The Annals of Statistics
1989, Vol. 17, No. 3, 1070–1086

ON THE RELATIONSHIP BETWEEN STABILITY OF EXTREME ORDER STATISTICS AND CONVERGENCE OF THE MAXIMUM LIKELIHOOD KERNEL DENSITY ESTIMATE

BY MICHEL BRONIATOWSKI, PAUL DEHEUVELS AND LUC DEVROYE[1]

Université Pierre et Marie Curie, Université Pierre et Marie Curie and McGill University

Let f be a density on the real line and let f_n be the kernel estimate of f in which the smoothing factor is obtained by maximizing the cross-validated likelihood product according to the method of Duin and Habbema, Hermans and Vandenbroek. Under mild regularity conditions on the kernel and f, we show, among other things that $\int |f_n - f| \to 0$ almost surely if and only if the sample extremes of f are strongly stable.

1. Introduction. Let $X_1, X_2, \dots$ be an i.i.d. sequence of random variables with distribution function F and density f and consider the kernel density estimate [Parzen (1962) and Rosenblatt (1956)]

$$f_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where the kernel K is a nonnegative function such that

$$\int_{-\infty}^{\infty} K(t)\, dt = 1 \tag{K1}$$

and

(K2) for some $r > 0$, $R > 0$, $m > 0$ and $M > 0$,
$$m 1_{[-r,r]}(x) \le K(x) \le M 1_{[-R,R]}(x) \quad \text{for all } -\infty < x < \infty.$$

Inspired by maximum likelihood theory, Habbema, Hermans and Vandenbroek (1974) and Duin (1976) proposed selecting the smoothing factor $h > 0$ by maximizing

$$L_n(h) = \frac{1}{n} \sum_{j=1}^{n} \log f_{n,h}^{(j)}(X_j), \tag{1.2}$$

where, for $j = 1, \dots, n$ and $n \ge 2$,

$$f_{n,h}^{(j)}(x) = \frac{1}{nh} \sum_{1 \le i \ne j \le n} K\left(\frac{x - X_i}{h}\right). \tag{1.3}$$

Received December 1987; revised July 1988.
[1]Research supported by NSERC Grant A3456 and FCAR Grant EQ1679.
AMS 1980 *subject classification*. Primary 62G05.
Key words and phrases. Density function, nonparametric estimation, consistency, order statistics, strong stability, kernel estimate.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

To avoid any problem concerning the existence of a maximizing value of h, we will denote in the sequel by h_n any positive number such that

$$L_n(h_n) \geq \sup_{h>0} L_n(h) - \frac{C}{n}, \tag{1.4}$$

where $C > 0$ is a fixed constant.

The resulting $f_n = f_{n,h_n}$ is the so-called *cross-validated maximum likelihood kernel density estimate*. In this paper, we are concerned with the L_1-consistency of f_n. It is known [Chow, Geman and Wu (1983) and Devroye and Györfi [DG] (1985), pages 153–154] that whenever f has compact support, then almost surely as $n \to \infty$,

$$\int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx \to 0. \tag{1.5}$$

On the other hand [Schuster and Gregory (1981) and Hall (1982)], some evidence has been found that the validity of (1.5) depends heavily upon the tail behavior of f. Our first theorem stated below confirms this phenomenon by showing that (1.5) may be characterized in terms of the *stability* of the extreme values of the sample.

Following Geffroy (1958), Barndorff-Nielsen (1963) and Resnick and Tomkins (1973), we say that a random sequence Y_n is *stable* (resp. *strongly stable*) iff there exists a nonrandom sequence y_n with $Y_n - y_n \to 0$ in probability (resp. almost surely) as $n \to \infty$. We will consider the case where $Y_n = X_{1,n}$ (resp. $Y_n = X_{n,n}$) and $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics of the sample $X_1, \ldots, X_n$.

THEOREM 1. *Under* (K1) *and* (K2):

(i) *Assume that almost surely* (*resp. in probability*) *as* $n \to \infty$,

$$\int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx \to 0. \tag{1.6}$$

Then the sample extremes $X_{1,n}$ *and* $X_{n,n}$ *are strongly stable* (*resp. stable*).

(ii) *Conversely, if the sample extremes* $X_{1,n}$ *and* $X_{n,n}$ *are strongly stable* (*resp. stable*) *and if* $(1-F)/f$ *is monotone in the upper tail and* F/f *is monotone in the lower tail, then almost surely* (*resp. in probability*) *as* $n \to \infty$, (1.6) *holds.*

In view of Theorem 1 and of the simple form of f_n, it is tempting to recommend this estimate for all distributions with strongly stable sample extremes (see Lemma 5 below). However, there is evidence that cross-validated maximum likelihood kernel density estimates may have worse asymptotic performances in terms, for instance of $E(\int |f_n - f|)$, than those achieved by nonrandom choices of $h = h(n)$ in (1.1) [see, e.g., Hall (1982)]. It is therefore interesting to evaluate more closely what kind of asymptotic rate may be achieved by f_n under some general regularity assumptions on the tail behavior of f. It turns out that a suitable class for our needs is provided by all distributions whose extremes belong to the domain of attraction of a Gumbel law, i.e., for the maximum $X_{n,n}$, such that there exist nonrandom sequences $a_n > 0$ and b_n with

$$\lim_{n\to\infty} P\big(a_n^{-1}(X_{n,n} - b_n) \le x\big) = \exp(-e^{-x}). \tag{1.7}$$

As we shall see in the sequel, this implies that, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} n^{\varepsilon} E\left(\int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx\right) = \infty, \tag{1.8}$$

which is disastrous to say the least (this is in particular the case of the *normal distributions* which have strongly stable extremes in the domain of attraction of Gumbel laws; see, e.g., Remarks 1 and 2 in the sequel).

The main conclusion of these results is that in such nonparametric estimates where a smoothing factor is chosen as a function of the data itself, the tail behavior of the underlying distribution may have crucial effects. This leads to the idea that one should transform or truncate the data in order to eliminate these disturbances.

Moreover, the above technique of cross-validated maximum likelihood does not give a viable estimate in general. In fact, this method tries to optimize the Kullback–Leibler norm [Hall (1987a, b)] which is quite pathological in a number of ways and not adapted to L_p-consistency. It should be therefore used only for distributions with compact support [see also Marron (1985)] and densities having positive limits in the tails.

In the remainder of our paper, we prove Theorem 1, jointly with some technical results of independent interest. In our proofs, we choose $C = 0$ in (1.4). The general case of $C \ge 0$ follows after routine modifications.

