POINTWISE CONSISTENCY OF THE HERMITE SERIES DENSITY ESTIMATE

Włodzimierz GREBLICKI and Mirosław PAWLAK
Technical University of Wrocław

printed in:
Statistics & Probability Letters
Vol. 3, No. 2, pp. 65–69, April 1985

Abstract: The Hermite series estimate of a density $f \in L_p$, $p > 1$, converges in the mean square to $f(x)$ for almost all $x \in R$, if $N(n) \to \infty$ and $N(n)/n^2 \to 0$ as $n \to \infty$, where N is the number of the Hermite functions in the estimate while n is the number of observations. Moreover, the mean square and weak consistency are equivalent. For m times differentiable densities, the mean square convergence rate is $O(n^{-(2m-1)/2m})$. Results for complete convergence are given.

Key words: density estimate, nonparametric, orthogonal series, Hermite series.

1. Introduction

We examine asymptotic pointwise properties of the Hermite series estimate of probability density functions. The usage of an orthogonal system to estimate densities was proposed by Čencov (1962) and developed by a number of authors. The Fourier series was employed by e.g. Kronmal and Tarter (1968), Wahba (I975), Bleuez and Bosq (1976, 1979) as well as Hall (1981). Although the trigonometric series estimate is very easy to compute, the Hermite series one is attractive due to some reasons. First of all we can estimate densities defined on the whole real line. When the unknown density is close to normal it is reasonable to hope that the density quickly goes to the density and a small number of terms in the estimate is desirable. In fact, the Hermite functions have the main contribution from the normal density and their tails behave normally. Finally, the Hermite series estimate is more robust, in some sense, then the Fourier one as it was observed by Good and Gaskins (1980).

A number of authors, in particular Schwartz (1967), Walter (1977, 1980), Bleuez and Bosq (1976, 1979), Hall (1980) as well as Greblicki and Pawlak (1985), have investigated the behavior of the estimate based on the Hermite series. In this paper we examine pointwise convergence of the estimate and assuming that the unknown density belongs to L_p, $p > 1$, we give conditions under which the estimate converges to the density in probability, in the mean square as well as completely at almost all points. Combining this with results given earlier by Bleuez and Bosq (1976) we formulate conditions which are both necessary and sufficient for the in probability, the mean square and complete convergence, for densities in L_p, $p > 1$. The first two types of consistency are, moreover, equivalent. The bound on p is as wide as possible, i.e., the restriction $p > 1$ can not be weakened to $p = 1$. We also considerably improve the rate of the convergence given by Walter (1971). The rate is the same as the optimal one obtained by Wahba (1975) for the estimate with the cosine series despite the fact that instead her smooth conditions near the endpoints we improve some growth conditions on the density and its derivatives.

2. The estimate

Let $(X_1, \ldots, X_n)$, be a sample of independent observations of a random variable X having the Lebesgue density f. Let, for $k = 0, 1, 2 \ldots$,

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

where

$$H_k(x) = e^{x^2} \frac{d^k}{dx^k} e^{-x^2}$$

is the k-th Hermite polynomial. The estimate of $f(x)$ considered in this paper is of the following form:

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

where N depends on n and

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

is the estimate of $a_k = E h_k(X)$. Denoting

$$d_N(x, y) = \sum_{k=0}^{N} h_k(x) h_k(y),$$

we can rewrite the estimator in the following equivalent form:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} d_N(x, X_i).$$

Estimating the mean square consistency we assume that

$$N(n) \xrightarrow{n} \infty \tag{1}$$

and

$$\frac{N^{1/2}(n)}{n} \xrightarrow{n} 0, \tag{2}$$

while complete convergence needs another condition:

$$\sum_{n=1}^{\infty} \exp\left(-\alpha \frac{n}{N^{1/2}(n)}\right) < \infty, \text{ all } \alpha > 0. \tag{3}$$

3. Lemmas

In our considerations we apply the following three lemmas:

Lemma 1.

$$\lim_{n \to \infty} a_k h_k(x) = f(x)$$

at every differentiability point of f. If $f \in L_p$, $p > 1$, the convergence holds for almost all $x \in \mathbb{R}$.

Proof. By the equiconvergence theorem, see Szegö (1959, p. 247), the deference between the n-th partial sum $s_n(x)$ of the expansion of f in the Hermite series and the n-th partial sum of the trigonometric expansion of f taken any finite interval containing x in its interior converges to zero as n tends to infinity. By this and Dini's theorem on pointwise convergence of trigonometric expansions, see, e.g., Sansone (1959), $s_n(x) \to f(x)$ as $n \to \infty$ at every differentiability point of f. The convergence holds also at almost all $x \in \mathbb{R}$. This is, in turn, implied by the equiconvergence theorem and a theorem which says that the trigonometric expansion of any function belonging to L_p, $p > 1$, converges to the function at almost all points, see Carleson (1966) for $p = 2$ and Hunt (1968) for $p > 1$.

Lemma 2. *For every $x \in \mathbb{R}$*

$$\max_y |d_N(x, y)| \leq c(x)(N + 1)^{1/2}.$$

Proof. Let $a > 0$. For any fixed $x \in \mathbb{R}$, we have

$$\max_{|x-y| \leq a} |d_N(x, y)| = O(n^{1/2}). \tag{5}$$

It is implied by an inequality

$$\max_{|x| \leq a} |h_k(x)| \leq c_1(\alpha)(k + 1)^{-1.4} \tag{6}$$

true for any nonegative α, see Szegö (1959, p. 242).

In turn, by virtue of Christoffel's formula

$$d_N(x, y) = (N + 1)^{1/2} \frac{h_{N+1}(x) h_N(y) - h_{N+1}(y) h_N(x)}{2^{1/2}(x - y)}.$$

By this and the inequality

$$\max_x |h_k(x)| \leq c(k + 1)^{-1/12}, \tag{7}$$

see Szegö (1959, p. 242), we get

$$\max_{|x-y|>a} |d_N(x,y)| \le c^2 a^{-2}(N+1)^{1/6} = O(N^{1/6}). \tag{8}$$

The lemma is a consequence of (5) and (8).

Lemma 3. var $d_n(x, X) \le d(x)(N+1)^{1/2}$ *at every Lebesgue point of* f *(in particular, at every continuity point of* f *and also at almost every* $x \in \mathbb{R}$*).*

Proof. Clearly,

$$n\text{var }\hat{f}_n(x) \le \int d_N^2(x,y)f(y)dy \tag{9}$$

for any fixed $x \in \mathbb{R}$. Using (6), (7) and Chrostoffel's formula yields

$$\begin{aligned}&\int_{|x-y|>\delta} d_N^2(x,y)f(y)dy \\ &\le \max_{|x-y|>\delta} d_N(x,y) \le \frac{c^2 c_1(|x|)}{\delta^2}(N+1)^{1/3} \\ &= O(N^{1/3})\end{aligned} \tag{10}$$

for any $\delta > 0$. In Sansone (1959, p. 337) we find

$$\pi d_N(x,y) = \frac{\sin[\mu(x-y)]}{x-y} + \frac{1}{\mu}T_N(x,y), \tag{11}$$

where $2\mu = (2N+1)^{1/2} + (2N+3)^{1/2}$ and where T_N is bounded by a constant independent of N, for x and y varying in finite intervals. Thus

$$\int_{|x-y|\le\delta} T_N^2(x,y)f(y)dy \le c_2, \tag{12}$$

c_2 independent of N. Moreover,

$$\begin{aligned}&\frac{1}{\mu\pi}\int_{|x-y|\le\delta} \frac{\sin^2[\mu(x-y)]}{(x-y)^2}f(y)dy \\ &\le \frac{1}{\mu\pi}\int_{|x-y|\le\delta}\left[\frac{\sin^2[\mu(x-y)]}{\sin^2(x-y)}\right]f(y)dy\end{aligned} \tag{13}$$

for $\delta \le \pi/2$. Let us observe that the function in brackets is the Fejér kernel. Therefore be virtue of the Lebesgue theorem on Cesaro summability of trigonometric series, the quantity in (13) converges to $f(x)$ as $n \to \infty$ at all Lebesgue points of f. From this (11), and (12),

$$\int_{|x-y|\le\delta} d_N^2(x,y)f(y)dy = O(N^{1/2})$$

at all Lebesgue points of f, whenever $\delta \le \pi/2$. Recalling (9) and (10) we complete the proof.

4. Consistency

From Lemmas 1 and 3 we easily get

Theorem 1. *Let* $f \in L_p$, $p > 1$. *If (1) and (2) hold, then*

$$E(\hat{f}(x) - f(x))^2 \stackrel{n}{\to} 0 \textit{ for almost all } x \in \mathbb{R}.$$

It should be mentioned that Bleuez and Bosq (1976) as well as Földes and Revész (1974) obtained a similar result but only for $p = \infty$. Condition $p >1$ in Theorem 1 can not be weakened to $p > 1$, due to Kolmogorov's counterexample which says that there exist densities whose Fourier and consequently Hermite series diverge for all $x \in \mathbb{R}$, see Zygmund (1959, Section 8.4). For these densities, the bias unfortunately does not converge to zero as n tends to infinity for all $x \in \mathbb{R}$.

Using Theorem 1, Lemma 1 and the Lebesgue density theorem it is easy to see that Lemma 1 in Bleuez and Bosq (1979) holds for almost all x, provided that $f \in L_p$, $p > 1$. From this we get a corollary on the equivalence of weak and mean square consistency.

Corollary 1. *The following conditions are equivalent:*

(i) $\hat{f}(x) \stackrel{n}{\to} f(x)$ *in probability for almost all* $x \in \mathbb{R}$ *and all* $f \in L_p$, $p > 1$,

(ii) $E(\hat{f}(x) - f(x))^2 \stackrel{n}{\to} 0$ *for almost all* $x \in \mathbb{R}$ *and all* $f \in L_p$, $p > 1$,

(iii) conditions (1) and (2).

In the next theorem we give conditions for complete convergence.

Theorem 2. *Let* $f \in L_p$, $p > 1$. *If (1) and (3) hold, then*

$$\hat{f}(x) \stackrel{n}{\to} f(x) \textit{completely for almost all } x \in \mathbb{R}.$$

It is clear that for reasons as in Theorem 1, restriction $p>1$ can not be weakened to $p \geq 1$. For $p=\infty$, a similar result was obtained by Bleuez and Bosq (1976). Combining Theorem 2 with Proposition 5 in that paper and making use of Lemma 1, we get

Corollary 2. *$\hat{f}(x) \overset{n}{\to} f(x)$ completely for almost all* $x \in \mathbb{R}$ and all $f \in L_p$, *$p>1$, if and only id (1) and (3) are satisfied.*

Proof of Theorem 2. Taking into account Lemmas 2 and 3 and using Bernstein's inequality, see e.g. Hoeffding (1963), we get

$$P\left\{|\hat{f}(x) - E\hat{f}(x)| > t\right\}$$
$$\leq 2\exp\left(-\frac{nt^2}{2\,(d(x)+tc(x))\,(N+1)^{1/2}}\right).$$

In view of this and Theorem 1, the proof has been completed.

5. Rates of convergence

Let us introduce the following notation:

$$t_m(x;f) = \left(x - \frac{d}{dx}\right)^m f(x).$$

Walter (1977) has shown that if $t_m(\cdot;f) \in L_2$,

$$a_k^2 \leq \frac{b_{k+m}^2}{(k+1)^m},$$

where b_k is the k-th coefficient of the expansion of $t_m(\cdot;f)$.

Theorem 3. *Let $f \in L_2$ and $t_m(\cdot;f) \in L_2$. If*

$$N(n) \sim n^{1/m}, \tag{15}$$

then

$$E(\hat{f}(x) - f(x))^2 = O(n^{-(2m-1)/2m}) \tag{16}$$

and

$$|\hat{f}(x) - f(x)| = O(n^{-(2m-1)/4m}\log n) \tag{17}$$

completely.

The rate in (16) considerably improves results given by Walter (1977) who, under conditions as in Theorem 3, showed that the mean square error converges to zero as rapidly as $O(n^{-(m-1)/m})$. Thus, for e.g. $m=1$, the rate presented by him is useless, while that obtained by us equals $O(n^{-1/2})$.

Proof of Theorem 3. By virtue of Lemma 1, (6) and (14), the bias is not greater than

$$\sum_{k=N+1}^{\infty} a_k h_k(x)$$
$$\leq c_1(|x|) \sum_{k=N+1}^{\infty} |b_{k+m}|(k+1)^{-m/2-1/4}$$
$$\leq c_1(|x|)\,\|t_m(\cdot;f)\| \left[\sum_{k=N+1}^{\infty} (k+1)^{-m-1/2}\right]^{1/2}$$
$$= O(N^{-m/2+1/4}),$$

where $\|\cdot\|$ is the L_2 norm. In turn, by virtue of Lemma 3, $\mathrm{var}\hat{f}_n(x) = O(N^{1/2}/n)$. Now, (16) can be verified with ease. Since one can prove (17) in the same way, the proof has been completed.

References

[1] J. Bleuez and D. Bosq (1976),.Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales, *C. R. Sc. Paris* **282**, 1023-1026.

[2] J. Bleuez and D. Bosq (1979),.Conditions nécessaires et suffisantes de convergence de la densité par la méthode des fonctions orthogonales, *Rev. Roumaine Math. Pures et Appl.* **24**, 869-886.

[3] N. N. Čencov (1962), Evaluation of an unknown distribution density from observations, *Soviet. Math.* **3**, 1559-1562.

[4] L. Carleson (1966), On convergence and growth of partial sums of Fourier series. *Acta Math.* **116**, 135-157.

[5] A. Földes and P. Révesz (1974), A general method for density estimation, *Sudia Sci. Math. Hungar.* **9**, 81-92.

[6] I. J. Good and R. A. Gaskins (1980), Density estimation and bumphunting by the penalized likelihood method exemplified by scatteling and meteorite data, *J. Amer. Statist. Assoc.* **75**, 42-73.

[7] W. Greblicki and M. Pawlak (1985), Hermite series estimates of a probability density and its derivatives, *J. Multivariate Anal.*, to appear.

[8] P. Hall (l980), Estimating a density on the positive half line by the method of orthogonal series, *Ann. Inst. Statist. Math.* **32**, 351-362.

[9] P. Hall (1981), On trigonometric series estimates of densities, *Ann. Statist.* **9**, 683-685.

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

[11] R. A. Hunt (1968), On the convergence of Fourier series, in: *Orthogonal Expansions and their Continuous Analogues*, Proc. Conf. Edwardsville (Southern Illinois Univ. Press., Carbondale) pp. 13-30.

[12] R. Kronmal and M. Tarter (l968), The estimation of probability densities and cumulatives by Fourier series methods, *J. Amer. Statist. Assoc.* **63**, 925-952.

[13] G. Sansone (1959), *Orthogonal Functions* (Interscience Publishers Inc.).

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

[15] G. Szegö (1959), *Orthogonal Polynomials* (Amer. Math. Soc. Coll. Pub.).

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

[17] G. G. Walter (1977), Properties of Hermite series estimates of probability densities, *Ann. Statist.* **8**, 1258-1264.

[18] G. G. Walter (1980), Addendum to "Properties of Hermite series estimates of probability densities", *Ann. Statist.* **8**, 454-455.

[19] A. Zygmund (1959), *Trigonometric Series, Vol. 1, 2* (Cambridge University Press).