The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1665–1671

ON CHOOSING A DELTA-SEQUENCE

BY MICHAEL WOODROOFE
University of Michigan

1. Introduction. We will be concerned with estimates of the density f from which a random sample $X_1, \cdots, X_n$ has been drawn. In particular, we will consider some modifications of the following type of estimate:

$$f_n(x; t_n) = t_n \int K(t_n(x-y))\, dF_n(y) = (1/n) \sum_{i=1}^{n} t_n K(t_n(x-X_i)). \tag{1.1}$$

Here K is a real-valued, bounded, symmetric, absolutely integrable function on $\boldsymbol{R}^1$ for which

$$\int K(y)\, dy = 1; \tag{1.2}$$

t_n is an increasing sequence of positive real numbers for which $t_n \to \infty$ with $t_n = o(n)$ as $n \to \infty$; and F_n denotes the sample distribution function of $X_1, \cdots, X_n$. Such estimates were originally proposed by Rosenblatt [3] and were studied in some detail by Parzen [2].

It is known ([1] and [2]) that the asymptotic behavior of (1.1) depends on the smoothness of f near x and on the sequence t_n. Moreover, the optimal choice of t_n in the sense of minimizing the asymptotic expression for mean square error also depends on the smoothness of f near x and is therefore unknown to the statistician. Here we will consider some modifications of (1.1) which may be described as follows: first estimate f and its derivatives using (1.1) with a t_n sequence as described above; next, use these initial estimates to estimate the optimal t_n sequence, $t_n = \tau_n = \tau_n(f, x, K)$ say, by $\hat{\tau}_n = \hat{\tau}_n(x, K, X_1, \cdots, X_n)$ say; and finally, estimate f by (1.1) with $\hat{\tau}_n$ replacing t_n. Two such modifications are considered; and in both cases we are able to show that under the appropriate regularity conditions

$$E[f_n(x; \hat{\tau}_n) - f(x)]^2 \sim E[f_n(x; \tau_n) - f(x)]^2 \tag{1.3}$$

as $n \to \infty$ where $\sim$ means that the ratio of the two sides tends to one. As may be expected, the proofs of (1.3) constitute rather involved exercises in large sample theory. In order to shorten them, we have developed some special methods and notation which we hope will be of methodological interest in its own right. Briefly, we have developed an algebra of o_E and O_E for handling mean convergence of random variables. This algebra is analogous to the algebra of o_p and O_p.

The paper consists of five sections. In Section two we collect some facts about sample densities of the form (1.1) and state precisely the effect of the smoothness of f on their asymptotic behavior; in Section three we present the algebra of o_E and O_E; and in Sections four and five we present the main theorems.

Received October 3, 1969.

2. Preliminaries. We will call a real-valued function g defined on an open interval I smooth of order α, $\alpha > 0$, at $x \in I$ iff g has m continuous derivatives on an open interval J with $x \in J \subset I$ and

$$|y|^{-\alpha}[g(x+y) - \sum_{i=0}^{m} g^{(i)}(x) y^i / i!] \to g_\alpha^{+}(x) \colon y \to 0^+$$
$$\to g_\alpha^{-}(x) \colon y \to 0^-$$

where m is the greatest integer strictly less than α and $|g_\alpha^{+}(x)| + |g_\alpha^{-}(x)| < \infty$. Thus, if g is smooth of order $\alpha > 0$ at x, then g is smooth of all orders β, $0 < \beta < \alpha$, at x, but not conversely. For example, the function g defined by $g(x) = -|x| \log |x|$, $x \in R^1$, is smooth of all orders less than one at zero but is not smooth of order one there. A function with $p \geqq 1$ continuous derivatives near a point x is, of course, smooth of all orders α, $0 < \alpha \leqq p$, at x. Finally, if g is smooth of order $\alpha > 0$ at x, and if $g_\alpha(x) = g_\alpha^{+}(x) + g_\alpha^{-}(x) \neq 0$, then g cannot be smooth of any order $\beta > \alpha$ unless α is an even integer and g possesses an αth derivative at x, in which case $g_\alpha^{+}(x) = g_\alpha^{-}(x) = g^{(\alpha)}(x)/\alpha!$.

