2). Minimax Approximation

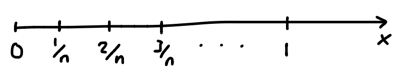
2.1). Weierstrass Approximation Theorem

Thm 2.1 — Let $f \in C[0,1]$. Then for any $\varepsilon > 0$ there exists a polynomial p such that
(Weierstrass, 1885) aged 70
$$\|f - p\|_\infty < \varepsilon.$$

- For any other interval $[a,b]$ we can just change variables.
- This is stronger than our previous error bounds (e.g. Taylor's Thm or Lagrange interpolation) because it doesn't require differentiability of f, only continuity.

We will give Sergei Bernstein's proof (ca. 1910) which provides an explicit sequence of converging polynomials.

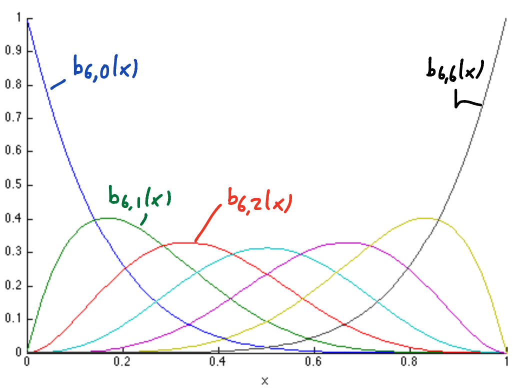
The n^{th} Bernstein polynomial for f is defined as
$$B_n(f,x) = \sum_{k=0}^{n} f(k/n)\binom{n}{k} x^k (1-x)^{n-k}.$$



usual Binomial coefficient $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Note that
$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \overset{\text{binomial thm.}}{=} (x + (1-x))^n = 1^n = 1$$
so $B_n(f,x)$ at any point $x \in [0,1]$ is a weighted average of the $n+1$ function values $f(k/n)$.

e.g. the functions $b_{n,k}(x) := \binom{n}{k} x^k (1-x)^{n-k}$ for $n = 6$:



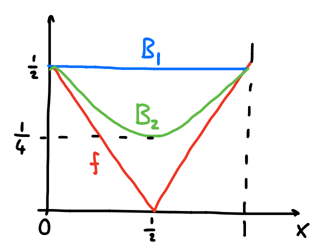
Notice that each $b_{n,k}(x)$ peaks near a different point of $[0,1]$.

Example: Approximate $f(x) = |x - \frac{1}{2}|$ on $[0,1]$ with Bernstein polynomials.

$n=1$: $B_1(f,x) = f(0)\binom{1}{0}(1-x) + f(1)\binom{1}{1}x = \frac{1}{2}(1-x) + \frac{1}{2}x = \frac{1}{2}.$

$n=2$: $B_2(f,x) = f(0)\binom{2}{0}(1-x)^2 + f(\frac{1}{2})\binom{2}{1}x(1-x) + f(1)\binom{2}{2}x^2$
$= \frac{1}{2} - x + x^2.$

Convergence is slow $(= n^{-1/2})$ at $x = \frac{1}{2}$.



This slow convergence makes Bernstein approximation impractical.

Key idea of Bernstein's proof :- the weights $b_{n,k}(x)$ become more and more localised around $x = k/n$ as $n \to \infty$.

We will use the fact that $b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$ is the probability mass fuction of the binomial distribution – ie. the probability of getting k successes in n trials, each with probability x. recall from first year!

– the mean per trial is $E(k/n) = x$ and the variance is $Var(k/n) = \frac{x(1-x)}{n}$.

Since $Var(k/n)$ shrinks as n grows, each distribution clusters more tightly around $x = k/n$, and $B_n(f, k/n) \to f(k/n)$.

Proof of Thm 2.1 –

The error is

$$|f(x) - B_n(f,x)| = \left| \sum_{k=0}^{n} \overbrace{b_{n,k}(x)}^{=1} f(x) - \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f(k/n) \right|$$

$$\leq \sum_{k=0}^{n} |f(x) - f(k/n)| \, b_{n,k}(x).$$

Now divide the sum into k/n near x (major contribution), say $|x - k/n| \leq \delta$, and those with $|x - k/n| > \delta$.

The first part is

$$\sum_{|x - k/n| \leq \delta} |f(x) - f(k/n)| \, b_{n,k}(x).$$

We simply choose δ small enough so that $|f(x) - f(k/n)| < \varepsilon/2$ whenever $|x - k/n| \leq \delta$.

The second sum is

$$\sum_{|x - k/n| > \delta} |f(x) - f(k/n)| \, b_{n,k}(x) \leq M \underbrace{\sum_{|x - k/n| > \delta} b_{n,k}(x)}_{S} \qquad \text{where } M = \| f(x) - f(k/n) \|_\infty .$$

We hope that for n large enough we will have $S \leq \frac{\varepsilon}{2M}$. This is reasonable since $b_{n,k}(x)$ is sizeable over a narrower domain as n increases.

To prove it rigorously, we can use

Chebyshev's Inequality – If X is a discrete random variable with mean $E(x)$ and variance $Var(x)$, then

$$\text{Prob}\left\{ |X - E(x)| \geq s\sqrt{Var(x)} \right\} \leq \frac{1}{s^2}.$$

Here $E(k/n) = x$, $Var(k/n) = \frac{x(1-x)}{n}$, and S is the probability of being more than δ from the mean.

To ensure $S \leq \frac{\varepsilon}{2M}$, Chebyshev would do it if $\frac{1}{s^2} = \frac{\varepsilon}{2M}$, so

$$s\sqrt{Var(x)} \leq \delta \iff \delta \geq \sqrt{\frac{x(1-x)}{n} \frac{2M}{\varepsilon}} \iff n \geq \frac{2Mx(1-x)}{\varepsilon \delta^2}.$$

Thus it works for n large enough. ▨