the possibility of a "jump hop", associated with impulsive forces, approximated by delta functions in "stick-slip" dynamical problems such as chattering chalk on a blackboard. This phenomenon appears to be the most likely candidate for the origin of the real hop of real hoops, but raises more complicated questions about the physics, and makes the analysis commensurately more difficult.

This analysis of the hopping hoop leads to several conclusions. First, and most important, is that Newton's laws and the kinematical constraint for "rough" contact are in general inconsistent when the normal force is zero. Second, real hoops that hop must skid first, and the subsequent hop cannot be smooth nor semi-smooth. Third, there is a rich structure in the behavior of real hoops: vary λ and I, vary the initial conditions, let ϑ be unbounded, follow the bounce(s) after the hop. Finally, with respect to this isolated singularity in Mr. Littlewood's *Miscellany*, he did say that in practice, "the hoop skids", but seemed to imply this to be due to a realistic friction law rather than a necessary consequence even with an unbounded coefficient of static friction. The answer to his query whether the behavior of the hoop is intuitive is given by the following

Theorem. *The behavior of hopping hoops is not intuitive.*

Proof: By inspection.

REFERENCES

1. J. E. Littlewood, *Littlewood's Miscellany*, Cambridge University Press., Cambridge, 1986.
2. T. F. Tokieda, The Hopping Hoop, *Amer. Math. Monthly* **104** (1997) 152–154.

Physiology Program, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115
jbutler@hsph.harvard.edu

Approximation of Hölder Continuous Functions by Bernstein Polynomials

Peter Mathé

In a recent MONTHLY **[5]**, a special instance of the Weierstraß approximation theorem attracted attention: approximation of real Lipschitz functions on [0, 1] by Bernstein polynomials

$$B_n(f,x) := \sum_{j=0}^{n} \binom{n}{j} x^j (1-x)^{n-j} f\left(\frac{j}{n}\right).$$

The authors of **[5]** provided a rate of uniform convergence of $B_n(f,\cdot)$ to f using large deviations techniques. It is the aim of this note to discuss the optimal rate of approximation with some historical remarks. More generally we consider the class $\mathrm{Lip}_\alpha(L)$ of Hölder continuous functions with exponent α for some $0 < \alpha \le 1$ and constant L, i.e., functions that obey

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

We present two simple proofs of the following

Theorem 1. *If the function $f:[0,1]\to\mathbb{R}$ is Hölder continuous with exponent α and constant L then*

$$|f(x)-B_n(f,x)|\le L\left(\frac{x(1-x)}{n}\right)^{\alpha/2} \quad \textit{for all } n\in\mathbb{N} \textit{ and all } x\in[0,1]. \tag{1}$$

In contrast to Gzyl and Palacios [**5**], who refine the original Bernstein argument to treat the specialization to Lipschitz functions, we use more direct arguments that are elementary and improve the rate to the optimal one.

The history of this result is worth recounting. Bernstein first used his eponymous polynomials to prove the Weierstraß approximation theorem in 1912 [**2**]. It took more than twenty years before results concerning the rate of convergence of $B_n(f,\cdot)$ to f appeared, by Popoviciu [**9**] and by Kac ([**6**] and [**7**]). While Popoviciu established speed of convergence in terms of the modulus of continuity, Kac originally proved exactly Theorem 1. The first proof we present is in the spirit of Kac and exploits the probabilistic meaning of the Bernstein polynomials. The second version uses Korovkin–type arguments [**1**], as in the nice exposition of the Weierstraß approximation theorem in [**10**, Chapter 1.2]. Other approaches are reviewed in Lorentz's monograph [**8**].

The probabilistic nature of the Bernstein polynomials can be recognized when interpreting the weights $\binom{n}{j}x^j(1-x)^{n-j}$ as point probabilities of a binomial distribution with parameters n and x. There are, however, several ways to realize this distribution. To establish a connection with empirical distribution functions, given n, let $u_1,\dots,u_n$ be random variables that are independent and uniformly distributed on $[0,1]$. Consider the random function $S_n:[0,1]\to[0,1]$ defined by

$$S_n(x):=\frac{1}{n}\sum_{j=1}^{n}\chi_{[0,x)}(u_j),\quad x\in[0,1],$$

where $\chi_{[0,x)}$ denotes the characteristic function of the interval $[0,x)$. S_n takes only the values j/n, $j=0,\dots,n$, with probabilities $P(S_n(x)=j/n)=\binom{n}{j}x^j(1-x)^{n-j}$. Thus at any point x the random function value $nS_n(x)$ is binomially distributed and we have expectation $\mathbf{E}S_n(x)=x$ and variance

