Elemente der Mathematik (Online first)
DOI 10.4171/EM/548

Published by EMS Press

Bernstein approximation and beyond: Proofs by means of elementary probability theory

Tiangang Cui and Friedrich Pillichshammer

Tiangang Cui received his doctorate in engineering science from the University of Auckland in 2010. Since 2023, he has been a senior lecturer at the University of Sydney in Australia. Previously, he held lecturer and senior lecturer positions at Monash University in Australia. His research interests include stochastic computation, dimension reduction, and inverse problems.

Friedrich Pillichshammer received his doctorate in mathematics from the University of Salzburg in 1999. Since the year 2003, he has been an associated professor at the Johannes Kepler University in Linz, Austria, where he is currently heading the Institute for Financial Mathematics and Applied Number Theory. His main research interests are in discrepancy theory, quasi-Monte Carlo methods, and information based complexity.

1 Bernstein approximation

For $n \in \mathbb{N}$ (the set of positive integers) and $k \in \{0, 1, \ldots, n\}$, the k-th Bernstein basis polynomial of degree n is defined as

$$\beta_n(k,x) := \binom{n}{k} x^k (1-x)^{n-k} \quad \text{for } x \in [0,1].$$

Bernstein-Polynome liefern einen konstruktiven Beweis für den aus der Analysis bekannten Approximationssatz von Karl Weierstraß. Dieser besagt, dass jede stetige Funktion auf einem beschränkten, abgeschlossenen Intervall beliebig genau und gleichmäßig durch Polynome angenähert werden kann. Interessanterweise kann der Beweis dieses Resultats auch mit Hilfe elementarer Wahrscheinlichkeitstheorie durchgeführt werden. Auf diese Weise erhält man darüber hinaus sogar Abschätzungen des Approximationsfehlers für Lipschitz-Funktionen. In diesem Beitrag wird diese Beweismethode vorgestellt und aufgezeigt, wie sie auf natürliche Weise auf andere interessante Situationen übertragen werden kann, bei denen der Fokus mehr auf Funktionen mit unbeschränktem Definitionsbereich liegt. Damit kann man auf elementarem Weg interessante Ergebnisse für den Szász-Mirakjan-Operator und den Baskakov-Operator erzielen.

For a continuous function $f\colon[0,1]\to\mathbb{R}$ and $n\in\mathbb{N}$, Bernstein [2] constructed an approximation scheme of the form

$$B_n(f;x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\beta_n(k,x)$$

to prove the Weierstrass approximation theorem. His proof is based on methods from elementary probability theory (see also [4, Proposition 5.2]). Kac [3] gave a formula for the approximation error for Lipschitz continuous functions. These results are also discussed by Mathé in [5], which is worth reading also because of the interesting historical comments.

Theorem 1. *Assume a function $f\colon[0,1]\to\mathbb{R}$ is α-Hölder continuous, i.e., $f\in C^{0,\alpha}([0,1])$ with $\alpha\in(0,1]$, such that*

$$|f(x)-f(y)|\le L|x-y|^{\alpha} \quad \text{for all } x,y\in[0,1],$$

for some real constant $L>0$. Then, for all $n\in\mathbb{N}$ and all $x\in[0,1]$, we have

$$|f(x)-B_n(f;x)|\le L\left(\frac{x(1-x)}{n}\right)^{\alpha/2}.$$

We briefly revisit the proof of this result according to Mathé [5] which uses elementary arguments from probability theory. Thereby, we make some minor reformulations that allow us to get a clearer picture of the overall situation, which in turn makes it clear how the method can be expanded.

Proof of Theorem 1. Let $x\in[0,1]$ be the success probability of a Bernoulli experiment and let $n\in\mathbb{N}$ be the total number of experiments. The number of successful Bernoulli experiments, which is a discrete random variable denoted by $K_{n,x}$, follows the binomial distribution with the probability mass function

$$\beta_n(k,x)=\binom{n}{k}x^k(1-x)^{n-k} \quad \text{for } k\in\{0,1,\dots,n\}.$$

The probability of $K_{n,x}=k$ is given by the k-th Bernstein basis polynomial. We have the expectation $\mathbb{E}[K_{n,x}]=nx$ and the variance $\mathrm{Var}[K_{n,x}]=nx(1-x)$.