We will call a real-valued, bounded, symmetric, (absolutely) integrable function K defined on R^1 a kernel, and we will call a kernel proper if it satisfies (1.2). Also, we will write $K \in A_r$ where $r \geqq 0$ is an even integer to mean that K is a kernel for which

$$\int y^i K(y)\, dy = 0, \qquad i = 1, \cdots, r-1, \tag{2.1}$$

$$\int y^r K(y)\, dy \neq 0, \tag{2.2}$$

$$\int y^r |K(y)|\, dy < \infty. \tag{2.3}$$

For $r \geqq 4$, the class A_r contains no nonnegative kernels, and its elements will therefore lead to possibly negative density estimates if used in (1.1). While negative density estimates are obviously undesirable, it is sometimes possible to obtain a higher rate of consistency with kernels in A_r for $r \geqq 4$ than with kernels in A_2 (cf. [1] and Corollary 2.2 below).

LEMMA 2.1. *Let g be a bounded, measurable function on R^1 which is smooth of order $\alpha > 0$ at x and let K be a kernel which satisfies* (2.1) *and* (2.3) *with $r \geqq \alpha$; then as $t \to \infty$*

$$t^\alpha [t \int K(t(x-y)) g(y)\, dy - g(x) \int K(y)\, dy] \to g_\alpha(x) k(\alpha), \tag{2.4}$$

where $g_\alpha(x) = g_\alpha^{+}(x) + g_\alpha^{-}(x)$ and $k(\alpha) = \int_0^\infty y^\alpha K(y)\, dy$.

PROOF. Because of (2.1), the left and right-hand sides of (2.4) differ by

$$\begin{aligned} &\int_{-\infty}^{0} t^\alpha [g(x+yt^{-1}) - \sum_{i=0}^{m} g^{(i)}(x)(yt^{-1})^i/i! - g_\alpha^{-}(x)(-yt^{-1})^\alpha] K(y)\, dy \\ &+ \int_0^\infty t^\alpha [g(x+yt^{-1}) - \sum_{i=0}^{m} g^{(i)}(x)(yt^{-1})^i/i! - g_\alpha^{+}(x)(yt^{-1})^\alpha] K(y)\, dy. \end{aligned} \tag{2.5}$$

Given $\varepsilon > 0$, there exists a δ, $0 < \delta < 1$, such that the integrand in the first integral is less in absolute value than $\varepsilon |y|^\alpha$ for $-\delta t < y < 0$, and it follows that the absolute value of the first integral in (2.5) does not exceed

$$\varepsilon \int_{-\delta t}^{0} |y|^{\alpha} |K(y)|\, dy + 2rB\delta^{-r} \int_{-\infty}^{-\delta t} y^{r} |K(y)|\, dy \tag{2.6}$$

where B is an upper bound for g on R^1, its derivatives at x, and $g_\alpha^{-}(x)$. Since (2.6) may be made arbitrarily small by choosing t sufficiently large and ε sufficiently small, the first integral in (2.5) tends to zero as $t \to \infty$. A similar treatment may be given to the second to complete the proof.

COROLLARY 2.1. *Let f be bounded on R^1 and smooth of order α at $x \in R^1$ and let $K \in A_r$, $r \geqq \alpha$, be a proper kernel; then*

$$t_n^{\alpha}(E[f_n(x; t_n)] - f(x)) \to f_\alpha(x) k(\alpha) \qquad \text{as} \quad n \to \infty.$$

If, in addition, K has a bounded, continuous, integrable rth *derivative on R^1 which satisfies* (2.3), *then*

$$t_n^{\alpha - r} E[f_n^{(r)}(x; t_n)] \to f_\alpha(x) k_r(\alpha)$$

as $n \to \infty$ where $k_r(\alpha) = \int_0^\infty y^\alpha K^{(r)}(y)\, dy$.

PROOF. Since $E[f_n(x; t_n)] = t_n \int K(t_n(x-y)) f(y)\, dy$, the first assertion is clear. So is the second if one observes that $K^{(r)}$ is again symmetric and also satisfies (2.1) and $\int K^{(r)}(y)\, dy = 0$, so that $E[f_n^{(r)}(x; t_n)]$ is equal to $t_n^r [t_n \int K^{(r)}(t_n(x-y)) f(y)\, dy - f(x) \int K^{(r)}(y)\, dy]$.

We will also need the following lemma, the proof of which may be found in [2].

