
DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT No. 315

November 1972

OPTIMAL CONVERGENCE PROPERTIES OF THE POLYNOMIAL ALGORITHM FOR DENSITY ESTIMATION

by

Grace Wahba

Typist: Bernice R. Weitzel

This work was supported in part by the Wisconsin Alumni Research Foundation and by the Air Force Office of Scientific Research under Grant AFOSR-72-2363.

ABSTRACT

Convergence properties of the polynomial algorithm for density estimation are determined for densities f satisfying $f^{(m)} \varepsilon \mathcal{L}_p$. It is shown that

$$E|f(x) - \hat{f}_{n,m}(x)|^2 = O(n^{-(2m-2/p)/(2m+1-2/p)})$$

$$m = 1, 2, \ldots$$

$$p = 1, 2, \ldots$$

where $\hat{f}_{n,m}(x)$ is the estimate of $f(x)$ based on n independent observations from the density f. This result was previously known for $p = 2$ and $p = \infty$. By applying a theorem of Farrell, it is shown that this rate is the best obtainable for the class of f's satisfying

$$\left[\int_{-\infty}^{\infty} [f^{(m)}(u)]^p du\right]^{1/p} \leq M.$$

Let $t_1, t_2, \ldots, t_n$ be the order statistics from a random sample of size n from a population with unknown density $f(x)$. We are interested in estimating the density $f(x)$. Let $F_n(x)$ be $n/(n+1)$ times the sample cumulative distribution function, and 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 t_{ik_n}, $i = 1, 2, \ldots, [\frac{n}{k_n}]$ by a smooth function, call it $\hat{F}_n$, and letting 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 "order statistic methods". The only examples of this method that we know of in the literature are [3] and Van Ryzin's histogram method [2], of which [3] is a generalization. The method described in [3] uses local polynomial interpolation and is as follows:

Suppose f possess $m-1$ continuous derivatives and $f^{(m)} \varepsilon \mathcal{L}_p(-\infty,\infty)$, for some integer $p \geq 1$. (f is then said to be in the Sobolev space $W_p^{(m)}$). Let ℓ be the greatest integer in $(n-1)/k_n$. Let

$$\begin{aligned} \hat{f}_{n,m}(x) &= 0 \quad , \quad x < t_{2k_n} \\ &= \frac{d}{dx} \hat{F}_{n,m}(x) \; , \quad t_{2k_n} \leq x < t_{(\ell-m+1)k_n} \qquad (1) \\ &= 0 \quad , \quad t_{(\ell-m+1)k_n} \leq x \end{aligned}$$

where $\hat{F}_{n,m}(x)$ is defined as follows:

For $m = 1$,

$$\hat{F}_{n,1}(x) = F_n(t_{ik_n}) + x\,\frac{F_n(t_{(i+1)k_n})-F_n(t_{ik_n})}{t_{(i+1)k_n}-t_{ik_n}}\,,\quad t_{ik_n} \leqq x < t_{(i+1)k_n};\ i=2,3,\ldots,\ell-1.$$

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

In [3], convergence properties of this algorithm are studied for $p = 2$ and $p = \infty$. More precisely, the following theorem concerning the mean square convergence of $\hat{f}_{n,m}(x)$ to $f(x)$ is proved in [3].

Theorem 1.[1/]

Let $f(u) \leq \Lambda$, all u, let $f(u) \geq \lambda$ for u in a neighborhood of x, let $|u(1 - F(u))|$ and $|uF(u)|$ be bounded respectively for $u \geq x$ and $u \leq x$. Let $f\varepsilon W_p^{(m)}$. Let $\hat{f}_{n,m}(x)$ be given by (1) with k_n given by

$$k_n = \left[\frac{1}{(2m-\frac{2}{p})}\ \frac{B}{A}\right]^{1/(2m+1-2/p)} (n+1)^{(2m-2/p)/(2m+1-2/p)}$$

where

[1/] This theorem is the content of Theorems 1 and 2 of [3].