For a function $f\colon[0,1]\to\mathbb{R}$, we have the identity

$$\mathbb{E}\left[f\left(\frac{K_{n,x}}{n}\right)\right]=\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}=B_n(f;x).$$

Therefore, by treating $f(x)$ as a deterministic term in the expectation, in particular,

$$\mathbb{E}[f(x)]=f(x) \quad \text{for all } x\in[0,1],$$

we obtain

$$|f(x)-B_n(f;x)|=\left|\mathbb{E}\left[f(x)-f\left(\frac{K_{n,x}}{n}\right)\right]\right|\le\mathbb{E}\left[\left|f(x)-f\left(\frac{K_{n,x}}{n}\right)\right|\right].$$

Assuming that $f \in C^{0,\alpha}([0,1])$ and using the above inequality, we then have

$$|f(x) - B_n(f;x)| \le L\mathbb{E}\left[\left|x - \frac{K_{n,x}}{n}\right|^{\alpha}\right] = \frac{L}{n^{\alpha}}\mathbb{E}[|nx - K_{n,x}|^{\alpha}]. \tag{1}$$

Applying Jensen's inequality, we obtain

$$\mathbb{E}[|nx - K_{n,x}|^{\alpha}] \le (\mathbb{E}[|nx - K_{n,x}|^2])^{\alpha/2} = \mathrm{Var}[K_{n,x}]^{\alpha/2} = \big(nx(1-x)\big)^{\alpha/2}. \tag{2}$$

Substituting (2) into (1), the result follows. ■

The key insight we gain here is that one can exploit the equivalence between discrete probability distributions and certain basis functions to construct approximation schemes. This note aims to demonstrate how this method can be easily adapted to other discrete probability distributions and thereby to other approximation schemes rather than that based on Bernstein basis polynomials. This very elementary approach leads in an easy way to results for the Szász–Mirakjan operator and the Baskakov operator, which are usually obtained by analytic methods (suitable references will be given below).

2 Other approximation schemes

We aim to approximate a continuous and bounded function $f\colon [0,\infty) \to \mathbb{R}$ over the unbounded domain $[0,\infty)$ using a family of discrete random variables $K_{n,x}$ with parameters $n \in \mathbb{N}$ and $x \in [0,\infty)$ taking values k in $\mathbb{N}_0$ (the set of non-negative integers). Let the probability mass function of $K_{n,x}$ be $p_n(k,x)$. Assuming $\mathbb{E}[K_{n,x}] = nx$ and $\mathrm{Var}[K_{n,x}] < \infty$, we define the approximation scheme

$$S_n(f;x) := \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) p_n(k,x). \tag{3}$$

Note that $S_n(f,x)$ is well-defined for bounded functions f because $p_n(\cdot\,,x)$ is a probability mass function.

The following theorem estimates the approximation error of $S_n(f;x)$.

Theorem 2. *Let $K_{n,x}$ with parameters $n \in \mathbb{N}$ and $x \in [0,\infty)$ be $\mathbb{N}_0$-valued random variables with probability mass function $p_n(k,x)$ for $k \in \mathbb{N}_0$ and with $\mathbb{E}[K_{n,x}] = nx$ and* $\mathrm{Var}[K_{n,x}] < \infty$. *Let $S_n(\cdot\,,x)$ be the associated approximation scheme given in (3). Assume that a bounded function $f\colon [0,\infty) \to \mathbb{R}$ satisfies a general Hölder-type condition,*

$$|f(x) - f(y)| \le L\frac{|x-y|^{\alpha}}{(\gamma + x + y)^{\beta}} \quad \textit{for all } x, y \in [0,\infty) \textit{ with } x \neq y, \tag{4}$$

for real constants $L > 0$, $\alpha \in (0,1]$ and $\beta, \gamma \in [0,\infty)$; then, for all $n \in \mathbb{N}$ and all $x \in (0,\infty)$, we have

$$|f(x) - S_n(f;x)| \le \frac{L}{n^{\alpha}}\frac{\mathrm{Var}[K_{n,x}]^{\alpha/2}}{(\gamma + x)^{\beta}}.$$