$$\mathbf{E}(S_n(x)-x)^2=\frac{x(1-x)}{n}. \tag{2}$$

Proof of Theorem 1: probabilistic version. By construction $B_n(f,x)=\mathbf{E}f(S_n(x))$. Using the triangle inequality and Hölder continuity we obtain

$$|f(x)-B_n(f,x)|\le\mathbf{E}|f(x)-f(S_n(x))|\le L\mathbf{E}|x-S_n(x)|^{\alpha}.$$

Apply the Hölder inequality with parameters $2/\alpha$ and $2/(2-\alpha)$, and arrive at

$$|f(x)-B_n(f,x)|\le L\left(\mathbf{E}|x-S_n(x)|^2\right)^{\alpha/2}.$$

Finally, use the variance formula given in (2) to obtain (1). ■

Proof of Theorem 1: analytic version Here we consider the mapping

$$f\in C([0,1])\to B_n(f,\cdot)\in C([0,1]),$$

which is a bounded linear operator that maps nonnegative functions into nonnegative polynomials (and thus is *positive*). Moreover, letting $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2$ denote the monomials of degree at most 2, it is easy to check that

$$B_n(p_0, x) = 1, \quad B_n(p_1, x) = x, \quad B_n(p_2, x) = x^2 + x(1-x)/n. \tag{2.5}$$

Denote $f_x(y) := L|x-y|^\alpha$. Thus for any function $f \in \mathrm{Lip}_\alpha(L)$ we have $-f_x(y) \le f(x) - f(y) \le f_x(y)$ for all $y \in [0,1]$. Applying B_n and using its positivity we obtain $|f(x) - B_n(f, y)| \le B_n(f_x, y)$. Hence, for $y = x$ we arrive at $|f(x) - B_n(f, x)| \le B_n(f_x, x)$.

As a substitute for the Hölder inequality the operator B_n obeys

$$B_n(|g|, x) \le \left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2} \tag{3}$$

for any function g. Indeed, the geometric-arithmetic-mean inequality

$$\sigma^{1/p}\tau^{1-1/p} \le \frac{1}{p}\sigma + (1 - 1/p)\tau, \quad p \ge 1, \sigma, \tau \ge 0,$$

with

$$p := \frac{2}{\alpha}, \ \tau := \left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2}, \quad \text{and } \sigma := |g(y)|^{2/\alpha}\tau^{1-2/\alpha} \text{ for } y \in [0,1]$$

yields

$$|g(y)| \le \frac{1}{p}|g(y)|^{2/\alpha}\left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2-1} + \frac{p-1}{p}\left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2}.$$

Using the positivity of B_n we finally arrive at

$$\begin{aligned} B_n(|g|, x) &\le 1/p\left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2} + (1 - 1/p)\left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2} \\ &= \left(B_n\left(|g|^{2/\alpha}, x\right)\right)^{\alpha/2}, \end{aligned}$$

which proves (3). Thus, for $g := f_x$ we obtain

$$|f(x) - B_n(f, x)| \le \left(B_n\left(f_x^{2/\alpha}, x\right)\right)^{\alpha/2}. \tag{3.5}$$

Using the linearity of B_n and its values at p_0, p_1, and p_2 given in (2.5), and observing that $f_x^{2/\alpha}(y) = L^{2/\alpha}(x^2 - 2xy + y^2)$, we conclude that $B_n(f_x^{2/\alpha}, x) = L^{2/\alpha}x(1-x)/n$. Substituting this equality into (3.5) completes the proof. ∎

For completeness we derive the asymptotically exact behavior of the error. The accuracy of the approximation of Hölder continuous functions by Bernstein polynomials is quantified by

$$e_n(L, \alpha) := \sup_{f \in \mathrm{Lip}_\alpha(L)} \sup_{x \in [0,1]} |f(x) - B_n(f, x)|.$$

Theorem 1 can thus be rewritten as $e_n(L, \alpha) \le L(1/(4n))^{\alpha/2}$. This is refined in

Theorem 2. $\lim_{n\to\infty} n^{\alpha/2} e_n(L, \alpha) = L2^{-\alpha/2}\Gamma((\alpha+1)/2)/\sqrt{\pi}$.

Kac proved that the rate of convergence cannot be improved on the class of Hölder continuous functions by establishing an appropriate lower bound for the specific function $f_{1/2} \in \mathrm{Lip}_\alpha(L)$. For this particular choice we have

$$e_n(L, \alpha) \ge \left|f_{1/2}(1/2) - B_n\left(f_{1/2}, 1/2\right)\right| = L\mathbf{E}\left|S_n(1/2) - 1/2\right|^\alpha. \tag{4}$$

Popoviciu derived the exact asymptotics of the mean absolute deviation $\mathbf{E}|S_n(x) - x|$, i.e., the case $\alpha = 1$, by providing the explicit representation

$$\mathbf{E}|S_n(x) - x| = 2\binom{n-1}{j} x^{j+1}(1-x)^{n-j}, \tag{5}$$

where j is the unique integer for which $x \in [j/n, (j+1)/n)$. This representation has a long history; see **[4]** for a proof and historical details, beginning in 1730, when De Moivre gave a similar explicit formula for the case $x = 1/2$. The derivation of (5) is combinatorial and the analysis of the asymptotic behavior is non-elementary. Our arguments do not rely on the representation (5). Instead we again try to keep close to probabilistic arguments and use the De Moivre–Laplace Theorem on the binomial approximation of the normal distribution.

Our proof of Theorem 2 requires basic knowledge of probability theory and real analysis. For convenience we split the proof of Theorem 2 into two lemmas. As the lower bound in (4) suggests we are concerned with the asymptotic behavior of $n^{\alpha/2}\mathbf{E}|S_n(1/2) - 1/2|^{\alpha}$, and more generally we study the functions $g_n(x) := n^{\alpha/2}\mathbf{E}|S_n(x) - x|^{\alpha}$. Given α we define $K_\alpha := 2^{\alpha/2}\Gamma((\alpha+1)/2)/\sqrt{\pi}$ and $g(x) := (x(1-x))^{\alpha/2}K_\alpha$.

Lemma 1. $\lim_{n\to\infty} g_n(x) = g(x)$ *for every* $x \in [0,1]$.

Moreover the functions g_n belong to $\mathrm{Lip}_{\alpha/2}(1)$ as can be seen from

Lemma 2. $|g_n(x) - g_n(y)| \le |x-y|^{\alpha/2}$ *for all* $n \in \mathbb{N}$ *and all* $x, y \in [0,1]$.

Suppose for a moment that both lemmas have been proved. Pointwise convergence from Lemma 1 and uniform Hölder continuity from Lemma 2 imply uniform convergence as stated in

Proposition 1. $\lim_{n\to\infty}\sup_{x\in[0,1]}|g_n(x) - g(x)| = 0$.

Proof: Our arguments follow the usual proof of the Arzelà–Ascoli Theorem. In fact, for any $m \in \mathbb{N}$ the set $\{j/m, j = 0, \dots, m\}$ is finite so that Lemma 1 provides some n_0 such that

$$|g_n(j/m) - g(j/m)| \le \left(\frac{1}{2m}\right)^{\alpha/2}, \quad j = 0, \dots, m, n \ge n_0.$$

Thus for an arbitrary $x \in [0,1]$ let j_x be chosen such that $|x - j_x/m| \le 1/(2m)$. Then we conclude, using Lemma 2,

$$\begin{aligned}|g_n(x) - g(x)| &\le |g_n(x) - g_n(j_x/m)| + |g_n(j_x/m) - g(j_x/m)| \\ &\quad + |g(j_x/m) - g(x)| \\ &\le 3\left(\frac{1}{2m}\right)^{\alpha/2}\end{aligned}$$

for $n \ge n_0$ and all $x \in [0,1]$. Since m may be chosen arbitrarily large we have established uniform convergence. ∎

Now the proof of Theorem 2 follows easily.

