The Annals of Statistics
1975, Vol. 3, No. 3, 698–706

CONSISTENCY IN NONPARAMETRIC ESTIMATION OF THE MODE[1]

BY THOMAS W. SAGER

The University of Iowa

Let X be an absolutely continuous real-valued random variable with additional restrictions to be imposed later. Venter (1967) ("On estimation of the mode," *Ann. Math. Statist.* **37** 1446–1455) estimated the mode of X by a point from the shortest interval containing a specified number $r = r(n)$ of observations. Venter demonstrated that such an estimator is strongly consistent under appropriate conditions on the distribution of X and on $r(n)$. It is the purpose of this paper to show that strong consistency actually holds under very general conditions on the distribution of X. Convergence rates are also obtained which are, in some cases, much faster than those reported by Venter.

1. Introduction. Venter (1967) considered estimating the mode of a univariate distribution by the midpoint (or either endpoint) of the shortest interval containing $r = r(n)$ observations. Under appropriate conditions on the distribution and on $r(n)$, Venter obtained almost sure convergence of such an estimator. It is our purpose to obtain more general conditions which still guarantee almost sure convergence. We shall retain Venter's assumptions on $r(n)$ but relax the conditions imposed on the distribution.

In our view, the mode is a characteristic of the distribution, viz., the location of the greatest concentration of probability. So we need to specify carefully which version of the density function is used to define the mode.

DEFINITION 1.1. Let $F(x)$ be an absolutely continuous distribution function on the real line. Define the density by $f(x) = \max\{(DF)^+(x), (DF)^-(x)\}$, $f(-\infty) = f(\infty) = 0$, where $(DF)^+(x)$ is the upper derivate on the right of F at x and $(DF)^-(x)$ is the upper derivate on the left of F at x.

We first consider the slightly more general problem of estimating the location of "local modes" in a closed interval.

DEFINITION 1.2. The subset M of $[c, d]$ is called the modal set of F on $[c, d]$ if

(a) f is constant on M, and
(b) $f(M) > f(x)$ for each $x \in [c, d] - M$, and
(c) for each open set U containing M, there exists $\varepsilon = \varepsilon(U) > 0$ such that $f(x) + \varepsilon \leqq f(M)$ for each $x \in [c, d] - U$.

Received October 1973; revised June 1974.

[1] This research forms part of the author's doctoral dissertation. It was sponsored in part by the National Science Foundation under Grant GP-30966X.

AMS 1970 *subject classifications*. Primary 62G05; Secondary 60F15.

Key words and phrases. Mode, consistency, convergence rates, estimation.

DEFINITION 1.3. We say that an absolutely continuous distribution function F satisfies the standard conditions on $[c, d] \subset [-\infty, \infty]$, $c < d$, if there is a nonempty modal set M in $[c, d]$ such that for some $\theta \in M$, either

$$f(\theta) = (DF)^+(\theta) = (DF)_+(\theta)\,, \qquad \theta < d\,, \tag{1.1}$$

or

$$f(\theta) = (DF)^-(\theta) = (DF)_-(\theta)\,, \qquad \theta > c \tag{1.2}$$

where $(DF)_+(\theta)$ is the lower derivate on the right of F at θ and $(DF)_-(\theta)$ is the lower derivate on the left of F at θ.

We shall henceforth assume that F is an absolutely continuous distribution function which satisfies the standard conditions on $[c, d]$. No other assumptions on the distribution will be required.

Let $Y_1, \cdots, Y_n$ denote the order statistics corresponding to a random sample of size n from F. Let $\{r(n)\}$ be a nonrandom sequence of integers to be specified further. For each n, let $K(n)$ be a discrete random variable defined by the following: If there are at least $r(n) + 1$ observations in $[c, d]$, let

$$\begin{aligned} &Y_{K(n)+r(n)} - Y_{K(n)} \\ &\quad = \min\{Y_{j+r(n)} - Y_j;\ j = 1, \cdots, n - r(n),\ c \leqq Y_j \leqq Y_{j+r(n)} \leqq d\}\,. \end{aligned} \tag{1.3}$$

If $[c, d]$ contains fewer than $r(n) + 1$ observations, let $K(n) = 1$ (for consistency $K(n)$ can be chosen arbitrarily in this case).

