36
The Weierstrass Approximation Theorem

Recall that the fundamental idea underlying the construction of the real numbers is approximation by the simpler rational numbers. Firstly, numbers are often determined as the unknown roots of some equation and when we cannot solve the equation explicitly, as is most often the case, then we must compute approximate solutions. But even if we write down a real number symbolically, like $\sqrt{2}$, for example, we cannot specify its numerical value completely in general. In this case, we approximate the real number to any desired accuracy using rational numbers with finite decimal expansions.

The situation for functions is completely analogous. In general, functions that are specified as the solutions of differential equations cannot be written down explicitly in terms of known functions. Instead, we must look for good approximations. Moreover, most of the functions that we can write down, i.e., those involving exp, log, sin, and so on, are "complicated" in the sense that they take on real values that cannot be written down explicitly. To use these functions in practical computations, we must resort to using good approximations of their values. Put it this way; when we press the e^x key on a calculator, we do not get e^x, rather we get a good approximation.

This raises one of the fundamental problems of analysis, which is figuring out how to approximate a given function using simpler functions. In this chapter, we begin the study of this problem by proving a fundamental result which says that any continuous function can be approximated arbitrarily well by polynomials. This is an important result because polynomials are relatively simple. In particular, a polynomial is specified completely by a finite set of coefficients. In other words, the relatively simple polynomials play the same role with respect to continuous functions that rational numbers play with real numbers.

The result is due to Weierstrass and it states:

Theorem 36.1 Weierstrass Approximation Theorem *Assume that f is continuous on a closed bounded interval I. Given any $\epsilon > 0$, there is a polynomial P_n with sufficiently high degree n such that*

$$|f(x) - P_n(x)| < \epsilon \text{ for } a \le x \le b. \tag{36.1}$$

There are many different proofs of this result, but in keeping with our constructivist tendencies, we present a constructive proof based on Bernstein[1] polynomials. The motivation for this approach rests in probability theory. We do not have space in this book to develop probability theory, but we describe the connection in an intuitive way. Later in Chapter 37 and Chapter 38, we investigate other polynomial approximations of functions that arise from different considerations.

Before beginning, we note that it suffices to prove Theorem 36.1 for the interval $[0, 1]$. The reason is that the arbitrary interval $a \le y \le b$ is mapped to $0 \le x \le 1$ by $x = (a-y)/(a-b)$ and vice versa by $y = (b-a)x + a$. If g is continuous on $[a, b]$, then $f(x) = g((b-a)x + a)$ is continuous on $[0, 1]$. If the polynomial P_n of degree n approximates f to within ϵ on $[0, 1]$, then the polynomial $\tilde{P}_n(y) = P_n((a-y)/(a-b))$ of degree n approximates $g(y)$ to within ϵ on $[a, b]$.

36.1 The Binomial Expansion

One ingredient needed to construct the polynomial approximations is an important formula called the binomial expansion. For natural numbers $0 \le m \le n$, we define the **binomial coefficient** $\binom{n}{m}$, or n **choose** m, by

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

Example 36.1.

$$\binom{4}{2} = \frac{4!}{2!2!} = 6, \quad \binom{6}{1} = \frac{6!}{1!5!} = 6, \quad \binom{3}{0} = \frac{3!}{3!0!} = 1$$

We can interpret n choose m as the number of distinct subsets with m elements that can be chosen from a set of n objects, or the number of combinations of n objects taken m at a time.

---

[1] The Russian mathematician Sergi Natanovich Bernstein (1880–1968) studied in France before returning to Russia to work. He proved significant results in approximation theory and probability.

EXAMPLE 36.2. We compute the probability $\mathcal{P}$ of getting an ace of diamonds in a poker hand of 5 cards chosen at random from a standard deck of 52 cards. Recall the formula

$$\begin{aligned}\mathcal{P}(\text{ event }) &= \text{probability of an event}\\ &= \frac{\text{number of outcomes in the event}}{\text{total number of possible outcomes}}\end{aligned}$$

that holds if all outcomes are equally likely. The total number of 5 card poker hands is $\binom{52}{5}$. Obtaining a "good" hand amounts to choosing any 4 cards from the remaining 51 cards after getting an ace of diamonds. So there are $\binom{51}{4}$ good hands. This means

$$\mathcal{P} = \frac{\binom{51}{4}}{\binom{52}{5}} = \frac{51!}{4!47!}\,\frac{5!47!}{52!} = \frac{5}{52}.$$

It is straightforward (Problem 36.3) to show the following identities,

$$\binom{n}{m} = \binom{n}{n-m}, \quad \binom{n}{1} = \binom{n}{n-1}, \quad \binom{n}{n} = \binom{n}{0} = 1. \tag{36.2}$$

An important application of the binomial coefficient is the following theorem.

