The Annals of Statistics
1995, Vol. 23, No. 3, 990–999

DENSITY ESTIMATION UNDER LONG-RANGE DEPENDENCE

BY SÁNDOR CSÖRGŐ[1] AND JAN MIELNICZUK[2]

University of Michigan and Polish Academy of Sciences

Dehling and Taqqu established the weak convergence of the empirical process for a long-range dependent stationary sequence under Gaussian subordination. We show that the corresponding density process, based on kernel estimators of the marginal density, converges weakly with the same normalization to the derivative of their limiting process. The phenomenon, which carries on for higher derivatives and for functional laws of the iterated logarithm, is in contrast with independent or weakly dependent situations, where the density process cannot be tight in the usual function spaces with supremum distances.

1. Introduction and results. Let $Z_1, Z_2, \dots$ be a stationary Gaussian process with mean $E(Z_1) = 0$, variance $E(Z_1^2) = 1$ and covariance function $r(k) = E(Z_1 Z_{k+1}) = k^{-\alpha} L(k)$, $k \in \mathbb{N} = \{1, 2, \dots\}$, where $0 < \alpha < 1$ and L is a function on $[1, \infty)$ that is slowly varying at infinity and is positive in some neighborhood of infinity. Such a sequence $\{Z_j\}_{j=1}^{\infty}$ exhibits long-range dependence in the sense that $\sum_{k=1}^{\infty} |r(k)| = \infty$.

Let G be an arbitrary real-valued Borel measurable function on the real line $\mathbb{R}$, and consider the subordinate stationary process $X_1 = G(Z_1)$, $X_2 = G(Z_2), \dots$ with marginal distribution function $F(x) = P\{X \le x\}$, $x \in \mathbb{R}$, where $X = G(Z)$ and Z is a standard normal random variable. The asymptotic distribution of the sums $\sum_{j=1}^{n} X_j$ is now well understood due to the principal contributions by Rosenblatt (1961), Taqqu (1975, 1979) and Dobrushin and Major (1979); compare also the references in Rosenblatt (1991).

On the other hand, Dehling and Taqqu (1989) considered estimating F using the sample distribution function $F_n(x) = n^{-1} \sum_{j=1}^{n} I\{X_j \le x\}$, $x \in \mathbb{R}$, where I is the indicator function. Let φ denote the standard normal density. Write $F_n(x) - F(x) = n^{-1} \sum_{j=1}^{n} B_x(Z_j)$, where $B_x(\cdot) = I\{G(\cdot) \le x\} - F(x)$, and in the weighted $\mathscr{L}^2$-space $\mathscr{L}^2(\mathbb{R}, \varphi)$ consider the Fourier–Hermite expansion

$$B_x(\cdot) = \sum_{k=m(x)}^{\infty} J_k(x) \frac{H_k(\cdot)}{k!},$$

Received December 1992; revised October 1994.

[1]Partially supported by NSF Grant DMS-92-08067.

[2]Work done while this author was Visiting Associate Professor, Department of Statistics, University of Michigan.

AMS 1991 *subject classifications*. Primary 62G07, 62M99; secondary 60F17.

Key words and phrases. Long-range dependence, Gaussian subordination, kernel density estimators, weak convergence in supremum norm, degenerate limiting processes, Hermite polynomials.

where

$$H_k(y) = (-1)^k \exp\left(\frac{y^2}{2}\right) \frac{d^k \exp(-y^2/2)}{dy^k}, \qquad y \in \mathbb{R},$$

is the k-th Hermite polynomial, $J_k(x) = E(H_k(Z)B_x(Z))$, $k = 0, 1, \dots,$ and $m(x) \in \mathbb{N}$ if $0 < F(x) < 1$, $x \in \mathbb{R}$. Then the Hermite rank of the class $\{B_x(\cdot): x \in \mathbb{R}\}$ is

$$m := \min\{m(x): J_{m(x)}(x) \neq 0 \text{ for some } x \in \mathbb{R}, 0 < F(x) < 1\} \in \mathbb{N},$$

and the long-range dependence condition for the sequence $\{X_j\}_{j=1}^{\infty}$ becomes $\alpha < 1/m$. Under this condition, for $d_{m,n}^2 = \mathrm{Var}(\sum_{j=1}^n H_m(Z_j))$ one has the asymptotic equality $d_{m,n}^2 \sim n^{2-m\alpha}L^m(n)/C^2(m, \alpha)$, with a convergence relation meant as $n \to \infty$, and hence

$$(1.1) \qquad \frac{n}{d_{m,n}} \sim C(m, \alpha)\frac{n^{m\alpha/2}}{L^{m/2}(n)} \quad \text{where} \quad C(m, \alpha) = \sqrt{\frac{(1 - m\alpha)(2 - m\alpha)}{2m!}}$$

