The Annals of Statistics
1985, Vol. 13, No. 3, 1050–1077

KERNEL ESTIMATES OF THE TAIL INDEX OF A DISTRIBUTION

BY SÁNDOR CSÖRGŐ, PAUL DEHEUVELS AND DAVID MASON

Szeged University, Université Paris VI, and University of Wisconsin

We introduce a new estimate of the exponent of a distribution whose tail varies regularly at infinity. This estimate is expressed as the convolution of a kernel with the logarithm of the quantile function, and includes as particular cases the estimates introduced by Hill and by De Haan.

Under very weak conditions, we prove asymptotic normality, consistency and discuss the optimal choices of the kernel and of the bandwidth parameter.

1. Introduction and results. Let $X_1, X_2, \cdots$ be independent positive random variables with common distribution function F. We assume that $1 - F$ is regularly varying of order $-a$ in the upper tail, or equivalently that, for any $\lambda > 0$,

$$\lim_{t\uparrow\infty}(1 - F(t\lambda))/(1 - F(t)) = \lambda^{-a}.$$

In the following, we shall study two nonparametric estimates $\tilde{a}_n$ and a_n of a, based on the order statistics $X_{1,n} \le \cdots \le X_{n,n}$ of $X_1, \cdots, X_n$. We first select a nonnegative nonincreasing kernel $\{K(u), u > 0\}$ which satisfies the condition

$$\int_0^\infty K(u)\,du = 1.$$

Next, we choose a bandwidth parameter $\lambda = \lambda_n > 0$. Our first estimate is

$$a_n = a_{n,\lambda} = \left(\int_0^{1/\lambda} \{\log^+ Q_n(1 - v\lambda)\} d\{vK(v)\}\right)^{-1}\left(\int_0^{1/\lambda} K(v)\,dv\right),$$

where $\{Q_n(s), 0 < s \le 1\}$ is the empirical quantile function, defined for $n \ge 1$ by

$$Q_n(s) = X_{k,n} \quad \text{if} \quad (k-1)/n < s \le (k/n), \quad 1 \le k \le n, \quad \text{and} \quad \log^+ x = \log(x \vee 1).$$

If we make the assumption that $K(\cdot)$ is right continuous, routine manipulations show that a_n can be written in the following equivalent form:

$$a_n = a_{n,\lambda} = \left(\sum_{j=1}^n \frac{j}{n\lambda} K\left(\frac{j}{n\lambda}\right)\{\log^+ X_{n-j+1,n} - \log^+ X_{n-j,n}\}\right)^{-1}\left(\int_0^{1/\lambda} K(v)\,dv\right),$$

where we shall put by convention $X_{0,n} = 1$.

We will also consider the following modified version of a_n in which the scale

Received May 1984; revised February 1985.
AMS 1980 *subject classifications.* Primary 62G05; secondary 62G30.
Key words and phrases. Pareto distribution, order statistics, nonparametric estimation, empirical processes, quantile process, regularly varying functions.

term has been corrected for finite samples:

$$\tilde{a}_n = \tilde{a}_{n,\lambda} = \left(\sum_{j=1}^{n} \frac{j}{n\lambda} K\left(\frac{j}{n\lambda}\right)\{\log^+ X_{n-j+1,n} - \log^+ X_{n-j,n}\}\right)^{-1}\left(\sum_{j=1}^{n} \frac{1}{n\lambda} K\left(\frac{j}{n\lambda}\right)\right).$$

There has been considerable recent interest in the problem of estimating a, starting with the pioneering work of Hill (1975) who introduced the estimate

$$\hat{a}_{n,k} = ((1/k)\ \textstyle\sum_{j=1}^{k} \log^+ X_{n-j+1,n} - \log^+ X_{n-k,n})^{-1}.$$

Another estimate was proposed by De Haan (1981) and studied further by De Haan and Resnick (1980):

$$\bar{a}_{n,k} = ((1/\log k)(\log^+ X_{n,n} - \log^+ X_{n-k+1,n}))^{-1}.$$

These estimates are both special cases of a_n(or $\tilde{a}_n$) corresponding, respectively, to $K(u) = 1_{\{0<u<1\}}$ and $H(u) = u^{-1}1_{\{0<u<1\}}$, and $\lambda = k/n$.

The asymptotic normality of $\hat{a}_{n,k}$ was established by Hall (1982) under the assumptions that, for some positive constants C and b,

$$1 - F(t) = Ct^{-a}(1 + O(t^{-b})) \quad \text{as} \quad t \to \infty.$$

Davis and Resnick (1984), S. Csörgő and Mason (1984), and Haeusler and Teugels (1984) have also discussed the asymptotic normality of the Hill estimate. Related work on the subject is to be found in Teugels (1981, 1982) and De Meyer and Teugels (1982), Gawronski and Stadtmüller (1984), Hall and Welsh (1984, 1985), and Welsh (1984).

Because of Mandelbrot's stimulating hypotheses on various speculative prices and related economic quantities (refer to Mandelbrot, 1963), the closely related problem of estimating the characteristic exponent $a \in (0, 2]$ of a stable law has received enormous interest (for an extensive survey see S. Csörgő, 1984; and DuMouchel, 1983). DuMouchel (1983) has shown that even slight changes in the middle of a stable distribution may yield nonrobust properties of the maximal likelihood and related estimators based on the whole sample, and suggests to "let the tails speak for themselves." This is exactly our point in this paper (see Example B below).

Before stating our results, it is useful to give some explanation about the structure of the above estimates. It can be proved (see Hall, 1978; Weissman, 1978), that, for any fixed $k \geq 1$, as $n \to \infty$,

$$\{\log X_{n-j+1} - \log X_{n-j,n},\ 1 \leq j \leq k\} \to \{\omega_j/ja,\ 1 \leq j \leq k\},$$

where the $\{\omega_j,\ 1 \leq j \leq k\}$ are independent exponential random variables with mean one.

It follows that, for a fixed $k \geq 1$, $\hat{a}_{n,k}$ is asymptotically the maximum likelihood estimate of a based on $\{X_{n-j+1,n},\ 1 \leq j \leq k\}$, corresponding to the reciprocal of the mean of the random variables $\{\omega_j/a,\ 1 \leq j \leq k\}$. Likewise De Haan's estimate corresponds to the inverse of the sum of the $\{\omega_j/ja,\ 1 \leq j \leq k\}$ divided by its expectation.

One cannot achieve consistency without letting $k = k_n$ tend to infinity. If this is not the case, the optimal character of Hill's estimate is not justified any more by the preceding arguments. We shall discuss this point later.

On the other hand, the justification of the construction of a_n will follow from the consistency of the empirical quantile function $Q_n(s)$ as an estimator of the quantile function

$$Q(s) = F^{-1}(s) = \inf\{x;\, F(x) \geq s,\ 0 < s < 1\}.$$

It is well known (see, e.g., De Haan, 1970) that $1 - F$ is regularly varying of order $-a$ in the upper tail if and only if $Q(1 - s)$ is regularly varying of order $-1/a$ at 0. By Karamata's theorem (see, e.g., Seneta, 1975), this is equivalent to the existence of functions $c(s)$ and $b(s)$, such that $c(s) \to c$ as $s \to 0$ and $b(s) \to 0$ as $s \to 0$, and that

$$\text{(Q)} \qquad Q(1 - s) = s^{-1/a} c(s) \exp\left(\int_s^1 \frac{b(u)}{u}\, du \right), \quad 0 < s < 1.$$

For suitable K's, it will be seen that, as $\lambda \to 0$, we have

$$\int_0^{1/\lambda} \{\log Q(1 - v\lambda)\}\, d\{vK(v)\} \to 1/a.$$

It follows that it is reasonable to hope that the same occurs when Q is replaced by Q_n with additional conditions on the rate of convergence of $\lambda = \lambda_n$ to zero.

The main problems to be solved are then the following:

1. Prove the consistency of a_n.
2. Obtain asymptotic normality of a_n.
3. Choose the optimal kernel K and bandwidth λ.

We shall give answers to these questions in the sequel. To begin with, we shall make the following basic assumptions on $K(\cdot)$.

(H1) $K(u) \geq 0$ for $0 < u < \infty$.
(H2) $K(\cdot)$ is nonincreasing and right continuous on $(0, \infty)$.
(H3) $\int_0^\infty K(t)\, dt = 1$.
(H4) $\int_0^\infty v^{-1/2} K(v)\, dv < \infty$.

We shall require these conditions for *asymptotic normality* of a_n. It is therefore not surprising that De Haan's estimate has a kernel which does not satisfy (H4), since his estimator has a limiting extreme value distribution (see De Haan and Resnick, 1980). In addition, we may use at times the assumptions:

(H5) There exists a $\Lambda < \infty$ such that $K(u) = 0$ for $u > \Lambda$.
(H6) There exists a $\Lambda < \infty$ such that $dK(u)/du = k(u)$ is defined for $u > \Lambda$ and such that $\lim_{u \uparrow \infty} u^{3/2} k(u) = 0$.
(H7) $\int_0^\infty K^2(u)\, du = 1$.

Next, we shall assume:

(D1) (i) $1 - F$ is regularly varying of order $-a$ in the upper tail, or equivalently, there exists $b(s)$ and $c(s)$ such that $c(s) \to c \in (0, \infty)$ and $b(s) \to 0$

as $s \to 0$, such that the representation given in (Q) above holds; and without loss of generality that

(ii) $Q(0) = \inf\{x; F(x) > 0\} = 1$.

(D2) (i) In the representation for $Q(1-s)$ given in (Q), one has either (H5) is satisfied and $c(s) = c$ (constant) for $0 < s < \varepsilon$ for some $\varepsilon > 0$; or $c(s) = c$ (constant) for $0 < s \le 1$.

(ii) One has either (H6) is satisfied; or the function $\{b(u), 0 < u < 1\}$ may be chosen such that $b(\cdot)$ is bounded on (0, 1).

Condition (D2)(i) is equivalent to assuming that $\log Q(1-s)$ is absolutely continuous (at least) in a right neighborhood of zero. It implies that $b(u)/u$ is then uniquely defined a.e. in this neighborhood.