Theorem 36.2 Binomial Expansion *For any natural number n,*

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m}. \tag{36.3}$$

EXAMPLE 36.3.

$$\begin{aligned}(a+b)^2 &= a^2 + 2ab + b^2\\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3\\ (a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\end{aligned}$$

The proof is by induction. For $n = 1$,

$$(a+b)^1 = a + b = \binom{1}{0}a + \binom{1}{1}b.$$

We assume the formula is true for $n-1$, so that

$$(a+b)^{n-1} = \sum_{m=0}^{n-1} \binom{n-1}{m} a^m b^{n-1-m},$$

and prove it holds for n.

We multiply out

$$\begin{aligned}(a+b)^n &= (a+b)(a+b)^{n-1} \\ &= \sum_{m=0}^{n-1} \binom{n-1}{m} a^{m+1} b^{n-1-m} + \sum_{m=0}^{n-1} \binom{n-1}{m} a^m b^{n-m}.\end{aligned}$$

Now changing variables in the sum,

$$\sum_{m=0}^{n-1} \binom{n-1}{m} a^{m+1} b^{n-1-m} = \sum_{m=1}^{n-1} \binom{n-1}{m-1} a^m b^{n-m} + a^n b^0,$$

while

$$\sum_{m=0}^{n-1} \binom{n-1}{m} a^m b^{n-m} = a^0 b^n + \sum_{m=1}^{n-1} \binom{n-1}{m} a^m b^{n-m}.$$

Hence,

$$(a+b)^n = a^0 b^n + \sum_{m=1}^{n-1} \left(\binom{n-1}{m-1} + \binom{n-1}{m} \right) a^m b^{n-m} + a^n b^0. \quad (36.4)$$

It is a good exercise (Problem 36.5) to show that

$$\binom{n-1}{m-1} + \binom{n-1}{m} = \binom{n}{m}. \quad (36.5)$$

Using this in (36.4) proves the result.

We use the binomial expansion to drive two other useful formulas. We differentiate both sides of

$$(x+b)^n = \sum_{m=0}^{n} \binom{n}{m} x^m b^{n-m} \quad (36.6)$$

to get

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

Setting $x = a$ and multiplying through by a/n,

$$a(a+b)^{n-1} = \sum_{m=0}^{n} \frac{m}{n} \binom{n}{m} a^m b^{n-m}. \quad (36.7)$$

Differentiating (36.6) twice (Problem 36.6) gives

$$\left(1 - \frac{1}{n}\right) a^2 (a+b)^{n-2} = \sum_{m=0}^{n} \left(\frac{m^2}{n^2} - \frac{m}{n^2} \right) \binom{n}{m} a^m b^{n-m}. \quad (36.8)$$

36.2 The Law of Large Numbers

The approximating polynomials used to prove Theorem 36.1 are constructed by taking linear combinations of more elementary polynomials called binomial polynomials. In this section, we explore the properties of the binomial polynomials and their connection to probability.

We set $a = x$ and $b = 1 - x$ in the binomial expansion (36.3) to get

$$1 = (x + (1-x))^n = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m}. \tag{36.9}$$

We define the $m+1$ **binomial polynomials** of degree n as the terms in the expansion, so

$$p_{n,m}(x) = \binom{n}{m} x^m (1-x)^{n-m}, \quad m = 0, 1, \cdots, n.$$

Example 36.4.

$$p_{2,0}(x) = \binom{2}{0} x^0 (1-x)^2 = (1-x)^2$$
$$p_{2,1}(x) = \binom{2}{1} x^1 (1-x)^1 = 2x(1-x)$$
$$p_{2,2}(x) = \binom{2}{2} x^2 (1-x)^0 = x^2$$

If $0 \le x \le 1$ is the probability of an event E, then $p_{n,m}(x)$ is the probability that E occurs exactly m times in n independent trials.

