The Annals of Statistics
1996, Vol. 24, No. 3, 1423–1431

ROBUST ESTIMATION OF THE LOCATION OF A VERTICAL TANGENT IN DISTRIBUTION

BY R. V. ERICKSON

Michigan State University

It is shown that the location of the set of $m + 1$ observations with minimal diameter, within local data, is a robust estimator of the location of a vertical tangent in a distribution function. The rate of consistency of these estimators is shown to be the same as that of asymptotically efficient estimators for the same model. Robustness means (1) only properties of the distribution local to the vertical tangent play a role in the asymptotics, and (2) these asymptotics can be proven given approximate information about just two parameters, the shape and quantile of the vertical tangent.

1. Introduction. For each integer $n \geq 1$ let $\hat{\theta}_n$ be a measurable, real function of n independent observations from the distribution $F(\cdot, \theta)$, θ real. We call such functions *estimators* of θ, and we say their sequence has *consistency rate* c_n if $\{c_n(\hat{\theta}_n - \theta),\ n \geq 1\}$ is tight. Under certain regularity assumptions, there are $\sqrt{n}$ consistent estimators of θ, but in some irregular cases it is possible to find estimators with $c_n/\sqrt{n} \to \infty$.

One of the first general results of this kind is given by Chernoff and Rubin [1] in their study of the location of a discontinuity in density. Under some conditions, they prove consistency rate n for the maximum likelihood estimator (MLE). Perhaps of more interest is the fact that "quasi-maximum likelihood estimators" (qMLE's) are show to have the same asymptotic behavior as MLE's, where the qMLE is based only on local data, known to be close to the point of discontinuity, and the likelihood that is maximized uses a simple density that is a local approximation to the true density.

Hall [4] considers the problem of estimating the endpoint of a distribution. He, too, introduces qMLE's and shows them to have consistency rate n^r, some $r > 1/2$. Here, again, the asymptotic properties of the estimators are shown to depend only on a few parameters associated with local properties of the underlying density. Hall calls such estimators *robust*.

These two papers seem to indicate that, in the presence of irregularities defined by local properties, robust local estimators may exist. The purpose of this paper is to show this to be so for the *location model*, $F(\cdot - \theta)$, when F has a vertical tangent at the origin. Let us look at what is known in this case.

Received March 1993; revised July 1995.
AMS 1991 *subject classifications*. Primary 62F12; secondary 62E20.
Key words and phrases. Robust, hyperefficient estimation, singularity, asymptotic distribution.

Ibragimov and Has'minskii [5] suppose that F has a known Lebesgue density f with a *singularity of order* α, $-1 < \alpha < 0$, a condition that forces $f(x)$ to spike to ∞ near $x = 0$ like $|x|^{\alpha}$, giving F a vertical tangent at the origin; they show there exist asymptotically efficient estimators of location with consistency rate n^r, $r := (1/1 + \alpha) > 1$. (See pages 283, 314 and 394 for definition, asymptotics and history.) But there are no robustness results, and, in addition, it seems very difficult to evaluate the efficient estimators with accuracy of order $1/n^r$. This raises the *main problem*: are there robust, easily evaluated estimators of location, with the same consistency rate as the optimal ones, when F has a vertical tangent?

A little thought leads to a possible solution: since the empirical distribution uniformly converges to the actual distribution, data points should be very close together near a vertical tangent. The purpose of this paper is to show that the location of the set of $m + 1$ observations with the smallest diameter, within the local data, is a robust estimator with the appropriate consistency rate. Since such estimators are very easy to evaluate, we have a positive answer to the question above.

Here is the structure of the paper: In Section 2, we define weak singularity and the relevant estimators, and then state the main theorem. In Sections 3 and 4, we recall a standard representation for order statistics and then prove two conditions equivalent to weak singularity. Section 5 provides a key decomposition, after which the main theorem is easily proved in the last section.

