The Annals of Statistics
1977, Vol. 5, No. 6, 1258–1264

PROPERTIES OF HERMITE SERIES ESTIMATION OF PROBABILITY DENSITY

BY GILBERT G. WALTER

University of Wisconsin—Milwaukee

An unknown density function $f(x)$, its derivatives, and its characteristic function are estimated by means of Hermite functions $\{h_j\}$. The estimates use the partial sums of series of Hermite functions with coefficients $\hat{a}_{jn} = (1/n) \sum_{i=1}^{n} h_j(X_i)$ where $X_1 \cdots X_n$ represent a sequence of i.i.d. random variables with the unknown density function f. The integrated mean square rate of convergence of the pth derivative of the estimate is $O(n^{(p/r)+(5/6r)-1})$. The same is true for the Fourier transform of the estimate to the characteristic function. Here the assumption is made that $(x - D)^r f \in L^2$ and $p < r$. Similar results are obtained for other conditions on f and uniform mean square convergence.

1. Introduction. The problem of estimating an unknown density f using a sequence $X_1, X_2, \cdots, X_n$ of i.i.d. random variables has received considerable attention during the last decades. A number of different methods have been proposed including the kernel method of Rosenblatt [4] and Parzen [3], the orthogonal series method of Kronmal–Tarter [2], the histogram method of Van Ryzin [7], the polynomial interpolation of Wahba [8], [10], the Fourier transform method of Blum and Susarla [1], and the Hermite series method of Schwartz [5]. The first four methods have been summarized and their mean square rates of convergence calculated by Wahba [9] who showed that the "best possible" rate was approached in certain cases with each of these methods.

The Hermite series method which is the subject of this work was not considered, however, even though its rate of convergence compares to the others. We shall adopt the notations of Schwartz in which the unknown density function is estimated by approximating the partial sums of its Hermite series with a sum $\sum \hat{a}_{jn} h_j(x)$ where $\hat{a}_{jn} = (1/n) \sum_{i=1}^{n} h_j(X_i)$. He obtained bounds for the rates of convergence which we shall improve slightly by using better bounds on the Hermite functions. Moreover we calculate the rates of convergence of the derivatives of the estimator to derivatives of the density. By using the fact that the Hermite functions are eigenfunctions of the Fourier transform we are able to get estimates for the characteristic function and their rates of convergence as well.

This method has much to commed it particularly for densities which do not have compact support. For many applications, e.g., life testing, it is more desirable to consider densities which tail off to zero at infinity. Some of these, the rapidly decreasing functions, are naturally associated with Hermite functions

Received February 1976; revised January 1977.
AMS 1970 *subject classification*. 62G05.
Key words and phrases. Density estimates, Hermite functions.

since they share this property. They may be characterized as those functions for which $(x - D)^r f \in L^2$ for each r. While most other estimators require special hypotheses to be used for such densities, the Hermite series estimator does not (as will be shown in Theorem 1).

Another advantage is the ease with which estimates may be calculated numerically. The calculations require only simple algorithms based on the recurrence relations for the Hermite functions. This is true as well for the estimates of the derivatives and the characteristic functions. No approximate differentiation or integration is needed.

Finally, as Schwartz observed, the extension to the multivariate case is immediate, and the error is the same.

In Section 2 we review some of the standard properties of the Hermite functions. In Section 3 we derive the rates of convergence of the estimate and its derivatives in the sense of MISE, while in Section 4 we do the same for the estimate of the characteristic function. In Section 5 the MSE and MISE rates are derived for densities with compact support. In Section 6 the results are extended to the multivariate case.

2. Properties of Hermite functions. The Hermite functions $\{h_j\}$ are the complete orthonormal system in $L^2(-\infty, \infty)$ which satisfy the equations

(1) $$(x^2 - D^2)h_j = (2j + 1)h_j \qquad j = 0, 1, 2, \cdots .$$

They may be expressed in terms of the Hermite polynomials H_j as

$$h_j(x) = \frac{H_j(x)e^{-x^2/2}}{(2^j j!\, \pi^{\frac{1}{2}})^{\frac{1}{2}}} \qquad j = 0, 1, 2, \cdots .$$

They satisfy the recurrence formulae (see [6], page 106),

(2) $$xh_j = (j/2)^{\frac{1}{2}}h_{j-1} + ((j + 1)/2)^{\frac{1}{2}}h_{j+1} \qquad j = 1, 2, \cdots$$

and

(3) $$Dh_j = (j/2)^{\frac{1}{2}}h_{j-1} - ((j + 1)/2)^{\frac{1}{2}}h_{j+1} \qquad j = 1, 2, \cdots .$$

They satisfy the following inequalities (see [6], page 242):

(4) $$|h_j(x)| \leqq C_\infty(j + 1)^{-(1/12)} \qquad x \in (-\infty, \infty), j = 0, 1, \cdots$$

and

(5) $$|h_j(x)| \leqq C_M(j + 1)^{-\frac{1}{4}} \qquad x \in (-M, M), j = 0, 1, \cdots .$$

The constant C_M satisfies $C_M \leqq 1 + \frac{1}{2}M^{\frac{5}{2}}$. (See the Appendix for the method of calculating this.) An expression for C_∞ may be found in [6], page 242.

