DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT No. 337

June 1973

INTERPOLATING SPLINE METHODS FOR DENSITY ESTIMATION II. VARIABLE KNOTS

by

Grace Wahba

University of Wisconsin
Madison, Wisconsin

Typist: Mary Ann Clark

This report was supported by the United States Army under Contract No. DA-31-124-ARO-D-462.

ABSTRACT

The variable knot interpolating spline method for density estimation is introduced. It is assumed that the true cumulative distribution function is in the Sobolev Hilbert space $W_2^{(2)}$ and has compact support. The sample c. d. f. is interpolated at a subset of the order statistics, by the smoothest function in $W_2^{(2)}$, where the smoothness criteria is small $\mathcal{L}_2$ - norm of the second derivative. The density estimate is the derivative of this interpolating function. The interpolating function is a cubic spline, and the density estimate is a quadratic spline. It is shown how to choose the optimal subset of the order statistics. A bound for the expected mean square error at a point is obtained for this estimate and is found to share the convergence rate of several well known density estimates. This estimate should compare favorably with other density estimates if the true c. d. f. is a smooth function in $W_2^{(2)}$.

1. Introduction and Summary

This note is the second of two reports on density estimates formed by differentiation of a smooth function interpolating the sample cumulative distribution function at specified points. The interpolating function is a spline function, which is the smoothest function interpolating the sample c. d. f. when the criteria for smoothness is the $\mathcal{L}_2$ norm of some derivative. In these notes we use as the criteria the second derivative, and the resulting interpolating function is then a cubic spline. Both these reports consider the case where the density is known to have compact support, say [0, 1]. In [12], the sample c. d. f. is interpolated at equally spaced points (called "knots"), where the distance between knots is a parameter to be chosen. This is the "histospline estimate" introduced by Boneva-Kendall-Stefanov [3]. In the present report the interpolation takes place at every k_n th order statistic, where n is the sample size, and $k_n << n$ is a parameter to be chosen.

Let $W_p^{(m)}$ be the Sobolev space of functions

$$\{f: \ f^{(\nu)} \text{ abs. cont.}, \quad \nu = 0, 1, \ldots, m-1, \ f^{(m)} \in \mathcal{L}_p[0, 1]\}$$

and let $W_p^{(m)}(M)$ be given by

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

where

$$\|f^{(m)}\|_p = [\int |f^{(m)}(\xi)|^p d\xi]^{\frac{1}{p}}, \quad p \ge 1$$

$$\|f^{(m)}\|_\infty = \sup_\xi |f^{(m)}(\xi)| .$$

Let $\phi(m, p) = (2m-(2/p))/(2m+1-(2/p))$. It is shown in [11], based on a more general result of Farrell [4], that if

$\hat{f}_n(x)$, $n = 1, 2, \ldots$ is any sequence of estimates of the true density f at the point x, based on a random sample of size n from f and if

$$\sup_{f \in W_p^{(m)}(M)} E(\hat{f}_n(x) - f(x))^2 = b_n n^{-\phi(m, p+\epsilon)} \tag{1.1}$$

for fixed, but arbitrary $\epsilon > 0$, then there exists some $D_o > 0$ such that $b_n \geq D_o$ for infinitely many n.

It was proven in [12], for $m = 1, 2, 3$ and certain values of p, that the rate $n^{-\phi(m, p)}$ is achieved for the histospline density estimate, when the spacing between knots is chosen optimally. That is, there exists a constant D such that

$$\sup_{f \in W_p^{(m)}(M)} E(\hat{f}_n(x) - f(x))^2 \leq D\, n^{-(2m-2/p)/(2m+1-2/p)}, \tag{1.2}$$

where $\hat{f}_n(x)$ is the histospline density estimate, and D depends only on m, p, M.

