Uniform in Bandwidth Estimation of Integral Functionals of the Density Function

EVARIST GINÉ and DAVID M. MASON

Department of Mathematics, University of Connecticut and Statistics Program, University of Delaware

ABSTRACT. We apply recent results on local U–statistcs to obtain uniform in bandwidth consistency and central limit theorems for some commonly used estimators of integral functionals of density functions.

key words: kernel density estimator, uniform in bandwidth, U–statistics

running headline: Estimation of integral functionals

1 Introduction

We shall study uniform in bandwidth consistency and central limit theorems [clt] for estimators of integral functionals of the density function. Here are some motivating examples.

$$T_1(F) = \int_{\mathbb{R}} (f'(x))^2 \, dx, \; T_2(F) = -\int_{\mathbb{R}} a(F(x)) f'(x) dx \text{ and } T_3(F) = \int_{\mathbb{R}} (f(x))^2 \, dx. \tag{1}$$

Note that the functional $T_3(F)$ is a special case of $T_2(F)$. The functionals $T_1(F)$ and $T_3(F)$ appear in plug-in data driven bandwidth selection procedures in density estimation (refer to Hall & Marron (1987) and the references therein) and the functional $T_2(F)$ arises as part of the variance in nonparametric location and regression estimation based on linear rank statistics (refer especially to Jurečková (1969)).

We shall consider the following general class of integral functionals of the density:

$$T(F) = \int_{\mathbb{R}} \varphi(x, F(x), F^{(1)}(x), ..., F^{(r)}(x)) \, dF(x), \tag{2}$$

where F is a cumulative distribution function on $\mathbb{R}$ with $r \geq 1$ derivatives $F^{(m)}$. At first sight, the first two examples in (1) are not of the form $T(F)$, however, under smoothness, after integrating by parts we can write

$$T_1(F) = -\int_{\mathbb{R}} f^{(2)}(x) f(x)\, dx \text{ and } T_2(F) = \int_{\mathbb{R}} a'(F(x)) f^2(x) dx.$$

Levit (1978), following Dmitriev *et al.* (1974), considers estimators of integral functionals of the form (1) by plugging in kernel density estimators for the $F^{(m)}$ in $T(F)$ and proves asymptotic normality as well as local asymptotic minimaxity with respect to power loss functions. Given that estimators of this type have optimality properties, it should be of interest to obtain their asymptotic normality uniformly in the bandwidth of the kernel estimators used to estimate the $F^{(m)}$. This would permit the insertion of bandwidth estimators into these kernel estimators and still have asymptotic normality of the resulting estimator of $T(F)$. This is the main goal of this article. We shall discuss in the next section what we mean by a *uniform in bandwidth clt* and its importance in the use of adaptive bandwidth estimators.

We shall then apply the general result (or its proof) to several examples from the literature (Bickel & Ritov (1988) -that we simplify along the lines of Giné & Nickl (2007b)-, Hall & Marron (1987), Grübel (1994), Cheng & Serfling (1981)). We shall also propose a practical criterion to adaptively choose the bandwidth. The main new ingredients relative to the work of these authors are the use of moment bounds for U-processes derived in Giné & Mason (2007a,b) and, for almost sure results, an exponential bound of Major (2006). For the reader's convenience, we collect the Giné & Mason (2007a,b) results in an Appendix

The estimation of integral functionals of a density, such as (2) or (1) is a subject of much current research. Aside from their practical value, these functionals are of theoretical interest in that they are tractable non linear functionals which have different behaviors according to the smoothness of the density. Aside from kernel based estimators such as those of Levit, Hall and Marron, Bickel and Ritov, etc., estimators based on L_2 expansions and projections, which satisfy the same optimality properties as those of Bickel & Ritov, have been considered by Birgé & Massart (1995), and Laurent (1997, 2005), among others. Estimators of this type are not treated in this article.

In Section 2 we shall establish a uniform in bandwidth central limit theorem for the class of estimators of (2). We will also indicate how to obtain laws of the iterated logarithm uniform in bandwidth for this estimator. In Sections 3 and 4 we shall consider several examples. Finally, in Section 5, we shall propose and study a way to adaptively choose bandwidths based on the results of Section 2.

Standard definitions and results from modern empirical process theory will be used freely. Unless otherwise indicated, these can be found in van der Vaart & Wellner (1996).

2 Uniform in bandwidth clt for the Levit estimator

We shall study the uniform in bandwidth behavior of estimators of $T(F)$ of the form (Levit (1978)):

$$T_n(\mathbf{h}) = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i, F_{n,i}(X_i), F^{(1)}_{n,h_1,i}(X_i), ..., F^{(r)}_{n,h_r,i}(X_i)) \,, \tag{3}$$

where $\mathbf{h} = (h_1, \ldots, h_r)$, $h_i > 0$, is a vector of bandwidths, X, X_1, X_2, $\ldots$, are i.i.d. F and for $i \geq 1$ and $n \geq 2$,

$$F_{n,i}\left(X_i\right) = \frac{1}{n-1}\sum_{1\leq i\neq j\leq n} K_0^{(-1)}\left(X_i - X_j\right),$$

with $K_0^{(-1)}(x) = I\left\{x > 0\right\}$, and for $1 \leq m \leq r$,

$$F^{(m)}_{n,h,i}(X_i) = \frac{1}{n-1}\frac{1}{h^m}\sum_{1\leq i\neq j\leq n} K_m^{(m-1)}\left(\frac{X_i - X_j}{h}\right), \; i = 1, \ldots, n,$$

where $K_m^{(0)} = K_m$ is an L_1 kernel, $m-1$ times differentiable with $K_m^{(m-1)}\left(x\right) = \frac{d^{m-1}K_m(x)}{dx^{m-1}}$, satisfying

$$\int_{\mathbb{R}} K_m\left(u\right) du = 1. \tag{4}$$

We note that the results below (Theorems 1 and 2) also apply if we change the estimator $T_n(\mathbf{h})$ a bit: a) of course the kernels K_m do not have to be different; b) $F_{n,i}(X_i)$ could be replaced by the cumulative distribution function of the kernel density estimator based on the sample with X_i deleted; then, since this cdf and the empirical cdf are asymptotically equivalent under the conditions of the theorems below (e.g., Bickel & Ritov (2003), p. 1036), they would also hold with this choice of $F_{n,i}$; c) finally, we remark that we take 'delete one' observation estimators for $F^{(m)}_{n,h,i}(X_i)$ because they are natural in that they have more simply expressed expected values and lead to U-statistics rather than to V-statistics, but the bias introduced if we did not delete the observation X_i from $F^{(m)}_{n,h,i}(X_i)$ would make no difference whatsoever (due to Condition VIII below).

From now on, to unburden our notation a bit, we shall write, for $n \geq 2$ and $i = 1, \ldots, n$,

$$F_n\left(X_i\right) = F_{n,i}\left(X_i\right) \text{ and } F^{(m)}_{n,h}(X_i) = F^{(m)}_{n,h,i}(X_i) \text{ for } 1 \leq m \leq r.$$

Define the sequence of processes in $\overrightarrow{\lambda} = (\lambda_1, \ldots, \lambda_r)$, $0 < a \leq \lambda_m \leq b < \infty$, for $m = 1, \ldots, r$, by

$$\nu_n\left(\overrightarrow{\lambda}\right) := \sqrt{n}\left\{T_n(\overrightarrow{\lambda} \otimes \mathbf{h}_n) - T(F)\right\}, \tag{5}$$

where $\{\mathbf{h}_n\}$ is a sequence of vectors $\mathbf{h}_n = (h_{1,n}, \ldots, h_{r,n})$ with positive coordinates converging to zero and

$$\overrightarrow{\lambda} \otimes \mathbf{h}_n = (\lambda_1 h_{1,n}, \ldots, \lambda_r h_{r,n}).$$

To ease the notation we shall write $\mathbf{h} = \mathbf{h}_n$, $h_m = h_{m,n}$, $m = 1, \ldots, r$, and even $h = h_{m,n}$ if no confusion is possible.

We will show, under suitable regularity conditions, that there exist i.i.d. mean zero and finite variance random variables $Y_1, Y_2, \ldots$, such that

$$\sup_{\overrightarrow{\lambda} \in [a,b]^r} \left|\nu_n\left(\overrightarrow{\lambda}\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right| = o_p\left(1\right) \tag{6}$$

and

$$\sup_{\overrightarrow{\lambda}\in[a,b]^r}\left|\nu_n\left(\overrightarrow{\lambda}\right)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i\right|=o\left(\sqrt{\log\log n}\right)\text{ a.s.}\tag{7}$$

Statements (6) and (7) imply that uniformly in $\overrightarrow{\lambda}\in[a,b]^r$ the estimators $T_n(\overrightarrow{\lambda}\otimes\mathbf{h}_n)$ based on the bandwidth vectors $\overrightarrow{\lambda}\otimes\mathbf{h}_n$ converge in distribution to a normal random variable or almost surely to $T(F)$ at the law of the iterated logarithm rate.

Towards defining the random variables $Y_1,Y_2,\ldots,$ let

$$\varphi(x):=\varphi(x,F(x),F^{(1)}(x),...,F^{(r)}(x))$$

and, subject to smoothness assumptions, set for $m=0,\ldots,r$,

$$\varphi_m(x):=\frac{\partial}{\partial y_m}\varphi(x,y_0,y_1,...,y_r)\,|_{(y_0,y_1,...,y_r)=\left(F(x),F^{(1)}(x),...,F^{(r)}(x)\right)}.\tag{8}$$

Let

$$\xi(X_i)=\varphi(X_i)-E\varphi(X)=\varphi(X_i)-T(F),$$

$$\xi_0(X_i)=\int_{X_i}^{\infty}\varphi_0(y)f(y)\,dy-\int_{\mathbb{R}}F(y)\varphi_0(y)f(y)\,dy$$

and, for $m=1,\ldots,r$, let

$$\chi_m(y):=\frac{d^{m-1}}{dy^{m-1}}\left(\varphi_m(y)f(y)\right)=-\frac{d^m}{dy^m}\int_y^{\infty}\varphi_m(x)f(x)\,dx,\tag{9}$$

where $\frac{d^0}{dy^0}g(y)=g(y)$, and set for $i\geq 1$,

$$\xi_m(X_i)=(-1)^{m-1}\left\{\chi_m(X_i)-E\chi_m(X)\right\}.$$

Define

$$Y_i=\xi(X_i)+\sum_{m=0}^{r}\xi_m(X_i)\,,\text{ for }i\geq 1.\tag{10}$$

For a given kernel K, $s > 0$ and non-negative measurable function H, let $\mathcal{G}_{s,K}(H)$ denote the class of all measurable real valued functions g on $\mathbb{R}$ for which there is a positive constant $M_K(g)$ such that for all $h > 0$ sufficiently small and for all $x \in \mathbb{R}$,

$$\left| \frac{1}{h} \int_{\mathbb{R}} g(u) K\left(\frac{x-u}{h}\right) du - g(x) \right| \le h^s M_K(g) H(x).$$

The following two propositions show that the classes $\mathcal{G}_{s,K}$ are not empty.

