JOURNAL OF APPROXIMATION THEORY 17, 359–365 (1976)

Another Look at the Korovkin Theorems

BENNETT EISENBERG*

Department of Mathematics, Lehigh University, Bethlehem, Pennsylvania 18015

Communicated by Oved Shisha

Received January 10, 1975

INTRODUCTION

If u is a continuous function on [0, 1], then $L_n u(p) = \sum_{k=0}^{n} u(k/n)\binom{n}{k} p^k(1-p)^{n-k}$ converges to $u(p)$ uniformly on [0, 1]. Although probabilistic terminology is not necessary for the proof, it is valuable in that it makes the result obvious. Namely, in probabilistic language this says that $L_n u(p)$ is the expected value of $u(X/n)$, where X is a binomial random variable. For large n, X/n will, with high probability, be very close to p and hence the expected value of $u(X/n)$ will be close to $u(p)$. The details of this simple probabilistic argument are in Feller [4].

What often distinguishes probability theory from pure measure theory is that probabilistic terminology can make difficult measure theoretic results intuitive. In this paper we present an approach to positive linear approximation based on probabilistic notation and methods. Korovkin [5] was apparently aware of connections between his subject and probability theory when he wrote "Linear Operators and Approximation Theory," but choose to phrase his results in the language of analysis. This practice has continued. One purpose of this paper is to remove the language barrier and thus to draw the attention of those in probability and those in approximation theory to results of common interest.

Section 1 is devoted to proving some old and new Korovkin type theorems. Although the main tool is a version of Tchebycheff's inequality, the concept of a random variable as either a function on an abstract probability space or as a coordinate function on R^n plays a central role in simplifying results and suggesting new ones.

In Section 2 we use these Korovkin type theorems to prove results related to the characterization of conditional expectation operators due to Moy [6]

* Research supported by N.S.F. Grant GP-37771.

and Bahadur [1]. We find that these results in turn have an interesting interpretation in approximation theory.

1. Bohman and Korovkin proved that if L_n is a sequence of positive linear operators on $C[a, b]$ such that $L_n(1) \to 1$, $L_n(x) \to x$, and $L_n(x^2) \to x^2$ uniformly, then for all f in $C[a, b]$, $L_n(f) \to f$ uniformly. A related result states that if L_n is a sequence of positive linear operators on the continuous periodic functions on $[0, 2\pi]$ such that $L_n(1) \to 1$, $L_n(\cos\theta) \to \cos\theta$, and $L_n(\sin\theta) \to \sin\theta$ uniformly, then for all continuous periodic f, $L_n(f) \to f$ uniformly.

Since then there have been many papers devoted to refining and extending these results. Yet despite the variety of methods the Riesz representation theorem has not been exploited. It gives the representation $L_n f(\theta) = \int_a^b f(t)\, d\mu_{n,\theta}(t)$, where $\mu_{n,\theta}$ are positive measures. Combined with some elementary probabilistic arguments it makes the Bohman–Korovkin theorems transparent and leads to some interesting extensions as well.

In this section L_n is a sequence of positive linear operators from the space $C(X)$ of continuous real or complex-valued functions on a compact Hausdorff space into the space of all functions of X. These assumptions are needed in order to employ the Riesz representation theorem in describing L_n. Alternatively we could allow L_n to act on a more general space, but assume L_n is of the form above. This is the approach of Stancu [7]. A third possibility is to assume L_n acts on the bounded measurable functions on some probability space. Then any $f_1, \ldots, f_n$, are random variables and can be identified with the coordinate functions on a compact subset X of R^n (endowed with an induced probability distribution). The continuous functions of $f_1, \ldots, f_n$, then are identified with $C(X)$. L_n restricted to such functions would then have a form as given by the Riesz representation. This is the approach taken in Section 2.

If μ is a probability measure on some space and $f = (f_1, \ldots, f_n)$ is a measurable vector-valued function, f is called a random vector and $\int f\, d\mu = (\int f_1\, d\mu, \ldots, \int f_m\, d\mu)$ is written $E(f)$ and called the expectation of f. $\int |f - E(f)|^2\, d\mu$ is written $\mathrm{Var}(f)$ and called the variance of f. Note that $\mathrm{Var}(f) = E(|f|^2) - |E(f)|^2$. The following lemma is a generalization of Tchebycheff's inequality.