It is also known (see [11]) that (1.2) holds for the Parzen kernel-type estimates [6] and certain Kronmal-Tartar orthogonal series estimates [5]. For $m = 1$, it holds for histogram methods where the "bins" are of the same size (chosen optimally). For the variable knot methods (Van Ryzin's histogram method [8, 9], see also [10], and the polynomial algorithm [10]) a restricted version of (1.2) has been found such that the bound on the right holds uniformly over sets $W_p^{(m)}(M) \cap \mathcal{S}(x, \lambda, \epsilon)$, where

$$\mathcal{S}(x, \lambda, \epsilon) = \{f : f(u) \geq \lambda \text{ for } |x - u| \leq \epsilon\}.$$

It is the purpose of this note to introduce the variable knot interpolating spline density estimate and to prove a restricted version of (1.2) for this estimate. It is assumed at least that $f \in W_2^{(1)}$, which entails that the c.d.f. $F \in W_2^{(2)}$. The results are demonstrated for $m=1$, $p=2$; $m = 2$, $1 < p \leq 2$; and for $m=3$, $1 \leq p \leq 3$. To obtain (1.2) for $m > 3$ with a spline method it is doubtless both necessary and sufficient that quintic or higher degree splines be used. This is equivalent to changing the smoothness criteria to third or higher order derivatives.

What is the best density estimate? We do not know. (See Wegman [13] for a comparative Monte Carlo study). D's satisfying (1.2) for the methods mentioned are given in [11] and [12] and this report. These D's are not necessarily the smallest possible. It remains to find the smallest possible D for each method. We hope these reports are a first step. Why choose an interpolating cubic spline method? The function interpolating to the c.d.f. (of which the density estimate is the derivative) has the smallest $\mathcal{L}_2$ norm of its second derivative among all interpolating functions in $W_2^{(2)}$. Whenever the true c.d.f. is in $W_2^{(2)}$ and smooth by this criteria, then the interpolating cubic spline methods should compare favorably with other (non-parametric) methods. Should one use the equi-spaced or variable knot method? We do not have a theoretical answer to this question. Intuitively, one feels that the variable knot method may be using more information. Preliminary Monte Carlo runs indicate that the variable knot method reproduces multi-modal densities very well.

In Section 2 we give a complete definition of the variable knot interpolating spline density estimate, and in Section 3, we prove (1.2).

2. Definition of the Variable Knot Interpolating Spline Density Estimate.

Let

$\bar{s} = (s_1, s_2, \ldots, s_\ell)$, where $0 < s_1 < s_2 < \ldots < s_\ell < 1$.

$\bar{y} = (y_1, y_2, \ldots, y_\ell)$

$\bar{a} = (a_1, a_o)$

$\bar{b} = (b_o, b_1)$.

Let $S(x) = S(x; \overline{s}; \overline{a}, \overline{y}, \overline{b})$ be the (unique) solution to the problem: Find $S \epsilon W_2^{(2)}$ to minimize

$$\int_0^1 (G''(u))^2 \, du$$

subject to

$$G'(0) = a_1$$

$$G(0) = a_o$$

$$G(s_i) = y_i, \quad i=1, 2, \ldots, \ell$$

$$G(1) = b_o$$

$$G'(1) = b_1$$

It is well known [7], that the solution S is a cubic spline with knots $s_1, s_2, \ldots, s_\ell$, that is, S is piecewise a polynomial of degree ≤ 3 in each interval $[s_j, s_{j+1}]$, $j = 0, 1, \ldots, \ell$, where $s_o = 0$, $s_{\ell+1} = 1$, with the pieces joined at the knots so that S, S' and S'' are continuous. An explicit formula for $S(x)$ may be found in various places, see [2], [12]. Efficient computational routines which deliver $S(x)$ and $S'(x)$ given $\overline{s}$, $\overline{a}$, $\overline{y}$ and $\overline{b}$ are commonly available, see [1].

Let F_n be $n/(n+1)$ times the sample c. d. f. based on a sample of size n from f. Without loss of generality, we hence forth suppose n satisfies

$$(n+1) = (\ell+1)k_n,$$

where k_n (an integer) is to be chosen, and ℓ is an integer. We assume that f is supported on $[0, 1]$. Let $t_1, t_2 \ldots, t_n$ be the order statistics and let $\overline{t}$, $\overline{F}_n$, $\overline{\alpha}$ and $\overline{\beta}$ be defined by

$$\overline{t} = (t_{k_n}, t_{2k_n}, \ldots, t_{\ell k_n}) \tag{2.1a}$$

$$\overline{F}_n = (F_n(t_{k_n}), F_n(t_{2k_n}), \ldots, F_n(t_{\ell k_n})) \tag{2.1b}$$

$$\overline{\alpha} = (\hat{f}_n(0), 0) \tag{2.1c}$$

$$\overline{\beta} = (1, \hat{f}_n(1)), \tag{2.1d}$$

where $\hat{f}_n(0)$ and $\hat{f}_n(1)$ are estimates of $f(0)$ and $f(1)$ to be described shortly. (If $f(0)$ and $f(1)$ are known, then the true values are substituted). Define the interpolating c.d.f. $\hat{F}_n(x)$ by

$$\hat{F}_n(x) = S(x; \overline{t}, \overline{\alpha}, \overline{F}_n, \overline{\beta}).$$