We note that if $F(d) - F(c) > 0$ and $\{r(n)\}$ is chosen so that $n^{-1}r(n) \to 0$, then the strong law of large numbers guarantees that $[c, d]$ will eventually contain $r(n) + 1$ observations with probability one. Further, since F is absolutely continuous, then $K(n)$ is unique and $Y_{K(n)+r(n)} - Y_{K(n)} > 0$ with probability one.

Choose the estimator $\theta(n)$ so that $Y_{K(n)} \leqq \theta(n) \leqq Y_{K(n)+r(n)}$.

2. Consistency.

THEOREM 2.1. *Let $F(x)$ be an absolutely continuous distribution function which satisfies the standard conditions on $[c, d] \subset [-\infty, \infty]$, $c < d$, with associated modal set M. Suppose*

$$n^{-1}r(n) \to 0 \quad \textit{as} \quad n \to \infty\,, \tag{2.1}$$

$$\textstyle\sum_{n=1}^{\infty} n\lambda^{r(n)} < \infty \quad \textit{for all} \quad \lambda\,, \qquad 0 < \lambda < 1\,; \tag{2.2}$$

then $\inf M \leqq \liminf_{n\to\infty} \theta(n) \leqq \limsup_{n\to\infty} \theta(n) \leqq \sup M$ *with probability one.*

PROOF. We give the proof when (1.1) holds. The proof for (1.2) is similar.

Choose and fix $\theta \in M$ which satisfies (1.1). For each n, let $J(n)$ be a discrete random variable defined by the following: If $[\theta, d]$ contains at least $r(n) + 1$ observations, let

$$Y_{J(n)+r(n)} = \min\{Y_{j+r(n)};\ j = 1, \cdots, n - r(n);\ \theta \leqq Y_j \leqq Y_{j+r(n)} \leqq d\}\,. \tag{2.3}$$

If $[\theta, d]$ contains fewer than $r(n) + 1$ observations, let $J(n) = 1$.

Consider the following events:

$$
\begin{aligned}
&\Omega_0 = [\lim_{n\to\infty} Y_{J(n)+r(n)} = \theta,\ \lim_{n\to\infty} Y_{J(n)} = \theta]\,,\\
&\Omega_1 = [\lim_{n\to\infty} \{F(Y_{J(n)+r(n)}) - F(Y_{J(n)})\}nr(n)^{-1} = 1]\,,\\
(2.4)\quad &\Omega_2 = [\lim_{n\to\infty} \{F(Y_{K(n)+r(n)}) - F(Y_{K(n)})\}nr(n)^{-1} = 1]\,,\\
&\Omega_3 = [\liminf_{n\to\infty} \{F(Y_{J(n)+r(n)}) - F(Y_{J(n)})\}/(Y_{J(n)+r(n)} - Y_{J(n)}) = f(\theta)]\,,\\
&\Omega_4 = [\liminf_{n\to\infty} \{F(Y_{K(n)+r(n)}) - F(Y_{K(n)})\}/(Y_{K(n)+r(n)} - Y_{K(n)}) = f(\theta)]\,,\\
&\Omega_5 = [\inf M \leqq \liminf_{n\to\infty} \theta(n) \leqq \limsup_{n\to\infty} \theta(n) \leqq \sup M]\,.
\end{aligned}
$$

The method of proof will be to show that $\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \subset \Omega_0 \cap \Omega_4 \subset \Omega_0 \cap \Omega_5$ and that $P(\Omega_0) = P(\Omega_1) = P(\Omega_2) = P(\Omega_3) = 1$.

We first establish the containment relationships. Let $\omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3$. Using (2.4) and the fact that (eventually) $0 < Y_{K(n)+r(n)} - Y_{K(n)} \leqq Y_{J(n)+r(n)} - Y_{J(n)}$, we have

$$
\begin{aligned}
&\frac{\liminf\,[F(Y_{J(n)+r(n)}) - F(Y_{J(n)})]/(Y_{J(n)+r(n)} - Y_{J(n)})}{\liminf\,[F(Y_{K(n)+r(n)}) - F(Y_{K(n)})]/(Y_{K(n)+r(n)} - Y_{K(n)})}\\
&\qquad \leqq \limsup \frac{Y_{K(n)+r(n)} - Y_{K(n)}}{Y_{J(n)+r(n)} - Y_{J(n)}}\\
&\qquad\qquad \times \frac{\limsup\,[F(Y_{J(n)+r(n)}) - F(Y_{J(n)})]n \cdot r(n)^{-1}}{\liminf\,[F(Y_{K(n)+r(n)}) - F(Y_{K(n)})]n \cdot r(n)^{-1}} \leqq 1\,.
\end{aligned}
$$