2. Strong stability is necessary. Let $D_n = \max_{1\le i\le n} \min_{1\le j\ne i\le n} |X_i - X_j|$ for $n \ge 2$. We prove in this section the following theorem.

THEOREM 2. *Assume that K satisfies* (K1) *and vanishes outside of* $[-1,1]$. *If*

$$\int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx \to 0 \quad a.s.\ as\ n \to \infty, \tag{2.1}$$

then $X_{n,n}$ is strongly stable and $h_n \to 0$ and $D_n \to 0$ almost surely as $n \to \infty$.

PROOF. We note (see Lemma 1 below) that $h_n \to 0$ a.s. Moreover [Schuster and Gregory (1981)] the inequality

$$h_n \ge D_n \tag{2.2}$$

implies that $D_n \to 0$ a.s. This, in turn, requires that $X_{n,n} - X_{n-1,n} \to 0$ a.s., which by a result of Geffroy (1965) is equivalent to strong stability of $X_{n,n}$. □

The following lemma states a general result of independent interest.

LEMMA 1. *Let $f_{n,\tilde{h}}$ be a kernel density estimate where K satisfies* (K1), *f is an arbitrary density and $\tilde{h} = \tilde{h}(n; X_1, \dots, X_n) > 0$ is a measurable function of n and of the data. Then, if almost surely (resp. in probability) as $n \to \infty$,*

$$\int_{-\infty}^{\infty} |f_{n,\tilde{h}}(x) - f(x)|\, dx \to 0, \tag{2.3}$$

$\tilde{h} \to 0$ almost surely (resp. in probability) as $n \to \infty$.

PROOF. Denote by $\bar{f}_{n,h}$ the estimate given in (1.1) with $X_{n+1}, \dots, X_{2n}$ replacing $X_1, \dots, X_n$ and set $\bar{f}_n = \bar{f}_{n,\tilde{h}}$. Lemma S2 in Devroye (1987) jointly with the observation that $\tilde{h}$ is independent of $X_{n+1}, \dots, X_{2n}$, implies that $I_n - E(\bar{I}_n | X_1, \dots, X_n) \to 0$ a.s. as $n \to \infty$, where $I_n := \int |f_n - f|$ and $\bar{I}_n := \int |\bar{f}_n - f|$. Since $E(I_n) - E(\bar{I}_n) \to 0$ as $n \to \infty$, our assumptions imply that $E(\bar{I}_n | X_1, \dots, X_n) \to 0$. The proof now proceeds as in Theorem 4 in DG [(1985, page 8). We limit ourselves to the a.s. part. Let f have characteristic function (chf) ϕ, let K have chf ψ and denote by $\bar{\phi}_n$ the empirical chf based on $X_{n+1}, \dots, X_{2n}$. Then $\bar{f}_{n,h}$ has chf

$$\frac{1}{n} \sum_{j=n+1}^{2n} e^{itX_j} \psi(th) = \bar{\phi}_n(t) \psi(th).$$

Thus

$$\begin{aligned}\int_{-\infty}^{\infty} |\bar{f}_n(x) - f(x)|\, dx &\geq \sup_t |\bar{\phi}_n(t)\psi(t\tilde{h}) - \phi(t)| \\ &\geq \sup_{|t| \leq C} |1 - \psi(t\tilde{h})|\, |\phi(t)| - \sup_{|t| \leq C} |\bar{\phi}_n(t) - \phi(t)|\end{aligned}$$

for arbitrary C. Since the Glivenko–Cantelli theorem implies [see, e.g., Laha and Rohatgi (1979), page 153] that $\sup_{|t| \leq C} |\bar{\phi}_n(t) - \phi(t)| \to 0$ a.s. and in the mean as $n \to \infty$, the fact that $E(\bar{I}_n | X_1, \dots, X_n) \to 0$ a.s. implies that $\psi(t\tilde{h}) \to 1$ a.s. for all t in a small neighborhood of the origin. We conclude from this and the fact that ψ is a chf of a density [which implies that ψ is continuous, $|\psi(s)| < 1$ for $s \neq 0$ and $\psi(s) \to 0$ as $|s| \to \infty$] that $\tilde{h} \to 0$ a.s. □

Theorem 2 proves the easy half of the almost sure version of Theorem 1. The second half of the proof is captured in the following section.

3. Strong stability is sufficient. We will make use of the following result [see, e.g., Devroye and Penrod (1984), page 1232 and DG (1985)].

LEMMA 2. *Assume that* (K1) *holds and that $\tilde{h} = \tilde{h}(n; X_1, \dots, X_n) \to 0$ is a measurable function of n and $X_1, \dots, X_n$. Then $\tilde{h} \to 0$ and $n\tilde{h} \to \infty$ a.s. as $n \to \infty$ imply that $\int_{-\infty}^{\infty} |f_{n,\tilde{h}}(x) - f(x)|\, dx \to 0$ a.s.*

Having Lemma 2, and Theorem 2, the proof of the almost sure version of Theorem 1 boils down to the verification that $h_n \to 0$ and $nh_n \to \infty$ a.s. This is done in Theorems 3 and 4 below. Basically, the condition $nh_n \to \infty$ always holds and has nothing to do with the tail behavior of f. The stability of the extremes is required to show that $h_n \to 0$, which represents the most difficult part of the proof.

THEOREM 3. *Assume* (K1) *and* (K2). *For any density f, the cross-validated choice h_n of h satisfies*

$$\liminf_{n\to\infty} \frac{nh_n}{\log n} > 0 \; a.s. \tag{3.1}$$

PROOF. Take a large constant A and define

$$D_n^\star = \max_{i: |X_i| \le A} \; \min_{1 \le j \ne i \le n: |X_j| \le A} |X_j - X_i|,$$

if at least two X_i's fall in $[-A, A]$, and $D_n^\star = 0$ otherwise (for $1 \le i, j \le n$). Let N be the number of X_i's in $[-A, A]$ (for $1 \le i \le n$). Obviously $N/n \to \int_{-A}^{A} f$ a.s. as $n \to \infty$, which (by assuming A large enough) is strictly positive. Since the (conditional on N) density of these N r.v.'s is proportional to $f 1_{[-A, A]}$, we may apply Lemma 16 in DG [(1985), pages 181–183] to obtain that on the event $\{N \to \infty\}$,

$$\liminf_{N\to\infty} \frac{ND_N^*}{\log N} > 0 \text{ a.s.},$$

where $D_N^* := \max_{1 \le i \le n} \min_{1 \le j \ne i \le n} |X_i - X_j|$, with the X_i's and X_j's restricted to $[-A, A]$. This, jointly with the inequalities $D_n \ge D_n^\star = D_N^*$, implies (3.1) as sought. □

THEOREM 4. *Assume that K satisfies* (K1) *and* (K2). *Then, for any density f such that the sample extremes $X_{1,n}$ and $X_{n,n}$ are strongly stable together with $(1 - F)/f$ being monotone in the upper tail and F/f monotone in the lower tail, the cross-validated choice h_n of h satisfies $h_n \to 0$ almost surely as $n \to \infty$.*

PROOF. We follow the proof of Theorem 6.4 of DG (1985) corresponding to the case where the support $S := \{x: f(x) > 0\}$ of f is bounded. We use the same notation, with the exception of $T := S \cap [-A, A]$, where A is a large constant. We see that:

(i) Lemmas 6.10 and 6.12 remain valid if all integrals are taken over T.