Thus $\hat{F}_n(x)$ is the cubic spline of interpolation to $F_n(x)$ at the points t_{ik_n}, $i = 0, 1, 2, \ldots \ell+1$, where $t_o = 0$, $t_{\ell+1} = 1$, and satisfying $\hat{F}_n'(0) = \hat{f}_n(0)$, $\hat{F}_n'(1) = \hat{f}_n(1)$. The variable knot cubic spline density estimate $\hat{f}_n(x)$ is given by

$$\hat{f}_n(x) = \frac{d}{dx} \hat{F}_n(x). \tag{2.2}$$

To define $\hat{f}_n(0)$ and $\hat{f}_n(1)$ we must choose $m=1, 2$ or 3, where it is assumed that $f \in W_p^{(m)}(M)$. Let $\ell_{o,\nu}(x)$ be the polynomial of degree m satisfying

$$\ell_{o,\nu}(x) = 1, \quad x = t_{\nu k_n}$$
$$= 0, \quad x = t_{jk_n}, \quad j \neq \nu, j = 0, 1, \ldots, m$$

and let $\ell_{1,\nu}(x)$ be the polynomial of degree m satisfying

$$\ell_{1,\nu}(x) = 1, \quad x = 1 - t_{\nu k_n}$$
$$= 0, \quad x = 1 - t_{jk_n}, \quad j \neq \nu, \quad j = 0, 1, \ldots, m.$$

Let

$$\hat{f}_n(0) = \frac{d}{dx} \sum_{\nu=0}^{m}{}' \ell_{o,\nu}(x) \Big|_{x=0} F_n(t_{\nu k_n}) = \frac{d}{dx} \sum_{\nu=0}^{m}{}' \ell_{o,\nu}(x) \frac{\nu k_n}{n+1} \tag{2.3}$$

$$\hat{f}_n(1) = \frac{d}{dx} \sum_{\nu=0}^{m}{}' \ell_{1,\nu}(x) \Big|_{x=1} F_n(1-t_{\nu k_n}) = \frac{d}{dx} \sum_{\nu=0}^{m}{}' \ell_{o,\nu}(x) \frac{n+1-\nu k_n}{n+1} \tag{2.4}$$

$\hat{f}_n(0)$ is the derivative at 0 of the m th degree polynomial interpolating $F_n(x)$ at $0, t_{k_n}, \ldots, t_{mk_n}$, and similarly for $\hat{f}_n(1)$.

3. Bounds for $E(f(x) - f_n(x))^2$.

In the remainder of this note, which consists of the proof of (1.2) for $\hat{f}_n$ given by (2.2), we will assume that $f(0)$ and $f(1)$ are known, and $\hat{f}_n(0)$ and $\hat{f}_n(1)$ in (2.1c) and (2.1d) are replaced by $f(0)$ and $f(1)$. However our result (1.2) is true without this assumption, as can be seen from the results of [10] and the argument in [12], since $E(f(0) - \hat{f}_n(0))^2$ converges to 0 at the same rate as $E(f(x) - \hat{f}_n(x))^2$ for $x \neq 0$. We omit the details of this argument in the interest of brevity.

