JOURNAL OF MULTIVARIATE ANALYSIS 15, 174–182 (1984)

Hermite Series Estimates of a Probability Density and Its Derivatives

WŁODZIMIERZ GREBLICKI AND MIROSŁAW PAWLAK

Institute of Engineering Cybernetics, Technical University of Wrocław, Wrocław, Poland

Communicated by P. R. Krishnaiah

The following estimate of the pth derivative of a probability density function is examined: $\sum_{k=0}^{N} \hat{a}_k h_k(x)$, where h_k is the kth Hermite function and $\hat{a}_k = ((-1)^p/n) \sum_{i=1}^{n} h_k^{(p)}(X_i)$ is calculated from a sequence $X_1, \ldots, X_n$ of independent random variables having the common unknown density. If the density has r derivatives the integrated square error converges to zero in the mean and almost completely as rapidly as $O(n^{-\alpha})$ and $O(n^{-\alpha} \log n)$, respectively, where $\alpha = 2(r-p)/(2r+1)$. Rates for the uniform convergence both in the mean square and almost complete are also given. For any finite interval they are $O(n^{-\beta})$ and $O(n^{-\beta/2} \log n)$, respectively, where $\beta = (2(r-p)-1)/(2r+1)$. © 1984 Academic Press, Inc.

1. INTRODUCTION

Among a number of methods for estimating a density, the nearest neighbour, the kernel, the orthogonal series, the polynomial interpolation and the histogram ones seem to be most popular. In this paper we examine the estimate of a density and its derivatives using the orthogonal Hermite series. We estimate the density as in Schwartz [7] and Bleuez and Bosq [1], as well as in Walter [9], and suggest an estimate of derivatives of the density. For the estimate of the pth derivative, assuming that the density has r derivatives, we show that the mean integrated square error converges to zero as rapidly as $O(n^{-\alpha})$, where $\alpha = 2(r-p)/(2r+1)$, whereas the rate we obtain for the integrated square error is $O(n^{-\alpha} \log n)$ almost completely. We also study the uniform consistency in the mean square and almost complete and give the rates of the convergence. They are $O(n^{-\beta})$ and $O(n^{-\beta/2} \log n)$ a.c., respectively, where $\beta = (2(r-p)-1)/(2r+1)$. Rates of the convergence of the mean integrated square error as well as the uniform convergence in the mean square given by us are better than those known in the literature. Concerning the a.c. convergence no similar results are known to the authors.

Received January 28, 1983; revised May 23, 1983.
AMS 1980 subject classification: Primary 62G05.
Key words and phrases: Hermite series, orthogonal series, density estimate.

0047-259X/84 $3.00

Let $\{h_k\}$, $k=0, 1, 2,\ldots$, be the Hermite orthonormal system over the real line R, i.e., let

$$h_k(x) = (2^k k! \pi^{1/2})^{-1/2} H_k(x) e^{-x^2/2},$$

where

$$H_k(x) = (-1)^k e^{x^2} (d^k/dx^k) e^{-x^2}$$

is the kth Hermite polynomial.

Let $X_1,\ldots, X_n$ be a sequence of independent identically distributed random variables having the Lebesgue density f. For $p \geqslant 0$, let $f^{(p)} \in L_2$ and let $f^{(p)}$ (the p-th derivative of f) have the representation

$$f^{(p)}(x) \sim \sum_{k=0}^{\infty} a_k h_k(x),$$

where

$$a_k = \int f^{(p)}(x) h_k(x) dx = (-1)^p \int h_k^{(p)}(x) f(x) dx.$$

The second equality holds if, e.g., $f^{(p)}$ is bounded. This suggests the following estimate of $f^{(p)}(x)$:

$$\hat{f}_p(x) = \sum_{k=0}^{N} \hat{a}_k h_k(x),$$

where N depends on n and where

$$\hat{a}_k = (-1)^p n^{-1} \sum_{i=1}^{n} h_k^{(p)}(X_i)$$

unbiasedly estimates a_k.

For $p=0$, the estimate studied in this paper is the same as that examined by Schwartz [7], Walter [9] and Greblicki [3]. Walter [9] also estimated derivatives of the density, but as the estimate he took the appropriate derivative of $f_0(x)$. Defining a new estimate of derivatives we manage to get better rates of the convergence.

