Ergodic Theory Paper – The Stone-Weierstrass Theorem

Agastya Goel

March 12, 2024

1 The Weierstrass Approximation Theorem

The Stone-Weierstrass theorem is an important theorem in real analysis, and by extension, ergodic theory. In this paper, we will prove a weaker version of it with Bernstein polynomials and demonstrate applications to ergodic theory and beyond.

Theorem 1.1 (Weierstrass Approximation Theorem)**.** *For any continuous function $f : [0, 1] \to \mathbb{R}$ and any $\epsilon > 0$, we can find a polynomial $P : [0, 1] \to \mathbb{R}$ such that $||f - P||_\infty < \epsilon$.*

The key tool we will use to prove this theorem is the Bernstein polynomial.

Definition 1.2. The n^{th} Bernstein polynomial for a continuous function $f : [0, 1] \to \mathbb{R}$ is

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

Bernstein polynomials have a few key properties. First off, they are linear in f, so $B_n(cf, x) = cB_n(f, x)$ and $B_n(f + g, x) = B_n(f, x) + B_n(g, x)$. Furthermore, $B_n(1, x) = 1$ by the binomial theorem.

The key idea of our proof is that the Bernstein polynomial approximates f very well because $x^k x^{n-k}$ is sharply peaked at $\frac{k}{n}$, so, for large n, our polynomial will look like spikes of the appropriate height added together. Stating this idea more formally:

Lemma 1.3. *For any $\epsilon > 0$, there exists some N such that for $n \geq N$, $|B_n(f) - f|_\infty < \epsilon$.*

We can now prove this form of the theorem. Consider some value $s \in [0, 1]$. Then,

$$\begin{aligned} |B_n(f, s) - f(s)| &= |B_n(f, s) - f(s)B_n(1, s)| \\ &= |B_n(f, s) - B_n(f(s), s)| \\ &= |B_n(f - f(s), s)|. \end{aligned}$$

Note that if we have some function g such that $g > |f|$, $B_n(g)$ is clearly greater than $B_n(f)$. Thus, our goal will be to bound $|f - f(s)|$ above and show that the Bernstein polynomial evaluated on this bound goes to 0 uniformly.

To bound f above, we use the idea of uniform continuity. That is, for any $\epsilon > 0$, we can find a δ such that for all $x, y \in [0, 1]$ with $|x - y| < \delta$, $|f(x) - f(y)| < \epsilon$. This idea generalizes that of regular continuity, in which the value of δ is allowed to depend on x and ϵ. For example, the function $x \sin(x)$ grows bigger and bigger and so we requires smaller and smaller δ values for larger and larger x. A classic results from real analysis is that any continuous function is uniformly continuous over any compact set, such as the unit interval.

Now, say we want to establish the bound $|B_n(f, s) - f(s)| < \epsilon$ for $s \in [0, 1]$. Then, we will choose a δ such that for all $x, y \in [0, 1]$ with $|x - y| < \delta$, $|f(x) - f(y)| < \frac{\epsilon}{2}$. Then, if $M = ||f||_\infty$. we may observe the following bounds on $|f(x) - f(s)|$ for all s:

$$\begin{cases} |f(x) - f(s)| < \frac{\epsilon}{2} & |x-s| < \delta \\ |f(x) - f(s)| \leq 2M \leq 2M\frac{(x-s)^2}{\delta^2} & |x-s| \geq \delta. \end{cases}$$

Now, we can add these two cases to obtain

$$|f(x) - f(s)| < \frac{\epsilon}{2} + 2M\frac{(x-s)^2}{\delta^2}$$

for all x. The reason that we chose to use the expression $2M\frac{(x-s)^2}{\delta^2}$ rather than $2M$ is because the former takes on a lower value for $|x-s| < \delta$. With this bounding function, which we will denote by $g(x) = \frac{\epsilon}{2} + 2M\frac{(x-s)^2}{\delta^2}$, we will now show that the Bernstein polynomial evaluated at s with this function goes to 0. Intuitively, we should expect this, because the polynomial evaluated at s depends strongly on the values near it for large n. More formally,

