Approximations in Bernstein Form

This version of the document is dated 2023-05-27.

Peter Occil

This page describes how to compute a polynomial in Bernstein form that comes close to a known function $f(\lambda)$ with a user-defined error tolerance, so that the polynomial's coefficients will lie in the closed unit interval if f's values lie in that interval. The polynomial is often simpler to calculate than the original function f and can often be accurate enough for an application's purposes.

The goal of these approximations is to avoid introducing transcendental and trigonometric functions to the approximation method. (For this reason, although this page also discusses approximation by so-called *Chebyshev interpolants*, that method is relegated to the appendix.)

> **Note:** This page was originally developed as part of a section on *approximate Bernoulli factories*, or algorithms that toss heads with probability equal to a polynomial that comes close to a continuous function. However, the information in this page is of much broader interest than the approximate Bernoulli factory problem.

1 Contents

2 About This Document

This is an open-source document; for an updated version, see the source code or its rendering on GitHub. You can send comments on this document on the GitHub issues page, especially if you find any errors on this page.

My audience for this article is **computer programmers with mathematics knowledge, but little or no familiarity with calculus**.

3 Definitions

This section describes certain math terms used on this page for programmers to understand.

The *closed unit interval* (written as [0, 1]) means the set consisting of 0, 1, and every real number in between.

For definitions of *continuous*, *derivative*, *convex*, *concave*, *Hölder continuous*, and *Lipschitz continuous*, see the definitions section in "**Supplemental Notes for Bernoulli Factory Algorithms**".

Any polynomial can be written in *Bernstein form* as—

$${n\choose 0}\lambda^0 (1-\lambda)^{n-0} a[0] + {n\choose 1}\lambda^1 (1-\lambda)^{n-1} a[1] + ... + {n\choose n}\lambda^n (1-\lambda)^{n-n} a[n],$$

where *n* is the polynomial's *degree* and *a*[0], *a*[1], ..., *a*[*n*] are its *n* plus one *coefficients*.[1]

4 Approximations by Polynomials

Suppose $f(\lambda)$ is continuous and maps the closed unit interval to the closed unit interval.

Then, a polynomial of a high enough degree (called n) can be used to approximate $f(\lambda)$ with an error no more than ϵ, as long as the polynomial's Bernstein coefficients can be calculated and an explicit upper bound on the approximation error is available. See my **question on MathOverflow**. Examples of these polynomials (all of degree n) are given in the following table.

Name	Polynomial	Its Bernstein coefficients are found as follows:	Notes
Bernstein polynomial.	$B_n(f)$.	$f(j/n)$, where $0\le j\le n$.	Originated with S.N. Bernstein (1912).
Order-2 iterated Boolean sum.	$U_{n,2} = B_n(W_{n,2})$.	$W_{n,2}(j/n)$, where $0\le j\le n$ and $W_{n,2}(\lambda) = 2 f(\lambda) - B_n(f)(\lambda)$.	Micchelli (1973)[2], Guan (2009)[3], Güntürk and Li (2021, sec. 3.3)[4].
Order-3 iterated Boolean sum.	$U_{n,3} = B_n(W_{n,3})$.	$W_{n,3}(j/n)$, where $0\le j\le n$ and $W_{n,3}(\lambda) = B_n(B_n(f)(\lambda))$ + $3 (f(\lambda)$ − $B_n(f)(\lambda))$.	Same.
		(First, define the following operation: **Get coefficients for n given m**: Treat the coefficients [$f(0/m)$, $f(1/m)$, ..., $f(m/m)$] as representing a polynomial in Bernstein form of degree m,	Tachev

Butzer's linear combination (order 1).	$L_{2,n/2} = 2 B_{n}(f(\lambda))$ − $B_{n/2}(f(\lambda))$.	then rewrite that polynomial to one of degree n with $n+1$ Bernstein coefficients [see note 3 at the end of this section], then return those coefficients.) **Get coefficients for n given $n/2$**, call them *a*[0], ..., *a*[*n*], then set the final Bernstein coefficients to $2 f(j/n) - a[j]$ for each j.	(2022)[5], Butzer (1955)[6]. $n\ge 6$ must be even.
Butzer's linear combination (order 2).	$L_{3,n/4} = B_{n/4}(f)/3$ + $B_{n}(f)\cdot 8/3$ − $2 B_{n/2}(f)$	**Get coefficients for n given $n/4$**, call them *a*[0], ..., *a*[*n*], then **get coefficients for n given $n/2$**, call them *b*[0], ..., *b*[*n*], then set the final Bernstein coefficients to $a[j]/3-2 b[j]+8 f(j/n)/3$ for each j.	Butzer (1955)[7]. $n\ge 4$ must be divisible by 4.
Lorentz operator (order 2).	$Q_{n-2,2}=B_{n-2}(f)-x(1-x)\cdot$ $B_{n-2}(f'')/(2(n-2))$.	**Get coefficients for n given $n-2$**, call them *a*[0], ..., *a*[*n*]. Then for each integer j with $1\le j\lt n$, subtract z from *a*[*j*], where $z=(((f''((j-1)/(n-2)))$ / $(4(n-2)))\cdot 2j(n-j)/((n-1)\cdot(n))$. The final Bernstein coefficients are now *a*[0], ..., *a*[*n*].	Holtz et al. (2011)[8]; Bernstein (1932)[9]; Lorentz (1966)[10]. $n\ge 4$; f'' is the second derivative of f.

The goal is now to find a polynomial of degree n such that—

1. the polynomial is within ϵ of $f(\lambda)$, and
2. each of the polynomial's Bernstein coefficients is not less than 0 or

greater than 1 (assuming none of f's values is less than 0 or greater than 1).

For some of the polynomials given above, a degree n can be found so that the degree-n polynomial is within ϵ of f, if f is continuous and meets other conditions. In general, to find the degree n, solve the error bound's equation for n and round the solution up to the nearest integer. See the table below, where:

- M_r is not less than the maximum of the absolute value of f's r-th derivative.
- H_r is not less than f's r-th derivative's Hölder constant (for the given Hölder exponent α).
- L_r is not less than f's r-th derivative's Lipschitz constant.