The condition that $c(s) = c$ (constant) is implied by the following condition on F (see Lemma 6 in the sequel):

(D2′) There exists an $x_0 \ge 0$ such that $dF(x)/dx = f(x)$ exists and $f(x) > 0$ for all $x \ge x_0$. Furthermore,

$$\lim_{x\uparrow\infty} xf(x)/(1 - F(x)) = a.$$

If (D2′) holds, we may take $b(\cdot)$ as

$$b(1 - F(x)) = ((1 - F(x))/xf(x)) - 1/a, \quad x \ge x_0.$$

We shall reduce the proofs to the following case which implies (D2)(i):

(D3) There exists a function $b(u)$ on (0, 1) such that $b(u) \to 0$ as $u \downarrow 0$, and

$$Q(1 - s) = s^{-1/a}\exp\left(\int_s^1 \frac{b(u)}{u}\,du\right), \quad 0 < s \le 1.$$

Before giving our main results, let us remark that the necessity of (D1) for the consistency of a_n has been proved in the case of Hill's estimate by Mason (1982a) In addition, for Hill's estimate, if one requires the asymptotic normality of $\hat{a}_{n,k}$ as $k \to \infty$ and $k/n \to 0$, one has to impose regularity conditions on $Q(1-s)$ in the neighborhood of zero, as shown in Davis and Resnick (1984) and more generally by S. Csörgő and Mason (1984). It turns out that the simplest such regularity condition is (D2)(i).

We shall prove the following theorems:

THEOREM 1. *Let* (H1–2–3–4) *and* (D1) *be satisfied. Then, as* $\lambda = \lambda_n \to 0$ *and* $n\lambda_n \to \infty$, *we have*

$$a_n \to_P a, \quad \tilde{a}_n \to_P a.$$

In addition, let (D2) *be satisfied. Set, for* $\lambda > 0$,

$$\beta_n = \beta_{n,\lambda} = \left\{\int_0^{1/\lambda} b(\lambda v)K(v)\,dv\right\} \Big/ \int_0^{1/\lambda} K(v)\,dv.$$

Then, we have, as $\lambda = \lambda_n \to 0$ *and* $n\lambda_n \to \infty$,

(i) $\sqrt{n\lambda}\,(a_n - \tilde{a}_n) = o_P(1)$, $\beta_n \to 0$, *and*

(ii) $\{\sqrt{n\lambda}/a\}\,\{\int_0^\infty K^2(v)\,dv\}^{-1/2}(a_n - a/(1 + a\beta_n)) \to_w N(0, 1)$.

THEOREM 2. *Let* (H1–2–3–4) *and* (D1–2) *be satisfied. Then, in order that there exists a nonrandom sequence* C_n *such that* $C_n(a_n - a)$ *converges weakly toward a limiting* $N(0, 1)$ *distribution for some sequence* $\lambda = \lambda_n \to 0$ *with* $n\lambda_n \to \infty$, *it is necessary and sufficient that*

$$\text{(C1)} \qquad \lim_{n\to\infty} \sqrt{n\lambda}\, \beta_n = \lim_{n\to\infty} \sqrt{n\lambda} \int_0^{1/\lambda} b(\lambda v) K(v)\, dv = 0.$$

If this condition is satisfied, then, as $\lambda = \lambda_n \to 0$ *and* $n\lambda_n \to \infty$, *we have*

$$\frac{\sqrt{n\lambda}}{a} \left\{ \int_0^\infty K^2(v)\, dv \right\}^{-1/2} (a_n - a) \to_w N(0, 1).$$

THEOREM 3. *Let* (H1–2–3–4) *and* (D1–2) *be satisfied. Assume further that*

$$\text{(C2)} \qquad \lim_{n\to\infty} \sqrt{n\lambda}\, \beta_n = \infty.$$

Then, we have, as $\lambda = \lambda_n \to 0$ *and* $n\lambda_n \to \infty$,

$$a_n - a = -a^2 \beta_n (1 + o_P(1)).$$

REMARK 1. Let a_n be the estimate corresponding to a kernel K and to a bandwidth λ. If we replace $K(t)$ by $\mu K(\mu t)$ and λ by $\lambda\mu$ for some $\mu > 0$, then a_n remains unchanged while $\int_0^\infty K^2(v)\, dv$ becomes $\mu \int_0^\infty K^2(v)\, dv$. It follows that, as far as we are concerned with efficiency criterion related to the variance of the limiting distribution of a_n, we can assume that (H7) holds.

REMARK 2. It can be seen (see, e.g., Mason, 1982b, Section 4) that (H1–2–3–4) imply

$$\text{(H4')} \qquad \int_0^\infty K^2(u)\, du < \infty.$$

On the other hand, (H1–2–3–4′) do not imply (H4).

The only real superiority of Hill's estimate follows from:

THEOREM 4. *Let* (H1–2–3–4–5–7) *be satisfied. Then, the minimum possible value* Λ *for which* $K(u) = 0$ *for* $u > \Lambda$ *is* $\Lambda = 1$, *and the unique possible kernel which satisfies these conditions is* $K(u) = 1_{\{0<u<1\}}$ *(corresponding to Hill's estimate).*

In other words, if we let, as in Hill's definition, $\lambda = k/n$, for a given asymptotic variance, Hill's estimate uses the minimum possible number k of upper order statistics of the sample.

A logical criterion for an optimal choice of $\lambda = \lambda_n$ and K is that which minimizes the expected mean square error of $1/a_n$. For this, we note that Theorem 1(ii) can be put equivalently as

$$a\sqrt{n\lambda} \left\{ \int_0^\infty K^2(v)\, dv \right\}^{-1/2} \left(\frac{1}{a_n} - \frac{1}{a} - \beta_n \right) \to_w N(0, 1).$$

Thus, it follows that a reasonable method of optimization consists in minimizing

$$M(n, \lambda, K) = \frac{1}{n\lambda}\int_0^\infty K^2(v)\,dv + \left\{\int_0^{1/\lambda} b(\lambda v)K(v)\,dv\right\}^2.$$

In this sum, the first term corresponds to the limiting variance of $1/a_n$, while the second corresponds to the limiting square of the bias.

We present in the following some results corresponding to the optimal choices of $\lambda = \lambda_n$ and K for the model introduced by Hall (1982), assuming that

$$\text{(D4)} \qquad 1 - F(x) = C_1 x^{-a}\{1 + C_2 x^{-b}(1 + o(1))\} \quad \text{as} \quad x \to \infty,$$

where $C_1 \neq 0$, $C_2 \neq 0$, $a > 0$ and $b > 0$ are constants.

This models covers a wide range of distributions. For instance, all nonnormal stable and all Frechet Φ_a distributions belong to this family. Routine computations as in Hall (1982) show that (D4) is equivalent to

$$\text{(D4}'\text{)} \qquad Q(1 - s) = s^{-1/a}D_1\{1 + D_2 s^{\alpha}(1 + o(1))\} \quad \text{as} \quad s \to 0,$$

where $D_1 = C_1^{1/a}$, $D_2 = C_2/(aC_1^{b/a})$ and $\alpha = b/a$.

We shall prove the following result:

THEOREM 5. *Let* (H1–2–3–4–5) *and* (D-1–2–4) *be satisfied. Then, there exists a* $\Lambda_1 > 0$, *depending upon* F *only, such that the following choice of* λ:

$$\lambda = \hat{\lambda}_n = \left\{\frac{1}{n}(2\alpha^3 D_2^2)^{-1}\left(\int_0^\infty K^2(v)\,dv\right)\left(\int_0^\infty v^{\alpha}K(v)\,dv\right)^{-2}\right\}^{1/(2\alpha+1)}$$

realizes

$$M(n, \hat{\lambda}_n, K) \sim \inf_{0<\nu<\Lambda_1} M(n, \nu, K) \quad \textit{as} \quad n \to \infty.$$

We have then, as $n \to \infty$,

$$M(n, \hat{\lambda}_n, K) \sim n^{-2\alpha/(2\alpha+1)}(2\alpha^3 D_2^2)^{1/(2\alpha+1)} \cdot\left(1 + \frac{1}{2\alpha}\right)\left(\int_0^\infty K^2(v)\,dv\right)^{2\alpha/(2\alpha+1)}\left(\int_0^\infty v^{\alpha}K(v)\,dv\right)^{2/(2\alpha+1)}.$$

Furthermore, the optimal choice of K *given by*

$$K(v) = \hat{K}_\alpha(v) = \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{2\alpha+1}{2\alpha+2}\right)^{\alpha+1}\left\{\left(\frac{2\alpha+2}{2\alpha+1}\right)^{\alpha} - v^{\alpha}\right\}$$

$$\textit{if} \quad 0 < v < \frac{2\alpha+2}{2\alpha+1}, \quad \hat{K}_\alpha(v) = 0 \quad \textit{otherwise},$$

is such that, as $n \to \infty$,

$$M(n, \hat{\lambda}_n, \hat{K}_\alpha) \sim \inf\left\{M(n, \nu, K);\ 0 < \nu < \Lambda_1,\ \int_0^\infty K(v)\,dv = \int_0^\infty K^2(v)\,dv = 1\right\}.$$

REMARK 3. The choice $\lambda = \hat{\lambda}_n$ in Theorem 5 is such that

$$\lim_{n\to\infty}\sqrt{n\lambda}\beta_n = -(2\alpha)^{-1/2}\left(\int_0^\infty K^2(v)\,dv\right)^{1/2}.$$

Hence, this optimal choice does not yield for $\sqrt{n\lambda}(a_n - a)$ a limiting centered normal distribution.

REMARK 4. For Hall's model, we get asymptotic centered normal limiting distributions for $\sqrt{n\lambda}(a_n - a)$ if and only if $\lambda_n = o(n^{-a/(2b+a)})$. The "if" part has been proved by Hall (1982) in the special case of the Hill estimator (a different proof is given later by S. Csörgő and Mason, 1984), and we see that his condition is sharp.

REMARK 5. The results of this paper can be extended without difficulty to the case where $K(\cdot)$ is a function of bounded variation on any interval (e, f), $0 < e < f < \infty$. It suffices to consider K as $K_d - K_i$ where both K_d and K_i are nonincreasing functions. In so doing, we may gain asymptotic efficiency. The price to be paid, however, is that it may be possible to obtain negative estimates of the index a when the sample size is small.

REMARK 6. Theorem 5 is a justification in itself for the introduction of kernel estimates of a. For Hall's family of tail behaviors of $1 - F$, it shows that Hill's estimate is far from being optimal. The optimal kernels correspond to intuition in the sense that they give more weight to the upper order statistics.