[cf. Taqqu (1975), Lemma 3.1]. Let $\mathscr{D}[-\infty, +\infty]$ be the nonseparable metric space of all real functions defined on $[-\infty, +\infty]$ that are right-continuous in $\{-\infty\} \cup \mathbb{R}$ and have left-side limits in $\mathbb{R} \cup \{+\infty\}$ such that $\sup\{|g(x) - h(x)|: x \in \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}\}$ is the distance between $g, h \in \mathscr{D}[-\infty, +\infty]$, and let $\to_{\mathscr{D}}$ denote convergence in distribution with respect to the σ-algebra generated by the set of open balls for this metric. The statistically relevant special case of Theorem 1.1 of Dehling and Taqqu (1989) then states that

$$(1.2) \qquad t_n(\cdot) = \frac{n}{d_{m,n}}\{F_n(\cdot) - F(\cdot)\} \to_{\mathscr{D}} J_m(\cdot)Y_m \quad \text{in } \mathscr{D}[-\infty, +\infty]$$

if $\alpha < 1/m$, where Y_m is $1/m!$ times the value at $t = 1$ of a Hermite process of rank m, given for each argument $t \in [0, 1]$ as a multiple Wiener–Itô–Dobrushin integral. We have $E(Y_m) = 0$ and $E(Y_m^2) = 1/(m!)^2$ for all $m \in \mathbb{N}$, the random variable Y_1 is normal, but $Y_2, Y_3, \dots$ are not normally distributed.

Suppose F has a Lebesgue density f on $\mathbb{R}$, and we want to estimate it from the observations $X_1, \dots, X_n$, using the standard kernel estimator $f_n(x) = \sum_{j=1}^n K((x - X_j)/b_n)/(nb_n)$, $x \in \mathbb{R}$, where $\{b_n\}$ is a sequence of positive bandwidths converging to zero and K is some, not necessarily positive integrable function such that $\int_{-\infty}^{\infty} K(s)\,ds = 1$. Also, *we assume throughout that* K *has a continuous derivative* K' *on* $\mathbb{R}$ *and* $K(s) = 0$ *for all* $s \notin (-A, A)$, *for some* $0 < A < \infty$, *and that the derivative* J_m' *is bounded and uniformly continuous* *on* $\mathbb{R}$. The relation

$$
(1.3)\quad \begin{aligned} D_n^0(x) &:= \frac{n}{d_{m,n}}\{f_n(x) - E(f_n(x))\} \\ &= \frac{1}{b_n}\int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) dt_n(y), \qquad x \in \mathbb{R}, \end{aligned}
$$

and (1.2) suggest that D_n^0 itself converges weakly to the derivative $J_m' Y_m$ in $\mathscr{C}(-\infty, +\infty)$. This is the nonseparable metric space of continuous bounded functions defined on $\mathbb{R}$ such that the distance between $g, h \in \mathscr{C}(-\infty, +\infty)$ is $\sup\{|g(x) - h(x)|: x \in \mathbb{R}\}$, and $\to_{\mathscr{D}}$ will denote again convergence in distribution with respect to the σ-algebra generated by the class of open balls for this metric. The goal of the present note is to show that this is so indeed, under two sets of conditions.

For a fixed $k = 0, 1, 2, \dots,$ we say that condition $\mathbb{C}_m^\alpha(k)$ holds if $0 < \alpha < 1/m$ and $n^\beta b_n^{k+1} \to \infty$ for some $0 < \beta < \min(\alpha, 1 - m\alpha)/2$, and we say that condition $\mathbb{C}_{m+1}^{\alpha, f}(k)$ holds if f is bounded on $\mathbb{R}$, $0 < \alpha < 1/(m+1)$ and $n^\beta b_n^{k+1} \to \infty$ for some $0 < \beta < \min(2\alpha, 1 - m\alpha)/2$. Condition $\mathbb{C}_{m+1}^{\alpha, f}(k)$ provides a somewhat bigger room for the choice of the bandwidth b_n than $\mathbb{C}_m^\alpha(k)$: heavier dependence may actually help. Setting

$$
Y_{mn} = \frac{1}{d_{m,n}} \sum_{j=1}^{n} \frac{H_m(Z_j)}{m!} \quad \text{and}
$$

$$
J_{mn}'(x) = \frac{1}{b_n}\int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) J_m'(y)\, dy, \qquad x \in \mathbb{R},
$$

the result for the unbiased empirical density process D_n^0 in (1.3) is the following.

