Proceedings of the 3rd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/2025.25318

Probabilistic Perspective of Weierstrass Approximation Theorem via Bernstein Polynomial

Boyi Xie

School of Mathematics, Shanghai University of Finance and Economics, Shanghai, China
2023111178@stu.sufe.edu.cn

Abstract. In view of the importance of Weierstrass approximation theorem, this paper presents a probabilistic approach to establishing the Weierstrass approximation theorem via Bernstein polynomials, emphasizing the interplay between probability theory and mathematical analysis. By interpreting Bernstein polynomials through the lens of Bernoulli trials, the author associates the Bernstein polynomial function with the expectation $E[f(S_n/n)]$, where S_n follows a binomial distribution with parameter x. Leveraging the law of large numbers, the author demonstrates the uniform convergence of the Bernstein polynomial function to the original function on $[0,1]$. The key technical contributions involve bounding the approximation error through variance analysis and employing piecewise estimation to address non-uniform continuity. This approach not only circumvents traditional analytical complexities but also highlights how probabilistic tools, such as particularly moment calculations and distributional convergence, naturally unify discrete approximations with continuous function spaces. The methodology simplifies computational procedures while maintaining rigorous convergence guarantees, offering an instructive example of cross-domain synergy between probability and classical analysis.

Keywords: Bernstein polynomial, Weierstrass approximation theorem, Probability theory

1. Introduction

The synergy between probability theory and mathematical analysis has profoundly enriched both fields, offering unified frameworks for tackling complex problems. A striking example lies in the probabilistic interpretation of Bernstein polynomials, where binomial distributions and expectation operators $E[f(S_n/n)]$ translate function approximation into probabilistic convergence analysis. By leveraging variance bounds $Var(S_n/n) = x(1-x)/n$ and the law of large numbers, discrete probabilistic models converge uniformly to continuous functions, directly proving the Weierstrass theorem. Such methods bypass traditional analytical complexities while highlighting how probability theory—via moment calculations and distributional limits—provides intuitive tools for approximating function spaces like $C(0,1)$. Conversely, mathematical analysis underpins foundational probability concepts, such as measure-theoretic integration and density functions, ensuring rigorous continuity and convergence proofs. This cross-disciplinary interplay not only

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.25318

simplifies computational workflows but also deepens insights into stochastic processes and deterministic analysis alike.

The probability theory has such a close relationship with mathematical analysis that applying probability theory to mathematical analysis has its unique importance. Probability theory is based on measure theory, which is also a core tool for modern mathematical analysis such as Lebesgue integrals and function space theory. The introduction of probability measures is the unmeasurable set-in analysis. Generalized integration and other problems provide more rigorous mathematical descriptions, while expanding the research scope of function spaces. Mathematical analysis mainly studies deterministic phenomena, while probability theory incorporates randomness into mathematical frameworks through tools such as random variables and distribution functions. The combination of the two can handle complex problems such as stochastic differential equations and stochastic integrals, enriching the research objects of analysis. Techniques in probability theory, such as randomization and Monte Carlo methods, are used in theorem proofs for mathematical analysis. For example, $1/\sqrt{2\pi}\int_{-\infty}^{x} e^{-t^2/2}dt \sim 1/\sqrt{2\pi}\left(e^{-x^2/2}\right)/x$ has already been proved [1], but it can also be proved by probability theory.

The development of approximation theory has promoted the development of functional and other branches of mathematics. At the same time, Bernstein polynomials are very important in approximation theory. The functions which can be approximated by Bernstein polynomial have very good properties. For instance, if the Bernstein polynomial belongs to the class of Lipschitz functions, then the function it approximates also belongs to the class of Lipschitz functions [2]. In this article, the author will mainly talk about how to use Bernstein polynomial to prove the Weierstrass approximation theorem.

2. Method and theory

2.1. Weierstrass approximation theorem

Theorem: If f is a continuous complex function on $[a,b]$, there exists a sequence of polynomials P_n such that

$$P_n(x) = f(x) \tag{1}$$

uniformly on $[a,b]$, and P_n can be taken real if f is real. The theorem was firstly found by Weierstrass. The theorem is proved by some techniques in mathematical analysis which are quite difficult [3]. Below is the proof.

Proof: First the author assumes that, $[a,b] = [0,1]$, since $[a,b]$ is a random interval. Else, assume $f(0) = f(1) = 0$. Because if the theorem is proven in this situation, set $g(x) = f(x) - f(0) - x[f(1) - f(0)] \quad (0 \le x \le 1)$. In this case $g(0) = g(1) = 0$, and if g can be derived as the limit of polynomial uniform convergence sequence, it can be easily proved that the same situation is also correct for f, since $f - g$ is also a polynomial.

