The Annals of Statistics
1975, Vol. 3, No. 1, 15–29

OPTIMAL CONVERGENCE PROPERTIES OF VARIABLE KNOT, KERNEL, AND ORTHOGONAL SERIES METHODS FOR DENSITY ESTIMATION[1]

BY GRACE WAHBA

University of Wisconsin

Let $W_p^{(m)}(M) = \{f: f^{(\nu)}$ abs. cont., $\nu = 0, 1, \cdots, m-1$, $f^{(m)} \in \mathscr{L}_p$, $||f^{(m)}||_p \leqq M\}$, where $||\cdot||_p$ is the norm in $\mathscr{L}_p$, m is a positive integer and p is a real number, $p \geqq 1$. Let $\{\hat{f}_n(x)\}$, $n = 1, 2, \cdots$ be any sequence of estimates of a density at the point x where $\hat{f}_n(x)$ depends on n independent observations from some density $f \in W_p^{(m)}(M)$. It is shown that if $\sup_{f\in W_p^{(m)}(M)} E_f(f(x) - \hat{f}_n(x))^2 = b_n n^{-\phi(m,p+\varepsilon)}$, where $\phi(m, p) = (2m - 2/p)/(2m + 1 - 2/p)$, and $\varepsilon > 0$, then there exists a $D_0 > 0$ such that $b_n \geqq D_0$ for infinitely many n. Thus the best possible mean square convergence rate for a density estimate, which is uniform over $W_p^{(m)}(M)$, is not better than $n^{-\phi(m,p+\varepsilon)}$ for arbitrarily small ε. The following types of density estimates are shown to have mean square error at a point bounded above by $Dn^{-\phi(m,p)}$, provided that a certain parameter, usually depending on m, p and M, is chosen optimally: the polynomial algorithm, kernel-type estimates, certain orthogonal series estimates, and the ordinary histogram. D's for each method are given.

1. Introduction. Let $W_p^{(m)}$ be the Sobolev space of functions whose first $m-1$ derivatives are absolutely continuous, and whose mth derivative is in $\mathscr{L}_p$. Let

$$||f^{(m)}||_p \equiv \left(\int_{-\infty}^{\infty} |f^{(m)}(\xi)|^p\, d\xi\right)^{1/p} \leqq M\,, \qquad \text{if} \quad p \geqq 1\,,$$
$$||f^{(m)}||_\infty \equiv \sup_\xi |f^{(m)}(\xi)| \leqq M\,, \qquad \text{if} \quad p = \infty\,,$$

and let

$$W_p^{(m)}(M) = \{f: f \in W_p^{(m)},\ ||f^{(m)}||_p \leqq M\}\,.$$

The functions in $W_p^{(m)}(M)$ may be thought of as possessing a certain minimal degree of smoothness, characterized by the parameters m, p and M. In this paper, m, p and M are fixed, $m = 1, 2, \cdots$, $p \geqq 1$, and $M > 0$.

In a recent paper [6] estimates of a density at a point were studied, for densities assumed to be in $W_p^{(m)}(M)$. The choice of estimate there depends m, p, and M. If $\hat{f}_n(x)$ is the estimate of $f(x)$ based on n independent observations from the density f, it was shown, for the estimates in [6],[2] that

$$E(f(x) - \hat{f}_n(x))^2 = O(n^{-\phi(m,p)}) \tag{1.1}$$

Received November 1972; revised September 1973.

[1] This research was sponsored by the Air Force Office of Scientific Research under grant AFOSR 72-2363.

AMS 1970 *subject classifications.* 6270, 6215, 6299.

Key words and phrases. Density estimates, optimal convergence rates.

[2] Only $p = 2$ and $p = \infty$ are considered in [6].

where

$$\phi(m,p) = (2m - 2/p)/(2m + 1 - 2/p) . \tag{1.2}$$

Subsequently, the author was led to conjecture that the convergence rate of (1.1) cannot be improved upon simultaneously for all $f \in W_p^{(m)}(M)$. Indeed, it is known that the Parzen kernel estimates [4] achieve this rate for $m = 1, 2, \cdots$, and $p = \infty$. Serendipitously, a paper by Farrell [1] appeared shortly thereafter with a theorem concerning the best available rates, which allows the question to be answered.

The purpose of this paper is twofold. First, it is shown, as a modified version of Farrell's theorem that, if

$$\sup_{f \in W_p^{(m)}(M)} E_f(f(x) - \hat{f}_n(x))^2 = b_n n^{-\phi(m,p+\varepsilon)} ,$$

where $\hat{f}_n(x)$ is any sequence of estimates of $f(x)$ based on n independent observations from f, and $\varepsilon > 0$ is fixed but may be arbitrarily small, then $\liminf_{n\to\infty} b_n = D_0(\varepsilon) > 0$. (Note that $\phi(m, p)$ is an increasing function of p.)

Secondly, several types of density estimates achieving the rate $n^{-\phi(m,p)}$ are compared on the basis of mean square error. The result (1.1) in [6] for the polynomial algorithm for density estimation is extended to all $p \geqq 1$. Next, it is shown that the Parzen kernel estimates achieve the rate $n^{-\phi(m,p)}$, for $m = 1, 2, \cdots, p \geqq 1$. Then, it is shown that the Kronmal–Tarter orthogonal series method [3] achieves this rate for $m = 1, 2, \cdots$, and $p = 2$ if f has compact support. (The result is probably not true for arbitrary orthogonal series, however.) Finally, it is shown that, for $m = 1$, $p \geqq 1$, the ordinary histogram method for f with compact support achieves the best obtainable rate if the size of the "bins" is allowed to vary appropriately with n.

For each method except the polynomial algorithm, we exhibit a D such that, for all $f \in W_p^{(m)}(M)$,

$$E(f(x) - f_n(x))^2 \leqq Dn^{-\phi(m,p)}(1 + o(1)) . \tag{1.3}$$

D is of the form

$$D = \theta[M^2 A B^{2m-2/p}]^{1/(2m+1-2/p)}$$

where

$$\theta = \theta(m, p) = \frac{(2m + 1 - 2/p)}{(2m - 2/p)^{(2m-2/p)}} ,$$

and A and B are constants given in terms of m, p and Λ where Λ satisfies

$$\sup_\xi f(\xi) \leqq \Lambda .$$

It can be shown that if f is a density there exists $\Lambda = \Lambda(m, p, M) < \infty$ such that

$$\sup_{f \in W_p^{(m)}(M)} \sup_\xi f(\xi) \leqq \Lambda(m, p, M) ,$$

but this demonstration is omitted. In the case of the polynomial algorithm, (1.3) holds uniformly only for the set of f's in $W_p^{(m)}(M)$ satisfying $f(u) \geqq \lambda > 0$, all u in a neighborhood of x. In this case, A and B are inversely dependent on λ.