This implies

$$
\begin{aligned}
f(\theta) &\leqq \liminf \frac{F(Y_{K(n)+r(n)}) - F(Y_{K(n)})}{Y_{K(n)+r(n)} - Y_{K(n)}}\\
&\leqq \liminf \frac{\int_{[Y_{K(n)}, Y_{K(n)+r(n)}]} f(\theta)\,dx}{Y_{K(n)+r(n)} - Y_{K(n)}} = f(\theta)\,.
\end{aligned}
$$

Thus $\Omega_1 \cap \Omega_2 \cap \Omega_3 \subset \Omega_4$.

To show $\Omega_0 \cap \Omega_4 \subset \Omega_0 \cap \Omega_5$, it suffices to show $\Omega_0 \cap \Omega_5^c \subset \Omega_0 \cap \Omega_4^c$. Let $\omega \in \Omega_0 \cap \Omega_5^c$. Thus there is a subsequence $\{n(j)\}$ such that $\theta(n(j))$ lies outside of $(\inf M - 2\delta, \sup M + 2\delta)$ for all j, for some $\delta > 0$. Since $\omega \in \Omega_0$, then $Y_{K(n)+r(n)} - Y_{K(n)}$ converges to zero. So $[Y_{K(n(j))}, Y_{K(n(j))+r(n(j))}]$ lies outside of $(\inf M - \delta, \sup M + \delta)$ for all large j. By Definition 1.2(c) this implies

$$
\frac{F(Y_{K(n(j))+r(n(j))}) - F(Y_{K(n(j))})}{Y_{K(n(j))+r(n(j))} - Y_{K(n(j))}} \leqq f(\theta) - \varepsilon
$$

for all large j, for some $\varepsilon > 0$. But this implies $\omega \in \Omega_4^c$.

We now establish the probability statements.

Since $(DF)^+(\theta) > 0$, then F assigns positive probability to every interval $[\theta, \theta + \varepsilon]$, $\varepsilon > 0$. So by (2.1) and the strong law of large numbers, $Y_{J(n)+r(n)}$ and $Y_{J(n)}$ converge almost surely to θ.

It remains to show that $P(\Omega_1) = P(\Omega_2) = P(\Omega_3) = 1$. To this end, we need the following lemma.

LEMMA 2.1. *Let $S_1, S_2, \cdots; T_1, T_2, \cdots$ be sequences of random variables such that $S_n \leqq T_n$ for each n and $[S_n, T_n]$ contains $r(n) + 1$ observations. Then $\{F(T_n) - F(S_n)\}\{F_n(T_n) - F_n(S_n)\}^{-1}$ converges almost surely to one, where F_n denotes the empirical distribution function.*

PROOF. Let $0 < \varepsilon < \frac{1}{2}$. If Z is a $\Gamma(m, 1)$ variable, then the well-known inequalities $P[Z > m(1+\varepsilon)] < E(e^{tZ})e^{-tm(1+\varepsilon)}$ and $P[Z < m(1-\varepsilon)] < E(e^{-tZ})e^{tm(1-\varepsilon)}$ with $t = \varepsilon/(1+\varepsilon)$ and $t = \varepsilon/(1-\varepsilon)$, respectively, yield

$$P[Z > m(1+\varepsilon)] < \{(1+\varepsilon)e^{-\varepsilon}\}^m < \exp(-\tfrac{1}{4}m\varepsilon^2)\,, \tag{2.5}$$

and

$$P[Z < m(1-\varepsilon)] < \{(1-\varepsilon)e^{\varepsilon}\}^m < \exp(-\tfrac{1}{4}m\varepsilon^2)\,.$$

Thus, if $Z_1, \cdots, Z_n$ are $\Gamma(m, 1)$ variables, (2.5) implies

$$P[\max\{Z_1, \cdots, Z_n\} > m(1+\varepsilon)] < n\exp(-\tfrac{1}{4}m\varepsilon^2)\,, \tag{2.6}$$

and

$$P[\min\{Z_1, \cdots, Z_n\} < m(1-\varepsilon)] < n\exp(-\tfrac{1}{4}m\varepsilon^2)\,.$$