To continue, $f(x)$ is defined to be 0 as x is not concluded in $[0,1]$. In order let f to be uniformly continuous on the interval. Put

$$Q_n(x) = c_n(1 - x^2)^n (n = 1, 2, 3, \ldots) \tag{2}$$

DOI: 10.54254/2753-8818/2025.25318

where c_n is chosen so that $\int_{-1}^{1} Q_n(x)dx = 1$. More information about the order of magnitude of c_n should be known, as

$$\int_{-1}^{1} \left(1-x^2\right)^n dx = 2\int_{0}^{1} \left(1-x^2\right)^n dx \geq 2\int_{0}^{1/\sqrt{n}} \left(1-x^2\right)^n dx$$

$$\geq 2\int_{0}^{\frac{1}{\sqrt{n}}} (1-nx^2)dx = \frac{4}{3\sqrt{n}} > \frac{1}{\sqrt{n}} \tag{3}$$

It follows from Eq. (3) that $c_n < \sqrt{n}$.

The inequality $\left(1-x^2\right)^n \geq 1-nx^2$ which is used the author above can be proved by taking the consideration of the function $\left(1-x^2\right)^n - 1 + nx^2$. The function equals to 0 when x equals to 0 and its derivative is non-negative on the interval $(0,1)$. For any $\delta > 0$, Eq. (3) indicates

$$Q_n(x) \leq \sqrt{n}(1-\delta^2)^n (\delta \leq abs(x) \leq 1) \tag{4}$$

so that $Q_n \to 0$ uniformly in $\delta \leq |x| \leq 1$. Set

$$P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)dt (0 \leq x \leq 1) \tag{5}$$

The assumptions listed above about f show that

$$P_n(x) = \int_{-x}^{1-x} f(x+t)Q_n(t)dt = \int_{0}^{1} f(t)Q_n(t-x)dt \tag{6}$$

What is more, the last integral is obviously a polynomial included in x. Therefore $\{P_n\}$ is a sequence of polynomials. If f is real, then the polynomials are real.

Given $\varepsilon > 0$, let $\delta > 0$ such that $|y-x| < \delta$ indicates $|f(y) - f(x)| < \frac{\varepsilon}{2}$. Let $M = sup(|f(x)|)$. Using (1), (3), and the fact that $Q_n(x) \geq 0$, for $0 \leq x \leq 1$,

$$|P_n(x) - f(x)| = \left| \int_{-1}^{1} [f(x+t) - f(x)]Q_n(t)dt \right|$$

$$\leq \int_{-1}^{1} abs\left(f(x+t) - f(x)\right)Q_n(t)dt$$

$$\leq 2M\int_{-1}^{-\delta} Q_n(t)dt + \frac{\varepsilon}{2}\int_{-\delta}^{\delta} Q_n(t)dt + 2M\int_{\delta}^{1} Q_n(t)dt$$

$$\leq 4M\sqrt{n}(1-\delta^2)^n + \frac{\varepsilon}{2} < \varepsilon \tag{7}$$

for all n which is large enough. Now the theorem is proved.

DOI: 10.54254/2753-8818/2025.25318

Obviously, this method is very complicated. So, can the polynomial approximation of continuous functions be transformed into a problem of probabilistic expectation, and use probability theory methods to prove this theorem? This can be achieved using Bernstein polynomial and probabilities methods.

2.2. Bernstein polynomial

Definition: For a continuous function $f \in C[0,1]$, the Bernstein polynomial of degree n is defined as $B_n(f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{1-k}$ [4].

This can be reinterpreted as the expectation of a binomial random variable: $B_n(f)(x) = E\left[f\left(\frac{x}{n}\right)\right]$, where X~Binomial(n,x). Assume that S_n~Binomial(n,x), and standardized random variable $\frac{S_n}{n}$, its expectation is $E\left[\frac{S_n}{n}\right] = x$, variance is $Var\left(\frac{S_n}{n}\right) = \frac{x(1-x)}{n}$ .Next, to build the realtion between Bernstein polynomial and the expectation, the polynomial coefficients can be interpreted as probability mass function of the binomial distribution $B_n(f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)P(S_n = k) = E\left[f\left(\frac{S_n}{n}\right)\right]$ .So for any $x \in [0,1]$,

$$|B_n(f)(x) - f(x)| = |E[f(\tfrac{S_n}{n}) - f(x)]| \le E[|f(\tfrac{S_n}{n}) - f(x)|] \tag{8}$$

This step is achieved by Jensen inequality in order to control the absolute value of the expectation. With the continuity of a function, $\forall \varepsilon > 0,\ \exists \delta > 0,$

$$s.t. |\tfrac{S_n}{n} - x| < \delta \to |f(\tfrac{S_n}{n}) - f(x)| < \varepsilon \tag{9}$$