2. Farrell's theorem for $f \in W_p^{(m)}(M)$. Let k be a positive integer and η a differentiable function on the real line. The function f is said to be in Farrell's class $C_{k\eta}$ if

1. $f^{(\nu)}$ continuous, $\nu = 0, 1, \cdots, k$
2. there exists a polynomial s of degree k such that, for all y, $|f(y) - s(y)| \leqq 2(k!)^{-1}y^k\eta'(y)$.

For our purposes we will take $\eta(y) = Ky^\tau$ for some positive constants K and τ. (See [1] page 172.)

We next show that $f \in W_p^{(m)}(M)$ implies that $f \in C_{m-1,\eta}$ with $\eta(y) = K_0 My^\tau$, $\tau = 2 - 1/p$. For $1 \leqq p < \infty$, $K_0 = \{2\tau[(m-1)q + 1]^{1/q}\}^{-1}$, $1/q + 1/p = 1$, and, for $p = \infty$, $K_0 = \frac{1}{2}$. This follows upon taking $s(y) = \sum_{\nu=0}^{m-1} f^{(\nu)}(0)(y^\nu/\nu!)$ since, with $1/p + 1/q = 1$, using a Hölder inequality on Taylor's formula with remainder,

$$\begin{aligned}\left|f(y) - \textstyle\sum_{\nu=0}^{m-1} f^{(\nu)}(0)\dfrac{y^\nu}{\nu!}\right| &\leqq \left|\int_0^y \frac{(y-u)^{m-1}}{(m-1)!} f^{(m)}(u)\,du\right| \\ &\leqq \frac{1}{(m-1)!}\left[\int_0^{|y|} (|y| - u)^{(m-1)q}\,du\right]^{1/q} \\ &\quad \cdot \left[\int_0^y |f^{(m)}(u)|^p\,du\right]^{1/p} \\ &\leqq 2\frac{|y|^{m-1}}{(m-1)!} \cdot K\tau|y|^{\tau-1}\end{aligned}$$

with

$$K = K_0\|f^{(m)}\|_p \,.$$

Let $X_1, X_2, \cdots, X_n$ be n independent random variables with common unknown density f. Without loss of generality, suppose we are estimating f at the point $x = 0$. Then

THEOREM 2.1 (Farrell, [1], Theorem 1.1). *Suppose $\{a_n, n \geqq 1\}$ is a sequence of nonnegative real numbers such that*

$$\liminf_{n\to\infty} \inf_{f \in C_{m-1,\eta}} P_f(|\gamma_n(X_1, X_2, \cdots, X_n) - f(0)| \leqq a_n) = 1 \tag{2.1}$$

with $\eta(y) = Ky^{2-1/p}$, (and where γ_n is an estimate of $f(0)$ based on $X_1, X_2, \cdots, X_n$). Then

$$\liminf_{n\to\infty} n^{\phi(m,p)}a_n^2 = \infty \,. \tag{2.2}$$

We have the following

COROLLARY. *If*

$$\sup_{f \in C_{m-1,\eta}} E_f(\gamma_n - f(0))^2 = b_n n^{-\phi(m,p)} \tag{2.3}$$

with $\eta(y) = Ky^{2-1/p}$, then there exists $D_0 > 0$ such that $b_n \geqq D_0$ for infinitely many n.

PROOF OF COROLLARY. By Tchebycheff's inequality and the hypothesis,

$$P_f(|\gamma_n - f(0)| \leqq a_n) \geqq 1 - \frac{E_f(\gamma_n - f(0))^2}{a_n^2} \geqq 1 - \frac{b_n n^{-\phi(m,p)}}{a_n^2}.$$

Taking $a_n = n^{-\phi(m,p)/2}$, if $\lim\sup b_n = 0$, then (2.1) is satisfied but (2.2) is not.

We would like to have the Corollary for $f \in W_p^{(m)}(M)$. However, $W_p^{(m)}(M)$ is strictly contained in $C_{m-1,\eta}$ with $\eta(y) = Ky^{2-1/p}$ for any $K > 0$ and so this is too much to ask. (The functions h defined in the Appendix are in $C_{m-1,\eta}$ with $\eta(y) = y^{2-1/p}$, but are not in $W_p^{(m)}$.) However, we come close with the following Theorem, by noting that $\phi(m, p)$ is a continuous, monotone increasing function of p for fixed m.

THEOREM 2.2. *Suppose $\{a_n, n \geqq 1\}$ is a sequence of nonnegative real numbers such that*

$$\liminf_{n\to\infty} \inf_{f \in W_p^{(m)}(M)} P_f(|\gamma_n - f(0)| \leqq a_n) = 1 . \tag{2.4}$$

Then, for every fixed $\varepsilon > 0$,

$$\liminf_{n\to\infty} n^{\phi(m,p+\varepsilon)} a_n^2 = \infty .$$

A proof of Theorem 2.2 is given in the Appendix. The Corollary is immediate, as before.

COROLLARY. *For any fixed $\varepsilon > 0$, suppose*

$$\sup_{f \in W_p^{(m)}(M)} E_f(\gamma_n - f(0))^2 = b_n n^{-\phi(m,p+\varepsilon)} .$$

Then there exists $D_0 > D$ such that $b_n > D_0$ for infinitely many n.

Thus, mean square convergence can take place uniformly over $W_p^{(m)}(M)$ at a rate which is no better than $\phi(m, p + \varepsilon)$ where $\varepsilon > 0$ is arbitrarily small.