If $f(\lambda)$:	**Then the following polynomial:**	**Is close to *f* with the following error bound:**	
Has Hölder continuous second derivative (see "Definitions").	$U_{n, 2}$.	ε = $(5H_2+4M_2)$ / $(32 n^{1+\alpha/2})$.	n=max(3, $(32\epsil
Has Lipschitz continuous second derivative.	$U_{n, 2}$.	ε = $(5L_2+4M_2)$ / $(32 n^{3/2})$.	n=max(3, $(32\epsil
Has Lipschitz continuous second derivative.	$Q_{n-2,2}$.	$\varepsilon = 0.098585 L_2/((n-2)^{3/2})$.	n=max(4, / $(\epsilo
Has continuous	$L_{2	$\varepsilon = (3*\text{sqrt}(3-4/n)/4)*M_3/n^2 <$ (3*sqrt(3)/4)*M	n=max(6, \sqrt{M_3 max(6,cei sqrt(M

continuous third derivative.	$L_{2, n/2}$.	$(3*\mathrm{sqrt}(3)/4)*M_3/n^2 < 1.29904*M_3/n^2 \leq 1.29904*M_3/n^{3/2}$.	$\mathrm{sqrt}(M_3/\varepsilon$ ceil($((1.2 $\epsilon) odd, add 1
Has Hölder continuous third derivative.	$U_{n, 2}$.	ε = $(9H_3+8M_2+8M_3)$ / $(64 n^{(3+\alpha)/2})$.	n=max(6, ceil($((9H $(64\epsil
Has Lipschitz continuous third derivative.	$U_{n, 2}$.	ε = $(9H_3+8M_2+8M_3)$ / $(64 n^2)$.	n=max(6, ceil($((9H $(64\epsil
Has Lipschitz continuous third derivative.	$L_{3, n/4}$.	$\varepsilon = L_3/(8*n^2)$.	n=max(4, sqrt(L_3/ε) max(4,cei sqrt(L_3/ε) nearest m
Has Lipschitz continuous derivative.	$B_n(f)$.	$\varepsilon = L_1/(8*n)$.	n = ceil(L
Has Hölder continuous derivative.	$B_n(f)$.	$\varepsilon = H_1/(4*n^{(1+\alpha)/2})$.	n = ceil((

Is Hölder continuous.	$B_n(f)$.	$\epsilon = H_0$*(1/(4*n))$^{\alpha/2}$.	n = ceil((
Is Lipschitz continuous.	$B_n(f)$.	$\epsilon = L_0$*sqrt(1/(4*n)).	n = ceil((
Is Lipschitz continuous.	$B_n(f)$.	ϵ = $\frac{4306+837\sqrt{6}}{5832} L_0/n^{1/2}$ < $1.08989 L_0/n^{1/2}$.	n=ceil((

> **Note:** In addition, by analyzing the proof of Theorem 2.4 of Güntürk and Li (2021, sec. 3.3)[17], the following error bounds for $U_{n, 3}$ *appear* to be true:
>
> - If $f(\lambda)$ has continuous fifth derivative: ϵ = 4.0421*max(M_0,...,M_5)/$n^{5/2}$.
> - If $f(\lambda)$ has continuous sixth derivative: ϵ = 4.8457*max(M_0,...,M_6)/n^3.

Bernstein polynomials ($B_n(f)$) have the advantages that only one Bernstein coefficient has to be found per run and that the coefficient will be bounded by 0 and 1 if $f(\lambda)$ is. But their disadvantage is that they approach f slowly in general, at a rate no faster than a rate proportional to $1/n$ (Voronovskaya 1932)[18].

On the other hand, polynomials other than Bernstein polynomials ($B_n(f)$) can approach f faster in many cases than $B_n(f)$, but are not necessarily bounded by 0 and 1. For these polynomials, the following process can be used to calculate the required degree n, given an error tolerance of ϵ, assuming none of f's values is less than 0 or greater than 1.

1. Determine whether f is described in the table above. Let A be the minimum of f on the closed unit interval and let B be the maximum of f there.
2. If $0 < A \le B < 1$, calculate n as given in the table above, but with $\epsilon=\min(\epsilon, A, 1-B)$, and stop.
3. Propositions B1, B2, and B3 in the **appendix** give conditions on f so that $W_{n,2}$ or $W_{n,3}$ (as the case may be) will be

nonnegative. If *B* is less than 1 and any of those conditions is met, calculate n as given in the table above, but with $\epsilon=\min(\epsilon, 1-B)$. (For B3, set n to max(n, m), where m is given in that proposition.) Then stop; $W_{n,2}$ or $W_{n,3}$ will now be bounded by 0 and 1.

4. Calculate n as given in the table above. Then, if any Bernstein coefficient of the resulting polynomial is less than 0 or greater than 1, double the value of n until this condition is no longer true.

The resulting polynomial of degree n will be within ϵ of $f(\lambda)$.

> **Notes:**
>
> 1. A polynomial's Bernstein coefficients can be rounded to multiples of δ (where $0 \lt\delta\le 1$) by setting either—
>
>   - c=floor(c/δ) * δ (rounding down), or
>   - c=floor($c/\delta + 1/2$) * δ (rounding to the nearest multiple),
>
>   for each coefficient c. The new polynomial will differ from the old one by at most δ. (Thus, to find a polynomial with multiple-of-δ coefficients that approximates f with error ϵ [which must be greater than δ], first find a polynomial with error $\epsilon - \delta$, then round that polynomial's coefficients as given here.)
>
> 2. *Gevrey's hierarchy* is a class of "smooth" functions with known bounds on their derivatives. A function $f(\lambda)$ belongs in *Gevrey's hierarchy* if there are $B\ge 1$, $l\ge 1$, $\gamma\ge 1$ such that f's n-th derivative's absolute value is not greater than $Bl^n n^{\gamma n}$ for every $n\ge 1$ (Kawamura et al. 2015)[19]; see also (Gevrey 1918)[20]). In this case, for each $n\ge 1$—
>
>   - the n-th derivative of f is continuous and has a maximum absolute value of at most $Bl^n n^{\gamma n}$, and
>   - the $(n-1)$-th derivative of f is Lipschitz continuous with Lipschitz constant at most $Bl^n n^{\gamma n}$.

> *Gevrey's hierarchy* with $\gamma=1$ is the class of functions equaling power series (see next section).
>
> 3. Tsai and Farouki (2001, section 3.2)[21], describes how to rewrite a polynomial of degree m in Bernstein form to a polynomial of higher degree in Bernstein form. This is also known as *degree elevation*.

4.1 Taylor Polynomials for "Smooth" Functions

If $f(\lambda)$ is "smooth" enough on the closed unit interval and if ϵ is big enough, then Taylor's theorem shows how to build a polynomial that comes within ϵ of f. In this section f may but need not be writable as a power series.

Let $n\ge 0$ be an integer, and let $f^{(i)}$ be the i-th derivative of $f(\lambda)$. Suppose that—

1. f is continuous on the closed unit interval, and
2. f satisfies $\epsilon\le f(0)\le 1-\epsilon$ and $\epsilon\le f(1)\le 1-\epsilon$, and
3. f satisfies $\epsilon\lt f(\lambda)\lt 1-\epsilon$ whenever $0\lt\lambda\lt 1$, and
4. f's $(n+1)$-th derivative is continuous and satisfies $\epsilon\ge M/((n+1)!)$, where M is not less than the maximum of the absolute value of that derivative, and
5. $f(0)$ is known as well as $f^{(1)}(0), ..., f^{(n)}(0)$.

Then the n-th *Taylor polynomial* centered at 0, given below, is within ϵ of f: $$P(\lambda) = a_0 \lambda^0 + a_1 \lambda^1 + ... + a_n \lambda^n,$$

where $a_0 = f(0)$ and $a_i = f^{(i)}(0)/(i!)$ for $i\ge 1$.

Items 2 and 3 above are not needed to find a polynomial within ϵ of f, but they *are* needed to ensure the Taylor polynomial's Bernstein coefficients will lie in the closed unit interval, as described after the note.