2. Weak singularities and local estimators.

Let F be a distribution function and let $r > 1$, $\rho := 1/r$. We say that F has a *weak singularity of degree* r (at the origin) if there are positive ε and bounded, measurable ϕ so that

$$F(x) - F(0) = \int_0^x \phi(t)\rho|t|^{\rho-1}\,dt, \qquad |x| \le \varepsilon,$$

and

$$c_{\pm 1} := \lim_{0<t\downarrow 0} \phi(\pm t) \text{ exists}, \qquad c_{-1} + c_1 > 0.$$

With such a distribution we associate *length* ε and *limits* $c_{\pm 1}$.

This definition of singularity is quite similar to that given by Ibragimov and Has'minskii [5], page 283, except there, in addition, ϕ must be continuous in the intervals $(-\varepsilon, 0)$ and $(0, \varepsilon)$ and must satisfy a further integral smoothness condition. They use the term *order* α while we use the term *degree* r, where the relation is $r = 1/(1 + \alpha)$, $\alpha = (1/r) - 1$. We relate distributions with a singularity to those with a vertical tangent in Section 4.

Our problem is to estimate θ in the location model, $F(\cdot - \theta)$, when F is an element of $\mathscr{F}(q_{-1}, q_1, r_0)$, *the set of all* F *such that* F *is a distribution with a weak singularity of some degree* r *and length* ε *for which*

$$F(-\varepsilon) < q_{-1} < F(0) < q_1 < F(\varepsilon), \qquad r > r_0 > 1.$$

Members of the family are said to have a singularity that is *locally observable* and *bounded below*. In general, the parameters $c_{\pm 1}$, $r, q := F(0)$ and ε are unknown, but we estimate many of them. Notice that local observability implies $0 < q < 1$.

All of our estimators are defined in terms of $X_{n:\,1}, \dots, X_{n:\,n}$, the order statistics of n independent observations from $F(\cdot - \theta)$. Fix any positive integer $m > 1/(r_0 - 1)$ and define the *m-diameters* $\delta_{n,k} := X_{n:\,k+m} - X_{n:\,k}$ and the *local index of minimal m-diameter*

$$K_n := \min\{k\colon \Delta_{n,k} \le \Delta_{n,j},\ nq_{-1} < k, j < nq_1\}.$$

Our estimators of θ, r and q are given by

$$\hat{\theta}_n := X_{n:\,K_n}, \qquad \hat{r}_n := -\ln(\Delta_{n,K_n})/\ln n, \qquad \hat{q}_n := K_n/n.$$

The length parameter, ε, has no asymptotic significance and is not estimated.

The parameters $C_{\pm 1} := 1/c^r_{\pm 1}$, with $1/0 := \infty$, play a fundamental role in various asymptotic distributions obtained here, and they have interesting estimators. To define them, introduce some notation: for any fixed τ, $0 < \tau < 1$, and $n > 1$, let $\beta_n := (\ln n)^{-\tau}$, $\kappa_n := n^{1-\beta_n}$, let k_n denote the integer part of κ_n and let

$$A_{-1,n} := X_{n:\,K_n} - X_{n:\,K_n - k_n}, \qquad A_{1,n} := X_{n:\,K_n + k_n} - X_{n:\,K_n},$$
$$\hat{C}_{\pm 1, n} := n^{\beta_n \hat{r}_n} A_{\pm 1, n}.$$

THEOREM 2.1 (Robust consistency). *Let $F \in \mathscr{F}(q_{-1}, q_1, r_0)$ and construct estimators as above. If the data arise from $F(\cdot - \theta)$, $q = F(0)$ and F has a weak singularity of degree r with limits $c_{\pm 1}$, then the three sequences,*

$$n^r(\hat{\theta}_n - \theta), \quad \ln n(\hat{r}_n - r) \quad \textit{and} \quad \sqrt{n}\,(\hat{q}_n - q),$$

each converge in law to nontrivial variables; the distributions of the first two depend only on r, c_{-1} and c_1, while the latter is normal with mean 0 and variance $q(1-q)$. Also, $\hat{C}_{\pm 1, n}$ converges to $C_{\pm 1}$ in probability; in addition, if $C_{\pm 1} < \infty$ and there exist positive M, σ and $\varepsilon_0 < \min(\varepsilon, 1/2)$ such that $|\phi(\pm t) - c_{\pm 1}| < M(-\ln t)^{-\sigma}$, for $0 < t \le \varepsilon_0$, and if we choose $0 < \tau < \sigma/(1+\sigma)$, then

$$(\ln n)^{\tau}(\hat{C}_{\pm 1, n} - C_{\pm 1})$$

converges in law to a distribution depending only on r, c_{-1} and c_1.