If, additionally, Z is $\Gamma(j, 1)$ then the inequalities $1 - 4\varepsilon < (1-\varepsilon)/(1+\varepsilon)$ and $(1+\varepsilon)/(1-\varepsilon) < 1 + 4\varepsilon$ together with (2.5) and (2.6) imply

$$\begin{aligned} P\Big[\max\Big\{\frac{Z_1}{Z}, \cdots, \frac{Z_n}{Z}\Big\} &> \frac{m(1+4\varepsilon)}{j}\,, \text{ or} \\ \min\Big\{\frac{Z_1}{Z}, \cdots, \frac{Z_n}{Z}\Big\} &< \frac{m(1+4\varepsilon)}{j}\Big] \\ &< 2n\exp(-\tfrac{1}{4}m\varepsilon^2) + 2\exp(-\tfrac{1}{4}j\varepsilon^2)\,. \end{aligned} \tag{2.7}$$

Now if $Y_1, \cdots, Y_n$ are the order statistics from a continuous distribution, it is well known that the coverages $F(Y_{i+r(n)}) - F(Y_i)$, $i = 0, 1, \cdots, n+1-r(n)$ (with $Y_0 = -\infty$, $Y_{n+1} = \infty$) can be represented as ratios W_i/W where W_i is $\Gamma(r(n), 1)$ and W is $\Gamma(n+1, 1)$. Thus, if $I(n)$ is a random variable assuming values in $\{0, 1, \cdots, n+1-r(n)\}$, we have from (2.7) that

$$\begin{aligned} 1 - P\Big[\Big|\{F(Y_{I(n)+r(n)}) - F(Y_{I(n)})\}\frac{n+1}{r(n)} - 1\Big| < 4\varepsilon\Big] \\ < 2(n+1)\exp(-\tfrac{1}{4}r(n)\varepsilon^2) + 2\exp(-\tfrac{1}{4}(n+1)\varepsilon^2)\,. \end{aligned} \tag{2.8}$$

Thus $\{F(Y_{I(n)+r(n)}) - F(Y_{I(n)})\}(n+1)r(n)^{-1}$ converges almost surely to one by the Borel–Cantelli lemma, since $\sum_{n=1}^{\infty}\{2(n+1)\exp(-\frac{1}{4}r(n)\varepsilon^2) + 2\exp(-\frac{1}{4}(n+1)\varepsilon^2)\} < \infty$ by setting $\lambda = \exp(-\frac{1}{4}\varepsilon^2)$ in (2.2).

Now specialize $I(n)$ so that $Y_{I(n)-1} < S_n \leqq Y_{I(n)} \leqq Y_{I(n)+r(n)} \leqq T_n < Y_{I(n)+r(n)+1}$. We have

$$\begin{aligned} \frac{F(Y_{I(n)+r(n)}) - F(Y_{I(n)})}{(r(n)+1)/n} &\leqq \frac{F(T_n) - F(S_n)}{F_n(T_n) - F_n(S_n)} \\ &\leqq \frac{F(Y_{I(n)+r(n)+1}) - F(Y_{I(n)-1})}{r(n)/n}\,. \end{aligned}$$

Thus $\{F(T_n) - F(S_n)\}\{F_n(T_n) - F_n(S_n)\}^{-1}$ converges almost surely to one.

From this lemma we immediately have $P(\Omega_1)=P(\Omega_2)=1$. To see that $P(\Omega_3)=1$, let $U_n = (Y_{J(n)+r(n)} - \theta)/[F(Y_{J(n)+r(n)}) - F(\theta)]$ and $V_n = (Y_{J(n)} - \theta)/[F(Y_{J(n)}) - F(\theta)]$ and write

$$\frac{Y_{J(n)+r(n)} - Y_{J(n)}}{F(Y_{J(n)+r(n)}) - F(Y_{J(n)})} = U_n \frac{[F(Y_{J(n)+r(n)}) - F(\theta)]nr(n)^{-1}}{[F(Y_{J(n)+r(n)}) - F(Y_{J(n)})]nr(n)^{-1}} + V_n - V_n \frac{[F(Y_{J(n)+r(n)}) - F(\theta)]nr(n)^{-1}}{[F(Y_{J(n)+r(n)}) - F(Y_{J(n)})]nr(n)^{-1}} . \tag{2.9}$$