LEMMA 2.2. *Let f be bounded on R^1 and continuous at x, and let $K \in A_0$ be a proper kernel; then $(n/t_n)\operatorname{Var}(f_n(x; t_n)) \to \bar{k} f(x)$, where $\bar{k} = \int K(y)^2\, dy$.*

If $K \in A_r$, $r \geqq 2$, then in view of (2.1) there can be at most one value of α, $0 < \alpha \leqq r$, for which $f_\alpha(x) \neq 0 \neq k(\alpha)$; and if there is such an α, Lemma 2.2 and Corollary 2.1 combine to give

COROLLARY 2.2. *If, in addition to the hypotheses of Lemma 2.2, f is smooth of order α at x, $f_\alpha(x) \neq 0 \neq f(x)$, $K \in A_r$, $r \geqq \alpha$, and $k(\alpha) \neq 0$, then*

$$E[f_n(x; t_n) - f(x)]^2 \sim (t_n/n) f(x) \bar{k} + t_n^{-2\alpha} (f_\alpha(x) k(\alpha))^2. \tag{2.7}$$

The asymptotically optimal choice of t_n (in the sense of minimizing (2.7)) *is τ_n where $\tau_n^{2\alpha+1} = 2\alpha n (f_\alpha(x) k(\alpha))^2 / f(x)\bar{k}$. With this choice of t_n,* (2.7) *is equal to $[2\alpha (f_\alpha(x) k(\alpha))^2 (f(x)\bar{k}/n)^{2\alpha}]^{\gamma} (2\alpha\gamma)^{-1}$ where $\gamma = 1/(2\alpha + 1)$.*

Finally, we will need the following lemma which follows easily from the results of [4].

LEMMA 2.3. *Let K be a kernel having $q \geqq 0$ bounded, integrable derivatives on R^1, and let f be bounded on R^1; then for $p \geqq 1$, the* 2pth *central moment of $f_n^{(q)}(x; t_n)$ is $O((t_n^{2q+1}/n)^p)$ as $n \to \infty$.*

3. The algebra of O_E. In this section X_n, Y_n, and Z_n, $n \geqq 1$, with or without further subscripts will denote random variables having moments of all orders, and a_n, b_n, and c_n will denote positive real numbers. All limits are taken as $n \to \infty$.

We will say that X_n is of small (large) expected order a_n and write $X_n = o_E(a_n)(X_n = O_E(a_n))$ iff $E|X_n|^p = o(a_n^p)(E|X_n|^p = O(a_n^p))$ for every $p \geqq 1$. Our immediate goal is to develop an algebra of O_E which we will use in the following two sections to establish (1.3). We begin by remarking that if $X_n = O_E(a_n)$ and $Y_n = O_E(b_n)$, then by the Hölder and Minkowski inequalities $X_n Y_n = O_E(a_n b_n)$ and $X_n + Y_n = O_E(a_n \vee b_n)$ where $\vee$ denotes maximum; moreover, we have $P[X_n \geqq \varepsilon] = O(a_n^p)$ for every $\varepsilon > 0$ and $p \geqq 1$ by Markov's inequality. From these simple properties applied to the inequalities, $|xx' - yy'| \leqq |x||x'-y'| + |y'||x-y|$ and $|x^k - y^k| \leqq k(|x|+|y|)^{k-1}|x-y|$, $k \geqq 1$, follows

LEMMA 3.1. *Let* $X_{ni} - Y_{ni} = O_E(a_n)$ *and* $X_{ni} = O_E(b_n)$, $i = 1, 2$; *then* $X_{n1}X_{n2} - Y_{n1}Y_{n2} = O_E(a_n[a_n \vee b_n])$, *and* $X_{n1}^k - Y_{n1}^k = O_E(a_n[a_n \vee b_n]^{k-1})$, $k \geqq 1$.

LEMMA 3.2. *Let* $X_n \geqq b_n > 0$, $Y_n \geqq 2\delta > 0$, *and* $-m \leqq Z_n \leqq 1$ w.p. *one for sufficiently large n. If* $X_n - Y_n = O_E(a_n)$ *where* $a_n^k = O(b_n)$ *for some* $k > 0$, *then* (i) $X_n^{Z_n} - Y_n^{Z_n} = O_E(a_n)$; (ii) $X_n^{-1} - Y_n^{-1} = O_E(a_n)$; *and* (iii) $\log X_n - \log Y_n = O_E(a_n)$.