Proposition 1 (Motivated by Proposition 1 of Levit (1978)) *Assume that K is integrable, has compact support, and for some integer $s \ge 1$ and $0 < \alpha \le 1$,*

$$\int_{\mathbb{R}} K(u)du = 1, \ \int_{\mathbb{R}} u^k K(u)du = 0 \text{ for } k = 1, \dots, s. \tag{11}$$

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 that, for some $h_0 > 0$, $L_g > 0$ and every $x \in \mathbb{R}$,

$$\sup_{|h| \le h_0} |h|^{-\alpha} \left| g^{(s)}(x+h) - g^{(s)}(x) \right| =: L_g H(x), \tag{12}$$

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

$$\left| \frac{1}{h} \int_{\mathbb{R}} g(u) K\left(\frac{x-u}{h}\right) du - g(x) \right| \le h^{s+\alpha} C_K L_g H(x). \tag{13}$$

Proof. Choose any $0 < h \le h_0$. By a change of variable and then Taylor's theorem and (11)

$$\left| \frac{1}{h} \int_{\mathbb{R}} g(u) K\left(\frac{x-u}{h}\right) du - g(x) \right| = \left| \int_{\mathbb{R}} \{g(x-hv) - g(x)\} K(v)\, dv \right|$$

$$= \frac{h^s}{(s-1)!} \left| \int_{\mathbb{R}} \int_0^1 (1-\xi)^{s-1} g^{(s)}(x - \xi h v) v^s d\xi K(v)\, dv \right|,$$

which since K has compact support is for some $c > 0$,

$$= \frac{h^s}{(s-1)!} \left| \int_{-c}^{c} \int_0^1 (1-\xi)^{s-1} g^{(s)}(x - \xi h v)\, d\xi v^s K(v)\, dv \right|.$$

By (11) and (12) this last quantity is

$$\leq \frac{h^{s+\alpha}L_g H(x)}{(s-1)!}\int_{-c}^{c}\int_0^1 (1-\xi)^{s-1}\xi^{\alpha}d\xi\,|v|^{s+\alpha}|K(v)|\,dvL_g =: h^{s+\alpha}C_K H(x).$$

q.e.d.

For integers $s \geq 1$, let $\mathcal{G}_s$ denote the class of all measurable real valued functions g on $\mathbb{R}$ such that g is s times continuously differentiable and $\int_{\mathbb{R}} |g^{(s)}(u)|du < \infty$.

Proposition 2 (Devroye & Györfi (1985) Lemma 22, p 122) *Assume that K is symmetric about 0 and for some integer $s \geq 1$,*

$$\int_{\mathbb{R}} K(u)du = 1,\ \int_{\mathbb{R}} u^k K(u)du = 0 \text{ for } k = 1,\ldots,s-1,\ \textit{when } s \geq 2,\ \textit{and } \int_{\mathbb{R}} |u|^s |K(u)|\,du < \infty. \tag{14}$$

Then for some constant $M_K > 0$, for all $h > 0$ and $g \in \mathcal{G}_s$

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{h}\int_{\mathbb{R}} g(u)K\left(\frac{x-u}{h}\right)du - g(x)\right| \leq h^s M_K \int_{\mathbb{R}}\left|g^{(s)}(u)\right|du.$$

Proposition 2 shows that $\mathcal{G}_s \subseteq \mathcal{G}_{s,K}(H)$, with $H := 1$, whenever K satisfies (14).

Here are the assumptions that we need for our main result.

Condition I. For each $m = 1,\ldots,r$, $f^{(m-1)}$ is bounded.

Condition II. For some $0 < \alpha \leq 1$, and each $m = 1,\ldots,r$, $f^{(m-1)}$ is in $\mathcal{G}_{2r+\alpha-1,K_m}(H)$, where H is a non-negative measurable function satisfying $EH^2(X) < \infty$, which may depend on $f^{(m-1)}$.

Condition III. Uniformly in $x \in \mathrm{supp}(f)$, the function $\varphi(x, y_0, y_1, \ldots, y_r)$ and its partial derivatives $\partial\varphi/\partial y_i$, $i = 0,\ldots,r$, satisfy a global Lipschitz condition with respect to the variables $y_0,\ldots,y_r$ on a bounded open convex subset D of $\mathbb{R}^{r+1}$ containing the closure of the range

$$\left\{\left(F(x), f(x), f^{(1)}(x), \ldots, f^{(r-1)}(x)\right) : x \in \mathbb{R}\right\}.$$

Condition IV. The functions φ_m, $m = 0, 1, \ldots, r$, as defined in (8) are bounded on supp(f).

Condition V. For each $m = 1, \ldots, r$, the function χ_m defined in (9) is Lipschitz of order $0 < \beta \leq 1$ and

$$\int_{\mathbb{R}} |u|^{\beta} |K_m(u)| \, du < \infty. \tag{15}$$

Condition VI. The kernels $K_m^{(m-1)}$, $m = 1, \ldots, r$, are in $L_2(\mathbb{R})$ and of bounded variation in $\mathbb{R}$.

Condition VII. For $m = 1, \ldots, r$, all $h > 0$ and $x \in \mathbb{R}$,

$$\frac{1}{h^m} \int_{\mathbb{R}} K_m^{(m-1)} \left(\frac{x-y}{h} \right) f(y) \, dy = \frac{1}{h} \int_{\mathbb{R}} K_m \left(\frac{x-y}{h} \right) f^{(m-1)}(y) \, dy \tag{16}$$

and with φ_m as (8),

$$\int_{\mathbb{R}} \varphi_m(y) \frac{1}{h^m} K_m^{(m-1)} \left(\frac{y-x}{h} \right) f(y) \, dy = (-1)^{m-1} \int_{\mathbb{R}} \chi_m(y) \frac{1}{h} K_m \left(\frac{y-x}{h} \right) dy. \tag{17}$$

(These two identities will hold whenever K_m is smooth and has bounded support.)

Condition VIII. With $0 < \alpha \leq 1$ as in Condition II, for $m = 1, \ldots, r$, $\sqrt{n} h_{m,n}^{2r-1} / \log(1/h_{m,n}) \to \infty$ and $\sqrt{n} h_{m,n}^{2r+\alpha-1} \to 0$.

Theorem 1 *Under conditions Conditions I-VIII, for all $0 < a < b < \infty$, (6) holds. Moreover, if $0 < Var(Y) < \infty$, where Y is defined as in (10), then the processes $\left\{ \nu_n \left(\overrightarrow{\lambda} \right) : \overrightarrow{\lambda} \in [a, b]^r \right\}$ converge in law in $\ell_\infty([a, b]^r)$ (in the sense of Hoffmann-Jørgensen) to the constant Gaussian process $G \left(\overrightarrow{\lambda} \right) = G$ where G is $N(0, Var(Y_1))$, that is,*

$$\sqrt{n} \left\{ T_n(\overrightarrow{\lambda} \otimes \mathbf{h}_n) - T(F) \right\} \to_d N(0, Var(Y_1)) \tag{18}$$

uniformly in $\overrightarrow{\lambda} \in [a, b]^r$.

Proof. Set

$$S_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, F(X_i), F^{(1)}(X_i), ..., F^{(r)}(X_i));$$

$$S_n^{(2)}\left(\overrightarrow{\lambda}\right)=\sum_{m=0}^{r}\frac{1}{n}\sum_{i=1}^{n}\varphi_m\left(X_i\right)\left\{F_{n,\lambda h_m}^{(m)}(X_i)-F^{(m)}(X_i)\right\}.$$

Next set

$$S_n^{(3)}\left(\overrightarrow{\lambda}\right)=\sum_{m=1}^{r}\frac{1}{n}\sum_{i=1}^{n}\varphi_m\left(X_i\right)\left\{k_m^{(m-1)}\left(X_i,\lambda_m h_m\right)-F^{(m)}(X_i)\right\},$$

where for any $1\leq m\leq r$, $h>0$ and $a\leq\lambda\leq b$,

$$k_m^{(m-1)}\left(X_i,\lambda h\right)=E\left(F_{n,\lambda h}^{(m)}(X_i)|X_i\right)=\frac{1}{\left(\lambda h\right)^m}\int_{\mathbb{R}}K_m^{(m-1)}\left(\frac{X_i-y}{\lambda h}\right)f\left(y\right)dy,$$

and although they do not depend on λ or h, we set

$$k_0^{(-1)}\left(X_i,\lambda h\right)=E\left(F_n(X_i)|X_i\right)=F(X_i)\text{ and }F_{n,\lambda h}^{(0)}\left(X_i\right)=F_n\left(X_i\right).$$

Finally denote

$$S_n^{(4)}\left(\overrightarrow{\lambda}\right)=\sum_{m=0}^{r}\frac{1}{n}\sum_{i=1}^{n}\varphi_m\left(X_i\right)\left\{F_{n,\lambda h_m}^{(m)}(X_i)-k_m^{(m-1)}\left(X_i,\lambda_m h_m\right)\right\}.$$

Notice that: 1) $S_n^{(1)}$ is the naive estimator of $T_n(\overrightarrow{\lambda}\otimes\mathbf{h}_n)$ if F is known, and that $S_n^{(1)}(\overrightarrow{\lambda})-T(F)=\frac{1}{n}\sum_{i=1}^{n}\xi(X_i)$, which is part of $\frac{1}{n}\sum_{i=1}^{n}Y_i$; 2) $S_n^{(2)}(\overrightarrow{\lambda})$ is the linear term in the Taylor expansion of the summands of $T_n(\overrightarrow{\lambda}\otimes\mathbf{h}_n)$ about the corresponding ones of $S_n^{(1)}$; and 3) $S_n^{(2)}(\overrightarrow{\lambda})=S_n^{(3)}(\overrightarrow{\lambda})+S_n^{(4)}(\overrightarrow{\lambda})$ decomposes $S_n^{(2)}$ into two intermediate steps.

We will show that

$$\gamma_n^{(1)}=T_n-S_n^{(1)}-S_n^{(2)},\quad\gamma_n^{(2)}=S_n^{(3)}\text{ and }\gamma_n^{(3)}=S_n^{(4)}-\frac{1}{n}\sum_{i=1}^{n}\sum_{m=0}^{r}\xi_m(X_i)$$

are small, which will give the theorem since, obviously, by the definition of Y_i,

$$T_n(\overrightarrow{\lambda}\otimes\mathbf{h}_n)-T(F)-\frac{1}{n}\sum_{i=1}^{n}Y_i=\gamma_n^{(1)}\left(\overrightarrow{\lambda}\right)+\gamma_n^{(2)}\left(\overrightarrow{\lambda}\right)+\gamma_n^{(3)}\left(\overrightarrow{\lambda}\right).\tag{19}$$

In order to handle γ_1, first we remark that Condition III implies that for some constant $C>0$

$$\left|\varphi\left(x,y+\Delta\right)-\varphi\left(x,y\right)-\Delta\dot{\varphi}\left(x,y\right)\right|\leq C\left\|\Delta\right\|^2,\tag{20}$$

for all $x \in \text{supp}(f)$ and all y and $y+\Delta$ in D, where $\dot{\varphi}(x,y)$ is the gradient of the mapping $y \to \varphi(x,y)$. In order to make use of Condition III, we must make sure that the y and $y+\Delta$ variables of interest are in D. Towards this aim we will show that

$$\left\{\left(F_n(X_i), F^{(1)}_{n,\lambda_1 h_1}(X_i), ..., F^{(r)}_{n,\lambda_r h_r}(X_i)\right) : i = 1, \dots, n,\ \overrightarrow{\lambda} \in [a,b]^r\right\} \subset D, \tag{21}$$

with probability tending to 1 as $n \to \infty$, and that

$$\left\{\left(F(X_i), k^{(0)}_1(X_i, \lambda_1 h_1)), ..., k^{(r-1)}_r(X_i, \lambda_r h_r)\right) : i = 1, \dots, n,\ \overrightarrow{\lambda} \in [a,b]^r\right\} \subset D, \tag{22}$$

for all n large enough.