> **Note:** If $f(\lambda)$ can be rewritten as a power series, namely $f(\lambda) = c_0 \lambda^0 + c_1 \lambda^1 + ... + c_i \lambda^i + ...$ whenever $0\le\lambda\le 1$ (so that f has a

> continuous k-th derivative for every k), and if the coefficients c_i—
>
> - are each greater than 0,
> - form a nowhere increasing sequence (example: (1/4, 1/8, 1/8, 1/16, ...)), and
> - meet the so-called "ratio test",
>
> then the algorithms in Carvalho and Moreira (2022)[22] can be used to find the first n+1 coefficients such that $P(\lambda)$ is within ϵ of f (see also the appendix).

Given the Taylor polynomial P, the algorithm to find the polynomial's Bernstein coefficients is as follows:

- Rewrite $P(\lambda)$ as a polynomial in Bernstein form. (One way to transform a polynomial to Bernstein form, given the "power" coefficients $a_0, ..., a_n$, is the so-called "matrix method" from Ray and Nataraj (2012)[23].) Let $b_0, ..., b_n$ be the Bernstein-form polynomial's coefficients. If any of those coefficients is less than 0 or greater than 1, double the value of n and rewrite P to Bernstein form of degree n, until none of the coefficients is less than 0 or greater than 1.

The result will be a polynomial of degree n with $(n+1)$ Bernstein coefficients.

4.2 Approximating an Integral

Roughly speaking, the *integral* of *f*(*x*) on an interval [*a*, *b*] is the "area under the graph" of that function when the function is restricted to that interval. If *f* is continuous there, this is the value that $\frac{1}{n} (f(a+(b-a)(1-\frac{1}{2})/n)+f(a+(b-a)(2-\frac{1}{2})/n)+...+f(a+(b-a)(n-\frac{1}{2})/n))$ approaches as n gets larger and larger.

If a polynomial is in Bernstein form of degree n, and is defined on the closed unit interval:

- The polynomial's integral on the closed unit interval is equal to the average of its $(n+1)$ Bernstein coefficients; that is, the integral is found by adding those coefficients together, then dividing by $(n+1)$ (Tsai and Farouki 2001)[24].
- If the polynomial is within ϵ of a function $f(\lambda)$ on

the closed unit interval, then its integral on that interval will be within ϵ of the integral of f.

> **Note:** A pair of articles by Konečný and Neumann discuss approximating the integral (and maximum) of a class of functions efficiently using polynomials or piecewise functions with polynomials as the pieces: Konečný and Neumann (2021)[25], Konečný and Neumann (2019)[26].

5 Approximations by Rational Functions

Consider the class of rational functions $p(\lambda)/q(\lambda)$ that map the closed unit interval to itself, where $q(\lambda)$ is in Bernstein form with non-negative coefficients. Then rational functions of this kind are not much better than polynomials in approximating $f(\lambda)$ when—

- the k-th derivative of f is continuous on the open interval (0, 1), but not the $(k+1)$-th derivative (Borwein 1979, section 3)[27], *or*
- $f(\lambda)$ is writable as $a_0 \lambda^0 + a_1 \lambda^1 + ...$, where $a_k\ge(k+1) a_{k+1}\ge 0$ whenever $k\ge 0$ (Borwein 1980)[28].

In addition, rational functions are not much better than polynomials in approximating $f(\lambda)$ when—

- the k-th derivative of f is continuous on the half-open interval (0, 1], but not the $(k+1)$-th derivative, *and*
- the rational function has no root that is a complex number whose real part is between 0 and 1 (Borwein 1979, theorem 29)[29].

6 Request for Additional Methods

Readers are requested to let me know of additional solutions to the following problem:

Let $f(\lambda)$ be continuous and map the closed unit interval to itself. Given $\epsilon\gt 0$, and given that $f(\lambda)$ belongs to a large class of functions (for example, it has a continuous, Lipschitz *continuous, concave, or nowhere decreasing k-th derivative for some integer k, or any combination of these), compute the Bernstein coefficients of a polynomial or rational function (of some degree n) that is within ϵ of $f(\lambda)$.*

The approximation error must be no more than a constant times $1/n^{r/2}$ if the given class has only functions with continuous r-th derivative.

Methods that use only integer arithmetic and addition and multiplication of rational numbers are preferred (thus, Chebyshev methods and other methods that involve cosines, sines, π, $\exp$, and $\ln$ are not preferred).

See also the **open questions**.

7 Notes

8 Appendix

8.1 Results Used in Approximations by Polynomials

Lemma A1: Let— $$f(x)=a_0 x^0 + a_1 x^1 + ...,$$ where the a_i are constants each 0 or greater and sum to a finite value and where $0\le x\le 1$ (the domain is the closed unit interval). Then f is convex and has a maximum at 1.

Proof: By inspection, $f(x)$ is a power series and is nonnegative wherever $x\ge 0$ (and thus wherever $0\le x\le 1$). Each of its terms has a maximum at 1 since—

- for $n=0$, $a_0 x^0=a_0$ is a non-negative constant (which trivially reaches its maximum at 1), and
- for each n where $a_0 = 0$, $a_0 x^n$ is the constant 0 (which trivially reaches its maximum at 1), and
- for each other n, x^n is a strictly increasing function and multiplying that by a_n (a positive constant) doesn't change whether it's strictly increasing.

Since all of these terms have a maximum at 1 on the domain, so does their sum.

The derivative of f is— $$f'(x) = 1\cdot a_1 x^0 + ... + i\cdot a_i x^{i-1} + ...,$$ which is still a power series with nonnegative values of a_n, so the proof so far applies to f' instead of f. By induction, the proof so far applies to all derivatives of f, including its second derivative.

Now, since the second derivative is nonnegative wherever $x\ge 0$, and thus on its domain, f is convex, which completes the proof. □

Proposition A2: For a function $f(x)$ as in Lemma A1, let— $$g_n(x)=a_0 x^0 + ... + a_n x^n,$$ and have the same domain as f. Then for every $n\ge 1$, $g_n(x)$ is within ϵ of $f(x)$, where $\epsilon = f(1) - g_n(1)$.

Proof: g_n, consisting of the first $n+1$ terms of f, is a power series with nonnegative coefficients, so by Lemma A1, it has a maximum at 1. The same is true for $f-g_n$, consisting of the remaining terms of f. Since the latter has a maximum at 1, the maximum error is $\epsilon = f(1)-g_n(1)$. □

For a function f described in Lemma A1, $f(1)=a_0 1^0 + a_1 1^1 + ... = a_0 + a_1+...$, and f's error behavior is described at the point 1, so the algorithms given in Carvalho and Moreira (2022)[30] — which apply to infinite sums — can be used to "cut off" f at a certain number of terms and do so with a controlled error.

Proposition B1: Let $f(\lambda)$ map the closed unit interval to itself and be continuous and concave. Then $W_{n,2}$ and $W_{n,3}$ (as defined in "For Certain Functions") are nonnegative on the closed unit interval.

Proof: For $W_{n,2}$ it's enough to prove that $B_n(f)\le f$ for every $n\ge 1$. This is the case because of Jensen's inequality and because f is concave.

