On the Approximation Properties of Bernstein Polynomials via Probabilistic Tools

Henryk Gzyl and José Luis Palacios

Abstract

We study two loosely related problems concerning approximation properties of Bernstein polynomials $B_n f(x)$ of some function f on $[0,1]$: the absence of Gibbs phenomenon at points at which f has jumps, and the convergence of $(B_n f)'(x)$ towards $f'(x)$. Both results are obtained using classical probabilistic tools. In particular the proof of the second statement relies on the representation of the derivative $(B_n f)'(x)$ as the expectation of the functional of a random variable.

1 Introduction.

Let f be a real function defined on the interval $[0,1]$. Let $S_{n,x}$ be a Binomial random variable with parameters n and x, and let $E[X]$ denote the expected value of the random variable X. In our previous paper [3], we showed how to use the theory of large deviations to derive rates of convergence of the Bernstein polynomials, defined as:

$$B_n f(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right) = Ef\left(\frac{S_{n,x}}{n}\right) \tag{1}$$

to the limit function f, when the f being approximated is Lipschitz continuous. The rate $O(n^{-1/3})$ obtained with Berstein's classic probabilistic proof, where all that is used is Chebyschev's inequality, was improved to $O((\ln n/n)^{1/2})$.

M. K. Khan ([4]) brought to our attention the fact that the optimal rate among Lipschitz functions is indeed $n^{-1/2}$. Moreover, P. Mathé, in a nice historical framework, showed in [5] that if the function is Hölder continuous with exponent α for some $0 < \alpha \leq 1$, the rate of convergence is $n^{-\alpha/2}$.

Some obvious questions come up when we compare the approximation of f given by (1) with approximations by, say, Fourier series or other orthogonal expansions. For starters: how does $B_n f(x_0)$ behave when f has a jump discontinuity at x_0? And more important, does the Gibbs-Wilbraham phenomenon (see [Hewitt-Hewitt]) take place as well?

In this paper we continue using probabilistic tools to further study approximation properties of $B_n f(x)$ when f is not continuous. In fact, we show that the Gibbs phenomenon does not occur for the approximation of monotone, piecewise smooth functions, having both left and right derivatives at every point, by Bernstein polynomials, contrary to what happens with the Fourier series of such a function. Specifically, we consider a simple jump function (see (2) below) and prove that $B_n f(x_0) \to \frac{1}{2}(f(x_0+0)+f(x_0-0))$ as $n \to \infty$, and also that the convergence is monotone on both sides of the discontinuity, as a consequence of which the characteristic overshoot of the Gibbs phenomenon does not occur. We also provide estimates of the size of the approximation error and the rate of convergence at the discontinuity point. Later on we prove that for a bounded function f, $(B_n f)'(x)$ converges to $f'(x)$ at all x at which $f'(x)$ exists.

Let f_n be a sequence of approximants to a piecewise smooth f that has right and left derivatives at every point, say a partial sum of a trigonometric series, Bessel functions or Gegenbauer polynomials. In general, the Gibbs phenomenon can be described by the following behavior:

(i) If x_0 is a discontinuity point of f, then

$$\lim_{n\to\infty} f_n(x) = \frac{f(x+0)+f(x-0)}{2}.$$

(ii) On any subinterval $[x_1, x_2]$ for which the function is continuous , we have uniform convergence:

$$\lim_{n\uparrow\infty} \max_{x_1\le x\le x_2} |f_n(x) - f(x)| = 0.$$

(iii) On any subinterval containing a single discontinuity x_0 of the function, we have Gibbs phenomenon: for small $\delta > 0$

$$\lim_{n\uparrow\infty} \left(\max_{|x_0-x|\le\delta} f_n(x) - \min_{|x_0-x|\le\delta} f_n(x) \right) = C|f(x_0+0) - f(x_0-0)|.$$

where

$$C = \frac{2}{\pi}\int_0^\pi \left(\frac{\sin x}{x}\right) dx \approx 1.18.$$

See the work by Bachman-Narici-Beckensterin, Dym-Mc Kean or Gray-Pinsky for precise statements and examples in which the Gibbs phenomenon occurs. Gibbs phenomenon arises when approximating discontinuities using smooth approximants. In addition to the more recent paper by Gray and Pinky, Hewitt-Hewitt provide us with an excellent survey with many historical details. In [12], Gottlieb and Shiu provide a short complementary historical review, among other interesting results which we mention below.