For the second inclusion (22), we just note that, by Conditions VII and II, for $1 \le m \le r$, we get for some constant $C > 0$ and all h_m small enough,

$$\begin{aligned} &\max_{1\le i\le n} \sup_{a \le \lambda_m \le b} \left| k^{(m-1)}_m(X_i, \lambda_m h_m) - F^{(m)}(X_i) \right| \\ &= \max_{1\le i\le n} \sup_{a \le \lambda_m \le b} \left| \frac{1}{\lambda_m h_m} \int_{\mathbb{R}} K_m\left(\frac{X_i - y}{\lambda_m h_m}\right) f^{(m-1)}(y) dy - f^{(m-1)}(X_i) \right| \\ &\le C\sqrt{n} h_m^{2r+\alpha-1} \max_{1\le i\le n} \left\{ H(X_i)/\sqrt{n} \right\}. \end{aligned} \tag{23}$$

Next $EH^2(X) < \infty$ implies

$$\max_{1\le i\le n} \left\{ H(X_i)/\sqrt{n} \right\} \to 0, \text{ a.s.} \tag{24}$$

Now (23) and (24) when combined with Condition VIII gives

$$\max_{1\le i\le n} \sup_{a \le \lambda_m \le b} \left| k^{(m-1)}_m(X_i, \lambda_m h_m) - F^{(m)}(X_i) \right| \to 0 \text{ a.s.}$$

Towards proving the first inclusion (21), we first notice that $\|F_n(x) - F(x)\|_\infty \to 0$ a.s. by Glivenko-Cantelli, which takes care of the first coordinate. To handle the remaining coordinates it will be helpful to introduce the following kernel estimator: set for $1 \le m \le r$, $h > 0$ and $x \in \mathbb{R}$,

$$\widehat{f}^{(m-1)}_{n,h}(x) = \frac{1}{nh^m} \sum_{i=1}^{n} K^{(m-1)}_m\left(\frac{x - X_j}{h}\right),\ i = 1, \dots, m,$$

Our aim is to apply Theorem 8 in Giné & Mason (2007a) (see Theorem 1A in the Appendix). First we note that, as shown by Nolan & Pollard (1987), the class of functions

$$\{K_m^{(m-1)}(c(x-\cdot)) : c \geq 0, x \in \mathbb{R}\} \tag{25}$$

is of VC type (or VC for short) and uniformly bounded by Condition VI (refer to the Appendix for the definition of VC type.) Next observe that

$$\sup_{x\in\mathbb{R}} E\left(K_m^{(m-1)}\left(\frac{x-X}{\lambda_m h_m}\right)\right)^2 \leq \lambda_m h_m \|f\|_\infty \|K_m^{(m-1)}\|_2^2,$$

which gives (note that we assume f is bounded and $K_m^{(m-1)} \in L_2(\mathbb{R})$)

$$\sup_{a\leq\lambda\leq b, x\in\mathbb{R}} \mathrm{Var}\left(K_m^{(m-1)}\left(\frac{x-X}{\lambda h_m}\right)\right) \leq C h_m =: \sigma^2,$$

for some constant $1 < C < \infty$. So, $n\sigma^2 \geq nh_m > \sqrt{n} h_m^{2r-1} >> \log \frac{1}{h_m}$ by Condition VIII. Hence, that theorem for $k = 1$ gives that there exist $C_1, C_2 < \infty$, independent of n and m, such that

$$E \sup_{a\leq\lambda\leq b, x\in\mathbb{R}} \left|\widehat{f}_{n,\lambda h_m}^{(m-1)}(x) - (\lambda h_m)^{-m}\, E K_m^{(m-1)}\left(\frac{x-X}{\lambda h_m}\right)\right| \leq C_1 \sqrt{\frac{\log(C_2/h_m)}{n h_m^{2m-1}}}.$$

Note that this sup is measurable because, the map $(x, y, h) \mapsto K_m^{(m-1)}((x-y)/h)$ being jointly measurable, the measurability arguments in Pollard (1984), Appendix C (permissible classes), do apply. This observation applies to other uncountable suprema below, and will not be repeated. Since for each $1 \leq i \leq n$ and $h > 0$,

$$\max_{1\leq i\leq n} \left|\widehat{f}_{n,h}^{(m-1)}(X_i) - F_{n,h}^{(m)}(X_i)\right| \leq \frac{1}{n-1} \sup_{x\in\mathbb{R}} \left|\widehat{f}_{n,h}^{(m-1)}(x)\right| + \frac{\left|K_m^{(m-1)}(0)\right|}{(n-1)\, h^m} \leq \frac{C}{(n-1)h^m} \tag{26}$$

for some $C < \infty$, we get, for some $C' < \infty$,

$$E\left(\max_{1\leq i\leq n} \sup_{a\leq\lambda\leq b} \left|F_{n,\lambda h_m}^{(m)}(X_i) - k_m^{(m-1)}(X_i, \lambda h_m)\right|\right) \leq C\sqrt{\log(1/h_m)/\left(n h_m^{2m-1}\right)} = o(1)$$

by Condition VIII. This when combined with (21) and the observation above about the first coordinate proves the first inclusion (21).

Step 1. We will estimate $\sup_{\overrightarrow{\lambda}\in[a,b]^r}\left|\gamma_n^{(1)}\left(\overrightarrow{\lambda}\right)\right|$. When inclusion (21) holds, by convexity of D, we can apply Condition III and Taylor's theorem (see (20)) to get that for some $M_1>0$

$$\sup_{\overrightarrow{\lambda}\in[a,b]^r}\left|\gamma_n^{(1)}\left(\overrightarrow{\lambda}\right)\right| = \sup_{\overrightarrow{\lambda}\in[a,b]^r}\left|T_n(\overrightarrow{\lambda}\otimes\mathbf{h}_n) - S_n^{(1)}\left(\overrightarrow{\lambda}\right) - S_n^{(2)}\left(\overrightarrow{\lambda}\right)\right|$$

$$\le M_1\left\{\sum_{m=1}^r \sup_{a\le\lambda\le b}\frac{1}{n}\sum_{i=1}^n\left(F_{n,\lambda h_m}^{(m)}(X_i) - F^{(m)}(X_i)\right)^2 + \frac{1}{n}\sum_{i=1}^n\left(F_n(X_i) - F\left(X_i\right)\right)^2\right\}. \tag{27}$$

First by Kolmogorov-Smirnov,

$$\frac{1}{n}\sum_{i=1}^n\left(F_n(X_i)-F(X_i)\right)^2 \le \|F_n - F\|_\infty^2 = O_p\left(\frac{1}{n}\right). \tag{28}$$

Next, for $m=1,\ldots,r$, the trivial inequality $(a+b)^2\le 2a^2+2b^2$ yields

$$\sum_{m=1}^r\sup_{a\le\lambda\le b}\frac{1}{n}\sum_{i=1}^n\left(F_{n,\lambda h_m}^{(m)}(X_i)-F^{(m)}(X_i)\right)^2 \le 2\sum_{m=1}^r\delta_{n,1}^{(m)} + 2\sum_{m=1}^r\delta_{n,2}^{(m)}, \tag{29}$$

where

$$\delta_{n,1}^{(m)} = \sup_{a\le\lambda\le b}\frac{1}{n}\sum_{i=1}^n\left(F_{n,\lambda h_m}^{(m)}(X_i) - k_m^{(m-1)}\left(X_i,\lambda h_m\right)\right)^2$$

and

$$\delta_{n,2}^{(m)} = \sup_{a\le\lambda\le b}\frac{1}{n}\sum_{i=1}^n\left(k_m^{(m-1)}\left(X_i,\lambda h_m\right) - F^{(m)}(X_i)\right)^2.$$

It follows from Conditions II and VII that

$$\delta_{n,2}^{(m)} = O\left(\|\mathbf{h}_n\|^{4r+2\alpha-2}\right)\frac{1}{n}\sum_{i=1}^n H^2\left(X_i\right) = O\left(\|\mathbf{h}_n\|^{4r+2\alpha-2}\right),\text{ a.s.} \tag{30}$$

Regarding $\delta_{n,1}^{(m)}$, as noted above, if $K_m^{(m-1)}$ is of bounded variation, the class of functions (25) is of VC type and uniformly bounded; therefore, Corollary 1 in Giné & Mason (2007b) (see Theorem 1A in the Appendix) applies. In that theorem, take $k=1$, notice that, since

$$E\left(K_m^{(m-1)}((x-X)/\lambda h_m)\right)^2 \le \lambda h_m\|f\|_\infty\|K_m^{(m-1)}\|_2^2,$$

we can take $\sigma = Ch_m^{1/2}$ for some $C < \infty$, and then that, by Condition VIII, $nh_m > L\log(1/h_m)$ (eventually, for any $L < \infty$). Hence, via the bound (26), it gives for $1 \le m \le r$ that

$$E\delta_{n,1}^{(m)} \le E \sup_{a\le\lambda\le b, x\in\mathbb{R}} \left(F_{n,\lambda h_m}^{(m)}(x) - k_m^{(m-1)}(x, \lambda h_m)\right)^2$$

$$\le C_1 \frac{h_m \log(1/h_m)}{(ah_m)^{2m}(n-1)} \le C_2 \frac{\log(1/h_m)}{nh_m^{2r-1}}. \tag{31}$$

Putting everything together we get from (27), (28), (29), (30) and (31) that

$$\sup_{\overrightarrow{\lambda}\in[a,b]^r} \left|\gamma_n^{(1)}\left(\overrightarrow{\lambda}\right)\right| = O_p\left(\|\mathbf{h}_n\|^{4r+2\alpha-2}\right) + O_p\left(\max_{1\le m\le r} \frac{\log(1/h_m)}{nh_m^{2r-1}}\right). \tag{32}$$

Step 2. It follows from an application of the Cauchy-Schwarz inequality, (31) and Condition IV that

$$\sup_{\overrightarrow{\lambda}\in[a,b]^r} \left|\gamma_n^{(2)}\left(\overrightarrow{\lambda}\right)\right|^2 \le \frac{1}{n}\sum_{m=1}^r \sum_{i=1}^n \varphi_m^2(X_i) \sum_{m=1}^r \delta_{n,2}^{(m)} = O_p\left(\|\mathbf{h}_n\|^{4r+2\alpha-2}\right). \tag{33}$$