PROOF. Let A_n be the event, $X_n > \delta$; then by Markov's inequality, $P(A_n^c) = O(b_n^k)$ for all $k > 0$, and therefore

$$\begin{aligned} E|X_n^{Z_n} - Y_n^{Z_n}|^p &\leqq (m \vee 1)^p \delta^{-p(m+1)} E[|X_n - Y_n|^p I_{A_n}] \\ &\quad + (m \vee 1)^p b_n^{-p(m+1)} |E[|X_n - Y_n|^{2p}] P(A_n^c)|^{\frac{1}{2}} \\ &= O(a_n^p) + o(a_n^p) = O(a_n^p) \end{aligned}$$

where I_A denotes the indicator of A. This establishes (i) of which (ii) is a special case. The proof of (iii) is similar to that of (i) and will be omitted.

LEMMA 3.3. *Let* $|X_n| \leqq M$ w.p. *one for sufficiently large n and let* $b_n + b_n^{-1} = O(n^b)$ *for some* $b > 0$. *If* $X_n = O_E(a_n)$ *where* $a_n = O(n^{-a})$ *for some* $a > 0$, *then* $b_n^{X_n} - 1 = O_E(a_n \log n)$.

PROOF. Let A_n be the event, $|X_n| < (\log n)^{-1}$; then, as above, $P(A_n^c) = O((b_n + b_n^{-1})^{-k})$ for all $k > 0$. Therefore, from the inequality, $|e^x - 1| \leqq |x| e^{|x|}$, we have

$$\begin{aligned} E|b_n^{X_n} - 1|^p &\leqq E|X_n \log b_n|^p \exp(|p(\log n)^{-1} \log b_n|) \\ &\quad + [E|X_n \log b_n|^{2p} P(A_n^c)]^{\frac{1}{2}} \exp(pM|\log b_n|) \\ &= O_E(a_n^p \log^p n)[1 + P(A_n^c)^{\frac{1}{2}}(b_n + b_n^{-1})^{pM}] \\ &= O(a_n^p \log^p n). \end{aligned}$$

4. The optimal constant. In this section we will suppose that f is known to be bounded on R^1 and to have $r \geqq 2$ continuous derivatives near x with $f(x) \neq 0 \neq f^{(r)}(x)$. We also suppose that $K \in A_r$ is a proper kernel possessing r bounded, continuous, integrable derivatives on R^1; then by Corollary 2.2, the asymptotically optimal t_n sequence will be $\tau_n = cn^\gamma$ where $\gamma = (2r+1)^{-1}$ and

$$c^{2r+1} = \frac{8r}{(r!)^2}(k(r)f^{(r)}(x))^2/\bar{k}f(x).$$

Thus, to estimate τ_n it suffices to estimate c. Let $0 < t_{ni} \to \infty$ with $t_{ni} = o(n^{\gamma^i})$ as $n \to \infty$, $i = 1, 2$, and let $f_n = f_n(x; t_{n1}), f_{nr} = f_n^{(r)}(x; t_{n2})$, $\mu_n = E[f_n]$, and $\mu_{nr} = E[f_{nr}]$. Define

$$c_n^{2r+1} = \frac{8r}{(r!)^2}[(k(r)\mu_{nr})^2 + b_n]/(|\mu_n| + b_n)\bar{k},$$

$$\hat{c}_n^{2r+1} = \frac{8r}{(r!)^2}[(k(r)f_{nr})^2 + b_n]/(|f_n| + b_n)\bar{k},,$$

$$\sigma_n = c_n n^\gamma, \quad \text{and} \quad \hat{\tau}_n = \hat{c}_n n^\gamma \tag{4.1}$$

where $0 < b_n \to 0$ with nb_n bounded away from zero. The theorem to be proved in this section is

THEOREM 4.1. *Let $r \geqq 2$ be an even integer; let f be bounded on R^1 and have r continuous derivatives near $x \in R^1$ with $f(x) \neq 0 \neq f^{(r)}(x)$; and let $K \in A_r$ be a proper kernel with a bounded, continuous, integrable* rth *derivative which satisfies* (2.3). *Define $\hat{\tau}_n$ by* (4.1) *and J by $J(y) = yK'(y)$, $y \in R^1$. If $|J| + |K|$ is dominated by a kernel K_1, which is non-decreasing on* $[0, \infty)$, *then* (1.3) *holds.*