To obtain the mean square error at a point, we note, as usual that it is convenient to look at the so-called bias and variance terms separately. We have

$$f(x) - \hat{f}_n(x) = \frac{d}{dx}(F(x) - \tilde{F}(x)) + \frac{d}{dx}(\tilde{F}(x) - \hat{F}_n(x)), \tag{3.1}$$

where

$$\tilde{F}(x) = S(x; \bar{t}, \bar{\alpha}, \bar{F}, \bar{\beta}) \tag{3.2}$$

with

$$\bar{F} = (F(t_{k_n}), F(t_{2k_n}), \ldots, F(t_{\ell k_n})).$$

Thus $\widetilde{F}$ is the cubic spline of interpolation to $F(x)$ at t_{ik_n}, $i = 0, 1, \ldots, \ell+1$, which also matches the first derivative of F at the boundaries. Then

$$E(f(x) - \hat{f}_n(x))^2 \leq 2E\left(\frac{d}{dx}(F(x) - \widetilde{F}(x)\right)^2 + 2E\left(\frac{d}{dx}(\widetilde{F}(x) - \hat{F}_n(x)\right)^2 . \quad (3.3)$$

The bulk of the work in obtaining bounds on both the bias term, $E(\frac{d}{dx}(F(x) - \widetilde{F}(x)))^2$, and the variance term, $E(\frac{d}{dx}(\widetilde{F}(x) - \hat{F}_n(x)))^2$ has already been done in [10] and in [12]. We state the results as lemmas here. Lemmas 1 and 2 give bounds on the bias and variance terms in terms of the order statistics.

Lemma 1.

Let $i = i(x)$ be the random integer (with values $i = 0, 1, \ldots \ell,$) which satisfies $x \in [t_{ik_n}, t_{(i+1)k_n})$. Let $\Delta_j = (t_{(j+1)k_n} - t_{jk_n})$, $j = 0, 1, \ldots, \ell$, $t_o \equiv 0$, $t_{\ell+1} \equiv 1$.

i) Suppose $f \in W_2^{(1)}$. Then

$$\left(\frac{d}{dx}(F(x) - \widetilde{F}(x))\right)^2 \leq \frac{1}{3}\Delta_i \int_0^1 |f'(\xi)|^2 \, d\xi . \quad (3.4)$$

ii) Suppose $f \in W_p^{(2)}$, for $p \geq 1$. Then

$$\left(\frac{d}{dx}(F(x) - \widetilde{F}(x))\right)^2 \leq \frac{1}{9}\Delta_i \sum_{j=0}^{\ell} \Delta_j^{3-2/p} \left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f''(\xi)|^p d\xi\right]^{2/p} . \quad (3.5)$$

iii) Suppose $f \in W_p^{(3)}$, for $p \geq 1$. Then

$$\left(\frac{d}{dx}(F(x) - \widetilde{F}(x))\right)^2 \leq \frac{1}{3}\Delta_i \sum_{j=0}^{\ell} \Delta_j^{5-2/p} \left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f'''(\xi)|^p \, d\xi\right]^{2/p} . \quad (3.6)$$

Proof: Lemma 1 is a direct consequence of Lemmas 1, 2, 3 and 4 of [12], see the proofs in the Appendix to [12].

We have a similar result for the variance term:

Lemma 2.

Let i be as in Lemma 1. Then

$$\frac{d}{dx}(\tilde{F}(x) - \hat{F}_n(x)) \leq 8\,\Delta_i \sum_{j=o}^{\ell} C_j \frac{|\psi_j|}{\Delta_j^2} \tag{3.7}$$

where

$$\psi_j = F(t_{(j+1)k_n}) - F(t_{jk_n}) - \frac{k_n}{n+1}, \tag{3.8}$$

$$C_j = \frac{1}{2^{|i-j|+1}} + \frac{1}{2^{|i+1-j|+1}}, \qquad j = 0, 1, \ldots, \ell, \quad i \neq j \tag{3.9}$$

$$= \frac{1}{2} + \frac{1}{2^2} + \frac{1}{8}, \qquad i = j\,.$$

Proof: Noting that

$$\tilde{F}(x) - \hat{F}_n(x) = S(x;\, \alpha_o, \bar{t}, \bar{\epsilon}, \beta_o)$$

where

$$\alpha_o = (0, 0)$$

$$\beta_o = (0, 0)$$

$$\bar{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_\ell), \quad \epsilon_j = F(t_{jk_n}) - \frac{jk_n}{n+1},$$

Lemma 2 becomes Lemma 5 of [12].

It remains to take expectations of the terms (3.4), (3.5), (3.6), (3.7), Lemmas 3, 4, and 5 below will provide all the tools we need.

Lemma 3.