(ii) Lemma 6.11 remains valid without change.

(iii) Lemma 6.13 is in general false when S is not compact. It is replaced by Lemma 3 below.

(iv) Lemma 6.14 is crucial to the proof. It is restated and proved in a more general setting in Lemma 4 below.

We now conclude the proof of Theorem 4 by mimicking the proof of Lemma 6.15 without change. Note that this proof requires parts (C), (E), (G) and (H) of Lemma 3, the monotonicity condition on $(1 - F)/f$ and F/f, and Lemma 4 presented below. We have to verify that property (H) of Lemma 3 can be applied. To do so, we note that the ultimate monotonicity of $(1-F)/f$ in the right tail and the stability of $X_{n,n}$ together imply that $(1-F)/f \to 0$ as $x \to \infty$ (see Lemma 5 and Remark 1 below).

The following Lemma 3 states some useful properties of the entropy needed in our proofs. In the sequel, we assume without loss of generality that $R = 1$ in (K2).

LEMMA 3. *Assume* (K1) *and* (K2) *and that* f *is a density with strongly stable extremes. Then*:

(A) $\int f \log_-(f \star K_h) > -\infty$ *and* $\int f \log_-(f \star u_h) > -\infty$ *for all* $h > 0$, *where* $\log_- = \min(\log, 0)$, $K_h = h^{-1}K(\cdot/h)$ *and* u_h *is the uniform density on* $[-h, h]$.

(B) $\int f \log_- f > -\infty$.

(C) $\int f \log(f \star K_h) < \int f \log f$ *for all* $h > 0$.

(D) *For a fixed* $h > 0$ *and a sequence* $h_n \to h$, *we have*

$$\sup_x |f \star K_{h_n} - f \star K_h| \to 0$$

as $n \to \infty$.

(E) $\int f \log(f \star K_h)$ *is continuous in* h *on* $(0, \infty)$.

(F) $\lim_{h \downarrow 0} \int f \log(f \star K_h) = \int f \log f$ *whenever*

$$\lim_{A \uparrow \infty} \liminf_{h \to 0} \inf\left(\int_A^\infty + \int_{-\infty}^{-A}\right) f \log_-(f \star K_h) = 0.$$

(G) $\lim_{h \to \infty} \int f \log(f \star K_h) = -\infty$.

(H) *Property* (F) *holds when conditions* (i) *and* (ii) *below hold simultaneously*: (i) f *has bounded support on the right, or the right tail of* f *is infinite and* $f/(1-F)$ *is ultimately nondecreasing.* (ii) f *has bounded support on the left, or the left tail of* f *is infinite and* f/F *is ultimately nonincreasing.*

PROOF. If S is bounded, then Lemma 3 is contained in Lemma 6.13 of DG (1985). So we assume without loss of generality that $F(x) < 1$ for all x.

(A) The first statement of (A) follows from the second one and (K2). For the second one, we will make use of the fact (see Lemma 6 in the sequel) that if $Q(u) = \inf\{x: 1 - F(x) \le u\}$ for $0 < u < 1$, the stability of $X_{n,n}$ implies that for all $C > 1$,

$$Q(1/Cx) - Q(1/x) \to 0 \quad \text{as } x \to \infty. \tag{3.2}$$

Let $A = kh$, where $k \ge 1$ is an arbitrary integer. By partitioning the interval $[-A, A]$ in $2k$ disjoint intervals of length h having probabilities p_i, $i = 1, \ldots, 2k$, we see that

$$\int_{-A}^{A} f \log(f \star u_h) \ge \sum_{i=1}^{2k} p_i \log\left(\frac{p_i}{2h}\right) \ge -\frac{2k}{e} - \log(2h) > -\infty, \tag{3.3}$$

where we have used the fact that $\inf_{x>0} x \log x = -1/e$. Choose by (3.2), $m \ge 1$ so large that $Q(2^{-i-1}) - Q(2^{-i}) \le h$ for all $i \ge m$, where $m \ge 1$ is such that

$Q(2^{-m}) \le A \le Q(2^{-m-1})$. We have

$$\int_A^\infty f \log(f \star u_h) = \int_A^\infty f(x) \log\left(\frac{F(x+h) - F(x-h)}{2h}\right) dx$$
$$\ge \sum_{i=m}^{\infty} \int_{Q(2^{-i})}^{Q(2^{-i-1})} f(x) \log_-(F(x+h) - F(x-h))\, dx$$
$$- \log(2h) \int_A^\infty f(x)\, dx.$$

Since $F(x+h) - F(x-h) \ge F(Q(2^{-i-1})) - F(Q(2^{-i})) = 2^{-i-1}$ for $Q(2^{-i}) \le x \le Q(2^{-i-1})$, this last expression is greater than or equal to

$$\sum_{i=m}^{\infty} 2^{-i-1} \log_- 2^{-i-1} - \log(2h) = -\sum_{i=m}^{\infty} \frac{(i+1)\log 2}{2^{i+1}} - \log(2h) > -\infty.$$

This, jointly with (3.3) and a similar argument used in the lower tail, completes the proof of (A).

(B), (C) By Jensen's inequality, $\int_S f \log((1/f) f \star K_h) \le \log(\int_S (f/f) f \star K_h) = \log(\int_S f \star K_h)$. We are done if $\int_S f \star K_h < 1$, so assume that $\int_S f \star K_h = 1$. Since equality in Jensen's inequality occurs iff $(f \star K_h)/f = f \star K_h$ a.e. in S, we must have $f(x) = 1$ whenever $f \star K_h(x) > 0$. Thus $\int |f - f \star K_h| = 0$, which by the arguments used in the proof of Lemma 1 implies $h = 0$, a contradiction.

In view of (A) and of the inequality (C) so obtained, we have $\int f \log f \ge \int f \log(f \star K_h) \ge \int f \log_-(f \star K_h) > -\infty$, and hence $\int f \log_- f > -\infty$. This establishes (B).

