(x))K_{h_n}^{(j)}\left(x - X_i\right)\mathrm{d}x. \tag{3.4}$$

and for future reference write for any $h > 0$ and X with density f,

$$Z\left(h\right) := \sum_{j=0}^{k}\int_{\mathbb{R}} \Phi_j(f_h(x), f_h^{(1)}(x), \ldots, f_h^{(k)}(x))K_h^{(j)}\left(x - X\right)\mathrm{d}x. \tag{3.5}$$

In this notation we can write

$$S_n(h_n) = n^{-1} \sum_{i=1}^{n} \{Z_i(h_n) - EZ_i(h_n)\}. \tag{3.6}$$

Keeping in mind that (2.5) implies

$$\left\{\left(f_h(x), f_h^{(1)}(x), \ldots, f_h^{(k)}(x)\right) : x \in \mathbb{R}\right\} \subset [-\kappa M, \kappa M]^{k+1} \tag{3.7}$$

and that we can infer from the assumptions on Φ that for some $D_\Phi > 0$,

$$\sup\left\{|\Phi_j|(y_0, y_1, \ldots, y_k) : (y_0, y_1, \ldots, y_k) \in [-\kappa M, \kappa M]^{k+1}\right\} \le D_\Phi$$

we get that for $1 \le i \le n$,

$$|Z_i(h_n)| \le \sum_{j=0}^{k} \int_{\mathbb{R}} |\Phi_j|(f_{h_n}(x), f_{h_n}^{(1)}(x), \ldots, f_{h_n}^{(k)}(x)) \left|K_{h_n}^{(j)}\right|(x - X_i)\,\mathrm{d}x$$

$$\le D_\Phi \sum_{j=0}^{k} \int_{\mathbb{R}} \left|K_{h_n}^{(j)}\right|(x - X_i)\,\mathrm{d}x = D_\Phi \sum_{j=0}^{k} h_n^{-j-1} \int_{\mathbb{R}} \left|K^{(j)}\right|\left(\frac{x - X_i}{h_n}\right)\mathrm{d}x$$

$$= D_\Phi \sum_{j=0}^{k} h_n^{-j} \int_{\mathbb{R}} \left|K^{(j)}\right|(u)\,\mathrm{d}u \le L h_n^{-k}$$

for some $L > 0$. Therefore we can apply Hoeffding's inequality [6] to get,

$$P\left\{|S_n(h_n)| > \frac{2\sqrt{\log n}L}{\sqrt{n}h_n^k}\right\} \le 2\exp(-2\log n),$$

from which we readily conclude using the Borel–Cantelli lemma that, with probability 1,

$$S_n(h_n) = O\left(\sqrt{\log n/(nh_n^{2k})}\right). \tag{3.8}$$

Thus whenever $\sqrt{\frac{\log n}{nh_n^{2k}}} = o(1)$, then, with probability 1,

$$S_n(h_n) = o(1). \tag{3.9}$$

Next we shall show that $T(f_h) \to T(f)$. Recall by (2.4), for each $j = 0, \ldots, k$,

$$f_{h_n}^{(j)}(x) = \int_{\mathbb{R}} K(v) f^{(j)}(x - h_n v)\,\mathrm{d}v,$$

which by (F.ii), (K.i) and the dominated convergence theorem implies that for each $j = 0, \ldots, k$,

$$f_{h_n}^{(j)}(x) \to f^{(j)}(x) \text{ for a.e. } x \in \mathbb{R}.$$

Thus for a.e. $x \in \mathbb{R}$, as $\to \infty$,

$$\Phi(f_{h_n}(x), f_{h_n}^{(1)}(x), \ldots, f_{h_n}^{(k)}(x)) \to \Phi(f(x), f^{(1)}(x), \ldots, f^{(k)}(x)). \tag{3.10}$$

Write for each $j = 0, \ldots, k$,

$$g_{h_n}^{(j)}(x) = \int_{\mathbb{R}} |K| \ (v) \left| f^{(j)} \right| (x - hv) \, \mathrm{d}v \text{ and } g^{(j)} = \kappa \left| f^{(j)} \right|,$$

where κ is as in (K.i). Clearly for each $j = 0, \ldots, k$, $\left| f_{h_n}^{(j)} \right| \leq g_{h_n}^{(j)}$, and

$$g_{h_n}^{(j)}(x) \to g^{(j)}(x) \text{ for a.e. } x \in \mathbb{R}. \tag{3.11}$$