PROOF. By Corollaries 2.1 and 2.2 we have $c_n \to c$ and

$$E[f_n(x; \sigma_n) - f(x)]^2 \sim E[f_n(x; \tau_n) - f(x)]^2 \tag{4.2}$$

as $n \to \infty$. Therefore, it will suffice to show that

$$E[f_n(x; \hat{\tau}_n) - f_n(x; \sigma_n)]^2 = o(n^{-2r\gamma}) \tag{4.3}$$

as $n \to \infty$. Let $a_n^2 = (t_{n1} \vee t_{n2}^{2r+1})/n = o(n^{-2r\gamma})$; then by Lemma 2.3 $(f_n - \mu_n)$ and $(f_{nr} - \mu_{nr})$ are $O_E(a_n)$. Moreover, $\mu_n \to f(x) > 0$, $(|f_n| + b_n) \geqq b_n$ where $a_n^4 = o(b_n)$, and $\mu_{nr} = O(1)$ so that

$$f_{nr}^2 - \mu_{nr}^2 = O_E(a_n),$$

$$(|f_n| + b_n)^{-1} - (|\mu_n| + b_n)^{-1} = O_E(a_n)$$

by Lemmas 3.1 and 3.2. Therefore, $\hat{c}_n^{2r+1} - c^{2r+1} = O_E(a_n)$ by Lemma 3.1. Finally, since $c_n \to c > 0$ and $\hat{c}_n \geqq 8rb_n/(r!)^2(Mt_{n1} + b_n)\bar{k}$ where M is an upper bound for $|K|$, we have $\hat{c}_n - c_n = O_E(a_n)$ by Lemma 3.2. Returning to (4.3), we have

$$E[f_n(x; \hat{\tau}_n) - f_n(x; \sigma_n)]^2 \leqq \left(E\left[n^\gamma \frac{\partial}{\partial t} f_n(x; t)_{\hat{\sigma}_n}\right]^4 E[\hat{c}_n - c_n]^4\right)^{\frac{1}{2}}$$

where $\hat{\sigma}_n$ lies between $\hat{\tau}_n$ and σ_n w.p. one; and since $(E[\hat{c}_n - c_n]^4)^{\frac{1}{2}} = O(a_n^2) = o(n^{-2r\gamma})$ by the choice of t_{n1} and t_{n2}, (4.3) would follow from the boundedness of $E[n^\gamma(\partial/\partial t)f_n(x; t)_{\hat{\sigma}_n}]^4$. Let A_n be the event $\hat{c}_n \geqq c/2$ and $s_n = (c/2)n^\gamma$; then for n large

$$\begin{aligned}E[(\partial/\partial t)f_n(x;t)_{\hat{\sigma}_n}]^4 &= E[(1/n)\textstyle\sum_{i=1}^n K(\hat{\sigma}_n(x-X_i))+J(\hat{\sigma}_n(x-X_i))]^4\\ &\leqq E[(1/n)\textstyle\sum_{i=1}^n K_1(\hat{\sigma}_n(x-X_i))]^4\\ &\leqq (2/cn^\gamma)^4E[(s_n/n)(\textstyle\sum_{i=1}^n K_1(s_n(x-X_i)))I_{A_n}]^4\\ &\quad + M^4P(A_n{}^c)\end{aligned} \tag{4.4}$$

where M is an upper bound for K_1. Now $(s_n/n)\sum_{i=1}^n K_1(s_n(x-X_i))$ is, aside from a constant factor, a sample density of the form (1.1), and therefore is $O_E(1)$ by Lemma 2.3. Moreover, $P(A_n{}^c) = O(n^{-k})$ for every $k>0$ since $\hat{c}_n - c_n = O_E(a_n)$. It follows that (4.3) is $O(n^{-2r\gamma})$, thus completing the proof of Theorem 4.1.