For $W_{n,3}$ it must also be shown that $B_n(B_n(f)(\lambda))$ is nonnegative. For this, using only the fact that f maps the closed unit interval to itself, $B_n(f)$ will have Bernstein coefficients in that interval (each coefficient is a value of f) and so will likewise map the closed unit interval to itself (Qian et al. 2011)[31]. Thus, by induction, $B_n(B_n(f)(\lambda))$ is nonnegative. The discussion for $W_{n,2}$ also shows that $(f - B_n(f))$ is nonnegative as well. Thus, $W_{n,3}$ is nonnegative on the closed unit interval. □

Proposition B2: Let $f(\lambda)$ map the closed unit interval to itself, be continuous, nowhere decreasing, and subadditive, and equal 0 at 0. Then $W_{n,2}$ is nonnegative on the closed unit interval.

Proof: The assumptions on f imply that $B_n(f)\le 2 f$ (Li 2000)[32], showing that $W_{n,2}$ is nonnegative on the closed unit interval. □

> **Note:** A subadditive function f has the property that $f(a+b) \le f(a)+f(b)$ whenever a, b, and $a+b$ are in f's domain.

Proposition B3: Let $f(\lambda)$ map the closed unit interval to itself and have a Lipschitz continuous derivative with Lipschitz constant L. If $f(\lambda) \ge \frac{L \lambda(1-\lambda)}{2m}$ on f's domain, for some $m\ge 1$, then $W_{n,2}$ is nonnegative there, for every $n\ge m$.

Proof: Let $E(\lambda, n) = \frac{L \lambda(1-\lambda)}{2n}$. Lorentz (1963)[33] showed that with this Lipschitz derivative assumption on f, B_n differs from $f(\lambda)$ by no more than $E(\lambda, n)$ for every $n\ge 1$ and wherever $0\lt\lambda\lt 1$. As is well known, $B_n(0)=f(0)$ and $B_n(1)=f(1)$. By inspection, $E(\lambda, n)$ is biggest when $n=1$ and decreases as n increases. Assuming the worst case that $B_n(\lambda) = f(\lambda) + E(\lambda, m)$, it follows that $W_{n,2}=2 f(\lambda) - B_n(\lambda)\ge 2 f(\lambda) - f(\lambda) - E(\lambda, m) = f(\lambda) - E(\lambda, m)\ge 0$ whenever $f(\lambda)\ge E(\lambda, m)$. Because $E(\lambda, k+1)\le E(\lambda,k)$ for every $k\ge 1$, the preceding sentence holds true for every $n\ge m$. □

The following results deal with a useful quantity when discussing the error in approximating a function by Bernstein polynomials. Suppose a coin shows heads with probability p, and n independent tosses of the coin are made. Then the total number of heads X follows a *binomial distribution*, and the r-th central moment of that distribution is as follows: $$T_{n,r}(p) = \mathbb{E}[(X-\mathbb{E}[X])^r] = \sum_{k=0}^n (k-np)^r{n \choose k}p^k (1-p)^{n-k},$$ where $\mathbb{E}[.]$ is the expected value ("long-run average"). (Traditionally, another central moment, that of X/n or the ratio of heads to tosses, is denoted $S_{n,r}(p)=T_{n,r}(p)/n^r=\mathbb{E}

[(X/n-\mathbb{E}[X/n])^r]$. T and S are notations of S.N. Bernstein, known for Bernstein polynomials.) The following results bound the absolute value of T and S.[34]

Result B4 (Molteni (2022)[35]): If r is an even integer such that $0\le r\le 44$, then for every integer $n\ge 1$, $|T_{n,r}(p)|\le (r!)/(((r/2)!)8^{r/2}) n^{r/2}$ and $|S_{n,r}(p)| \le (r!)/(((r/2)!)8^{r/2})\cdot(1/n^{r/2})$.

Result B4A (Adell et al. (2015)[36]): For every odd integer $r\ge 1$, $T_{n,r}(p)$ is positive whenever $0\le p\lt 1/2$, and negative whenever $1/2\lt p\le 1$.

Lemma B5: For every integer $n\ge 1$:

- $|S_{n,0}(p)|=1=1\cdot(p(1-p)/n)^{0/2}$.
- $|S_{n,1}(p)|=0=0\cdot(p(1-p)/n)^{1/2}$.
- $|S_{n,2}(p)|=p(1-p)/n=1\cdot(p(1-p)/n)^{2/2}$.

The proof is straightforward.

Result B6 (Adell and Cárdenas-Morales (2018)[37]): Let $\sigma(r,t) = (r!)/(((r/2)!)t^{r/2})$. If $r\ge 0$ is an even integer, then—

- for every integer $n\ge 1$, $|T_{n,r}(p)|\le \sigma(r,6)n^{r/2}$ and $|S_{n,r}(p)| \le \sigma(r,6)/n^{r/2}$, and
- for every integer $n\ge 1$, $|T_{n,r}(1/2)|\le \sigma(r,8)n^{r/2}$ and $|S_{n,r}(1/2)| \le \sigma(r,8)/n^{r/2}$.

Lemma B9: Let $f(\lambda)$ have a Lipschitz continuous r-th derivative on the closed unit interval (see "**Definitions**"), where $r\ge 0$ is an integer, and let M be equal to or greater than the r-th derivative's Lipschitz constant. Then, for every $0\le x_0 \le 1$:

1. f can be written as $f(\lambda) = R_{f,r}(\lambda, x_0) + f(x_0) + \sum_{i=1}^{r} (\lambda-x_0)^i f^{(i)}(x_0)/(i!)$ where $f^{(i)}$ is the i-th derivative of f.
2. If r is odd, $|B_n(R_{f,r}(\lambda, x_0))(x_0)| \le M/((((r+1)/2)!)(\beta n)^{(r+1)/2})$ for every integer $n\ge 1$, where β is 8 if $r\le 43$ and 6 otherwise.
3. If $r=0$, $|B_n(R_{f,r}(\lambda, x_0))(x_0)| \le M/(2n^{1/2})$ for every integer $n\ge 1$.
4. If r is even and greater than 0, $|B_n(R_{f,r}(\lambda, x_0))(x_0)| \le \frac{M}{(r+1)!n^{(r+1)/2}}\left(\frac{2\cdot(r+1)!(r)!}

{\gamma^{r+1}((r/2)!)^2}\right)^{1/2}$ for every integer $n\ge 2$, where γ is 8 if $r\le 42$ and 6 otherwise.

Proof: The well-known result of part 1 says f equals the *Taylor polynomial* of degree r at x_0 plus the *Lagrange remainder*, $R_{f,r}(\lambda, x_0)$. A result found in Gonska et al. (2006)[38], which applies for any integer $r\ge 0$, bounds that Lagrange remainder [39]. By that result, because f's r-th derivative is Lipschitz continuous— $$|R_{f,r}(\lambda, x_0)|\le \frac{|\lambda-x_0|^r}{r!} M \frac{|\lambda-x_0|}{r+1}=M\frac{|\lambda-x_0|^{r+1}}{(r+1)!}.$$

The goal is now to bound the Bernstein polynomial of $|\lambda-x_0|^{r+1}$. This is easiest to do if r is odd.