Let $t_1, t_2, \ldots, t_n$ be the order statistics from a density f, which has support on $[0, 1]$. Let $t_o = 0$, $t_{(\ell+1)k_n} = 1$.

i) Let $f(u) \leq \Lambda$, all u. Then, for any fixed $q(1 \leq q \leq k_n)$,

$$E(t_{(j+1)k_n} - t_{jk_n})^{-q} \leq \Lambda^q \left(\frac{n+1}{k_n}\right)^q \left(1 + O(\frac{1}{k_n})\right), \quad j = 0, 1, \ldots, \ell. \tag{3.10}$$

ii) Let $i = i(x)$, $(i = 0, 1, \ldots, \ell)$ be the random integer which satisfies

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

and suppose that $f(u) \geq \lambda$ for $u \epsilon [x-\epsilon, x+\epsilon]$, some $\epsilon > 0$. Then, for fixed $q(1 \leq q \leq k_n/2)$,

$$E(t_{(i+1)k_n} - t_{ik_n})^q \leq \frac{1}{\lambda^q} (\frac{k_n}{n+1})^q \left(1 + O(\frac{1}{k_n})\right) \tag{3.11}$$

If $f(u) \geq \lambda$ on $[0, 1]$, then (3.11) holds for any i.

Proof: This Lemma is a special case of Lemma 1 of [10].

Lemma 4.

$$E^{\frac{1}{2}} \psi_i^4 \leq 3^{\frac{1}{2}} \frac{k_n}{(n+1)^2} \left(1 + O\left(\frac{k_n}{n+2}\right)\right) \tag{3.12}$$

Proof: This is Lemma 2 of [10].

Lemma 5.

Let $t_o, t_1, \ldots, t_{n+1}$ be as in Lemma 3 with f strictly positive on $[0, 1]$.

Let $\rho \epsilon \mathcal{L}_p[0, 1]$ and let

$$X_j = \int_{t_{jk_n}}^{t_{(j+1)k_n}} |\rho(\xi)|^p d\xi .$$

Then, for any $r > 0$, $\epsilon > 0$, and $k_n \equiv k$,

$$E X_j^r \le \left\{ \int_{F^{-1}(\frac{jk}{n+1} - \epsilon)}^{F^{-1}(\frac{(j+1)k}{n+1} + \epsilon)} |\rho(\xi)|^p \, d\xi \right\}^r + \frac{1}{2(n+1)\epsilon^2} \|\rho\|_p^{pr}. \qquad (3.13)$$

Proof: Let $g(u,v)$ be the joint density of t_{jk} and $t_{(j+1)k}$, and let

$$h(u,v) = \left[\int_u^v |\rho(\xi)|^p \, d\xi \right]^r \quad \text{for } 0 \le u \le v \le 1. \text{ Then}$$

$$E X_j^r = \iint_{A_1 \cup A_2} g(u,v)\, h(u,v)\, du\, dv$$

where

$$A_1 = \{u, v: F^{-1}(\frac{jk}{n+1} - \epsilon) \le u \le v \le F^{-1}(\frac{(j+1)k}{n+1} + \epsilon)\}$$

$$A_2 = \{[0,1] \times [0,1]\} \ominus A_1 .$$

On A_1,

$$h(u,v) \le \left\{ \int_{F^{-1}(\frac{jk}{n+1} - \epsilon)}^{F^{-1}(\frac{(j+1)k}{n+1} + \epsilon)} |\rho(\xi)|^p \, d\xi \right\}^r$$

and on A_2, $|h(u,v)| \le \|\rho\|_p^{pr}$. Now,

$$\Pr\{t_{jk}, t_{(j+1)k} \in A_2\} \leq \Pr\{t_{jk} \notin [F^{-1}(\frac{jk}{n+1} - \epsilon), F^{-1}(\frac{jk}{n+1} + \epsilon)]\}$$

$$+ \Pr\{t_{(j+1)k} \notin [F^{-1}(\frac{(j+1)k}{n+1} - \epsilon), F^{-1}(\frac{(j+1)k}{n+1} + \epsilon)]\} \quad . \tag{3.14}$$