Since $P(\Omega_0) = 1$ and in view of (1.1), we have $U_n \to f(\theta)^{-1}$ with probability one and $V_n \to f(\theta)^{-1}$ with probability one. Thus by setting $S_n = \theta$ and $T_n = Y_{J(n)+r(n)}$ in Lemma 2.1 and using $P(\Omega_1) = 1$, we conclude that (2.9) converges to $f(\theta)^{-1}$ with probability one. Thus $P(\Omega_3) = 1$. The proof of the theorem is complete.

COROLLARY 2.1. *If the mode is unique*, i.e., $M = \{\theta\}$, *then* $\theta(n) \to \theta$ *with probability one.*

It is worthwhile to note some of the gains embodied in Theorem 2.1. Venter (1967) imposed the following conditions on the distribution:

(a) The density $f(x)$ is continuous and strictly positive on its support, which is an interval (a, b) containing the mode; and

(b) $\alpha(\delta, 2, 2) > 1$ for all small positive δ (see Definition 3.1 of this paper).

Theorem 2.1 does not require continuity of the density; the distribution function may even be nondifferentiable on a set of Lebesgue measure zero (possibly including the mode). Other distributions included in the purview of Theorem 2.1 are those with gaps in their supports and those whose modes occur at the boundary of their support (e.g., the exponential: $f(x) = \exp(-x + \theta)$, $x > \theta$). Condition (b) is unnecessary for consistency; thus many highly asymmetric densities are also included (e.g., (3.1) of this paper).

Finally, we remark that Corollary 2.1 implies Theorem 1 of Venter (1967).

3. Rates of convergence. In this section, we retain all of the background and assumptions of Section 2, except that we suppose the modal set of F on $[c, d]$ is a singleton, $M = \{\theta\}$. Additionally, we impose a condition similar to (4.1) of Venter (1967).

DEFINITION 3.1. For $c \leqq \theta - R_1\delta < \theta - \delta < \theta$ and/or $\theta < \theta + \delta < \theta + R_2\delta \leqq d$, define $\alpha(\delta, R_1, R_2) = \min\{r^-(\delta), 1^-(\delta)\}/\max\{r^+(R_2\delta), 1^+(R_1\delta)\}$, where

$$r^-(\delta) = \inf\{f(x); \theta \leqq x \leqq \theta + \delta\},$$
$$r^+(R_2\delta) = \sup\{f(x); \theta + R_2\delta \leqq x \leqq d\},$$
$$1^-(\delta) = \inf\{f(x); \theta - \delta \leqq x \leqq \theta\},$$
$$1^+(R_1\delta) = \sup\{f(x); c \leqq x \leqq \theta - R_1\delta\}.$$

Also let $r(\delta, R_2\delta) = r^-(\delta)/r^+(R_2\delta)$ and $1(\delta, R_1\delta) = 1^-(\delta)/1^+(R_1\delta)$.

The assumption to be employed in this section will be that $\alpha \geqq 1 + \rho\delta^k$ for all small δ and for some $R_1 > 1$, $R_2 > 1$, $\rho > 0$, $k > 0$. We remark that Venter (1967) employed the same assumption about α but with $R_1 = R_2 = 2$. As we have seen in Section 2, no such assumption is necessary for consistency. Additionally, such an assumption with fixed R_1 and R_2 rules out of consideration many highly asymmetric densities. For instance, consider the densities

$$
\begin{aligned}
&\text{(a)}\quad f(x) = \tfrac{1}{2}x + \tfrac{1}{2}, && -1 \leqq x \leqq 0 \\
&\qquad\quad = -\tfrac{1}{6}x + \tfrac{1}{2}, && 0 < x \leqq 3 \\
(3.1)\quad &\qquad\quad = c, && x = 0 \\
&\text{(b)}\quad f(x) = cn/(n+1), && (2n+1)/2n(n+1) \leqq |x| < 1/n \\
&\qquad\quad = cn/(n+S_n), && 1/(n+1) \leqq |x| < (2n+1)/2n(n+1)
\end{aligned}
$$

where in (b) $\{S_n; n = 1, 2, \cdots\}$ is a sequence of positive integers with S_n tending to infinity and S_n/n tending to 0 and c is determined to make $f(x)$ a density on $[-1, 1]$.