PROPOSITION 1. *If either $\mathbb{C}_m^\alpha(0)$ or $\mathbb{C}_{m+1}^{\alpha, f}(0)$ holds, then*

$$
(1.4a)\qquad \sup_{-\infty < x < \infty} \left| D_n^0(x) - J_{mn}'(x) Y_{mn} \right| \to 0 \quad a.s.,
$$

$$
(1.4b)\qquad \sup_{-\infty < x < \infty} \left| D_n^0(x) - J_m'(x) Y_{mn} \right| \to_P 0
$$

and $D_n^0(\cdot) \to_{\mathscr{D}} J_m'(\cdot) Y_m$ in $\mathscr{C}(-\infty, +\infty)$.

In order to replace $E(f_n)$ by f, we consider a kernel K of order $\ell \in \mathbb{N}$, $\ell \geq 2$. This means that $\int_{-\infty}^{\infty} |s|^\ell |K(s)|\, ds < \infty$ along with $\int_{-\infty}^{\infty} s^j K(s)\, ds = 0$ for $j = 1, \dots, \ell - 1$ and $\int_{-\infty}^{\infty} s^\ell K(s)\, ds \neq 0$, but K is not necessarily symmetric about 0. The main result of the paper for the empirical density process $D_n := n\{f_n - f\}/d_{m,n}$ is as follows.

THEOREM 1. *Assume that f has $\ell - 1$ absolutely continuous derivatives for some $\ell \in \{2, 3, \dots\}$, and the ℓth derivative $f^{(\ell)}$ is bounded on $\mathbb{R}$. Let K be* *a kernel of order* ℓ. *Suppose that* $n^{\gamma} b_n \to 0$ *for some* $\gamma > (m\alpha)/(2\ell)$. *If either* $\mathbb{C}_m^{\alpha}(0)$ *or* $\mathbb{C}_{m+1}^{\alpha, f}(0)$ *holds, then*

$$\sup_{-\infty < x < \infty} |D_n(x) - J'_{mn}(x) Y_{mn}| \to 0 \quad a.s., \tag{1.5a}$$

$$\sup_{-\infty < x < \infty} |D_n(x) - J'_m(x) Y_{mn}| \to_P 0 \tag{1.5b}$$

and $D_n(\cdot) \to_{\mathscr{D}} J'_m(\cdot) Y_m$ *in* $\mathscr{C}(-\infty, +\infty)$.

Since $0 < \sup\{|J'_m(x)|: x \in \mathbb{R}\} < \infty$, the limiting process $J'_m(\cdot)Y_m$ is nonzero. [We also point out that if J'_m is Lipschitz(θ) on $\mathbb{R}$ for some $\theta \in (0, 1]$, i.e., if $|J'_m(x) - J'_m(y)| \le C|x - y|^{\theta}$, $x, y \in \mathbb{R}$, for a constant $C > 0$, then the convergence in (1.4b) and (1.5b) is also a.s.; see the end of the proof of Theorem 1.] To have nonempty statements in the theorem, under condition $\mathbb{C}_m^{\alpha}(0)$ we must require $\ell > m\alpha/\min(\alpha, 1 - m\alpha)$, while under $\mathbb{C}_{m+1}^{\alpha, f}(0)$ we need $\ell > m\alpha/\min(2\alpha, 1 - m\alpha)$. For such ℓ, if $b_n \equiv n^{-\delta}$ and $\alpha < m^{-1}$, the theorem allows $(m\alpha)/(2\ell) < \delta < \min(\alpha, 1 - m\alpha)/2$, while if $\alpha < (m+1)^{-1}$ it allows $(m\alpha)/(2\ell) < \delta < \min(2\alpha, 1 - m\alpha)/2$ for a bounded f. In a practical situation one would need to have information about α and m, besides the degree of smoothness, for permissible choices of b_n. The reader is referred to Cheng and Robinson (1991) and Hall, Lahiri and Truong (1994) for relevant considerations on the bandwidth problem.