Step 3. In order to treat $\gamma_n^{(3)}\left(\overrightarrow{\lambda}\right)$ we must introduce some additional terms. Consider for any $1 \le m \le r$ and $a \le \lambda \le b$,

$$S_{n,m}^{(4)}(\lambda) = \frac{1}{n}\sum_{i=1}^n \varphi_m(X_i)\left\{F_{n,\lambda h_m}^{(m)}(X_i) - k_m^{(m-1)}(X_i, \lambda h_m)\right\}$$

$$= \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \varphi_m(X_i)\left\{\frac{1}{(\lambda h_m)^m} K_m^{(m-1)}\left(\frac{X_i - X_j}{\lambda h_m}\right) - k_m^{(m-1)}(X_i, \lambda h_m)\right\}.$$

Set

$$L_{m,\lambda}(x, y) = \frac{\varphi_m(x)}{2}\left\{\frac{1}{(\lambda h_m)^m} K_m^{(m-1)}\left(\frac{x-y}{\lambda h_m}\right) - k_m^{(m-1)}(x, \lambda h_m)\right\}$$

$$+\frac{\varphi_m(y)}{2}\left\{\frac{1}{(\lambda h_m)^m} K_m^{(m-1)}\left(\frac{y-x}{\lambda h_m}\right) - k_m^{(m-1)}(y, \lambda h_m)\right\}.$$

Now for $1 \le m \le r$ and $a \le \lambda \le b$,

$$S_{n,m}^{(4)}(\lambda) = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} L_{m,\lambda}(X_i, X_j).$$

By the Hoeffding decomposition (see the Appendix for a description of the Hoeffding decomposition) we can write

$$S_{n,m}^{(4)}(\lambda) = \sum_{k=1}^{2} \binom{2}{k} U_n^{(k)}(\pi_k L_{m,\lambda}).$$

In particular,

$$\pi_1 L_{m,\lambda}(X_i) = \int_{\mathbb{R}} \frac{\varphi_m(y)}{2} \left\{ \frac{1}{(\lambda h_m)^m} K_m^{(m-1)}\left(\frac{y - X_i}{\lambda h_m}\right) - k_m^{(m-1)}(y, \lambda h_m) \right\} f(y)\, dy.$$

In view of (17) we obtain

$$\frac{1}{2}\int_{\mathbb{R}} \varphi_m(y) \frac{1}{(\lambda h_m)^m} K_m^{(m-1)}\left(\frac{y - X_i}{\lambda h_m}\right) f(y)\, dy = \frac{(-1)^{m-1}}{2} \int_{\mathbb{R}} \chi_m(y) \frac{1}{\lambda h} K_m\left(\frac{y - X_i}{\lambda h_m}\right) dy,$$

so that

$$\binom{2}{1} U_n^{(1)}(\pi_1 L_{m,\lambda}) = \frac{(-1)^{m-1}}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} \chi_m(y) \left\{ \frac{1}{\lambda h_m} K_m\left(\frac{y - X_i}{\lambda h_m}\right) - \frac{1}{\lambda h_m} E K_m\left(\frac{y - X}{\lambda h_m}\right) \right\} dy.$$

We have then that for $m = 1, \ldots, r$,

$$\left| \binom{2}{1} U_n^{(1)}(\pi_1 L_{m,\lambda}) - \frac{1}{n}\sum_{i=1}^{n} \xi_m(X_i) \right| = \left| \frac{1}{n}\sum_{i=1}^{n} \left\{ \omega_{m,\lambda h}(X_i) - E\omega_{m,\lambda h}(X_i) \right\} \right|,$$

where for $h > 0$

$$\omega_{m,h}(\cdot) = \int_{\mathbb{R}} \left\{ \chi_m(y) - \chi_m(\cdot) \right\} \frac{1}{h} K_m\left(\frac{y - \cdot}{h}\right) dy.$$

With this notation

$$\sup_{a \le \lambda \le b} \left| \binom{2}{1} U_n^{(1)}(\pi_1 L_{m,\lambda}) - \frac{1}{n}\sum_{i=1}^{n} \xi_m(X_i) \right| = \sup_{a \le \lambda \le b} \left| \frac{1}{n}\sum_{i=1}^{n} \left\{ \omega_{m,\lambda h}(X_i) - E\omega_{m,\lambda h}(X_i) \right\} \right| =: \Delta_n(m).$$

For each $1 \le m \le r$ introduce the class of functions $\mathcal{W}_m = \{\omega_{m,h}(\cdot) : 0 < h \le b\}$ (there is no loss of generality in assuming $\|h\| \le 1$). It is readily checked that each class $\mathcal{W}_m$ is of VC type: in fact, $\mathcal{W}_m$ is uniformly bounded and its covering numbers $N(\mathcal{W}_m, \|\cdot\|_\infty, \varepsilon)$ are at most of the order c/ε^β. To see this, just note that, by a change of variables and Condition V, for $h, h' > 0$ and for all $x \in \mathbb{R}$,

$$|\omega_{m,h}(x) - \omega_{m,h'}(x)| = \left| \int \left(\chi_m(x + uh) - \chi_m(x + uh')\right) K_m(u) du \right| \le A|h - h'|^\beta \int |u|^\beta |K_m(u)| du$$

for some $A < \infty$, and that, similarly, $|\omega_{m,h}(x)| \le Ah^\beta \int |u|^\beta |K_m(u)| du$, proving boundedness and the VC property for $\mathcal{W}_m$. Moreover, in particular, for each $a \le \lambda \le b$ and $h > 0$,

$$E\left(\omega_{m,\lambda h}(X)\right)^2 \le A^2 h^{2\beta} \left(\int_{\mathbb{R}} |u|^\beta |K_m(u)| \, du \right)^2 < \infty. \tag{34}$$

We can now apply Theorem 1A to get that for some constants A_1 and A_2 and for $m = 1, \ldots, r$,

$$E\Delta_n(m) = A_1 h_m^\beta \sqrt{\log\left(A_2/h_m\right)/n} \tag{35}$$

Next, to estimate $U_n^{(2)}(\pi_2 L_{m,\lambda})$, $m = 1, \ldots, r$ and $a \le \lambda \le b$, we will apply Theorem 1A once more. Since

$$\pi_2 \left(k_m^{(m-1)}(x, \lambda h_m) - k_m^{(m-1)}(y, \lambda h_m) \right) = 0,$$

this theorem applies if the class of functions of x and y,

$$\left\{ \varphi_m(x) K_m^{(m-1)}(c(x-y)) + \varphi_m(y) K_m^{(m-1)}(c(y-x)) : c \ge 0 \right\}$$

is uniformly bounded and of VC type. But this follows because φ_m is bounded, the class $\{K_m^{(m-1)}(c(x-y)) : c \ge 0\}$ is VC-subgraph assuming $K_m^{(m-1)}$ is of bounded variation by the arguments in Nolan & Pollard (1987), and then so is $\{\varphi_m(x) K_m^{(m-1)}(c(x-y)) : c \ge 0\}$ by e.g., Lemma 2.6.18, (vi), in van der Vaart & Wellner (1996), and from the fact that, as a simple computation on covering numbers shows, if two classes $\mathcal{F}$ and $\mathcal{G}$ are VC type, then so is the class $\{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$. So, we can apply Theorem 1A to the class of functions

$$\left\{ \phi_m(x) K_m^{(m-1)}((x-y)/\lambda h_m) - \phi_m(y) K_m^{(m-1)}((y-x)/\lambda h_m) : a \le \lambda \le b \right\}$$

with $k = 2$. By boundedness of the functions ϕ_m, as in Step 1, we can take σ to be a constant times $h_m^{1/2}$, and Theorem 1A then gives that, for $m = 1, \ldots, r$,

$$\sup_{a \le \lambda \le b} \left| U_n^{(2)}(\pi_2 L_{m,\lambda}) \right| = O_p\left(\frac{\log\left(1/h_m\right)}{n h_m^{m-1/2}} \right) = O_p\left(\frac{\log\left(1/h_m\right)}{n h_m^{r-1/2}} \right). \tag{36}$$

We now treat the case $m = 0$,

$$S_{n,0}^{(4)} = \frac{1}{n} \sum_{i=1}^{n} \varphi_0(X_i) \{F_n(X_i) - F(X_i)\}$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \varphi_0(X_i) \{I\{X_i - X_j > 0\} - F(X_i)\}.$$

Set for $m = 0$,

$$L_0(x, y) = \frac{\varphi_0(x)}{2} \{I\{x - y > 0\} - F(x)\} + \frac{\varphi_0(y)}{2} \{I\{y - x > 0\} - F(y)\}.$$

Now by the Hoeffding decomposition,

$$S_{n,0}^{(4)} = \sum_{k=1}^{2} \binom{2}{k} U_n^{(k)}(\pi_k L_0).$$

We see that

$$\pi_1 L_0(X_i) = \int_{\mathbb{R}} \frac{\varphi_0(y)}{2} \{I\{y - X_i > 0\} - F(y)\} f(y)\, dy,$$

which gives

$$\binom{2}{1} U_n^{(1)}(\pi_1 L_0) = \frac{1}{n} \sum_{i=1}^{n} \xi_0(X_i). \tag{37}$$

Since φ_0 is bounded on $\text{supp}(f)$ (Condition IV), $L_0(X_1, X_2)$ is square integrable and therefore we have

$$U_n^{(2)}(\pi_2 L_0) = O_p\left(\frac{1}{n}\right). \tag{38}$$

Noting that

$$S_n^{(4)}\left(\overrightarrow{\lambda}\right) = S_{n,0}^{(4)} + \sum_{m=1}^{r} S_{n,m}^{(4)}(\lambda_m), \tag{39}$$

we get from (35), (36), (37) and (38) that

$$\sup_{\overrightarrow{\lambda} \in [a,b]^r} \left|\gamma_n^{(3)}\left(\overrightarrow{\lambda}\right)\right| = O_p\left(\frac{1}{n} + \max_{1 \leq m \leq r} \frac{\log(1/h_m)}{n h_m^{r-1/2}} + \max_{1 \leq m \leq r} h_{m,n}^{\beta} \sqrt{\frac{\log(1/h_{m,n})}{n}}\right). \tag{40}$$

Combining (32), (33) and (40) with Condition VIII, we conclude that

$$\sup_{\overrightarrow{\lambda}\in[a,b]^r}\left|\nu_n\left(\overrightarrow{\lambda}\right)-\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i\right| =$$

$$O_p\left(\sqrt{n}\|\mathbf{h}_n\|^{4r+2\alpha-2}\right)+O_p\left(\max_{1\le m\le r}\frac{\log\left(1/h_m\right)}{\sqrt{n}h_m^{2r-1}}\right)+O_p\left(\sqrt{n}\|\mathbf{h}_n\|^{2r+\alpha-1}\right)$$

$$+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\max_{1\le m\le r}\frac{\log\left(1/h_{m,n}\right)}{\sqrt{n}h_{m,n}^{r-1/2}}\right)+O_p\left(\max_{1\le m\le r}h_{m,n}^{\beta}\sqrt{\log\left(1/h_{m,n}\right)}\right)=o_p\left(1\right),$$

which finishes the proof of (6).