If z_{jk} is the jk th order statistic out of n from a uniform distribution on $[0,1]$, then

$$\Pr\{t_{jk} \notin [F^{-1}(\frac{jk}{n+1} - \epsilon), F^{-1}(\frac{jk}{n+1} + \epsilon]\}$$

$$= \Pr\{z_{jk} \notin [\frac{jk}{n+1} - \epsilon, \frac{jk}{n+1} + \epsilon]\} \, .$$

Since $E z_{jk} = \frac{jk}{n+1}$, var $z_{jk} = \frac{p_j(1-p_j)}{(n+2)}$ where $p_j = \frac{jk}{n+1}$,

then by Chebychev's theorem, the right hand side of (3.14) is bounded by

$$\frac{p_j(1-p_j) + p_{j+1}(1-p_{j+1})}{(n+2)\epsilon^2} \leq \frac{1}{2(n+2)\epsilon^2} \, .$$

Since

$$E X_j^r \leq \left\{ \int_{F^{-1}(\frac{jk}{n+1} - \epsilon)}^{F^{-1}(\frac{(j+1)k}{n+1} + \epsilon)} |\rho(\xi)|^p \, d\xi \right\}^r P(A_1) + \|\rho\|_p^{pr} P(A_2)$$

the Lemma follows.

To obtain a bound on the bias term we note that, if $f \in W_2^{(1)}$ and $f \geq \lambda$ in a neighborhood of $\mathbf{x}$, then Lemma 1, i) and Lemma 3, ii) give directly

$$E\left(\frac{d}{dx}(F(x)-\widetilde{F}(x))\right)^2 \leq \frac{1}{3\lambda}\left(\frac{k_n}{n+1}\right)(1+O(\frac{1}{k_n}))\,. \qquad (3.15)$$

To consider the case $f \in W_p^{(2)}$ or $f \in W_p^{(3)}$, note that, if $1/r + 1/s = 1$, r, $s > 1$, then by a Hölder inequality and then the Cauchy-Schwartz inequality,

$$E\,\Delta_i\Delta_j^\theta\left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f^{(m)}(\xi)|^p\,d\xi\right]^{2/p} \leq E^{1/r}\,(\Delta_i\Delta_j^\theta)^r\;E^{1/s}\left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f^{(m)}(\xi)|^p d\xi\right]^{2s/p}$$

$$\leq \left[E\,\Delta_i^{2r}\,E\,\Delta_j^{2\theta r}\right]^{1/2r}\,.\;E^{1/s}\left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f^{(m)}(\xi)|^p d\xi\right]^{2s/p} \qquad (3.16)$$

for any $\theta \geq 1$.

By Lemma 3, ii), assuming now that $f(u) \geq \lambda$ on $[0, 1]$,

$$\left[E\,\Delta_i^{2r}\;E\Delta_j^{2\theta r}\right]^{1/2r} \leq \frac{1}{\lambda^{(\theta+1)}}\left(\frac{k_n}{n+1}\right)^{(\theta+1)}\left(1+O(\frac{1}{k_n})\right)\,. \qquad (3.17)$$

Next, define γ_j by

$$\gamma_j = \int_{F^{-1}((j-1/2)k_n/n)}^{F^{-1}((j+3/2)k_n/n)} |f^{(m)}(\xi)|^p\,, \quad j = 0, 1, \ldots, \ell.$$

By Lemma 5, with ϵ taken as $k_n/2(n+1)$, and $s > 1$,

$$E^{1/s}\left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f^{(m)}(\xi)|^p \, d\xi\right]^{2s/p} \le \left[\gamma_j^{\;2s/p} + 2\,\frac{(n+1)}{k_n^2}\,\|\rho\|_p^{2s}\right]^{1/s}$$

$$\le \gamma_j^{2/p} + \left(2\,\frac{n+1}{k_n^2}\right)^{1/s} \|\rho\|_p^2 . \qquad (3.18)$$

Combining (3.16), (3.17), and (3.18), gives, for $m = 2, 3$, and $s > 1$

$$E\,\Delta_i \sum_{j=0}^{\ell} \Delta_j^{2m-1-2/p}\left[\int_{t_{jk_n}}^{t_{(j+1)k_n}} |f^{(m)}(\xi)|^p d\xi\right]^{2/p}$$

$$\le \frac{1}{\lambda^{2m-2/p}}\left(\frac{k_n}{n+1}\right)^{2m-2/p}\left[\sum_{j=0}^{\ell}\gamma_j^{2/p} + \left(\frac{n+1}{k_n}\right)\left(2\frac{n+1}{k_n^2}\right)^{1/s}\right]\left(1+O\left(\frac{1}{k_n}\right)\right)$$