Clearly, $\hat{\theta}_n - \theta$, $\hat{q}_n$, $\hat{r}_n$ and $\hat{C}_{\pm 1, n}$ are independent of θ, so there is no loss of generality in assuming $\theta = 0$. This we do from now on.

Our estimators of location are robust: (1) the asymptotic distribution depends only on local parameters of the singularity, and (2) the asymptotics are proven when the shape (degree r) and quantile $[q = F(0)]$ are only approximately known and the unknown limits satisfy $c_{-1} + c_1 > 0$.

3. Representation of order statistics. In this section we present one well-known way of generating the order statistics $X_{n:1}, \dots, X_{n:n}$ of n independent observations from an arbitrary distribution F. This representation and the strong law of large numbers will allow us to prove almost sure convergence rather than convergence in law in most cases. Of course, if the order statistics are generated some other way, only convergence in law continues to hold.

Given F with $0 < q = 1 - p := F(0) < 1$, define associated distributions

$$F_{-1}(x) := (F(0) - F(-x-))/q, \quad F_1(x) := (F(x) - F(0))/p, \quad x \geq 0.$$

Let

$$F_{\pm 1}^{-1}(u) := \inf\{x \geq 0 : F_{\pm 1}(x) \geq u\}, \qquad u \geq 0,$$

denote the corresponding left-continuous inverses.

Construct $\{V_j,\ j = 0, \pm 1, \pm 2, \dots\}$ iid exponential random variables with mean 1, and define the partial sums

$$S_k := \sum_{j=-\infty}^{\infty} ([k \leq j \leq 0] + [1 \leq j \leq k]) V_j,$$

where $[\cdots]$ denotes the indicator function. Independently of all V_j construct random variables N_n which have binomial distribution with parameters n and q; let $M_n := n + 1 - N_n$.

LEMMA 3.1 (Representation). *Introduce the variables of the previous three paragraphs. The order statistics $X_{n:k}$ are given by $X_{n:k} = Y_{n:k-N_n}$, where*

$$Y_{n:j} := -F_{-1}^{-1}(U_{n:j})[1 - N_n \leq j \leq 0] + F_1^{-1}(U_{n:j})[1 \leq j \leq n - N_n],$$

and

$$U_{n:j} := (S_j / S_{-N_n})[1 - N_n \leq j \leq 0] + (S_j / S_{M_n})[1 \leq j \leq n - N_n]$$

give the order statistics of N_n and $n - N_n$ independent observations from the uniform distribution on (0, 1).

Such a representation is well known, and it, and a variant, are used by Chernoff and Rubin [1] and Hall [4]. See Feller [3], page 75, for a result concerning order statistics from uniform observations.

4. Equivalent forms of weak singularity. In this section we show an equivalence between singularity and the notion of a vertical tangent in distribution. This result is important because it expresses singularity in a way that can be used with the representation of order statistics given in the previous section.

LEMMA 4.1 (Equivalence). *Let G be a nondecreasing, continuous function on the interval $[0, \varepsilon]$, with $G(0) = 0$ and $\varepsilon > 0$. Form the left-continuous, strictly increasing inverse*

$$G^{-1}(u) := \inf\{x \geq 0 : G(x) \geq u\}, \qquad 0 \leq u \leq G(\varepsilon).$$

Let $r > 1$ and $\rho := 1/r$, and, for $0 < z \le \varepsilon$, suppose $a = a(z) < b = b(z)$ are nonnegative numbers. The following are equivalent:

There is a measurable γ so that

$$G(x) = \int_0^x \gamma(t)\rho t^{\rho-1}\,dt, \qquad a \le \gamma(t) \le b, \qquad 0 \le t \le z, \tag{1}$$

$$a^r \le \frac{G^r(y) - G^r(x)}{y - x} \le b^r, \qquad 0 \le x < y \le z, \tag{2}$$

$$a^r \le \frac{v^r - u^r}{G^{-1}(v) - G^{-1}(u)} \le b^r, \qquad 0 \le u < v \le G(z). \tag{3}$$