For (a) we compute $\alpha(\delta, 2, 2) = (1-\delta)/(1-2\delta/3) < 1$ for all small δ. However, $\alpha(\delta, 2, 6) = (1-\delta)/(1-2\delta) > 1 + \delta$ for all small δ. Example (b) is symmetric about its mode $x = 0$, and therefore $\alpha(\delta, R, R) = r(\delta, R\delta)$ for any $R > 1$. Now $r^-(\delta) = cm/(m + S_m)$ where $1/(m+1) \leqq \delta < 1/m$. Hence $R/(m+1) \leqq R\delta < R/m$ so that $r^+(R\delta) \geqq cK(K+1)$ with K approximately m/R. For all small δ, we have $S_m > R$ so that $r^-(\delta) < r^+(R\delta)$. Hence $r(\delta, R\delta) < 1$ for all small δ. Thus example (b) fails to satisfy the hypotheses of Theorem 3.1 for any R_1 and R_2. However, it is easy to verify that (b) satisfies the standard conditions. So Theorem 2.1 gives consistency in (b), but we cannot say anything about a convergence rate.

THEOREM 3.1. *Let $F(x)$ be an absolutely continuous distribution function which satisfies the standard conditions on $[c, d] \subset [-\infty, \infty]$, $c < d$, with modal set M in $[c, d]$. Let $r(n)$ be of the form $An^{2k/(1+2k)}$ for some $A > 0$ and let $\delta(n) = n^{-1/(1+2k)}(\log n)^{1/k}$ (for k specified below). Let $\theta(n)$ be as before.*

(1) *If $M = \{c\}$ and there are positive constants $R > 1$, ρ, k such that $r(\delta, R\delta) \geqq 1 + \rho\delta^k$ for all small positive δ, then $\theta(n) = c + o(\delta(n))$ with probability one.*

(2) *If $M = \{d\}$ and there are positive constants $R > 1$, ρ, k such that $1(\delta, R\delta) \geqq 1 + \rho\delta^k$ for all small positive δ, then $\theta(n) = d + o(\delta(n))$ with probability one.*

(3) *If $M = \{\theta\}$ with $c < \theta < d$ and there are positive constants $R_1 > 1$, $R_2 > 1$, ρ, k such that $\alpha(\delta, R_1, R_2) \geqq 1 + \rho\delta^k$ for all small positive δ, then $\theta(n) = \theta + o(\delta(n))$ with probability one.*

PROOF. First we need a convergence rate in Lemma 2.1 when $r(n)$ has the above form.

LEMMA 3.1. *Let $S_1, S_2, \cdots; T_1, T_2, \cdots$ be a sequence of random variables such that $S_n \leqq T_n$ for each n and $[S_n, T_n]$ contains $r(n) + 1$ observations where $r(n)$ is of the form An^ν, $0 < \nu < 1$. Then $\{F(T_n) - F(S_n)\}\{F_n(T_n) - F_n(S_n)\}^{-1} = 1 + o(r(n)^{-\frac{1}{2}} \log r(n))$ with probability one.*

PROOF. In (2.8) replace ε with $\varepsilon r(n)^{-\frac{1}{2}} \log r(n)$ to obtain

$$(3.2) \qquad 1 - P\left[\left|\{F(Y_{I(n)+r(n)}) - F(Y_{I(n)})\}\frac{n+1}{r(n)} - 1\right| < 4\varepsilon r(n)^{-\frac{1}{2}} \log r(n)\right]$$
$$< 2(n+1)\exp(-\tfrac{1}{4}(\log r(n))^2\varepsilon^2)$$
$$+ 2\exp(-\tfrac{1}{4}(n+1)r(n)^{-1}(\log r(n))^2\varepsilon^2) .$$

Since the series

$$(3.3) \qquad \sum_{n=1}^{\infty} \{2(n+1)(An^\nu)^{-\frac{1}{4}\varepsilon^2\nu\log(nA^{1/\nu})} + 2(An^\nu)^{-\frac{1}{4}\varepsilon^2(n+1)(An^\nu)^{-1}\log(An^\nu)}\}$$

is finite for $0 < \nu < 1$, then the Borel–Cantelli lemma implies that $\{F(Y_{I(n)+r(n)}) - F(Y_{I(n)})\}nr(n)^{-1} = 1 + o(r(n)^{-\frac{1}{2}} \log r(n))$ with probability one. Thus by specializing $I(n)$ as in the remarks following (2.8) we obtain the desired conclusion.