In fact we have shown that

$$\sup_{\overrightarrow{\lambda}\in[a,b]^r}\left|\nu_n\left(\overrightarrow{\lambda}\right)-n^{-1/2}\sum_{i=1}^n Y_i\right|=o_p^*\left(1\right),$$

meaning convergence to zero in outer probability (that is, the random quantities at the left hand side are dominated by random variables that converge to zero in probability). If $EY_1^2<\infty$, this in particular implies by general principles on convergence in law of bounded processes in the sense of Hoffmann-Jørgensen (e.g., Corollary 5.1.3 in de la Peña & Giné (1999) together with the triangle inequality for the dual bounded Lipschitz norm), that $\nu_n\to_{\mathcal{L}}G,\ \text{ in }\ell^\infty\left([a,b]^r\right)$, thus completing the proof of the theorem. q.e.d.

The above theorem has the following consequence that may be useful for bandwidth selection. The proof is standard and is omitted.

Corollary 1. *Under the conditions of Theorem 1, if* $\overrightarrow{\lambda}_n$ *are random vectors satisfying the condition that* $\min\left(\lambda_{n,1},\ldots,\lambda_{n,r}\right)$ *and* $\max\left(\lambda_{n,1},\ldots,\lambda_{n,r}\right)$ *are tight in* $(0,\infty)$*, then* $\nu_n\left(\overrightarrow{\lambda}_n\right)\to_d G$.

In order to prove (7), we still use the decomposition (19) of Theorem 1; then, to estimate the a.s. size of its different components we block as usual and, instead of using moment bounds for the expectation of empirical and U-processes, we use the maximal inequality in Theorem 4, Giné

& Mason(2007b) in combination with Major's (2006) exponential bound (Theorem 5 in Giné & Mason(2007b)) and the moment bound in Corollary 1 of the same reference (see Theorem 1A in the Appendix), just as was done in subsection 3.2.3 of the same article. The conclusion is:

Theorem 2. *Assume Conditions I-VII and (instead of Condition VIII) that, for* $1 \leq m \leq r$, $h_{m,n} \downarrow 0$ *as* $n \to \infty$, *that there exists* $c < \infty$ *such that* $h_{m,n} \leq c h_{m,2n}$ *and*

$$\log(1/h_{m,n})/\log\log n \to \infty, \quad \sqrt{n \log\log n}\, h_{m,n}^{2r-1}/\log(1/h_{m,n}) \to \infty, \quad \sqrt{n} h_{m,n}^{2r+\alpha-1}/\sqrt{\log\log n} \to 0$$

as $n \to \infty$. *Then,* (7) *holds and therefore, if* $0 < Var(Y_1) < \infty$, *by the law of the iterated logarithm for* Y_1,

$$\limsup_n \sqrt{\frac{n}{2\log\log n}} \sup_{\vec{\lambda}\in[a.b]^r} \left| T_n(\vec{\lambda} \otimes \mathbf{h}_n) - T(F) \right| = \sqrt{Var(Y_1)} \ \ a.s.$$

A compact LIL also holds when $0 < Var(Y_1) < \infty$, but we refrain from formulating it (see e.g. Giné & Mason (2007b) for a compact LIL in a similar setting).

3 Functionals of the form $\theta = \int \phi(f(x))dx$

Grübel (1994) considers estimation of integral functionals of the form:

$$\theta = \int_{\mathbb{R}} \phi\left(f\left(x\right)\right) dx,$$

and shows that his estimators are consistent at rate $n^{-1/2}$ and asymptotically normal, using empirical process methods. Under suitable assumptions his Theorem 2.2 says that

$$\sqrt{n}\left(\theta_n - \theta\right) \to_d \sqrt{Var\left(\phi'\left(f\left(X\right)\right)\right)}\, g, \tag{41}$$

where g is standard normal and

$$\theta_n = \int_{\mathbb{R}} \phi\left(\widehat{f}_{n,h_n}(x)\right) dx, \tag{42}$$

with

$$\widehat{f}_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_j}{h}\right). \tag{43}$$

At first glance it does not look like we can handle his setup. Here are the assumptions that Grübel (1994) needs to conclude the central limit theorem (41).

1. Assumptions on K: The kernel K is a symmetric density function with support in $[-1, 1]$.

2. Assumptions on f: The density f is bounded with derivative f', which satisfies that for some $\delta > 0$,

$$|f'(x+y) - f'(x)| \leq |y| H(x), \quad |y| < \delta, \ x \in \mathbb{R}, \tag{44}$$

where H is an integrable function. (It appears that he also needs H to be square integrable.)

3. Assumptions on ϕ: The function ϕ is a continuous function defined on $[0, \infty)$ with a continuous and bounded second derivative on $(0, \infty)$ such that $\phi(0) = 0$ and the integral θ exists.

4. Assumptions on h_n: As $n \to \infty$, $\sqrt{n}h_n \to \infty$ and $\sqrt{n}h_n^2 \to 0$.

To obtain a uniform in bandwidth version of Theorem 2.2 of Grübel (1994) we need to slightly strengthen his assumptions.

A. In addition to the kernel K being a symmetric density function with support $[-1, 1]$, assume that K is of bounded variation.

B. Assume that f has a bounded derivative f' that fulfills (44) for some $\delta > 0$ and some H satisfying $EH^2(X) < \infty$; in particular, f is bounded. [Trivially, H square integrable and f bounded imply $EH^2(X) < \infty$. Moreover H integrable implies that f' is bounded. (This is shown below in Remark 2.)]

C. Let ϕ be a continuous function defined on the interval $I = (-\varepsilon, D)$ for some $\varepsilon > 0$ and $D > 0$, where I contains the closure of the image of f. We assume that ϕ has second derivative on I which is Lipschitz and $\phi(0) = \phi'(0) = 0$.

D. As $n \to \infty$, $\sqrt{n}h_n / \log(1/h_n) \to \infty$ and $\sqrt{n}h_n^2 \to 0$.

The conditions in C imply that the integral θ exists. We shall be applying Theorem 1 to the function of the three variables x, y_0 and y,

$$\varphi(x, y_0, y) := 1\{y \in I\}\, \phi(y)/y,$$

with the natural convention $\phi(0)/0 := 0$. Notice that the requirement that $\phi'(0) = 0$ is not a real restriction, since otherwise we can replace $\phi(y)$ by $\phi_0(y) = \phi(y) - \phi'(0)\, y$. For this ϕ_0, $\theta - \phi'(0) = \int_{\mathbb{R}} \phi_0(f(x))\, dx$. So, after getting an estimate of $\theta - \phi'(0)$ based on ϕ_0 we just add $\phi'(0)$ to it to obtain an estimate of θ.

Formally φ is a function of the three variables x, y_0 and y, however it is clearly constant in the first two. Therefore we can simplify our notation by writing

$$\Psi(y) := \varphi(x, y_0, y).$$

The estimator of θ corresponding to this function given by (3) is

$$\tilde{\theta}_n(h) = \frac{1}{n} \sum_{i=1}^{n} \Psi(\hat{f}_{n,h,i}(X_i)), \tag{45}$$

where

$$\hat{f}_{n,h,i}(x) := \frac{1}{n-1} \frac{1}{h} \sum_{1 \le j \le n, j \ne i} K\left(\frac{x - X_j}{h}\right),$$

which may be considered a modification of Grübel's θ_n. Then, Theorem 1 gives:

Theorem 3. *Under the assumptions A-D, and with* $Y_i = \phi'(f(X_i)) - E\phi'(f(X_i))$, $i \in N$, *for any* $0 < a \le b < \infty$,

$$\sqrt{n} \sup_{a \le \lambda \le b} \left| \tilde{\theta}_n(\lambda h_n) - E\tilde{\theta}_n(\lambda h_n) - \frac{1}{n} \sum_{i=1}^{n} Y_i \right| = o_p(1).$$

Moreover, if $0 < Var(\phi'(X)) < \infty$, *then*

$$\sqrt{n}\left(\tilde{\theta}_n(\lambda h_n) - E\tilde{\theta}_n(\lambda h_n)\right) \to_d \sqrt{Var(\phi'(f(X)))}\, g$$

uniformly in $\lambda \in [a, b]$ in the sense of Theorem 1.

Proof. By Theorem 1, it suffices to verify Conditions I-VIII from Section 2. Clearly Condition I holds with $r = 1$ and by Proposition 1, Condition II does too with $s = \alpha = 1$.

To verify Condition III, first note that the assumptions that $\phi(0) = \phi'(0) = 0$ and ϕ has a second derivative on I which is Lipschitz imply that for some $C > 0$,

$$\left|\phi''(y) - \phi''(0)\right| \leq C|y|, \tag{46}$$

$$|\phi'(y) - \phi''(0)\, y| 1\{y \in I\} \leq Cy^2 \tag{47}$$

and

$$|\phi(y) - 2^{-1}\phi''(0)\, y^2| 1\{y \in I\} \leq C|y|^3. \tag{48}$$

From (47) and (48) we get that for $y \in I$

$$\left|\frac{d}{dy}\Psi(y)\right| = \left|y^{-1}\phi'(y) - y^{-2}\phi(y)\right| \leq 2^{-1}\left|\phi''(0)\right| + 2C|y| \tag{49}$$

and from (46), (47) and (48) we get that for $y \in I$

$$\left|\frac{d^2}{dy^2}\Psi(y)\right| = \left|y^{-1}\phi''(y) - 2y^{-2}\phi'(y) + 2y^{-3}\phi(y)\right| \leq 5C. \tag{50}$$

This implies that Condition III is satisfied.

Next observe that by (49) for some $C' > 0$

$$|\varphi_1(x)| = \left|\frac{\partial}{\partial F^{(1)}}\varphi\left(x, F(x), F^{(1)}(x)\right)\right| = \left|\frac{\phi'(f(x))}{f(x)} - \frac{\phi(f(x))}{f^2(x)}\right| \leq C',$$

which yields Condition IV.

For Condition V, note that, since

$$\varphi_1(x) = \frac{\phi'(f(x))}{f(x)} - \frac{\phi(f(x))}{f^2(x)},$$

we have

$$\chi_1(x) = f(x)\varphi_1(x) = \phi'(f(x)) - \frac{\phi(f(x))}{f(x)},$$

hence

$$\frac{d}{dx}\chi_1(x) = f'(x)\left[\phi''(f(x)) - \frac{\phi'(f(x))}{f(x)} + \frac{\phi(f(x))}{f^2(x)}\right].$$

Thus, since f and f' are bounded and (46) and (49) hold, we get for some $L > 0$ that $|d\chi_1/dx| \leq L$, which gives Condition V with $\beta = 1$.

Condition VII holds trivially and Conditions VI and VIII by assumption.

We get after a little calculation that $Y_1 = \phi'(f(X_1)) - E\phi'(f(X))$. Moreover, as long as ϕ' is not constant on I, our assumptions imply that $0 < Var\left(\phi'(f(X))\right) < \infty$.

In conclusion, Theorem 1 applies and gives Theorem 3. q.e.d.

Similarly, Theorem 2 gives a LIL for $\tilde{\theta}_n(\lambda h_n)$ uniform in λ. We omit the routine statement and derivation.

Remark 1. The results in this paper are not applicable to estimation of the negative entropy $\tau(f) = \int_{\mathbb{R}} f(x)\log(f(x))\,dx$, when f is not bounded away from 0 on its support. Assuming a smooth enough density, Eggermont & LaRiccia (1999), using entirely different methods, established best asymptotic normality and a law of the iterated logarithm for the kernel density function estimator

$$\tau(\widehat{f}_{n,h}) = \int_{\mathbb{R}} \widehat{f}_{n,h}(x)\log(\widehat{f}_{n,h}(x))\,dx, \tag{51}$$

where $\widehat{f}_{n,h}$ is defined as in (43) with $K(u) = 2^{-1}\exp(-|u|)$, $u \in \mathbb{R}$, is the double exponential kernel.