LEMMA 1. *If f is a random vector with respect to a parameterized family of probability measures $P_{n,\theta}$ such that $E_{n,\theta}(f) \to m(\theta)$ uniformly in θ as $n \to \infty$ and $\mathrm{Var}_{n,\theta}(f) \to 0$ uniformly in θ as $n \to \infty$, then for any $\delta > 0$, $P_{n,\theta}\{t \mid |f(t) - m(\theta)| \geqslant \delta\} \to 0$ uniformly in n. (In the proof we suppress the t in the notation.)*

Proof.

$$\begin{aligned}P_{n,\theta}(|f - m(\theta)| \geqslant \delta) &\leqslant (1/\delta^2)\, E_{n,\theta}(|f - m(\theta)|^2)\\ &= (1/\delta^2)\, E_{n,\theta}(|(f - E_{n,\theta}(f)) + (E_{n,\theta}(f) - m(\theta))|^2)\\ &= (1/\delta^2)(\mathrm{Var}_{n,\theta}(f) + |E_{n,\theta}(f) - m(\theta)|^2),\end{aligned}$$

since the cross terms have expectation zero. Thus if $\mathrm{Var}_{n,\theta}(f) \to 0$ uniformly in θ and $E_{n,\theta}(f) \to m(\theta)$ uniformly in θ, then $P_{n,\theta}(|f - m(\theta)| \geqslant \delta) \to 0$ uniformly. □

THEOREM 1. *Assume L_n is a sequence of positive linear operators from $C(X)$ into the space of all functions on X, where X is a compact Hausdorff space. If*

$$L_n(1) \to 1 \text{ uniformly} \tag{1}$$

and for f_i in $C(X)$, $i = 1,..., m$

$$L_n(f_i) \to f_i \text{ uniformly} \tag{2}$$

and

$$L_n(|f|^2) \to |f|^2 \text{ uniformly}, \tag{3}$$

where $f = (f_1 ,..., f_m)$ then for all g of the form $u(f)$ with u continuous on the range of f, $L_n(g) \to g$ uniformly.

Proof. Since $L_n(1) \to 1$ uniformly, $L_n(h) \to h$ uniformly if and only if $L_n(h)/L_n(1) \to h$ uniformly. We may therefore assume without loss of generality that $L_n(1) = 1$.

For each θ in X the map $h \to L_n(h)(\theta)$ is a positive linear functional and by the Riesz representation theorem

$$L_n(h)(\theta) = \int_X h(t)\, d\mu_{n,\theta}(t),$$

where $\mu_{n,\theta}$ is a sequence of probability measures for each θ. By hypothesis (2), $L_n(f)(\theta) = E_{n,\theta}(f) \to f(\theta)$ uniformly and since $|f(\theta)|$ is bounded $|E_{n,\theta}(f)|^2 \to |f(\theta)|^2$ uniformly. By hypothesis (3), $L_n(|f|^2) = E_{n,\theta}(|f|^2) \to |f(\theta)|^2$ uniformly. Combining these last two results, $\mathrm{Var}_{n,\theta}(f) = E_{n,\theta}(|f|^2) - |E_{n,\theta}(f)|^2$ converges to zero uniformly.

Now let $g = u(f)$ with u continuous on the range of f. Since X is compact, u is bounded and uniformly continuous on the range of f. There then exists an M and δ such that

$$|u(f)| \leqslant M \quad \text{and} \quad |f(t) - f(t')| \leqslant \delta \Rightarrow |u(f(t)) - u(f(t'))| \leqslant \epsilon/2.$$

Then

$$\begin{aligned} | L_n g(\theta) - g(\theta)| &= | E_{n,\theta}(u(f)) - u(f(\theta))| \\ &= | E_{n,\theta}(u(f) - u(f(\theta)))| \\ &\leqslant \epsilon/2 + 2M\mu_{n,\theta}\{|f - f(\theta)| \geqslant \delta\}. \end{aligned}$$

Since $E_{n,\theta}(f) \to f(\theta)$ and $\mathrm{Var}_{n,\theta}(f) \to 0$ uniformly by Lemma 1, $\mu_{n,\theta}\{|f - f(\theta)| \geqslant \delta\} \to 0$ uniformly in $\to \theta$. It can be made less than $\epsilon/4M$ uniformly in θ for n large enough. Thus $L_n(g) \to g$ uniformly. □

Next notice that functions of the form $u(f)$ have a simple interpretation.

LEMMA 2. *If g is continuous on a compact set and $f(x) = f(y) \Rightarrow g(x) = g(y)$, where f is a continuous vector-valued function then $g = u(f)$, where u is continuous on the range of f.*

Proof. Since $f(x) = f(y) \Rightarrow g(x) = g(y)$, g can be written as some function of f, say $g = u(f)$. If u is not continuous on the range of f there is a closed set F for which $u^{-1}(F)$ is not closed in the range of f. But then $f^{-1}(u^{-1}(F))$ is not closed since the domain of f is compact. Thus $u(f)$ is not continuous contradicting the hypothesis. □

The lemma leads to a restatement of the theorem.