In the next section we shall show that the Bernstein polynomial approximant satisfies (i) and (ii) above, and that it does not overshoot at a point of discontinuity, in other words, that (iii) holds with $C = 1$ whenever f is a finite sum of jump functions, that is functions that change only at jump points. We extend these results to a monotone f with a finite number of jumps with a minor modification on (iii). In the last section we show how to write $(B_n f)'(x)$ as an expectation of a functional of a binomial variable, and study the convergence of this expectation to $f'(x)$.

2 No Gibbs phenomenon for Bernstein polynomials.

Let us consider a simple jump function

$$f(x) = \begin{cases} a & x < x_0 \\ b(> a) & x \geq x_0 \end{cases} . \tag{2}$$

The Bernstein polynomial for this function is

$$B_n f(x) = aP\left(\frac{S_{n,x}}{n} < x_0\right) + bP\left(\frac{S_{n,x}}{n} \geq x_0\right) = a + (b-a)P\left(\frac{S_{n,x}}{n} \geq x_0\right),$$

for which the following bounds obviously hold:

$$a = B_n f(0) \leq B_n f(x) \leq B_n f(1) = b. \tag{3}$$

It can also be proved that $B_n f$ is an increasing function on account of the fact that if $x \leq y$ then for $0 \leq k \leq n$ one has

$$P\left(S_{n,x} \geq k\right) = \sum_{i \geq k} \binom{n}{i} x^i (1-x)^{n-i} \leq \sum_{i \geq k} \binom{n}{i} y^i (1-y)^{n-i} = P\left(S_{n,y} \geq k\right). \tag{4}$$

Both sides of (4) are equal to 1 when $k = 0$. For other values of k, one subtracts form each side the appropriate terms that preserve the inequality.

A more elegant way to prove (4) is to consider n independent copies of the bivariate $0-1$-valued variables (X_i, Y_i), $1 \leq i \leq n$, with joint distribution dictated by the probabilities $P(X_i = Y_i = 1) = x$, $P(X_i = 1, Y_i = 0) = 0$ and $P(X_i = Y_i = 0) = 1 - y$. If we define $Z = \sum_{i=1}^n X_i$ and $W = \sum_{i=1}^n Y_i$, then the distributions of Z and W are those of, respectively, $S_{n,x}$ and $S_{n,y}$, and by construction $\{Z \geq k\} \subset \{W \geq k\}$, so that $P\left(S_{n,x} \geq k\right) = P(Z \geq k) \leq P(W \geq k) = P\left(S_{n,y} \geq k\right)$.

We will deal now with uniform convergence on an interval $[x_1, x_2] \subset [0,1] - \{x_0\}$. Consider first $x_2 < x_0$. By Chebyschev's inequality

$$P\left(\frac{S_{n,x}}{n} \geq x_0\right) \leq P\left(\left|\frac{S_{n,x}}{n} - x\right| \geq x_0 - x\right)$$

$$\leq \frac{x(1-x)}{n(x_0-x)^2} \leq \frac{1}{4n(x_0-x_2)^2}, \tag{5}$$

and therefore

$$B_n f(x) \to a = f(x)$$

uniformly in the interval $[x_1, x_2]$. A similar argument holds for $x \in [x_1, x_2]$ such that $x_0 < x_1$.