If r is odd, then $(\lambda-x_0)^{r+1} = |\lambda-x_0|^{r+1}$, so by Results B4 and B6, the Bernstein polynomial of $|\lambda-x_0|^{r+1}$ can be bounded as follows: $$|B_n((\lambda-x_0)^{r+1})(x_0)| \le \frac{(r+1)!}{(((r+1)/2)!)\beta^{(r+1)/2}}\frac{1}{n^{(r+1)/2}} = \sigma(r,n),$$ where β is 8 if $r\le 43$ and 6 otherwise. Therefore— $$|B_n(R_{f,r}(\lambda, x_0))(x_0)| \le \frac{M}{(r+1)!} |B_n((\lambda-x_0)^{r+1})(x_0)|$$ $$\le \frac{M}{(r+1)!}\frac{(r+1)!}{(((r+1)/2)!)\beta^{(r+1)/2}}\frac{1}{n^{(r+1)/2}} = \frac{M}{(((r+1)/2)!)(\beta n)^{(r+1)/2}}.$$

If r is 0, then the Bernstein polynomial of $|\lambda-x_0|^{1}$ is bounded by $\sqrt{x_0(1-x_0)/n}$ for every integer $n\ge 1$ (Cheng 1983)[40], so— $$|B_n(R_{f,r}(\lambda, x_0))(x_0)| \le \frac{M}{(r+1)!}\sqrt{x_0(1-x_0)/n}\le \frac{M}{(r+1)!}\frac{1}{2n^{1/2}}=\frac{M}{2n^{1/2}}.$$

If r is even and greater than 0, the Bernstein polynomial for $|\lambda-x_0|^{r+1}$ can be bounded as follows for every $n\ge 2$, using **Schwarz's inequality** (see also Bojanic and Shisha [1975][41] for the case $r=4$): $$B_n(|\lambda-x_0|^{r+1})(x_0)=B_n((|\lambda-x_0|^{r/2}|\lambda-x_0|^{(r+2)/2})^2)(x_0)$$ $$\le\sqrt{|S_{n,r}(x_0)|}\sqrt{|S_{n,r+2}(x_0)|}\le\sqrt{\sigma(r,n)}\sqrt{\sigma(r+2,n)}$$ $$\le\frac{1}{n^{(r+1)/2}}\left(\frac{2\cdot(r+1)!(r)!}{\gamma^{r+1}((r/2)!)^2}\right)^{1/2},$$where γ is 8 if $r\le 42$ and 6 otherwise. Therefore— $$|B_n(R_{f,r}(\lambda, x_0))(x_0)| \le \frac{M}{(r+1)!\cdot n^{(r+1)/2}}\left(\frac{2\cdot(r+1)!(r)!}{\gamma^{r+1}((r/2)!)^2}\right)^{1/2}. $$□

> **Notes:**
>
> 1. If a function $f(\lambda)$ has a continuous r-th derivative on its domain (where $r\ge 0$ is an integer), then by Taylor's theorem for real variables, $R_{f,r}(\lambda, x_0)$, is writable as $f^{(r)}(c)\cdot (\lambda-x_0)^r /(r!),$ for some c between λ and x_0 (and thus on f's domain) (DLMF [42] **equation 1.4.36**). Thus, by this estimate, $|R_{f,r}(\lambda, x_0)| \le \frac{M}{r!} (\lambda-x_0)^r.$
> 2. It would be interesting to strengthen this lemma, at least for $r\le 10$, with a bound of the form $MC\cdot\max(1/n, (x_0(1-x_0)/n)^{1/2})^{r+1}$, where C is an explicitly given constant depending on r, which is possible because the Bernstein polynomial of $|\lambda-x_0|^{r+1}$ can be bounded in this way (Lorentz 1966)[43].

Corollary B9A: Let $f(\lambda)$ have a Lipschitz continuous r-th derivative on the closed unit interval, and let M be that r-th derivative's Lipschitz constant or greater. Then, for every $0\le x_0 \le 1$:

If r is:	Then $\text{abs}(B_n(R_{f,r}(\lambda, x_0))(x_0)) \le$ ...
0.	$M/(2 n^{1/2})$ for every integer $n\ge 1$.
0.	$M\cdot\sqrt{x_0(1-x_0)/n}$ for every integer $n\ge 1$.
1.	$M/(8 n)$ for every integer $n\ge 1$.
2.	$\sqrt{3}M/(48 n^{3/2}) < 0.03609 M/n^{3/2}$ for every integer $n\ge 2$.
3.	$M/(128 n^2)$ for every integer $n\ge 1$.
4.	$\sqrt{5}M/(1280 n^{5/2}) < 0.001747 M/n^{5/2}$ for every integer $n\ge 2$.
5.	$M/(3072 n^3)$ for every integer $n\ge 1$.

Proposition B10: Let $f(\lambda)$ have a Lipschitz continuous third derivative on the closed unit interval. For each $n\ge 4$ that is divisible by 4, let $L_{3,n/4}(f) = (1/3)\cdot B_{n/4}(f) - 2\cdot

B_{n/2}(f) + (8/3)\cdot B_{n}(f)$. Then $L_{3,n/4}(f)$ is within $M/(8 n^2)$ of f, where M is the maximum of the absolute value of that fourth derivative.

Proof: This proof is inspired by the proof technique in Tachev (2022)[44].

Because f has a Lipschitz continuous third derivative, f has the Lagrange remainder $R_{f,3}(\lambda, x_0)$ given in Lemma B9 and Corollary B9A.

It is known that $L_{3,n/4}$ is a linear operator that preserves polynomials of degree 3 or less, so that $L_{3,n/4}(f) = f$ whenever f is a polynomial of degree 3 or less (Ditzian and Totik 1987)[45], Butzer (1955)[46], May (1976)[47]. Because of this, it can be assumed without loss of generality that $f(x_0)=0$.

Therefore—$$|L_{3,n/4}(f(\lambda))(x_0) - f(x_0)| = |L_{3,n/4}(R_{f,3}(\lambda, x_0))|.$$ Now denote σ_n as the maximum of $|B_n(R_{f,3}(\lambda, x_0))(x_0)|$ over $0\le x_0\le 1$. In turn (using Corollary B9A)— $$|L_{3,n/4}(R_{f,3}(\lambda, x_0))| \le(1/3)\cdot\sigma_{n/4} + 2\cdot\sigma_{n/2}+(8/3)\cdot\sigma_n$$ $$\le (1/3)\frac{M}{128 (n/4)^2} + 2\frac{M}{128 (n/2)^2} + (8/3)\frac{M}{128 n^2} =M/(8 n^2).$$ □

Proposition B10A: Let $f(\lambda)$ have a Lipschitz continuous second derivative on the closed unit interval. Let $Q_{n,2}(f)=B_n(f)(x)-\frac{x(1-x)}{2n} B_n(f'')(x)$ be the *Lorentz operator* of order 2 (Holtz et al. 2011)[48], (Lorentz 1966)[49], which is a polynomial in Bernstein form of degree $n+2$. Then if $n\ge 2$ is an integer, $Q_{n,2}(f)$ is within $\frac{M(\sqrt{3}+3)}{48 n^{3/2}} \lt 0.098585 M/(n^{3/2})$ of f, where M is that second derivative's Lipschitz constant or greater.