Proof of Theorem 2. We infer from (4) that $n^{\alpha/2}e_n(L, \alpha) \geq Lg_n(1/2)$. Thus Lemma 1 yields $\liminf_{n\to\infty} n^{\alpha/2}e_n(L, \alpha) \geq Lg(1/2)$. On the other hand, interchanging $\sup_{f \in \mathrm{Lip}_\alpha(L)}$ and $\sup_{x\in[0,1]}$ in the definition of $e_n(L, \alpha)$ gives

$$n^{\alpha/2}e_n(L, \alpha) = n^{\alpha/2} \sup_{x\in[0,1]} \sup_{f\in\mathrm{Lip}_\alpha(L)} |f(x) - B_n(f, x)|$$

$$\leq Ln^{\alpha/2} \sup_{x\in[0,1]} \mathbf{E}|x - S_n(x)|^\alpha = L \sup_{x\in[0,1]} g_n(x).$$

Since

$$\left| \sup_{x\in[0,1]} g_n(x) - \sup_{x\in[0,1]} g(x) \right| \leq \sup_{x\in[0,1]} |g_n(x) - g(x)|,$$

Proposition 1 implies that $\sup_{x\in[0,1]} g_n(x) \to \sup_{x\in[0,1]} g(x)$ as $n \to \infty$. Thus $\limsup_{n\to\infty} n^{\alpha/2}e_n(L, \alpha) \leq L\sup_{x\in[0,1]} g(x) = Lg(1/2)$. ■

We turn to proofs of the lemmas.

Proof of Lemma 1. Since the random variable $nS_n(x)$ is binomially distributed for every x, we may apply the De Moivre–Laplace Theorem, which asserts that

$$P\left(a < \frac{nS_n(x) - nx}{\sqrt{nx(1-x)}} < b\right) \to \sqrt{\frac{1}{2\pi}} \int_a^b e^{-u^2/2}\, du$$

for all $a < b$. Especially for all $t > 0$ this yields

$$P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > t\right) \to \sqrt{\frac{2}{\pi}} \int_t^\infty e^{-u^2/2}\, du. \tag{6}$$

Furthermore, we may use Chebyshev's inequality to conclude that

$$P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > t\right) \leq \frac{\mathbf{E}\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right|^2}{t^2} = \frac{1}{t^2},$$

hence

$$P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > t\right) \leq \min\left\{1, \frac{1}{t^2}\right\}. \tag{7}$$

We recall that $g_n(x) = n^{\alpha/2}\mathbf{E}|S_n(x) - x|^\alpha$. Stieltjes integration allows us to compute

$$g_n(x) = \alpha \int_0^\infty t^{\alpha-1} P\left(\sqrt{n}\,|S_n(x) - x| > t\right) dt$$

$$= \alpha(x(1-x))^{\alpha/2} \int_0^\infty s^{\alpha-1} P\left(\left|\frac{nS_n(x) - nx}{\sqrt{nx(1-x)}}\right| > s\right) ds.$$

Since $\int_0^\infty t^{\alpha-1}\min\{1, t^{-2}\}\,dt < \infty$ and we have pointwise convergence as in (6), the Lebesgue Dominated Convergence Theorem implies

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \alpha \int_0^\infty t^{\alpha-1} P\big(\sqrt{n}\,|S_n(x) - x| > t\big)\,dt$$

$$= \big(x(1-x)\big)^{\alpha/2} \int_0^\infty s^\alpha \sqrt{\frac{2}{\pi}}\, e^{-s^2/2}\,ds$$

$$= \big(x(1-x)\big)^{\alpha/2} \sqrt{\frac{2}{\pi}}\, 2^{(\alpha-1)/2} \int_0^\infty u^{(\alpha+1)/2-1} e^{-u}\,du$$

$$= \big(x(1-x)\big)^{\alpha/2} 2^{\alpha/2} \frac{\Gamma((\alpha+1)/2)}{\sqrt{\pi}} = g(x). \qquad \blacksquare$$

It is evident from the proof that K_α is the αth absolute moment of the standard normal distribution, i.e., $K_\alpha = \mathbf{E}|\gamma|^\alpha$, where γ is standard normal. This could also be guessed, since the De Moivre–Laplace Theorem may be viewed as a special instance of the Central Limit Theorem.

Proof of Lemma 2. Since $\big||a|^\alpha - |b|^\alpha\big| \le |a-b|^\alpha$ for all $a, b \in \mathbb{R}$ and $0 < \alpha \le 1$, we have

$$|g_n(x) - g_n(y)| \le n^{\alpha/2}\mathbf{E}\big|(S_n(x) - x) - (S_n(y) - y)\big|^\alpha$$

$$\le n^{\alpha/2}\Big(\mathbf{E}\big|(S_n(x) - x) - (S_n(y) - y)\big|^2\Big)^{\alpha/2}$$

$$= \big(n\mathrm{Var}(S_n(x) - S_n(y))\big)^{\alpha/2}.$$

Now suppose $x < y$ and recall that $S_n(x) - S_n(y) = -\sum_{j=1}^n \frac{1}{n}\chi_{[x,y)}(u_j)$, which is the sum of n independent random variables. It is well known that then the variance of the sum equals the sum of the variances, which implies

$$|g_n(x) - g_n(y)| \le \left(n^2 \mathrm{Var}\left(\frac{1}{n}\chi_{[x,y)}(u)\right)\right)^{\alpha/2} = \big(\mathrm{Var}(\chi_{[x,y)}(u))\big)^{\alpha/2}$$

$$= \big((y-x)(1-(y-x))\big)^{\alpha/2} \le |y-x|^{\alpha/2} \qquad \blacksquare$$