II. Mean Integrated Square Error

In this paper the quality of the estimate is measured by the integrated square error, i.e., with

$$\int (\hat{f}_p(x) - f^{(p)}(x))^2 dx = \sum_{k=0}^{N} (\hat{a}_k - a_k)^2 + \sum_{k=N+1}^{\infty} a_k^2. \tag{1}$$

Defining $h_{-k} = h_k$, we have

$$h_k'(x) = (|k|/2)^{1/2}\, h_{k-1}(x) - ((|k|+1)/2)^{1/2}\, h_{k+1}(x),$$

$k = \pm 1, \ \pm 2, \ldots,$ and

$$h_0'(x) = \tfrac{1}{2}^{1/2} h_{-1}(x)$$

(see Szegö [8, p. 106]). From this follows

$$h_k^{(p)}(x) = \sum_{j=-p}^{p} \alpha_{kpj} h_{k+j}(x), \tag{2}$$

where

$$|\alpha_{kpj}| \leqslant K_p(|k|+p)^{p/2}. \tag{3}$$

By (2), $\hat{a}_k$ can be rewritten in the form

$$\hat{a}_k = (-1)^p\, n^{-1} \sum_{j=-p}^{p} \alpha_{kpj} \sum_{i=1}^{n} h_{k+j}(X_i). \tag{4}$$

THEOREM 1. *Let $f^{(p)} \in L_2$, $p \geqslant 0$. If*

$$N \xrightarrow{n} \infty, \tag{5}$$

$$N^{p+5/6}/n \xrightarrow{n} 0, \tag{6}$$

then

$$E \int (\hat{f}_p(x) - f^{(p)}(x))^2\, dx \xrightarrow{n} 0.$$

Proof. By (3),

$$\operatorname{var} \hat{a}_k \leqslant (2p+1)\, K_p^2 (k+p)^p\, n^{-1} E \left\{ \sum_{j=-p}^{p} h_{k+j}^2(X) \right\}. \tag{7}$$

Hence, by an inequality

$$\max_x |h_k(x)| \leqslant C(k+1)^{-1/12} \tag{8}$$

(see Szegö [8, Theorem 8.91.3]),

$$\operatorname{var} \hat{a}_k \leqslant (2p+1)^2\, K_p^2 n^{-1} (k+p)^p\, (k+p+1)^{-1/6}.$$

Finally,

$$\sum_{k=0}^{N} E(\hat{a}_k - a_k)^2 = O(N^{p+5/6}/n).$$

Recalling (1) and (5) we complete the proof.

The next theorem says how rapidly the mean integrated square error converges to zero. Let us introduce the following notation:

$$t_r(x) = (x - d/dx)^r f(x).$$

THEOREM 2. *Let* $f^{(p)} \in L_2$, $r > p$ *and* $t_r \in L_2$. *Moreover, let* $E|X|^{2/3} < \infty$. *If*

$$N \sim n^{2/(2r+1)}, \tag{9}$$

then

$$E \int (\hat{f}_p(x) - f^{(p)}(x))^2 \, dx = O(n^{-2(r-p)/(2r+1)}).$$

Proof. Let us observe that if $E|X|^{s/3} < \infty$, $s > 0$,

$$E|h_k(X)|^s \leqslant c(k+1)^{-s/4}. \tag{10}$$

This is implied by the two inequalities

$$\max_{|x| \leqslant a} |h_k(x)| \leqslant C_a (k+1)^{-1/4}, \tag{11}$$

for any nonnegative a, and

$$\max_{|x| \geqslant a} |x^{-1/3} h_k(x)| \leqslant D_a (k+1)^{-1/4},$$

for any positive a. The inequalities are, in turn, implied by Theorem 8.91.3 in Szegö [8].

By (7) and (10),

$$\sum_{k=0}^{N} E(\hat{a}_k - a_k)^2 = O(N^{p+1/2}/n). \tag{12}$$

On the other hand, by virtue of Walter's [9] result,

$$\sum_{k=N+1}^{\infty} a_k^2 = O(N^{-(r-p)}). \tag{13}$$

The theorem is now a consequence of (1), (9), (12) and (13).

The rate given by us is considerably better than $O(n^{-(6(r-p)-5)/6r})$ reported by Walter [9]. For example, for $p=0$ and $r=1$, our $O(n^{-2/3})$ is better than his $O(n^{-1/6})$. The rate in Theorem 2 is even better than $O(n^{-(2(r-p)-1)/2r})$, i.e., that derived by Walter [9] for densities having bounded support. Hall's [4] results are not comparable since he estimated densities on the half real line.

III. Integrated Square Error

In this section we examine the almost complete convergence of the integrated square error and study the rate of the convergence. The authors would like to mention that no result concerning the almost sure convergence of the error for the estimate is known to them.

THEOREM 3. *Let $f^{(p)} \in L_2$. If, in addition to (5),*

$$\sum_{n=1}^{\infty} \exp(-\alpha N^{p+5/6}/n) < \infty, \tag{14}$$

for all positive α, then

$$\int (\hat{f}_p(x) - f^{(p)}(x))^2 \, dx \xrightarrow{n} 0 \qquad \text{a.c.}$$

Remark 1. Condition (14) is satisfied if

$$N^{p+5/6} \log n/n \xrightarrow{n} 0.$$

Now imposing some smoothness restrictions on the density and assuming an appropriate moment to exist, we give the rate of the convergence. For a sequence $\{Y_n\}$ of random variables, we say that $Y_n = O(a_n)$ a.c. if $\beta_n Y_n/a_n \to 0$ a.c. as $n \to \infty$, for all sequences $\{\beta_n\}$ convergent to zero.