3. Convergence properties of the polynomial algorithm for density estimation. Let $t_1, t_2, \cdots, t_n$ be the order statistics for $X_1, X_2, \cdots, X_n$ and let $F_n(x)$ be $n/(n+1)$ times the sample cumulative distribution function, based on $t_1, t_2, \cdots, t_n$. Let $k_n \ll n$ be an appropriately chosen sequence depending on n. An estimate for $f(x)$ may be obtained by interpolating F_n at every k_nth order statistic, t_{ik_n}, $i = 1, 2, \cdots, [n/k_n]$, by a smooth function. Call this smooth function $\hat{F}_n$, and let the density estimate $\hat{f}_n$ be given by

$$\hat{f}_n(x) = \frac{d}{dx} \hat{F}_n(x) .$$

We call this class of methods "variable knot interpolating methods." The "knots" are the points of interpolation. The only examples of these methods that we know of in the literature are the polynomial algorithm [6] and Van Ryzin's histogram method [5], of which [6] is a generalization. The method described in [6] uses local polynomial interpolation and is as follows:

Suppose $f \in W_p^{(m)}$. Let l be the greatest integer in $(n-1)/k_n$. Let

$$\begin{aligned} \hat{f}_n(x) &= 0\,, \qquad x < t_{2k_n} \\ &= \frac{d}{dx}\hat{F}_n(x)\,, \qquad t_{2k_n} \leqq x < t_{(l-m+1)k_n} \\ &= 0\,, \qquad t_{(l-m+1)k_n} \leqq x \end{aligned}$$

where $\hat{F}_n(x)$ is defined as follows: For $m = 1$,

$$\hat{F}_n(x) = F_n(t_{ik_n}) + (x - t_{ik_n})\frac{F_n(t_{(i+1)k_n}) - F_n(t_{ik_n})}{t_{(i+1)k_n} - t_{ik_n}}\,,$$

$$t_{ik_n} \leqq x < t_{(i+1)k_n}\,; \qquad i = 2, 3, \cdots, l-1\,.$$

For $m \geqq 2$, let $\hat{F}_{n,i}(x)$, $i = 1, 2, \cdots, l-m-1$, be the mth degree polynomial which interpolates to $F_n(x)$ at the $m+1$ points $x = t_{ik_n}, t_{(i+1)k_n}, \cdots, t_{(i+m)k_n}$. For $x \in [t_{(i+1)k_n}, t_{(i+2)k_n})$, define $\hat{F}_n(x)$ to coincide with $\hat{F}_{n,i}(x)$, $i = 1, 2, \cdots, l-m-1$.

More explicitly, for any given numbers $x_0 < x_1 < \cdots < x_m$, let $l_\nu(x) = l_\nu(x; x_0, x_1, \cdots, x_m)$ be the mth degree polynomial with $l_\nu(x_\mu) = 1$, $\nu = \mu = 0, 1, \cdots, m$, $l_\nu(x_\mu) = 0$, $\mu \neq \nu$. Let $l_{i,\nu}(x) = l_\nu(x; t_{ik_n}, t_{(i+1)k_n}, \cdots, t_{(i+m)k_n})$. Then

$$f_n(x) = \frac{d}{dx}\sum_{\nu=0}^m l_{i,\nu}(x)\frac{k_n}{(n+1)}\,, \qquad i = i(x)\,, \quad x \in [t_{2k_n}, t_{(l-m+1)k_n}) \tag{3.1}$$

$$f_n(x) = 0 \qquad \text{otherwise,}$$

where it is understood that d/dx applies to the polynomial $l_{i,\nu}(x)$ with i fixed, and where $i(x)$ is defined for $x \in [t_{2k_n}, t_{(l-m+1)k_n})$ as that value i which satisfies

$$t_{(i+1)k_n} \leqq x < t_{(i+2)k_n}$$

for $m \geqq 2$, and by that value i which satisfies

$$t_{ik_n} \leqq x < t_{(i+1)k_n}$$

when $m = 1$. Thus,

$$\begin{aligned} f(x) - \hat{f}_n(x) &= \left\{f(x) - \sum_{\nu=1}^m \frac{d}{dx}l_{i,\nu}(x)\int_{t_{ik_n}}^{t_{(i+\nu)k_n}} f(\xi)\,d\xi\right\} \\ &\quad + \left\{\sum_{\nu=1}^m \frac{d}{dx}l_{i,\nu}(x)\left(F(t_{(i+\nu)k_n}) - F(t_{ik_n}) - \frac{\nu k_n}{n+1}\right)\right\}\,, \\ &\qquad x \in [t_{2k_n}, t_{(l-m+1)k_n})\,, \\ f(x) - \hat{f}_n(x) &= f(x)\,, \qquad x \notin [t_{2k_n}, t_{(l-m+1)k_n})\,. \end{aligned}$$

Therefore,

$$\begin{aligned} &E(f(x) - \hat{f}_n(x))^2 \\ &\leqq 2E\left\{f(x) - \sum_{\nu=1}^m \frac{d}{dx}l_{i,\nu}(x)\int_{t_{ik_n}}^{t_{(i+\nu)k_n}} f(\xi)\,d\xi\right\}^2 \\ &\quad + 2E\left\{\sum_{\nu=1}^m \frac{d}{dx}l_{i,\nu}(x)\left(F(t_{(i+\nu)k_n}) - F(t_{ik_n}) - \frac{\nu k_n}{n+1}\right)\right\}^2 \end{aligned} \tag{3.2}$$

where $i = i(x)$ is a random integer, and, if $x \notin [t_{2k_n}, t_{(l-m+1)k_n})$, then $l_{i,\nu}(x)$ is defined as 0. We will call the first term on the right the bias term, the second, the variance term.

Letting $F(x) = \int_{-\infty}^{x} f(u)\,du$, the variance term may be viewed as the error in approximating $F(t_{ik_n})$ by $\hat{F}(t_{ik_n}) = ik_n/(n+1)$. Under some additional conditions to be stated later, it is shown in [6], that the variance term has the bound

$$(3.3\text{a}) \qquad 2E\left\{\sum_{\nu=1}^{m} \frac{d}{dx} l_{i,\nu}(x)\left(F(t_{(i+\nu)k_n}) - F(t_{ik_n}) - \frac{\nu k_n}{n+1}\right)\right\}^2 \leqq B_1 \frac{1}{k_n}\left(1 + O\left(\frac{1}{k_n}\right) + O\left(\frac{k_n}{n}\right)\right)$$

where

$$(3.3\text{b}) \qquad B_1 = 2m^{2m+3\frac{1}{2}} \frac{\Lambda^{2m}}{\lambda^{2(m-1)}} 3^{\frac{1}{2}}$$ [3]

and $\lambda > 0$ satisfies

$$(3.3\text{c}) \qquad \lambda \leqq f(u)$$

for u in a neighborhood of x. We remark that B_1 is probably not the best constant.