If $X_1, X_2, \dots$ are independent and identically distributed [and in many weakly dependent situations; cf. Bradley (1983)], the density process $D_n^{\diamond} = \sqrt{nb_n}\{f_n - f\}/\sqrt{f}$ is asymptotically pointwise normal, and the estimation problem has a local character: the values at different points become asymptotically independent. The empirical process $\sqrt{n}\{F_n - F\}$ has nondegenerate limits in distribution both pointwise and under the supremum norm. However, as Bickel and Rosenblatt (1973) show, the process $D_n^{\diamond}$ has to be renormalized to exhibit the extreme-value theoretic nature, discovered by Woodroofe (1967), of the density estimation problem under the supremum norm. In contrast, under long-range dependence the stochastic order of the supremum norms of the two processes $F_n - F$ and $f_n - f$ is the same, independent of b_n: the density estimation problem becomes global.

Let

$$Q(\alpha, m) = C(m, \alpha)[\Gamma((1+\alpha)/2)/\{\Gamma(\alpha)\Gamma((1-\alpha)/2)\}]^{m/2}/m!,$$

and introduce

$$H'_{\alpha, m} = \left\{ h(\cdot) = J'_m(\cdot) Q(\alpha, m) \int_0^1 \left[\int_{-\infty}^{s} g(y)(s-y)^{-(\alpha+1)/2}\, dy \right]^m ds : \int_{-\infty}^{\infty} g^2(y)\, dy \le 1 \right\},$$

where $C(m, \alpha)$ is as in (1.1). Accompanying (1.2), a special case of Theorem 2 of Dehling and Taqqu (1988) says that the sequence $t_n(\cdot)/(\log\log n)^{m/2}$ is almost surely relatively compact in $\mathscr{D}[-\infty, +\infty]$ with cluster set $H_{\alpha,m}$, which is defined upon replacing $J_m'(\cdot)$ by $J_m(\cdot)$ in the definition of the class $H_{\alpha,m}'$. The "derivative" of their result is as follows.

THEOREM 2. *Under the conditions of Theorem* 1, *the sequence of the processes* $D_n(\cdot)/(\log\log n)^{m/2}$ *is almost surely relatively compact in* $\mathscr{C}(-\infty, \infty)$, *and the set of its limit points is* $H_{\alpha,m}'$.

Let $k = 0, 1, 2, \dots$ be an integer, and consider the problem of estimating $f^{(k)}(x)$ by $f_n^{(k)}(x) = \sum_{j=1}^n K^{(k)}((x - X_j)/b_n)/(nb_n^{k+1})$, $x \in \mathbb{R}$. Define $J_{mn}^{(k+1)}$ as J_{mn}', replacing there J_m' by $J_m^{(k+1)}$. Set $D_n^{(k)} = n\{f_n^{(k)} - f^{(k)}\}/d_{m,n}$. The last theorem carries on the phenomenon encountered for higher derivatives. The analogous proof is omitted.

THEOREM 3. *Assume that, for some* $k \in \{0, 1, 2, \dots\}$ *and* $\ell \in \{2, 3, \dots\}$, *the density function* f *has* $\ell + k - 1$ *absolutely continuous derivatives and that the* $(\ell + k)$*th derivative* $f^{(\ell+k)}$ *is bounded, while the* $(k+1)$*st derivative* $J_m^{(k+1)}$ *is bounded and uniformly continuous on* $\mathbb{R}$. *Let* K *be a kernel of order* ℓ *that has* $k + 1$ *continuous derivatives on* $\mathbb{R}$ *such that* $K(s) = 0$ *for all* $s \notin (-A, A)$ *and* $K^{(j)}(-A) = 0 = K^{(j)}(A)$ *for all* $j = 1, \dots, k$, *for some* $A \in (0, \infty)$. *If* $n^\gamma b_n \to 0$ *for some* $\gamma > (m\alpha)/(2\ell)$, *and if either* $\mathbb{C}_m^\alpha(k)$ *or* $\mathbb{C}_{m+1}^{\alpha,f}(k)$ *holds, then* $\sup_{-\infty<x<\infty}|D_n^{(k)}(x) - J_{mn}^{(k+1)}(x)Y_{mn}| \to 0$ *a.s.*, $\sup_{-\infty<x<\infty}|D_n^{(k)}(x) - J_m^{(k+1)}(x)Y_{mn}| \to_P 0$ *and* $D_n^{(k)}(\cdot) \to_{\mathscr{D}} J_m^{(k+1)}(\cdot)Y_m$ *in* $\mathscr{C}(-\infty, +\infty)$.