Example 36.5. In particular, consider tossing an coin with probability x that a head (H) occurs and, correspondingly, probability $1 - x$ that a tail (T) occurs. The coin is "unfair" if $x \ne 1/2$. The probability of the occurrence of a particular sequence of n tosses containing m heads, e.g.,

$$\underbrace{HTTHHTHTHTHTTHHHTHTHTTT \cdots T}_{m \text{ heads in } n \text{ tosses}},$$

is $x^m (1-x)^{m-n}$ by the multiplication rule for probabilities. There are $\binom{n}{m}$ sequences of n tosses with exactly m heads. By the addition rule for probabilities, $p_{n,m}(x)$ is the probability of getting exactly m heads in n tosses.

The binomial polynomials have several useful properties, some of which follow directly from the connection to probability. For example, we interpret

$$\sum_{m=0}^{n} p_{n,m}(x) = 1 \tag{36.10}$$

as saying that event E with probability x occurs either exactly 0, 1, $\cdots$, or n times in n independent trials with probability 1. Since $p_{n,m}(x) \geq 0$ for $0 \leq x \leq 1$, (36.10) implies that $0 \leq p_{n,m}(x) \leq 1$ for $0 \leq x \leq 1$, as it must since it is a probability.

A couple more useful properties: (36.7) implies

$$\sum_{m=0}^{n} m p_{n,m}(x) = nx \tag{36.11}$$

and (36.8) implies

$$\sum_{m=0}^{n} m^2 p_{n,m}(x) = (n^2 - n)x^2 + nx. \tag{36.12}$$

An important use of the binomial polynomials is an application to the Law of Large Numbers. Suppose we have an event E that has probability x of occurring, such as the unfair coin from Example 36.5. But suppose we don't know the probability. How might we determine x? If we conduct a single trial, e.g., flip the coin once, we might see event E or might not. One trial does not give much information for determining x. However, if we conduct a large number $n \gg 1$ of trials, then intuition suggests that E should occur approximately nx times out of n trials, at least "most of the time."

Example 36.6. The connection between the probability of occurrence in one trial and the frequency of occurrence in many trials is not completely straightforward to determine. Consider coin tossing again. If we flip a fair coin $100,000$ times, we expect to see *around* $50,000$ heads *most of the time*. Of course, we could be very unlucky and get all tails. But the probability of this occurring is

$$\left(\frac{1}{2}\right)^{100000} \approx 10^{-30103}.$$

On the other hand, it is also unlikely that we will see heads in exactly half of the tosses. In fact, one can show that the probability of getting heads exactly half of the time is approximately $1/\sqrt{\pi n}$ for n large, and therefore also goes to zero as n increases.

A Law of Large Numbers encapsulates in some way the intuitive connection between the probability of an event occurring in one trial and the frequency that the event occurs in a large number of trials. A mathematical expression of this intuition is a little tricky to state, however, as we saw in Example 36.6. We prove the following version that is originally due to Jacob Bernoulli.

Theorem 36.3 Law of Large Numbers *Assume that event E occurs with probability x and let m denote the number of times E occurs in n trials. Let $\epsilon > 0$ and $\delta > 0$ be given. The probability that m/n differs from x by less than δ is greater than $1-\epsilon$, i.e.,*

$$\mathcal{P}\left(\left|\frac{m}{n} - x\right| < \delta\right) > 1 - \epsilon, \tag{36.13}$$

for all n sufficiently large.

Note that we can choose $\epsilon > 0$ and $\delta > 0$ arbitrarily small at the cost of making n possibly very large, hence the name of the theorem. Also note that while this result says that it is likely that event E will occur approximately xn times in n trials, *it does not say that event E will occur exactly xn times in n trials nor does it say that event E must occur approximately xn times in n trials*. Thus, this result does not contradict the computations in Example 36.6.

Phrased in terms of the binomial polynomials, we want to show that given ϵ, $\delta > 0$,

$$\sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|<\delta}} p_{n,m}(x) > 1 - \epsilon \tag{36.14}$$

for n sufficiently large.