EXAMPLES. In general, when $\alpha = b/a = 1$ in (D4′), then the optimal kernel is

$$\hat{K}_1(v) = \tfrac{9}{8}\,(\tfrac{4}{3} - v), \quad 0 < v < \tfrac{4}{3}.$$

Since

$$\int_0^\infty v\hat{K}_1(v)\,dv = \frac{4}{9},$$

the corresponding optimal bandwidth would be

$$\hat{\lambda}_n = (1/n^{1/3})(81a^2C_1^2/32C_2^2)^{1/3},$$

depending, of course, on unknown parameters.

(a) In the case of the Frechet extreme value distribution $\Phi_a(x) = e^{-x^{-a}}$, $x > 0$, we have $\alpha = 1$, $C_1 = 1$ and $C_2 = \tfrac{1}{2}$. In practice, one would use the optimal kernel $\hat{K}_1(\cdot)$ and the bandwidth $\hat{\hat{\lambda}}_n = \hat{a}_n^{2/3}(81/8)^{1/3}n^{-1/3}$, where $\hat{a}_n$ is some initial consistent estimate of a.

(b) Let $F(x) = F_{a,\beta,\gamma}(x)$ be the distribution function of the stable distribution with location zero, scale parameter $\gamma > 0$, skewness parameter $-1 \le \beta \le 1$, and characteristic exponent $0 < a < 2$. This means that the characteristic function of F is

$$C_{a,\beta,\gamma}(t) = \begin{cases} \exp\{-\gamma\,|t|^a(1 - i\,\beta\,\mathrm{sgn}(t)\tan(\pi\alpha/2))\}, & a \neq 1, \\ \exp\{-\gamma\,|t|\,(1 + i(2/\pi)\beta\,\mathrm{sgn}(t)\log|t|)\}, & a = 1, \end{cases}$$

(refer to Hall (1981) to avoid any sign problem). Excluding the case $\beta = -1$ and the Cauchy distribution ($a = 1$, $\beta = 0$), making the obvious transformation for scale and integrating the well-known expansions for the density given in Skorohod (1954), for $1 - F$ one obtains (D4) with $b = a$ (so that again $\alpha = 1$) and

$$C_1 = C_1(a, \beta, \gamma) = (\pi a)^{-1}\gamma^{1-(1/a)}A_1(a, \beta)$$

and

$$C_2 = C_2(a, \beta, \gamma) = \gamma A_2(a, \beta)/2A_1(a, \beta),$$

where $A_1(a, \beta)$ and $A_2(a, \beta)$ are some nonzero constants depending on a and β. (In the Cauchy case, $A_2(1, 0) = 0$, but (D4) holds with $C_1 = 1$, $C_2 = (3\pi)^{-1}$, $a = 1$ and $b = 2$. If $\beta = -1$, then (D4) holds for $F(-x)$ instead of $1 - F(x)$ with the same C_1, C_2 and $b = a$.)

Suppose first that $\beta = 1$, i.e, that the distribution is completely asymmetric and skewed to the right. (If $a < 1$, then this distribution is concentrated on $[0, \infty)$.) In this case,

$$A_1(a, 1) = \begin{cases} \Gamma(a + 1)(1 + \tan^2(\pi a/2))^{1/2}\sin \pi a, & a \neq 1, \\ 2 & a = 1, \end{cases}$$

and

$$A_2(a, 1) = \begin{cases} -\dfrac{1}{2}\,\Gamma(2a + 1)(1 + \tan^2(\pi a/2)\sin 2\pi a, & a \neq 1, \\[2ex] -\dfrac{8}{\pi}\displaystyle\int_0^\infty e^{-t}t^2 \log t\,dt, & a = 1. \end{cases}$$

Our estimator of a is a_n, based on the observations which are not less than one and on the optimal kernel $\hat{K}_1(\cdot)$. The optimal bandwidth would be

$$\hat{\lambda}_n = \frac{1}{n^{1/3}}\left(\frac{81A_1^4(a, 1)}{8\pi^2\gamma^{2/a}A_2^2(a, 1)}\right)^{1/3} = \frac{1}{n^{1/3}}\,V(a, \gamma),$$

and again in practice one would perhaps use $\hat{\hat{\lambda}}_n = V(\hat{a}_n, \hat{\gamma}_n)n^{-1/3}$, where $\hat{a}_n$ and $\hat{\gamma}_n$ are some consistent initial estimates of a and γ obtained by any of the many procedures described by DuMouchel (1983) and S. Csörgő (1984).

Adaption of this method in the nonasymmetric case ($|\beta| < 1$) will be considered elsewhere.

(c) Consider the generalized Pareto family

$$1 - F(x) = (1 + x/(a\sigma))^{-a}, \quad x > 0,$$

put forward by DuMouchel (1983), where $\sigma > 0$ and we assume that $a > 0$.

Elementary calculus gives that $1 - F(x)$ satisfies (D4) and (D4′) with $b = 1$, $C_1 = (a\sigma)^a$, $C_2 = -a^2\sigma$, $D_1 = a\sigma$, $D_2 = -1$. Since $\alpha = a^{-1}$, here the optimal kernel also depends on the index a.

(d) Throughout the examples above, we have seen that the optimal kernel and bandwidth may eventually depend on some parameters of the distribution, including the tail index a itself. We propose for practical purposes to construct pseudo-optimal estimates where these parameters are replaced by preliminary estimations. This idea has proved itself successful in the field of density estimation, under the name of the "semi-parametric method" (see, e.g., Devroye and Penrod, 1984, Section 2.1). We do not offer here consistency results for these modified estimates, which will be discussed elsewhere.

2. Proof of the theorems. A rough outline of the proof is as follows. Let us make a heuristic calculus by assuming that Q is differentiable and by putting

$$Q_n(1 - s) = Q(1 - s) + n^{-1/2}B_n(1 - s)Q'(1 - s) + (\mathrm{SET})_1,$$

where, for each $n \geq 1$, B_n is a Brownian bridge and (SET) stands for "small error term."

Then, by taking logarithms, we get

$$\log Q_n(1 - s) = \log Q(1 - s) + n^{-1/2}B_n(1 - s)(Q'(1 - s)/Q(1 - s)) + (\mathrm{SET})_2.$$

Next, we note that, as $s \to 0$,

$$Q'(1 - s)/Q(1 - s) = -(1/as)(1 + o(1)).$$

By integrating with respect to $d\{vK(v)\}$, this leads to the expansion

$$A_n = \int_0^{1/\lambda} \{\log Q_n(1 - v\lambda)\}\, d\{vK(v)\} = A_{n1} + A_{n2} + A_{n3},$$

where

$$A_{n1} = \int_0^{1/\lambda} \{\log Q(1 - v\lambda)\}\, d\{vK(v)\} \quad \text{is a bias term},$$

$$A_{n2} = -\int_0^{1/\lambda} \frac{B_n(1 - v\lambda)}{av\lambda\sqrt{n}}\, d\{vK(v)\} \quad \text{is a random term},$$

and

$$A_{n3} = (\mathrm{SET})_3 = A_n - A_{n1} - A_{n2} \quad \text{is an error term}.$$

In the proof, we shall have to evaluate successively each of these terms. The results we seek will be obtained by showing that, as $\lambda \to 0$ and $n\lambda \to \infty$, A_{n3} is asymptotically negligible with respect to A_{n1} and A_{n2}.

The rest of this paper is organized as follows. In Section 2.1 we establish preliminary lemmas and study A_{n1}. In Section 2.2 we treat the random term A_{n2}. In Section 2.3 we deal with the error term A_{n3}. Finally in Section 2.4 we discuss the optimal choices of K and λ.

2.1 *Preliminaries—the bias term.*

LEMMA 1. *Let K satisfy* (H1–2–3–4). *Then we have*

(i) $$K(u) = o(u^{-1/2}) \quad \text{as} \quad u \to 0;$$

(ii) $$K(u) = o(u^{-1}) \quad \text{as} \quad u \to \infty.$$

PROOF. (i) We have $uK^2(u) \le \int_0^u K^2(v)\,dv \to 0$ as $u \to 0$.

(ii) Likewise $(u/2)K(u) \le \int_{u/2}^u K(v)\,dv \to 0$ as $u \to \infty$.

LEMMA 2. *Let K satisfy* (H1–2–3–4). *Then*

(i) $$\int_0^\infty d\{vK(v)\} = 0, \int_0^\infty v\,dK(v) = -\int_0^\infty K(v)\,dv = -1,$$

and for any $\lambda > 0$

(ii) $$\int_0^{1/\lambda} (\log v\lambda)\,d\{vK(v)\} = -\int_0^{1/\lambda} K(v)\,dv.$$

PROOF. (i) For $0 < a < b < \infty$, we have

$$\int_a^b d\{vK(v)\} = bK(b) - aK(a) \to 0 \quad \text{as} \quad a \to 0 \quad \text{and} \quad b \to \infty, \text{ by Lemma 1.}$$

Likewise

$$\int_a^b v\,dK(v) = bK(b) - aK(a) - \int_a^b K(v)\,dv \to -1 \quad \text{as} \quad a \to 0 \quad \text{and} \quad b \to \infty.$$

(ii) By Lemma 1 and integrating by parts, we have

$$\int_0^{1/\lambda} (\log v\lambda)\,d\{vK(v)\} = [(\log v\lambda)vK(v)]_0^{1/\lambda} - \int_0^{1/\lambda} K(v)\,dv = -\int_0^{1/\lambda} K(v)\,dv.$$

The next lemmas correspond to the study of the bias term A_{n1}.