Hence, if on interval $[0, \varepsilon]$, G has density γ with respect to the measure $\rho t^{\rho-1}\,dt$, then

$$\left(\gamma_* \circ G^{-1}(v)\right)^r \le \frac{v^r - u^r}{G^{-1}(v) - G^{-1}(u)} \le \left(\gamma^* \circ G^{-1}(v)\right)^r, \qquad 0 \le u < v \le G(\varepsilon), \tag{4}$$

where

$$\gamma_*(t) := \inf_{0 < s \le t} \gamma(s), \qquad \gamma^*(t) := \sup_{0 < s \le t} \gamma(s).$$

If a_n and $0 \le u_n < v_n$ satisfy $v_n \to 0$, $a_n u_n \to L_u$, $a_n v_n \to L_v > L_u \ge 0$ and

$$0 \le \gamma_0 := \gamma_*(0+) = \gamma^*(0+) < \infty$$

then

$$\lim_n \frac{L_v^r - L_u^r}{a_n^r\left(G^{-1}(v_n) - G^{-1}(u_n)\right)} = \gamma_0^r. \tag{5}$$

PROOF. The properties attributed to G^{-1} are well known and follow easily from its definition.

To prove the equivalence between (1) and (2), use the change of variable formula in [2], 2.9.21, on the right sides of these expressions:

$$G^r(y) - G^r(x) = \int_{G(x)}^{G(y)} rt^{r-1}\,dt, \qquad G(y) - G(x) = \int_{G^r(x)}^{G^r(y)} \rho t^{\rho-1}\,dt.$$

Specifically, assuming (1), it is clear that $ax^\rho \le G(x) \le bx^\rho$, and then (2) follows from the change of variable formula which gives

$$G^r(y) - G^r(x) = \int_x^y rG^{r-1}(t)\gamma(t)\rho t^{\rho-1}\,dt.$$

Conversely, if (2) holds, then $\Gamma := G^r$ is absolutely continuous, so that the derivative Γ' exists a.e., with bounds

$$a^r x \le \Gamma(x) \le b^r x, \qquad a^r \le \Gamma'(x) \le b^r \quad \text{a.e.}$$

A change of variable gives

$$G(y) - G(x) = \int_x^y \rho\Gamma^{\rho-1}(t)\Gamma'(t)\,dt = \int_x^y \gamma(t)\rho t^{\rho-1}\,dt,$$

where a.e.

$$\gamma(t) := (\Gamma(t)/t)^{\rho-1}\Gamma'(t)$$

is bounded between a and b, and (1) follows.

Next, assume (2). For each u, v, $0 \le u < v \le G(\varepsilon')$, let $x := G^{-1}(u)$ and $y := G^{-1}(v)$. Since G is continuous, $x < y$ and $u = G(x)$, $v = G(y)$, and (3) follows by substitution in (2).

Finally, assume (3). If $a = a(\varepsilon') > 0$, then (3) implies G^{-1} is continuous, so that G is strictly increasing, and (2) follows from substitution, as above. So assume $a = 0$; then the lower bound in (2) is clear, and it remains to prove the upper bound. Fix x, y, $0 \le x < y \le \varepsilon'$, and set $u = G(x)$, $v = G(y)$. Then (2) is clear if $u = v$, so assume $u < v$ and choose any u' so that $u \le u' < v' := v$. As above, (2) holds at $x' := G^{-1}(u')$ and $y' := G^{-1}(v')$. It also holds at x' and y, since $y \ge y'$ and $v = v'$. Finally, let u' decrease to u, so that x' decreases to some x''. Then (2) holds at x'' and y, and then also x and y, since $u = G(x'')$ and $x \le x''$.

Part (4) is now immediate, and (5) follows from (4). □

5. Decomposition of $K_n = J_n + N_n$. In the general location problem where θ is unknown, we cannot observe N_n, the number of observations less than θ. But we show now that N_n is finitely close to the observable statistic K_n. This is the key result, and it makes the proof of the main theorem fairly easy.