Remark 2. To compare our conditions with those of Grübel (1994) (and also with those of Bickel & Ritov (1988), that we consider below), it is interesting to remark the following fact (see also

Lemma 1 in the last mentioned article): *Suppose f and H are measurable real valued functions, both defined on $\mathbb{R}$, that H is integrable, and that for some $\delta > 0$ for all $x \in \mathbb{R}$,*

$$|f(x+y) - f(x)| \leq |y| H(x) \tag{52}$$

whenever $|y| \leq \delta$. Then, f is bounded. The proof is as follows. Since f is continuous it is bounded by some M on $[-\delta, \delta]$. Since H is Lebesgue integrable, there exists a double sequence of numbers $\{\xi_n\}_{n=-\infty}^{\infty}$ such that $\xi_n \in \left(\frac{n\delta}{2}, \frac{(n+1)\delta}{2}\right]$ and $\delta \sum_{n=-\infty}^{\infty} H(\xi_n) < \infty$. Choose any $x \in \mathbb{R}$. Without loss of generality we can assume that $x > \delta$. Now for some ξ_n, $n > 0$, $x \in (\xi_{n-1}, \xi_n]$. We get

$$|f(x)| \leq |f(x) - f(\xi_0)| + |f(\xi_0)| = |f(\xi_n - (\xi_n - x)) - f(\xi_0)| + |f(\xi_0)|$$

$$\leq |f(\xi_n - (\xi_n - x)) - f(\xi_n)| + \sum_{k=1}^{n} |f(\xi_k) - f(\xi_{k-1})| + |f(\xi_0)|,$$

which by (52) and $\xi_0 \in [-\delta, \delta]$ is $\leq \delta \sum_{n=-\infty}^{\infty} H(\xi_n) + M$.

Other similar functionals that fall within the framework of Section 2 have been considered in the literature. For instance, Cheng & Serfling (1981) considered $T(f) = \int \phi(x)\psi(F(x))f^2(x)dx$. Obviously, Theorems 1 and 2 apply and give a CLT and a LIL uniform in bandwidth for $T(f)$. Besides the Grübel example, the only other example that we shall consider will be the integrated squares of derivatives of a density.

4 The functionals $T(F) = \int_{\mathbb{R}} \left(f^{(k)}(x)\right)^2 dx$.

If f is smooth enough, e.g., if the derivatives $f^{(s)}$ exist and $\lim f^{(s)}(x) = 0$ as $x \to \pm\infty$, $0 \leq s \leq 2k$, then, by integration by parts,

$$T(F) = \int_{\mathbb{R}} \left(f^{(k)}(x)\right)^2 dx = (-1)^k \int_{\mathbb{R}} f^{(2k)}(x) f(x) dx, \tag{53}$$

which falls within the general scheme considered above. In this case, the function ϕ is

$$\phi(x, y_0, \ldots, y_{2k+1}) = (-1)^k y_{2k+1} \tag{54}$$

and the estimators $T_n(h)$ corresponding to this function are

$$T_n(h) = \frac{(-1)^k}{n(n-1)} \frac{1}{h^{2k+1}} \sum_{1 \le i \ne j \le n} K^{(2k)} \left(\frac{X_i - X_j}{h} \right). \tag{55}$$

Actually, these estimators were proposed by Hall & Marron (1987), who computed their variance and bias under different smoothness assumptions on f and K. Direct application of Theorem 1 turns out to be too crude in this case because the fact that the function ϕ is linear (instead of simply Lipschitz) implies that $\gamma_n^{(1)} = 0$ in (19) and allows for a better estimation of the bias than in the proof of Theorem 1, Step 2. Following the proof of this theorem with a modification in Step 2, we do obtain a central limit theorem for $T_n(\lambda h_n) - T(F)$ which, in view of a result of Bickel & Ritov (1988), is essentially best possible. The result we obtain generalizes to derivatives of the density and to uniformity in bandwidth a result of Giné & Nickl (2007b) who show, in particular, that, in order to achieve the bias improvement of Bickel & Ritov (1988) over Hall & Marron (1987) for the estimation of the integral of the square of a density, it is not necessary to combine two estimators. We treat the bias as in Giné & Nickl (2007b), by applying a Fourier analytic lemma from Giné & Nickl (2007a) and this allows for weakest smoothness conditions on f (see the comment below Theorem 4). The variance part can be treated in a number of ways, in particular as a consequence of results in Giné & Mason (2007a) or in Du & Schick (2007). However, it seems more appropriate in this article to demonstrate how the proof of Theorem 1, with some variations, specializes to the present particular case.

Here is how the general scheme of Section 2 applies and what we get. In the notation of previous sections, for ϕ as in (54), and with $r = 2k + 1$, we have:

$$\varphi_m(x) = 0 \quad \text{for} \quad m = 0, 1, \ldots, 2k, \quad \varphi_{2k+1}(x) = (-1)^k,$$

$$\xi(X_i) = (-1)^k f^{(2k)}(X_i) - T(F), \quad \chi_m(y) = 0, \quad \text{for} \quad m = 1, \dots, 2k,$$

$$\chi_{2k+1}(y) = (-1)^k f^{(2k)}(y),$$

$$\xi_m(X_i) = 0, \quad \text{for} \quad m = 1, \dots, 2k, \quad \xi_{2k+1}(X_i) = (-1)^k f^{(2k)}(X_i) - T(F) = \xi(X_i),$$

$$Y_i = \xi(X_i) + \xi_{2k+1}(X_i) = 2(-1)^k \left[f^{(2k)}(X_i) - E f^{(2k)}(X_i) \right]. \tag{56}$$

This gives $\gamma_n^{(1)}(\lambda) = 0$, as mentioned above, and also

$$(-1)^k \gamma_n^{(2)}(\lambda) = \frac{1}{n} \sum_{i=1}^n \frac{1}{(\lambda h_n)^{2k+1}} \int K^{(2k)} \left(\frac{X_i - y}{\lambda h_n} \right) f(y) dy - \frac{1}{n} \sum_{i=1}^n f^{(2k)}(X_i),$$

$$(-1)^k \gamma_n^{(3)}(\lambda)$$

$$= (-1)^k T_n(\lambda h_n) - \frac{1}{n} \sum_{i=1}^n \frac{1}{(\lambda h_n)^{2k+1}} \int K^{(2k)} \left(\frac{X_i - y}{\lambda h_n} \right) f(y) dy - \frac{1}{n} \sum_{i=1}^n (f^{(2k)}(X_i) - E f^{(2k)}(X)),$$

and

$$T_n(\lambda h_n) - T(F) - \frac{1}{n} \sum_{i=1}^n Y_i = \sum_{i=2}^3 \gamma_n^{(i)}(\lambda) = \gamma_n^{(2)}(\lambda) + \gamma_n^{(3)}(\lambda).$$

Note that $E\gamma_n^{(3)}(\lambda) = 0$, so that the bias is contained in $\gamma_n^{(2)}(\lambda)$. A close look at Step 3 in the proof of Theorem 1 shows that if Conditions V with $\beta = \alpha$, VI and VII hold, $\|f\|_\infty < \infty$ and $nh_n >> \log h_n^{-1}$, then statements (35) and (36) hold for $m = 2k + 1$ (for other values of m everything vanishes), and we conclude

$$\sup_{a \le \lambda \le b} \left| \gamma_n^{(3)}(\lambda) \right| = O_p \left(h_n^\alpha \sqrt{\frac{\log h_n^{-1}}{n}} \right) + O_p \left(\frac{\log h_n^{-1}}{n h_n^{2k+1/2}} \right). \tag{57}$$

Instead of proceeding as in Step 2, we consider separately $E\gamma_n^{(2)}(\lambda) = ET_n(F, \lambda h_n) - T(F)$ (the bias) and then estimate $E \sup_{a \le \lambda \le b} \left| \gamma_n^{(2)}(\lambda) - E\gamma_n^{(2)}(\lambda) \right|$ by means of Theorem 1A. Proceeding by analogy with the proof of (35), we obtain, for this sup (recall that we take $\beta = \alpha$),

$$\sup_{a \le \lambda \le b} \left| \gamma_n^{(2)}(\lambda) - E\gamma_n^{(2)}(\lambda) \right| = O_p \left(h_n^\alpha \sqrt{\frac{\log h_n^{-1}}{n}} \right). \tag{58}$$

For the bias, assume that K satisfies condition (11) with $s = 2k$, that $f^{(2k)} \in L_p(\mathbb{R})$ for some p and that for $|h| \le h_0$ ($h_0 > 0$) and all x,

$$\left|f^{(2k)}(x+h) - f^{(2k)}(x)\right| \le |h|^\alpha H(x) \quad \text{with} \quad \int_{\mathbb{R}} H^2(x)dx < \infty, \quad \|H\|_\infty < \infty. \tag{59}$$

Then, by Lemma 12 in Giné & Nickl (2007a) and the comment after its proof, with $\bar{f}(x) = f(-x)$, we have $\bar{f}^{(2k)} * f^{(2k)} \in C^{2\alpha}(\mathbb{R})$, assuming 2α is not an integer, and $\bar{f}^{(2k)} * f^{(2k)} \in C^{2\alpha-\delta}(\mathbb{R})$ for any $\delta > 0$ if 2α is an integer. (Or, applying the simpler Lemma 1 in Giné & Nickl (2007b), $\bar{f}^{(2k)} * f^{(2k)} \in C^{2\alpha-\delta}(\mathbb{R})$ for any $\delta > 0$.) Here, $*$ denotes convolution, and C^β, the Banach space of continuous functions with $\{\beta\} \ge 0$ continuous derivatives and with the derivative of order $\{\beta\}$ Lipschitz of order $\beta - \{\beta\}$, where $\{\beta\}$ denotes the largest integer strictly smaller than β. This allows us to deal with the bias in a way similar to Giné & Nickl (2007b), proof of Theorem 1: setting $\bar{f}(x) = f(-x)$, we get, for $0 < \alpha < 1/2$,

$$\begin{aligned}
\left|E\gamma_n^{(2)}(\lambda)\right| &= \frac{1}{(\lambda h_n)^{2k+1}} \left| \int\int K^{(2k)}\left(\frac{x-y}{\lambda h}\right) f(y)f(x)dydx - \int f^{(2k)}(x)f(x)dx \right| \\
&= \left| \int\int K(u)\left(f(x-\lambda hu) - f(x)\right) f^{(2k)}(x)dudx \right| \\
&= \left| \int\int K(u)\frac{(\lambda hu)^{2k}}{(2k)!} f^{(2k)}(x-\theta\lambda hu) f^{(2k)}(x)dudx \right| \\
&= \left| \int\int K(u)\frac{(\lambda hu)^{2k}}{(2k)!} \left(f^{(2k)}(x-\theta\lambda hu) - f^{(2k)}(x)\right) f^{(2k)}(x)dudx \right| \\
&= \frac{(\theta\lambda h)^{2k}}{(2k)!} \left| \int K(u)u^{2k} \left[(\bar{f}^{(2k)} * f^{(2k)})(\theta\lambda hu) - (\bar{f}^{(2k)} * f^{(2k)})(0)\right] du \right| \\
&\le C\frac{(\theta\lambda h)^{2k}}{(2k)!} \left(\int |K(u)u^{2(k+\alpha)}|du\right) |\theta\lambda h|^{2\alpha},
\end{aligned}$$

where $0 \le \theta \le 1$. Moreover, we can replace 2α by $1-\delta$ if $\alpha = 1/2$ and by 1 if $\alpha > 1/2$, to conclude

$$\sup_{a \le \lambda \le b} \left|E\gamma_n^{(2)}(\lambda)\right| = O\left(h_n^{2(k+\alpha')}\right) \tag{60}$$

where for any $\delta > 0$,

$$\alpha' = \alpha \text{ if } 0 < \alpha < 1/2, \ \alpha' = \alpha - \delta \text{ if } \alpha = 1/2 \text{ and } \alpha' = 1/2 \text{ if } \alpha > 1/2. \tag{61}$$

Bickel & Ritov (1988), Lemma 1, show that condition (59) implies that f and its first $2k$ derivatives are uniformly bounded, in particular then, that $f^{(2k)}$ is Lipschitz of order α, that is, condition (59) implies the part of Condition V corresponding to χ_m. Then, (57), (58) and (60), yield the following theorem.