$$A = 2a(m) \; ||f^{(m)}||_p^2 \; (\tfrac{m}{\lambda})^{2m-2/p} \; (1 + O(\tfrac{1}{k_n}))$$

$$B = m^{2m+3\frac{1}{4}} \; \frac{\Lambda^{2m}}{\lambda^{2(m-1)}} \; 3^{\frac{1}{2}} \; (1 + O(\tfrac{1}{k_n}) + O(\tfrac{k_n}{n}))$$

with

$$\begin{aligned} a(m) &= 1, \; m = 1 \\ &= (\tfrac{5}{2})^2, \; m = 2 \\ &= \left[\frac{2(m+3)}{(m-1)!}\right]^2, \; m = 3, 4, \ldots \end{aligned}$$

and

$$\begin{aligned} ||f^{(m)}||_p &= \left[\int_{-\infty}^{\infty} [f^{(m)}(u)]^p \, du\right]^{1/p}, \; p = 1, 2, \ldots \\ &= \sup_{\xi} |f^{(m)}(\xi)| \;, \; p = \infty . \end{aligned}$$

Then, for $p = 2$ and $p = \infty$,

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

where

$$D = \frac{(2m+1-2/p)}{(2m-2/p)^{(2m-2/p)}} \; (AB^{2m-2/p})^{1/(2m+1-2/p)}.$$

The purpose of this communication is to demonstrate the truth of Theorem 1 for $p = 1, 3, 4, \ldots$, and to apply a theorem of Farrell to show that the rate of convergence of (2) is the best obtainable for the estimation of a density at a point for $f \varepsilon W_p^{(m)}$.

The proof of Theorem 1 in [3] relies on the analysis of the so-called bias and variance parts of the error. Letting $F(x) = \int_{-\infty}^{x} f(u)du$, the variance part is due to the error comitted in approximating $F(t_{ik_n})$ by $\hat{F}_n(t_{ik_n}) = \frac{ik_n}{n+1}$. The bias part is then due to the error comitted in approximating $f(x)$ using only values of $F(t_{ik_n})$. The bias part of the error for the density estimate of (1) is studied in [3] as follows:

For any given numbers $x_0 < x_1 < \ldots < x_m$, let $\ell_\nu(x) =$ $\ell_\nu(x;\, x_0, x_1, \ldots x_m)$ be the mth degree polynomial with $\ell_\nu(x_\mu) = \delta_{\mu,\nu}$ $\mu,\nu = 0, 1, \ldots, m$. Then the mth degree polynomial $\tilde{F}(x)$ interpolating to $F(x)$ at $x_0, x_1, \ldots, x_m$ is given by

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

For $x \varepsilon [x_0, x_m]$, let $\tilde{f}(x) = \frac{d}{dx}\tilde{F}(x)$,

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

The following theorem is given in [3].

Theorem 2.[2] Let $f \varepsilon W_p^{(m)}$ for $p = 2$. Then

$$|f(x) - \tilde{f}(x)|^2 \le a(m) \left(\int_{x_0}^{x_m} [f^{(m)}(u)]^p \right)^{2/p} (x_m - x_0)^{2m-2/p}, \quad x \varepsilon [x_0, x_m],\ m = 1, 2$$

$$x \varepsilon [x_1, x_{m-1}],\ m \ge 3. \tag{3}$$

Theorem 2 is immediately extended to $p = 1, 3, 4, \ldots$ by replacing the Cauchy-Schwartz inequality in (3.9) of [3] by a Hölder inequality with 2 replaced by p. Theorem 1 then follows for $p = 1, 3, 4, \ldots$ from Theorem 2 by following the steps in [3] exactly, simply replacing 2 by p in (3) whenever it occurrs.

To show that the rate of (2) is the best obtainable, we will apply a theorem of Farrell. f is said to be in Farrell's class $C_{k\eta}$ if

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

where, for our purposes we take $\eta(x) = Kx^{\tau}$ for some positive constants K and τ. (See [1] p. 172).