(D) By (K2), and using the same "$o(1)$" convention as in Lemma 1 in DG (1985), page 156,

$$|f \star (K_{h_n} - K_h)| \le C \int |K_{h_n} - K_h| + \sup_x (K_{h_n}(x) + K_h(x)) \int_{f \ge C} f$$
$$\le o(1) + \frac{2M + o(1)}{h} \int_{f \ge C} f,$$

where $C > 0$ is an arbitrary fixed constant. (D) follows from the fact that we can make the last term in the bound above as small as desired.

(E) This part follows from Lebesgue's dominated convergence theorem if, for some $0 < \delta < 1$,

$$\text{(3.4)} \qquad \int f \log_+ \left(\sup_{\nu \in H} f \star K_\nu \right) \le \infty \quad \text{and} \quad \int f \log_- \left(\sup_{\nu \in H} f \star K_\nu \right) > -\infty,$$

where $\log_+ = \max(\log, 0)$ and $H = [h(1-\delta), h(1+\delta)]$. By (K2), routine arguments show that

$$\text{(3.5)} \qquad a f \star u_b \le \inf_{\nu \in H} f \star K_\nu \le \sup_{\nu \in H} f \star K_\nu \le A f \star u_B,$$

for some suitable choices of positive constants a, A, b and B. A joint application of (A) and (3.5) completes the proof of (3.4) as sought.

(F) Lemma 6.13 in DG (1985) shows that, for any finite interval $[-A, A]$,

$$\lim_{h \downarrow 0} \int_{-A}^{A} f \log(f \star K_h) = \int_{-A}^{A} f \log f. \tag{3.6}$$

We also have, by Fatou's lemma,

$$\liminf_{h \downarrow 0} \int_{A}^{\infty} f \log_+(f \star K_h) \geq \int f \log_+ f. \tag{3.7}$$

In view of (C), routine arguments based on (3.6) and (3.7) complete the proof of (F).

(G) Is a consequence of Fatou's lemma by which

$$\liminf_{h \to \infty} \int \{-f \log_-(f \star K_h)\} \geq \int f \liminf_{h \to \infty} (-\log_-(f \star K_h)) = \infty.$$

(H) We limit ourselves to show that $\lim_{A \uparrow \infty} \liminf_{h \downarrow 0} \int_A^\infty f \log_-(f \star K_h) = 0$ under (i). A similar proof holds for $\int_{-\infty}^{-A}$ under (ii). The case of bounded support is again proved in DG (1985). In the second case, we have, for A large enough and $y \geq A - 1$, $1 - F(y) \leq f(y)/C$ for some constant C. Also, we can choose A such that $(1 - F)/f \downarrow$ above $A - 1$. Thus by (K2),

$$\int_A^\infty f \log(f \star K_h) \geq \int_A^\infty f \log\left(\frac{mrhC(1 - F(x))}{2h}\right)$$

$$= \log\left(\frac{mrC}{2}\right) \int_A^\infty f + \int_A^\infty f \log(1 - F), \quad \text{for } 0 < h < 1 = R.$$

The first term of this last expression can be made small by choosing A large enough. The second term is equal to $(1 - F(A))(\log(1 - F(A)) - 1) \to 0$ as $A \to \infty$, hence the result. □

LEMMA 4. *Let $I = [c_1, c_2] \subset (0, \infty)$ be fixed and let $L_n(h)$ be as in* (1.2). *Then, under the assumptions of Theorem* 4,

$$\limsup_{n \to \infty} \sup_{h \in I} \left| L_n(h) - \int f \log(f \star K_h) \right| = 0 \quad a.s. \tag{3.8}$$

PROOF. Let $L_{n1}(h) = (1/n)\sum_{j=1}^n 1_{[-A, A]}(X_j) \log f_{n,h}^{(j)}(X_j)$ and $L_{n2}(h) = (1/n)\sum_{j=1}^n 1_{[A, \infty)}(X_j) \log f_{n,h}^{(j)}(X_j)$. It is easily verified that for any fixed constant $A > 0$, we have

$$\sup_{h \in I} \left| L_{n1}(h) - \int_{-A}^{A} f \log(f \star K_h) \right| \to 0 \quad \text{a.s. as } n \to \infty.$$

Hence, ignoring the left tails without loss of generality, we are done if we can show that for any fixed $\varepsilon > 0$ there exists an A large enough so that

$$\inf_{h \in I} \int_A^\infty f \log(f \star K_h) > -\varepsilon \quad \text{and} \quad \sup_{h \in I} \int_A^\infty f \log(f \star K_h) < \varepsilon, \tag{3.9}$$

together with

$$\limsup_{n\to\infty}\sup_{h\in I} L_{n2}(h) < \varepsilon \quad\text{and}\quad \liminf_{n\to\infty}\inf_{h\in I} L_{n2}(h) > -\varepsilon \text{ a.s.} \tag{3.10}$$

In the first place, (3.9) follows from parts (A) and (E) of Lemma 3, jointly with (3.5) (in which we replace H by suitable intervals in terms of c_1 and c_2).

Next, we note by (K2) (recall that $R = 1$) that $f_{n,h}^{(j)}(X_j) \le (MN_j)/(nc_1)$, where N_j is the number of points (not including X_j) falling in $[X_j - c_2, X_j + c_2]$ and $h \in I$. Since $N_j < n$, we have

$$\begin{aligned} L_{2n}(h) &\le \frac{1}{n}\sum_{j=1}^{n} 1_{[A,\infty)}(X_j)\log\left(\frac{M}{nc_1}N_j\right) \\ &\le \frac{1}{n}\log\left(\frac{M}{c_1}\right)\sum_{j=1}^{n} 1_{[A,\infty)}(X_j) \quad \text{for all } h \in I, \end{aligned} \tag{3.11}$$

which in turn can be made almost surely less than ε in the upper tail if we choose A in such a way that $\int_A^\infty f \le \frac{1}{2}\varepsilon/\log(M/c_1)$. This proves the first statement in (3.10).

To complete our proof, assume without loss of generality that $r = 1$ in (K2) (R being now arbitrary). By a similar argument as used for (3.11) we have

$$L_{2n}(h) \le \frac{1}{h}\sum_{j=1}^{n} 1_{[A,\infty)}(X_j)\log\left(\frac{m}{nc_2}N_j\right). \tag{3.12}$$

By choosing A in such a way that $\int_A^\infty f \le \frac{1}{2}\varepsilon/\log(m/c_2)$, we see that all we need is to prove that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} 1_{[A,\infty)}(X_j)\log\left(\frac{n_j}{n}\right) \ge -\tfrac{1}{2}\varepsilon \quad \text{a.s.} \tag{3.13}$$

In view of (3.13) and (3.12), the proof of (3.10) completes the proof of Lemma 4. Observe that in the proofs of Lemmas 3 and 4, we have used the stability of extremes in (3.2) and (3.13). In the remainder of this section, we prove these two statements. □

Our next lemma captures some useful properties of distributions with stable extremes. Its proof follows from routine Karamata-type representations used jointly with characterizations such as given in de Haan and Hordijk (1972) and Deheuvels (1984). We omit details [see also Seneta (1975), Barndorff-Nielsen (1963) and Geffroy (1958)].