Theorem 4. *Let $k \geq 0$, let $H \in L_2(\mathbb{R})$, $\|H\|_\infty < \infty$, let $1/4 < \alpha \leq 1$ and let α' be given by (61). Assume*

a') f has $2k$ derivatives, $f^{(2k)} \in L_2(\mathbb{R})$, and there is $h_0 > 0$ such that

$$\left| f^{(2k)}(x+h) - f^{(2k)}(x) \right| \leq |h|^\alpha H(x) \tag{62}$$

for all $x \in \mathbb{R}$ and all $|h| \leq h_0$.

b') K is a $2k$ times differentiable symmetric kernel of bounded support that satisfies condition (11) with $s = 2k$, and such that $K^{(2k)}$ is of bounded variation.

c') $\sqrt{n} h_n^{2k+1/2} / \log h_n^{-1} \to \infty$ and $\sqrt{n} h_n^{2k+2\alpha'} \to 0$.

Let T_n be as in (55) and Y_i as in (56). Then, the conclusion of Theorem 1 holds, that is,

$$\sqrt{n} \sup_{a \leq \lambda \leq b} \left| T_n(\lambda h_n) - \int_{\mathbb{R}} (f^{(k)})^2(x) dx - \frac{1}{n} \sum_{i=1}^n Y_i \right| \to 0 \tag{63}$$

in probability, in particular, the processes $\sqrt{n} \left(\theta_n(\lambda h_n) - \int_{\mathbb{R}} (f^{(k)})^2(x) dx \right)$, $\lambda \in [a, b]$, converge in law in $\ell^\infty[a, b]$ to the constant process G, where G is $N\left(0, \mathrm{Var}(2 f^{(2k)}(X_1))\right)$.

The smoothness assumption $\alpha > 1/4$ is the weakest for which efficient estimation of the functionals considered in this section is possible (e.g., Bickel & Ritov (1988), Theorem 2(ii)).

Again, we omit the LIL for this simplified Bickel-Ritov estimator.

5 A suggested practical choice of bandwidths

We assume that each estimator for $1 \le m \le r$,

$$F_{n,h}^{(m)}(X_i) = \frac{1}{n-1}\frac{1}{h_m^m}\sum_{j\neq i} K_m^{(m-1)}\left(\frac{X_i - X_j}{h_m}\right),\ i = 1, \ldots, m,$$

has its own bandwidth $h_m := h_{m,n}$. Ideally we would choose $h_m := h_{m,n}$, $1 \le m \le r$, that satisfies Condition VIII from Section 2 and minimizes

$$E\left(T_n(h_1, \ldots, h_r) - T(F)\right)^2,$$

where $T_n(h_1, \ldots, h_r) = T_n(\mathbf{h})$ is as in (3). However, except in special simple cases, (see Hall & Marron (1987)) this is not feasible. Here is a practical solution. Notice that

$$E\left(T_n(\mathbf{h}) - T(F)\right)^2$$

$$\le 2E\left(T_n(\mathbf{h}) - n^{-1}\sum_{i=1}^{n}\varphi(X_i, F(X_i), F^{(1)}(X_i), ..., F^{(r)}(X_i))\right)^2 + \frac{2}{n}Var\varphi\left(X\right),$$

which assuming that the function $\varphi\left(x, y_0, y_1, \ldots, y_r\right)$ satisfies a global Lipschitz condition with respect to the variables $y_0, \ldots, y_r$, is for some constant $A > 0$,

$$\le A\left[E\left(F_n(X_1) - F(X_1)\right)^2 + \sum_{m=1}^{r} E\left(F_{n,h_m}^{(m)}(X_1) - F^{(m)}(X_1)\right)^2\right] + \frac{2}{n}Var\varphi\left(X\right).$$

We see from this inequality that a reasonable choice of $h_1, \ldots, h_r$ would be one that minimizes

$$E\left(F_{n,h_m}^{(m)}(X_1) - F^{(m)}(X_1)\right)^2$$

for each $1 \le m \le r$. Now with $h = h_m$,

$$E\left(F_{n,h}^{(m)}(X_1) - F^{(m)}(X_1)\right)^2 = E\left(Var\left(F_{n,h}^{(m)}(X_1)|X_1\right)\right) + E\left(E\left(F_{n,h}^{(m)}(X_1)|X_1\right) - F^{(m)}(X_1)\right)^2.$$

We will show that subject to regularity conditions we get that

$$E\left(Var\left(F_{n,h}^{(m)}(X_1)|X_1\right)\right) \sim n^{-1}h^{-2m+1}\left(\int_{\mathbb{R}}(K_m^{(m-1)}(u))^2du\int_{\mathbb{R}} f^2\left(x\right)dx\right) =: n^{-1}h^{-2m+1}C_{1,m}\left(f\right) \tag{64}$$

and

$$E\left(E(F_{n,h}^{(m)}(X_1)|X_1) - F^{(m)}(X_1)\right)^2$$
$$\sim h^{2s+2}\left(\frac{1}{(s+1)!}\int_{\mathbb{R}} u^{s+1}K_m(u)du\right)^2 \int_{\mathbb{R}} (f^{(m+s)}(x))^2 f(x)\,dx =: h^{2(s+1)}C_{2,m,s}(f)\,, \tag{65}$$

where s is to be chosen. Thus we see that

$$E\left(F_{n,h}^{(m)}(X_1) - F^{(m)}(X_1)\right)^2 \sim n^{-1}h^{-2m+1}C_{1,m} + h^{2(s+1)}C_{2,m,s}(f)\,.$$

Hence an approximate minimum of $E\left(F_{n,h}^{(m)}(X_1) - F^{(m)}(X_1)\right)^2$ is obtained by choosing h to minimize $n^{-1}h^{-2m+1}C_{1,m} + h^{2(s+1)}C_{2,m,s}(f)$, giving

$$h = \left(\frac{(2m-1)C_{1,m}(f)}{2(s+1)C_{2,m,s}(f)}\right)^{\frac{1}{2(m+s)+1}} \left(\frac{1}{n}\right)^{\frac{1}{2(m+s)+1}}. \tag{66}$$

In particular, by setting $s+m = 2r-1$ we get that

$$h^{4r-1} = \frac{(2m-1)C_{1,m}(f)}{2(s+1)C_{2,m,s}(f)}\frac{1}{n} =: \frac{c_m}{n}. \tag{67}$$

This choice of $h = h_{m,n}$ satisfies Condition VIII for $\alpha > 1/2$ and each $1 \le m \le r$. The unknown constants $C_{1,m}(f)$ and $C_{2,m,s}(f)$ can be estimated consistently using the estimators that we have just developed.

Here are some conditions that imply (64) and (65):

Proposition 4. *Let $s \in N$. If a) f is $m+s$ times continuously differentiable, b) $f, f^{(1)}, \ldots, f^{(m+s)}$ are bounded, and c) the kernel K is integrable, has compact support, satisfies condition (11) and has $m-1$ continuous derivatives. Then conditions (64) and (65) for this s are satisfied.*

In what follows we sketch the proof of Proposition 4. First we shall state a little lemma that will be helpful in our derivations.

Lemma 1 *Let τ be an almost everywhere differentiable function on $\mathbb{R}$ with derivative τ' such that*

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}}\left|\frac{\tau(x+\varepsilon)-\tau(x)}{\varepsilon} - \tau'(x)\right|^k f(x)\,dx = 0 \tag{68}$$

for $k \geq 1$. Then for any finite measure κ on $\mathbb{R}^2$ with compact support

$$\lim_{h\to 0}\int_{\mathbb{R}}\left|\int_{\mathbb{R}^2}\left(\frac{\tau\left(x-hu\xi\right)-\tau\left(x\right)}{-u\xi h}-\tau'\left(x\right)\right)\kappa\left(du,d\xi\right)\right|^k f\left(x\right)dx=0 \tag{69}$$

for the same k for which (68) holds.

After normalizing the measure here, this lemma follows directly from Jensen's inequality and the Lebesgue dominated convergence theorem.

Proposition 5 *Assume that K is integrable, has compact support and satisfies (11) for some integer $s \geq 1$. Then, for every s times continuously differentiable function g such that $g^{(s)}$ satisfies (68) for some $k \geq 1$ with almost everywhere existing derivative $g^{(s+1)}$ in $L_1(f)$, we have*

$$\lim_{h\to 0} h^{-(s+1)}\int_{\mathbb{R}}\left|\frac{1}{h}\int_{\mathbb{R}} g\left(y\right)K\left(\frac{x-y}{h}\right)dy-g\left(x\right)\right|^k f\left(x\right)dx$$

$$=\frac{\left(-1\right)^{s+1}}{(s+1)!}\left|\int_{\mathbb{R}} u^{s+1}K(u)du\right|^k\int_{\mathbb{R}}\left|g^{(s+1)}\left(x\right)\right|^k f\left(x\right)dx.$$

Proof Choose $k \geq 1$. As in the proof of Proposition 1 by a change of variable and then Taylor's theorem and (11)

$$\int_{\mathbb{R}}\left|\frac{1}{h}\int_{\mathbb{R}} g\left(y\right)K\left(\frac{x-y}{h}\right)dy-g\left(x\right)\right|^k f(x)dx$$

$$=\frac{\left(-1\right)^s h^s}{(s-1)!}\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\int_0^1\left(1-\xi\right)^{s-1}g^{(s)}\left(x-\xi hu\right)u^s d\xi K\left(u\right)du\right|^k f\left(x\right)dx. \tag{70}$$

Introducing the finite measure κ with compact support in $\mathbb{R}^2$ and total mass $1/\left(s(s+1)\right)$,

$$\kappa\left(d\xi,du\right)=\left(1-\xi\right)^{s-1}\xi u^{s+1}K\left(u\right)d\xi du,$$

we see that the last integral is

$$=\frac{\left(-1\right)^{s+1}h^{s+1}}{(s-1)!}\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\int_0^1\left[\left(\frac{g^{(s)}\left(x-\xi hu\right)-g^{(s)}\left(x\right)}{-h\xi u}\right)-g^{(s+1)}\left(x\right)\right]\kappa\left(d\xi,du\right)\right.$$

$$\left.+\frac{1}{s(s+1)}\left(\int_{\mathbb{R}} u^{s+1}K(u)du\right)g^{(s+1)}\left(x\right)\right|^k f\left(x\right)dx.$$

Applying Lemma 1 we get (70). q.e.d.