The bias term is due to the error committed in approximating $f(x)$ from values of $F(t_{ik_n})$, $i = 1, 2, \cdots, [n/k_n]$. The mth degree polynomial $\tilde{F}(x)$ interpolating to $F(x)$ at $x_0, x_1, \cdots, x_m$ is given by

$$\tilde{F}(x) = \sum_{\nu=0}^{m} l_\nu(x) \int_{-\infty}^{x_\nu} f(\xi)\,d\xi$$

and its derivative $\tilde{f}(x) = (d/dx)\tilde{F}(x)$ is given by

$$\tilde{f}(x) = \sum_{\nu=0}^{m} \frac{d}{dx} l_\nu(x) \int_{-\infty}^{x_\nu} f(\xi)\,d\xi = \sum_{\nu=1}^{m} \frac{d}{dx} l_\nu(x) \int_{x_0}^{x_\nu} f(\xi)\,d\xi\,.$$

To analyze $f(x) - \tilde{f}(x)$, the following lemma was given in ([6], Theorem 3), for $p = 2$.

LEMMA 3.1. *Let $f \in W_p^{(m)}$ for $p = 2$. Then*

$$(f(x) - \tilde{f}(x))^2 \leqq a(m)\left(\int_{x_0}^{x_m} |f^{(m)}(u)|^p\right)^{2/p}(x_m - x_0)^{2m-2/p}\,,$$
$$x \in [x_0, x_m],\ m = 1, 2;\ x \in [x_1, x_{m-1}],\ m \geqq 3$$

where

$$a(1) = 1\,, \qquad a(2) = (\tfrac{5}{2})^2\,, \qquad a(m) = \left[\frac{2(m+3)}{(m-1)!}\right]^2, \qquad m \geqq 3\,.$$

Lemma 3.1 is immediately extended to $p \geqq 1$ by replacing the Cauchy–Schwarz inequality in (3.9) of [6] by a Hölder inequality with 2 replaced by p.

[3] The factor 2 was erroneously omitted in [6], Equation (2.26 b).

The entire argument of [6] (including the use of Lemma 1) now goes through exactly for $p \geqq 1$, simply by replacing 2 by p in Theorem 3 of [6].
The result (from [6]) is then

$$(3.4a) \qquad 2E\left(f(x) - \sum_{\nu=1}^{m} \frac{d}{dx} l_{i,\nu}(x) \int_{tik_n}^{t(i+\nu)k_n} f(\xi)\, d\xi\right)^2 \leqq M^2 A_1 \left(\frac{k_n}{n+1}\right)^{2m-2/p} \left(1 + O\left(\frac{1}{k_n}\right)\right), \qquad p \geqq 1$$

where $||f^{(m)}||_p \leqq M$ and

$$(3.4b) \qquad A_1 = 2a(m) \cdot m \left(\frac{m}{\lambda}\right)^{2m-2/p}.$$

Thus, ignoring a factor $(1 + O(1/k_n) + O(k_n/n))$, if $f \in W_p^{(m)}(M)$ and $f(u) \geqq \lambda$ for u in a neighborhood of x, then

$$(3.5) \qquad E(f(x) - \hat{f}_n(x))^2 \leqq M^2 A_1 \left(\frac{k_n}{n+1}\right)^{2m-2/p} + B_1 \frac{1}{k_n}.$$

The right-hand side of (3.5) is minimized (see Lemma 4a of [4]) by taking

$$(3.6) \qquad k_n = \left[\frac{1}{(2m - 2/p)} \frac{B_1}{M^2 A_1}\right]^{1/(2m+1-2/p)} (n+1)^{(2m-2/p)/(2m+1-2/p)},$$

in which case

$$E(f(x) - \hat{f}_n(x))^2 \leqq D_1 n^{-\phi(m,p)}$$

where

$$D_1 = \theta(M^2 A_1 B_1^{2m-2/p})^{1/(2m+1-2/p)}$$

and θ is the constant given following (1.3).
For completeness we state the extended version of Theorems 1 and 2 of [6], as now obtains for $p \geqq 1$.

THEOREM 3.1. *Let $f(u) \leqq \Lambda$, all u, let $f(u) \geqq \lambda$ for u in a neighborhood of x, let $|u(1 - F(u))|$ and $|uF(u)|$ be bounded respectively for $u \geqq x$ and $u \leqq x$. Let m be an integer, $m \geqq 1$. Let p be a real number, $p \geqq 1$ and let $f \in W_p^{(m)}(M)$. Let $\hat{f}_n(x)$ be given by* (3.1) *with k_n given by* (3.6). *Then*

$$E(f(x) - \hat{f}_n(x))^2 \leqq D_1 n^{-(2m-2/p)/(2m+1-2/p)}(1 + o(1))$$

where

$$D_1 = \theta(M^2 A_1 B_1^{2m-2/p})^{1/(2m+1-2/p)}$$

and

$$A_1 B_1^{2m-2/p} = \left[2a(m) \cdot m \cdot \left(\frac{m}{\lambda}\right)^{2m-2/p}\right]\left[2m^{2m+3\frac{1}{4}} \frac{\Lambda^{2m}}{\lambda^{2(m-1)}} 3^{\frac{1}{2}}\right]^{2m-2/p}.$$

4. Convergence properties of the Parzen kernel-type density estimates. The argument of this section was graciously suggested to the author by Professor Farrell. Suppose $f \in W_p^{(m)}(M)$. Let $K(y)$ be a real-valued function on $(-\infty, \infty)$

satisfying

$$\begin{array}{lll} \text{(i)} & \sup_{-\infty<y<\infty} |K(y)| < \infty & \\ \text{(ii)} & \int_{-\infty}^{\infty} |K(y)| < \infty & \\ \text{(iii)} & \lim_{y\to\infty} |yK(y)| = 0 & \\ \text{(iv)} & \int_{-\infty}^{\infty} K(y)\, dy = 1 & \\ \text{(v)} & \int_{-\infty}^{\infty} y^i K(y) = 0 & i = 1, 2, \cdots, m-1 \\ \text{(vi)} & \int_{-\infty}^{\infty} |y|^{m-1/p} |K(y)|\, dy < \infty\,. & \end{array}$$

The kernel-type density estimate $\hat{f}_n(x)$ is then given by

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - t_j}{h}\right) \tag{4.1}$$

where $h > 0$ is to be chosen so that $h \to 0$, $nh \to \infty$.