Proof: Since $Q_{n,2}(f)$ preserves polynomials of degree 2 or less (Holtz et al. 2011, Lemma 14)[50] and since f has a Lipschitz continuous second derivative, f has the Lagrange remainder $R_{f,2}(\lambda, x_0)$ given in Lemma B9, and f'', the second derivative of f, has the Lagrange remainder $R_{f'',0}(\lambda, x_0)$. Thus, using Corollary B9A, the error bound can be written as— $$|Q_{n,2}(f(\lambda))(x_0) - f(x_0)|\le|B_n(R_{f,2}(\lambda, x_0))| +

\frac{x_0(1-x_0)}{2n} |B_n(R_{f'',0}(\lambda,x_0))|$$ $$\le \frac{\sqrt{3}M}{48 n^{3/2}} + \frac{1}{8n} \frac{M}{2 n^{1/2}} = \frac{M(\sqrt{3}+3)}{48 n^{3/2}} \lt 0.098585 M/(n^{3/2}).$$ □

Corollary B10B: Let $f(\lambda)$ have a continuous second derivative on the closed unit interval. Then $B_n(f)$ is within $\frac{M}{8n}$ of f, where M is the maximum of that second derivative's absolute value or greater.

Proof: Follows from Lorentz (1963)[51] and the well-known fact that M is an upper bound of f's first derivative's (minimal) Lipschitz constant. □

In the following propositions, $f^{(r)}$ means the r-th derivative of the function f and $\max(|f|)$ means the maximum of the absolute value of the function f.

Proposition B10C: Let $f(\lambda)$ have a Hölder continuous second derivative on the closed unit interval, with Hölder exponent α ($0\lt\alpha\le 1$) and Hölder constant L or less. Let $U_{n,2}(f)=B_n(2f-B_n(f))$ be f's iterated Boolean sum of order 2 of Bernstein polynomials. Then if $n\ge 3$ is an integer, the error in approximating f with $U_{n,2}(f)$ is as follows: $$|f-U_{n,2}(f)|\le \frac{M_2}{8 n^{2}} + 5 L/(32 n^{1+\alpha/2}) \le ((5L+4M_2)/32)/n^{1+\alpha/2},$$ where M_2 is the maximum of that second derivative's absolute value or greater.

Proof: This proof is inspired by a result in Draganov (2004, Theorem 4.1)[52].

The error to be bounded can be expressed as $|(B_n(f)-f)(B_n(f)-f)|$. Following Corollary B10B: $$|(B_n(f)-f)(B_n(f)-f)|\le \frac{1}{8n} \max(|(B_n(f))^{(2)}-f^{(2)}|).\tag{B10C-1}$$ It thus remains to estimate the right-hand side of the bound. A result by Knoop and Pottinger (1976)[53], which works for every $n\ge 3$, is what is known as a *simultaneous approximation* error bound, showing that the second derivative of the Bernstein polynomial approaches that of f as n increases. Using this result: $$|(B_n(f))^{(2)}-f^{(2)}| \le \frac{1}{n} M_2+(5/4) L/n^{\alpha/2},$$ so— $$|(B_n(f)-f)(B_n(f)-f)|\le \frac{1}{8n} \left(\frac{1}{n} M_2+(5/4) L/n^{\alpha/2}\right)$$ $$\le \frac{M_2}{8 n^{2}} + \frac{5L}{32 n^{1+\alpha/2}}\le \frac{5L+4M_2}{32}\frac{1}{n^{1+\alpha/2}}.$$ □

> **Note**: The error bound $0.75 M_2/n^2$ for $U_{n,2}$ is false in general if $f(\lambda)$ is assumed only to be non-negative, concave, and have a continuous second derivative on the closed unit interval. A counterexample is $f(\lambda)=(1-(1-2\lambda)^{2.5})/2$ if $\lambda <1/2$ and $(1-(2\lambda-1)^{2.5})/2$ otherwise.

Proposition B10D: Let $f(\lambda)$ have a Hölder continuous third derivative on the closed unit interval, with Hölder exponent α ($0\lt\alpha\le 1$) and Hölder constant L or less. If $n\ge 6$ is an integer, the error in approximating f with $U_{n,2}(f)$ is as follows: $$|f-U_{n,2}(f)|\le \frac{\max(|f^{(2)}|)+\max(|f^{(3)}|)}{8n^2}+9L/(64 n^{(3+\alpha)/2})$$ $$\le \frac{9L+8\max(|f^{(2)}|)+8\max(|f^{(3)}|)}{64n^{(3+\alpha)/2}}.$$

Proof: Again, the goal is to estimate the right-hand side of (B10C-1). But this time, a different simultaneous approximation bound is employed, namely a result from Kacsó (2002)[54], which in this case works if $n\ge\max(r+2,(r+1)r)=6$, where $r=2$. By that result: $$|(B_n(f))^{(2)}-f^{(2)}| \le \frac{r(r-1)}{2n} M_2+\frac{r M_3}{2n}+\frac{9}{8}\omega_2(f^{(2)},1/n^{1/2})$$ $$\le \frac{1}{n} M_2+M_3/n+\frac{9}{8} L/n^{(1+\alpha)/2},$$ where $r=2$, $M_2 = \max(|f^{(2)}|)$, and $M_3=\max(|f^{(3)}|)$, using properties of ω_2, the second-order modulus of continuity of $f^{(2)}$, given in Stancu et al. (2001)[55]. Therefore— $$|(B_n(f)-f)(B_n(f)-f)|\le \frac{1}{8n} \left(\frac{1}{n} M_2+M_3/n+\frac{9}{8} L/n^{(1+\alpha)/2}\right)$$ $$\le \frac{M_2+M_3}{8n^2} + \frac{9L}{64 n^{(3+\alpha)/2}}\le \frac{9L+8M_2+8M_3}{64n^{(3+\alpha)/2}}.$$ □

In a similar way, it's possible to prove an error bound for $U_{n,3}$ that applies to functions with a Hölder continuous fourth or fifth derivative, by expressing the error bound as $|(B_n(f)-f)((B_n(f)-f)(B_n(f)-f))|$ and replacing the values for M_2, M_3, and L in the bound proved at the end of Proposition B10D with upper bounds for $|(B_n(f))^{(2)}-f^{(2)}|$, $|(B_n(f))^{(3)}-f^{(3)}|$, and $|(B_n(f))^{(4)}-f^{(4)}|$, respectively.

8.2 Chebyshev Interpolants

The following is a method that employs *Chebyshev interpolants* to compute the Bernstein coefficients of a polynomial that comes within ϵ of $f(\lambda)$, as long as f meets certain conditions. Because the method introduces a trigonometric function (the cosine function), it appears here in the appendix and it runs too slowly for real-time or "online" use; rather, this method is more suitable for pregenerating ("offline") the approximate version of a function known in advance.