Now suppose that the hypothesis of part (1) of the theorem holds. Define $J(n)$ by (2.3) (with $\theta = c$).

LEMMA 3.2. *$Y_{J(n)+r(n)} = c + o(\delta(n))$ with probability one.*

PROOF. First, we assert that $r^+(R\delta) \to f(c)$ as $\delta \to 0$. If not, then from the definition of $r^+(R\delta)$ we have that $f(x) + \varepsilon < f(c)$ for some $\varepsilon > 0$ and for all x, $c < x < d$. But this cannot be, for it implies $(DF)^+(c) + \varepsilon < f(c)$.

Hence, since $r(\delta, R\delta) \geqq 1 + \rho\delta^k$, there is a positive constant B such that $B < r^+(R\delta) < r^-(\delta) \leqq r^-(\varepsilon\delta)$ for all small δ and for each ε, $0 < \varepsilon < 1$. Thus $F(c + \varepsilon\delta(n)) - F(c) \geqq \int_{[c,c+\varepsilon\delta(n)]} r^-(\varepsilon\delta(n))\,dx > B\varepsilon\delta(n)$ for $0 < \varepsilon < 1$ and for all large n. Therefore,

$$(3.4) \qquad \liminf_{n\to\infty} \{F(c + \varepsilon\delta(n)) - F(c)\}/(B\varepsilon\delta(n)) \geqq 1 .$$

By Lemma 2.1, we have

$$(3.5) \qquad \{F(Y_{J(n)+r(n)}) - F(c)\}nr(n)^{-1} \to 1 \qquad \text{with probability one.}$$

But $n^{-1}r(n)/(B\varepsilon\delta(n)) \to 0$ as $n \to \infty$; so by (3.5) we have

$$(3.6) \qquad \{F(Y_{J(n)+r(n)}) - F(c)\}/(B\varepsilon\delta(n)) \to 0 \qquad \text{with probability one.}$$

From (3.4) and (3.6) we deduce that $F(c + \varepsilon\delta(n)) > F(Y_{J(n)+r(n)})$ for all large n with probability one. And since ε is arbitrary, this implies $Y_{J(n)+r(n)} = c + o(\delta(n))$ with probability one.

LEMMA 3.3. *$Y_{K(n)} = c + o(\delta(n))$ with probability one.*

PROOF. Let Ω_0 be the event where Lemma 3.4 holds. Let Ω_1 be the event where Lemma 3.1 holds with $S_n = Y_{K(n)}$ and $T_n = Y_{K(n)+r(n)}$. Let Ω_2 be the event where Lemma 3.1 holds with $S_n = Y_{J(n)}$ and $T_n = Y_{J(n)+r(n)}$. Let Ω_3 be the event where Lemma 3.2 holds.

To prove Lemma 3.3, it suffices to show that $\Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_0^c = \emptyset$. Suppose $\omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_0^c$. Since $\omega \in \Omega_0^c$, then there is an ε and a subsequence $\{n(j)\}$ such that

$$(3.7) \qquad Y_{K(n(j))} > c + R\varepsilon\delta(n(j)) \qquad \text{for all } j .$$

Using the hypothesis of (1), Lemma 3.2, (3.7), and Lemma 3.1, we have

$$
\begin{aligned}
1 + \rho(\varepsilon\delta(n(j)))^k &\leqq r(\varepsilon\delta(n(j)), R\varepsilon\delta(n(j))) \\
&\leqq \frac{r^-(\varepsilon\delta(n(j)))}{r^+(R\varepsilon\delta(n(j)))} \cdot \frac{Y_{J(n(j))+r(n(j))} - Y_{J(n(j))}}{Y_{K(n(j))+r(n(j))} - Y_{K(n(j))}} \\
&\leqq \frac{[F(Y_{J(n(j))+r(n(j))}) - F(Y_{J(n(j))})]n(j)r(n(j))^{-1}}{[F(Y_{K(n(j))+r(n(j))}) - F(Y_{K(n(j))})]n(j)r(n(j))^{-1}} \qquad (3.8) \\
&= 1 + o(r(n(j))^{-\frac{1}{2}} \log r(n(j))) \\
&= 1 + o(n(j)^{-k/(1+2k)} \log n(j)) \, .
\end{aligned}
$$

However, $1 + \rho(\varepsilon\delta(n(j)))^k = 1 + \rho\varepsilon^k n(j)^{-k/(1+2k)} \log n(j)$, which contradicts the above inequality for large j.