LEMMA 3. *Let K satisfy* (H1–2–3–4) *and Q satisfy* (D3). *Then*:

$$A_{n1} = \int_0^{1/\lambda} \{\log Q(1 - v\lambda)\}\,d\{vK(v)\} = \frac{1}{a}\int_0^{1/\lambda} K(v)\,dv + \int_0^{1/\lambda} b(\lambda v)K(v)\,dv.$$

Furthermore, if $\lambda \to 0$, *then*

$$\int_0^{1/\lambda} K(v)\,dv \to 1 \quad \text{and} \quad \int_0^{1/\lambda} b(\lambda v)K(v)\,dv \to 0.$$

PROOF. By (D3), we have

$$\log Q(1 - v\lambda) = -\frac{1}{a}\log v\lambda + \int_{v\lambda}^1 \frac{b(u)}{u}\,du,$$

and hence, by Lemma 2(ii),

$$A_{n1} = \frac{1}{a}\int_0^{1/\lambda} K(v)\,dv + \int_0^{1/\lambda}\left\{\int_{v\lambda}^1 \frac{b(u)}{u}\,du\right\} d\{vK(v)\}.$$

For $0 < a < 1/\lambda$, we have

$$\int_a^{1/\lambda}\left\{\int_{v\lambda}^1 \frac{b(u)}{u}\,du\right\} d\{vK(v)\} = -aK(a)\int_{a\lambda}^1 \frac{b(u)}{u}\,du + \int_a^{1/\lambda} b(\lambda v)K(v)\,dv.$$

Since $b(u) \to 0$ as $u \to 0$, we have, as $a \to 0$, for a fixed λ,

$$aK(a)\int_{a\lambda}^1 \frac{b(u)}{u}\,du = o(a^{1/2}K(a)) = o(1).$$

This proves the first assertion of Lemma 3.

Next, we have evidently

$$\lim_{\lambda\to 0}\int_0^{1/\lambda} K(v)\,dv = \int_0^\infty K(v)\,dv = 1.$$

For the last term, and for any given $\varepsilon > 0$, we choose $\delta > 0$ such that $|b(u)| \le \varepsilon$ for $0 < u \le \delta$. This gives

$$\left|\int_0^{1/\lambda} b(\lambda v)K(v)\,dv\right| \le \varepsilon\int_0^\infty K(v)\,dv + \left|\int_{\delta/\lambda}^{1/\lambda} b(\lambda v)K(v)\,dv\right|.$$

Next, δ being fixed, we use the bound

$$\left|\int_{\delta/\lambda}^{1/\lambda} b(\lambda v)K(v)\,dv\right| \le \frac{\delta}{\lambda}K\left(\frac{\delta}{\lambda}\right)\int_\delta^1 |b(v)|\,\frac{dv}{\delta} \to 0 \quad \text{as} \quad \lambda \to 0,$$

by Lemma 1.

This completes the proof of Lemma 3.

Lemma 4. *Let (H1–2–3–4) be satisfied. Then, as $\lambda \to 0$ and $n\lambda \to \infty$, we have*

$$\sqrt{n\lambda}\left\{\sum_{j=1}^n \frac{1}{n\lambda}K\left(\frac{j}{n\lambda}\right) - \int_0^{1/\lambda} K(v)\,dv\right\} \to 0.$$

Proof. Since K is nonincreasing, we have

$$0 \le \int_0^{1/\lambda} K(v)\,dv - \sum_{j=1}^n \frac{1}{n\lambda}K\left(\frac{j}{n\lambda}\right) \le \int_0^{1/n\lambda} K(v)\,dv + \int_{1/\lambda}^{(n+1)/n\lambda} K(v)\,dv$$

$$\le \frac{1}{n\lambda}K\left(\frac{1}{\lambda}\right) + \int_0^{1/n\lambda} K(v)\,dv = \frac{o(1)}{n} + o((n\lambda)^{-1/2}),$$

hence the result.

REMARK 7. Assuming that, as $\lambda \to 0$ and $n\lambda \to \infty$, $a_n \to_P$ a, Lemma 4 proves that $\sqrt{n\lambda}\,(a_n - \tilde{a}_n) \to_P 0$.

We shall now discuss the Karamata type representation of $Q(1-s)$ given in (Q). We first note that this representation is not unique since we may add to $b(\cdot)$ any function $b_1(\cdot)$ such that $b_1(s) \to 0$ as $s \to 0$ and $\int_0^1 b_1(u)/u\,du = c_1$ to obtain another representation of the same kind.

Next, using the fact that log $Q(1-s)$ is nonincreasing and assuming that $Q(0) = 1$, we can see that log $c(\cdot)$ is of bounded variation. This implies the existence of the following representation:

$$Q(1-s) = s^{-1/a}\exp\left(\int_s^1 dM(u)\right),$$

where $dM(\cdot)$ is a Radon measure. If we denote the Lebesgue decomposition of dM by $dM(u) = m(u)\,du + dM_d(u) + dM_s(u)$, where dM_d is the discrete component of dM and dM_s is the singular component of dM, then we must have

$$\int_0^1 dM_d(u) \quad \text{and} \quad \int_0^1 dM_s(u) \quad \text{finite,}$$

and

$$m(u) = \frac{b(u)}{u} + m_1(u) \quad \text{with} \quad \int_0^1 m_1(u)\,du \quad \text{finite.}$$

It is not difficult to see that these conditions are sufficient under weak additional restrictions to give the following lemma:

LEMMA 5. *$Q(1-s)$ is regularly varying of order $-1/a$ at 0, and $Q(0) = 1$, if and only if there exists a discrete measure M_d, a singular measure M_s and a measurable function m on* (0, 1), *such that*

(i) $M_d \geq 0$, $M_s \geq 0$, $m(u) \geq -1/au$, $0 < u < 1$.
(ii) $\int_0^1 dM_d(u)$, $\int_0^1 dM_s(u)$ *and* $\int_0^1 um(u)\,du$ *are finite.*
(iii) *For any* $0 < s < 1$, $\int_s^1 m(u)\,du$ *is finite.*
(iv) *There exists a function $b(u)$ such that $b(u) \to 0$ as $u \to 0$ and that*

$$\int_0^1 \left(m(u) - \frac{b(u)}{u}\right) du \text{ is finite.}$$

(v) *We have*

$$Q(1-s) = s^{-1/a}\exp\left(\int_s^1 \{dM_d(u) + dM_s(u) + m(u)\,du\}\right).$$

PROOF. Straightforward.

It follows from Lemma 5 that if $m(\cdot)$ satisfies (i–ii–iii) and, in addition, is such that

$$\lim_{u\to 0} um(u) = 0,$$

then we may take in (iv) $b(u) = um(u)$ to obtain a $Q(1-s)$ which is a regularly varying function of order $-1/a$. In particular, we have:

LEMMA 6. *Let $F(\cdot)$ be such that, for some x_0, there exists a density $f(x) = dF(x)/dx > 0$ for $x \geq x_0$. Assume, in addition, that*

$$\lim_{x\to\infty} xf(x)/(1-F(x)) = a \in (0, \infty).$$

Then there exists an s_0, $0 < s_0 \leq 1$ such that, for any $0 < s \leq s_0$, we have

$$Q(1-s) = Q(1-s_0)(s/s_0)^{-1/a}\exp\left(\int_s^{s_0} \frac{b(u)}{u}\,du\right),$$

where $b(\cdot)$ is defined for $x \geq x_0$ by

$$b(1-F(x)) = ((1-F(x))/xf(x)) - (1/a).$$

PROOF. By direct integration, we get, for $x \geq x_0$,

$$1 - F(x) = (1 - F(x_0))(x/x_0)^{-a}\exp\left(\int_x^{x_0} \frac{\varepsilon(t)}{t}\,dt\right),$$

where

$$\varepsilon(t) = (tf(t)/(1-F(t))) - a.$$

Hence $1 - F$ is regularly varying of order $-a$. The result now follows by inverting $1 - F$.

REMARK 8. The function $b(u)$, $0 < u \leq 1 - F(x_0)$ defined in Lemma 6 is scale free in the sense that it does not change if one replaces $F(x)$ by $F(\rho x)$, for any $\rho > 0$.

LEMMA 7. *Let K satisfy* (H1–2–3–4) *and let Q satisfy* (D1). *Then, we have*

$$A_{n1} = \int_0^{1/\lambda} \{\log Q(1 - v\lambda)\}\, d\{vK(v)\} \to 1/a \quad as \quad \lambda \to 0.$$

PROOF. By Lemma 5, we may assume that

$$Q(1-s) = s^{-1/a}c(s)\exp\left(\int_s^1 \frac{b(u)}{u}\,du\right),$$

where $c(s) \to c \in (0, \infty)$ and $b(s) \to 0$ as $s \to 0$, and where

$$0 < \inf_{s\geq 0} c(s) \leq \sup_{s\geq 0} c(s) < \infty.$$

By Lemma 2 and Lemma 3, the proof can be accomplished by showing that

$$\lim_{\lambda\to 0} \int_0^{1/\lambda} \{\log c(v\lambda)\}K(v)\,dv = \int_0^\infty \{\log c\}K(v)\,dv,$$

and that

$$\lim_{\lambda\to 0}\int_0^{1/\lambda}\{\log c(v\lambda)\}v\,dK(v) = \int_0^{\infty}\{\log c\}v\,dK(v),$$

which follow from Lebesgue's theorem and from Lemma 2.

2.2 *The random term.* In this paragraph we determine the limiting distribution of

$$A_{n2} = -\int_0^{1/\lambda}\frac{B_n(1-v\lambda)}{av\lambda\sqrt{n}}\,d\{vK(v)\},\quad \text{as}\quad \lambda\to 0,$$

where $B_n(\cdot)$ is a Brownian bridge. We first note that A_{n2} is equal in distribution to $a^{-1}n^{-1/2}(C_\lambda - D_\lambda)$, where

$$C_\lambda = \int_0^{1/\lambda}\frac{W(v\lambda)}{v\lambda}\,d\{vK(v)\},\quad D_\lambda = \int_0^{1/\lambda}W(1)\,d\{vK(v)\},$$

and $\{W(u), u\ge 0\}$ is a Wiener process. We get easily from Lemma 1

$$D_\lambda = W(1)\{(1/\lambda)K(1/\lambda)\} = o_P(1)\quad\text{as}\quad\lambda\to 0.$$

For the first term, we need the following lemma:

LEMMA 8. *Let K satisfy* (H1–2–3–4). *Then, for any $\lambda > 0$,*

$$C_\lambda = \int_0^{1/\lambda}\frac{W(v\lambda)}{v\lambda}\,d\{vK(v)\}$$

follows a $N(0,\sigma_\lambda^2)$ distribution, where

$$\sigma_\lambda^2 = \left\{\frac{1}{\lambda}K\left(\frac{1}{\lambda}\right)\right\}^2 + \frac{1}{\lambda}\int_0^{1/\lambda}K^2(v)\,dv.$$

PROOF. Choose any $0<\delta<1$ and consider the random variable

$$\gamma_\lambda(\delta) = \int_{\delta/\lambda}^{1/\lambda}\frac{W(v\lambda)}{v\lambda}\,d\{vK(v)\}.$$