On the other hand, if $x = x_0$ we have

$$B_n f(x_0) = a + (b-a) P\left(\frac{S_{n,x_0}}{n} \geq x_0\right).$$

Now, if we consider $X_i, 1 \leq i \leq n$, to be independent Bernoulli variables with parameter x_0 we can write

$$P\left(\frac{S_{n,x_0}}{n} \geq x_0\right) = P\left(\sum_{i=1}^{n}(X_i - x_0) \geq 0\right) = P\left(\frac{\sum_{i=1}^{n}(X_i - x_0)}{\sqrt{n x_0(1-x_0)}} \geq 0\right) \to \frac{1}{2},$$

as $n \to \infty$ by the central limit theorem (CLT), and

$$\left| B_n f(x_0) - \frac{a+b}{2} \right| = |a-b| \left| P\left(\frac{S_{n,x_0}}{n} \geq x_0\right) - \frac{1}{2} \right| \to 0,$$

as $n \to \infty$, that is, $B_n f(x_0)$ converges to the average of the right and left limits of f at x_0.

These calculations take care of properties (i) and (ii) for a jump function as described in (2).

To verify that condition (iii) holds with C=1, i. e., that the approximation by Bernstein polynomials fits well, note that on account of $B_n f$ being increasing we have

$$\max_{|x_0-x|\leq\delta} B_n f(x) - \min_{|x_0-x|\leq\delta} B_n f(x) = B_n f(x_0+\delta) - B_n f(x_0-\delta),$$

and letting n tend to infinity and using property (ii) on $x_0+\delta$ and $x_0-\delta$ we obtain

$$\lim_{n\uparrow\infty}\left(\max_{|x_0-x|\leq\delta} B_n f(x) - \min_{|x_0-x|\leq\delta} B_n f(x)\right) = |f(x_0+\delta) - f(x_0-\delta)|$$

$$= |f(x_0+0) - f(x_0-0)|.$$

Comments These results can be extended obviously to a finite sum of simple jump functions. The extension to any monotone function with finitely many jumps in $[0,1]$ -which can be written as a sum of a continuous monotone function plus a finite sum of simple jump functions- is now straightforward. Notice that when dealing with approximations by Bernstein polynomials, items (i) and (ii) mentioned in the introduction, remain valid, But item (iii) has to be replaced by

(iii)' If x_0 is a discontinuity of f then

$$\lim_{\delta\to 0}\lim_{n\to\infty}\left(\max_{|x_0-x|\le\delta} B_nf(x) - \min_{|x_0-x|\le\delta} B_nf(x)\right) = |f(x_0+0) - f(x_0-0)|.$$

which can be rephrased as saying that there is no Gibbs phenomenon for approximation by Berstein polynomials.

This connects with the following general problem, for which there seems to be no general answer: Suppose we want to approximate a jump by a sequence of approximants. Which properties of the approximating sequence cause the overshoot phenomenon? Typically, as in approximation by trigonometric polynomials, the approximating sequence consists of partial sums of an orthogonal series. Is it orthogonality which causes the overshooting? If so, then one should not wonder that Bernstein polynomials do not show it. Is it smoothness? There seems to be no general theorem in this direction, but seemingly Gibbs phenomenon arises from approximating a jump by a system which is designed to allow approximations of arbitrary order. The latter statement is supported by results of V. L. Velikin [7], who showed that spline approximation yields the same phenomenon, if the smoothness of the splines tends to ∞. Bernstein polynomials, although smooth, allow approximation only up to order n^{-1}, no matter how smooth the function is, see for example the results in section 1.6 of the book by G. G. Lorentz. In section 4.16 of [10] it is shown how approximating by Cesaro sums, kills the overshooting. And this issue is explored further in section 3 of [12] as part of the problem of devising rapidly convergent methods for reconstructing local behavior (value of a function at a point) from global data (Fourier coefficients).

In what we did above, we did not pay attention to issues related to speed of convergence. As we did in [2], we can make some precise statements invoking the following large deviations result found for instance in [9]

Lemma. For a binomial random variable $S_{n,x}$ and $a > 0$ arbitrary

$$P\left(|S_{n,x} - nx| > a\right) \le 2e^{-\frac{2a^2}{n}}. \tag{6}$$

Then, for example, the bound $O(\frac{1}{n})$ of the uniform convergence in (ii) obtained with Chebyschev inequality in (5) may be improved to the exponential bound $2e^{-2n(x_0-x)^2}$. Likewise, when verifying that condition (iii) above holds with C=1, we can argue that the speed of convergence to the limit is exponential. The lemma will also be used in the proof of proposition 2 below.