- f must be continuous on the interval $[a, b]$ and must have an r-th derivative of *bounded variation*, as described later.
- Suppose f's domain is the interval $[a, b]$. Then the *Chebyshev interpolant* of degree n of f (Wang 2023)[56], (Trefethen 2013)[57] is— $$p(\lambda)=\sum_{k=0}^n c_k T_k(2\frac{\lambda-a}{b-a}-1),$$ where—
    - $c_k=\sigma(k,n)\frac{2}{n}\sum_{j=0}^n \sigma(j,n) f(\gamma(j,n))T_k(\cos(j\pi/n))$,
    - $\gamma(j,n) = a+(b-a)(\cos(j\pi/n)+1)/2$,
    - $\sigma(k,n)$ is 1/2 if k is 0 or n, and 1 otherwise, and
    - $T_k(x)$ is the k-th **Chebyshev polynomial of the first kind** (`chebyshevt(k,x)` in the SymPy computer algebra library).
- Let $r\ge 1$ and $n\gt r$ be integers. If f is continuous on the interval $[a, b]$ and f has an r-th derivative of *bounded variation*, then the degree-n Chebyshev interpolant of f is within $\left(\frac{(b-a)}{2}\right)^r\frac{4V}{\pi r(n-r)^r}$ of f, where V is the r-th derivative's *total variation* or greater. This relies on a theorem in chapter 7 of Trefethen (2013)[58] as well as a statement in note 1 at the end of this section.
    - If the r-th derivative is nowhere decreasing or nowhere increasing on the interval $[a, b]$, then V can equal abs($f(b)-f(a)$).
    - If the r-th derivative is Lipschitz continuous with Lipschitz constant M or less, then V can equal $M\cdot(b-a)$ (Kannan and Kreuger 1996)[59].
    - The required degree is thus $n=\text{ceil}(r+\frac{(b-a)}{2}(4V/(\pi r\epsilon))^{1/r})$ ≤ $\text{ceil}(r+\frac{(b-a)}{2}(1.2733 V/(r\epsilon))^{1/r})$, where $\epsilon>0$ is the desired error tolerance.

- If f is real analytic on $[a, b]$, a better error bound is possible, but describing it requires ideas from complex analysis that are too advanced for this article. See chapter 8 of Trefethen (2013)[60].

---

1. Compute the required degree n as given above, with error tolerance $\epsilon/2$.
2. Compute the (n plus one) coefficients of f's degree-n Chebyshev interpolant, call them $c_0, ..., c_n$.
3. Compute the $(n+1)\times(n+1)$ matrix M described in Theorem 1 of Rababah (2003)[61].
4. Multiply the matrix by the transposed coefficients $(c_0, ..., c_n)$ to get the polynomial's Bernstein coefficients $b_0, ..., b_n$. (Transposing means turning columns to rows and vice versa.)
5. For each i, replace the Bernstein coefficient b_i with $\text{floor}(b_i / (\epsilon/2) + 1/2) \cdot (\epsilon/2)$.
6. Return the Bernstein coefficients $b_0, ..., b_n$.

> **Notes:**
>
> 1. The following statement can be shown. Let $f(x)$ be continuous on the interval $[a, b]$. If the r-th derivative of f, call it FR, has total variation V, where $r\ge 1$, then $g(x) = FR(a+(b-a) (x+1)/2)$ has total variation $V\left(\frac{b-a}{2}\right)^r$ on the interval $[-1, 1]$.
> 2. The method in this section doesn't require $f(\lambda)$ to have a particular minimum or maximum. If f must map the closed unit interval to itself and the Bernstein coefficients must lie on that interval, the following changes to the method are needed:
>   - $f(\lambda)$ must be continuous on the closed unit interval ($a=0$, $b=1$) and take on only values in that interval.
>   - If any Bernstein coefficient returned by the method is less than 0 or greater than 1, double the value of n and repeat the method starting at step 2 until that condition is no longer true.
> 3. It would be of interest to build Chebyshev-like interpolants that sample $f(\lambda)$ at *rational* values of λ that get closer to the Chebyshev points (e.g., $\cos(j\pi/n)$) with increasing n, and to find results that provide explicit bounds (with no hidden constants) on the approximation error that are close to those for Chebyshev interpolants.

9 License

---

1. choose(n, k) = (1*2*3*...*n)/((1*...*k)*(1*...*(n−k))) = n!/(k! * (n − k)!) $={n \choose k}$ is a *binomial coefficient*, or the number of ways to choose k out of n labeled items. It can be calculated, for example, by calculating i/(n−i+1) for each integer i satisfying n−k+1 ≤ i ≤ n, then multiplying the results (Yannis Manolopoulos. 2002. "Binomial coefficient computation: recursion or iteration?", SIGCSE Bull. 34, 4 (December 2002), 65–67. DOI: **https://doi.org/10.1145/820127.820168**). For every m>0, choose(m, 0) = choose(m, m) = 1 and choose(m, 1) = choose(m, m−1) = m; also, in this document, choose(n, k) is 0 when k is less than 0 or greater than n.
   n! = 1*2*3*...*n is also known as n factorial; in this document, (0!) = 1.↩

2. Micchelli, Charles. "The saturation class and iterates of the Bernstein polynomials." Journal of Approximation Theory 8, no. 1 (1973): 1-18.↩

3. Guan, Zhong. "**Iterated Bernstein polynomial approximations**", arXiv:0909.0684 (2009).↩

4. Güntürk, C.S., Li, W., "**Approximation of functions with one-bit neural networks**", arXiv:2112.09181 [cs.LG], 2021.↩

5. Tachev, Gancho. "**Linear combinations of two Bernstein polynomials**", *Mathematical Foundations of Computing*, 2022.↩

6. Butzer, P.L., "Linear combinations of Bernstein polynomials", Canadian Journal of Mathematics 15 (1953).↩

7. Butzer, P.L., "Linear combinations of Bernstein polynomials", Canadian Journal of Mathematics 15 (1953).↩

8. Holtz, O., Nazarov, F., Peres, Y., "**New Coins from Old, Smoothly**", *Constructive Approximation* 33 (2011).↩

9. Bernstein, S. N. (1932). "Complément a l'article de E. Voronovskaya." CR Acad. URSS, 86-92.↩

10. G.G. Lorentz, "The degree of approximation by polynomials with positive coefficients", 1966.↩

11. Tachev, Gancho. "**Linear combinations of two Bernstein polynomials**", *Mathematical Foundations of Computing*, 2022.↩

12. G.G. Lorentz, "Inequalities and saturation classes for Bernstein polynomials", 1963.↩

13. Qian et al. suggested an *n* which has the upper bound *n*=ceil(1+max($2n$,$n^2 (2^{n}C)/\epsilon$)), where C is the maximum of f on its domain, but this is often much worse and works only if f is a polynomial (Qian, W., Riedel, M. D., & Rosenberg, I. (2011). Uniform approximation and Bernstein polynomials with coefficients in the unit interval. European Journal of Combinatorics, 32(3), 448-463).↩