Evidently $\gamma_\lambda(\delta)$ is well defined as a Gaussian random variable with expectation zero and whose variance may be computed by integration by parts from

$$E(\gamma_\lambda^2(\delta)) = \int_\delta^1 d\left\{\frac{t}{\lambda}K\left(\frac{t}{\lambda}\right)\right\}\int_\delta^1\frac{\min(s,t)}{st}\,d\left\{\frac{s}{\lambda}K\left(\frac{s}{\lambda}\right)\right\},$$

giving, after some routine calculus,

$$E(\gamma_\lambda^2(\delta)) = \left\{\frac{1}{\lambda}K\left(\frac{1}{\lambda}\right)\right\}^2 - 2\left\{\frac{1}{\lambda}K\left(\frac{1}{\lambda}\right)\right\}\left\{\frac{\delta}{\lambda}K\left(\frac{\delta}{\lambda}\right)\right\} + \left\{\frac{1}{\lambda}K\left(\frac{\delta}{\lambda}\right)\right\}\left\{\frac{\delta}{\lambda}K\left(\frac{\delta}{\lambda}\right)\right\}$$
$$+\int_\delta^1\frac{1}{\lambda}K^2\left(\frac{t}{\lambda}\right)\frac{dt}{\lambda} - 2\,\frac{\delta}{\lambda}K\left(\frac{\delta}{\lambda}\right)\int_\delta^1\frac{1}{t}K\left(\frac{t}{\lambda}\right)\frac{dt}{\lambda}.$$

If we let $\delta \to 0$, using Lemma 1, it can be seen easily that

$$\lim_{\delta\to 0} E(\gamma_\lambda^2(\delta)) = \left\{\frac{1}{\lambda} K\left(\frac{1}{\lambda}\right)\right\}^2 + \frac{1}{\lambda}\int_0^{1/\lambda} K^2(v)\,dv.$$

It follows that C_λ is well defined as the limit in expected mean square of $\gamma_\lambda(\delta)$ as $\delta \to 0$, hence the result.

A direct consequence of Lemma 8 is the result we seek:

LEMMA 9. *Let K satisfy* (H1–2–3–4). *Then, as $\lambda \to 0$ and $n\lambda \to \infty$,*

$$\sqrt{n\lambda} A_{n2} = -\int_0^{1/\lambda} \frac{B_n(1 - v\lambda)}{av\sqrt{\lambda}}\, d\{vK(v)\} \to_w N\left(0,\, a^{-2}\int_0^\infty K^2(v)\,dv\right).$$

PROOF. We let $\lambda \to 0$ in Lemma 8 and use Lemma 1.

2.3 *The error term.* The main tool we shall make use of in this section is a recent result of M. Csörgő, S. Csörgő, Horváth and Mason [Cs-Cs-H-M] (1984), where a probability space (Ω, A, P) is constructed carrying a sequence $U_1, U_2, \cdots$ of independent random variables uniformly distributed on (0, 1) and a sequence of Brownian bridges $\{B_n(s), 0 \le s \le 1\}$, which has, among others, the following property.

Let $U_{1,n} < \cdots < U_{n,n}$ denote the order statistics of $U_1, \cdots, U_n$ and define the uniform quantile function $U_n(s)$ as

$$U_n(s) = U_{k,n} \quad \text{if} \quad (k-1)/n < s \le k/n, \quad k = 1, \cdots, n,$$

and

$$U_n(0) = 0.$$

LEMMA 10. *On the probability space of* Cs-Cs-H-M (1984), *for any* $0 \le \nu < 1/2$, *we have, as* $n \to \infty$,

$$\sup_{1/(n+1)\le s\le 1} |\sqrt{n}(1 - s - U_n(1-s)) - B_n(1-s)|/s^{-\nu+1/2} = O_P(n^{-\nu}).$$

PROOF. This is an easy consequence of Corollary 2.1, page 24, of Cs-Cs-H-M (1984).

We shall assume in the sequel that the uniform (0, 1) random variables $U_1, U_2, \cdots$ and the Brownian bridges $B_n(\cdot)$, $n = 1, 2, \cdots$, are defined on the probability space of Lemma 10 and that, without loss of generality, we have

$$X_n = Q(U_n), \quad Q_n(1-s) = Q(U_n(1-s)), \quad 0 < s \le 1, \quad n = 1, 2, \cdots.$$

We shall assume in the sequel (unless otherwise specified) that (D3) holds. This is because we see from the above representation for Q_n and from the sum representation of our estimator a_n in the introduction that either asymptotically (the first case in (D2)(i)) or always (the second case in (D2)(i)) the logarithms of $c(s) = c$ in (Q) cancel. Therefore we may assume without loss of generality that (D3) holds. Our aim is to evaluate the limiting behavior as $\lambda \to 0$ and $n\lambda \to \infty$ of the integral

$$\int_0^{1/\lambda} \Delta_n(v\lambda)\, d\{vK(v)\},$$

where

$$\Delta_n(s) = \Delta_n^{(1)}(s) + \Delta_n^{(2)}(s) = \log Q_n(1-s) - \log Q(1-s),$$

$$\Delta_n^{(1)}(s) = -(1/a)\{\log(1 - U_n(1-s)) - \log s\},$$

and

$$\Delta_n^{(2)}(s) = \int_{1-U_n(1-s)}^{s} \frac{b(u)}{u}\, du, \quad 0 < s \le 1.$$

Let

$$T_n = \sqrt{n\lambda} \int_0^{1/\lambda} \left\{\Delta_n(v\lambda) - \frac{B_n(1 - v\lambda)}{av\lambda\sqrt{n}}\right\} d\{vK(v)\}.$$

We intend in the following to show that, as $\lambda \to 0$ and $n\lambda \to \infty$,

$$T_n = o_P(1).$$

For this, we shall split T_n into three parts:

$$T_n = R_n + R_n' + R_n'',$$

where

$$R_n = \sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{1/\lambda} \left\{\Delta_n(v\lambda) - \frac{B_n(1 - v\lambda)}{av\lambda\sqrt{n}}\right\} d\{vK(v)\},$$

$$R_n' = \sqrt{n\lambda} \int_0^{1/(\lambda(n+1))} \Delta_n(v\lambda)\, d\{vK(v)\},$$

and

$$R_n'' = -\sqrt{n\lambda} \int_0^{1/(\lambda(n+1))} \frac{B_n(1 - v\lambda)}{av\lambda\sqrt{n}}\, d\{vK(v)\}.$$

We first consider R_n' and R_n''.

LEMMA 11. *Let K satisfy* (H1–2–3–4), *and assume that* (D3) *holds. Then, as $\lambda \to 0$ and $n\lambda \to \infty$, we have $R_n' = o_P(1)$ and $R_n'' = o_P(1)$.*

PROOF. a) For $0 < v \le 1/(\lambda(n+1))$, $Q_n(1 - v\lambda) = Q(U_{n,n})$ and

$$\Delta_n(v\lambda) = -\frac{1}{a}\log(1 - U_{n,n}) + \frac{1}{a}\log v\lambda + \int_{1-U_{n,n}}^{v\lambda} \frac{b(u)}{u}\, du.$$

Next, we have, for any x,

$$\lim_{n\to\infty} P(-\log(1 - U_{n,n}) - \log(n + 1) < x) = e^{-e^{-x}}.$$

It follows that, as $n\lambda \to \infty$ *and* $n \to \infty$,

$$\sqrt{n\lambda} \int_0^{1/(\lambda(n+1))} \{-\log(1 - U_{n,n}) - \log(n + 1)\}\, d\{vK(v)\}$$

$$= \sqrt{n\lambda}\, O_P(1) \int_0^{1/(\lambda(n+1))} d\{vK(v)\} = o_P(1) \quad \text{by Lemma 1.}$$

In the remaining term, we have by (ii) of Lemma 2

$$\sqrt{n\lambda} \int_0^{1/(\lambda(n+1))} \frac{1}{a} \log(v\lambda(n + 1))\, d\{vK(v)\} = -\frac{\sqrt{n\lambda}}{a} \int_0^{1/(\lambda(n+1))} K(v)\, dv,$$

which by Lemma 1 equals $o(1)$ as $n\lambda \to \infty$.

Finally, we consider

$$\sqrt{n\lambda} \int_0^{1/(\lambda(n+1))} \left\{\int_{1-U_{n,n}}^{v\lambda} \frac{b(u)}{u}\, du\right\} d\{vK(v)\}$$

$$= \sqrt{n\lambda} \left\{\frac{1}{(n + 1)\lambda} K\left(\frac{1}{(n + 1)\lambda}\right) \int_{1-U_{n,n}}^{1/(n+1)} \frac{b(u)}{u}\, du - \int_0^{1/(\lambda(n+1))} b(\lambda v)K(v)\, dv\right\},$$

after integrating by parts. By Lemma 1, we have, as $\lambda \to 0$ and $n\lambda \to \infty$,

$$\sqrt{n\lambda} \int_0^{1/(\lambda(n+1))} b(\lambda v)K(v)\, dv \to 0.$$

Finally, we have to deal with the term

$$\frac{\sqrt{n\lambda}}{(n + 1)\lambda} K\left(\frac{1}{(n + 1)\lambda}\right) \int_{1-U_{n,n}}^{1/(n+1)} \frac{b(u)}{u}\, du.$$

If we use the fact that $b(u) \to 0$ as $u \to 0$ and that, for any $x > 0$,

$$\lim_{n\to\infty} P((n + 1)(1 - U_{n,n}) > x) = e^{-x},$$

we get evidently, as $n \to \infty$,

$$\int_{1-U_{n,n}}^{1/(n+1)} \frac{b(u)}{u}\, du = o_P(1) \log((n + 1)(1 - U_{n,n})) = o_P(1).$$

Finally, using again Lemma 1, we have, as $\lambda \to 0$ and $n\lambda \to \infty$,

$$\frac{\sqrt{n\lambda}}{(n + 1)\lambda} K\left(\frac{1}{(n + 1)\lambda}\right) \to 0.$$

This proves that $R_n' = o_P(1)$, as claimed.

b) We first note that, if $\{W(u),\, u \geq 0\}$ is a Wiener process, then R''_n is equal in distribution to $R''_{n0} - R''_{n1}$, where

$$R''_{n0} = \frac{\sqrt{\lambda}}{a} \int_0^{1/(\lambda(n+1))} \frac{W(v\lambda)}{v\lambda}\, d\{vK(v)\},$$

and

$$R''_{n1} = \frac{\sqrt{\lambda}}{a}\, W(1)\left\{\frac{1}{(n+1)\lambda}\, K\left(\frac{1}{(n+1)\lambda}\right)\right\}.$$

By Lemma 1, $R''_{n1} = o_P(1)$ as $\lambda \to 0$ and $n\lambda \to \infty$.