(3.19)

For $1 \le p \le 2$, $\sum_{j=0}^{\ell} \gamma_j^{\,2/p} \le \left(\sum_{j=0}^{\ell}\gamma_j\right)^{2/p} = 2^{2/p}\|f^{(m)}\|_p^2$. Thus, by combining

(3.19) with Lemma 1, ii) or iii) gives, for $m=2, 3$, $1 \le p \le 2$, and any $s > 1$,

$$E\left(\frac{d}{dx}\,(F(x) - \tilde{F}(x))\right)^2 \le$$

$$\frac{d_m}{\lambda^{2m-2/p}}\left(\frac{k_n}{n+1}\right)^{2m-2/p} 2^{2/p}\|f^{(m)}\|_p^2\left(1+2^{-(2/p)+(1/s)}\left[\frac{n+1}{k_n^{(s+2)/(s+1)}}\right]^{1+1/s}\right)\left(1+O\left(\frac{1}{k_n}\right)\right)$$

(3.20)

where $d_2 = 1/9$ and $d_3 = 1/3$. We will later take $k_n = O(n^{(2m-2/p)/(2m+1-2/p)})$. If k_n is chosen this way and $s < 2m-1-2/p$, then $(n+1)/k_n^{(s+2)/(s+1)} \to 0$. Since $2m-1-2/p > 1$ for $m=3$, or $m=2$ and $p > 1$, in these cases this term is asymptotically negligible.

A bound for the variance term follows easily below: From (3.7),

$$\left[E \frac{d}{dx} (\tilde{F}(x) - \hat{F}_n(x))\right]^2 \leq 64 \sum_{j=0}^{\ell}{}' \sum_{k=0}^{\ell}{}' C_j C_k \, E \, \Delta_i^2 \frac{|\psi_j \psi_k|}{\Delta_j^2 \Delta_k^2} .$$

Let $a_{jk} = E \, \Delta_i^2 \dfrac{|\psi_j \psi_k|}{\Delta_j^2 \Delta_k^2}$. Now

$$a_{jj} = E \, \Delta_i^2 \Delta_j^{-4} \psi_j^2 \leq E^{1/4} \Delta_i^8 \, E^{1/4} \Delta_j^{-16} \, E^{1/2} \psi_j^4 \leq \frac{\Lambda^4}{\lambda^2} \, 3^{1/2} \frac{k_n}{(n+1)^2} \left(1 + O(\frac{k_n}{n+2}) \right) .$$

But $a_{jk} \leq \sqrt{a_{jj}} \sqrt{a_{kk}}$ and so

$$\left[E \frac{d}{dx} (\tilde{F}(x) - \hat{F}_n(x))\right]^2 \leq 64 \left(\sum_{j=0}^{\ell}{}' C_j \sqrt{a_{jj}} \right)^2 \leq 64 \sup_j a_{jj} \left(\sum_{j=0}^{\ell}{}' C_j \right)^2$$

$$\leq 64 \sup_j a_{jj} \left(\sum_{j=0}^{\ell}{}' C_j \right)^2$$

$$\leq B \frac{1}{k_n} \left(1 + O\left(\frac{k_n}{n+2}\right) \right) \qquad (3.21a)$$

where, (with the aid of (3.9) and (3.12),

$$B = (8 \cdot 3\tfrac{1}{8})^2 \, 3^{1/2} \frac{\Lambda^4}{\lambda^2} . \qquad (3.21b)$$

Here $f(u) \geq \lambda$ for $u \in [x-\epsilon, x+\epsilon]$.