When $k = 0$, Theorem 3 reduces to Theorem 1. For meaningful statements in this theorem we must have either $\ell > (k+1)m\alpha/\min(\alpha, 1 - m\alpha)$ or $\ell > (k+1)m\alpha/\min(2\alpha, 1 - m\alpha)$, respectively. Theorem 2 also extends for $D_n^{(k)}/(\log\log n)^{m/2}$. Under the conditions of Theorem 3, in obvious notation, the set of limit points is $H_{\alpha,m}^{(k+1)}$.

We note that an arbitrary marginal distribution function F can be obtained in the instantaneous Gaussian subordination model considered in this paper by choosing the transforming function $G = G_F$ as $G_F(\cdot) = F^{-1}(\Phi(\cdot))$, where Φ is the standard normal distribution function. [Here and below $g^{-1}(y) = \inf\{x \in \mathbb{R}\colon\ s_g g(x) \geq s_g y\}$, $\min(g(-\infty), g(\infty)) < y < \max(g(-\infty), g(\infty))$, for a nonconstant monotone function g on $\mathbb{R}$, where $s_g = 1$ if g is nondecreasing and $s_g = -1$ if g is nonincreasing.] However, G_F is nondecreasing, and $m = 1$ necessarily for any monotone G. On the other hand, Dehling and Taqqu (1992) show in a constructive fashion that an arbitrarily high Hermite rank m may be achieved even with a continuous G.

Something of the nature of Theorem 1 can also be conjectured from the results of Hall and Hart (1990) for linear processes. Our results are closer in spirit to those of Cheng and Robinson (1991) on the pointwise asymptotic distribution of $f_n - f$ when f_n is based on $X_j = G_*(Z_{p+j}, Z_{q+j})$, $j = 1, \dots, n$, where $p, q \geq 0$ are fixed integers and $G_*\colon \mathbb{R}^2 \mapsto \mathbb{R}$ is a Borel function satisfying additional sets of conditions. The technique used by Cheng and Robinson (1991) is similar to that of Rosenblatt (1991). Assuming $L(\cdot) \equiv 1$

and that, in the present instantaneous subordination model, G is a continuously differentiable monotone function for which $G'(G^{-1}(\cdot)) \neq 0$, Rosenblatt (1991) proves (under some conditions on K and b_n) that the finite-dimensional distributions of the process $n^{\alpha/2}\{f_n - E(f_n)\}$ are asymptotically normal. Of course, the monotonicity and continuity of G forces $F(\cdot) = \Phi(s_G G^{-1}(\cdot))$, and then the smoothness conditions on F imply those on G. The main point is that, in view of (1.1), Rosenblatt's result follows from Theorem 1 (under our conditions on K and b_n) for any monotone G, since then $m = 1$ as noted above. Our technique seems to be superior on the whole, and it is based on a precision version of Theorem 3.1 of Dehling and Taqqu (1989) (the names hereafter abbreviated as DT), which appears in (2.1) below, and a variant of (2.1) in (2.3).

2. Proofs. Assuming $0 < \alpha < m^{-1}$ and using all the notation above, define

$$S_n(k, x) = \frac{1}{d_{m,n}} \left\{ \sum_{j=1}^{k} \left[I\{X_j \le x\} - F(x) \right] - J_m(x) \sum_{j=1}^{k} \frac{H_m(Z_j)}{m!} \right\},$$
$$x \in \mathbb{R},\ k = 1, \dots, n.$$

The symbol ε will denote a positive constant, the concrete value of which may change from line to line, but any such value can be made as close to zero as we wish. Also, C and C_ε will denote positive constants, possibly depending on α and m and the second also depending on ε, whose values may be different at each appearance. Let $\kappa = \min(\alpha, 1 - m\alpha)$. A careful study of the proof of Theorem 3.1 of DT (1989) yields the inequality

$$(2.1) \qquad P\left\{ \max_{k \le n} \sup_{-\infty < x < \infty} |S_n(k, x)| > \eta \right\} \le C_\varepsilon n^{-\kappa + \varepsilon} \left\{ 1 + \frac{1}{\eta^{2+\varepsilon}} \right\}$$
$$\text{for all } 0 < \eta \le 1,$$

specifying their κ and replacing their 3 by $2 + \varepsilon$.