THEOREM 1′. *Under the hypotheses of Theorem* 1, $L_n(g) \to g$ *uniformly for all g in $C(X)$ satisfying $g(x) = g(y)$ whenever $f(x) = f(y)$. In particular, if f is* 1–1, $L_n(g) \to g$ *for all g in $C(X)$.*

COROLLARY 1. *The Second Bohman–Korovkin Theorem. Let* $X = [0, 2\pi]$. *If L_n is a sequence of positive linear operators on $C(X)$,*

(1) $L_n(1) \to 1$ *uniformly,*
(2) $L_n(\cos \lambda) \to \cos \lambda$ *uniformly,*
(3) $L_n(\sin \lambda) \to \sin \lambda$ *uniformly,*

then $L_n(g) \to g$ uniformly for all g periodic on $[0, 2\pi]$.

Proof. $f = (\cos \lambda, \sin \lambda)$ is 1–1 except that $f(0) = f(2\pi)$ and $|f^2| = 1$ so that $L_n(1) \to 1 \Rightarrow L_n(|f|^2) \to |f|^2$. □

Furthermore we have the obvious generalization,

COROLLARY 2. *Let X be a compact Hausdorff space and L_n a sequence of positive linear operators on $C(X)$, then if* (1) $L_n(1) \to 1$ *uniformly and for f_i in $C(X)$, $i = 1,\ldots, m$ with $|f| =$ constant,* (2) $L_n(f_i) \to f_i$ *uniformly, then for all g in $C(X)$ satisfying $g(x) = g(y)$ whenever $f(x) = f(y)$, $L_n(g) \to g$ uniformly.*

For example, if X is the sphere $x^2 + y^2 + z^2 = 1$ and $L_n 1 \to 1$, $L_n x \to x$, $L_n y \to y$, $L_n z \to z$ uniformly on X then for all g continuous on the sphere $L_n g \to g$ uniformly.

COROLLARY 3. *Let X be any compact set in R^n. Let L_n be a sequence of positive linear operators on $C(X)$. If*

(1) $L_n(1) \to 1$ *uniformly,*

(2) $L_n(x_i) \to x$ *uniformly, where x_i are the coordinate functions, and*

(3) $L_n(\sum x_i^2) \to \sum x_i^2$ *uniformly,*

then for all g in $C(X)$, $L_n(g) \to g$ uniformly.

Proof. $(x_1, \ldots, x_n)$ is 1–1. □

Corollaries 2 and 3 extend results of Volkov and Morozov (see Censor [2]).

2. In this section we combine concepts of uniform approximation with those of mean-square approximation. We assume L is a positive linear operator on $\mathscr{L}^2(\mu)$, where μ is a probability measure. An example of such an operator would be convolution with a Fejer kernel,

$$L_n f(x) = (1/2\pi) \int_{-\pi}^{\pi} f(t)\, K_n(x - t)\, dt, \qquad \text{where } K_n(x) = \frac{1}{n}\left[\frac{\sin(nx/2)}{\sin(x/2)}\right]^2.$$

Here, $L_n f$ is the average of Fourier series of f of increasing degree. L_n is then a sequence of positive linear operators such that for all f in $\mathscr{L}^2(dt)$, $L_n f$ converges to f in mean square and for all continuous periodic f, $L_n f$ converges to f uniformly. For other examples, see Dzjalyk [3].

In the spirit of Korovkin we find conditions on L and f so that $Lf = f$. We then use these results to give a simple proof of an important result in probability theory.

THEOREM 2. *Assume L is a positive linear operator on $\mathscr{L}^2(\mu)$ and assume $L1 = 1$ and $Lg_i = g_i$ for $i = 1, \ldots, k$. Then for any convex function u with $u(g_1, \ldots, g_k)$ in $\mathscr{L}^2(\mu)$, $Lu(g_1, \ldots, g_k) \geqslant u(g_1, \ldots, g_k)$.*

Proof. For any point $x_0 = (g_1(\omega), \ldots, g_k(\omega))$ there is a linear function l_0 with graph through $(x_0, u(x_0))$ and lying below the graph of $u(x)$. (In fact, this is how we define a convex function.) Then $Lu(g_1, \ldots, g_k) \geqslant L(l_0(g_1, \ldots, g_k)) = l_0(g_1, \ldots, g_k)$, the first inequality due to the positivity of L and the second inequality due to the linearity of l_0 and the assumptions on L. In particular, $Lu(g_1, \ldots, g_k)(\omega) \geqslant l_0(g_1, \ldots, g_k)(\omega) = u(g_1, \ldots, g_k)(\omega)$ and ω is arbitrary. □