3 Convergence of $(B_n f)(x)'$ towards $f'(x)$

In the previous section we showed that $B_n f$ is increasing when f is an increasing jump function by direct computations with the binomial distribution. We can show in general that when $f(x)$ is increasing (resp. decreasing), then the derivative $(B_n f)'(x)$ of $B_n f(x)$ is positive (resp. negative), and therefore $B_n f$ is also increasing (resp. decreasing). In fact, we can provide the following

Proposition 1. The derivative of the Bernstein polynomial $B_n f$ can be expressed as

$$(B_n f)'(x) = E\left[\left(\frac{S_{n,x} - nx}{\sqrt{nx(1-x)}}\right)^2 D(n,f)(x)\right], \tag{7}$$

where

$$D(n,f)(x) = \frac{f\left(\frac{S_{n,x}}{n}\right) - f(x)}{\frac{S_{n,x}}{n} - x}.$$

Proof. Deriving (1) with respect to x we obtain

$$\begin{aligned}
(B_n f)'(x) &= \sum_{i=0}^{n} \binom{n}{i} i f\left(\frac{i}{n}\right) x^{i-1}(1-x)^{n-i} \\
&\quad - \sum_{i=0}^{n} \binom{n}{i} (n-i) f\left(\frac{i}{n}\right) x^{i}(1-x)^{n-i-1} \\
&= \frac{1}{x} E S_{n,x} f\left(\frac{S_{n,x}}{n}\right) - \frac{1}{1-x} E\left(n - S_{n,x}\right) f\left(\frac{S_{n,x}}{n}\right) \\
&= \frac{1}{x(1-x)} E\left[(S_{n,x} - nx) f\left(\frac{S_{n,x}}{n}\right)\right] \\
&= \frac{1}{x(1-x)} E\left[(S_{n,x} - nx)\left(f\left(\frac{S_{n,x}}{n}\right) - f(x) + f(x)\right)\right] \\
&= \frac{1}{x(1-x)} E\left[(S_{n,x} - nx)\left(f\left(\frac{S_{n,x}}{n}\right) - f(x)\right)\right] \\
&= E\left[\frac{(S_{n,x} - nx)^2}{nx(1-x)}\left(\frac{f\left(\frac{S_{n,x}}{n}\right) - f(x)}{\frac{S_{n,x}}{n} - x}\right)\right] \\
&= E\left[\left(\frac{S_{n,x} - nx}{\sqrt{nx(1-x)}}\right)^2 D(n,f)(x)\right],
\end{aligned}$$

as claimed •

From this is clear that $(B_n f)'(x)$ is positive (resp. negative) in case $D(n,f)(x)$ is positive (negative), and this occurs whenever f is increasing (resp. decreasing). Furthermore, if x is a point where the derivative exists, $D(n,f)(x) \to f'(x)$ as $n \to \infty$, and the squared term in the integral in (7), by the CLT, converges to the square of a standard normal variable, so we should have $(B_n f)'(x) \approx f'(x)$ when n is large. This can be made rigorous via a probabilistic proof which in fact does not use the CLT, as follows:

Proposition 2. Assume $\sup_{x\in[0,1]} f(x) = M < \infty$. Then for any x such that $f'(x)$ exists we have

$$\lim_{n\to\infty} (B_n f)'(x) = f'(x).$$

Proof. Define $A(n,f)(x) = D(n,f)(x) - f'(x)$. Then we can write

$$\begin{aligned}(B_n f)'(x) &= E\left[\left(\frac{S_{n,x}-nx}{\sqrt{nx(1-x)}}\right)^2 D(n,f)(x)\right] \\ &= f'(x) + E\left[\left(\frac{S_{n,x}-nx}{\sqrt{nx(1-x)}}\right)^2 A(n,f)(x)\right], \qquad (8)\end{aligned}$$

since $E\left(\frac{S_{n,x}-nx}{\sqrt{nx(1-x)}}\right)^2 = 1$.