Applying this to $n = n_l = \lceil e^l \rceil = \min\{j \in \mathbb{N}\colon\ j \ge e^l\}$ and $\eta = \eta_l = \exp[-l(\kappa - 2\varepsilon)/(2 + \varepsilon)]$, $l = 0, 1, 2, \dots$, by the Borel–Cantelli lemma there exists an a.s. finite random integer l_0 such that

$$\max_{k \le n_l} \sup_x |S_{n_l}(k, x)| \le \exp[-l(\kappa/2 - \varepsilon)]$$

if $l \ge l_0$. Let $n \ge \exp(l_0)$ and let $l = l(n)$ be the unique integer such that $n_{l-1} \le n < n_l$. Then $l = l(n) \to \infty$, and, since $\exp(-l) \le n^{-1}$,

$$|S_n(n, x)| \le \frac{d_{m,n_l}}{d_{m,n}} \max_{k \le n_l} |S_{n_l}(k, x)| \le \frac{d_{m,n_l}}{d_{m,n}} \exp\left[-l\left(\frac{\kappa}{2} - \varepsilon \right) \right]$$
$$\le \frac{d_{m,n_l}}{d_{m,n}} n^{-(\kappa/2 - \varepsilon)}, \qquad x \in \mathbb{R}.$$

Also, using the asymptotic equality above (1.1),

$$\frac{d_{m,n_l}}{d_{m,n}} \le C\frac{L^{m/2}(n_l)n_l^{1-m\alpha/2}}{L^{m/2}(n)n^{1-m\alpha/2}} \le C\frac{L^{m/2}(n_l)}{L^{m/2}(n)} \le Cn^{\varepsilon}n_l^{\varepsilon} \le C_{\varepsilon}n^{2\varepsilon},$$

since $L(n)n^{\varepsilon} \to \infty$ and $L(n)n^{-\varepsilon} \to 0$. Combining the two bounds, we obtain

$$\sup_{-\infty<x<\infty} |S_n(n,x)| = \mathscr{O}(n^{-\kappa/2+\varepsilon}) \quad \text{a.s.} \tag{2.2}$$

Next we assume $0 < \alpha < (m+1)^{-1}$, let $\lambda = \min(\alpha, 1-(m+1)\alpha)$ and introduce

$$S_n^*(k,x) = \frac{1}{d_{m+1,n}}\sum_{j=1}^{k}\left[I\{X_j \le x\} - F(x) - J_m(x)\frac{H_m(Z_j)}{m!} - J_{m+1}(x)\frac{H_{m+1}(Z_j)}{(m+1)!}\right],$$

for $k = 1,\dots,n$ and $x \in \mathbb{R}$. Then, setting $g(x,y) = g(y) - g(x)$ for any real-valued function g on $\mathbb{R}$, the proof of Lemma 3.1 in DT (1989), with m replaced by $m+1$, gives

$$E[S_n^*(k,x,y)]^2 \le C_{\varepsilon}\frac{kF(x,y)}{n^{\lambda+1-\varepsilon}} \quad \text{for all } -\infty < x \le y < \infty.$$

Replacing now the function $\Lambda(x)$ of the proof of Lemma 3.2 in DT (1989) by

$$\Lambda^*(x) = F(x) + \int_{\{s:\, G(s)\le x\}}\left\{\frac{|H_m(s)|}{m!} + \frac{|H_{m+1}(s)|}{(m+1)!}\right\}\varphi(s)\,ds, \qquad x \in \mathbb{R},$$

and using the last moment inequality, a laborious modification of that proof, for S_n^* replacing S_n, yields

$$P\left\{\sup_{-\infty<x<\infty}|S_n^*(l,x)| > \eta\right\} \le C_{\varepsilon}n^{-\lambda+\varepsilon}\left\{\frac{l}{n}\frac{1}{\eta^{2+\varepsilon}} + \left(\frac{l}{n}\right)^{2-m\alpha}\right\}, \qquad l = 1,\dots,n;\ \eta \in (0,1].$$