Consider the complementary sum

$$\sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|\ge\delta}} p_{n,m}(x) = 1 - \sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|<\delta}} p_{n,m}(x),$$

which we estimate simply as

$$\sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|\ge\delta}} p_{n,m}(x) \le \frac{1}{\delta^2} \sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|\ge\delta}} \left(\frac{m}{n} - x\right)^2 p_{n,m}(x) \le \frac{1}{n^2\delta^2} S_n$$

where

$$\begin{aligned} S_n &= \sum_{m=0}^{n} (m - nx)^2 p_{n,m}(x) \\ &= \sum_{m=0}^{n} m^2 p_{n,m}(x) - 2nx \sum_{m=0}^{n} m p_{n,m}(x) + n^2x^2 \sum_{m=0}^{n} p_{n,m}(x). \end{aligned} \tag{36.15}$$

Using (36.10), (36.11), and (36.12), we find S_n simplifies (Problem 36.9) to $S_n = nx(1-x)$. Since $x(1-x) \le 1/4$ for $0 \le x \le 1$, $S_n \le n/4$. Therefore,

$$\sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|\ge\delta}} p_{n,m}(x) \le \frac{1}{4n\delta^2} \tag{36.16}$$

and

$$\sum_{\substack{0\le m\le n\\ \left|\frac{m}{n}-x\right|<\delta}} p_{n,m}(x) \ge 1-\frac{1}{4n\delta^2}.$$

In particular, for fixed ϵ, $\delta > 0$, we can insure that $(4n\delta^2)^{-1} < \epsilon$ by choosing $n > 1/(4\delta^2\epsilon)$.

36.3 The Modulus of Continuity

In order to prove a strong version of Theorem 36.1, we introduce a useful generalization of Lipschitz continuity.

First note that by Theorem 32.11, the continuous function f on $[a, b]$ in Theorem 36.1 is actually uniformly continuous on $[a, b]$. That is given $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x)-f(y)| < \epsilon$ for all x, y in $[a, b]$ with $|x-y| < \delta$.[2] Now a Lipschitz continuous function f with constant L is uniformly continuous because $|f(x) - f(y)| \le L|x - y| < \epsilon$ for all x, y with $|x - y| < \delta = \epsilon/L$. On the other hand, uniformly continuous functions are not necessarily Lipschitz continuous. They do, however, satisfy a generalization of the condition that defines Lipschitz continuity called the modulus of continuity.

The generalization is based on the observation that if f is uniformly continuous on a closed, bounded interval $I = [a, b]$, then for any $\delta > 0$, the set of numbers

$$\{|f(x) - f(y)| \text{ with } x, y \text{ in } I,\ |x - y| < \delta\} \tag{36.17}$$

is bounded. Otherwise, f could not be uniformly continuous (Problem 36.10). But, Theorem 32.15 then implies that the set of numbers (36.17) has a least upper bound. Turning this around, we define the **modulus of continuity** $\omega(f, \delta)$ of a general function f on a general interval I by

$$\omega(f,\delta) = \sup_{\substack{x,y \text{ in } I\\ |x-y|<\delta}} \{|f(x) - f(y)|\}.$$

Note that $\omega(f, \delta) = \infty$ if the set (36.17) is not bounded. We can guarantee that $\omega(f, \delta)$ is finite if f is uniformly continuous and I is a closed interval, but if f is not uniformly continuous and/or I is open or unbounded, then $\omega(f, \delta)$ might be infinite.

EXAMPLE 36.7. We know x^2 is uniformly continuous on $[0, 1]$. Now consider the difference $|x^2 - y^2| = |x - y|\,|x + y|$, where $|x - y| < \delta$.

[2] Uniformity refers to the fact that δ can be chosen independently of x and y.

The values of $|x-y|$ increases monotonically from 0 to δ, while the corresponding largest values of $|x+y|$ decrease monotonically from 2 to $2-\delta$. The largest value of their product occurs when $|x-y|=\delta$ so that $\omega(x^2,\delta)=2\delta-\delta^2$.

EXAMPLE 36.8. $\omega(x^{-1},\delta)$ on $(0,1)$ is infinite.