Notice that for each $n \geq 1$ and $j = 0, \ldots, k$,

$$\int_{\mathbb{R}} g_{h_n}^{(j)} (x) \, dx = \int_{\mathbb{R}} \left| g^{(j)} \right| (x) \, \mathrm{d}x. \tag{3.12}$$

Also since $\Phi(0, \ldots, 0) = 0$ and Φ is assumed to be differential with continuous derivatives Φ_j on C, where C satisfies (2.8), we get by (3.7) and the mean value theorem that for some $M_\Phi > 0$,

$$|\Phi| \left(f_{h_n}(x), f_{h_n}^{(1)}(x), \ldots, f_{h_n}^{(k)}(x) \right) \leq M_\Phi \sum_{j=0}^{k} g_{h_n}^{(j)}(x), \text{ for all } x \in \mathbb{R}. \tag{3.13}$$

From (3.10), (3.11), (3.12) and (3.13), we readily that as $n \to \infty$,

$$T(f_h) = \int_{\mathbb{R}} \Phi(f_{h_n}(x), f_{h_n}^{(1)}(x), \ldots, f_{h_n}^{(k)}(x)) \, \mathrm{d}x \to T(f),$$

using a standard convergence result that is stated, for instance, as problem 12 on p. 102 of Dudley [3]. It says that if f_n and g_n are integrable functions for a measure μ with $|f_n| \leq g_n$, such that as $n \to \infty$, $f_n(x) \to f(x)$ and $g_n(x) \to g(x)$ for almost all x. Then $\int g_n \, \mathrm{d}\mu \to \int g \, \mathrm{d}\mu < \infty$, implies that $\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu$.

Therefore whenever $T\left(\widehat{f}_h\right) - T(f_h) = o(1)$ a.s., we have

$$T\left(\widehat{f}_h\right) - T(f) \to 0 \text{ a.s.} \tag{3.14}$$

Now (3.1), i. e., $\frac{\log n}{n h_n^{2k+1}} \to 0$, implies $\frac{\log n}{n h_n^{2k}} \to 0$. Thus both (3.3) and (3.9) hold, which imply (3.14). □

Remark 3.2. *In the case $k = 0$, Theorem 3.1 generalizes the first part of Theorem 2 in [8] from $k = 0$ to $k \geq 0$ and to a larger class of functions Φ. Moreover, the proof of Theorem 3.1 completes that of the first part of Theorem 2 of [8]. A final easy step showing that $T(f_h) \to T(f)$ is missing there.*

3.2. Central limit theorem

In this section we shall use Theorem 2.1 to establish a central limit theorem for $T\left(\widehat{f}_h\right)$. Before stating and proving our result, we must first introduce some additional assumptions and then derive a limiting variance needed in its formulation.

Assumptions on the density f.

(F.iv) Assume that for some $0 < M < \infty$, $|f(x)| \le M$ for $x \in \mathbb{R}$, and if $k \ge 1$ then f is $2k$-times continuously differentiable and its derivatives $f^{(j)}$ satisfy for $x \in \mathbb{R}$, $|f^{(j)}(x)| \le M < \infty$, $j = 1, \ldots, 2k$.

Assumptions on the kernel K.

We assume conditions (K.i)-(K.v) on the kernel.

Assumptions on Φ.

$\Phi : \mathbb{R}^{k+1} \to \mathbb{R}$, $k \ge 0$, such that $\Phi(0, \ldots, 0) = 0$ and all of its partial derivatives in $y_0, \ldots, y_k$,

$$\frac{\partial^{m_0}}{\partial y_0^{m_0}} \cdots \frac{\partial^{m_k}}{\partial y_k^{m_0}} \Phi(y_0, \ldots, y_k),$$

where

$$m_0 + m_1 + \cdots + m_k = j,\ 0 \le j \le k+1,$$

are continuous on an open convex set C containing $[-\kappa M, \kappa M]^{k+1}$ and they are uniformly bounded on C by a constant $B_\Phi > 0$.