An analogous variant of the proof of Theorem 3.1 in DT (1989) then leads to

$$P\left\{\sup_{k\le n}\sup_{-\infty<x<\infty}|S_n^*(k,x)| > \eta\right\} \le C_{\varepsilon}n^{-\lambda+\varepsilon}\left\{1 + \frac{1}{\eta^{2+\varepsilon}}\right\} \quad \text{for all } 0 < \eta \le 1. \tag{2.3}$$

The reasoning that led from (2.1) to (2.2) now gives $\sup_x|S_n^*(n,x)| = \mathscr{O}(n^{-\lambda/2+\varepsilon})$ a.s. Thus, since $|Y_{m+1,n}| = \mathscr{O}((\log\log n)^{(m+1)/2})$ a.s. by the law of the iterated logarithm of Mori and Oodaira (1987) as stated in DT (1988), the three inequalities

$$\sup_{|x-y|\le\delta} |S_n(n,x,y)|$$
$$\le \frac{d_{m+1,n}}{d_{m,n}}\Big\{\sup_{|x-y|\le\delta} |J_{m+1}(x,y)||Y_{m+1,n}| + \sup_{|x-y|\le\delta} |S_n^*(n,x,y)|\Big\},$$

$$J_{m+1}^2(x,y) \le (m+1)!|F(x,y)|, \qquad x,y \in \mathbb{R},$$

and

$$\frac{d_{m+1,n}}{d_{m,n}} \le \frac{C_\varepsilon}{n^{\alpha/2-\varepsilon}}$$

imply that

$$\sup_{|x-y|\le\delta} |S_n(n,x) - S_n(n,y)| = \mathcal{O}\big(n^{-\alpha/2+\varepsilon}\{\sqrt{\delta} + n^{-\lambda/2}\}\big) \tag{2.4}$$
$$\text{a.s. for each } \delta > 0,$$

provided the density f of F is bounded on $\mathbb{R}$.

PROOF OF PROPOSITION 1. Starting from (1.3) and using the assumptions on K and J_m, integration by parts gives $D_n^0(x) = J'_{mn}(x)Y_{mn} + R_n(x)$ for every $x \in \mathbb{R}$, where

$$R_n(x) = \frac{\int_{-\infty}^{\infty} K'((x-y)/b_n)S_n(n,y)\,dy}{b_n^2}$$
$$= \frac{\int_{-\infty}^{\infty} K'(t)[S_n(n,x-tb_n) - S_n(n,x)]\,dt}{b_n}$$

since $\int_{-\infty}^{\infty} K'(t)\,dt = 0$. Setting $C' = \int_{-\infty}^{\infty}|K'(t)|\,dt < \infty$, from the first expression, $R_n := \sup\{|R_n(x)|: x \in \mathbb{R}\} \le C' \sup\{|S_n(n,y)|: y \in \mathbb{R}\}/b_n$, so that $R_n \to 0$ a.s., by (2.2), and (1.4a) follows under $\mathbb{C}_m^\alpha(0)$. For a bounded f, the second form of $R_n(x)$ and (2.4) yield $R_n \le C' \sup\{|S_n(n,x) - S_n(n,y)|: |x-y| \le Ab_n\}/b_n \to 0$ a.s., under $\mathbb{C}_{m+1}^{\alpha,f}(0)$, so that (1.4a) follows again. Also, $\Delta_n := \sup\{|J'_{mn}(x) - J'_m(x)|: x \in \mathbb{R}\} \to 0$, by the conditions on J'_m and K, so that (1.4a) implies (1.4b) since $Y_{mn} = \mathcal{O}_P(1)$. This is because, as shown by Dobrushin and Major (1979) and Taqqu (1979), $Y_{mn} \to_{\mathcal{D}} Y_m$ in $\mathbb{R}$, which together with (1.4b) also implies the last statement in a routine fashion. □

PROOF OF THEOREM 1. By Proposition 1, we only have to show that

$$\Delta_n^0 := \frac{n}{d_{m,n}} \sup_{-\infty<x<\infty} |E(f_n(x)) - f(x)| \to 0 \tag{2.5}$$
$$\text{if } n^\gamma b_n \to 0 \text{ for some } \gamma > \frac{m\alpha}{2\ell}.$$