For a function f satisfying the general Hölder-type condition (4), we denote it by $f \in C^{0,\alpha,\beta,\gamma}([0,\infty))$. The constant γ is used to avoid singularities at the origin in error estimates. With increasing β, we impose a stronger decaying rate on f toward its tail. With $\beta = 0$, the general Hölder-type condition reduces to the plain Hölder condition

$$|f(x) - f(y)| \leq L|x-y|^{\alpha} \quad \text{for all } x, y \in [0,\infty),$$

for $L > 0$ and $\alpha \in (0,1]$. In this case, the error estimate in Theorem 2 becomes

$$|f(x) - S_n(f;x)| \leq \frac{L}{n^{\alpha}} \operatorname{Var}[K_{n,x}]^{\alpha/2}.$$

Proof of Theorem 2. Let $f\colon [0,\infty) \to \mathbb{R}$ be a continuous and bounded function and let $K_{n,x} \in \mathbb{N}_0$ be a discrete random variable with probability mass function $p_n(k,x)$, where $n \in \mathbb{N}$ and $x \in (0,\infty)$. We have

$$\mathbb{E}\left[f\left(\frac{K_{n,x}}{n}\right)\right] = \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) p_n(k,x) = S_n(f;x),$$

and thus

$$|f(x) - S_n(f;x)| = \left|\mathbb{E}\left[f(x) - f\left(\frac{K_{n,x}}{n}\right)\right]\right| \leq \mathbb{E}\left[\left|f(x) - f\left(\frac{K_{n,x}}{n}\right)\right|\right].$$

Assuming $f \in C^{0,\alpha,\beta,\gamma}([0,\infty))$ and using the above inequality, we have

$$|f(x) - S_n(f;x)| \leq L\mathbb{E}\left[\frac{|x - \frac{K_{n,x}}{n}|^{\alpha}}{(\gamma + x + \frac{K_{n,x}}{n})^{\beta}}\right] \leq \frac{L}{n^{\alpha}} \frac{\mathbb{E}[|nx - K_{n,x}|^{\alpha}]}{(\gamma + x)^{\beta}}.$$

Following the argument in (2), we have $\mathbb{E}[|nx - K_{n,x}|^{\alpha}] \leq \operatorname{Var}[K_{n,x}]^{\alpha/2}$, and therefore the result follows. ■

As an application, we obtain results for the Szász–Mirakjan operator and the Baskakov operator.

Szász–Mirakjan operator

The Szász–Mirakjan operator [6] is based on the Poisson distribution P_{λ} with parameter $\lambda > 0$, which has the probability mass function

$$p(k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k \in \mathbb{N}_0.$$

If K_{λ} is distributed according to P_{λ}, which is written as $K_{\lambda} \sim P_{\lambda}$, then it is well known that $\mathbb{E}[K_{\lambda}] = \operatorname{Var}[K_{\lambda}] = \lambda$. Now, for $n \in \mathbb{N}$ and $x \in [0,\infty)$, we define the parameter $\lambda = nx$ and let $K_{n,x} \sim P_{nx}$. This defines the basis functions

$$p_n(k,x) = \mathrm{e}^{-nx} \frac{(nx)^k}{k!} \quad \text{for } k \in \mathbb{N}_0.$$

Then, for $f \in C^{0,\alpha,\beta,\gamma}([0,\infty))$ and $n \in \mathbb{N}$, we have the approximation scheme

$$S_n(f;x) := \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) \mathrm{e}^{-nx} \frac{(nx)^k}{k!},$$

which is known as Szász–Mirakjan operator; see, e.g., [6, 7] for a detailed study using analytic methods. From Theorem 2, we obtain its error estimate

$$|f(x) - S_n(f;x)| \le \frac{L}{n^{\alpha/2}} \frac{x^{\alpha/2}}{(\gamma + x)^{\beta}} \le \frac{L}{n^{\alpha/2}} (\gamma + x)^{\alpha/2-\beta}.$$

If $\beta = \alpha/2$, then the bound is uniform in x, and if $\beta > \alpha/2$, the bound even decays toward the tail as x increases. The convergence rate $n^{-\alpha/2}$ is also obtained in the above mentioned papers [6, 7]. Furthermore, the convergence rate is best possible for $\alpha = 1$ as shown in [6, p. 241].

Baskakov operator

The Baskakov operator [1] is based on the Pascal distribution $\mathrm{PC}_{n,x}$ (also known as negative binomial distribution) with parameters $n \in \mathbb{N}$ and $x \in [0,\infty)$, which has the probability mass function

$$p_n(k,x) = \binom{n+k-1}{k} \frac{x^k}{(1+x)^{n+k}} \quad \text{for } k \in \mathbb{N}_0.$$

If $K_{n,x} \sim \mathrm{PC}_{n,x}$, then we have $\mathbb{E}[K_{n,x}] = nx$ and $\mathrm{Var}[K_{n,x}] = nx(1+x)$. Then, for bounded $f \in C^{0,\alpha,\beta,\gamma}([0,\infty))$ and $n \in \mathbb{N}$, we have the approximation scheme

$$S_n(f;x) = \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) \binom{n+k-1}{k} \frac{x^k}{(1+x)^{n+k}}.$$

This operator is known as Baskakov operator in literature; see, e.g., [1,7]. From Theorem 2, we obtain its error estimate

$$|f(x) - S_n(f;x)| \le \frac{L}{n^{\alpha/2}} \frac{(x(x+1))^{\alpha/2}}{(\gamma + x)^{\beta}}.$$

With a decay rate $\beta \geq \alpha$, we can also extend the convergence rate of the Baskakov operator uniformly for all $x \in [0,\infty)$. For example, with $\gamma = 1$ and $\beta = \alpha$, for all $n \in \mathbb{N}$ and uniformly for all $x \in (0,\infty)$, we have

$$|f(x) - S_n(f;x)| \le \frac{L}{n^{\alpha/2}}.$$

The convergence rate $n^{-\alpha/2}$ can be also deduced from [7, Theorem 3.2].

Final remark and examples

For both Szász–Mirakjan and Baskakov operators, it is possible to bound the error $|f(x) - S_n(f;x)|$ uniformly for all $x \in [0,\infty)$ for bounded functions with a sufficiently large decay rate β. However, their convergence still follows a rate of $n^{-\alpha/2}$, and thus we can recover a convergence rate of $n^{-1/2}$ with $\alpha = 1$ as the best-case scenario.

To numerically implement these operators, we also require a truncation in k in the approximation. For $m \in \mathbb{N}$, we have the truncated approximation scheme

$$\tilde{S}_{n,m}(f;x) := \sum_{k=0}^{m} f\left(\frac{k}{n}\right) p_n(k,x). \tag{5}$$

The error of the approximation (5) has an apparent upper bound

$$|f(x) - \tilde{S}_{n,m}(f;x)| \leq |f(x) - S_n(f;x)| + |S_n(f;x) - \tilde{S}_{n,m}(f;x)|,$$

in which the first term of the bound is given by Theorem 2 and the second term of the bound (the *truncation error*) satisfies

$$\begin{aligned} |S_n(f;x) - \tilde{S}_{n,m}(f;x)| &= \left|\sum_{k>m} f\left(\frac{k}{n}\right) p_n(k,x)\right| \\ &\leq \sup_{x\in(m/n,\infty)} |f(x)| \sum_{k>m} p_n(k,x) \\ &= \sup_{x\in(m/n,\infty)} |f(x)| \mathbb{P}[K_{n,x} > m]. \end{aligned} \tag{6}$$

We first consider a function $f \in C^{0,\alpha,\beta,\gamma}([0,\infty))$ with a sufficiently large β that leads to a uniform error bound $|f(x) - S_n(f;x)| \leq Ln^{-\alpha/2}$ in the untruncated approximation. For $x \in (m/n,\infty)$, we further assume that f satisfies a tail condition $|f(x)| = O(g(x))$ for $x \to \infty$ for some strictly decreasing function $g\colon (0,\infty) \to \mathbb{R}^+$. Applying the bound in (6) and $\mathbb{P}[K_{n,x} > m] \leq 1$, the truncation error satisfies

$$|S_n(f;x) - \tilde{S}_{n,m}(f;x)| = O(g(m/n)).$$

Then, using $m = \lceil n g^{-1}(n^{-\alpha/2})\rceil$, the error of the truncated approximation scheme satisfies $|f(x) - \tilde{S}_{n,m}(f;x)| = O(n^{-\alpha/2})$. The following are some examples.

(1) The function $f(x) = (1+x^2)^{-1}$ satisfies the general Hölder condition (4) with $\alpha = 1$ and $\beta = 1$. We also have $f(x) = O(x^{-2})$ for $x \to \infty$. Thus, we can choose $m = \lceil n^{5/4}\rceil$ so that the truncated approximation satisfies $|f(x) - \tilde{S}_{n,m}(f;x)| = O(n^{-1/2})$.

(2) The function $f(x) = \exp(-x)$ satisfies the general Hölder condition (4) with $\alpha = 1$ and $\beta = 1$. Thus, we can choose $m = \lceil n\log(n)/2\rceil$ so that the truncated approximation satisfies $|f(x) - \tilde{S}_{n,m}(f;x)| = O(n^{-1/2})$.

If the function $f\colon [0,\infty) \to \mathbb{R}$ does not necessarily satisfy a tail condition to guide the truncation in k, but is at least bounded, then we can use an alternative argument from elementary probability theory. Assuming $\|f\|_\infty := \sup_{x\in[0,\infty)}|f(x)| < \infty$, the bound of the truncation error in (6) also leads to

$$|S_n(f;x) - \tilde{S}_{n,m}(f;x)| \le \|f\|_\infty \, \mathbb{P}[K_{n,x} > m],$$

in which $\mathbb{P}[K_{n,x} > m]$ can be estimated using Chebyshev's inequality. Choosing $m \ge 2\mathbb{E}[K_{n,x}]$, we have

$$\mathbb{P}[K_{n,x} > m] \le \mathbb{P}\big[|K_{n,x} - \mathbb{E}[K_{n,x}]| \ge \mathbb{E}[K_{n,x}]\big] \le \frac{\mathrm{Var}[K_{n,x}]}{\mathbb{E}[K_{n,x}]^2}.$$

In the case of the Szász–Mirakjan operator, we have $\mathbb{E}[K_{n,x}] = \mathrm{Var}[K_{n,x}] = nx$, and thus, choosing $m = 2\lceil nx \rceil$, we obtain

$$|S_n(f;x) - \tilde{S}_{n,m}(f;x)| \le \|f\|_\infty \frac{1}{nx}.$$

In the case of the Baskakov operator, we have $\mathbb{E}[K_{n,x}] = nx$ and $\mathrm{Var}[K_{n,x}] = nx(1+x)$. Hence, choosing again $m = 2\lceil nx \rceil$, we obtain

$$|S_n(f;x) - \tilde{S}_{n,m}(f;x)| \le \|f\|_\infty \frac{1}{n}\frac{1+x}{x}.$$

References

[1] V. A. Baskakov, An instance of a sequence of linear positive operators in the space of continuous functions. *Dokl. Akad. Nauk SSSR (N.S.)* **113** (1957), 249–251 Zbl 0080.05201 MR 94640

[2] S. Bernstein, Démonstration du théorème de Weierstraß, fondée sur le calcul des probabilités. *Charkow Ges. (2)* **13** (1912), 1–2 Zbl 43.0301.03

[3] M. Kac, Une remarque sur les polynômes de M. S. Bernstein. *Stud. Math.* **7** (1938), 49–51 Zbl 0018.20704

[4] E. Lesigne, *Heads or tails*. Stud. Math. Libr. 28, American Mathematical Society, Providence, RI, 2005 Zbl 1085.60001 MR 2139534

[5] P. Mathé, Approximation of Hölder continuous functions by Bernstein polynomials. *Amer. Math. Monthly* **106** (1999), no. 6, 568–574 Zbl 0983.41007 MR 1699224

[6] O. Szasz, Generalization of S. Bernstein's polynomials to the infinite interval. *J. Research Nat. Bur. Standards* **45** (1950), 239–245 Zbl 1467.41005 MR 45863

[7] V. Totik, Approximation by Bernstein polynomials. *Amer. J. Math.* **116** (1994), no. 4, 995–1018 Zbl 0812.41018 MR 1287945

Tiangang Cui
School of Mathematics and Statistics
The University of Sydney
Sydney, NSW 2006, Australia
tiangang.cui@sydney.edu.au

Friedrich Pillichshammer
Institut für Finanzmathematik und Angewandte Zahlentheorie
Johannes Kepler Universität Linz
Altenbergerstr. 69
4040 Linz, Austria
friedrich.pillichshammer@jku.at