Part (1) of the theorem now follows easily since $\theta(n) - c \leqq Y_{K(n)+r(n)} - c \leqq Y_{J(n)+r(n)} - Y_{J(n)} + Y_{K(n)} - c = o(\delta(n))$ with probability one by Lemmas 3.2 and 3.3. The proof of part (2) is similar to that of part (1).

Suppose the hypothesis of part (3) holds. Since $\alpha(\delta, R_1, R_2) \leqq \min\{r(\delta, R_2\delta), 1(\delta, R_1\delta)\}$, then $r(\delta, R_2\delta) \geqq 1 + \rho\delta^k$ and $1(\delta, R_1\delta) \geqq 1 + \rho\delta^k$ for all small δ. Thus parts (1) and (2) of the theorem are applicable.

Define $I(n)$ and $J(n)$ by the following:
If $[c, \theta]$ contains at least $r(n) + 1$ observations, let

$$
\begin{aligned}
(3.9) \qquad & Y_{I(n)+r(n)} - Y_{I(n)} \\
&= \min\{Y_{j+r(n)} - Y_j; j = 1, \cdots, n - r(n), c \leqq Y_j \leqq Y_{j+r(n)} \leqq \theta\} \, .
\end{aligned}
$$

If $[c, \theta]$ contains fewer than $r(n) + 1$ observations, let $I(n) = 1$.
If $[\theta, d]$ contains at least $r(n) + 1$ observations, let

$$
\begin{aligned}
(3.10) \qquad & Y_{J(n)+r(n)} - Y_{J(n)} \\
&= \min\{Y_{j+r(n)} - Y_j; j = 1, \cdots, n - r(n), \theta \leqq Y_j \leqq Y_{j+r(n)} \leqq d\} \, .
\end{aligned}
$$

If $[\theta, d]$ contains fewer than $r(n) + 1$ observations, let $J(n) = 1$.

By parts (1) and (2) of the theorem, we have

$$
(3.11) \qquad Y_{J(n)+r(n)} = \theta + o(\delta(n)) \qquad \text{with probability one,}
$$

and

$$
Y_{I(n)} = \theta + o(\delta(n)) \qquad \text{with probability one.}
$$

Now if $[Y_{K(n)}, Y_{K(n)+r(n)}] \subset [c, \theta]$, then $Y_{K(n)} = Y_{I(n)}$ and $Y_{K(n)+r(n)} = Y_{I(n)+r(n)}$; if $[Y_{K(n)}, Y_{K(n)+r(n)}] \subset [\theta, d]$, then $Y_{K(n)} = Y_{J(n)}$ and $Y_{K(n)+r(n)} = Y_{J(n)+r(n)}$; if $Y_{K(n)} \leqq \theta \leqq Y_{K(n)+r(n)}$, then $Y_{I(n)} \leqq Y_{K(n)} = Y_{K(n)+r(n)} \leqq Y_{J(n)+r(n)}$. Consideration of these cases together with (3.11) shows that $Y_{K(n)} = \theta + o(\delta(n))$ with probability one and $Y_{K(n)+r(n)} = \theta + o(\delta(n))$ with probability one. The proof of the theorem is complete.

REMARK. For $k < \frac{1}{2}$, the convergence rates given above are a substantial improvement over those reported by Venter (1967). As k tends to zero, with $r(n)$ of the form $An^{2k/(1+2k)}$, the convergence becomes quite rapid indeed.

4. Acknowledgments. I wish to thank Drs. Tim Robertson and F. T. Wright for their guidance during the writing of my dissertation which led to this paper. I thank the referee for pointing out Example 3.1(b).

REFERENCE

[1] VENTER, J. H. (1967). On estimation of the mode. *Ann. Math. Statist.* **37** 1446–1455.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305