Math 104, Summer 2019
Lecture: Bernstein Polynomials, 8/5/2019

1 Motivation for Bernstein: Better Than Polynomial Interpolation?

Why do we need Bernstein polynomials when we can simply interpolate a function at a set of points?

Suppose given a continuous function f on $[a, b]$, we want a sequence of polynomials $(p_n) \to f$ **uniformly**. Naively, we may take p_n to be unique polynomial of degree n interpolating precisely at a collection of $n + 1$ points.

Example: Suppose $f(x) := |x|$ on $[-1, 1]$. Suppose we take $p_0 = 0$ and p_1 is the line passing through $(-1, 1)$ and $(1, 1)$. Take p_2 to be the quadratic passing through $(-1, 1), (0, 0), (1, 1)$. Take p_3 to be the cubic through $(-1, 1), \left(-\frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{1}{3}\right), (1, 1)$.

Because we construct our interpolation very carelessly, the actual approximation error for this example grows unboundedly. There is certainly no hope of using this method for approximation. We note that our collection of points is inherently suboptimal, but certainly this example illustrates that simple naive interpolation can perform poorly.

2 Power Series for Approximation

A question was posed in lecture: why don't we use power series to approximate a function?

Recall that last time we mentioned that any power series that has a nondegenerate radius of convergence **has a derivative**. The convergence of that derivative's power series is the same, modulo maybe the endpoints. Take for instance any power series centered about $x = 0$. We can differentiate a power series an infinite amount of times (we call these **analytic** as they are smooth).

The problem is that there are many continuous functions that are not smooth, and it turns out that 'most' continuous functions are not smooth has **no** derivative at **any** point. It's not bad to think of behavior in the middle of an interval somewhat as Brownian motion, and that behavior at the endpoints is more like (wild) oscillation. The idea here is that if we take any closed subset $[a', b'] \subset (a, b)$, then f is bounded on any fixed interval; however, at points $\xi \in (a, b) \setminus [a', b']$, $f(\xi)$ can behave wildly.

We say that any interval $(a, b) \subset \mathbb{R}$ is topologically the same as $\mathbb{R}$, so we simply consider an interval $[0, 1]$ with the idea that we can easily transform to any interval $(a, b) \subset \mathbb{R}$.

3 Bernstein Polynomials

First we post our claim (definition) then work through some motivations and intuition that are otherwise hidden in the proof supplied by Ross.

Definition: Bernstein Polynomial -

Take f to be a continuous function on $[0,1]$ (and because f is continuous on a closed interval, f is uniformly continuous). Define the nth Bernstein polynomial by

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

Notice that the only dependence on x is in the $x^k(1-x)^{n-k}$ factors.

Example: Take $f(x) = x^2$ as a simple example (assuming we choose not to take a sequence of simply x^2.) We then write that the Bernstein polynomial is

$$B_3 f(x) := \frac{1}{9}\binom{3}{1}x(1-x)^2 + \frac{4}{9}\binom{3}{2}x^2(1-x) + \binom{3}{3}x^3,$$

with degree 3. The binomial coefficients come from the value of f itself.

Example: Take $f(x) = \cos x$. Then we write:

$$\begin{aligned} B_3 f(x) = &\binom{3}{0}(1-x)^3 + \cos\left(\frac{1}{3}\right)\binom{3}{1}x(1-x)^2 \\ &+ \cos\left(\frac{2}{3}\right)\binom{3}{2}x^2(1-x) + \cos(1)\binom{3}{3}x^3 \end{aligned}$$

4 Demystifying Bernstein Polynomials

Max gives that this is truly elegant and notes that Ross fails to include the intuition at all. We do not regurgitate the proof supplied in the book but rather employ the following techniques.
We invent new notation for our purposes. Define the function

$$g_{n,k}(x) := \binom{n}{k} x^k (1-x)^{n-k}.$$

We investigate this function further to understand what is truly going on. Taking an arbitrary $g_{n,k}$, we want to 'plot' its behavior. Unless $k = 0$ or $k = n$, $g_{n,k}(0) = g_{n,k}(1) = 0$. Recall that the standard single-variable to find the local max and min is to set a function's derivative to 0 and solving. We spare this computation and note that

$$g_{n,k}\left(\frac{k}{n}\right) = \binom{n}{k}\frac{k^k(n-k)^{n-k}}{n^n}$$

We will use Stirling's Approximation to deal with the $\binom{n}{k}$ factor, which we have essentially seen before.

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

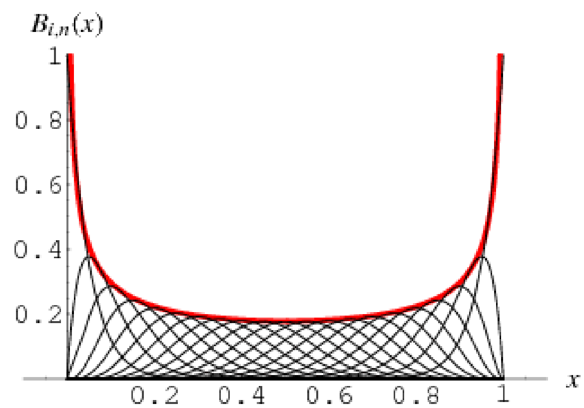
Now we try to approximate $\binom{n}{k}\frac{k^k(n-k)^{n-k}}{n^n}$ in terms of factorials:

$$\begin{aligned}\max\{g_{n,k}(x), x \in [0,1]\} &\approx \binom{n}{k}\frac{\left(\frac{e^k k!}{\sqrt{2\pi k}}\right)\left(\frac{e^{n-k}(n-k)!}{\sqrt{2\pi(n-k)}}\right)}{\frac{e^n e!}{\sqrt{2\pi n}}} \\ &= \frac{\sqrt{2\pi n}}{\sqrt{2\pi k}\sqrt{2\pi(n-k)}}\end{aligned}$$

If we imagine n is fixed and let k range from 0 to n, then for some large n, k/n acts as equispaced points from 0 to 1. We can get a sense for the peaks for each k/n.