Let

$$f_n(x) = E\hat{f}_n(x) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x - \xi}{h}\right) f(\xi)\, d\xi\,.$$

From [4], Theorem 2A, the variance term is

$$\begin{aligned} E(\hat{f}_n(x) - f_n(x))^2 &= \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(y)\, dy \left(1 + O\left(\frac{1}{nh}\right)\right) \\ &\leqq B_2 \frac{1}{nh}\left(1 + O\left(\frac{1}{nh}\right)\right) \end{aligned} \tag{4.2a}$$

where

$$B_2 = \Lambda \int_{-\infty}^{\infty} K^2(y)\, dy\,. \tag{4.2b}$$

The bias term may be established for $m = 1, 2, \cdots, p \geqq 1$, by noting that

$$E(f_n(x) - f(x)) = \int_{-\infty}^{\infty} K(-\xi) f(x + \xi h)\, d\xi - f(x)\,.$$

Now

$$\begin{aligned} f(x + \xi h) &= f(x) + \sum_{j=1}^{m-1} \frac{(\xi h)^j}{j!} f^{(j)}(x) \\ &\quad + \int_x^{x+\xi h} \frac{(x + \xi h - u)^{m-1}}{(m-1)!} f^{(m)}(u)\, du\,. \end{aligned} \tag{4.3}$$

Using (iv)—(vi) in (4.3) gives

$$E(f_n(x) - f(x)) = \int_{-\infty}^{\infty} K(-\xi)\, d\xi \int_x^{x+\xi h} \frac{(x + \xi h - u)^{m-1}}{(m-1)!} f^{(m)}(u)\, du\,,$$

and an application of a Hölder inequality to the inner integral gives

$$[E(f_n(x) - f(x))]^2 \leqq M^2 A_2 h^{2m-2/p} \tag{4.4a}$$

where

$$A_2 = \frac{1}{[(m-1)!]^2} \frac{1}{[(m-1)q + 1]^{2/q}} \left[\int_{-\infty}^{\infty} |K(\xi)|\, |\xi|^{m-1/p}\, d\xi\right]^2 \tag{4.4b}$$

with $1/q + 1/p = 1$.

Thus, ignoring a factor $(1 + O(1/nh))$,

$$E(f(x) - \hat{f}_n(x))^2 \leqq M^2 A_2 h^{2m-2/p} + B_2 \frac{1}{nh}. \tag{4.5}$$

Define $k_n = nh$, and choose

$$k_n = \left[\frac{1}{(2m - 2/p)} \frac{B_2}{M^2 A_2}\right]^{1/(2m+1-2/p)} \cdot n^{(2m-2/p)/(2m+1-2/p)}, \tag{4.6}$$

which minimizes the right-hand side of (4.5).

We have the following

THEOREM 4.1. *Let m be an integer, $m \geqq 1$. Let p be a real number, $p \geqq 1$ and let $f \in W_p^{(m)}(M)$. Let $\hat{f}_n(x)$ be given by* (4.1) *where K satisfies* (i)—(vi) *and $h = k_n/n$ with k_n given by* (4.6). *Then*

$$E(f(x) - \hat{f}_n(x))^2 \leqq D_2 n^{-(2m-2/p)/(2m+1-2/p)}(1 + o(1))$$

with

$$D_2 = \theta(M^2 A_2 B_2^{2m-2/p})^{1/(2m+1-2/p)}$$

and

$$A_2 B_2^{(2m-2/p)} = \frac{1}{[(m-1)!\,[((m-1)/(1-1/p)) + 1]^{1-1/p}]^2} \times [\textstyle\int_{-\infty}^{\infty} |K(\xi)|\,|\xi|^{m-1/p}\, d\xi]^2 \cdot [\Lambda \int_{-\infty}^{\infty} K^2(y)\, dy]^{(2m-2/p)}.$$

From the point of view of minimizing the bound on the mean square error here, to optimize the choice of kernel, one should choose K subject to (i)—(vi) to minimize

$$\int_{-\infty}^{\infty} |K(\xi)|\,|\xi|^{m-1/p}\, d\xi [\textstyle\int_{-\infty}^{\infty} K^2(\xi)\, d\xi]^{m-1/p}.$$

5. Convergence properties of the Kronmal–Tarter orthogonal series density estimate. Suppose that $f(\xi) = 0$ for $\xi \notin [0, 1]$ and $f \in W_p^{(m)}(M)$ on $[0, 1]$. Let $\psi_k(x) = \cos \pi kx$, $k = 0, 1, 2, \cdots$. Then the Kronmal–Tarter orthogonal series density estimate [3] is given by

$$\hat{f}_n(x) = \textstyle\sum_{k=0}^{r} \hat{a}_k \psi_k(x) \tag{5.1}$$

where r is to be chosen, and

$$\hat{a}_k = \frac{2}{n} \textstyle\sum_{j=1}^{n} \psi_k(t_j), \qquad k = 0, 1, 2, \cdots \tag{5.2}$$

$\hat{a}_k$ is an unbiased estimator of a_k, where

$$a_k = 2 \textstyle\int_0^1 f(\xi)\psi_k(\xi)\, d\xi, \qquad k = 0, 1, 2, \cdots.$$

Since $\{\psi_k\}_{k=0}^{\infty}$ are complete on $\mathscr{L}_2[-1, 1]$ with respect to even functions on $[-1, 1]$ and we can define $f(-\xi) = f(\xi)$, the density f has the Fourier expansion

$$f(x) = \frac{a_0}{2} + \textstyle\sum_{k=1}^{\infty} a_k \cos \pi kx.$$

Thus

$$f(x) - f_n(x) = \sum_{k=1}^{r} (a_k - \hat{a}_k)\phi_k(x) + \sum_{k=r+1}^{\infty} a_k \phi_k(x) .$$

Here $a_0 = \hat{a}_0 = 2$.