LEMMA 5. *Let $Q(u) = \inf\{x: 1 - F(x) \le u\}$ for $0 < u < 1$. The stability of $X_{n,n}$ is equivalent to:*

(A) *Q can be represented in a right neighborhood of zero by*

$$Q(u) = \eta(u) + \int_C^{1/u}\frac{\varepsilon(s)}{s}\,ds, \tag{3.14}$$

where $\eta(u)$ is bounded with finite limit η_0 as $u \downarrow 0$, $C > 0$ is a constant and $\varepsilon(s)$ is a continuous nonnegative function with limit zero as $s \to \infty$.

Assume further that f is a density such that $f/(1 - F)$ is ultimately monotone and that $F(x) < 1$ for all x. Then if $X_{n,n}$ is strongly stable, $\lim_{x\uparrow\infty} f(x)/(1 - F(x)) = \infty$, *and:*

(B) *Q can be represented in a right neighborhood of zero by*

$$Q(u) = \theta(u) + \int_C^{1/u} \frac{\alpha(s)}{s \log\log s} ds, \tag{3.15}$$

where $\theta(u)$ is bounded with finite limit θ_0 as $u \downarrow 0$, $C > e$ is a constant and $\alpha(s)$ is a continuous nonnegative function with limit zero as $s \to \infty$.

Conversely, if (B) *holds, then $X_{n,n}$ is strongly stable.*

REMARK 1. Since $Q'(u) = 1/f(Q(u))$, the change of variable $u = 1 - F(x)$ used jointly with (3.14) leads to the sufficient condition for stability of $X_{n,n}$ [Geffroy (1958)],

$$\lim_{x\to\infty} \frac{1 - F(x)}{f(x)} = 0. \tag{3.16}$$

Likewise, (3.15) gives the the sufficient condition for strong stability of $X_{n,n}$ [de Haan and Hordijk (1972)],

$$\lim_{x\to\infty} \frac{1 - F(x)}{f(x)} \log\log\left(\frac{1}{1 - F(x)}\right) = 0. \tag{3.17}$$

Using (3.17) it is easily verified that the normal distributions have strongly stable extremes. Moreover (3.16) motivates the monotone-failure-rate-type assumptions in Theorem 1.

PROOF OF (3.2). This statement follows directly from (3.14). □

PROOF OF (3.13). We make use of the representation in (3.15), assuming without loss of generality that $Q(1) = 0$, $Q(0) = \infty$ and that for $i = 1, \dots, n$, $X_{n-i+1,n} = Q(U_{i,n})$ where $U_{1,n} < \cdots < U_{n,n}$ are the order statistics of i.i.d. uniform $(0,1)$ random variables with empirical distribution function $U_n(x) = n^{-1}\#\{1 \le i \le n: U_{i,n} \le x\}$.

Fix an arbitrary $t > 0$ and let $\rho(u) = 1 - F(Q(u) - t)$ and $\sigma(u) = 1 - F(Q(u) + t)$ for $0 < u < 1$. If (3.14) holds, then for any $0 < \xi < 1$, there exists a $u_0 > 0$ such that for all $0 < u < u_0$,

$$\rho(u) < \xi u < u < u/\xi < \sigma(u) \le \tfrac{1}{2}. \tag{3.18}$$

Moreover, by (3.15), for any $\lambda > 0$, there exists a $0 < u_1 < u_0$ such that for all $0 < u < u_1$,

$$\begin{aligned} \rho(u) &< u\left(\log\left(\frac{1}{u}\right)\right)^{-2\lambda} < u < 2u\left(\log\left(\frac{1}{u}\right)\right)^{\lambda} \\ &< u\left(\log\left(\frac{1}{u}\right)\right)^{2\lambda} < \sigma(u) \quad \text{and} \quad 4\rho(u) < \sigma(u). \end{aligned} \tag{3.19}$$

Denote by $N_n(a,b) = \#\{a < X_j \le b: 1 \le j \le n\}$ the number of X_j's falling in $(a, b]$. Obviously

$$N_n(Q(u)-t, Q(u)+t) = N_n(Q(\sigma(u)), Q(\rho(u))) = nU_n(\sigma(u)) - nU_n(\rho(u)). \tag{3.20}$$

To evaluate this last expression, we use the fact [Csáki (1975, 1982)] that for any nondecreasing sequence l_n of positive constants such that $\sum_{n=1}^{\infty} 1/(nl_n^2) < \infty$, we have

$$\limsup_{n\to\infty} n^{1/2} \sup_{0\le t\le 1} \frac{|U_n(t)-t|}{l_n\sqrt{t(1-t)}} = 0 \quad \text{a.s.} \tag{3.21}$$

If we take $l_n = \log n$ in (3.21), we see that, almost surely as $n \to \infty$,

$$\sup_{0\le t\le 1} |U_n(t)-t|/\sqrt{t(1-t)} = o(n^{-1/2}\log n). \tag{3.22}$$

Note that $\sigma(u)\downarrow 0$ as $u \downarrow 0$. Moreover, by (3.18), $u^{-1}\sigma(u) \to \infty$ as $u \downarrow 0$. It follows that $(\log n)/\sqrt{n\sigma(u)} \to 0$ as $n \to \infty$, uniformly in $u \ge 1/(n\log^2 n)$. By (3.22), this implies that

$$\sup_{1/(n\log^2 n)\le u\le u_0} \left| \frac{U_n(\sigma(u)) - U_n(\rho(u))}{\sigma(u)} - 1 + \frac{\rho(u)}{\sigma(u)} \right| = o\left(\frac{\log n}{\sqrt{n\sigma(u)}}\right) \to 0 \quad \text{a.s. as } n\to\infty. \tag{3.23}$$

We will now take in (3.23) $u = U_{j,n}$ and consider the following two cases.