EXAMPLE 36.9. $\omega(\sin(x^{-1}),\delta)=2$ on $(0,1)$ since for any $\delta>0$ we can find x and y within δ of 0, and hence within δ of each other, such that $\sin(x^{-1})=1$ and $\sin(y^{-1})=-1$.

Note that the functions in Example 36.8 and Example 36.9 are not uniformly continuous on the indicated intervals. In fact, if f is uniformly continuous on $[a,b]$, then $\omega(f,\delta)\to 0$ as $\delta\to 0$ (Problem 36.14).

If f is Lipschitz continuous on $[a,b]$ with constant L, then $\omega(f,\delta)\leq L\delta$. In this sense, the modulus of continuity is a generalization of the idea of Lipschitz continuity.

36.4 The Bernstein Polynomials

To construct the approximating polynomial, we partition $[0,1]$ by a uniform mesh with $n+1$ nodes

$$x_m=\frac{m}{n},\quad m=0,\cdots,n.$$

The **Bernstein polynomial** of degree n for f on $[0,1]$ is

$$B_n(f,x)=B_n(x)=\sum_{m=0}^{n} f(x_m)p_{n,m}(x). \tag{36.18}$$

Note that the degree of B_n is at most n.

The reason that the Bernstein polynomials become increasingly accurate approximations as the degree n increases is rather intuitive. The formula for $B_n(x)$ decomposes into two sums,

$$B_n(x)=\sum_{x_m\approx x} f(x_m)p_{n,m}(x)+\sum_{|x_m-x|\text{ large}} f(x_m)p_{n,m}(x).$$

The first sum converges to $f(x)$ as n becomes large, since we can find nodes $x_m=m/n$ arbitrarily close to x by taking n large.[3] The second sum converges to zero by the Law of Large Numbers. This is exactly what we prove below.

Before stating a convergence result, we consider a couple of examples.

[3]Recall that any real number can be approximated arbitrarily well by rational numbers.

EXAMPLE 36.10. The Bernstein polynomial B_n for x^2 on $[0,1]$ with $n \geq 2$ is given by

$$B_n(x) = \sum_{m=0}^{n} \left(\frac{m}{n}\right)^2 p_{n,m}(x).$$

By (36.12), this means

$$B_n(x) = \left(1 - \frac{1}{n}\right) x^2 + \frac{1}{n} x = x^2 + \frac{1}{n} x(1-x).$$

We see that $B_n(x^2, x) \neq x^2$ and in fact the error

$$|x^2 - B_n(x)| = \frac{1}{n} x(1-x)$$

decreases like $1/n$ as n increases.

EXAMPLE 36.11. We compute B_1, B_2, and B_3 for $f(x) = e^x$ on $[0,1]$,

$$B_1(x) = e^0(1-x) + e^1 x = (1-x) + ex$$
$$B_2(x) = (1-x)^2 + 2e^{1/2} x(1-x) + ex^2$$
$$B_3(x) = (1-x)^3 + 3e^{1/2} x(1-x)^2 + 3e^{2/3} x^2(1-x) + ex^3.$$

We plot these functions in Fig. 36.1.

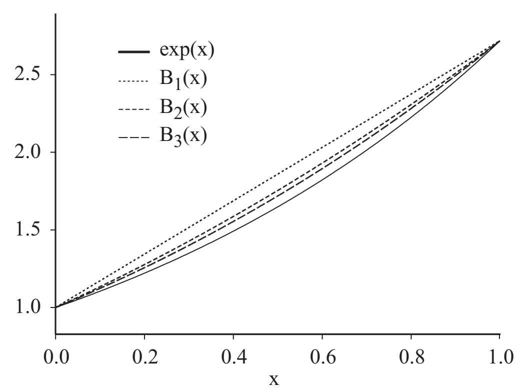

FIGURE 36.1. The first three Bernstein polynomials for e^x.