THEOREM 4. *Let $f^{(p)} \in L_2$, $p \geqslant 0$. Let $r > p$, and let $t_r \in L_2$. Let, moreover, $E|X|^s < \infty$, $s > 8(r+1)/3(2r+1)$. If (9) is satisfied, then*

$$\int (\hat{f}_p(x) - f^{(p)}(x))^2 \, dx = O(n^{-2(r-p)/(2r+1)} \log n) \qquad a.c.$$

Remark 2. The restriction in Theorem 4 concerning the existence of $E|X|^s$ is fulfilled for $s > 16/9$ independently of r.

We shall now prove Theorem 4. The proof of Theorem 3 is omitted since it is similar. The only difference is that S_{Np} should be defined as $\sum_{k=0}^{N} \sum_{j=-p}^{p} (k+|j|+1)^{-1/6} \alpha_{kpj}^2$ and that (8) instead of (10) and Hoeffding's [5] instead of Fuc and Nagaev's inequality should be used.

Proof of Theorem 4. By virtue of (9) and (13) it suffices to verify that, for all positive t and all sequences $\{\beta_n\}$ convergent to zero,

$$\sum_{n=1}^{\infty} P\left\{\sum_{k=0}^{N} (\hat{a}_k - a_k)^2 > t/\gamma_n\right\} < \infty, \tag{15}$$

where $\gamma_n = \beta_n n^{\alpha}/\log n$ and $\alpha = 2(r-p)/(2r+1)$. It is clear that the probability in (15) is not greater than

$$P\left\{\sum_{k=0}^{N} \left[\sum_{j=-p}^{p} \alpha_{kpj} n^{-1} \sum_{i=1}^{n} Y_{k+j}(X_i)\right]^2 > t/\gamma_n\right\}$$
$$\leqslant P\left\{\sum_{k=0}^{N} \sum_{j=-p}^{p} \left[\alpha_{kpj} n^{-1} \sum_{i=1}^{n} Y_{k+j}(X_i)\right]^2 > t/(2p+1)\gamma_n\right\}, \tag{16}$$

where $Y_k(X_i) = h_k(X_i) - Eh_k(X_i)$. In turn, the probability in (16) does not exceed

$$\sum_{k=0}^{N} \sum_{j=-p}^{p} P\left\{n^{-1}\left[\sum_{i=1}^{n} Y_{k+j}(X_i)\right]^2 > t(k+|j|+1)^{-1/2}/\gamma_n(2p+1)^2 S_{Np}\right\}, \tag{17}$$

where, by (3),

$$S_{Np} = \sum_{k=0}^{N} \sum_{j=-p}^{p} (k+|j|+1)^{-1/2} \alpha_{kpj}^2 \leqslant c(N+1)^{p+1/2}. \tag{18}$$

Applying Fuc and Nagaev's inequality [2, Corollary 4] we find the probability in (17) dominated by $A_{kjn} + B_{kjn}$, where

$$A_{kjn} = c_1 \gamma_n^{q/2} n^{1-q} (k+|j|+1)^{q/4} S_{Np}^{q/2} E\,|h_{k+j}(X)|^q / t^{q/2}$$

and

$$B_{kjn} = 2\exp\{-c_2 tn/\gamma_n (k+|j|+1)^{1/2} S_{Np} E h_{k+j}^2(X)\},$$

c_1 and c_2 positive and independent of t, k and j, where $q = 3s$. Using (10) and (18), we get

$$A_n = \sum_{k=0}^{N} \sum_{j=-p}^{p} A_{kjn} = O(n^{\gamma}),$$

where $\gamma = (4r - 2rq - q + 6)/2(2r + 1)$. As $\gamma < -1$, $\sum_{n=1}^{\infty} A_n < \infty$. In order to verify $\sum_{n=1}^{\infty} B_n < \infty$, where $B_n = \sum_{k=0}^{N} \sum_{j=-p}^{p} B_{kjn}$, it suffices to notice that, by (10) and (18), $B_{kjn} \leqslant 2\exp(-c_3 tn \log n/\beta_n)$, c_3 positive and independent of t, k and j. The proof has been completed.