Next, using Lemma 8, we get

$$\begin{aligned} R''_{n0} &=_d N\left(0,\, a^{-2}\left\{\frac{1}{(n+1)\lambda}\, K^2\left(\frac{1}{(n+1)\lambda}\right) + \int_0^{1/(\lambda(n+1))} K^2(v)\, dv\right\}\right) \\ &= o_P(1) \quad \text{as} \quad \lambda \to 0 \quad \text{and} \quad n\lambda \to \infty. \end{aligned}$$

This completes the proof of Lemma 11.

Lemma 11 shows that, as $\lambda \to 0$ and $n\lambda \to \infty$, $T_n - R_n = o_P(1)$. Let us now split R_n into $R_n^{(1)} + R_n^{(2)}$, where

$$R_n^{(1)} = \sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{1/\lambda} \left\{\Delta_n^{(1)}(v\lambda) - \frac{B_n(1 - v\lambda)}{av\lambda\sqrt{n}}\right\} d\{vK(v)\},$$

and

$$R_n^{(2)} = \sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{1/\lambda} \Delta_n^{(2)}(v\lambda)\, d\{vK(v)\}.$$

We shall first concentrate on $R_n^{(1)}$ and use the Taylor expansion of $\Delta_n^{(1)}$:

$$\begin{aligned} \Delta_n^{(1)}(s) &= -(1/a)\{\log(1 - U_n(1 - s)) - \log s\} \\ &= -(1/as)\{1 - s - U_n(1 - s)\} + (\{1 - s - U_n(1 - s)\}^2/2a\theta_n^2(s)), \end{aligned}$$

where

$$\min\{s,\, 1 - U_n(1 - s)\} < \theta_n(s) < \max\{s,\, 1 - U_n(1 - s)\}.$$

LEMMA 12. *Let K satisfy* (H1–2–3–4). *Then, as $\lambda \to 0$ and $n\lambda \to \infty$, we have* $R_n^{(1)} = o_P(1)$.

PROOF. The proof will be made in seven steps.

STEP 1. Let $0 < \nu < 1/2$ be given. We consider, for $d\mu = K(v)\,dv$ or $d\mu = -v dK(v)$,

$$\sqrt{n\lambda}\int_{1/(\lambda(n+1))}^{1/\lambda}\left|\frac{1}{a}(1 - v\lambda - U_n(1 - v\lambda))\frac{1}{v\lambda} - \frac{B_n(1 - v\lambda)}{a\sqrt{n}\,v\lambda}\right| d\mu(v)$$
$$= \frac{\sqrt{\lambda}}{a}\int_{1/(\lambda(n+1))}^{1/\lambda} |\sqrt{n}(1 - v\lambda - U_n(1 - v\lambda)) - B_n(1 - v\lambda)|\,\frac{d\mu(v)}{v\lambda}$$
$$\le \sup_{1/(n+1)\le s\le 1}\frac{|\sqrt{n}(1 - s - U_n(1 - s)) - B_n(1 - s)|}{s^{-\nu+1/2}}\,\frac{\sqrt{\lambda}}{a}$$
$$\cdot\int_{1/(\lambda(n+1))}^{1/\lambda}(v\lambda)^{-\nu-1/2}\,d\mu(v).$$

By Lemma 10, this expression is, as $n \to \infty$,

$$O_P(n^{-\nu})\frac{\lambda^{1/2}}{a}\int_{1/(\lambda(n+1))}^{1/\lambda}(v\lambda)^{-\nu-1/2}\,d\mu(v) = O_P((n\lambda)^{-\nu})\int_{1/(\lambda(n+1))}^{1/\lambda} v^{-\nu-1/2}\,d\mu(v) \equiv I_n.$$

STEP 2. Take $d\mu(v) = K(v)\,dv$. For any $\delta > 0$, we have

$$\int_{1/(\lambda(n+1))}^{\delta} v^{-\nu-1/2}K(v)\,dv \le \{\sup_{0<s\le\delta}\sqrt{s}\,K(s)\}\int_{1/(\lambda(n+1))}^{\infty} v^{-1-\nu}\,dv.$$

Hence, by taking the corresponding part of I_n, we get

$$O_P((n\lambda)^{-\nu})\int_{1/(\lambda(n+1))}^{\delta} v^{-\nu-1/2}K(v)\,dv \le \{\sup_{0<s\le\delta}\sqrt{s}\,K(s)\}O_P(1),$$

where the $O_P(1)$ is independent of δ.

STEP 3. Take $d\mu(v) = -v\,dK(v)$. For any $\delta > 0$, we get likewise

$$\int_{1/(\lambda(n+1))}^{\delta} v^{-\nu-1/2}\{-v\,dK(v)\}$$
$$= \int_{1/(\lambda(n+1))}^{\delta} v^{-\nu+1/2}\{-dK(v)\}$$
$$\le (\lambda(n+1))^{\nu-1/2}K\left(\frac{1}{(n+1)\lambda}\right)$$
$$+\left(\frac{1}{2} - \nu\right)\{\sup_{0<s\le\delta}\sqrt{s}\,K(s)\}\int_{1/(\lambda(n+1))}^{\infty} v^{-1-\nu}\,dv$$
$$= \{\sup_{0<s\le\delta}\sqrt{s}\,K(s)\}O((n\lambda)^{\nu}).$$

By taking the corresponding part of I_n, we get in the same manner

$$O_P((n\lambda)^{-\nu})\int_{1/(\lambda(n+1))}^{\delta} v^{-\nu-1/2}\{-v\,dK(v)\} \le \{\sup_{0<s\le\delta}\sqrt{s}\,K(s)\}O_P(1),$$

where the $O_P(1)$ is independent of δ.

STEP 4. It is easy to see now that Lemma 1 in combination with Steps 2 and 3 allows us to choose for any $\varepsilon > 0$ a $\delta > 0$ such that

$$P\left(\left| O_p((n\lambda)^{-\nu}) \int_{1/(\lambda(n+1))}^{\delta} v^{-\nu-1/2}\, d\mu(v) \right| > \varepsilon \right) < \varepsilon$$

for all n sufficiently large, where $d\mu = K(v)$ or $d\mu = -v dK(v)$.

Let us choose $\delta > 0$ and consider the remaining part of I_n, namely

$$O_P((n\lambda)^{-\nu}) \int_{\delta}^{1/\lambda} v^{-\nu-1/2} d\mu(v) \le O_P((n\lambda)^{-\nu}) \int_{\delta}^{\infty} v^{-\nu-1/2}\, d\mu(v).$$

Since for either $d\mu = K(v)\, dv$ or $d\mu = -v\, dK(v)$, we have

$$\int_{\delta}^{\infty} v^{-\nu-1/2}\, d\mu(v) < \infty,$$

it follows that the above expression is $o_P(1)$ as $n\lambda \to \infty$. Thus, we see that, as $\lambda \to 0$ and $n\lambda \to \infty$, we have for $d\mu = K(v)\, dv$ or $d\mu = -v\, dK(v)$

$$\sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{1/\lambda} \left| \frac{1}{a} (1 - v\lambda - U_n(1 - v\lambda)) \frac{1}{v\lambda} - \frac{B_n(1 - v\lambda)}{a\sqrt{n} v\lambda} \right| d\mu(v) = o_P(1).$$

STEP 5. It remains to study the term

$$J_n = \frac{\sqrt{n\lambda}}{2a} \int_{1/(\lambda(n+1))}^{1/\lambda} \frac{|1 - v\lambda - U_n(1 - v\lambda)|^2}{\theta_n^2(v\lambda)}\, d\mu(v),$$

where $d\mu = K(v)\, dv$ or $d\mu = -v\, dK(v)$.

Here, we shall need the following lemma:

LEMMA 13. *For any $\rho > 1$, let $A_n(\rho)$ be the event $A_n(\rho) = \{s/\rho \le 1 - U_n(1 - s) \le \rho s,\ 1/(n + 1) \le s \le 1\}$. Then*

$$\lim_{\rho \uparrow \infty} \{\liminf_{n\to\infty} P(A_n(\rho))\} = 1.$$

PROOF. The proof follows easily from the inequalities in Remark 1 of Wellner (1978).

PROOF OF LEMMA 12 (continued). Let us chose by Lemma 13 an n_0 and a $\rho > 1$ such that, for $n \ge n_0$, we have $P(A_n(\rho)) > 1 - \varepsilon$, $\varepsilon > 0$ being fixed in advance.

This being the case, we have

$$s/\rho \le \theta_n(s) \le \rho s,$$

and hence

$$J_n \le \frac{(n\lambda)^{1/2}}{2a} \int_{1/(\lambda(n+1))}^{1/\lambda} \frac{\rho^2 |1 - v\lambda - U_n(1 - v\lambda)|^2}{(v\lambda)^2}\, d\mu(v) \equiv J_n(\rho).$$

We shall now make use of the following lemma:

LEMMA 14. *For any* $0 \le \nu < 1/2$, *we have, as* $n \to \infty$,

$$\sup_{1/(n+1)\le s\le 1}(n^{\nu}|1 - s - U_n(1 - s)|/s^{1-\nu}) = O_P(1).$$

PROOF. It follows from Theorem 2.1 of Mason (1983).

PROOF OF LEMMA 12 (continued). We choose ν in Lemma 14 such that $1/4 < \nu < 1/2$. This gives, as $n \to \infty$,

$$J_n(\rho) = O_P((n\lambda)^{(1/2)-2\nu}) \int_{1/(\lambda(n+1))}^{1/\lambda} v^{-2\nu}\, d\mu(v).$$

STEP 6. In the expression above, put $1/2 - 2\nu = -\gamma$, noting that $0 < \gamma < 1/2$. We get

$$J_n(\rho) = O_P((n\lambda)^{-\gamma}) \int_{1/(\lambda(n+1))}^{1/\lambda} v^{-\gamma-1/2}\, d\mu(v),$$

in which we recognize the same expression which has been evaluated in Steps 2, 3 and 4. The same proofs show that, as $\lambda \to 0$ and $n \to \infty$, we have $J_n(\rho) = o_P(1)$.