5. The optimal rate. In the previous section we had to assume that the unknown density f had at least $r \geqq 2$ derivatives at the point in question. In this section we will show that this assumption may be weakened by using a more complicated estimation scheme. Specifically, we will assume only that for some unknown value α_0 of α, $0<\alpha_0\leqq 2$, f is smooth of order α_0 and $f_{\alpha_0}(x) \neq 0 \neq f(x)$. There can, of course, be at most one such α_0. Moreover, if f really is smooth near x, say $f''(x)\neq 0$ exists, then we have $\alpha_0 = 2$. However, we are not requiring the existence of even one derivative. We will also assume that $K \in A_2$ is a proper kernel with a bounded, continuous, integrable second derivative K'' and that

$$k_2(\alpha) = \int_0^\infty y^\alpha K''(y)\,dy \neq 0, \qquad 0<\alpha\leqq 2, \tag{5.1a}$$

$$k_2'(0) \neq 0 \neq k(\alpha_0), \qquad \text{and} \tag{5.1b}$$

$$\int_0^\infty y^2 \log(1+y)\,|K''(y)|\,dy < \infty. \tag{5.1c}$$

$k_2(\cdot)$ will then satisfy a uniform Lipschitz condition on $[0, 2]$. For example, the standard normal density satisfies the assumptions placed on K.

Under the assumptions of the preceding paragraph, we have (from Corollary 2.2) that $\tau_n = cn^{\gamma_0}$ where both $\gamma_0 = (2\alpha_0+1)^{-1}$ and $c = [2\alpha_0(f_{\alpha_0}(x)k(\alpha_0))^2/f(x)\bar{k}]^{\gamma_0}$ are unknown. Therefore, we will have to estimate first α_0 and then c. Let $0 < t_{ni} = A_i n^{\delta_i}$, $n \geqq 1$, $i = 1, 2, 3$, where $\delta_1 < \delta_2 < 1/25$, $\delta_3 < 1/5$, and $A_1 = A_2$. Also let $f_n = f_n(x; t_{n3})$, $f_{ni}'' = f_n''(x; t_{ni})$, $i = 1, 2$, $\mu_n = E[f_n]$, and $\mu_{ni}'' = E[f_{ni}'']$, $i = 1, 2$. Define α_n and $\hat{\alpha}_n$ by

$$2-\alpha_n = \left|\frac{\log(|\mu_{n2}''|+b_n)-\log(|\mu_{n1}''|+b_n)}{(\delta_2-\delta_1)\log n}\right| \wedge (2-b_n)$$

$$2-\hat{\alpha}_n = \left|\frac{\log(|f_{n2}''|+b_n)-\log(|f_{n1}''|+b_n)}{(\delta_2-\delta_1)\log n}\right| \wedge (2-b_n)$$

where $0 < b_n \to 0$ with nb_n bounded away from zero. Also let

$$\begin{aligned}e_n &= \mu_{n2}''/k_2(\alpha_n)t_{n2}^{2-\alpha_n}, \qquad \hat{e}_n = f_{n2}''/k_2(\hat{\alpha}_n)t_{n2}^{2-\hat{\alpha}_n},\\ c_n &= [(2\alpha_n e_n{}^2 k(\alpha_n)^2 + b_n)/(|\mu_n| + b_n)\bar{k}]^{\gamma_n},\\ \hat{c}_n &= [(2\hat{\alpha}_n \hat{e}_n{}^2 k(\hat{\alpha}_n)^2 + b_n)/(|f_n| + b_n)\bar{k}]^{\hat{\gamma}_n},\\ \sigma_n &= c_n n^{\gamma_n}, \qquad \text{and} \quad \hat{\tau}_n = \hat{c}_n n^{\hat{\gamma}_n}\end{aligned} \tag{5.2}$$

where $\gamma_n = (2\alpha_n+1)^{-1}$ and $\hat{\gamma}_n = (2\hat{\alpha}_n+1)^{-1}$. The theorem to be proved in this section is

THEOREM 5.1. *Let f be bounded on R^1 and smooth of order α_0, $0 < \alpha_0 \leqq 2$, at x with $f_{\alpha_0}(x) \neq 0 \neq f(x)$, and let $K \in A_2$ be a proper kernel which has a bounded, continuous, integrable second derivative and satisfies* (5.1). *Define $\hat{\tau}_n$ by* (5.2) *and J as in Theorem* 4.1. *If $|J|+|K|$ is dominated by a kernel K_1 which is non-decreasing on* $[0, \infty)$, *then* (1.3) *holds.*