IV. Uniform Convergence

Let $f^{(p)} \in L_\infty \cap L_2$ and let $f^{(p)}$ be of bounded variation on every finite interval. Hence, by virtue of equiconvergence Theorem 9.1.6 in Szegö [8] and the Dirichlet–Jordan theorem on the convergence of the Fourier series, see Sansone [6], $\sum_{k=0}^{n} a_k h_k(x) \to f^{(p)}(x)$ as $n \to \infty$ uniformly on every finite interval. Therefore, an application of (11) leads to

$$\begin{aligned} |\hat{f}^p(x) - f^{(p)}(x)| &\leqslant \left| \sum_{k=0}^{N} (\hat{a}_k - a_k)\, h_k(x) \right| + \left| \sum_{k=N+1}^{\infty} a_k h_k(x) \right| \\ &\leqslant \left[\sum_{k=0}^{N} (\hat{a}_k - a_k)^2 \sum_{k=0}^{N} h_k^2(x) \right]^{1/2} + \left| \sum_{k=N+1}^{\infty} a_k h_k(x) \right| \\ &\leqslant 2C_a(N+1)^{1/4} \left(\sum_{k=0}^{N} (\hat{a}_k - a_k)^2 \right)^{1/2} \\ &\quad + C_a \sum_{k=N+1}^{\infty} |a_k|\,(k+1)^{-1/4} \end{aligned} \tag{19}$$

on the interval $[-a, a]$. In order to estimate the second term in (19) we use the inequality $a_k^2 \leqslant k^{-(r-p)} b_{k+r-p}^2$ derived by Walter [9] under the condition $t_r \in L_2$; b_k is the kth coefficient of the expansion of t_r in the Hermite series. Thus, the second term in (19) is upper bounded by

$$\begin{aligned} &c_1 \sum_{k=N+1}^{\infty} |b_{k+r-p}|\,(k+1)^{-(r-p)/2-1/4} \\ &\quad \leqslant c_2 \left[\sum_{k=N+1}^{\infty} (k+1)^{-(r-p)-1/2} \right]^{1/2} \sum_{k=N+1}^{\infty} b_k^2 = O(N^{-(r-p)/2+1/4}), \end{aligned}$$

where c_1 and c_2 are some positive constants.

The next theorem can now be verified by using tricks as in previous sections and will be given without proof.

THEOREM 5. *Let $f^{(p)} \in L_\infty \cap L_2$, $p \geqslant 0$, and let $f^{(p)}$ be of bounded variation on finite intervals. Let $r > p$, and let $t_r \in L_2$. Let, moreover, $\{N\}$ be selected according to* (9). *Let $E|X|^s < \infty$. Then, for $s = \frac{2}{3}$,*

$$E\{\sup_{|x| \leqslant a} |\hat{f}_p(x) - f^{(p)}(x)|^2\} = O(n^{-(2(r-p)-1)/(2r+1)}),$$

and, for $s > 8(r+1)/3(2r+1)$,

$$\sup_{|x| \leqslant a} |\hat{f}_p(x) - f^{(p)}(x)| = O(n^{-(2(r-p)-1)/2(2r+1)} \log n) \qquad \text{a.c.}$$

The uniform almost sure convergence of densities has been studied by Bleuez and Bosq [1]. As far as the rate for the almost sure convergence, the result given by us is better than Walter's [10] $O(n^{-((r-p)-3/4)/2(r+3)})$. For example, for $p = 0$ and $r = 1$, his $O(n^{-1/32})$ is worse than our $O(n^{-1/6} \log n)$.

V. MULTIDIMENSIONAL GENERALIZATION

Schwartz [7] has observed that generalization results concerning orthogonal series estimates on higher dimensions are easy. Denoting by d the dimension of X and recognizing k in the definition of the definition of the estimate as a d-vector index one can verify that, e.g., Theorem 1 remains valid if (6) is replaced by

$$N^{(p+5/6)d}/n \xrightarrow{n} 0.$$

REFERENCES

[1] BLEUEZ, J., AND BOSQ, D. (1976). Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales. *C. R. Acad. Sci. Paris* **282** 1023–1026.

[2] FUC, D. H., AND NAGAEV, S. V. (1971). Probability inequalities for sums of independent random variables. *Teor. Veroyatnost. i Primenen.* **16** 660–675.

[3] GREBLICKI, W. (1981). Asymptotic efficiency of classifying procedures using the Hermite series estimate of multivariate probability densities. *IEEE Trans. Inform. Theory* **27** 364–366.

[4] HALL, P. (1980). Estimating a density on the positive half line by the method of orthogonal series. *Ann. Inst. Statist. Math.* **32** 351–362.

[5] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[6] SANSONE, G. (1959). *Orthogonal Functions*. Interscience, New York.

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

[8] SZEGÖ, G. (1939). *Orthogonal Polynomials*. Amer. Math. Soc. Coll. Publ., Providence, R. I.

[9] WALTER, G. G. (1977). Properties of Hermite series estimation of probability density. *Ann. Statist.* **5** 1258–1264.

[10] WALTER, G. G. (1980). Addendum to "Properties of Hermite series estimation of probability density." *Ann. Statist.* **8** 454–455.