Case 1. Let $1/(n\log^2 n) \le U_{j,n} \le n^{-1/4}$. We note [Barndorff-Nielsen (1961) and Geffroy (1958)] that $P(U_{1,n} \le 1/(n\log^2 n) \text{ i.o.}) = 0$. Moreover, if n is so large that $n^{-1/4} < u_1 < u_0$, by (3.19), we have $1 - (\rho(u)/\sigma(u)) \ge \frac{3}{4} > \frac{1}{2}$. It follows from (3.19), (3.20) and (3.23) that, almost surely for n sufficiently large, uniformly over all $U_{j,n} \le n^{-1/4}$, we have

$$N_n(X_{j,n}-t, X_{j,n}+t) \ge \frac{n}{2}\sigma(U_{j,n}) \ge \frac{n}{2}U_{j,n}\left(\log\frac{1}{U_{j,n}}\right)^{2\lambda} \ge nU_{j,n}\left(\log\frac{1}{U_{j,n}}\right)^{\lambda}. \tag{3.24}$$

Choose now $\lambda = 4$. Using again the fact that $P(U_{1,n} \le 1/(n\log^2 n) \text{ i.o.}) = 0$, it follows from (3.24) that, ultimately with probability 1, for all $U_{j,n} \le n^{-1/4}$,

$$N_n(X_{j,n}-t, X_{j,n}+t) \ge \frac{\log^4(n\log^2 n)}{\log^2 n} > \log n. \tag{3.25}$$

Case 2. Let $n^{-1/4} \le U_{j,n} \le u_1$. Another application of (3.22) shows that in this case, $nU_{j,n} \sim j$ uniformly in j. Hence by (3.19), using again the fact that

$\rho(u) < \frac{1}{4}\sigma(u)$, we have ultimately with probability 1, for all $n^{-1/4} \le U_{j,n} \le u_1$,

$$N_n(X_{j,n} - t, X_{j,n} + t) > j/2 > \log n. \tag{3.26}$$

Moreover, for any fixed $0 < \alpha < 1$, we have ultimately with probability 1, $1 \le j \le 2n\alpha$ whenever $U_{j,n} \le \alpha$, and $j \le n^{4/5}$ whenever $U_{j,n} \le n^{-1/4}$.

Let now $t = c_2$, and consider (3.13). Recall that

$$N_j = N_n(X_{j,n} - t, X_{j,n} + t) - 1.$$

Set $\alpha = 1 - F(A) < u_1$ and observe, by (3.24) and (3.26), that almost surely as $n \to \infty$,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{\infty} 1_{[A,\infty)}(X_j)\log\left(\frac{N_j}{n}\right) \\
&\ge \frac{1}{n}\sum_{j=1}^{[n^{4/5}]}\log\left(\frac{\log n}{n}\right) + \frac{1}{n}\sum_{j=[n^{4/5}]+1}^{[2n\alpha]}\log\left(\frac{j}{2n}\right) \\
&\to 2\int_0^{\alpha}(\log s)\,ds = 2\alpha(\log\alpha - 1) = -D(\alpha), \quad \text{for } \alpha < u_1.
\end{aligned}$$

The observation that $D(\alpha)$ may be chosen as small as desired when A is sufficiently large completes the proof of (3.13). □

The proof of the strong part of Theorem 1 is now completed. In the following sections, we consider the weak limiting behavior of $\int|f_n - f|$.

4. Weak laws. In this section, we give the proof of the weak version of Theorem 1. First, we argue as in the proof of Theorem 2 that $\int|f_n - f| \to_P 0$ implies $D_n \to_P 0$, and hence that $X_{n,n} - X_{n-1,n} \to_P 0$. Obviously, this implies (3.2) which being equivalent to (3.14), implies the stability of $X_{n,n}$. A similar argument holds for $X_{1,n}$.

A close look at the proof of Theorem 3 shows that it remains valid with "in probability" replacing "almost surely." The only change is to replace (3.13) by the statement that, for any fixed $t > 0$ and $\varepsilon > 0$, there exists a finite constant A such that

$$\limsup_{n\to\infty} P\left(\frac{1}{n}\sum_{j=1}^{n} 1_{[A,\infty)}(X_j)\log\left(\frac{N_j}{n}\right) \ge -\varepsilon\right) \le \varepsilon. \tag{4.1}$$

The proof of (4.1) is similar to that of (3.13) with slight modifications. Let $0 < \xi < 1$ be such that for all $0 < u < u_0 = u_0(\xi)$, (3.18) holds. By Wellner (1978), we have

$$\lim_{n\uparrow\infty}\liminf_{n\to\infty} P(u/\eta < U_n(u) < u\eta : U_{1,n} \le u \le 1) = 1. \tag{4.2}$$

Since $U_n(u) = 0$ for $0 \le u < u_{1,n}$, (4.2) ensures that, for any $\varepsilon > 0$, there exist $\eta > 1$ and n_0 such that $n \ge n_0$ implies

$$(4.3) \quad P\left(\frac{U_n(\sigma(u)) - U_n(\rho(u))}{\sigma(u)} > \frac{1}{\eta}\left(1 - \frac{\eta^2\rho(u)}{\sigma(u)}\right) : U_{1,n} \le \sigma(u) \le 1 \right) > 1 - \frac{1}{3}\varepsilon.$$

Choose ξ in (3.19) in such a way that $\eta^2\xi^2 = \frac{1}{4}$. We have by (4.3),

$$P\left(\frac{U_n(\sigma(u)) - U_n(\rho(u))}{\sigma(u)} > \frac{3}{4\eta} : U_{1,n} \le \sigma(u),\, u \le u_0 \right) > 1 - \frac{1}{3}\varepsilon,$$

which, in view of (3.18) and (3.20) and by setting $u = U_{j,n}$, implies that

$$P\left(N_n(X_{j,n} - t,\, X_{j,n} + t) > \tfrac{3}{2} n U_{j,n} : 1 \le j \le nU_n(u_0) \right) > 1 - \tfrac{1}{3}\varepsilon,$$

where we have used the fact that $\sigma(u)/\eta > u/\xi\eta = 2u$. By (4.2) this implies in turn that

$$(4.4) \quad P\left(N_n(X_{j,n} - t,\, X_{j,n} + t) > \frac{3j}{2\eta} : 1 \le j \le nU_n(u_0) \right) > 1 - \frac{2}{3}\varepsilon.$$

Set now $\alpha = 1 - F(A) < u_0$ and $n_1 \ge n_0$ such that for all $n \ge n_1$, $P(U_n(u_0) < 1 - F(A)) < \frac{1}{3}\varepsilon$. We have by (4.4),

$$P\left(\frac{1}{n} \sum_{j=1}^{n} 1_{[A,\infty)}(X_j) \log\left(\frac{N_j}{n} \right) \ge -\varepsilon \right)$$
$$\le \varepsilon + P\left(\frac{1}{n} \sum_{j=1}^{[n\alpha]} \log\left(\frac{3j}{2\eta n} \right) \ge -\varepsilon \right) \quad \text{for } n \ge n_1,$$

which is equal to ε for n sufficiently large and all $0 < \alpha < u_0$ such that $\int_0^\alpha \log(3x/\eta)\,dx \ge -\frac{1}{2}\varepsilon$.