$$\begin{aligned} B_n(g, x) &= \frac{\epsilon}{2} + \frac{2M}{\delta^2} B_n((x-s)^2, x) \\ &= \frac{\epsilon}{2} + \frac{2M}{\delta^2} \sum_{k=0}^{n} \left(\frac{k}{n} - s\right)^2 \binom{n}{k} x^k (1-x)^{n-k} \\ &= \frac{\epsilon}{2} + \frac{2M}{\delta^2} \left(s^2 \left(\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \right) - \frac{2s}{n} \left(\sum_{k=0}^{n} k \binom{n}{k} x^k (1-x)^{n-k} \right) + \frac{1}{n^2} \left(\sum_{k=0}^{n} k^2 \binom{n}{k} x^k (1-x)^{n-k} \right) \right). \end{aligned}$$

Now:

$$\sum_{k=0}^{n} k^a \binom{n}{k} x^k (1-x)^{n-k} = (1-x)^n \sum_{k=0}^{n} k^a \binom{n}{k} \left(\frac{x}{1-x}\right)^k.$$

Now, we will take a moment to derive an expression for $\sum_{k=0}^{n} k^a \binom{n}{k} z^k$. Let $D(P) = z\frac{\mathrm{d}P}{dz}$ where P is a polynomial. Essentially, we multiply the z^k term by k. Then, we have:

$$\sum_{k=0}^{n} \binom{n}{k} z^k = (1+z)^n$$

$$\sum_{k=0}^{n} k \binom{n}{k} z^k = D\left((1+z)^n\right) = nz(1+z)^{n-1}$$

$$\sum_{k=0}^{n} k^2 \binom{n}{k} z^k = D\left(nz(1+z)^{n-1}\right) = nz(1+z)^{n-1} + n(n-1)z^2(1+z)^{n-2}.$$

Thus, we have:

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

$$(1-x)^n \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{x}{1-x}\right)^k = (1-x)^n n \left(\frac{x}{1-x}\right) \left(1 + \frac{x}{1-x}\right)^{n-1} = nx$$

$$(1-x)^n \sum_{k=0}^{n} k^2 \binom{n}{k} \left(\frac{x}{1-x}\right)^k = (1-x)^n n \left(\frac{x}{1-x}\right) \left(\frac{1}{1-x}\right)^{n-2} \left(\frac{1}{1-x} + (n-1)\left(\frac{x}{1-x}\right)\right) = nx\left((n-1)x + 1\right).$$

Plugging this back into our equation for $B_n(g, x)$ and evaluating it at $x = s$:

$$
\begin{aligned}
B_n(g, x) &= \frac{\epsilon}{2} + \frac{2M}{\delta^2}\left(s^2 - \frac{2s}{n}nx + \frac{1}{n^2}nx\left((n-1)x+1\right)\right) \\
&= \frac{\epsilon}{2} + \frac{2M}{\delta^2}\left(s^2\left(1 - 2 + 1 - \frac{1}{n}\right) + \frac{s}{n}\right) \\
&= \frac{\epsilon}{2} + \frac{2M}{n\delta^2}\left(s - s^2\right).
\end{aligned}
$$

Since $s - s^2$ is bounded, for large enough n, we can make this value less than ϵ for all s simultaneously, which shows that for any ϵ, we can find an n such that the approximation is within ϵ everywhere. See here for an example of this approximation in action.

Note that the proof we gave above applies to continuous functions over $[0, 1]$, but the proof can be easily extended to any closed interval.

