10.2.2. Weierstrass Approximation Theorem.
Let f be any continuous real-valued function on $[a, b]$. Then there is a sequence of polynomials p_n that converges uniformly to f on $[a, b]$.

In the language of normed vector spaces, this theorem says that the polynomials are dense in $C[a, b]$ in the max norm.

In fact, this theorem is sufficiently important that many different proofs have been found. The proof we give was found in 1912 by Bernstein, a Russian mathematician. It explicitly constructs the approximating polynomial. This algorithm is not the most efficient, but the problem of finding efficient algorithms can wait until we have proved that the theorem is true.

from <u>Real Analysis with Real Applications</u>

by Davidson and Donsig.

10.3. Bernstein's Proof of the Weierstrass Theorem

Recall the binomial formula, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$. If we set $y = 1-x$, then we obtain

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

Bernstein started by considering the functions

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

now called **Bernstein polynomials**. They have several virtues. They are all polynomials of degree n. They take only nonnegative values on $[0,1]$. And they add up to 1. Moreover, P_k^n is a "bump" function with a maximum at k/n, as a routine calculus calculation shows. For example, the four functions P_k^3 for $0 \le k \le 3$ are given in Figure 10.3.

Given a continuous function f on $[0,1]$, define a polynomial $B_n f$ by

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

This is a linear combination of the polynomials P_k^n; and so $B_n f$ is a polynomial of degree at most n. We think of B_n as a function from the vector space $C[0,1]$ into itself. This map has several easy but important properties. If $f, g \in C[0,1]$, we say that $f \ge g$ if $f(x) \ge g(x)$ for all $0 \le x \le 1$.

10.3.1. PROPOSITION. *The map B_n is linear and monotone. That is, for all $f, g \in C[0,1]$ and $\alpha \in \mathbb{R}$,*

(1) $B_n(f+g) = B_n f + B_n g$

(2) $B_n(\alpha f) = \alpha B_n f$

(3) $B_n f \ge 0$ *if* $f \ge 0$

(4) $B_n f \ge B_n g$ *if* $f \ge g$

(5) $|B_n f| \le B_n g$ *if* $|f| \le g$.

The only part that requires any cleverness is the monotonicity. However, since each $P_k^n \ge 0$, it follows that when $f \ge 0$, then $B_n f$ is also positive. In particular, $|f| \le g$ means that $-g \le f \le g$; and hence $-B_n g \le B_n f \le B_n g$. The details are left to the reader.

Next let us compute $B_n f$ for three basic polynomials: 1, x, and x^2.

10.3.2. LEMMA. $B_n 1 = 1$, $\quad B_n x = x$, $\quad$ *and*

$$B_n x^2 = \frac{n-1}{n} x^2 + \frac{1}{n} x = x^2 + \frac{x - x^2}{n}.$$

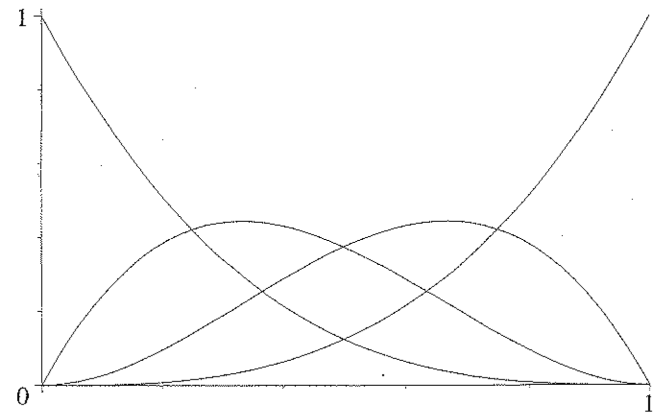

FIGURE 10.3. The Bernstein polynomials of degree 3.

PROOF. For the first equation, we use the Binomial Theorem to get

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

Next, notice that

$$\frac{k}{n}\binom{n}{k} = \frac{k}{n}\frac{n!}{k!(n-k)!} = \frac{(n-1)!}{(k-1)!(n-k)!} = \binom{n-1}{k-1}.$$

Using this result and the Binomial Theorem, we have

$$\begin{aligned} B_n x &= \sum_{k=0}^{n} \frac{k}{n}\binom{n}{k} x^k (1-x)^{n-k} \\ &= x \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1}(1-x)^{n-k} \\ &= x\big(x + (1-x)\big)^{n-1} = x. \end{aligned}$$

Finally, notice that

$$\begin{aligned} \frac{k^2}{n^2}\binom{n}{k} &= \frac{k^2}{n^2}\frac{n!}{k!(n-k)!} \\ &= \frac{(k-1)+1}{n}\frac{(n-1)!}{(k-1)!(n-k)!} \\ &= \frac{n-1}{n}\frac{k-1}{n-1}\frac{(n-1)!}{(k-1)!(n-k)!} + \frac{1}{n}\frac{(n-1)!}{(k-1)!(n-k)!} \\ &= \frac{n-1}{n}\binom{n-2}{k-2} + \frac{1}{n}\binom{n-1}{k-1}. \end{aligned}$$

Note that $\binom{m}{-2} = \binom{m}{-1} = 0$. Hence

$$\begin{aligned} B_n x^2 &= \sum_{k=0}^{n} \frac{k^2}{n^2} \binom{n}{k} x^k (1-x)^{n-k} \\ &= \frac{n-1}{n} \sum_{k=2}^{n} \binom{n-2}{k-2} x^k (1-x)^{n-k} + \frac{1}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} x^k (1-x)^{n-k} \\ &= \frac{n-1}{n} x^2 \sum_{k=2}^{n} \binom{n-2}{k-2} x^{k-2} (1-x)^{n-k} + \frac{x}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} \\ &= \frac{n-1}{n} x^2 \big(x + (1-x)\big)^{n-2} + \frac{1}{n} x \big(x + (1-x)\big)^{n-1} \\ &= \frac{n-1}{n} x^2 + \frac{1}{n} x = x^2 + \frac{x - x^2}{n}. \end{aligned}$$

■

PROOF OF WEIERSTRASS'S THEOREM. By Exercise 10.2.A, it suffices to prove the theorem for the interval $[0, 1]$. Fix a continuous function f in $C[0, 1]$. We will prove that for each $\varepsilon > 0$, there is some $N > 0$ so that

$$\|f(x) - B_n f(x)\| < \varepsilon \quad \text{for all} \quad n \geq N.$$

Since $[0, 1]$ is compact, f is uniformly continuous on $[0, 1]$ by Theorem 5.5.9. Thus for our given $\varepsilon > 0$, there is some $\delta > 0$ so that

$$|f(x) - f(y)| \leq \frac{\varepsilon}{2} \quad \text{for all} \quad |x - y| \leq \delta, \;\; x, y \in [0, 1].$$

Also, f is bounded on $[0, 1]$ by the Extreme Value Theorem (Theorem 5.4.4). So let

$$M = \|f\|_\infty = \sup_{x \in [0,1]} |f(x)|.$$

Fix any point $a \in [0, 1]$. We claim that

$$|f(x) - f(a)| \leq \frac{\varepsilon}{2} + \frac{2M}{\delta^2} (x - a)^2.$$

Indeed, if $|x - a| \leq \delta$, then

$$|f(x) - f(a)| \leq \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{2M}{\delta^2} (x - a)^2$$

by our estimate of uniform continuity. And if $|x - a| \geq \delta$, then

$$|f(x) - f(a)| \leq 2M \leq 2M \Big(\frac{x - a}{\delta}\Big)^2 \leq \frac{\varepsilon}{2} + \frac{2M}{\delta^2} (x - a)^2.$$

Notice that by linearity and the fact that $B_n 1 = 1$, we obtain

$$B_n(f - f(a))(x) = B_n f(x) - f(a).$$

Now use the positivity of our map B_n to obtain

$$\begin{aligned}|B_nf(x) - f(a)| &\leq B_n\Big(\frac{\varepsilon}{2} + \frac{2M}{\delta^2}(x-a)^2\Big)\\ &= \frac{\varepsilon}{2} + \frac{2M}{\delta^2}\Big(x^2 + \frac{x-x^2}{n} - 2ax + a^2\Big)\\ &= \frac{\varepsilon}{2} + \frac{2M}{\delta^2}(x-a)^2 + \frac{2M}{\delta^2}\frac{x-x^2}{n}.\end{aligned}$$

Evaluate this at $x = a$ to obtain

$$|B_nf(a) - f(a)| \leq \frac{\varepsilon}{2} + \frac{2M}{\delta^2}\,\frac{a-a^2}{n} \leq \frac{\varepsilon}{2} + \frac{M}{2\delta^2 n}.$$

We use the fact that $\max\{a - a^2 : 0 \leq a \leq 1\} = \frac{1}{4}$.

This estimate does not depend on the point a. So we have found

$$\|B_nf - f\|_\infty \leq \frac{\varepsilon}{2} + \frac{M}{2\delta^2 n}.$$

So now choose $N \geq \dfrac{M}{\delta^2\varepsilon}$ so that $\dfrac{M}{2\delta^2 N} < \dfrac{\varepsilon}{2}$. Then for all $n \geq N$,

$$\|B_nf - f\|_\infty \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$ ■

As was already mentioned, using Bernstein polynomials is not an efficient way of finding polynomial approximations. However, Bernstein polynomials have other advantages, which are developed in the Exercises.