Next, the expectation can be divided into two parts:

$$E[|f(\tfrac{S_n}{n}) - f(x)|] \le \varepsilon \bullet P(|\tfrac{S_n}{n} - x| < \delta) + 2M \bullet P(|\tfrac{S_n}{n} - x| \ge \delta)(M = sup|f|) \tag{10}$$

The second part has an upper bound:

$$P(|\tfrac{S_n}{n} - x| \ge \delta) \le \frac{Var(\frac{S_n}{n})}{\delta^2} = \frac{x(1-x)}{n\delta^2} \le \frac{1}{4n\delta^2} \tag{11}$$

Finally,

$$|B_n(f)(x) - f(x)| \le E[|f(\tfrac{S_n}{n}) - f(x)|] \le \varepsilon + \frac{2M}{4n\delta^2} \tag{12}$$

$\forall \varepsilon > 0,$ take the δ which satisfies the continuity condition, then take $n > \frac{M}{2\varepsilon\delta^2}$,

$$|B_n(f)(x) - f(x)| \le 2\varepsilon \tag{13}$$

Now the theorem is proved by Bernstein polynomial and probabilities methods. The author also wants to make a comparison between traditional mathematic analysis proof and probabilities methods proof.

In traditional analytic proofs, explicit computation of Bernstein polynomial coefficients is required to approximate continuous functions. This involves cumbersome algebraic

DOI: 10.54254/2753-8818/2025.25318

manipulations and often leads to computational errors. What is more, proof relies heavily on intricate combinatorial summations. These steps obscure the intuition behind the Weierstrass approximation theorem, requiring advanced analysis skills to grasp. The complexity of coefficient manipulation and combinatorial arguments makes the proof inaccessible to beginners, as it prioritizes algebraic rigor over conceptual clarity.

In contrast, probabilistic methods circumvent the explicit construction of polynomial coefficients. By leveraging the linearity of expectation and applying Jensen's inequality, the proof becomes more natural. The error estimation is simplified using variance, avoiding advanced techniques. This approach not only enhances clarity for novices but also establishes a profound connection between probability theory and mathematical analysis.

3. Results and applications

3.1. Core methodologies and technical features

The author has already used the Bernstein polynomial to prove Weierstrass approximation theorem. To derive more general and detailed conclusions, it is necessary to summarize the core ideas in the procedure of proof.

The first step is about constructor. Bernstein polynomial approximates functions through discrete sampling points, its standard form is:

$$B_n(f)(x) = \sum_{k=0}^{n} f(\frac{k}{n})(\frac{n}{k})x^k(1-x)^{1-k} \tag{14}$$

where n is the degree of the polynomial, $f\left(\frac{x}{n}\right)$ is the sampling values at equally spaced nodes k/n . Bernstein polynamial can be regarded as the weighted average of function values, with weights determined by the binomial distribution $\binom{n}{k}x^k(1-x)^{1-k}$. Essentially, they represent an expectation approximation in a probabilistic sense.

Then, the second issue is the convergence and error analysis of Bernstein polynomial. According to Weierstrass theorem shown in Eq. (1), when $n \to \infty$, $B_n(f)(x)$ converge uniformly to a continuous function on the interval $(0,1)$. For the convergence speed, the error bound is related to the smoothness of the function, if $f(x)$ satisfies the Lipschitz condition. This condition states that for functions which are defined on a subset of the real number set $f: D \subseteq R \to R$, if there exists a constant K , s.t. $|f(a) - f(b)| \leq K|a-b|$ for $\forall a, b \in D$. The minimum constant K for f is called the Lipschitz constant for f . Then the error bound is $o\left(\frac{1}{\sqrt{n}}\right)$; if $f(x)$ is second-order continuously differentiable, then the error bound comes to $o\left(\frac{1}{n}\right)$. Finally, the third step is method characteristics and optimization.

Before closing, the author shall give a remark the advantages. Bernstein polynomial possesses the properties of preserving non-negativity and monotonicity, and they do not exhibit the Runge phenomenon due to uneven spacing between interpolation nodes. What is more, the values of Bernstein polynomial always lie within the convex hull of the function's sampling points, making them suitable for geometric modeling, such as in the case of Bezier curves.

Furthermore, to obtain deeper conclusions, talking about some limitations are necessary. When the computation of binomial coefficients involves large amounts, n needs to be combined with recursive or numerical optimization methods. Else, in the approximation of functions with non-

DOI: 10.54254/2753-8818/2025.25318

uniform nodes or high oscillations, it is necessary to combine with other basis functions(such as Chebyshev polynomials) to improve efficiency.

In conclusion, Bernstein polynomial achieves function approximation in the form of probability weighting, combining both theoretical completeness and practical engineering applicability. However, balancing its efficiency and accuracy requires adjusting parameters or combining with other approximation methods according to specific problems.