We show that $f \varepsilon W_p^{(m)}$ implies $f \varepsilon C_{m-1,\eta}$ with $\eta(x) = Kx^{\tau}$, $\tau = 2-1/p$, K a constant given below. This follows upon taking $s(x) = \sum_{\nu=0}^{m-1} f^{(\nu)}(0)\frac{x^{\nu}}{\nu!}$, since, with $\frac{1}{p} + \frac{1}{q} = 1$, using a Hölder inequality on Taylors formula with remainder,

---

[2] Theorem 3 of [3].

$$|f(x) - \sum_{\nu=0}^{m-1} f^{(\nu)}(0)\, \frac{x^\nu}{\nu!}| \leq |\int_0^x \frac{(x-u)^{m-1}}{(m-1)!}\, f^{(m)}(u)du|$$

$$\leq \frac{1}{(m-1)!}\left[\int_0^{|x|} (x-u)^{(m-1)q} du\right]^{1/q} \left|\int_0^x |f^{(m)}(u)|^p du\right|^{1/p}$$

$$\leq 2\, \frac{x^{m-1}}{(m-1)!} \cdot \quad K\tau x^{\tau-1}$$

with

$$\tau = 2-1/p$$

$$K = \frac{1}{2\tau}((m-1)q+1)^{-1/q} \cdot \left[\int_{-\infty}^{\infty} |f^{(m)}(u)|^p du\right]^{1/p} .$$

We will now apply the following

Theorem 3. (Farrell, [1], Thm. 1.1). Suppose $\{a_n, n \geq 1\}$ is a sequence of non-negative real numbers such that

$$\liminf_{n\to\infty} \inf_{f\varepsilon C_{m-1,\eta}} P_f(|\gamma_n(t_1, t_2, \ldots, t_n)| - f(0)| \leq a_n) = 1 . \tag{4}$$

with $\eta(x) = Kx^{(2-1/p)}$, (and where γ_n is an estimate of $f(0)$ based on $t_1, t_2, \ldots, t_n$). Then

$$\liminf_{n\to\infty} n^{(2m-2/p)/(2m+1-2/p)} a_n^2 = \infty . \tag{5}$$

Let $Y_n = |\gamma_n(t_1, t_2, \ldots, t_n) - f(0)|$ and let $\phi = (2m-2/p)/(2m+1-2/p)$. By Tchebycheff's inequality,

$$P(Y_n \leq a_n) \geq 1 - \frac{EY_n^2}{a_n^2} .$$

Thus, if $EY_n^2 = b_n O(n^{-\phi})$ for any sequence b_n tending to 0, then, upon taking $a_n = O(n^{-\phi/2})$, we have that (4) is satisfied but (5) is not. Thus, no sequence of estimates of $f(0)$ with a better convergence rate than that of (2) can be found, for $f \varepsilon W_p^{(m)}$.

We have also recently succeeded in showing, for the case $m = 1$, $p = 2$, that if we replace $\hat{F}_{n,1}(x)$ of (1) by an appropriate cubic polynomial spline of interpolation to $F_n(x)$ at t_{ik_n}, $i = 1, 2, \ldots [\frac{n}{k_n}]$, then

$$E|f(x) - \hat{f}_{n,1}(x)|^2 = O(n^{-\phi}) .$$

Cubic polynomial spline interpolation should prove to be a highly practical method. These results will appear separately.

References

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

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

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

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Department of Statistics University of Wisconsin Madison, Wisconsin 53706	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

A POLYNOMIAL ALGORITHM FOR DENSITY ESTIMATION

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Scientific Interim

5. AUTHOR(S) *(First name, middle initial, last name)*

Grace Wahba

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 1972	8	3

8a. CONTRACT OR GRANT NO. AFOSR-72-2363 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 315 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Air Force Office of Scientific Research 1400 Wilson Boulevard Arlington, Virginia

13. ABSTRACT

Convergence properties of the polynomial algorithm for density estimation are determined for densities f satisfying $f^{(m)} \varepsilon \mathcal{L}_p$. It is shown that

$$E|f(x) - \hat{f}_{n,m}(x)|^2 = O(n^{-(2m-2/p)/(2m+1-2/p)})$$

$$m = 1, 2, \ldots$$
$$p = 1, 2, \ldots$$

where $\hat{f}_{n,m}(x)$ is the estimate of $f(x)$ based on n independent observations from the density f. This result was previously known for $p = 2$ and $p = \infty$. By applying a theorem of Farrell, it is shown that this rate is the best obtainable for the class of f's satisfying

$$\left[\int_{-\infty}^{\infty} [f^{(m)}(u)]^p du\right]^{1/p} \leq M .$$

DD FORM 1 NOV 65 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT