JOURNAL OF APPROXIMATION THEORY 18, 307–315 (1976)

The L_2 Norm of the Approximation Error for Bernstein Polynomials

DOUGLAS H. JONES*

Department of Statistics, Rutgers University, New Brunswick, New Jersey 08903

Communicated by Oved Shisha

Received February 10, 1975

Wassily Hoeffding (*J. Approximation Theory* 4 (1971), 347–356) obtained a convergence rate for the L_1 norm of the approximation error, using Bernstein polynomials for a wide class of functions. Here, by a different method of proof, a similar result is obtained for the L_2 norm.

1. Introduction and Summary

Let φ be a real-valued function on (0, 1). Following Hoeffding [7], we define the related Bernstein polynomial of degree $N - 1$ as

$$B_N\varphi(t) = \sum_{i=0,N-1} \varphi((i+1)/(N+1))\, p_{N-1,i}(t), \tag{1.1}$$

where

$$p_{N-1,i}(t) = \binom{N-1}{i} t^i(i-t)^{N-1-i}. \tag{1.2}$$

This definition is a slight modification of the usual one and it allows functions that are unbounded at 0 and 1.

Throughout this paper we will assume that φ is absolutely continuous on (0, 1) with a continuous derivative ψ. The derivative must satisfy part (a) of the following condition T at 0 and 1.

CONDITION T. We say a function $\psi(u)$ satisfies condition T at a point $p \in [0, 1]$ if (a) for any $\epsilon > 0$ there exist $\tau > 0$ and $0 < q < 1$ such that for any $u_1, u_2 \in (0, 1)$ satisfying $0 < q(p - u_2) < p - u_1 < p - u_2 \leqslant \tau$ or $0 < q(u_2 - p) < u_1 - p < u_2 - p \leqslant \tau$, $|\psi(u_2)/\psi(u_1) - 1| \leqslant \epsilon$ holds, and

* Research supported by a Summer Fellowship from the Research Council, Rutgers University.

(b) there exist $\gamma > 0$, $M > 0$, $a \geqslant 0$ such that for $0 < | p - u | < \gamma$, $u \in (0, 1)$, $\psi(u) \geqslant M | u - p |^a$ holds.

Part (a) of the condition says that ψ may either approach infinity or zero at 0 and 1 provided it does not vary too widly. Part (b) says $\psi(u)$ cannot approach zero faster than some powre of u as $u \to 0$ or 1. In particular, (a) is satisfied if for some $\epsilon > 0$, $\psi(u) = | p - u |^b$ for $0 < | p - u | < \epsilon$ and some finite b for $p = 0, 1$; or if $\lim_{u \uparrow 1} \psi(u)$ and $\lim_{u \downarrow 0} \psi(u)$ exist and are finite and nonzero, or more generally if $\psi(u)$ is regularly varying in the sense of Karamata as $u \uparrow 1$ and $u \downarrow 0$ (see [2] or [4]). We have taken the above definition and comments of Stigler [10], in order that we may follow his proofs for ψ satisfying both parts (a) and (b), then upon making an observation we drop part (b).

We give separate results for bounded and unbounded functions φ. We give the bounded case first.

THEOREM 1. *Let φ be an absolutely continuous function on* $(0, 1)$ *with derivative ψ. Suppose φ and ψ satisfy the conditions*

(i) *φ is bounded on* $[0, 1]$ *and there exist $\delta > 0$ and $t_0 \in (0, 1)$ such that* $| \varphi(t) - \varphi(0)| \leqslant Kt^\delta$, $| \varphi(t) - \varphi(1)| \leqslant K(1 - t)^\delta$ *for* $0 \leqslant t \leqslant t_0$,

(ii) *ψ is continuous on* $(0, 1)$, *satisfies condition* T, *part* (a) *at* 0 *and* 1, *and*

(iii) $J_2(\varphi) = \int [\psi(t)]^2 \, t(1 - t) \, dt < \infty$.

Then

$$\int [B_N\varphi(t) - \varphi(t)]^2 \, dt \leqslant N^{-1} J_2(\varphi) + o(N^{-1}). \tag{1.3}$$

DEFINITION. Define Φ to be the class of functions φ specified in the hypothesis of Theorem 1.

Hoeffding [7, Theorem 3] proves the related result for the constant $C = (2/e)^{1/2}$, $\int | B_N\varphi(t) - \varphi(t)| \, dt \leqslant N^{-1/2} J_1(\varphi) C + O(N^{-1})$, where $J_1(\varphi) = \int [t(1 - t)]^{1/2} \, | d\varphi |$. In Theorem 4 of the same paper, he obtains an asymptotic equality sharpening the last inequality, with $C = (2/\pi)^{1/2}$, for any step function φ of bounded variation in $[0, 1]$ having finitely many steps in every closed subinterval of $(0, 1)$. Furthermore, the equality holds irrespective of whether $J_1(\varphi)$ is finite or not.

We cannot prove similar results for the L_2 norm of the error. That is, there are no functions φ contained in Φ for which there is equality in (1.3). Also, we do not know if the finiteness of $J_2(\varphi)$ is needed for the L_2 norm to be of order $N^{-1/2}$. It is interesting to note that our method of proof can be applied to the L_1 norm to produce the quantity $J_1(\varphi)$ in the asymptotic limit given by Hoeffding [7, Theorem 4]. For this reason, we feel that the quantity $J_2(\varphi)$ is the correct functional for the bound in (1.3). While our method of proof can give results for any L_p norm with $p \geqslant 1$, that of Hoeffding's paper apparently cannot be extended beyond the L_1 norm.

The following theorem shows that the rate of N^{-1} in (1.3) cannot be replaced by a faster one for the functions $\varphi \in \Phi$ with $\| \varphi \|_2 \leqslant 1$. (I thank Professor R. H. Berk of Rutgers University for kindly formulating and outlining the proof of Theorem 2.)

THEOREM 2. *Let $\varphi \in \Phi$ such that $\| \varphi \|_2 \leqslant 1$. Then*

$$\sup_{\varphi \in \Phi} \| B_N\varphi - \varphi \|_2 \geqslant (2\pi)^{-1/2} N^{-1/2} + o(N^{-1/2}). \tag{1.4}$$

The next theorem applies to unbounded functions.

THEOREM 3. *Let* (i′) *replace* (i) *of Theorem* 1.

(i′) *There exists an s such that $0 < s \leqslant 2$ for which there exists K, $\delta > 0$ such that*

$$[\varphi(t)]^2 \leqslant K[t(1-t)]^{-1+s/2+\delta}, \qquad 0 < t < 1.$$

Then

$$\int [B_N\varphi(t) - \varphi(t)]^2 \, dt = o(N^{-s/2}). \tag{1.5}$$

A theorem similar to Theorem 2 may be given for the class of functions examined in Theorem 3.

In Section 2 we prove that the error term is bounded above by a quantity that is much studied in the theory of rank and order statistics. In Section 3 we present some lemmas that give the formula for the asymptotic limit of the variance of order statistics. We give the proofs of Theorems 1, 2, and 3 in Section 4. In Section 5 we make some remarks about the extension of the method used for the L_2 norm to obtain a bound on the L_p norm with $p > 1$.

2. RELATION TO RANK AND ORDER STATISTICS

In this paper we wish to use certain probabilistic techniques peculiar to the theory of order and rank statistics. Suppose $U_1, U_2, \ldots$ is a sequence of independent random variables each having density $f(u) = 1$, $0 \leqslant u \leqslant 1$, $= 0$ otherwise. And let R_1 denote the rank of U_1 among the partial sequence $U_1, \ldots, U_N$, for each $N \geqslant 2$ (we suppress the notational dependence on N in R_1). Further, define $V_N(t)$ to be a binomial random variable with parameters N, t:

$$P(V_N(t) = i) = p_{N,i}(t),$$

where $p_{N,i}(t)$ is defined by (1.2).

For a probability space $(\Omega, \mathcal{A}, \mu)$ and a measurable real-valued function X we define the expectation of X as

$$EX = \int_{\Omega} X(\omega)\, \mu(d\omega)$$

whenever it exists. Consequently we may rewrite (1.1) as

$$B_N\varphi(t) = E\varphi((V_{N-1}(t) + 1)/(N + 1)). \tag{2.1}$$

Using the well-known fact (see [8])

$$P(R_1 = i \,\|\, U_1 = t) = P(V_{N-1}(t) + 1 = i), \qquad i = 1,\dots, N,$$

we may subsequently write (2.1) as a conditional expectation,

$$E\varphi((V_{N-1}(t) + 1)/(N + 1)) = E\{\varphi(R_1/(N + 1)) \,\|\, U_1 = t\}. \tag{2.2}$$

Conditional expectations and their properties are discussed in [3].

Thus properties of Bernstein polynomials may be determined by the use of techniques developed for rank and order statistics. In Theorem 4, the L_2-norm degree of approximation is seen to be bounded by a familiar quantity in statistics; for instance see [5].

THEOREM 4. *Let* $\int | \varphi |^2 < \infty$. *Then*

$$\int | B_N\varphi(t) - \varphi(t)|^2\, dt \leqslant E[\varphi(U_1) - \varphi(R_1/(N + 1))]^2. \tag{2.3}$$

Proof. From (2.1) and (2.2) above we see that upon using the properties of the conditional expectation

$$\begin{aligned}\int | B_N\varphi(t) - \varphi(t)|^2\, dt &= \int [E\{\varphi(R_1/(N + 1)) \,\|\, U_1 = t\} - \varphi(t)]^2\, dt \\ &= E[E\{\varphi(R_1/(N + 1)) - \varphi(U_1) \,\|\, U_1\}]^2.\end{aligned}$$

The inequality follows upon application of Jensen's inequality. ||

Note that a similar inequality will hold for L_p, $p \geqslant 1$.

In Section 4 we study the rate at which the right-hand quantity of (2.3) goes to zero.

3. SOME PRELIMINARY LEMMAS

We will state a lemma about the asymptotic behavior of the variance of an order statistic. The conditions imposed will involve the following proposition, due to Bickel [1] and restated by Stigler [10].

Let $U_{1N} \leqslant \cdots \leqslant U_{NN}$ be the ordered values of $U_1, ..., U_N$. And let $g_{iN}(u)$ denote the density of U_{iN}; that is,

$$g_{iN}(u) = N \binom{N-1}{i-1} u^{i-1}(1-u)^{N-i}, \qquad 0 < u < 1. \tag{3.1}$$

Consequently, $E\varphi(U_{iN}) = \int \varphi(u)\, g_{iN}(u)\, du$. Let us assume $\psi(t)$ is positive on (0, 1).

PROPOSITION 1. *The following three statements are equivalent.*

(i) *There exists a finite* $r > 0$ *such that*

$$\lim_{t \to 0} t^r \varphi(t) = \lim_{t \to 1} (1-t)^r \varphi(t) = 0.$$

(ii) *There exists a finite* $m \geqslant 0$ *such that*

$$\int \psi(t)[t(1-t)]^m\, dt < \infty.$$

(iii) *For any finite number* $k > 0$, *there exists a finite* $r = r(k, \varphi) \leqslant N/2$ *such that if* $r \leqslant i \leqslant N - r$, *then* $E \mid \varphi(U_{iN})|^k < \infty$.

Furthermore, the above is implied by $J_2(\varphi) < \infty$.

The last part of the proposition follows from Jensen's inequality. Now we state the following lemma and refer the reader to [10, Lemma 4, p. 775] for the proof.

LEMMA 1. *Let* $h(u)$ *be a positive function such that for some* $k \geqslant 0$, $\int h(u)[u(1-u)]^k\, du < \infty$. *Let* b_N *be any sequence of integers such that* $b_N \to \infty$, $b_N/N \to 0$ *as* $N \to \infty$. *Then for any* $m \geqslant 0$ *there exists* $\lambda = \lambda(m, k) > 0$ *such that*

$$N^m \int_{B_N{}^C(i)} h(u)\, g_{iN}(u)\, du \to 0 \tag{3.2}$$

uniformly for $b_N \leqslant i \leqslant N - b_N$, *where*

$$B_N(i) = [(i-1)/(N-1) - \lambda d_N (N-1)^{-1},\; (i-1)/(N-1) + \lambda d_N (N-1)^{-1}],$$

and

$$d_N = [\min(i-1, N-i-1) \log N]^{1/2}.$$

We now state the crucial lemma, which is due to Stigler [10]. We will give the proof, as it is special to our investigation.

LEMMA 2. *Let φ be absolutely continuous on* (0, 1) *with derivative ψ. Assume there exists $\tau > 0$ such that* $\lim_{t\to 0} t^\tau \varphi(t) = \lim_{t\to 1} (1-t)^\tau \varphi(t) = 0$. *If $\psi(t)$ is strictly positive on* (0, 1) *and satisfies condition* T *at* 0 *and* 1, *then for any sequence of integers (b_N) such that $b_N/\log N \to \infty$ and $b_N/N \to 0$ as $N \to \infty$,*

$$NE[\varphi(U_{iN}) - \varphi(p_i)]^2/\sigma^2(p_i) = 1 + o(1) \tag{3.3}$$

uniformly for $p_i \in [b_N/N, 1 - b_N/N]$ where (i) $p_i = i/(N+1)$ *and* (ii) $\sigma^2(p_i) = \psi^2(p_i)\, p_i(1-p_i)$.

Proof. Let $\epsilon > 0$ and let N be large enough such that $N^{-1/2}(\log N)^{1/2} < \epsilon$, and for $b_N \leqslant i \leqslant N - b_N$, $E[\varphi(U_{iN}) - \varphi(p_i)]^2$ exists. Let $B_N(i)$ be given as in Lemma 1.

Now we claim that

$$N \int_{B_N{}^C(i)} ([\varphi(u) - \varphi(p_i)]^2/\sigma^2(p_i))\, g_{iN}(u)\, du = O(N^{-1})$$
$$\text{uniformly for} \quad p_i \in [b_N/N, 1 - b_N/N]. \tag{3.4}$$

Now note that by condition T at 0,

$$\sigma^{-2}(p_i) \leqslant N^2[\psi(p_i)]^{-2}, \qquad [\psi(p_i)]^{-1} \leqslant CN^a$$

so

$$N \int_{B_N{}^C(i)} ([\varphi(u) - \varphi(p_i)]^2/\sigma^2(p_i))\, g_{iN}(u)\, du$$
$$\leqslant N^{2a+3} \int_{B_N{}^C(i)} [\varphi(u) - \varphi(p_i)]^2\, g_{iN}(u)\, du.$$

Thus it follows upon expansion of the integrand, Proposition 1(i) and (ii), and Lemma 1 that (3.4) is uniformly $O(N^{-1})$.

Now for large N, ψ exists and is continuous on $B_N(i)$ for $p_i \in [b_N/N, 1 - b_N/N]$, so by the mean-value theorem, $\varphi(u) - \varphi(p_i) = (u - p_i)\, \psi(\theta_i(u))$, where $\theta_i(u)$ is some point between u and p_i, for $u \in B_N(i)$. Let us denote $\psi_i(u) = \psi(\theta_i(u))$, and define $\psi_i(p_i) = \psi(p_i)$. We note that on $B_N(i)$, since ψ is strictly positive, $\psi_i(u)$ satisfies condition T uniformly in [0, 1], if $p_i \in [b_N/N, 1 - b_N/N]$. Also,

$$(N+2) \int (u - p_i)^2\, g_{iN}(u)\, du = p_i(1 - p_i), \tag{3.5}$$

and from (3.4) with $\varphi(x) = x$, it follows that

$$[p_i(1 - p_i)]^{-1}(N+2) \int_{B_N{}^C(i)} (u - p_i)^2\, g_{iN}(u)\, du = O(N^{-1}).$$

By (3.4) it is enough to show that

$$[p_i(1-p_i)]^{-1} N \int_{B_N(i)} (u-p_i)^2[[\psi_i(u)/\psi(p_i)]^2 - 1]\, g_{iN}(u)\, du \tag{3.6}$$

tends to zero. But it can be easily seen that (3.6) is smaller in absolute value than

$$2 \sup_{u \in B_N(i)} |[\psi_i(u)/\psi(p_i)]^2 - 1|,$$

which tends to zero uniformly for $p_i \in [b_N/N, 1 - b_N/N]$. ||

Remark 1. Note that the proof may be modified to obtain

$$E[\varphi(U_{iN}) - \varphi(p_i)]^2 = N^{-1}(1 + o(1))\, \sigma^2(p_i) \tag{3.7}$$

with ψ now only satisfying condition T, part (a), at 0 and 1. Consequently, ψ is not restricted to being strictly positive in the interior of [0, 1].

4. Proof of Theorems 1, 2, and 3

We first prove the results for $E[\varphi(U_1) - \varphi(R_1/(N+1))]^2$ in place of $\int [B_N\varphi(t) - \varphi(t)]^2\, dt$, then apply Theorem 4. Since φ is absolutely continuous, it is the difference between two nondecreasing functions. Thus without loss of generality we may assume φ to be nondecreasing. Hence ψ will be nonnegative. Let (b_N) satisfy Lemma 2 and be specified later. First we have

$$\begin{aligned} &E[\varphi(U_1) - \varphi(R_1/(N+1))]^2 \\ &\quad = N^{-1} \sum E[\varphi(U_{iN}) - \varphi(p_i)]^2 \\ &\quad = N^{-1}\Bigl(\sum_{1,b_N-1} + \sum_{N-b_N+1,N} + \sum_{b_N,N-b_N}\Bigr) E[\varphi(U_{iN}) - \varphi(p_i)]^2. \end{aligned} \tag{4.1}$$

Upon applying Lemma 2, condition (iii) of Theorem 1, and Remark 1, we see that the rightmost term satisfies

$$N^{-1} \sum_{b_N,N-b_N} E[\varphi(U_{iN}) - \varphi(p_i)]^2/N^{-1}J_2(\varphi) = 1 + o(1). \tag{4.2}$$

Now we will show that under condition (i′) of Theorem 3 the tail terms are $o(N^{-s/2})$, thus finishing the proof of Theorem 3. Then we note that condition (i) of Theorem 1 implies condition (i′) for $\varphi_1(t) = \varphi(t) - \varphi(0)$ and $\varphi_2(t) = \varphi(t) - \varphi(1)$ with $s = 2$. Thus we reapply the proof for the tail terms to these modified functions.

We claim that $[\varphi(t)]^2 \leqslant K[t(1-t)]^{-1+s/2+\delta}$ implies $E[\varphi(U_{iN}) - \varphi(p_i)]^2 \leqslant Cp_i^{-1+s/2+\delta}$ for some constant C, independent of N. In fact

$$E[\varphi(U_{iN}) - \varphi(p_i)]^2 \leqslant 2E[\varphi(U_{iN})]^2 + 2[\varphi(p_i)]^2$$

so it is sufficient to prove $E[\varphi(U_{iN})]^2 \leqslant C(p_i)^{-1+s/2+\delta}$.

$$\begin{aligned} E[\varphi(U_{iN})]^2 &= \int \varphi^2(u)\, g_{iN}(u)\, du \\ &\leqslant K^2 \int [u(1-u)]^{-1+s/2+\delta}\, g_{iN}(u)\, du \\ &= K^2 N \binom{N-1}{i-1} \Gamma(i-1+s/2+\delta) \\ &\quad \times \Gamma(N-i+s/2+\delta)\, \Gamma(N-1+s/2+4\delta). \end{aligned}$$

Then by Sterling's formula, for some constant C, $E[\varphi(U_{iN})]^2 \leqslant C(p_i)^{-1+s/2+\delta}$. Thus letting $b_N = (\log N)^2$, we have

$$\begin{aligned} N^{-1} \sum_{1,b_N} E[\varphi(U_{iN}) - \varphi(p_i)]^2 &\leqslant N^{-1} C \sum_{1,b_N} (p_i)^{-1+s/2+\delta} \\ &\leqslant C(N+1)\, N^{-1} \int_{0,(N+1)^{-1}b_N} t^{-1+s/2+\delta}\, dt \\ &= C(N+1)\, N^{-1} N^{-s/2} b_N^{s/2+\delta} / N^{\delta} \\ &= o(N^{-s/2}). \end{aligned} \tag{4.3}$$

Upon combining (4.2), (4.3), and a similar relationship for the other tail, we see that the result follows. ||

Proof of Theorem 2. Let $f(u) = 0$, $0 \leqslant u \leqslant \frac{1}{2}$, and $f(u) = 1$, $\frac{1}{2} < u \leqslant 1$. Then it follows from Hoeffding's Theorem 4 that $\| B_N f - f \|_1 = (2/\pi)^{1/2} J_1(f)\, N^{-1/2} + o(N^{-1/2}) = (2\pi)^{-1/2} N^{-1/2} + o(N^{-1/2})$. Further, for each integer $N \geqslant 1$ there exists $\varphi_N \in \Phi$ such that $\| f - \varphi_N \|_1 \leqslant N^{-1}$. This implies $\| B_N f - B_N \varphi_N \|_1 \leqslant N^{-1}$. Consequently we have $\| B_N f - f \|_1 \leqslant \| B_N \varphi_N - \varphi_N \|_1 + 2N^{-1}$. Using this and the Schwarz inequality,

$$\begin{aligned} \sup_{\varphi \in \Phi} \| B_N \varphi - \varphi \|_2 &\geqslant \| B_N \varphi_N - \varphi_N \|_1 \\ &\geqslant \| B_N f - f \|_1 - 2N^{-1} \\ &= (2\pi)^{-1/2} N^{-1/2} + o(N^{-1}). \quad \| \end{aligned} \tag{4.4}$$

5. Extension and Conjectures

The mechanics of the proof apply equally well to $\int | B_N\varphi(t) - \varphi(t)|^r\, dt$ for any $r \geqslant 1$, provided an asymptotic expression similar to (3.5) for $E | U_{iN} - p_i |^r$ is available. It is known (see [1]) that, for any $\alpha > 0$, $E | U_{iN} - p_i |^r = N^{-r/2}[p_i(1 - p_i)]^{r/2} \mu_r + o(N^{-r/2})$ uniformly for $p_i \in [\alpha, 1 - \alpha]$, where $\mu_r = \int_{-\infty}^{\infty} | x |^r (2\pi)^{-1/2} e^{-x^2/2}\, dx$. We conjecture that the uniformity extends to $p_i \in [b_N/N, 1 - b_N/N]$, where b_N is defined in Lemma 2. Thus we *should* have, for even r,

$$\int | B_N\varphi(t) - \varphi(t)|^r\, dt \leqslant N^{-r/2} J_r(\varphi)\, \mu_r + o(N^{-r/2})$$

for functions satisfying Theorem 1 where $J_r(\varphi) = \int \psi^r(t)[t(1 - t)]^{r/2}\, dt < \infty$.

References

1. P. J. Bickel, Some contributions to the theory of order statistics, *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **1** (1965), 575–591.
2. D. M. Čibisov, Some theorems on the limiting behavior of the empirical distribution function, *Trudy Mat. Inst. Steklov* **71** (1964), 104–112 (in Russian). Translated in "Selected Translations in Mathematical Statistics and Probability," Vol. 6, Amer. Math. Soc., Providence, R. I., 1966.
3. J. L. Doob, "Stochastic Processes," Wiley, New York, 1953.
4. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol. 2, Wiley, New York, 1966.
5. J. Hájek, Some extensions of the Wald–Wolfowitz–Noether Theorem, *Ann. Math. Statist.* **32** (1961), 506–523.
6. W. Hoeffding, On the distribution of the expected values of the order statistics, *Ann. Math. Statist.* **24** (1953), 93–100.
7. W. Hoeffding, The L_1 norm of the approximation error for Bernstein-type polynomials, *J. Approximation Theory* **4** (1971), 347–356.
8. W. Hoeffding, On the centering of a simple linear rank statistic, *Ann. Statist.* **1** (1973), 54–66.
9. G. G. Lorentz, "Bernstein Polynomials," University of Toronto Press, Toronto, 1953.
10. S. M. Stigler, Linear functions of order statistics, *Ann. Math. Statist.* **40** (1969), 770–788.