Preliminaries to calculating a variance

Let p and q be m times continuously differentiable functions such that for each $0 \le j \le m$

$$\lim_{v \to \infty} p^{(j)}(\pm v)\, q^{(m-j)}(\pm v) = 0. \tag{3.15}$$

We shall be use the formula following from integration by parts and (3.15):

$$\int_{\mathbb{R}} p^{(m)}(v)\, q(v)\, \mathrm{d}v = (-1)^m \int_{\mathbb{R}} p(v)\, q^{(m)}(v)\, \mathrm{d}v. \tag{3.16}$$

Set

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

which by the change of variable $v = \frac{y-t}{h} + u$ or $t = y + h(u - v)$

$$= \int_{\mathbb{R}} h^{-j} K^{(j)}(v)\, f(y + h(u-v))\, \mathrm{d}v.$$

Applying, in turn, the formula (3.16) we get

$$f_h^{(j)}(y + hu) = \int_{\mathbb{R}} K(v)\, f^{(j)}(y + h(u-v))\, \mathrm{d}v. \tag{3.17}$$

Notice from (3.17), (F.iv) and (K.i), we get from the bounded convergence theorem that for every $0 \le j \le 2k$ and a.e. $y \in \mathbb{R}$

$$f_h^{(j)}(y + hu) \to f^{(j)}(y). \tag{3.18}$$

Let Ψ be a function from $\mathbb{R}^{k+1} \to \mathbb{R}$ satisfying the assumptions on Φ and set

$$\Psi\left(y\right)=\Psi\left(f(y),f^{(1)}(y),\ldots,f^{(k)}(y)\right) \tag{3.19}$$

and

$$\Psi\left(y,h\right)=\Psi\left(f_h(y),f_h^{(1)}(y),\ldots,f_h^{(k)}(y)\right). \tag{3.20}$$

Notice that we have

$$\Psi\left(y+hu,h\right)=\Psi\left(f_h(y+hu),f_h^{(1)}(y+hu),\ldots,f_h^{(k)}(y+hu)\right). \tag{3.21}$$

Clearly by (3.18), $\Psi\left(y+hu,h\right)\to\Psi\left(y\right)$. Let for $j=0,\ldots,k$,

$$\Psi_j\left(y_0,y_1,\ldots,y_k\right)=\frac{\partial\Psi(y_0,y_1,\ldots,y_k)}{\partial y_j}.$$

Further set for $j=0,\ldots,k$,

$$\Psi_j\left(y\right)=\Psi_j\left(f(y),f^{(1)}(y),\ldots,f^{(k)}(y)\right)$$

and

$$\Psi_j\left(y,h\right)=\Psi_j\left(f_h(y),f_h^{(1)}(y),\ldots,f_h^{(k)}(y)\right).$$

Note that we have

$$\Psi_j\left(y+hu,h\right)=\Psi_j\left(f_h(y+hu),f_h^{(1)}(y+hu),\ldots,f_h^{(k)}(y+hu)\right).$$

We see that

$$\frac{d\Psi\left(y+hu,h\right)}{du}=h\left(\sum_{j=0}^{k}\Psi_j\left(y+hu,h\right)f_h^{(j+1)}(y+hu)\right).$$

Write

$$\Psi^{(1)}\left(y_0,y_1,\ldots,y_{k+1}\right)=\sum_{j=0}^{k}\Psi_j\left(y_0,y_1,\ldots,y_k\right)y_{j+1},$$

and observe that

$$\Psi^{(1)}\left(y\right):=\frac{d}{dy}\Psi\left(y\right)=\Psi^{(1)}\left(f\left(y\right),\ldots,f^{(k+1)}\left(y\right)\right).$$

We see that

$$\frac{d\Psi\left(y+hu,h\right)}{du}=h\Psi^{(1)}\left(y+hu,h\right),$$

where

$$\Psi^{(1)}\left(y+h,h\right)=\Psi^{(1)}\left(f_h\left(y+hu\right),\ldots,f_h^{(k+1)}\left(y+hu\right)\right).$$

We shall write

$$\Psi^{(1)}\left(y,h\right)=\Psi^{(1)}\left(f_h\left(y\right),\ldots,f_h^{(k+1)}\left(y\right)\right).$$

Now for $m\geq 1$ set

$$\Psi_j^{(m-1)}\left(y_0,y_1,\ldots,y_{k+m-1}\right)=\frac{d}{dy_j}\Psi^{(m-1)}\left(y_0,y_1,\ldots,y_{k+m-1}\right),\ 0\leq j\leq k+m-1.$$

Here $\Psi^{(0)} = \Psi$ and $\Psi_j^{(0)} = \Psi_j$. Also let

$$\Psi^{(m)}(y_0, y_1, \ldots, y_{k+m}) = \sum_{j=0}^{k+m-1} \Psi_j^{(m-1)}(y_0, y_1, \ldots, y_{k+m-1})\, y_{j+1},$$

and note that

$$\Psi^{(m)}(y) := \frac{d^m}{dy^m}\Psi(y) = \Psi^{(m)}\left(f(y), \ldots, f^{(k+m)}(y)\right).$$