2 Extensions

The Stone–Weierstrass Theorem as usually stated is much more general than the version given here. Intuitively, this is to be expected because the only properties of Bernstein polynomials that we used were that they are linear, sharply peaked, and perfectly approximate constant functions. Such properties allow us to use a wide variety of function types to uniformly approximate continuous functions. In particular, we can approximate continuous functions with trignometric polynomials, (i.e., sin, cos, and their powers). For a continuous function $f(x)$ over $[0, 1]$, and a large number n, we may approximate f as

$$
\frac{1}{n^2}\sum_{i=0}^{2n}\left(f\left(\frac{i}{2n}\right)\left(\sum_{j=0}^{n-1}\cos\left(2\pi j\left(x - \frac{i}{2n}\right)\right)\right)^2\right).
$$

(Note that this does not actually converge to f at 0 and 1. This issue can be taking a subinterval of $[0, 1]$, but I have chosen to leave it in this form because of its simplicity.) This approximation is demonstrated on the same Desmos as the Bernstein polynomial example. The fundamental reason it works is because the function

$$
\frac{2}{n}\left(\sum_{j=0}^{n-1}\cos\left(2\pi j\left(x - \frac{i}{2n}\right)\right)\right)^2
$$

is sharply peaked at $\frac{i}{2n}$, and it integrates to 1. Now, we cannot use the same proof as we did for the Bernstein polynomials because these polynomials do not perfectly approximate constant functions.

Intuitively, we should also expect exponential quadratics (i.e. functions of the form e^{ax^2+bx+c} to work since, for negative a, these functions are also sharply peaked. In particular, we may approximate f as

$$
\sum_{i=0}^{n}\left(f\left(\frac{i}{n}\right)\frac{1}{\sqrt{2\pi}}e^{-n^2\frac{\left(x-\frac{i}{n}\right)^2}{2}}\right).
$$

Once again, this shown on the same Desmos, and it works because

$$
\frac{n}{\sqrt{2\pi}}e^{-n^2\frac{\left(x-\frac{i}{n}\right)^2}{2}}
$$

is strongly peaked at $\frac{i}{n}$ and (almost) normalized. However, we cannot prove it in exactly the same way as we did for polynomials because it is difficult to make a sum of normal distributions exactly 1. Approximating functions with sums of normal distributions can be done formally, however. To demonstrate this, we will now introduce the more general Stone–Weierstrass theorem (over functions), although we do not have the tools to prove it.

Theorem 2.1 (Stone–Weierstrass)**.** *A set of real-valued functions that is closed under addition, multiplication, and scalar multiplication; includes all constant functions; and separates points (i.e., for any points x, y, there is a function for which the two points have different values) can uniformly approximate any continuous real-valued function over a compact Hausdorff space.*

We will not define a Hausdorff space in this paper, but most reasonable compact sets are also Hausdorff spaces. In particular, compact subsets of $\mathbb{R}^n$ are Hausdorff spaces.

3 Applications

In this section, we will apply the various forms of the Stone–Weierstrass theorem. First, we will review one of its applications to ergodic theory, and then we will discuss applications to other branches of mathematics.

First, we will prove the ergodicity of irrational rotations. This example is derived the Euler Circle ergodic theory class.

Theorem 3.1. *For an irrational number α, the transformation*

$$T(x) = x + \alpha$$

over $\mathbb{R}/\mathbb{Z}$ is ergodic.

Proof. We will show that T is uniquely ergodic via Oxtoby's theorem. In particular, consider the function $f = \cos(2\pi kx + \phi)$ for some values of k, ϕ. In particular, k must be an integer to ensure that f is continuous. Then, we want to show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j(x)\right) = \int_0^1 \cos(2\pi kx + \phi) \mathrm{d}x = \begin{cases} 1 & k = 0 \\ 0 & \end{cases}.$$

For the rest of this proof, we will neglect ϕ, since by proving that $\cos(2\pi kx)$ has the desired property, we prove the same for all shifts of this function. Now,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j(x)\right) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \cos\left(2\pi k\left(x + j\alpha\right)\right) \\
&= \lim_{n\to\infty} \frac{1}{2n} \sum_{j=0}^{n-1} \left(e^{i(2\pi k(x+j\alpha))} + e^{-i(2\pi k(x+j\alpha))}\right) \\
&= \lim_{n\to\infty} \frac{1}{2n} \left(e^{2\pi ikx} \sum_{j=0}^{n-1} e^{2\pi ikj\alpha} + e^{-2\pi ikx} \sum_{j=0}^{n-1} e^{-2\pi ikj\alpha}\right).
\end{aligned}$$