STEP 7. We may now choose ρ arbitrarily large to let $P(A_n(\rho))$ increase to 1. This suffices to show that $J_n = o_P(1)$ as $\lambda \to 0$ and $n\lambda \to \infty$.

If we put together the preceding results, we have proved that $R_n^{(1)} = o_P(1)$. The proof of Lemma 12 is now complete.

Up to now, we have proved that $T_n - R_n^{(2)} = o_P(1)$.

In the sequel, we shall assume that $b(u)$ is defined for all u, with $b(u) = 0$ for $u < 0$ or $u > 1$. If we use the assumption that $b(u) \to 0$ as $u \to 0$, then we see that there exists a $\delta > 0$ such that $b(\cdot)$ is bounded on the interval $(0, 2\delta)$. Put

$$b_n^{\star}(s) = \sup\{|b(u)|;\ \min(s, 1 - U_n(1 - s)) \le u \le \max(s, 1 - U_n(1 - s))\}.$$

By Glivenko-Cantelli, with probability one, there exists an n_0 such that, for $n \ge n_0$, $b_n^{\star}(s)$ is bounded for $0 \le s \le \delta$. This gives

$$|\Delta_n^{(2)}(s)| \le b_n^{\star}(s)\,|\log(1 - U_n(1 - s)) - \log s|.$$

Consider now

$$L_n = \sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{\delta/\lambda} |\Delta_n^{(2)}(v\lambda)|\, d\mu(v),$$

where $d\mu = K(v)\,dv$ or $d\mu = -v\,dK(v)$. We have, by the triangular inequality,

$$L_n \leq \sqrt{n\lambda}\,\{\sup_{1/(n+1)\leq s\leq \delta} b_n^\star(s)\} \cdot \int_{1/(\lambda(n+1))}^{1/\lambda} \left| \log(1 - U_n(1 - v\lambda)) - \log v\lambda - \frac{B_n(1 - v\lambda)}{v\lambda\sqrt{n}} \right| d\mu(v)$$

$$+ \sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{\delta/\lambda} b_n^\star(v\lambda)\, \frac{|B_n(1 - v\lambda)|}{v\lambda\sqrt{n}}\, d\mu(v) \equiv L_n' + L_n''.$$

In the proof of Lemma 12, we have shown that, as $\lambda \to 0$ and $n\lambda \to \infty$,

$$L_n' = \{\sup_{1/(n+1)\leq s\leq \delta} b_n^\star(s)\} o_P(1) = o_P(1).$$

Next, if we take expectations, we have in L_n''

$$\sqrt{\lambda} \int_{1/(\lambda(n+1))}^{1/\lambda} \frac{|B_n(1 - v\lambda)|}{v\lambda}\, d\mu(v) = O_P\left(\sqrt{n} \int_{1/(\lambda(n+1))}^{\infty} (v\lambda)^{-1/2}\, d\mu(v)\right)$$

$$= O_P\left(\int_0^\infty v^{-1/2}\, d\mu(v)\right) = O_P(1).$$

It follows that, as $\lambda \to 0$ and $n\lambda \to \infty$,

$$L_n = \{\sup_{1/(n+1)\leq s\leq \delta} b_n^\star(s)\} O_P(1).$$

Since $b(u) \to 0$ as $u \to 0$, it follows that, by choosing $\delta > 0$ arbitrarily small, we may make

$$\sqrt{n\lambda} \int_{1/(\lambda(n+1))}^{\delta/\lambda} |\Delta_n^{(2)}(v\lambda)|\, d\mu(v) \quad \text{as small as desired.}$$

Let us assume from now on that $\delta > 0$ is fixed, and consider the remaining terms

$$\left|\sqrt{n\lambda} \int_{\delta/\lambda}^{1/\lambda} \Delta_n^{(2)}(v\lambda)\, d\mu(v)\right| = \left| \sqrt{n\lambda} \int_{\delta/\lambda}^{1/\lambda} \left\{ \int_{1-U_n(1-v\lambda)}^{v\lambda} \frac{b(s)}{s}\, ds \right\} d\mu(v) \right|,$$

where $d\mu = K(v)\,dv$ or $d\mu = -v\,dK(v)$.

By Kolmogorov-Smirnov, with probability increasing to one as c increases to infinity, we can give an upper bound of this integral by

$$M_n = \sqrt{n\lambda} \int_{\delta/\lambda}^{1/\lambda} \left\{ \int_{v\lambda - c/\sqrt{n}}^{v\lambda + c/\sqrt{n}} \frac{|b(s)|}{s}\, ds \right\} d\mu(v).$$

Let us assume that $c > 0$ is fixed.

(a) Let $d\mu = K(v)\,dv$. We have

$$M_n = 2c\sqrt{\lambda}\int_\delta^1\left\{\frac{1}{2c/\sqrt{n}}\int_{u-c/\sqrt{n}}^{u+c/\sqrt{n}}\frac{|b(s)|}{s}\,ds\right\}K\left(\frac{u}{\lambda}\right)\frac{du}{\lambda}$$

$$\leq 2c\delta^{-1/2}\{\sup_{\delta\leq u\leq 1}(u/\lambda)^{1/2}K(u/\lambda)\}\int_\delta^1\left\{\frac{1}{2h}\int_{u-h}^{u+h}\frac{|b(s)|}{s}\,ds\right\}du,$$

where $h = c/\sqrt{n}$.

We shall make use of the following lemma.

LEMMA 15. *Let $\phi \geq 0$ be integrable on $(A-h, B+h)$, where $0 < h < (B-A)/2 < \infty$. Then we have*

$$\int_A^B\left\{\frac{1}{2h}\int_{u-h}^{u+h}\phi(s)\,ds\right\}du \leq \int_{A-h}^{B+h}\phi(u)\,du.$$

PROOF. By integration by parts.

Applying Lemma 15 we get for $n \geq 4c^2/\delta^2$,

$$M_n \leq 2c\delta^{-1/2}\{\sup_{\delta/\lambda\leq u\leq 1/\lambda}\sqrt{u}\,K(u)\}\int_{\delta/2}^1\frac{|b(s)|}{s}\,ds = o(1) \quad \text{as} \quad \lambda \to 0.$$

(b) Let $d\mu = -v\,dK(v)$, $h = c/\sqrt{n}$, and consider

$$M_n = -2c\sqrt{\lambda}\int_\delta^1\left\{\frac{1}{2h}\int_{u-h}^{u+h}\frac{|b(s)|}{s}\,ds\right\}\frac{u}{\lambda}\,dK\left(\frac{u}{\lambda}\right).$$

If we assume that, for some $\Lambda \geq 0$, $dK(u) = k(u)\,du$ for $u \geq \Lambda$, and if, in addition, we have

$$\lim_{u\uparrow\infty}u^{3/2}k(u) = 0,$$

then we can use the same argument as in (a), replacing $K(u)$ by $uk(u)$, to show that $M_n = o(1)$.

If we assume that $b(\cdot)$ is bounded on (0, 1), then we have directly

$$M_n = O\left(\sqrt{\lambda}\int_{\delta/\lambda}^\infty v\,dK(v)\right) = o(1) \quad \text{as} \quad \lambda \to 0.$$

Since $T_n = \sqrt{n\lambda}A_{n3}$, we have just proved the following lemma:

LEMMA 16. *Let K satisfy* (H1–2–3–4), *and assume that* (D1–2–3) *are satisfied. Then, as $\lambda \to 0$ and $n\lambda \to \infty$, we have*

$$\sqrt{n\lambda}\,A_{n3} = o_p(1).$$

PROOF OF THEOREM 1. According to the explanations given shortly after Lemma 10, if (H1–2–3–4) and (D1–2) are satisfied, then it follows from Lemma

7, Lemma 9 and Lemma 16 that, as $\lambda \to 0$ and $n\lambda \to \infty$,

$$a\sqrt{n\lambda}\left\{\int_0^\infty K^2(v)\,dv\right\}^{-1/2}(a_n^{-1} - a^{-1} - \beta_n) \to_w N(0, 1).$$

Since evidently this implies that $a_n \to_P a$, we have

$$\frac{\sqrt{n\lambda}}{a}\left\{\int_0^\infty K^2(v)\,dv\right\}^{-1/2}(a - a_n - aa_n\beta_n) \to_w N(0, 1),$$

or equivalently

$$\frac{\sqrt{n\lambda}}{a}\left\{\int_0^\infty K^2(v)\,dv\right\}^{-1/2}\left(a_n - \frac{a}{1 + a\beta_n}\right) \to_w N(0, 1),$$

which is the desired statement (ii) of the theorem. Statement (i) follows by the last statement of Lemma 3 and by Remark 7.

Next, we have to show that if we assume only (D1), we have still $a_n \to_P a$ if $\lambda \to 0$ and $n\lambda \to \infty$. A close look at the arguments developed from Lemma 7 to Lemma 16 shows that this will follows from:

LEMMA 17. *Under the conditions of Lemma* 7, *we have, as* $\lambda \to 0$ *and* $n\lambda \to \infty$,

$$\int_0^{1/\lambda}\{\log c(v\lambda) - \log c(1 - U_n(1 - v\lambda))\}\,d\{vK(v)\} = o_P(1).$$

PROOF. The proof of Lemma 17 is identical to the proof of Lemma 7 by Glivenko-Cantelli.

PROOF OF THEOREM 2 AND THEOREM 3. By Theorem 1(ii), we have

$$\frac{\sqrt{n\lambda}}{a}\left\{\int_0^\infty K^2(v)\,dv\right\}^{-1/2}(a_n - a + a^2\beta_n(1 + o(1))) \to_w N(0, 1).$$

The results follow directly.