Before stating the main theorem we remark that there exist

$\Lambda = \Lambda\,(m, p, M) < \infty$ such that

$$\sup_{\substack{f \in W_p^{(m)}(M) \\ f \text{ a density}}} \quad \sup_s \ f(s) \leq \Lambda\,.$$

We assume such a Λ is chosen. Let $\mathcal{S}(\lambda) = \{f\colon f(u) \geq \lambda,\ u \in [0,1]\}$. We now have the main

<u>Theorem:</u> Let $f(x) = 0$, $x \notin [0,1]$, let $f \in W_p^{(m)}(M)$ on $[0,1]$ for one of the following: $m=1$, $p=2$, or $m=2$, $1 < p \leq 2$, or $m=3$, $1 \leq p \leq 2$. If $m=1$, let $f \in \mathcal{S}(x, \lambda, \epsilon)$, if $m = 2$ or 3, let $f \in \mathcal{S}(\lambda)$. Then

$$E(f(x) - f_n(x))^2 \leq A\ M^2 \left(\frac{k_n}{n}\right)^{2m-2/p} \left(1 + I_m + O\left(\frac{1}{k_n}\right)\right) + B\,\frac{1}{k_n}\left(1 + O\left(\frac{k_n}{n+2}\right)\right). \tag{3.22}$$

where

$$A = \frac{2}{3\lambda}, \quad m = 1,\ p = 2 \tag{3.23}$$

$$= \frac{2 \cdot 2^{2/p}}{9\lambda^{4-2/p}}, \quad m = 2, \quad 1 < p \leq 2$$

$$= \frac{2 \cdot 2^{2/p}}{3\,\lambda^{6-2/p}}, \quad m = 3, \quad 1 \leq p \leq 2$$

$$I_M = 0, \quad m = 1$$

$$= 2^{-(2/p)+(1/s)} \left[\frac{n+1}{k_n^{1+1/(s+1)}}\right]^{1+1/s}, \quad \text{any } s > 1$$

$$B = 2\left(\frac{\Lambda}{\lambda}\right)^2 \Lambda^2 \left(8 \cdot 3\tfrac{1}{8}\right)^2 3^{1/2} \tag{3.24}$$

Proof: Combine (3. 3), (3. 15) or (3. 20) and (3. 21).

If I_m and lower order terms in $\frac{1}{k_n}$ and $\frac{k_n}{n}$ are ignored, the right hand side of (3. 22) is minimized by letting

$$k_n = \left[\frac{1}{(2m-2/p)} \quad \frac{B}{AM^2} \right]^{1/(2m+1-2/p)} n^{(2m-2/p)/(2m+1-2/p)} . \tag{3.25}$$

We have the

Corollary:

Let k_n be given by (3. 25) . Then

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

where

$$D = \frac{(2m+1-2/p)}{(2m-2/p)^{(2m-2/p)}} \quad (M^2 A B^{2m-2/p})^{1/(2m+1-2/p)} .$$

and A and B are given by (3. 23) and (3. 24) .

REFERENCES

[1] Academic Computing Center, "Approximation and Interpolation, Reference Manual for the 1108." Chapter 6. MACC, University of Wisconsin Madison.

[2] Ahlberg, J. H., Nilson, E. N. and Walsh, J. L. (1967). The Theory of Splines and their Applications, Academic Press, New York.

[3] Boneva, L., Kendall, D. and Stefanov, I. (1971). Spline Transformations:- Three new diagnostic aids for the statistical data analyst. J. Roy. Statistist. Soc. 33, 1-70.

[4] 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.

[5] Kronmal, R. and Tartar, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. JASA, 63, 925-952.

[6] Parzen, E. (1962). On the estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.

[7] Schoenberg, I. J. (1967). On Spline Functions, in Inequalities O. Shisha, ed., 255-291, Academic Press, Inc., New York.

[8] Van Ryzin, John (1970). On a histogram method of density estimation. University of Wisconsin, Madison, Statistics Department Technical Report # 226.

[9] Van Ryzin, John (1973). A histogram method of density estimation. To appear, Communications in Statistics.

[10] Wahba, Grace. (1971). A polynomial algorithm for density estimation. Ann. Math. Statist., 42, 6, 1870-1886.

[11] Wahba, Grace. (1972). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, submitted.

[12] Wahba, Grace. (1973). Interpolating spline methods for density estimation. I. Equi-spaced knots. University of Wisconsin, Madison, Statistics Department Technical Report #327.

[13] Wegman, E. J. (1972). Nonparametric probability density estimation: II. A comparison of density estimation methods. J. Statist. Comput. Simul., 1, 225-246.