The variance term is studied by observing that

$$E(\sum_{k=1}^{r} (a_k - \hat{a}_k)\phi_k(x))^2 = \frac{1}{n} \{\sum_{k,l=1}^{r} \phi_k(x)\phi_l(x)[4 \int_0^1 \phi_k(\xi)\phi_l(\xi)f(\xi)\, d\xi - a_k a_l]\}$$

$$= \frac{1}{n} [4 \int_0^1 (\sum_{k=1}^{r} \phi_k(x)\phi_k(\xi))^2 f(\xi)\, d\xi - (\sum_{k=1}^{r} a_k \phi_k(x))^2] .$$

Now

$$\sum_{k=1}^{r} \phi_k(x)\phi_k(\xi) = \sum_{k=1}^{r} \cos \pi k x \cos \pi k \xi = \tfrac{1}{2}[w_r(x + \xi) + w_r(x - \xi)]$$

where

$$w_r(\tau) = \cos (\tfrac{1}{2}(r + 1)\pi\tau) \frac{\sin (r\pi\tau/2)}{\sin (\pi\tau/2)} .$$

Therefore, for large r, and x in the interior of $[0, 1]$, the variance term "behaves like" $f(x)(r/n)$. For concreteness, we note that since

$$\int_0^1 (\sum_{k=1}^{r} \phi_k(x)\phi_k(\xi))^2\, d\xi \leqq \frac{r}{2}$$

(5.3 a) $$E(\sum_{k=1}^{r} (a_k - \hat{a}_k)\phi_k(x))^2 \leqq B_3 \frac{r}{n}$$

where

(5.3 b) $$B_3 = 2\Lambda .$$

To establish a bound on the bias term, we use the following

LEMMA 5.1. (Young and Hausdorff). *Suppose* $g(x) \in \mathscr{L}_2[-1, 1]$ *with Fourier series* $\sum_{-\infty}^{\infty} g_k e^{i\pi k x}$, $g_k = \frac{1}{2} \int_{-1}^{1} g(x) e^{i\pi k x}\, dx$. *If* $1 < p \leqq 2$, *and* $1/p + 1/q = 1$, *then*

$$(\sum_{-\infty}^{\infty} |g_k|^q)^{1/q} \leqq (\tfrac{1}{2} \int_{-1}^{1} |g(x)|^p\, dx)^{1/p} .$$

This result is stated in Hardy, Littlewood, and Pólya [2], equation (8.5.7); for the proof see [2], page 221. The limitation on p is essential; indeed, if $p \geqq 2$, the reverse inequality holds (see (8.5.6)).

Now,

(5.4) $$|\sum_{k=r+1}^{\infty} a_k \cos \pi k x| \leqq \sum_{k=r+1}^{\infty} |a_k| k^m \cdot \frac{1}{k^m}$$

$$\leqq (\sum_{r+1}^{\infty} |a_k k^m|^q)^{1/q} \left(\sum_{r+1}^{\infty} \frac{1}{k^{pm}}\right)^{1/p} .$$

Also,

(5.5) $$\sum_{r+1}^{\infty} \frac{1}{k^{pm}} \leqq \int_r^{\infty} \frac{1}{\xi^{pm}}\, d\xi = \frac{1}{(pm - 1)} \left(\frac{1}{r}\right)^{pm-1} .$$

Let $f \in W_2^{(m)}$, then $f^{(m)} \in \mathscr{L}_2$ and $f^{(m)}$ has a Fourier series expansion. If, furthermore, $f \in \mathscr{B} : \{f : f^{(2\nu-1)}(0+) = f^{(2\nu-1)}(1-), \nu = 1, 2, \cdots, [(m - 1)/2]\}$, then it can be shown that

$$f^{(m)}(x) = \pi^m \sum_{k=1}^{\infty} a_k(-1)^{m/2}k^m \cos \pi kx\,, \qquad m \quad \text{even}$$
$$= \pi^m \sum_{k=1}^{\infty} a_k(-1)^{(m+1)/2}k^m \sin \pi kx\,, \qquad m \quad \text{odd}\,.$$

Let $g(x) = f^{(m)}(x)$. Then the nonzero Fourier coefficients of g have absolute values $\pi^m|a_k|k^m$, $k = 1, 2, \cdots$, and by the Lemma of Young and Hausdorff, we have

$$\pi^m[\sum_{k=1}^{\infty} |a_k k^m|^q]^{1/q} \leqq (\int_0^1 |f^{(m)}(\xi)|^p)^{1/p}\,, \qquad 1 < p \leqq 2\,. \tag{5.6}$$

Putting together (5.4), (5.5) and (5.6) gives

$$|\sum_{k=r+1}^{\infty} a_k \cos \pi kx|^2 \leqq M^2 A_3 \left(\frac{1}{r}\right)^{2m-2/p} \tag{5.7a}$$

where

$$A_3 = \frac{1}{\pi^{2m}} \frac{1}{(pm-1)^{2/p}}\,. \tag{5.7b}$$

Thus

$$E(f(x) - \hat{f}_n(x))^2 \leqq M^2 A_3 \left(\frac{1}{r}\right)^{2m-2/p} + B_3 \frac{r}{n} \tag{5.8}$$

where

$$A_3 = \frac{1}{\pi^{2m}} \frac{1}{(pm-1)^{2/p}}$$
$$B_3 = 2\Lambda\,.$$

Define k_n by $k_n = n/r$, and choose $r = n/k_n$ with

$$k_n = \left[\frac{1}{(2m-2/p)} \frac{B_3}{M^2 A_3}\right]^{1/(2m+1-2/p)} n^{(2m-2/p)/(2m+1-2/p)}\,. \tag{5.9}$$

Then the right-hand side of (5.8) is minimized. We have the following

THEOREM 5.1. *Let m be an integer, $m \geqq 1$. Let $p = 2$. Let $f \in W_p^{(m)}(M)$ on* [0, 1] *and* 0 *elsewhere, and let $f \in \mathscr{B}$. Let $\hat{f}_n(x)$ be given by* (5.1) *where $r = n/k_n$ with k_n given by* (5.9). *Then*

$$E(f(x) - \hat{f}_n(x))^2 \leqq D_3 n^{-(2m-2/p)/(2m+1-2/p)}$$

with

$$D_3 = \theta(M^2 A_3 B_3^{2m-2/p})^{1/(2m+1-2/p)}$$

and

$$A_3 B_3^{2m-2/p} = \frac{1}{\pi^{2m}(pm-1)^{2/p}} (2\Lambda)^{2m-2/p}\,.$$

We remark here that there is some doubt as to the truth of this result for $2 < p < \infty$. Also, one cannot use an arbitrary orthonormal series and expect to obtain the same result, as $\sup_{x,k} |\cos \pi kx| \leqq 1$ was needed in the proof in (5.3a) and (5.4).