The Hermite functions are eigenfunctions of the Fourier transform operator. That is, they satisfy the equation

(6) $$(i)^j h_j(\omega) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{i\omega x}h_j(x)\, dx\,, \qquad \omega \in (-\infty, \infty),\ j = 0, 1, 2, \cdots .$$

We shall exploit these properties to obtain the convergence theorems in the next sections.

3. Rates of convergence of the estimate and its derivatives. We shall use the same approximating function $\hat{f}_n$ to the unknown df f as was used in [5],

$$\hat{f}_n(x) = \sum_{j=0}^{q(n)} \hat{a}_{jn} h_j(x) \qquad n = 1, 2, \cdots \tag{7}$$

where $q(n)$ is some increasing sequence of integers satisfying $q(n)/n \to 0$ and $\hat{a}_{jn}$ is the estimator for the Hermite coefficient a_j of the density $(= \int fh_j)$. The variance of the coefficient estimator is

$$\begin{aligned} E(\hat{a}_{jn} - a_j)^2 &= \frac{1}{n^2} E\{\textstyle\sum_{k=1}^{n} h_j^2(X_k)\} - \frac{a_j^2}{n} \\ &\leqq \frac{2}{n} C_\infty (j+1)^{-\frac{1}{6}} \qquad j = 0, 1, \cdots, n = 1, 2, \cdots \end{aligned} \tag{8}$$

for $f \in L^2 \cap L^1$ by inequality (4). If f has compact support as well, then inequality (5) can be used instead to obtain the bound

$$E(\hat{a}_{jn} - a_j)^2 \leqq \frac{2}{n} C_M (j+1)^{-\frac{1}{2}} \qquad j = 0, 1, \cdots, n = 1, 2, \cdots. \tag{9}$$

By equation (3) the derivatives of h_j may be expressed as linear combinations of other Hermite functions. This expression has the form

$$h_j^{(p)} = \sum_{s=m}^{p} \alpha_j^{s,p} h_{j+s} \qquad p, j = 0, 1, \cdots, m = \max(-j, -p) \tag{10}$$

where the coefficients satisfy

$$|\alpha_j^{s,p}| \leqq K_p (j+p)^{p/2} \qquad p, j = 0, 1, \cdots, |s| \leqq p. \tag{11}$$

We shall need this inequality in order to calculate the error of the derivatives. Indeed we see by (8), (10), and (11) that

$$\begin{aligned} E \int [\hat{f}_n^{(p)}(x) &- f^{(p)}(x)]^2\, dx \\ &= E \int [\textstyle\sum_{j=0}^{q} (\hat{a}_{jn} - a_j) h_j^{(p)}]^2 \\ &\quad - 2E \int \textstyle\sum_{j=0}^{q} (\hat{a}_{jn} - a_j) h_j^{(p)} \sum_{k=q+1}^{\infty} a_k h_k^{(p)} \\ &\quad + \int [\textstyle\sum_{j=q+1}^{\infty} a_j h_j^{(p)}]^2 \\ &= \textstyle\sum_{j=0}^{q} E(\hat{a}_{jn} - a_j)^2 \int [h_j^{(p)}]^2 + \int [\sum_{j=q+1}^{\infty} a_j h_j^{(p)}]^2 \\ &\leqq \frac{2}{n} C_\infty^2 \textstyle\sum_{j=0}^{q} (j+1)^{-\frac{1}{6}} K_p^2 (2p+1)(j+p)^p \\ &\quad + \int [\textstyle\sum_{j=q+1}^{\infty} a_j \sum_{s=-p}^{\infty} a_j^{s,p} h_{j+s}]^2 \\ &\leqq C_1 \frac{(q+p)^p (q+1)^{\frac{5}{6}}}{n} + C_2 \textstyle\sum_{j=q+1}^{\infty} a_j^2 (j+p)^p. \end{aligned} \tag{12}$$

The middle term drops out since $\hat{a}_{jn}$ is an unbiased estimator of a_j. The constants C_1 and C_2 are derived from the constants in the previous line and may be calculated directly. Both depend upon p but not upon n or q.