Throughout the remainder of this paper we assume (1) $F \in \mathscr{F}(q_{-1}, q_1, r_0)$, (2) order statistics from n independent observations are constructed as in the representation lemma and (3) the estimators and notations are as given in Section 2.

When we apply the equivalence lemma to $G = F_1$, we have $\gamma(t) = \phi(t)/p$, $\gamma_0 = c_1/p$ and $C_1 = 1/c_1^r$. For fixed $j \ge 1$,

$$U_{n:j} \to 0, \qquad nU_{n:j} \to S_j/p \quad \text{a.s.}$$

Thus, from part (5) of the equivalence lemma, we have

$$n^r(Y_{n:j+m} - Y_{n:j}) \to C_1(S_{j+m}^r - S_j^r), \qquad j \ge 1, \quad \text{a.s.,} \tag{6}$$

where the limit is ∞ if $c_1 = 0$. Similar results hold for $j \le 0$.

To decompose K_n, define

$$\delta_{n,j} := n^r \Delta_{n,j+N_n} = n^r(Y_{n:j+m} - Y_{n:j}),$$

so that $K_n = J_n + N_n$, where

$$J_n := \min\{j\colon \delta_{n,j} \le \delta_{n,i},\ nq_{-1} - N_n < j, i < nq_1 - N_n\}.$$

From the above limits (6), with extension to the case of $j \le 0$, it follows that, for each fixed j, $\delta_{n,j} \to \delta_j$ a.s., where

$$\delta_j := C_{-1}(S_j^r - S_{j+m}^r)[j \le -m] + (C_{-1}S_j^r + C_1 S_{j+m}^r)[-m < j \le 0] + C_1(S_{j+m}^r - S_j^r)[j \ge 1].$$

Note that in this and similar expressions we use the conventions $\infty \times 0 = 0$ and $\infty \times \pm a = \pm\infty$ for $a > 0$.

We now show that J_n converges to finite J, where J locates the minimal δ_j:

$$J := \min\{j: \delta_j \le \delta_i,\ j, i \text{ integer}\}.$$

LEMMA 5.1 (Decomposition). *Make the assumptions of this section.*

(a) *J is finite almost surely; if $C_1 = \infty$ ($C_{-1} = \infty$), then $J \le -m$ ($J \ge 1$);*
(b) *J_n converges to J almost surely.*

PROOF. (a) Define $D_j := |S^r_{j+m} - S^r_j|$. For $j > 0$, $D_j > r(S_j/j)^{r-1}W_j$, where $W_j := j^{r-1}\sum_{j+1}^{j+m} V_i \sim j^{r-1}S_m$. Since S_m has a gamma distribution, for any positive w, $P(W_j \le w) \le w^m/j^{m(r-1)}$, and by the Borel–Cantelli lemma, the strong law of large numbers and the choice of m such that $m(r-1) > 1$, we have $W_j \to \infty$, $D_j \to \infty$ and $\delta_j \to \infty$ a.s. as $j \to \infty$. A similar argument shows $\delta_j \to \infty$ a.s. as $j \to -\infty$. On the other hand, since $c_{-1} + c_1 > 0$, $\delta^* := \min(\delta_{-m}, \delta_1)$ is finite. Hence,

$$|J| \le J^* := \min\{t: \delta^* < \delta_j, |j| \ge t\} < \infty \quad \text{a.s.}$$

The last statement in (a) follows from the definitions of J and δ_j.

(b) From the definition of $F_{\pm 1}$ and the local observability hypothesis, there exists $\eta > 0$ for which

$$\eta_{-1} := (q - q_{-1})/q + \eta < F_{-1}(\varepsilon), \qquad \eta_1 := (q_1 - q)/p + \eta < F_1(\varepsilon).$$

Let $\lambda_{\pm 1}(k) := kq_{\pm 1} - N_k \pm m$, so that $\lambda_{\pm 1}(k)/k \to q_{\pm 1} - q$ a.s. By the strong law, $U_{k:\lambda_1(k)} \to (q_1 - q)/p$ a.s., with a similar result for the -1 case, and

$$\nu_1 := \min\{n \ge 1: \forall\, k \ge n, U_{k:\lambda_{-1}(k)} < \eta_{-1}, U_{k:\lambda_1(k)} < \eta_1\} < \infty \quad \text{a.s.}$$