Extending formula (4.5) in Bretagnolle and Huber (1979) to account for a possibly asymmetric kernel K, under the stated conditions on f and K, for all $x \in \mathbb{R}$ we have

$$E(f_n(x)) - f(x) = b_n^{\ell} E_n(x)$$
$$:= b_n^{\ell}\left\{\int_x^{\infty} M_{b_n}^{-}(x-y) f^{(\ell)}(y)\,dy + \int_{-\infty}^{x} M_{b_n}^{+}(x-y) f^{(\ell)}(y)\,dy\right\},$$

where, defining $M^{\pm}(u) = (-1)^{\ell}\int_x^{\infty}(s-u)^{\ell-1}K(\pm s)\,ds/(\ell-1)!$ for $u > 0$ and $M^{\pm}(u) = (-1)^{\ell}M^{\pm}(-u)$ for $u < 0$, $M_{b_n}^{\pm}(t) = M^{\pm}(t/b_n)/b_n$, $t \in \mathbb{R}$. Thus, since $f^{(\ell)}$ is bounded and the support of K is compact, we see that $\sup\{|E_n(x)|: x \in \mathbb{R}\} < \infty$. Hence $\Delta_n^0 \le C_\varepsilon n^{m\alpha/2+\varepsilon} b_n^{\ell}$ by (1.1), and this implies (2.5). [If J_m' is Lipschitz(θ) on $\mathbb{R}$ for some $\theta \in (0,1]$, then (1.4b) and (1.5b) also hold a.s. since, then, for Δ_n of the proof of the proposition we have $\Delta_n = \mathscr{O}(b_n^{\theta}) = o(n^{-\theta\gamma})$ and for $W_{mn} := Y_{mn}/(\log\log n)^{m/2}$ the lim sup of $|W_{mn}|$ is a.s. finite by the log log law of Mori and Oodaira (1987).] □

PROOF OF THEOREM 2. By (1.5a), the assertion follows if the same holds for $U_{mn}(\cdot) := J_{mn}'(\cdot)W_{mn}$. If $V_{mn}(\cdot) := J_m'(\cdot)W_{mn}$, then $\sup_{-\infty<x<\infty}|U_{mn}(x) - V_{mn}(x)| \le \Delta_n|W_{mn}| \to 0$ a.s., since $\Delta_n \to 0$ from the proof of the proposition. The assertion for $V_{mn}(\cdot)$ follows from the Mori–Oodaira law as in DT (1988) for the empirical process. □

Acknowledgments. The comments of three referees and an Associate Editor are appreciated. They helped the presentation of the final version and also spotted an oversight in the original. We thank Peter Hall and Young Truong for sending us the preprint of their paper with S. N. Lahiri.

REFERENCES

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

BRADLEY, R. C. (1983). Asymptotic normality of some kernel-type estimators of probability density. *Statist. Probab. Lett.* **1** 295–300.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137.

CHENG, B. and ROBINSON, P. M. (1991). Density estimation in strongly dependent non-linear time series. *Statist. Sinica* **1** 335–359.

DEHLING, H. and TAQQU, M. S. (1988). The functional law of the iterated logarithm for the empirical process of some long-range dependent sequences. *Statist. Probab. Lett.* **7** 81–85.

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783.

DEHLING, H. and TAQQU, M. S. (1992). Continuous functions whose level sets are orthogonal to all polynomials of a given degree. *Acta Math. Hungar.* **60** 217–224.

DOBRUSHIN, R. L. and MAJOR, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50** 27–52.

HALL, P. and HART, J. D. (1990). Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory Related Fields* **87** 253–274.

HALL, P., LAHIRI, S. N. and TRUONG, Y. K. (1994). On bandwidth choice for density estimation with dependent data. Statistics Report No. SR6/94, Centre for Mathematics and its Applications, Australian National Univ.

MORI, T. and OODAIRA, H. (1987). The functional iterated logarithm law for stochastic processes represented by multiple Wiener integrals. *Probab. Theory Related Fields* **76** 299–310.

ROSENBLATT, M. (1961). Independence and dependence. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* 431–443. Univ. California Press, Berkeley.

ROSENBLATT, M. (1991). *Stochastic Curve Estimation*. IMS, Hayward, CA.

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302.

TAQQU, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50** 53–83.

WOODROOFE, M. (1967). On the maximum deviation of the sample density. *Ann. Math. Statist.* **38** 475–481.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
1444 MASON HALL
ANN ARBOR, MICHIGAN 48109-1027

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
ORDONA 21
01-237 WARSAW
POLAND