It turns out that the peaks seem to be fitted (enveloped) under the graph of

$$\text{envelope} := \frac{1}{\sqrt{2\pi n}}\sqrt{\frac{1}{x-x^2}}$$



We consider the two cases $x = 0, 1$ separately, where $g_{n,0}(0) = g_{n,n}(1) = 1$. Now when we let n vary, this curve of h above gets depressed downwards. The following lemma is generally true (past our purposes).

Lemma 4.1. For any $x \in \mathbb{R}$ and $n \geq 0$,

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

To see quickly this is true, recall the Binomial expansion for $(a+b)^n$ with $a := x$ and $b := (1-x)$, where $(x + (1-x)^n = 1$.

Definition: Partition of Unity -

We say that the collection $g_{n,k}(x)$ is a **partition of unity** centered about x with interval width approximately $\frac{1}{\sqrt{n}}$, in that these partitions sum to 1 ('unity').

Remark: If we look at some value of x where $x \approx \frac{k}{n}$, then in the Bernstein polynomial when we vary the value of k' with n fixed, the term

$$\binom{n}{k'} x^{k'} (1-x)^{n-k'}$$

has the largest value when $k' = k$.

Now when $x \approx \frac{m}{n}$, then $g_{n,k} = \binom{n}{k} x^k (1-x)^{n-k}$ has the largest value when $k = m$. In fact, the values of $g_{n,k}(x)$ is proportional to $\frac{1}{\sqrt{n}}$, as seen by the expression for the envelope. That is, for some fixed x,

$$\frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1}{x - x^2}} = \underbrace{\frac{1}{\sqrt{2\pi}} \sqrt{\frac{1}{x(1-x)}}}_{\text{constant}} \frac{1}{\sqrt{n}}$$

We write

$$\left| \frac{k-m}{n} \right| <\approx \frac{1}{\sqrt{n}} \iff |k-m| <\approx \sqrt{n}$$

Remark: Essentially,

$$B_n f(x) \approx 2\sqrt{n} \int_{x - \frac{1}{\sqrt{n}}}^{x + \frac{1}{\sqrt{n}}} f(t)\, dt \to f(x).$$

Alternatively, we can reason that if x is the probability of 'heads' and $(1-x)$ is the probability of 'tails' and treat $f\left(\frac{k}{n}\right)$ as a random variable on a binomially distributed probability space, then we are calculating the expected value of f. With intuition on probability, we can expect $B_n f(x)$ to converge (at least pointwise) to $f(x)$.

5 Pseudo-Proof

Now we work through a pseudo-proof as inspired by Ross. We want to show $B_n f \to f$ uniformly on $[0, 1]$. We take

$$N := \frac{M}{\epsilon \delta^2},$$

where M takes on an eventual importance; however, for now we draw attention to the relationship between N and δ in that δ is proportional to $\frac{1}{\sqrt{N}}$ and particularly that N **is chosen independent of** x.
Applying the triangle inequality once, we get:

$$|B_n f(x) - f(x)| \leq \underbrace{\sum_{k=0}^{n} \left| f\left(\frac{k}{n}\right) - f(x) \right|}_{*} \underbrace{\binom{n}{k} x^k (1-x)^{n-k}}_{*'}.$$

Recall that in heuristic terms from earlier, we reasoned that the nonnegligible terms are large when $\frac{k}{n} \approx x$. We want to divide the error bound based (arbitrarily) on the condition:

$$\left| \frac{k-m}{n} \right| <\approx \frac{1}{\sqrt{n}}$$

where one of the above underbraced terms (*, *') is small. Precisely, Ross chooses to draw the distinction to break the set $\{0, 1, \ldots, n\}$ into two sets:

$$k \in A \text{ if } \left| \frac{k}{n} - x \right| < \delta$$
$$k \in B \text{ if } \left| \frac{k}{n} - x \right| \geq \delta,$$

and the number of elements of A is proportional to $\frac{1}{\sqrt{n}}$, and B is the set of remaining indices.
Now we write:

$$\sum_{k \in A} \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} \leq \sum_{k \in A} \frac{\epsilon}{2} \underbrace{\binom{n}{k} x^k (1-x)^{n-k}}$$
$$= \frac{\epsilon}{2} \underbrace{\left(\sum \cdots \right)}_{\leq 1}$$

We picked M as an upper bound for all values of f, so we have:

$$\sum_{k \in B} \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} \leq 2M \left[\sum_{k \in B} \binom{n}{k} x^k (1-x)^{n-k} \right]$$
$$\leq \frac{2M}{n^2 \delta^2} \underbrace{\left[\sum_{k \in B} \overbrace{(k - nx)^2}^{= n^2 \left(\frac{k}{n} - x\right)^2} \binom{n}{k} x^k (1-x)^{n-k} \right]}_{\leq n/4}.$$

This result in the book appears magical and unmotivated, but this lecture is designed to provide intuition behind this. Otherwise, understanding Ross' derivation and proof line by line can suffice if one accepts the lack of motivation.