PROOF. By Corollary 2.2 we have $\alpha_n = \alpha_0 + o((\log n)^{-1})$. It follows successively that $n^{\delta\alpha_n} \sim n^{\delta\alpha_0}$ for any δ, that $t_{n2}^{\alpha_n} \sim t_{n2}^{\alpha_0}$, that $e_n \to f_{\alpha_0}(x)$, that $c_n \to c$, that $\sigma_n \sim \tau_n$, and that (4.2) holds (with the new definitions of σ_n and τ_n). Therefore, it will suffice to demonstrate (4.3) with $r = 2$ (and the new definitions of σ_n and $\hat{\tau}_n$). Let $a_n^2 = (t_{n2}^5 \vee t_{n3})/n = o(n^{-4/5}(\log n)^{-6})$; then by Lemma 2.3 we have $(f_n - \mu_n) = O_E(a_n)$ and $(f_{ni}'' - \mu_{ni}'') = O_E(a_n)$, $i = 1, 2$. Since also $a_n^4 = o(b_n)$, it follows from Lemma 3.2 that $(\hat{\alpha}_n - \alpha_n) = O_E(a_n)$ and, thereafter, that $k_2(\hat{\alpha}_n) - k_2(\alpha_n) = O_E(a_n)$, that $(\hat{\gamma}_n - \gamma_n) = O_E(a_n)$, and by Lemma 3.3 that $t_{n2}^{\hat{\alpha}_n - \alpha_n} - 1 = O_E(a_n \log n)$. Since $k_2(\alpha_n) \to k_2(\alpha_0) \neq 0$, and for large n $|k_2(\hat{\alpha}_n)| \geqq |k_2'(0)b_n/2|$, an application of Lemma 3.2 gives $|\hat{e}_n| - |e_n| = O_E(a_n \log n)$. It now follows that $(\hat{c}_n^{2\hat{\alpha}_n+1} - c_n^{2\alpha_n+1}) = O_E(a_n \log n)$ by an argument similar to that of the previous section. Let $(d_n, \hat{d}_n) = (c_n^{2\alpha_n+1}, \hat{c}_n^{2\hat{\alpha}_n+1})$; then $d_n \to d > 0$ and $\hat{d}_n \geqq b_n/(Mt_{n3} + b_n)\bar{k}$ where M is an upper bound for K. Therefore,

$$\hat{c}_n - c_n = (\hat{d}_n^{\hat{\gamma}_n} - d_n^{\hat{\gamma}_n}) + d_n^{\gamma_n}(d_n^{\hat{\gamma}_n - \gamma_n} - 1)O_E(a_n \log n) + O_E(a_n \log^2 n) = O_E(a_n \log^2 n)$$

by Lemmas 3.2 and 3.3. It now follows that $n^{-\gamma_n}(\hat{\tau}_n - \sigma_n) = O_E(a_n \log^3 n)$ and therefore that

$$E[f_n(x; \hat{\tau}_n) - f_n(x; \sigma_n)]^2 \leqq \left(E\left[n^{\gamma_n} \frac{\partial}{\partial t} f_n(x; t)_{\hat{\sigma}_n}\right]^4\right)^{\frac{1}{2}} o(n^{-4/5}) \tag{5.3}$$

where $\hat{\sigma}_n$ lies between $\hat{\tau}_n$ and σ_n. The remainder of the proof of Theorem 5.1 consists of showing the expectation on the right side of (5.3) to be bounded and may be accomplished by repeating the argument given in (4.4).

6. Concluding remark. The referee has pointed out that the $\hat{\tau}_n$ sequences of Sections four and five depend on a b_n sequence which seems just as arbitrary as the t_n sequence of (1.1). The same may be said of the t_{ni} sequences. While the point is well taken, the determination of the b_n and t_{ni} sequences in Sections four and five is not as crucial as that of the t_n sequence in (1.1). Indeed, the former affects only the rate of convergence in (1.3), while the latter affects the rate of mean square consistency.

I would like to thank the referee for pointing out that certain portions of the original version of this paper were hard to follow. I hope that the revision is somewhat easier to read.

REFERENCES

[1] BARTLETT, M. S. (1963). Statistical estimation of density functions. *Sankhyā*, **25-A** 245–254.

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

[3] ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[4] VON BAHR, B. (1965). On the convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36** 808–818.