COROLLARY 1. *If we also assume the g_i are bounded and either $\| L \| = 1$ or $L = L^*$ then $Lf = f$ for all f in $\mathscr{L}^2(\mu)$ of the form $F(g_1, \ldots, g_k)$.*

Proof. In the theorem let $u(g_1, \ldots, g_k) = \sum g_i^2$. Then $L(\sum g_i^2) \geqslant \sum g_i^2$. But if $\| L \| = 1$ then there must be equality. If $L = L^*$ then $\int L(\sum g_i^2) 1 \, d\mu = \int \sum g_i^2 (L1) \, d\mu = \int \sum g_i^2 \, d\mu$ and again we must have $L(\sum g_i^2) = \sum_i g_i^2$. Now identify the random variables g_i with the coordinate random variables x_i on R^k and apply Corollary 3 of Section 1. Since the g_i's are bounded we may restrict the x_i's to a compact set in R^k. Then $L1 = 1$, $Lx_i = x_i$, and $L(\sum x_i^2) = \sum x_i^2$. Hence L is the identity on continuous functions of $x_1, \ldots, x_k$. Then by either the boundedness or self adjointness we conclude that L must be the identity on all $\mathscr{L}^2$ functions of the form $F(g_1, \ldots, g_k)$. □

To overcome the boundedness assumption on the g_i's we now assume that L is idempotent.

COROLLARY 2 (Moy–Bahadur). *If L is a positive orthogonal projection with $L1 = 1$, $Lg_i = g_i$, $i = 1, \ldots, k$ then $Lf = f$ for all f in $\mathscr{L}^2(\mu)$ of the form $F(g_1, \ldots, g_k)$.*

Proof. Let $g_{i,m} = g_i$ truncated so $| g_{i,m} | \leqslant M$. Then by positivity $| Lg_{i,m} | \leqslant M$ and $L(Lg_{i,m}) = Lg_{i,m}$ since L is a projection. Hence by Corollary 1, $Lf = f$ for f of the form $F(Lg_{1,m}, \ldots, Lg_{k,m})$. But

$$\| Lg_{i,m} - g_i \| = \| Lg_{i,m} - Lg_i \| \leqslant \| g_{i,m} - g_i \| \to 0.$$

Hence for bounded continuous F, $LF(g_1, \ldots, g_k) = F(g_1, \ldots, g_k)$. Since $\| L \| = 1$ this equality extends to all $F(g_1, \ldots, g_k)$ in $\mathscr{L}^2(\mu)$. □

A projection onto subspaces of functions of the form $F(g_1, \ldots, g_k)$ is a conditional expectation operator, hence the significance of this result in probability. However, it also has significance in approximation theory, mainly of a negative character. It says that approximation using projections and approximation by positive linear operators are essentially disjoint subjects. For if L projects on the span of l, $g_1, \ldots, g_k$ in $\mathscr{L}^2(\mu)$ and $L \geqslant 0$ then L projects on all functions $F(g_1, \ldots, g_k)$. If the g_i separate points this says L is the identity. From Corollary 1 alone we see that there can be no positive Hilbert Schmidt operator taking 1 into 1 and f into f, where f is a bounded nonsimple function. For then there would be an infinite-dimensional eigenspace.

ACKNOWLEDGMENT

I thank Jerry King for suggesting that Tchebycheff's inequality could be used to prove the Bohman–Korovkin theorem and for many stimulating conversations thereafter.

References

1. R. R. Bahadur, Measurable subspaces and subalgebras, *Proc. Amer. Math. Soc.* **6** (1955), 565–570.
2. E. Censor, Quantitative Results for positive linear approximation operators, *J. Approximation Theory* **4** (1971), 442–450.
3. V. Dzjadyk, Approximation of functions by positive linear operators and singular integrals, *Mat. Sb.* **112** (1966), 508–517.
4. W. Feller, "Introduction to Probability Theory and Its Applications II," Wiley, New York, 1968.
5. P. Korovkin, On convergence of positive linear operators in the space of continuous functions, *Dokl. Akad. Nayk. SSSR* **90** (1953), 961–964.
6. S. 6. S. Moy, Characterization of conditional expectation as a transformation on function spaces, *Pacific J. Math.* **4** (1954), 47.
7. D. Stancu, Use of probabilistic methods in the theory of uniform approximation of continuous functions, *Rev. Roumaine. Math. Pures Appl.* **14** (1969), 673–691.
8. D. Wulbert, Convergence of operators and Korovkin's Theorem, *J. Approximation Theory* **1** (1968), 381–390.
9. J. P. King, Probability and positive linear operators, *Rev. Roumaine Math. Pures Appl.* **20** (1975), 325–327.