Here is another easy lemma that will be good enough for our purposes.

Lemma 2 *Let ρ be a measurable real-valued function on $\mathbb{R}$ such that*

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}}\left|\rho\left(x+\varepsilon\right)-\rho\left(x\right)\right|f\left(x\right)dx=0 \tag{71}$$

Then, for any function H in $L_1(\mathbb{R})$ and with with compact support,

$$\lim_{h\to 0}\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\rho\left(x-uh\right)-\rho\left(x\right)\right|H\left(u\right)f\left(x\right)dudx=0. \tag{72}$$

Towards verifying (64), note that the hypotheses in Proposition 4 imply that K_m satisfies Condition VII, has compact support and $\left(K_m^{(m-1)}\right)^2\in L_1\left(\mathbb{R}\right)$. We get that

$$\begin{aligned}Var\left(F_{n,h}^{(m)}(X_1)|X_1\right)=\frac{1}{n-1}\Bigg(&\frac{1}{h^{2m}}\int_{\mathbb{R}}\left(K_m^{(m-1)}\right)^2\left(\frac{X_1-y}{h}\right)f\left(y\right)dy\\&-\left(\frac{1}{h^m}\int_{\mathbb{R}}K_m^{(m-1)}\left(\frac{X_1-y}{h}\right)f\left(y\right)dy\right)^2\Bigg),\end{aligned}$$

which by Condition VII is

$$\frac{1}{n-1}\left(\frac{1}{h^{2m}}\int_{\mathbb{R}}\left(K_m^{(m-1)}\right)^2\left(\frac{X_1-y}{h}\right)f\left(y\right)dy-\left(\frac{1}{h}\int_{\mathbb{R}}K_m\left(\frac{X_1-y}{h}\right)f^{(m-1)}\left(y\right)dy\right)^2\right).$$

Since the density f satisfies (71) by hypotheses a) and b) in Proposition 4, we get that, by Lemma 2 with $H=\left(K_m^{(m-1)}\right)^2$ and $\rho=f$,

$$\begin{aligned}\frac{1}{h}E\int_{\mathbb{R}}\left(K_m^{(m-1)}\right)^2\left(\frac{X_1-y}{h}\right)f\left(y\right)dy&=\int_{\mathbb{R}}\int_{\mathbb{R}}\left(K_m^{(m-1)}\left(u\right)\right)^2f\left(x-uh\right)duf\left(x\right)dx\\&\sim\left(\int_{\mathbb{R}}(K_m^{(m-1)}(u))^2du\int_{\mathbb{R}}f^2\left(x\right)dx\right),\ \text{as } h\searrow 0.\end{aligned}$$

Moreover, by boundedness of $f^{(m-1)}$, clearly,

$$\int_{\mathbb{R}}\left(\frac{1}{h}\int_{\mathbb{R}}K_m\left(\frac{x-y}{h}\right)f^{(m-1)}\left(y\right)dy\right)^2f\left(x\right)dx=O(1).$$

Therefore we readily see that (64) holds.

Finally, note that (65) follows directly from Proposition 5 with $K = K_m$, $g = f^{(m-1)}$ and $k = 2$, thus completing the proof of Proposition 4.

6 Conclusion

We have presented an approach based on the notion of a local U–statistic to establish the uniform in bandwidth asymptotic normality of a large class of estimators of integral functionals of the density function. This class includes those that arise in plug-in data driven bandwidth selection procedures in density estimation and appear as part of the variance in nonparametric location and regression estimators based on linear rank statistics. We also worked out in detail a couple of examples and suggested a potentially practical bandwidth selection procedure. We expect that our methods will prove to have many more applications.

Acknowledgements The authors are grateful to the two referees for pointing out a number of oversights as well as providing useful suggestions that improved the presentation of the results. Evarist Giné's research was partially supported by an NSA Grant that of David M. Mason by an NSF Grant.

References

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

Bickel, P. J. & Ritov, Y. (2003). Nonparametric estimators that can be 'plugged in'. *Ann. Statist.* **31**, 1033-1053.

Birgé, L. & Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Probab.* **23**, 11–29.

Cheng, K. F. & Serfling, R. J. (1981). On estimation of a class of efficacy-related parameters. *Scand. Actuar. J.* 83–92.

de la Peña, V. & Giné, E. (1999). *Decoupling: from dependence to independence.* Springer, New York.

Devroye, L. & Györfi, L. (1985). *Nonparametric density estimation: the L_1 View.* John Wiley and Sons, New York.

Dmitriev, Yu. G., Koshkin, G. M., Simakhin, V. A., Tarashenko, F. P. & Shulenin, V. P. (1974). *Nonparametric estimation of functionals on the basis of Stationary samples.* [in Russian.] Izdat. Tomsk. Gos. Univ., Tomsk.

Du, J. & Schick, A. (2007). Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *International J. Statistics and Management Systems* **2**, 67-87.

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

Giné, E. & Mason, D. (2007a). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.* **35**, 1105-1145.

Giné, E. & Mason, D. (2007b). Laws of the iterated logarithm for local U-statistic processes. *J. Theoret. Probab.* **20**, 457-485.

Giné, E. & Nickl, R. (2007a). Uniform central limit theorems for kernel density estimators. *Probab. Theory Rel. Fields*, to appear.

Giné, E & Nickl, R. (2007b). A simple adaptive estimator of the integrated square of a density. *Bernoulli*, to appear.

Grübel, R. (1994). Estimation of density functionals. *Ann. Inst. Statist. Math.* **46**, 67–75.

Hall, P & Marron, J. S. (1987). Estimation for integrated squared density derivatives. *Statist. Probab. Lett.* **6,** 109–155.

Jurečková, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **40**, 1889–1900.

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

Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM Probab.* Stat. **9**, 1–18 (electronic).

Levit, B. Ya. (1978). Asymptotically efficient estimation of nonlinear functionals. (Russian) P*roblems Inform. Transmission* **14**, 65–72

Major, P. (2006) An estimate on the supremum of a nice class of stochastic integrals and U-processes. *Probab. Theory Rel. Fields*, online first (2005).

Nolan, D. & Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15**, 780-799.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer, New York.

van der Vaart, A. & Wellner, J. (1996). *Weak convregence of stochastic processes with applications to statistics.* Springer, New York.

Appendix

We shall record here some definitions and results that are needed in the statements and proofs of our results. Let P be a probability measure on a measure space $(S, \mathcal{S})$. We say that a class of measurable P-square integrable functions $\mathcal{F}$ defined on $(S, \mathcal{S})$ is VC–type (or VC for short) with respect to an envelope F (meaning a measurable function F such that $|f| \leq F$ for all $f \in \mathcal{F}$) if the covering number $N(\mathcal{F}, L_2(Q), \varepsilon)$, defined as the smallest number of $L_2(Q)$ open balls of radius

ε required to cover $\mathcal{F}$, satisfies

$$N(\mathcal{F}, L_2(Q), \varepsilon) \le \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v, \quad 0 < \varepsilon \le 2\|F\|_{L_2(Q)}, \tag{73}$$

for some $A,\ v > 0$, for every probability measure Q on $\mathcal{S}$ for which $F \in L_2(Q)$. If (73) holds for $\mathcal{F}$, then we say that the VC class $\mathcal{F}$ admits the characteristics A and v. (VC is for Vapnik and Červonenkis. Nolan & Pollard (1987) use the term Euclidean for VC type.)

We introduce the following notation and facts. Let $X, X_i,\ i \in \mathbf{N}$, be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. For a kernel L of k variables we set

$$U_n^{(k)}(L) = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I_n^k} L(X_{i_1}, \ldots, X_{i_k}) \tag{74}$$

where $\mathbf{i} = (i_1, \ldots, i_m)$ and $I_n^m = \{(i_1, \ldots, i_m) : 1 \le i_j \le n, i_j \ne i_k \ \text{ if } \ j \ne k\}$. Assume now that L is a function of m variables, symmetric in its entries. Then, for $1 \le k \le m$, the Hoeffding projections with respect to P are defined as

$$\pi_k L(x_1, \ldots, x_k) = (\delta_{x_1} - P) \times \cdots \times (\delta_{x_k} - P) \times P^{m-k}(L) \tag{75}$$

and $\pi_0 L = EL(X_1, \ldots, X_m)$. The *Hoeffding decomposition* states the following:

$$U_n^{(m)}(L) - EL = \sum_{k=1}^{m} \binom{m}{k} U_n^{(k)}(\pi_k L). \tag{76}$$

Assuming that L is in $L_2(P^m)$, this is an orthogonal decomposition and $E\left(\pi_k L \,|X_2, \ldots, X_k\right) = 0$ for $k \ge 1$, that is, the kernels $\pi_k L$ are canonical for P. Also, $\pi_k,\ k \ge 1$, are nested projections in L_2^0 and in L^2, that is, $\pi_k \circ \pi_\ell = \pi_k$ if $k \le \ell$, and $E(\pi_k L)^2 \le E(L - EL)^2 \le EL^2$.

Moment bounds

For any functional Ψ defined on a class of functions $\mathcal{F}$ set $\|\Psi(f)\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\Psi(f)|$. We record the following moment bounds of Giné & Mason (2007a,b) for $\|U_n^{(k)}(\pi_k f)\|_{\mathcal{F}}$, when $\mathcal{F}$ is of VC type.

Theorem 1A. (Theorem 8 of Giné & Mason (2007a) and Corollary 1 of Giné & Mason (2007b)) *Let $\mathcal{F}$ be a collection of measurable functions $S^m \mapsto \mathbb{R}$ symmetric in their entries and with absolute values bounded by $M < \infty$, and let P be any probability measure on $(S, \mathcal{S})$ (with X_i i.i.d. P). Assume $\mathcal{F}$ is VC with respect to the envelope $F \equiv M$, with characteristics A and v, that without loss of generality we take so that $A \geq e^m$ and $v \geq 1$. Then there exist constants $D_1 := D_1(m, A, v, M)$ and $D_2 = D_2(m, A, v, M)$ such that,*

$$\left(E\left[n^{k/2}\left\|U_n^{(k)}(\pi_k f)\right\|_{\mathcal{F}}\right]^d\right)^{1/d} \leq D_1\sigma\left[\log\left(\frac{AM}{\sigma}\right)\right]^{k/2},\ k = 0, 1, \ldots, m,\ d = 1, 2, \qquad (77)$$

assuming $n\sigma^2 \geq D_2 \log\left(\frac{AM}{\sigma}\right)$, where σ^2 is any number satisfying $\|P^m f^2\|_{\mathcal{F}} \leq \sigma^2 \leq M^2$.

Evarist Giné, Departement of Mathematics, University of Connecticut, 196 Auditorium Road, Storrs, CT 06269-3009, USA, gine@math.uconn.edu