Hence, inequalities of the type given in the equivalence lemma (3) hold for all local data when $n \ge \nu_1$. To give these explicitly, first choose $\varepsilon' = \min(\varepsilon_{-1}, \varepsilon_1)$, where $\varepsilon_{\pm 1}$ are so small that

$$a_{\pm 1}(\varepsilon_{\pm 1}) := \inf\{\phi(\pm t), 0 < t \le \varepsilon_{\pm 1}\} \ge c_{\pm 1}/2;$$

take $\varepsilon_{\pm 1} = \varepsilon$ if $c_{\pm 1} = 0$. Next, let $b := \sup\{\phi(t), -\varepsilon \le t \le \varepsilon\}$. Finally, let

$$\nu_2 := \min\{n \ge 1: \forall\, k \ge n, U_{k:-m} < F_{-1}(\varepsilon'), U_{k:m} < F_1(\varepsilon')\},$$
$$\nu_3 := \min\{n \ge 1: \forall\, k \ge n, 1/2 < S_{-N_k}/kq < 2, 1/2 < S_{M_k}/kp < 2\}.$$

By the strong law, $\nu := \nu_1 + \nu_2 + \nu_3 < \infty$ a.s. Let $B := 1/b^r$ and recall the variables δ^*, D_j introduced in the proof of part (a). For $n > \nu$,

$$\delta_n^* := \min(\delta_{n,-m}, \delta_{n,1}) \le 4^r\delta^*,$$
$$\delta_{n,j} \ge 2^{-r}BD_j \quad \text{if } nq_{-1} - N_n < j < nq_1 - N_n.$$

Hence

$$n > \nu \Rightarrow |J_n| \le J^{**} := \min\{t: 8^r\delta^* < BD_j, |j| \ge t\} < \infty \quad \text{a.s.}$$

Let $\mu := \min\{\delta_j - \delta_J : j \neq J, |j| \leq J^{**}\}$. Since the S_j's are continuous, $\mu > 0$ a.s. Let

$$\nu_4 := \min\{n > \nu : |\delta_{n,j} - \delta_j| < \mu/3, |j| \leq J^{**}\}.$$

Then $\nu_4 < \infty$ a.s., and $n > \nu_4$ implies $J_n = J$. □

6. Proof of Theorem 2.1. Clearly, using the results above, we have

$$n^r(\hat{\theta}_n - \theta) = n^r Y_{n:J_n} \to -C_{-1} S_J^r [J \leq 0] + C_1 S_J^r [J \geq 1] \quad \text{a.s.,}$$

and this limit is finite since $C_{-1} = \infty$ implies $J \geq 1$, while $C_1 = \infty$ implies $J \leq -m$. Next,

$$r_n^* := \ln n(\hat{r}_n - r) = -\ln \Delta_{n,K_n} - r \ln n = -\ln \delta_{n,J_n} \to r^* := -\ln \delta_J \quad \text{a.s.,}$$

and this last variable is finite since δ_J is positive and finite almost surely. Notice that these limiting distributions depend only on r and $c_{\pm 1}$. Finally, $\sqrt{n}(\hat{q}_n - q) = (J_n + N_n - nq)/\sqrt{n}$ certainly converges in law to a normal variable with mean 0 and variance $q(1-q)$.

When proving the results about $\hat{C}_{\pm 1, n}$, we restrict attention to the "+1" case and, from now on, drop that subscript from C and A. Since

$$\hat{C}_n := \hat{C}_{+1,n} = \exp(\beta_n r_n^*) n^{\beta_n r} A_n \quad \text{and} \quad (\ln n)^\tau(\exp(\beta_n r_n^*) - 1) \to r^* \quad \text{a.s.,}$$

it suffices to study the convergence of $\mu_n^r A_n$, where $\mu_n := n^{\beta_n} = n/\kappa_n$.