3.2. Some applications of approximation theory

The author has already proved that Weierstrass approximation theorem can be proved by Bernstein polynomial. However, approximation theory also holds a significant position in other fields.

First, stable numerical interpolation, in numerical analysis, high-degree polynomial interpolation often suffers from the Runge phenomenon, where oscillations amplify near interval endpoints. Bernstein polynomials mitigate this issue through their shape-preserving and smoothing properties. By design, they avoid overshooting and maintain non-negativity, ensuring stable approximations even for noisy data. This stability is critical in engineering simulations, where robustness outweighs the need for rapid convergence [5].

Second, geometric modeling: Bézier Curves, Bernstein polynomials form the mathematical backbone of Bézier curves, the cornerstone of computer-aided design (CAD) systems. A Bézier curve of degree n is expressed as: $C(t) = \sum_{i=0}^{n} B_{i,n}(t) P_i \quad t \in [0,1]$, where $B_{i,n}(t)$ are Bernstein basis functions, and P_i are control points. The convex hull property (guaranteed by the non-negativity and partition of unity of Bernstein polynomials) ensures curves remain within the convex hull of control points, enabling intuitive shape manipulation in graphic design and animation [6].

Third, image processing and feature extraction, in image segmentation, Bernstein polynomials are employed to approximate object boundaries. By fitting polynomial curves to pixel data, they enable efficient representation of complex shapes with minimal control points. This technique is particularly valuable in medical imaging for reconstructing organ contours or tumor margins from discrete voxel data. Additionally, their smoothness reduces aliasing artifacts in image resizing algorithms [7].

Forth, emerging applications in machine learning, recent advances explore Bernstein polynomials as activation functions or basis functions in neural networks. Their bounded derivatives and smoothness enhance training stability for regression tasks, especially when modeling physical systems governed by continuous but non-analytic laws.

All in all, Bernstein polynomials bridge theoretical mathematics and practical engineering [8], offering a robust framework for approximating continuous functions. From foundational proofs in analysis to real-world applications in CAD and image processing, their unique blend of simplicity, stability, and geometric interpretability ensures enduring relevance. As computational demands grow, these polynomials will likely inspire new algorithms at the intersection of approximation theory and data-driven modeling [9]. Their legacy exemplifies how abstract mathematical constructs can evolve into indispensable tools for technological innovation.

4. Conclusion

This paper presents a novel probabilistic approach to establishing the Weierstrass approximation theorem via Bernstein polynomials, highlighting the synergy between probability theory and mathematical analysis. By interpreting Bernstein polynomials through the lens of Bernoulli trials

DOI: 10.54254/2753-8818/2025.25318

and leveraging the law of large numbers, the author demonstrates the uniform convergence of polynomials to continuous functions. The proposed methodology simplifies computational procedures while maintaining rigorous convergence guarantees, offering an instructive example of cross-domain synergy. The technical contributions involve bounding the approximation error through variance analysis and employing piecewise estimation to address non-uniform continuity. The results reveal that Bernstein polynomials achieve function approximation in the form of probabilistic weighting, combining theoretical completeness with practical engineering applicability. However, balancing efficiency and accuracy requires adjusting parameters or combining with other approximation methods according to specific problems. Overall, this work deepens insights into stochastic processes and deterministic analysis, enriching the research scope of function spaces.

References

[1] Mao Weibing. (2017). Simple Approximate Relationships Between the Normal Distribution Function and Its Density. Yinshan Academic Journal, 31(04), 8-9.
[2] Guo Cundi. (2000). Bernstein polynomials and the continuous functions they approximate. Journal of Xi'an University of Engineering and Technology, (03), 325-326.
[3] Walter Rudin. (2013). Principles of Mathematical Analysis. McGraw-Hill Education.
[4] A.N. Shiryaev. (2004). Probability. World Publishing Corporation.
[5] Kaur, J., et al. (2024). A generalization of modified α-Bernstein operators and its related estimations and errors. Arabian Journal of Mathematics, 13(3), 521–531.
[6] Zhang Jiuting. (2018). A class of inequalities based on Bernstein polynomials and their applications. Journal of Inner Mongolia Normal University, 47 (03), 199-202.
[7] Wei Yanjun & Feng Bojin & Wu Weiguo. (2016). Multi threshold image segmentation algorithm based on polynomial consistent approximation. Journal of Communications, 37 (10), 56-64.
[8] Liu Yong & Wang Changqin. (2005). The approximation degree of Bernstein polynomial on the convergence interval. Journal of Dalian Railway Institute, (04), 1-3.
[9] Bustamante, J., & Muñoz-Delgado, F. J. (2014). Bernstein polynomial and discontinuous functions. Journal of Mathematical Analysis and Applications, 411(2), 829–837.