Set

$$\Psi^{(m)}(y+h, h) = \Psi^{(m)}\left(f_h(y+hu), \ldots, f_h^{(k+m)}(y+hu)\right)$$

and

$$\Psi^{(m)}(y, h) = \Psi^{(m)}\left(f_h(y), \ldots, f_h^{(k+m)}(y)\right).$$

We readily get that

$$\frac{d^m \Psi(y+hu, h)}{du^m} = h^m \Psi^{(m)}(y+hu, h) \tag{3.22}$$

and, as $h \searrow 0$,

$$h^{-m}\frac{d^m \Psi(y+hu, h)}{du^m} = \Psi^{(m)}(y+hu, h) \to \Psi^{(m)}(y) = \frac{d^m}{dy^m}\Psi(y). \tag{3.23}$$

Computation of limit variance

We are now prepared to compute our limiting variance. Let $\Phi_j(x)$ and $\Phi_j(x,h)$ be defined exactly as $\Psi_j(x)$ and $\Psi_j(x,h)$. Recall the definition of $S_n(h)$ in (2.13) and that of $Z(h)$ in (3.5). We can write

$$S_n(h) = \sum_{j=0}^{k} \int_{\mathbb{R}} \Phi_j(x,h)\left(\widehat{f}_h^{(j)}(x) - f_h^{(j)}(x)\right) \mathrm{d}x.$$

and

$$Z(h) = \sum_{j=0}^{k} \int_{\mathbb{R}} \Phi_j(x,h)\, h^{-j-1} K^{(j)}\left(\frac{x-X}{h}\right) \mathrm{d}x.$$

Thus we see that if $Z_1(h), \ldots, Z_n(h)$ are i.i.d. $Z(h)$, then

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

Now

$$EZ(h) = \sum_{j=0}^{k} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \Phi_j(x,h)\, h^{-j-1} K^{(j)}\left(\frac{x-y}{h}\right) \mathrm{d}x\right] f(y)\,\mathrm{d}y$$

$$= \sum_{j=0}^{k} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \Phi_j(y+hu,h)\, h^{-j} K^{(j)}(u)\,\mathrm{d}u\right] f(y)\,\mathrm{d}y. \tag{3.24}$$

Note that we get from (3.16) and (3.22), the identity

$$\int_{\mathbb{R}} \Phi_j\left(y+hu,h\right)h^{-j}K^{(j)}\left(u\right)\mathrm{d}u = (-1)^j\int_{\mathbb{R}}\Phi_j^{(j)}(y+hu,h)K\left(u\right)\mathrm{d}u, \tag{3.25}$$

and from (3.23) we conclude that for a.e. $y\in\mathbb{R}$ and all u, as $h\searrow 0$,

$$\Phi_j^{(j)}(y+hu,h)\to\frac{\mathrm{d}^j}{\mathrm{d}y^j}\Phi_j\left(f(y),f_h^{(1)}(y),\ldots,f_h^{(k)}(y)\right) =: \Phi_j^{(j)}\left(y\right). \tag{3.26}$$

Set

$$\mu_k\left(y\right)=\sum_{j=0}^{k}\left(-1\right)^j\Phi_j^{(j)}\left(y\right). \tag{3.27}$$

Note that our assumptions imply that for some $B>0$ and all $h>0$ and $j=0,\ldots,k$,

$$\max_{0\le j\le k}\sup_{u,y}\left|\Phi_j^{(j)}\left(y+hu,h\right)\right|\le B \text{ and thus } \left|\Phi_j^{(j)}\left(y+hu,h\right)K(u)\right|\le B\left|K\right|(u).$$

Therefore by (3.26) and the dominated convergence theorem as $h\searrow 0$

$$H_h^{(j)}\left(y\right):=\int_{\mathbb{R}}\Phi_j^{(j)}\left(y+hu,h\right)K\left(u\right)\mathrm{d}u\to\Phi_j^{(j)}\left(y\right).$$