Introduce the notation $I_n := J_n + k_n$. From the convergence given in the first sentence of this section and the fact that $I_n > 0$ eventually a.s., we see that

$$(\ln n)^\tau \mu_n^r (A_n - B_n) \to 0 \quad \text{a.s.,} \quad \text{for } B_n := F_1^{-1}(U_n), U_n := U_{n:I_n}[I_n > 0].$$

Since $U_n \to 0$, $\mu_n U_n \to 1/p$ a.s., from the equivalence lemma (5) we have $\mu_n^r B_n \to C$ a.s., giving simple consistency.

Now assume $c := c_1 > 0$ and suppose there exist positive σ and ε_0 and finite M, so that $|\phi(t) - c| < M(-\ln t)^{-\sigma}$, $0 < t \leq \varepsilon_0$. Then

$$|\psi^r(t) - c^r| = O(|\psi(t) - c|) = O(-\ln t)^{-\sigma} \quad \text{as } t \to 0, \tag{7}$$

for $\psi = \phi^*$ or $\psi = \phi_*$. If we choose $0 < \tau < \sigma/(1 + \sigma)$, we claim

$$(\ln n)^\tau (\hat{C}_n - C) \to C r^* \quad \text{a.s.} \tag{8}$$

To prove this and thus complete the proof of the robust consistency theorem, by the above results it suffices to show

$$(\ln n)^\tau (\mu_n^r B_n - C) \to 0 \quad \text{a.s.}$$

To do this, introduce the notation $\eta_n := \mu_n^r B_n$, $\zeta := (1/p)^r$, $\zeta_n := (\mu_n U_n)^r - \zeta$, $\xi := \zeta/C$, $\xi_n := (\zeta_n + \zeta)/\eta_n - \xi$. Then

$$|\mu_n^r B_n - C| = |\eta_n - (\zeta/\xi)| = \left| \frac{\zeta_n + \zeta}{\xi_n + \xi} - \frac{\zeta}{\xi} \right| = O(|\zeta_n| + |\xi_n|).$$

Since the summands in S_n are exponentially distributed, it follows from the Marcinkiewicz strong law (see [6], page 126) that

$$\mu_n U_n - (1/p) = o\big(k_n^{-1/3}\big), \quad \text{thus } (\ln n)^{\tau}|\zeta_n| = o(1) \quad \text{a.s.}$$

To complete the proof of (8), apply part (4) of the equivalence lemma with $G = F_1$, $\gamma = \phi/p$, $\gamma_0 = c/p$ and $u = 0$. For small v, (4) implies $G^{-1}(v) \leq (2v/\gamma_0)^r$, and, because of the monotonicity of γ^* and γ_*, we can substitute this bound into the extreme members of (4) and still retain the inequality. Next, subtract $\xi = \gamma_0^r$ from all terms in this new expression and use (7) to see that, if $u_n \to 0$ and $\mu_n u_n \to 1/p$,

$$\left| \frac{(\mu_n u_n)^r}{\mu_n^r G^{-1}(u_n)} - \gamma_0^r \right| = O(-\ln u_n)^{-\sigma} = O(\ln \mu_n)^{-\sigma}$$

$$= O(\ln n)^{-\sigma(1-\tau)} = o(\ln n)^{-\tau},$$

since $\ln \mu_n = \beta_n \ln n = (\ln n)^{1-\tau}$ and $\tau - \sigma(1-\tau) < 0$. Thus,

$$(\ln n)^{\tau}|\xi_n| = o(1) \quad \text{a.s.} \qquad \square$$

Acknowledgments. We wish to thank Raoul LePage for reminding us of the representation of order statistics, a referee for suggesting stylistic changes and, most especially, an associate editor for pointing out the robustness possibilities inherent in our estimator.

REFERENCES

[1] Chernoff, H. and Rubin, H. (1956). The estimation of the location discontinuity in density. *Proc. Third Berkeley Symp. Statist. Probab.* **1** 19–37. Univ. California Press, Berkeley.
[2] Federer, H. (1969). *Geometric Measure Theory*. Springer, New York.
[3] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd ed. Wiley, New York.
[4] Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568.
[5] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation—Asymptotic Theory* (translated by Samuel Kotz). Springer, New York.
[6] Stout, W. (1974). *Almost Sure Convergence*. Academic Press, New York.

Department of Statistics and Probability
Michigan State University
East Lansing, Michigan 48824