The second term need not converge in general. Here we have assumed implicitly that the pth derivative of f exists but that is not sufficient for convergence of the series.

THEOREM 1. *Let $(x - D)^r f \in L^2(-\infty, \infty)$ for some integer $r > 0$; let p be an integer satisfying $0 \leqq p < r$; let $q(n) = O(n^{1/r})$; then the mean integrated square error in the pth derivative of the estimate* (7) *satisfies*

$$E \int |\hat{f}_n^{(p)} - f^{(p)}|^2 = O(n^{p/r+5/6r-1}) .$$

The asymptotic expression in the conclusion clearly holds for the first term in brackets in (12). In order to prove it for the second term, we observe that by hypothesis the series $\sum b_j^2$ converges where the b_j are the coefficients of $(x - D)^r f$. These coefficients are related to those of f by the expression

$$b_j = \int (x - D)^r f h_j = \int f(x + D)^r h_j = (2j)^{\frac{1}{2}}(2j - 2)^{\frac{1}{2}} \cdots (2j - 2r + 2)^{\frac{1}{2}} a_{j-r} , \tag{13}$$

by repeated application of integration by parts and the sum of (2) and (3). Hence we have

$$\sum_{j=q+1}^{\infty} a_j^2 (j + p)^p \leqq \sum_{j=q+1}^{\infty} b_{j+r}^2 (2j)^{-r} (j + p)^p \leqq \frac{(q + 1 + p)^p}{(2q + 2)^r} \sum_{j=q+r+1}^{\infty} b_j^2 = O(n^{-1+p/r}) , \tag{14}$$

which is more than we need. This proves the theorem.

REMARK. The case $p = 0$ of this theorem corresponds to the theorem of Schwartz mentioned in the introduction. The hypotheses are the same, but the conclusion has been extended to the case $r = 1$ while the asymptotic expression has been sharpened to $O(n^{5/6r-1})$ as compared to $O(n^{1/r-1})$.

It should be observed that for each p and r satisfying the hypothesis the exponent $p/r + 5/6r - 1$ is negative and hence the error approaches zero.

4. Rates of convergence of characteristic functions of the estimates. Because of the unique property that the Hermite functions have with respect to the Fourier transform (6), we are able to get an approximation to the characteristic function of f with no additional work. Indeed if we define $\tilde{\hat{f}}_n$ to be the Fourier transform of $\hat{f}_n$,

$$\tilde{\hat{f}}_n = \sum_{j=0}^{q(n)} (2\pi)^{\frac{1}{2}} i^j \hat{a}_{jn} h_j \qquad n = 1, 2, \cdots , \tag{15}$$

then the following holds.

COROLLARY 1. *Let r, p, f and $q(n)$ be as in Theorem 1, then the* MISE *in the pth derivative of the estimate* (15) *satisfies*

$$E \int |\tilde{\hat{f}}_n^{(p)} - \tilde{f}^{(p)}|^2 = O(n^{p/r+5/6r-1})$$

where $\tilde{f}$ is the characteristic function of f.

5. Rates of convergence of MISE and MSE for estimates of functions with compact support. If the function f we are trying to estimate has compact support the inequality (5) can be used and the hypothesis can be weakened a little.

THEOREM 2. *Let f have compact support and suppose $D^r f \in L^2$ for some integer $r > 0$; let p be an integer such that $0 \leqq p < r$; let $q(n) = O(n^{1/r})$; then the* MISE *in the* pth *derivative of the estimate* (7) *satisfies*

$$E \int |\hat{f}_n^{(p)} - f^{(p)}|^2 = O(n^{p/r+1/2r-1}) .$$

Since f has compact support and $D^r f \in L^2$, $x^p D^s f \in L^2$ for all integers $p \geqq 0$ and $0 \leqq s \leqq r$. Hence $(x - D)^r f \in L^2$, and the hypothesis of Theorem 1 is met. The first term in the brackets of (12) is now dominated by $3C_M^2(q+p)^p(q+1)^{\frac{1}{2}}/n$, which leads to the stronger conclusion.

When uniform mean square error is considered, similar results obtain.

THEOREM 3. *Let f have compact support and suppose $D^r f \in L^2$ for some integer $r > 1$; let p be an integer such that $0 \leqq p < r - 1$; let $q(n) = O(n^{1/r})$; then the mean square error in the* pth *derivative of the estimate* (7) *satisfies*

$$E(\hat{f}_n^{(p)}(x) - f^{(p)}(x))^2 = O(n^{p/r+1/r-1})$$

uniformly on compact sets.