This completes the proof of (4.1). The proof of Lemma 6.15 in DG (1985) requires small changes. We omit the details.

5. Bad performances of cross-validated estimates: The Gumbel case. In order to motivate this section, we consider a distribution with stable maximum $X_{n,n}$, i.e., such that $Q(u) = \inf\{x: 1 - F(x) \le u\}$ has the representation (3.14) of Lemma 4. We introduce the additional regularity condition that (3.16) holds, i.e., that

$$(5.1) \quad \lim_{x \uparrow \infty} \frac{1 - F(x)}{f(x)} = \lim_{u \downarrow 0} \frac{u}{f(Q(u))} = 0.$$

By (5.1), we see that the representation (3.14) may be stated as

$$(5.2) \quad Q(u) = \eta_0 + \int_C^{1/u} \frac{\varepsilon(s)}{s}\,ds \quad \text{for } 0 < u \le u_0,$$

where $C > 0$, η_0 and $0 < u_0 < 1$ are suitable constants and $\varepsilon(1/u) = u/f(Q(u))$.

In the case where the support of f is unbounded above, the positive function $\varepsilon(s)$ must satisfy simultaneously to the conditions

$$\varepsilon(s) \to 0 \quad \text{as } s \to \infty \quad \text{and} \quad \int_0^\infty \frac{\varepsilon(s)}{s}\, ds = \infty. \tag{5.3}$$

In view of (5.3), if we additionally assume that $\varepsilon(s)$ has *regular variation in the upper tail*, the only possibility is that $\varepsilon(s)$ is *slowly varying at infinity*, i.e.,

$$\text{(S)} \qquad \text{for all } \lambda > 0, \quad \lim_{s \uparrow \infty} \frac{\varepsilon(\lambda s)}{\varepsilon(s)} = 1, \quad \text{where} \quad \varepsilon(s) = \frac{1}{\{sf(Q(1/s))\}} \quad \text{for } s \geq s_0 = \frac{1}{u_0}.$$

It turns out [see, e.g., Gnedenko (1943), de Haan (1970) and Sweeting (1985)] that the condition (S) implies in general that $X_{n,n}$ belongs to the domain of attraction of a *Gumbel distribution*, i.e., that

$$\lim_{n \to \infty} P(X_{n,n} - b_n \leq a_n x) = \exp(-e^{-x}), \tag{5.4}$$

where

$$a_n = \varepsilon(n) = 1/\{nf(Q(1/n))\} \quad \text{and} \quad b_n = Q(1/n). \tag{5.5}$$

Under (S) (note that here, we do not necessarily assume that the support of f is bounded above), it is clear from (5.4) that a necessary and sufficient condition for $X_{n,n}$ to be stable is that $a_n \to 0$ which coincides with the (sufficient) condition (3.16). Moreover, straightforward computations [see, e.g., Lemmas 4–10 and Remark 2 in Deheuvels (1986)] show that

$$\begin{aligned} &\lim_{n \to \infty} P(X_{n,n} - X_{n-1,n} > a_n x,\, X_{n-1,n} - X_{n-2,n} > a_n y) \\ &\qquad = \exp(-x - 2y) \quad \text{for all } x \geq 0 \text{ and } y \geq 0. \end{aligned} \tag{5.6}$$

In view of (5.6) and of the slow variation of a_n, we see from (2.2) that

$$\text{for any } \varepsilon > 0, \quad \lim_{n \to \infty} n^{\varepsilon} h_n = \infty \quad \text{in probability}. \tag{5.7}$$

Thus, for all distributions in (S), the rate of convergence to zero of h_n is dramatically slow. The purpose of the following propositions is to show that the same holds for $\int |f_n - f|$.

THEOREM 5. *Under* (K1) *and* (K2) *and the assumption that K is Lipschitz, for all densities f such that $\int f \log(1 + f) < \infty$ jointly with* (S): *$u/f(Q(u))$ is slowly varying at zero, we have for all $\varepsilon > 0$,*

$$\lim_{n \to \infty} n^{\varepsilon} E\left(\int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx \right) = \infty. \tag{5.8}$$

PROOF. It follows from (5.7) and the remark that the stability of $X_{n,n}$ implies that $\int^{\infty} \log(1 + |x|) f(x)\, dx < \infty$, by an application of the following lemma which is of independent interest.

LEMMA 6. *Let $f_{n,\tilde{h}}$ be kernel density estimate, where K is a nonnegative Lipschitz kernel satisfying (K1), $\tilde{h} = \tilde{h}(n; X_1, \dots, X_n) > 0$ is a measurable function of n and of the data and f a density satisfying $\int f \log(1 + f) < \infty$ and $\int \log(1 + |x|) f(x)\,dx < \infty$. Then there exist positive constants A, B depending on K and f only, and a universal constant C, such that, for all n large enough,*

$$(5.9)\qquad E\left(\int |f_{n,\tilde{h}}(x) - f(x)|\,dx\right) \geq E\left(\min(A\tilde{h}^2, B)\right) - C\left\{\frac{\log n}{n}\right\}^{1/2}.$$

PROOF. We argue as in the proof of Lemma 1, using the same notation. We have by (2.4),

$$\int |\bar{f}_n - f| \geq \sup_s |\bar{\phi}_n(s)\psi(s\tilde{h}) - \psi(s)| \geq |1 - \psi(t\tilde{h})|\,|\phi(t)| - |\bar{\phi}_n(t) - \phi(t)|$$

for arbitrary t. By the Cauchy–Schwarz inequality, we have

$$E\left(|\bar{\phi}_n(t) - \phi(t)|\right) \leq n^{-1/2}\left\{E\left(|e^{itX_1} - \phi(t)|^2\right)\right\}^{1/2} \leq n^{-1/2}.$$

Let us choose s such that $|\phi(s)| \geq \frac{1}{2}$. We obtain the inequality

$$(5.10)\qquad E\left(\int |\bar{f}_n - f|\right) \geq \tfrac{1}{2}E\left(|1 - \psi(s\tilde{h})|\right) - n^{-1/2}.$$