Now $\left|H_h^{(j)}\left(y\right)\right|\le B\int_{\mathbb{R}}|K|\left(u\right)\mathrm{d}u=B\kappa$. Hence by the bounded convergence theorem

$$\int_{\mathbb{R}}H_h^{(j)}\left(y\right)f\left(y\right)\mathrm{d}y\to\int_{\mathbb{R}}\Phi_j^{(j)}\left(y\right)f\left(y\right)\mathrm{d}y.$$

This of course implies that as $h\to 0$,

$$EZ\left(h\right)\to\int_{\mathbb{R}}\left\{\sum_{j=0}^{k}\left(-1\right)^j\Phi_j^{(j)}\left(y\right)\right\}f(y)\,\mathrm{d}y=\int_{\mathbb{R}}\mu_k\left(y\right)f(y)\,\mathrm{d}y=E\mu_k\left(X\right).$$

Next write for $0\le j,m\le k$,

$$\gamma_{j,m}\left(y\right)=\int_{\mathbb{R}^2}\Phi_j\left(x,h\right)\Phi_m\left(z,h\right)h^{-2-j-m}K^{(j)}\left(\frac{x-y}{h}\right)K^{(m)}\left(\frac{z-y}{h}\right)\mathrm{d}x\mathrm{d}z$$

$$=\int_{\mathbb{R}^2}\Phi_m\left(y+hu,h\right)\Phi_j\left(y+hv,h\right)h^{-j-m}K^{(j)}\left(u\right)K^{(m)}\left(v\right)\mathrm{d}v\mathrm{d}u.$$

Similarly we see that

$$E\gamma_{j,m}\left(X\right)\to\left(-1\right)^{m+j}\int_{\mathbb{R}}\Phi_m^{(m)}\left(y\right)\Phi_j^{(j)}\left(y\right)f\left(y\right)\mathrm{d}y.$$

Therefore since

$$EZ^2\left(h\right)=\sum_{j=0}^{k}\sum_{m=0}^{k}\int_{\mathbb{R}}\gamma_{j,m}\left(y\right)f\left(y\right)\mathrm{d}y,$$

we conclude that as $h\to 0$,

$$EZ^2\left(h\right)\to\sum_{j=0}^{k}\sum_{m=0}^{k}\int_{\mathbb{R}}\Phi_m^{(m)}\left(y\right)\Phi_j^{(j)}\left(y\right)\left(-1\right)^{m+j}f\left(y\right)\mathrm{d}y=E\mu_k^2\left(X\right).$$

Clearly the same proof shows that as $h \to 0, EZ^4(h) \to E\mu_k^4(X)$.

Also it is readily verified that $E\mu_k(X)$, $E\mu_k^2(X)$ and $E\mu_k^4(X)$ are finite under the conditions on Φ and f. In summary, we get

Lemma *Under the above assumptions for any sequence of positive numbers $h_n \to 0$, as $n \to \infty$,*

$$nVar(S_n(h_n)) = Var(Z(h_n)) \to Var(\mu_k(f(X))) =: \sigma^2(f) < \infty \tag{3.28}$$

and

$$EZ^4(h_n) \to E\mu_k^4(X) < \infty. \tag{3.29}$$

Part (2.16) of Theorem 2.1 and the above lemma, combined with Lyapunov's central limit theorem, yield the next result.

Theorem 3.3. *Under the above assumptions imposed in this subsection on the density f, the kernel K and function Φ, if a positive sequence $h = h_n \le 1$ is chosen so that $1/\left(\sqrt{n}h_n^{2k+1}\right) \to 0$ then*

$$\sqrt{n}\left\{T(\widehat{f}_h) - T(f_h)\right\} \to_d N\left(0, \sigma^2(f)\right). \tag{3.30}$$

In the next subsection, we shall discuss smoothness conditions that permit the replacement of $T(f_h)$ by $T(f)$ in (3.30).

3.3. Three examples of the application of Theorem 3.3

In this subsection we apply Theorem 3.3 to the three examples (i), (ii) and (iii) in (1.2). In the first two $k = 0$, so in addition to the smoothness conditions in our central limit theorem, we require $\sqrt{n}h_n \to \infty$. In example (i), $\mu_0(f(x)) = \Phi_1(f(x))$, where $\Phi_1(x) = \frac{d}{dx}(\phi(x)x)$, giving $\sigma^2(f) = Var(\Phi_1(f(X)))$. This matches with the second part of Theorem 2 of [8]. In example (ii), one gets that $\mu_0(f(x)) = \Phi'(f(x))$ and $\sigma^2(f) = Var(\Phi'(f(X)))$. This agrees with Theorem 3 of [5]. Note that example (i) is a special case of (ii). To apply Theorem 3.3 to example (iii) we must choose h_n such that $\sqrt{n}h_n^{k+1} \to \infty$. In this case $\mu_k(f(x)) = 2f^{(2k)}(x)$, and $\sigma^2(f) = Var\left(\left(2f^{(2k)}(X)\right)\right)$, which is in agreement with Theorem 4 of [5].