If $k = 0$, we obtain the desired result. Otherwise, $e^{2\pi ik\alpha} \neq 1$, so we may apply the formula for the sum of a geometric series:

$$\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{2n} \left(e^{2\pi ikx} \frac{1 - e^{2\pi ikn\alpha}}{1 - e^{2\pi ik\alpha}} + e^{-2\pi ikx} \frac{1 - e^{-2\pi ikn\alpha}}{1 - e^{-2\pi ik\alpha}}\right) \\
&\leq \lim_{n\to\infty} \frac{1}{2n} \left(\left|e^{2\pi ikx}\right| \left|\frac{1 - e^{2\pi ikn\alpha}}{1 - e^{2\pi ik\alpha}}\right| + \left|e^{-2\pi ikx}\right| \left|\frac{1 - e^{-2\pi ikn\alpha}}{1 - e^{-2\pi ik\alpha}}\right|\right) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \left(\left|\frac{1}{1 - e^{2\pi ik\alpha}}\right| + \left|\frac{1}{1 - e^{-2\pi ik\alpha|}}\right|\right) \\
&= 0.
\end{aligned}$$

Hence, we have shown the desired result for all cos and sin functions. Clearly, the above proof also holds for linear combinations of these functions, and since any product of cos, sin functions can be expressed as a sum of those same functions, we may use the Weierstrass approximation theorem to show that for any ϵ,

$$\left| \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j(x)\right) \right) - c(f) \right| < \epsilon$$

for any $\epsilon > 0$, and thus, the left hand side is equal to 0, which shows the unique ergodicity of T. □

The above proof fails for rational numbers α because we might have

$$e^{2\pi i k \alpha} = 1$$

even when $k \neq 0$.

Now, we will show a classic result from differential topology. This example is taken from the Euler Circle Differential Topology class.

Theorem 3.2 (The Brouwer fixed-point theorem)**.** *Let B^n be the n dimensional ball defined by*

$$B^n = \{\mathbf{v} \in \mathbb{R}^n : ||\mathbf{v}|| \leq 1\}.$$

Then, any continuous map $f : B^n \to B^n$ has a fixed point (i.e., a point $\mathbf{x}$ where $f(x) = x$).

The differential topology background required to prove this is beyond the scope of this paper, so we will instead assume a simpler result which is due to differential topology.

Lemma 3.3. *A map is smooth if it is infinitely differentiable. Then, any smooth map $f : B^n \to B^n$ has a fixed point.*

Proof of Theorem 3.2. Consider the first coordinate of our map f. If we can uniformly approximate one coordinate with a smooth function, we can approximate the map overall with a smooth function. Thus, we will consider multivariate polynomials of the Cartesian coordinates of $\mathbf{x}$. To apply Stone–Weierstrass, we must show that

- B^n is a compact Hausdorff space,
- Multivariate polynomials are smooth,
- Multivariate polynomials separate points.

B^n is a compact Hausdorff space, since it is a subset of $\mathbb{R}^n$. Multivariate polynomials closed under addition, multiplication, and scalar multiplication, so they form an algebra. Finally, we can easily find smooth functions that separate points; in particular, the function

$$f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{u}$$

separates the points $\mathbf{u}, \mathbf{v}$ and can be rewritten as a multivariate polynomial. Thus, we may apply Stone–Weierstrass to conclude that we may uniformly approximate each of the coordinates of our map, and thus, the map itself.

Now, we will employ a proof by contradiction. Since B^n is compact, if $f(x) \neq x$, the minimum

$$\epsilon = \min_{\mathbf{x} \in B^n} ||\mathbf{x} - \mathbf{f}(\mathbf{x})||$$

exists and is positive. However, if we approximate f with a smooth function g to an error lower than ϵ, then at a fixed point of g, say, $\mathbf{y}$, we must have

$$||\mathbf{y} - f(\mathbf{y})|| < \epsilon,$$

which is a contradiction. □