Let m_i (resp. M_i) be the ith moment (resp. absolute moment) of K. By a truncated Taylor series expansion of e^{itx} we obtain

$$1 - \psi(t) = -itm_1 + \tfrac{1}{2}t^2 m_2 + \gamma(t),$$

where $|\gamma(t)| \leq \frac{1}{6}M_3|t|^3$. From this we see that

$$(5.11)\qquad |1 - \psi(th)| \geq \tfrac{1}{2}(th)^2 m_2 - \tfrac{1}{6}|th|^3 M_3 \geq \tfrac{1}{4}(th)^2 m_2 := 2Ah^2$$

for $|th| \leq 3m_2/(2M_3)$, while for $|th| > 3m_2/(2M_3)$, we have

$$(5.12)\qquad |1 - \psi(th)| \geq \inf_{|u| > 3m_2/(2M_3)} |1 - \psi(u)| := 2B.$$

By (5.10), (5.11) and (5.12), we can conclude that

$$(5.13)\qquad E\left(\int |\bar{f}_n - f|\right) \geq E\left(\min(A\tilde{h}^2, B)\right) - n^{-1/2}.$$

From the proof of Theorem 9 of Devroye (1988), we retain that, under our assumptions,

$$(5.14)\qquad \left|E\left(\int |f_n - f|\right) - E\left(\int |\bar{f}_n - f|\right)\right| \leq C\left\{\frac{\log n}{n}\right\}^{1/2}$$

for all n large enough where C is a universal constant ($C > \sqrt{10240}$ will do). A joint application of (5.13) and (5.14) completes the proof of Lemma 6. □

REMARK 2. It is easily verified that (1) the normal distributions have both extremes in the domain of attraction of a Gumbel distribution and (2) the exponential distributions have the upper extreme in the domain of attraction of a Gumbel distribution. Hence, in Cases 1 and 2 we have (5.8). Moreover, the upper tail of an exponential distribution is not stable, so that in this case f_n is not even L_1-consistent.

REMARK 3. Interestingly, the phenomenon (5.8) described in Theorem 5 occurs also for all densities f with bounded support having at least an extreme value in the domain of attraction of Gumbel's distribution. Hence, even in this case, one has to be very cautious in the use of f_n.

REFERENCES

BARNDORFF-NIELSEN, O. (1961). On the rate of growth of the partial maxima of a sequence of independent identically distributed random variables. *Math. Scand.* **9** 383–394.

BARNDORFF-NIELSEN, O. (1963). On the limit behavior of extreme order statistics. *Ann. Math. Statist.* **34** 992–1002.

CHOW, Y. S., GEMAN, S. and WU, L. D. (1983). Consistent cross-validated density estimation. *Ann. Statist.* **11** 25–38.

CSÁKI, E. (1975). Some notes on the law of the iterated logarithm for the empirical distribution function. In *Limit Theorems in Probability Theory. Colloq. Math. Soc. János Bolyai* (P. Révész, ed.) **11** 45–58. North-Holland, Amsterdam.

CSÁKI, E. (1982). On the standardized empirical distribution function for the empirical distribution function. In *Nonparametric Statistical Inference. Colloq. Math. Soc. János Bolyai* (I. Vincze, ed.) **32** 123–138. North-Holland, Amsterdam.

DE HAAN, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes.* Math. Centre Tracts, Amsterdam.

DE HAAN, L. and HORDIJK, A. (1972). The rate of growth of the sample maxima. *Ann. Math. Statist.* **43** 1185–1196.

DEHEUVELS, P. (1984). Strong limit theorems for maximal spacings from a general univariate distribution. *Ann. Probab.* **12** 1181–1193.

DEHEUVELS, P. (1986). On the influence of the extremes of an i.i.d. sequence on the maximal spacings. *Ann. Probab.* **14** 194–208.

DEVROYE, L. (1983). On arbitrarily slow rates of global convergence in density estimation. *Z. Wahrsch. verw. Gebiete* **62** 475–483.

DEVROYE, L. (1987). An L_1 asymptotically optimal kernel estimate. Technical Report, McGill Univ.

DEVROYE, L. (1988). The kernel estimate is relatively stable. *Probab. Theory Related Fields* **77** 521–536.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

DEVROYE, L. and PENROD, C. S. (1984). The consistency of automatic kernel density estimates. *Ann. Statist.* **12** 1231–1249.

DUIN, R. P. W. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Trans. Comput.* **C-25** 1175–1179.

GEFFROY, J. (1958). Contributions à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris* **7** 37–185.

GEFFROY, J. (1965). Sur une condition nécessaire et suffisante de stabilité presque sûre des valeurs extrêmes d'un échantillon. *C. R. Acad. Sci. Paris Ser. A–B* **260** 3271–3273.

GNEDENKO, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* (2) **44** 423–453.

HABBEMA, J. D. F., HERMANS, J. and VAN DEN BROEK, K. (1974). A stepwise discriminant analysis program using density estimation. In *Compstat 1974* (G. Bruckmann, ed.) 101–110. Physica, Vienna.

HALL, P. (1982). Cross-validation in density estimation. *Biometrika* **69** 383–390.

HALL, P. (1987a). On Kullback–Leibler loss and density estimation. *Ann. Statist.* **15** 1491–1519.

HALL, P. (1987b). On the use of compactly supported density estimates in problems of discrimination. *J. Multivariate Anal.* **23** 131–158.

LAHA, R. G. and ROHATGI, V. K. (1979). *Probability Theory.* Wiley, New York.

MARRON, J. S. (1985). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. *Ann. Statist.* **13** 1011–1023.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

RESNICK, S. I. and TOMKINS, R. J. (1973). Almost sure stability of maxima. *J. Appl. Probab.* **10** 387–401.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SCHUSTER, E. F. and GREGORY, G. G. (1981). On the nonconsistency of maximum likelihood nonparametric density estimators. In *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface* (W. F. Eddy, ed.) 295–298. Springer, New York.

SENETA, E. (1975). *Regularly Varying Functions. Lecture Notes in Math.* **508**. Springer, Berlin.

SWEETING, T. J. (1985). On the domain of local attraction in extreme value theory. *Ann. Probab.* **13** 196–205.

WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. verw. Gebiete* **44** 73–88.

MICHEL BRONIATOWSKI
LABORATOIRE DE STATISTIQUE THÉORIQUE ET APPLIQUÉE
UNIVERSITÉ PIERRE ET MARIE CURIE
TOUR 45-55, 3ÈME ETAGE
4 PLACE JUSSIEU
75252 PARIS CEDEX 05
FRANCE

PAUL DEHEUVELS
7 AVENUE DU CHÂTEAU
92340 BOURG-LA-REINE
FRANCE

LUC DEVROYE
SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
MONTRÉAL, QUÉBEC
CANADA H3A 2K6