We prove:

Theorem 36.4 Bernstein Approximation Theorem *Let f be a continuous function on $[0,1]$ and $n \geq 1$ a natural number. Then*

$$|f(x) - B_n(f,x)| \leq \frac{9}{4} \omega(f, n^{-1/2}). \tag{36.19}$$

If f is Lipschitz continuous with constant L, then

$$|f(x) - B_n(f,x)| \le \frac{9}{4} L n^{-1/2}. \tag{36.20}$$

Theorem 36.1 follows immediately since for $\epsilon > 0$, we simply choose n sufficiently large so that

$$|f(x) - B_n(f,x)| \le \frac{9}{4}\omega(f, n^{-1/2}) < \epsilon.$$

Using (36.10), we write the error as a sum involving the differences between $f(x)$ and the values of f at the nodes:

$$\begin{aligned} f(x) - B_n(x) &= \sum_{m=0}^{n} f(x) p_{n,m}(x) - \sum_{m=0}^{n} f(x_m) p_{n,m}(x) \\ &= \sum_{m=0}^{n} (f(x) - f(x_m)) p_{n,m}(x) \end{aligned}$$

We expect that the differences $f(x) - f(x_m)$ should be small when x is close to x_m by the continuity of f. To take advantage of this, for $\delta > 0$, we split the sum into two parts

$$\begin{aligned} f(x) - B_n(x) = &\sum_{\substack{0 \le m \le n \\ |x - x_m| < \delta}} (f(x) - f(x_m)) p_{n,m}(x) \\ &+ \sum_{\substack{0 \le m \le n \\ |x - x_m| \ge \delta}} (f(x) - f(x_m)) p_{n,m}(x). \end{aligned} \tag{36.21}$$

The first sum is small by the continuity of f, since

$$\begin{aligned} \left| \sum_{\substack{0 \le m \le n \\ |x - x_m| < \delta}} (f(x) - f(x_m)) p_{n,m}(x) \right| &\le \sum_{\substack{0 \le m \le n \\ |x - x_m| < \delta}} |f(x) - f(x_m)| p_{n,m}(x) \\ &\le \omega(f, \delta) \sum_{\substack{0 \le m \le n \\ |x - x_m| < \delta}}^{n} p_{n,m}(x) \\ &\le \omega(f, \delta) \sum_{m=0}^{n} p_{n,m}(x) = \omega(f, \delta). \end{aligned}$$

We can get a crude bound on the second sum in (36.21) easily. Since f is continuous on $[0,1]$ there is a constant C such that $|f(x)| \le C$ for $0 \le x \le 1$. Therefore,

$$\sum_{\substack{0 \le m \le n \\ |x - x_m| \ge \delta}} (f(x) - f(x_m)) p_{n,m}(x) \le 2C \sum_{\substack{0 \le m \le n \\ |x - x_m| \ge \delta}} p_{n,m}(x) \le \frac{C}{n\delta^2}$$

by (36.16). So we can make the second sum as small as desired by taking n large.

To get a sharper estimate on the second sum in (36.21), we use a trick similar to that used to prove Theorem 19.1. We let M be the largest integer less than or equal to $|x - x_m|/\delta$ and choose M uniformly spaced points $y_1, y_2, \cdots, y_M$ in the interval spanned by x and x_m so that each of the resulting $M+1$ intervals have length $|x - x_m|/(M+1) < \delta$.

Now, we can write

$$f(x) - f(x_m) = (f(x) - f(y_1)) + (f(y_1) - f(y_2)) + \cdots \\ + (f(y_M) - f(x_m)).$$

Therefore,

$$|f(x) - f(x_m)| \le (M+1)\omega(f,\delta) \le \left(1 + \frac{|x - x_m|}{\delta}\right)\omega(f,\delta).$$

We use this to estimate the second sum in (36.21),

$$\begin{aligned}\Bigg|\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} (f(x) - f(x_m))p_{n,m}(x)\Bigg| \\ \le \omega(f,\delta)\Bigg(\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} p_{n,m}(x) + \frac{1}{\delta}\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} |x - x_m|p_{n,m}(x)\Bigg).\end{aligned}$$