Now, since $f'(x)$ exists, given $\epsilon > 0$, there exists $\delta > 0$ such that if $\left|\frac{S_{n,x}}{n} - x\right| < \delta$ then $|A(n,f)(x)| < \epsilon$. We prove that the absolute value of the second summand in (8) goes to zero by splitting and bounding it by

$$\begin{aligned}&E\left[\left(\frac{S_{n,x}-nx}{\sqrt{nx(1-x)}}\right)^2 |A(n,f)(x)|\, 1_{\{|\frac{S_{n,x}}{n}-x|<\delta\}}\right] \\ &+E\left[\left(\frac{S_{n,x}-nx}{\sqrt{nx(1-x)}}\right)^2 |A(n,f)(x)|\, 1_{\{|\frac{S_{n,x}}{n}-x|\geq\delta\}}\right], \qquad (9)\end{aligned}$$

where 1_A stands for the indicator function of the set A. The first summand in (9) can be bound by

$$\epsilon E\left[\left(\frac{S_{n,x}-nx}{\sqrt{nx(1-x)}}\right)^2 1_{\{|\frac{S_{n,x}}{n}-x|<\delta\}}\right] \leq \epsilon, \qquad (10)$$

whereas the second summand in (9) can be bound by

$$\left(\frac{2M}{\delta}+f'(x)\right)\frac{n}{4}E\left[\left(\frac{S_{n,x}}{n}-x\right)^2 1_{\{|\frac{S_{n,x}}{n}-x|\geq\delta\}}\right]$$
$$\leq\left(\frac{2M}{\delta}+f'(x)\right)\frac{n}{4}P\left(\left|\frac{S_{n,x}}{n}-x\right|\geq\delta\right), \tag{11}$$

where the last inequality uses the fact that the square of the distance of the points $\frac{S_{n,x}}{n}$ and x in the interval $[0,1]$ is less than 1. Now (11) can be bound, using (6) by

$$\left(\frac{2M}{\delta}+f'(x)\right)\frac{n}{2}e^{-2n\delta^2},$$

which goes to zero as $n\to\infty$, finishing the proof •

We should mention that a similar result is stated in as a problem 2 of chapter VII of Feller's classic [13]. What is not clear is whether a representation like (7), from which our proof follows, was known to him.

Aknowledgements It is a pleasure to thank the referees for their comments. They helped us improve our presentation.

References

[1] E. Parzen, Modern Probability Theory and its Applications, New York, Wiley, 1960.

[2] H. Gzyl and J. L. Palacios, The Weierstrass approximation theorem and large deviations, *American Mathematical Monthly*, 104 (1997).

[3] M. K. Kahn, On the Weierstrass approximation theorem and large deviations, unpublished manuscript (1997).

[4] P. Mathé, Approximations of Hölder continuous functions by Bernstein polynomials, *American Mathematical Monthly*, 106 (1999).

[5] A. Gray and M. A. Pinsky, Gibbs phenomenon for Fourier-Bessel series, *Exposition. Math.*, 11 (1993) 123-135.

[6] E. Hewitt and R.E. Hewitt,The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis, *Arch. Hist. Exact Sci.*, 21:2 (1979/80) 129-160.

[7] V.L. Velikin, A limit relation for different methods of approximating periodic functions by splines and trigonometric polynomials, *Ann. Math.*, 13:1 (1987) 45-74.

[8] G.G. Lorentz, Bernstein Polynomials, Vol. 8 of Math. Expos., Univ. of Toronto Press, Toronto 1953.

[9] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, New York, 1992.

[10] G. Bachman, L. Narici and E. Beckenstein, *Fourier and Wavelet Analysis*, Springer-Verlag, Berlin, 2000.

[11] H. Dym and H. Mc Kean, *Fourier Series and Integrals*, Acad. Press, New York, 1972

[12] D. Gottlieb and C. Shiu, On the Gibbs phenomenon and its resolution, SIAM Rev, 39:4, 1997, 644-468.

[13] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, Wiley, New York, 1971.

Henryk Gzyl and José Luis Palacios
Depto. de Cómputo y Estadística
Universidad Simón Bolívar
Caracas, Venezuela
hgzyl@reacciun.ve; jopala@cesma.usb.ve