6. Convergence properties of the ordinary histogram. Suppose that $f(\xi) = 0$ for $\xi \notin [0, 1]$ and $f \in W_p^{(1)}(M)$ on $[0, 1]$ for some $p \geqq 1$. Let h be chosen so that $1/h = l$, an integer. Let I_j be the interval $[jh, (j+1)h)$, $j = 0, 1, \cdots, l-1$. Let

$$\hat{f}_n(x) = \frac{Y_j}{nh}, \qquad x \in I_j,\ j = 0, 1, \cdots, l-1, \tag{6.1}$$

where

$$Y_j = \text{number of} \quad t_1, t_2, \cdots, t_n \quad \text{in} \quad I_j .$$

Since Y_j is binomial $B(n, p_j)$ where $p_j = \int_{jh}^{(j+1)h} f(\xi)\, d\xi$,

$$E\hat{f}_n(x) = \frac{1}{h} p_j$$

$$\operatorname{Var} \hat{f}_n(x) = \frac{p_j(1-p_j)}{nh^2} \leqq \frac{\Lambda}{nh} . \tag{6.2}$$

Now, for $x \in I_j$,

$$E(f(x) - \hat{f}_n(x)) = \frac{1}{h} \int_{jh}^{(j+1)h} (f(x) - f(\xi))\, d\xi ,$$

and, for $f \in W_p^{(1)}(M)$, and $x, \xi \in I_j$,

$$|f(x) - f(\xi)| = |\textstyle\int_\xi^x f^{(1)}(u)\, du| \leqq n^{1-1/p} ||f^{(1)}||_p \leqq Mh^{1-1/p} .$$

Thus

$$[E(f(x) - \hat{f}_n(x))]^2 \leqq M^2 h^{2-2/p}$$

and

$$E(f(x) - \hat{f}_n(x))^2 \leqq M^2 A_4 h^{2m-2/p} + B_4 \frac{1}{nh} \tag{6.3}$$

where

$$A_4 = 1 , \qquad B_4 = \Lambda .$$

Define k_n by $k_n = nh$, and choose $h = k_n/n$ with

$$k_n = \left[\frac{1}{(2m - 2/p)} \frac{B_4}{M^2 A_4} \right]^{1/(2m+1-2/p)} n^{(2m-2/p)/(2m+1-2/p)} , \qquad m = 1 . \tag{6.4}$$

Then the right-hand side of (6.3) is minimized and we have the following

THEOREM 6.1. *Let p be a real number, $p \geqq 1$ and let $f \in W_p^{(1)}(M)$ on* $[0, 1]$ *and* 0 *elsewhere. Let $\hat{f}_n(x)$ be given by* (6.1) *where $h = k_n/n$ with k_n chosen as in* (6.4). *Then*

$$E(f(x) - f_n(x))^2 \leqq D_4 n^{-(2-2/p)/(3-2/p)}$$

with

$$D_4 = \theta(A_4 B_4^{2-2/p})^{1/(3-2/p)}$$

where

$$A_4 B_4^{2-2/p} = \Lambda^{2-2/p} .$$

7. Concluding remark. We conclude with a brief remark concerning the criteria we have been using, namely minimum mean square error at a point.

Firstly, there is no asymptotic distribution theory here, and it probably does not exist. In order for asymptotic normality to obtain, it is apparent that the bias (squared) term must be asymptotically negligible compared to the variance. If the parameter (respectively k_n, h, r and h here) is chosen so that this happens, then the rate $n^{-(2m-2/p)/(2m+1-2/p)} = n^{-\phi(m,p)}$ will not obtain. Thus, it seems preferable to choose the parameter to maximize the convergence rate, and use Tchebycheff's Theorem to construct confidence intervals.

Acknowledgment. We wish to acknowledge a number of helpful discussions with Professor R. H. Farrell.

APPENDIX

PROOF OF THEOREM 2.1. The proof consists of modifying certain arguments of [1]. We suppose, without loss of generality, that f is being estimated at the point $x = 0$.

Define

$$\eta(x) = x^{2-1/s} \tag{A.1}$$

where s is any given fixed real number greater than 1 and let $g_{0\delta}(x)$ be defined by

$$\begin{aligned} g_{0\delta}(x) &= -\eta'(x)\,, & 0 \leqq x \leqq \delta/2 \\ g_{0\delta}(x) &= -\eta'(\delta - x)\,, & \delta/2 \leqq x \leqq \delta \\ g_{0\delta}(x) &= 0 & \delta < x \\ g_{0\delta}(-x) &= -g_{0\delta}(x)\,, & x \leqq 0\,. \end{aligned}$$

Proceeding recursively, once $g_{(k-1)\delta}$ has been defined, then let

$$\begin{aligned} g_{k\delta}^*(x) &= \textstyle\int_{-2^{(k-1)}\delta}^{x} g_{(k-1)\delta}(t)\,dt\,, \qquad -2^{(k-1)}\delta \leqq x \leqq 2^{(k-1)}\delta\,, \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

and

$$g_{k\delta}(x) = g_{k\delta}^*(x + 2^{(k-1)}\delta) - g_{k\delta}^*(x - 2^{(k-1)}\delta)\,.^4$$

Let

$$\varepsilon_{k\delta}(x) = g_{k\delta}(x - 2^{(k-1)}\delta)\,. \tag{A.2}$$

$\varepsilon_{k\delta}$, $k = 1, 2, \cdots$, has the following properties:

$$\varepsilon_{k\delta}(x) = 0\,, \qquad |x| \geqq 3 \cdot 2^{k-1}\delta \tag{A.3}$$

$$\textstyle\int_{-\infty}^{\infty} \varepsilon_{k\delta}(x)\,dx = 0 \tag{A.4}$$

$$\max_x |\varepsilon_{k\delta}(x)| = \varepsilon_{k\delta}(0) = c_1\delta^{k+1-1/s} \geqq \delta^{k+1-1/s} \tag{A.5}$$

$$\textstyle\int_{-\infty}^{\infty} (e_{k\delta}(x))^2\,dx \leqq c_2\delta^{2k+3-2/s} \tag{A.6}$$

where $c_1 = 2^{(k-1)(k-2)/2-1+1/s}$, $c_2 = 3 \cdot 2^k\bar{c}_1^{\,2}$, $\bar{c}_1 = 2^{(k-1)(k-2)/2} > c_1$. Relation (A.5)