The proof is similar to the others except that cross product terms must be considered in calculating

$$E[\textstyle\sum_{j=0}^{q} (\hat{a}_{jn} - a_j)h_j^{(p)}(x)]^2 .$$

However, this can be reduced to the previous inequalities by means of Schwarz's inequality.

REMARK. The rate of convergence given here might be compared to that given by Wahba [9] for another orthogonal system, the trigonometric on [0, 1]. She shows that for $f \in W_2^{(m)}$, the MSE is $O(n^{-1+1/2m})$, while Theorem 3, in the case $p = 0$, gives us $O(n^{-1+1/m})$. Thus in this case for which the trigonometric system is natural, the Hermite does almost as well, while for the case considered in Theorem 1 which is natural for the Hermite, the trigonometric system cannot even be used.

6. Higher dimensions. One of the surprising things about this method of estimating is the ease to which it can be extended to higher dimensions. In fact, the observation made in [5], that the error estimates are *exactly* the same in the multivariate case, is equally valid for the types of convergence considered here. If, in Theorems 1, 2 or 3, the indices p, q and r (but not n) are considered to be m-tuples of nonnegative integers, the same conclusions are valid if interpreted properly. For example, in Theorem 1, the rates of convergence for the estimate would be

$$O(n^{-1+\sum_{i=1}^{m} p_i/r_i+5/6r_i}) .$$

This could be obtained by taking $q(n) = (q_1(n), q_2(n), \cdots, q_m(n))$ as

$$q_i(n) = n^{1/r_i} \qquad i = 1, 2, \cdots, m .$$

APPENDIX

In order to calculate C_M in equation (5) we use the formula ([6], page 218)

$$\text{(a-1)} \qquad h_n(x) = \lambda_n \cos\left((2n+1)^{\frac{1}{2}}x - \frac{n\pi}{2}\right) + \frac{1}{(2n+1)^{\frac{1}{2}}} \int_0^x \sin (2n+1)^{\frac{1}{2}}(x-t)t^2 h_n(t)\, dt ,$$

where $\lambda_n = |h_n(0)|$ if n is even and $|h_n'(0)|/(2n+1)^{\frac{1}{2}}$ if n is odd. The value of λ_n when $n = 2k$ may be obtained from the recurrence formula (2) (see [11], pages 163 and 165):

$$\text{(a-2)} \qquad |h_{2k}(0)|^2 = \frac{(2k)!}{2^{2k}(k!)^2} \frac{1}{\pi^{\frac{1}{2}}} < \frac{(2/k)^{\frac{1}{2}}}{\pi} , \qquad k = 1, 2, \cdots .$$

Since $h_n'(0) = (2n)^{\frac{1}{2}} h_{n-1}(0)$, the same sort of inequality holds for odd n and hence λ_n satisfies

$$\text{(a-3)} \qquad \lambda_n \leqq (n+1)^{-\frac{1}{4}} \qquad n = 0, 1, 2 \cdots .$$

Hence by Schwarz's inequality, we have

$$\text{(a-4)} \qquad |h_n(x)| \leqq (n+1)^{-\frac{1}{4}}[1 + (n+1)^{-\frac{1}{4}}\{\textstyle\int_0^{|x|} t^4\, dt\}^{\frac{1}{2}}\{\int_0^{|x|} h_n^2\}^{\frac{1}{2}}]$$
$$\leqq (n+1)^{-\frac{1}{4}}[1 + |x|^{\frac{5}{2}}/2] ,$$

from which equation (5) follows.

REFERENCES

[1] Blum, J. and Susarla, V. (1976). A Fourier inversion method for the estimation of a density and its derivatives. To appear in *Ann. Statist.*

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

[3] Parzen, E. (1962). On the estimation of probability density and mode. *Ann. Math. Statist.* **33** 1065–1076.

[4] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 932–937.

[5] Schwartz, S. C. (1967). Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.* **38** 1261–1265.

[6] Szëgo, G. (1959). *Orthogonal Polynomials.* Amer. Math. Soc. Coll. Publ. **32**, Providence.

[7] Van Ryzin, J. (1973). A histogram method of density estimation. *Comm. Stat.* **2** 493–506.

[8] Wahba, G. (1971). A polynomial algorithm for density estimation. *Ann. Math. Statist.* **42** 1870–1886.

[9] Wahba, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

[10] Wahba, G. (1975). Interpolating spline methods for density estimation I. Equi-spaced knots. *Ann. Statist.* **3** 30–48.

[11] WALTER, G. G. (1976). Hermite series as boundary values. *Trans. Amer. Math. Soc.* **218** 155–171.

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF WISCONSIN–MILWAUKEE
MILWAUKEE, WISCONSIN 53201