Let us now briefly discuss conditions under which we can replace $T(f_h)$ by $T(f)$ in (3.30). Towards this end, we cite here Proposition 1 of [5], which, in turn, was motivated by Proposition 1 of [7].

Proposition 3.4. *Assume that K is integrable, has compact support, and for some integer $s \ge 1$,*

$$\int_{\mathbb{R}} K(u)\,\mathrm{d}u = 1, \int_{\mathbb{R}} u^k K(u)\,\mathrm{d}u = 0 \text{ for } k = 1, \ldots, s, \tag{3.31}$$

and let H be a non-negative measurable function. Then there is a constant $C_K > 0$ such that, for every s times continuously differentiable function g satisfying for some $h_0 > 0$, $L_g > 0$, $0 < \alpha \le 1$,

$$\sup_{|h| \le h_0} |h|^{-\alpha}\left|g^{(s)}(x+h) - g^{(s)}(x)\right| =: L_g H(x), \text{ for every } x \in \mathbb{R}, \tag{3.32}$$

one has, for all $0 < h \le h_0$ *and every* $x \in \mathbb{R}$,

$$\left| \frac{1}{h} \int_{\mathbb{R}} g(u) K\left(\frac{x-u}{h}\right) \mathrm{d}u - g(x) \right| \le h^{s+\alpha} C_K L_g H(x). \tag{3.33}$$

Therefore if our kernel K also has compact support and satisfies (3.31) with $s = 1$ and our density f fulfills condition (3.32) with $\alpha = 1$ and $H \in L_1(\mathbb{R})$, then for all $h > 0$ small enough, for every Φ, which is Lipschitz on $[-\kappa M, \kappa M]$, there exists a constant $B > 0$ such that

$$\left| \int_{\mathbb{R}} \Phi(f_h(x))\, \mathrm{d}x - \int_{\mathbb{R}} \Phi(f(x))\, \mathrm{d}x \right| \le h^2 B \int_{\mathbb{R}} H(x)\, \mathrm{d}x.$$

Thus if we have both $\sqrt{n} h_n \to \infty$ and $\sqrt{n} h_n^2 \to 0$, we can conclude in examples (i) and (ii) that

$$\sqrt{n}\left(T\left(\widehat{f}_{h_n}\right) - T(f)\right) \to_d N\left(0, \sigma^2(f)\right). \tag{3.34}$$

We can also apply this proposition to example (iii) for any $k \ge 1$. Here, in order to be able to replace $T(f_h)$ by $T(f)$, we require that both $\sqrt{n} h_n^{2k+1} \to \infty$ and $\sqrt{n} h_n^{s+\alpha} \to 0$, where s and α satisfy (3.32) and (3.33).

Acknowledgements The authors thank the two referees for a careful reading of the manuscript.

References

[1] Devroye, L. (1991). Exponential inequalities in nonparametric estimation. *Nonparametric functional estimation and related topics* (*Spetses,* 1990), 31–44, *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.,* 335, *Kluwer Acad. Publ., Dordrecht.*

[2] Dony, J., Einmahl, U. and Mason, D. M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austrian J. Statistics* **35** 105–120.

[3] Dudley, R. M. (1989). *Real Analysis and Probability*, Chapman & Hall Mathematics Series, New York.

[4] Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403.

[5] Giné, E. and Mason, D. M.(2008) Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.* **35** 739–761.

[6] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** 13–30.

[7] Levit, B. Ya (1978). Asymptotically efficient estimation of nonlinear functionals. (Russian) *Problems Inform. Transmission* **14** 65–72.

[8] Mason, D. M. (2003). Representations for integral functionals of kernel density estimators. *Austrian J. Statistics* **32** 131–142.

[9] Nadaraya, E. A. (1989). Nonparametric estimation of probability densities and regression curves. (Translated from the Russian) *Mathematics and its Applications* (*Soviet Series*), 20. *Kluwer Academic Publishers Group, Dordrecht,* 1989; Russian original: *Tbilis. Gos. Univ., Tbilisi,* 1983.