[4] This is (2.4) of [1]. There is a typographical error there.

is given in [1], Equation 2.15d, and the others follow readily there. Now let $k = m - 1$. For fixed $p < s$, it is easy to see that $\varepsilon_{(m-1)\delta} \in W_p^{(m)}$ and

$$\text{(A.7)} \qquad \int_{-\infty}^{\infty} |\varepsilon_{(m-1)\delta}^{(m)}(x)|^p\,dx = 2^{m+2}(2 - 1/s)(1 - 1/s)\int_0^{\delta/2} x^{-p/s}\,dx \leqq 2^{m+3}\frac{\delta^{1-p/s}}{(1 - p/s)}\,.$$

Now, $\exists f \in W_p^{(m)}(M/2^{1/p})$ with $f(x) = a > 0$ for $|x| \leqq 3 \cdot 2^{k-1}\delta$.
Let

$$\text{(A.8)} \qquad h_\delta(x) = f(x) + \varepsilon_{(m-1)\delta}(x)\,.$$

Then, by (A.3) and (A.7)

$$\int_{-\infty}^{\infty} |h_\delta^{(m)}(x)|^p\,dx = \int_{-\infty}^{\infty} |f^{(m)}(x)|^p\,dx + \int_{-\infty}^{\infty} |\varepsilon_{(m-1)\delta}^{(m)}(x)|^p\,dx \leqq \frac{M^p}{2} + 2^{m+3}\frac{\delta^{1-p/s}}{(1 - p/s)}\,.$$

Thus, h_δ is a density in $W_p^{(m)}(M)$ whenever

$$\text{(A.9)} \qquad a \geqq c_1\delta^{m-1/s}$$

and

$$\text{(A.10)} \qquad \delta \leqq \left[\frac{M^p(1 - p/s)}{2^{m+4}}\right]^{1/(1-(p/s))}\,.$$

By [1], equations (3.3) and (3.4)

$$\text{(A.11)} \qquad P_{h_\delta}(|\gamma_n - h_\delta(0)| \leqq a_n) \leqq [P_f(|\gamma_n - h_\delta(0)| \leqq a_n)]^{\frac{1}{2}}(1 + c_3\delta^{2m+1-2/s})^{n/2}$$

where $c_3 = c_2/a$. Letting

$$\inf_{f \in W_p^{(m)}(M)} P_f(|\gamma_n - f(0)| \leqq a_n) = 1 - \theta_n\,,$$

then (A.11) gives

$$\frac{(1 - \theta_n)^2}{(1 + c_3\delta^{2m+1-2/s})^n} \leqq P_f(|\gamma_n - h_\delta(0)| \leqq a_n)$$

and, by hypothesis

$$(1 - \theta_n) \leqq P_f(|\gamma_n - f(0)| \leqq a_n)\,.$$

Since, by (A.5),

$$|h_\delta(0) - f(0)| = c_1\delta^{m-1/s}\,,$$

then

$$\text{(A.12)} \qquad \frac{(1 - \theta_n)^2}{(1 + c_3\delta^{2m+1-2/s})^n} + (1 - \theta_n) \geqq 1 \Rightarrow 2a_n \geqq c_1\delta^{m-1/s}\,.$$

Letting $\delta = (d_n/n)^{1/(2m+1-2/s)}$, (A.12) gives

$$\text{(A.13)} \qquad \frac{(1 - \theta_n)^2}{\theta_n} \geqq \left(1 + c_3\frac{d_n}{n}\right)^n \Rightarrow n^{\varphi(m,s)}a_n^2 \geqq \tfrac{1}{4}d_n^{\varphi(m,s)}$$

where $\varphi(m, s)$ is given by (1.2). Since the left-hand inequality in (A.13) is true whenever

$$d_n \leqq \frac{n}{c_3}\left[\left(\frac{(1-\theta_n)^2}{\theta_n}\right)^{1/n} - 1\right] = d_n'$$

and $d_n' \to \infty$ as $\theta_n \to 0$, we can always find $\{d_n\}$ so that $d_n \to \infty$ and (A.9) and (A.10) are satisfied for all n sufficiently large. Letting $s = p + \varepsilon$, the theorem is proved for $1 \leqq p < \infty$. For $p = \infty$, let $\eta(x) = (M/4)x^2$. Then

$$\sup_x |\varepsilon_{(m-1)\delta}^{(m)}(x)| = M/2$$

and $\varepsilon_{(m-1)\delta} \in W_\infty^{(m)}(M/2)$. Letting $s = p + \varepsilon = \infty$, (A.3)—(A.6) hold upon multiplying the right-hand sides of (A.5) and (A.6) by $M/4$, and h_δ is a density in $W_\infty^{(m)}$ whenever $f \in W_\infty^{(m)}(M/2)$ and $a \geqq (M/4)c_1\delta^m$. The remainder of the argument is unchanged, and the theorem follows for $p = \infty$.

REFERENCES

[1] FARRELL, R. H. (1972). On best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

[2] HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1934). *Inequalities.* Cambridge Univ. Press, (2nd ed. reprinted 1964).

[3] KRONMAL, R. and TARTER, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* **63** 925–952.

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

[5] VAN RYZIN, J. (1970). On a histogram method of density estimation. Univ. of Wisconsin, Department of Statistics Technical Report #226.

[6] WAHBA, GRACE (1971). A polynomial algorithm for density estimation. *Ann. Math. Statist.* **42** 1870–1886.

[7] WEGMAN, E. J. (1972). Nonparametric probability density estimation: I. A summary of available methods. *Technometrics* **14** 533–546.

[8] WEGMAN, E. J. (1972). Nonparametric probability density estimation: II. A comparison of density estimation methods. *J. Statist. Comput. Simul.* **1** 225–245.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706