Using the fact that $|x - x_m|/\delta = M \ge 1$,

$$\begin{aligned}&\Bigg|\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} (f(x) - f(x_m))p_{n,m}(x)\Bigg| \\ &\quad\le \omega(f,\delta)\Bigg(\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} p_{n,m}(x) + \frac{1}{\delta^2}\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} (x - x_m)^2 p_{n,m}(x)\Bigg) \\ &\quad\le \omega(f,\delta)\Bigg(\sum_{m=0}^{n} p_{n,m}(x) + \frac{1}{\delta^2}\sum_{m=0}^{n}(x - x_m)^2 p_{n,m}(x)\Bigg) \\ &\quad\le \omega(f,\delta)\left(1 + \frac{1}{4n\delta^2}\right)\end{aligned}$$

by (36.11) and (36.12). So

$$\Bigg|\sum_{\substack{0\le m\le n\\|x-x_m|\ge\delta}} (f(x) - f(x_m))p_{n,m}(x)\Bigg| \le \omega(f,\delta)\left(1 + \frac{1}{4n\delta^2}\right).$$

Putting the estimates on the sums back into (36.21),

$$|f(x) - B_n(x)| \le \omega(f, \delta) \left(2 + \frac{1}{4n\delta^2}\right).$$

Setting $\delta = n^{-1/2}$ proves the theorem.

36.5 Accuracy and Convergence

We can interpret Theorem 36.4 as saying that the Bernstein polynomials $\{B_n(f,x)\}$ converge uniformly to $f(x)$ on $[0,1]$ as $n \to \infty$. In other words, the errors of the Bernstein polynomials B_n for a given function f on $[0,1]$ tend to zero as n increases. This is a strong property; unfortunately, the price is that the convergence is very slow in general.

EXAMPLE 36.12. To demonstrate how slowly the Bernstein polynomials can converge, we plot the Bernstein polynomial of degree 4 for $\sin(\pi x)$ on $[0,1]$ in Fig. 36.2.

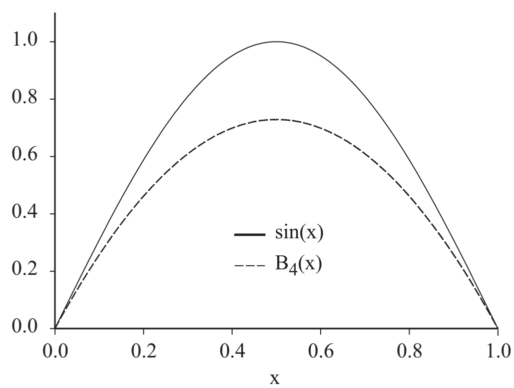

FIGURE 36.2. A plot of the Bernstein polynomial $B_4(x)$ for $\sin(\pi x)$.

If the error bound in (36.19) is accurate, i.e.,

$$|f(x) - B_n(x)| \approx \frac{9}{4}\omega(f, n^{-1/2}) \approx Cn^{-1/2} \text{ for some constant } C,$$

then we have to increase n by a factor of 100 in order to see an improvement of 10 (one additional digit of accuracy) in the error. This follows because from the computation

$$\frac{|f(x) - B_{n_1}(x)|}{|f(x) - B_{n_2}(x)|} \approx \frac{n_1^{-1/2}}{n_2^{-1/2}} = 10^{-1}$$

we need $n_2 = 100n_1$.

The error can decrease more quickly in some cases. Above, we saw that the error for x^2 decreases like $1/n$. But even this is relatively slow compared to some other polynomial approximations and for this reason the Bernstein polynomials are not often encountered in practice.

36.6 Unanswered Questions

We have shown that continuous functions can be approximated by polynomials. But we have not really explained why polynomials are well-suited for approximating functions. In other words, what are the properties of polynomials that make them good approximations? Are there other sets of functions that have similar approximation properties? Atkinson [2], Isaacson and Keller [15], and Rudin [19] have interesting material on these topics.

Chapter 36 Problems

36.1. Evaluate $\binom{8}{3}$.

36.2. Explain the claim that $\binom{n}{m}$ gives the number of ways that n objects can be arranged in groups of m.

36.3. Prove (36.2).

36.4. Expand $(a+b)^6$.

36.5. Prove (36.5).

36.6. Prove (36.8).

36.7. Verify (36.12).

36.8. Determine a formula for the probability of getting exactly $n/2$ heads when tossing a fair coin n times, where n is even. Make a plot of the formula for a n in the range of 1 to 100 and test the claim that it approaches $\sqrt{\pi n}$ for n large.