We have thus proved the lemmas and hence Theorem 2. We admit that deriving the exact asymptotics of the error is much more elaborate than the simple arguments leading to Theorem 1. The gain in accuracy is less than $\sqrt{2/\pi}$.

To complete our development of properties of the Bernstein polynomials using the probabilistic representation, we add the following result, originally proved by Brown, Elliott, and Paget [**3**].

Proposition 2. *If the function $f:[0,1] \to \mathbb{R}$ is Hölder continuous with exponent α and constant L then so are the corresponding Bernstein polynomials $B_n(f, \cdot)$.*

Proof: Use the triangle inequality, Hölder continuity, the representation of $S_n(x)$ as an empirical distribution function, and finally the Hölder inequality to derive for

$x \le y$ the estimates

$$|B_n(f,x) - B_n(f,y)| \le \mathbf{E}|f(S_n(x)) - f(S_n(y))| \le L\mathbf{E}|S_n(x) - S_n(y)|^{\alpha}$$
$$= L\mathbf{E}\left|\frac{1}{n}\sum_{j=1}^{n} \chi_{[x,y)}(u_j)\right|^{\alpha} \le L\left(\mathbf{E}\left|\frac{1}{n}\sum_{j=1}^{n} \chi_{[x,y)}(u_j)\right|\right)^{\alpha}$$
$$\le L|x-y|^{\alpha}. \qquad \blacksquare$$

This proof is brief and elementary, but heavily uses the specific realization of the $S_n(x)$ through an empirical distribution function, thereby correlating the random variables $S_n(x)$ and $S_n(y)$ properly. This was also crucial for proving Lemma 2. The simple arguments leading to Theorem 1 did not rely on any specific realization.

REFERENCES

1. H. Bauer, Approximation and abstract boundaries, *Amer. Math. Monthly* **85** (1978) 632–647.
2. S. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilities, *Comm. Soc. Math. Kharkov* **13** (1912) 1–2.
3. B. M. Brown, D. Elliott, and D. F. Paget, Lipschitz constants for the Bernstein polynomials of a Lipschitz continuous function, *J. Approx. Theory* **49** (1987) 196–199.
4. P. Diaconis and S. Zabell, Closed form summation for classical distributions: Variations on a theme of De Moivre, *Statist. Science* **6** (1991) 284–302.
5. H. Gzyl and J. L. Palacios, The Weierstrass approximation theorem and large deviations, *Amer. Math. Monthly* **104** (1997) 650–653.
6. M. Kac, Une remarque sur les polynomes de M. S. Bernstein, *Studia Math.* **7** (1938) 49–51.
7. M. Kac, Reconnaissance de priorité relative à ma note, "Une remarque sur les polynomes de M. S. Bernstein", *Studia Math.* **8** (1939) 170.
8. G. G. Lorentz, *Bernstein Polynomials, Math. Expo.* Vol. 8, Univ. of Toronto Press, Toronto, 1953.
9. T. Popoviciu, Sur l'approximation des fonctions convexes d'ordre supérieur, *Mathematica* **10** (1934) 49–54.
10. D. Werner, *Funktionalanalysis, 2nd Ed.* Springer Lehrbuch, Springer, Berlin, 1997.

Weierstrass-Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D–10117 Berlin, Germany
mathe@wias-berlin.de

An Extension of the Wallace-Simson Theorem: Projecting in Arbitrary Directions

Miguel de Guzmán

1. THE WALLACE-SIMSON LINE. The Wallace line has been a popular object of study for many geometers during the two past centuries. Let us start by recalling the theorem.

The Wallace-Simson Theorem. *Consider a triangle ABC. The locus of all those points P in its plane such that the orthogonal projections of P on the three sides of the triangle are collinear is the circumcircle of ABC. The line of the projections is called the Wallace-Simson line of P with respect to ABC.*