14. Schurer and Steutel, "On an inequality of Lorentz in the theory of Bernstein polynomials", 1975.↩

15. Kac, M., "Une remarque sur les polynômes de M. S. Bernstein", Studia Math. 7, 1938.↩

16. Sikkema, P.C., "Der Wert einiger Konstanten in der Theorie der Approximation mit Bernstein-Polynomen", 1961.↩

17. Güntürk, C.S., Li, W., "**Approximation of functions with one-bit neural networks**", arXiv:2112.09181 [cs.LG], 2021.↩

18. E. Voronovskaya, "Détermination de la forme asymptotique d'approximation des fonctions par les polynômes de M. Bernstein", 1932.↩

19. Kawamura, Akitoshi, Norbert Müller, Carsten Rösnick, and Martin Ziegler. "**Computational benefit of smoothness: Parameterized bit-complexity of numerical operators on analytic functions and Gevrey's hierarchy**." Journal of Complexity 31, no. 5 (2015): 689-714.↩

20. M. Gevrey, "Sur la nature analytique des solutions des équations aux dérivées partielles", 1918.↩

21. Tsai, Y., Farouki, R.T., "Algorithm 812: BPOLY: An Object- Oriented Library of Numerical Algorithms for Polynomials in Bernstein Form", ACM Transactions on Mathematical Software, June 2001.↩
22. Carvalho, Luiz Max, and Guido A. Moreira. "**Adaptive truncation of infinite sums: applications to Statistics**", arXiv:2202.06121 (2022).↩
23. S. Ray, P.S.V. Nataraj, "A Matrix Method for Efficient Computation of Bernstein Coefficients", Reliable Computing 17(1), 2012.↩
24. Tsai, Y., Farouki, R.T., "Algorithm 812: BPOLY: An Object- Oriented Library of Numerical Algorithms for Polynomials in Bernstein Form", ACM Transactions on Mathematical Software, June 2001.↩
25. Konečný, Michal, and Eike Neumann. "Representations and evaluation strategies for feasibly approximable functions." Computability 10, no. 1 (2021): 63-89. Also in arXiv: **https://arxiv.org/pdf/1710.03702**↩
26. Konečný, Michal, and Eike Neumann. "**Implementing evaluation strategies for continuous real functions**", arXiv:1910.04891 (2019).↩
27. Borwein, P. B. (1979). Restricted uniform rational approximations (Doctoral dissertation, University of British Columbia).↩
28. Borwein, Peter B. "Approximations by rational functions with positive coefficients." Journal of Mathematical Analysis and Applications 74, no. 1 (1980): 144-151.↩
29. Borwein, P. B. (1979). Restricted uniform rational approximations (Doctoral dissertation, University of British Columbia).↩
30. Carvalho, Luiz Max, and Guido A. Moreira. "**Adaptive truncation of infinite sums: applications to Statistics**", arXiv:2202.06121 (2022).↩
31. Qian, Weikang, Marc D. Riedel, and Ivo Rosenberg. "Uniform approximation and Bernstein polynomials with coefficients in the unit interval." European Journal of Combinatorics 32, no. 3 (2011): 448-463.↩
32. Li, Zhongkai. "Bernstein polynomials and modulus of continuity." Journal of Approximation Theory 102, no. 1 (2000): 171-174.↩

33. G.G. Lorentz, "Inequalities and saturation classes for Bernstein polynomials", 1963.↩

34. *Summation notation*, involving the Greek capital sigma (Σ), is a way to write the sum of one or more terms of similar form. For example, $\sum_{k=0}^n g(k)$ means $g(0)+g(1)+...+g(n)$, and $\sum_{k\ge 0} g(k)$ means $g(0)+g(1)+...$.↩

35. Molteni, Giuseppe. "Explicit bounds for even moments of Bernstein’s polynomials." Journal of Approximation Theory 273 (2022): 105658.↩

36. Adell, J. A., Bustamante, J., & Quesada, J. M. (2015). Estimates for the moments of Bernstein polynomials. Journal of Mathematical Analysis and Applications, 432(1), 114-128.↩

37. Adell, J.A., Cárdenas-Morales, D., "**Quantitative generalized Voronovskaja’s formulae for Bernstein polynomials**", Journal of Approximation Theory 231, July 2018.↩

38. Gonska, H.H., Piţul, P., Raşa, I., "On Peano's form of the Taylor remainder, Voronovskaja's theorem and the commutator of positive linear operators", In Numerical Analysis and Approximation Theory, 2006.↩

39. The result from Gonska et al. actually applies if the r-th derivative belongs to a broader class of continuous functions than Lipschitz continuous functions, but this feature is not used in this proof.↩

40. Cheng, F., "On the rate of convergence of Bernstein polynomials of functions of bounded variation", Journal of Approximation Theory 39 (1983).↩

41. G.G. Lorentz, *Bernstein polynomials*, 1953.↩

42. *NIST Digital Library of Mathematical Functions*, **https://dlmf.nist.gov/** , Release 1.1.9 of 2023-03-15.↩

43. G.G. Lorentz, "The degree of approximation by polynomials with positive coefficients", 1966.↩

44. Tachev, Gancho. "**Linear combinations of two Bernstein polynomials**", *Mathematical Foundations of Computing*, 2022.↩

45. Ditzian, Z., Totik, V., *Moduli of Smoothness*, 1987.↩

46. Butzer, P.L., "Linear combinations of Bernstein polynomials", Canadian Journal of Mathematics 15 (1953).↩

47. May, C.P., "Saturation and inverse theorems for a class of exponential-type operators", Canadian Journal of Mathematics 28 (1976).↩

48. Holtz, O., Nazarov, F., Peres, Y., "**New Coins from Old, Smoothly**", *Constructive Approximation* 33 (2011).↩

49. G.G. Lorentz, "The degree of approximation by polynomials with positive coefficients", 1966.↩

50. Holtz, O., Nazarov, F., Peres, Y., "**New Coins from Old, Smoothly**", *Constructive Approximation* 33 (2011).↩

51. G.G. Lorentz, "Inequalities and saturation classes for Bernstein polynomials", 1963.↩

52. Draganov, Borislav R. "On simultaneous approximation by iterated Boolean sums of Bernstein operators." Results in Mathematics 66, no. 1 (2014): 21-41.↩

53. Knoop, H-B., Pottinger, P., "Ein Satz vom Korovkin-Typ für C^k-Räume", Math. Zeitschrift 148 (1976).↩

54. Kacsó, D.P., "Simultaneous approximation by almost convex operators", 2002.↩

55. Stancu, D.D., Agratini, O., et al. *Analiză Numerică şi Teoria Aproximării*, 2001.↩

56. H. Wang, "**Analysis of error localization of Chebyshev spectral approximations**", arXiv:2106.03456v3 [math.NA], 2023.↩

57. Trefethen, L.N., ***Approximation Theory and Approximation Practice***, 2013.↩

58. Trefethen, L.N., ***Approximation Theory and Approximation Practice***, 2013.↩

59. R. Kannan and C.K. Kreuger, *Advanced Analysis on the Real Line*, 1996.↩

60. Trefethen, L.N., ***Approximation Theory and Approximation Practice***, 2013.↩

61. Rababah, Abedallah. "Transformation of Chebyshev–Bernstein polynomial basis." Computational Methods in Applied Mathematics 3.4 (2003): 608-622.↩