36.9. Prove that S_n defined in (36.15) is equal to $S_n = nx(1-x)$.

Problems 36.10–36.15 have to do with the modulus of continuity. Several of the proofs in this book could be generalized by using the modulus of continuity instead of Lipschitz continuity.

36.10. Prove that if f is uniformly continuous on $[a,b]$, then for any $\delta > 0$ the set of numbers (36.17) is bounded.

36.11. Evaluate

(a) $\omega(x^2,\delta)$ on $[0,2]$ (b) $\omega(1/x,\delta)$ on $[1,2]$ (b) $\omega(\log(x),\delta)$ on $[1,2]$.

36.12. Verify Example 36.8.

36.13. Verify Example 36.9.

36.14. Prove that if f is uniformly continuous on $[a,b]$, then $\omega(f,\delta) \to 0$ as $\delta \to 0$.

36.15. Prove that if f has a continuous derivative on $[a,b]$, then $\omega(f,\delta) \leq \max_{[a,b]} |f'|\delta$.

Computing Bernstein polynomial approximations can be tedious. You might want to use MAPLE© *, for example, to do Problems 36.16–36.21.*

36.16. Compute formulas for $p_{3,m}$, $m = 0, 1, 2, 3$.

36.17. Verify the computations in Example 36.11.

36.18. Compute the Bernstein polynomials for x on $[0,1]$.

36.19. Compute and plot the Bernstein polynomials for $\exp(x)$ on $[1,3]$ of degree 1, 2, and 3.

36.20. (a) Compute a summation formula for the Bernstein polynomial for x^3 on $[0,1]$ for degree ≥ 3. (b) Find an explicit formula for the Bernstein polynomial from (a) that does not involve summation. (c) Write down a formula for the error.

36.21. Compute and plot the Bernstein polynomials for $\sin(\pi x)$ on $[0,1]$ of degree 1, 2, 3, and 4.

We have shown that the Bernstein polynomials approximate a differentiable function, which is continuous of course, uniformly well. In Problem 36.22, we ask you to show that the derivative of the function is also approximated by the derivatives of the function's Bernstein polynomials.

36.22. If $f(x)$ has a continuous first derivative in $[0,1]$, prove that the derivatives of the Bernstein polynomials $\{P_n'(f,x)\}$ converge uniformly to $f'(x)$ on $[0,1]$.
Hint: First, verify the formulas

$$p_{n,m}' = n(p_{n-1,m-1} - p_{n-1,m}) \text{ for } m = 1, \cdots, m-1$$
$$p_{n,n}' = np_{n-1,n-1}, \quad p_{n,0}' = -np_{n-1,0}.$$

Then find a summation formula for the error $f'(x) - P_n'(x)$ and rearrange the sum in terms of $p_{n-1,m}$ for $m = 0, 1, \cdots, n-1$.

36.23. If f is continuous on $[0,1]$ and if

$$\int_0^1 f(x)x^n\,dx = 0 \text{ for } n = 0, 1, 2, \cdots,$$

, then prove that $f(x) = 0$ for $0 \leq x \leq 1$. *Hint:* This says that the integral of the product of f and *any* polynomial is zero. Use Theorem 36.1 to first prove that

$$\int_0^1 f^2(x)\,dx = 0.$$

We say that the real numbers $\mathbb{R}$ *are* **separable** *because any real number can be approximated to arbitrary accuracy by a rational number. The analogous property holds for the space of continuous functions on a closed, bounded interval, which is the content of the theorem we ask you to prove in Problem 36.24.*

36.24. Prove the following extension of the Weierstrass Approximation Theorem:

Theorem 36.5 *Assume that f is continuous on a closed bounded interval I. Given any $\epsilon > 0$, there is a polynomial P_n with rational coefficients with finite decimal expansions and of sufficiently high degree n such that*

$$|f(x) - P_n(x)| < \epsilon \text{ for } a \leq x \leq b.$$

Hint: Use Theorem 36.1 to first get an approximate polynomial and then analyze the effect of replacing its coefficients by rational approximations.

http://www.springer.com/978-0-387-95484-4

Practical Analysis in One Variable
Estep, D.
2002, XX, 623 p., Hardcover
ISBN: 978-0-387-95484-4