2.4 *Optimal choices of* K *and* λ. We shall assume in the sequel that (D4) holds and use the notations $D_1 = C_1^{1/a}$, $D_2 = C_2/(aC_1^{b/a})$, $\alpha = b/a$, for which namely,

$$Q(1 - s) = s^{-1/a}D_1\{1 + D_2 s^\alpha(1 + o(1))\} \quad \text{as} \quad s \to 0.$$

LEMMA 18. *Let* (D1–4) *hold, and assume that, for* $0 < s < s_0$,

$$Q(1 - s) = s^{-1/a}\exp\left(\int_s^1 \frac{b(u)}{u}\,du\right).$$

Then we have

$$\log D_1 = \int_0^1 \frac{b(u)}{u}\,du \quad \textit{and} \quad \int_0^s \frac{b(u)}{u}\,du = -D_2 s^{\alpha}(1 + o(1)) \quad \textit{as} \quad s \to 0.$$

PROOF. We have for $0 < s < s_0$,

$$\int_s^1 \frac{b(u)}{u}\,du = \log D_1 + \log(1 + D_2 s^{\alpha}(1 + o(1))),$$

hence, for $s \to 0$, we get

$$\int_0^1 \frac{b(u)}{u}\,du = \log D_1$$

and

$$\int_0^s \frac{b(u)}{u}\,du = \log(1 + D_2 s^{\alpha}(1 + o(1))) \sim D_2 s^{\alpha} \quad \text{as} \quad s \to 0.$$

Let us now consider a kernel K which satisfies (H1–2–3–4–5) which imply that, for some $\Lambda < \infty$, $K(t) = 0$ for $t > \Lambda$. Let us also assume that $b(t) = 0$ for $t \geq 1$. We have then

$$M(n, \lambda, K) = \frac{1}{n\lambda} \int_0^{\infty} K^2(v)\,dv + \left\{\int_0^{\infty} b(\lambda v) K(v)\,dv\right\}^2.$$

We shall make use of the following lemma:

LEMMA 19. *Let K satisfy* (H1–2–3–4–5) *and assume that* (D1–4) *hold. Then we have, as* $\lambda \to 0$,

$$\beta_\lambda^{\star} \equiv \int_0^{\infty} b(\lambda v) K(v)\,dv = -\alpha D_2 \left\{\int_0^{\infty} v^{\alpha} K(v)\,dv\right\} \lambda^{\alpha}(1 + o(1)).$$

PROOF. There exists a λ_0 such that, for $0 < \lambda < \lambda_0$, we have

$$\beta_\lambda^{\star} = \int_0^{\lambda\Lambda} \frac{b(u)}{u} \left\{\frac{u}{\lambda} K\left(\frac{u}{\lambda}\right)\right\} du.$$

On integrating by parts, we get

$$\beta_\lambda^{\star} = -\int_0^{\lambda\Lambda} \left\{\int_0^s \frac{b(u)}{u}\,du\right\} \left\{K\left(\frac{s}{\lambda}\right)\frac{ds}{\lambda} + \frac{s}{\lambda}\,dK\left(\frac{s}{\lambda}\right)\right\}.$$

It follows evidently from Lemma 18 that, as $\lambda \to 0$,

$$\begin{aligned}\beta_\lambda^{\star} &= D_2 \left\{\int_0^{\lambda\Lambda} s^{\alpha} K\left(\frac{s}{\lambda}\right)\frac{ds}{\lambda}\right\}(1 + o(1)) + D_2 \left\{\int_0^{\lambda\Lambda} s^{\alpha} \frac{s}{\lambda}\,dK\left(\frac{s}{\lambda}\right)\right\}(1 + o(1)) \\ &= -\alpha D_2 \left\{\int_0^{\lambda\Lambda} s^{\alpha} K\left(\frac{s}{\lambda}\right)\frac{ds}{\lambda}\right\}(1 + o(1)) = -\alpha D_2 \left\{\int_0^{\infty} v^{\alpha} K(v)\,dv\right\} \lambda^{\alpha}(1 + o(1)),\end{aligned}$$

after integrating by parts.

PROOF OF THEOREM 5. Under the hypotheses of Theorem 5 (we may again assume without loss of generality that (D3) is satisfied, as noted after Lemma 10), we have, using Lemma 19, as $\lambda \to 0$,

$$M(n, \lambda, K) = \frac{1}{n\lambda} \int_0^\infty K^2(v)\, dv + \lambda^{2\alpha}(\alpha D_2)^2 \left\{ \int_0^\infty v^\alpha K(v)\, dv \right\}^2 (1 + o(1)).$$

Let $\lambda = \hat{\lambda}_n$ be the value of λ which minimizes

$$\frac{1}{n\lambda} \int_0^\infty K^2(v)\, dv + \lambda^{2\alpha}(\alpha D_2)^2 \left\{ \int_0^\infty v^\alpha K(v)\, dv \right\}^2.$$

We have evidently

$$\hat{\lambda}_n = n^{-1/(2\alpha+1)} \{2\alpha^3 D_2^2\}^{-1/(2\alpha+1)} \left(\int_0^\infty K^2(v)\, dv \right)^{1/(2\alpha+1)} \left(\int_0^\infty v^\alpha K(v)\, dv \right)^{-2/(2\alpha+1)} \to 0$$

$$\text{as} \quad n \to \infty.$$

It follows that there exists a $\Lambda_1 > 0$ such that, as $n \to \infty$,

$$M(n, \hat{\lambda}_n, K) \sim \inf_{0 < \lambda \le \Lambda_1} M(n, \lambda, K).$$

This proves the first part of Theorem 5.

We now seek a kernel $K(\cdot)$ which minimizes

$$\int_0^\infty v^\alpha K(v)\, dv,$$

when K satisfies the constraints $K \ge 0$, $\int_0^\infty K(s)\, ds = 1$, and $\int_0^\infty K^2(s)\, ds =$ Constant.

It is straightforward to see that the optimal kernel must be of the form

$$K(s) = A - Bs^\alpha, \quad 0 < s < (A/B)^{1/\alpha}.$$

We can then chose the normalizing constants to get

$$\int_0^\infty K(s)\, ds = \int_0^\infty K^2(s)\, ds = 1.$$

This gives finally $K(s) = \gamma\{\Lambda^\alpha - s^\alpha\}$, $0 < s < \Lambda$, $K(s) = 0$ otherwise, where

$$\gamma = \left\{ \frac{\alpha + 1}{\alpha} \right\} \left\{ \frac{2\alpha + 1}{2\alpha + 2} \right\}^{\alpha+1} \quad \text{and} \quad \Lambda = \frac{2\alpha + 2}{2\alpha + 1}.$$

This completes the proof of Theorem 5.

Acknowledgements. This paper was written while we were attending the Oberwolfach conference on Order Statistics, Quantile Processes, and Extreme Value Theory, March 25–31, 1984. We would like to thank the organizers, Professors R.-D. Reiss and W. van Zwet for making this collaboration possible.

We thank the referee for his precise comments and discussions.

REFERENCES

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. (1984). Weighted empirical and quantile processes. Submitted. (Paper 1 in Tech. Report No. 24, Lab. Res. Statist. Probab., Carleton Univ., Ottawa).

CSÖRGŐ, S. (1984). Adaptative estimation of the parameters of stable laws. In *Colloquia Math. Soc. J. Bolyai 36. Limit Theorems in Probability and Statistics* (P. Révész, ed.). 305–368, North Holland, Amsterdam.

CSÖRGŐ, S. and MASON, D. (1984). Central limit theorems for sums of extreme values. (Paper 2 in Techn. Report No. 25, Lab. Res. Statist. Probab., Carleton Univ. Ottawa). To appear in *Math. Proc. Cambridge Philos. Soc.*

DAVIS, R. and RESNICK, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12** 1467–1487.

DE HAAN, L. (1970). On regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tracts 32*, Amsterdam.

DE HAAN, L. (1981). Estimation of the minimum of a function using order statistics. *J. Amer. Statist. Assoc.* **76** 467–469.

DE HAAN, L. and RESNICK, S. (1980). A simple asymptotic estimate for the index of a stable distribution. *J. Roy. Statist. Soc. Ser. B* **42** 83–87.

DE MEYER, A. and TEUGELS, J. L. (1982). Limit theorems for Pareto-type distributions. Preprint.

DEVROYE, L. and PENROD, C. S. (1984). The consistency of automatic kernel density estimates. *Ann. Statist.* **12** 1241–1259.

DUMOUCHEL, W. H. (1983). Estimating the stable index α in order to measure tail thickness: a critique. *Ann. Statist.* **11** 1019–1031.

GAWRONSKI, W. and STADTMÜLLER, U. (1984). Parameter estimation for distributions with regularly varying tails. *Statistics and Decisions.* To appear.

HAEUSLER, E. and TEUGELS, J. L. (1984). On the asymptotic normality of Hill's estimator for the exponent of regular variation. Unpublished manuscript.

HALL, P. (1978). Representations and limit theorems for extreme value distributions. *J. Appl. Probab.* **15** 639–644.

HALL, P. (1981). A comedy of errors: the canonical form for a stable characteristic function. *Bull. London Math. Soc.* **13** 23–27.

HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, **44** 37–42.

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084.

HALL, P. and WELSH, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 331–341.

HILL, B. M. (1975). A simple approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.

MANDELBROT, B. (1963). The variation of certain speculative prices. *J. Bus. Univ. Chicago* **36**, 394–419.

MASON, D. (1982a). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10** 754–764.

MASON, D. (1982b). Some characterization strong laws for linear functions of order statistics. *Ann. Probab.* **10** 1051–1057.

MASON, D. (1983) The asymptotic distribution of weighted empirical distribution functions. *Stoch. Processes Appl.* **15** 99–109.

SENETA, E. (1975). Regularly varying functions. *Lecture Notes in Mathematics* **508** Springer-Verlag, Berlin.

SKOROHOD, A. V. (1954). Asymptotic formulas for stable distribution laws. *Dokl. Akad. Nauk SSSR* **98** 731–734 (In Russian. English translation in: *Select. Transl. Math. Statist. Probab.* **1** (1961), 157–161).

TEUGELS, J. L. (1981). Limit theorems on order statistics. *Ann. Probab.* **9** 868–880.

TEUGELS, J. L. (1982). Estimating the index in Pareto-type distributions. Preprint.

WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73** 812–815.

WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function, *Z. Wahrsch. verw. Gebiete* **45** 73–88.

WELSH, A. H. (1984). On the use of the empirical distribution and characteristic function to estimate parameters of regular variation. Preprint.

BOLYAI INSTITUTE
SZEGED UNIVERSITY
ARADI VÉRTANUK TERE 1
H 6726 SZEGED
HUNGARY

UNIVERSITÉ PARIS VI
t.45-55, E3, L.S.T.A.
4 PLACE JUSSIEU
75230 PARIS CEDEX 05
FRANCE

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON ST.
MADISON, WISCONSIN 53705