COIN TOSSING, PROBABILITY, AND THE WEIERSTRASS APPROXIMATION THEOREM

ROBERT G. KULLER, University of Colorado

1. Introduction. The purpose of this article is to give a full account of the details and the motivation that are behind Bernstein's proof [1] of Weierstrass' famous approximation theorem. Although these matters are not beyond the reach of the undergraduate student in an advanced calculus course, the author knows of no easily accessible treatment that fully exposes the idea of the proof.

2. Statement of the theorem. The theorem of Weierstrass states: *Every function* $x(t)$, *continuous on an interval* $a \leqq t \leqq b$, *can be arbitrarily closely and uniformly approximated by polynomials.*

More precisely, for any $\epsilon > 0$ there is a polynomial $P(t)$ such that $|x(t) - P(t)| \leqq \epsilon$ for all t satisfying $a \leqq t \leqq b$.

Since the correspondence $x(t) \leftrightarrow x[a + t(b-a)]$ demonstrates the essential equivalence of functions on $a \leqq t \leqq b$ and functions on $0 \leqq t \leqq 1$ (geometrically, the graph of one is obtained from the graph of the other by the reversible operations of translation and stretching), we do not lose generality if we give a proof of the Weierstrass theorem for the interval $0 \leqq t \leqq 1$ only.

3. Motivation for the proof. Imagine a loaded coin which turns up "heads" with probability t, $0 \leqq t \leqq 1$. In a game consisting of n consecutive tosses, the probability that exactly k "heads" occur is given by

$$\frac{n!}{k!(n-k)!} t^k (1-t)^{n-k}, \qquad (k = 0, 1, 2, \cdots n).$$

Now suppose that the continuous function $x(t)$ assigns a value to the game as follows: the game shall pay off $x(\frac{k}{n})$ dollars if exactly k of the n tosses are "heads." We obtain for the "expected value" E_n (mean value) of one play consisting of n tosses

$$E_n = \sum_{k=0}^{n} V_k P_k,$$

where V_k = pay-off for k heads among n tosses, and P_k = probability of outcome k heads among n tosses. Since

$$V_k = x\left(\frac{k}{n}\right) \quad \text{and} \quad P_k = \frac{n!}{k!(n-k)!} t^k (1-t)^{n-k},$$

we obtain

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

for the average value of one play of the game.

Now, if n is very large, it is reasonable to expect that head will come up nearly nt times, since t is the probability for head. This means that $x(tn/n)=x(t)$ and $E_n(t)$ will be nearly the same for large values of n, i.e., we expect the difference $|x(t)-E_n(t)|$ to tend to zero as n increases to infinity. But $E_n(t)$ is a polynomial of nth order in t, so what we have established by a heuristic argument is already Weierstrass' approximation theorem, which we will now proceed to prove by essentially standard methods.

4. A start on the proof. We begin by attempting to estimate the size of the quantity $|x(t)-E_n(t)|$. Using the identity

$$1 = 1^n = [t + (1-t)]^n = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k}$$

we can write

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

and

$$\begin{aligned} |x(t) - E_n(t)| &= \left| \sum_{k=0}^{n} \left[x(t) - x\left(\frac{k}{n}\right)\right] \binom{n}{k} t^k (1-t)^{n-k} \right| \\ &\leqq \sum_{k=0}^{n} \left| x(t) - x\left(\frac{k}{n}\right)\right| \binom{n}{k} t^k (1-t)^{n-k}. \end{aligned}$$

We will now utilize the continuity of $x(t)$, and the fact that a function which is continuous on a closed interval is uniformly continuous there, i.e., for any $\epsilon>0$ there is a $\delta(\epsilon)$ such that $|x(t)-x(t')|\leqq\epsilon$ whenever $|t-t'|\leqq\delta(\epsilon)$. In view of this it is natural to divide the values of k into two groups,

$$S_1 = \text{all values of } k \text{ satisfying} \left| t - \frac{k}{n} \right| \leqq \delta(\epsilon)$$

and

$$S_2 = \text{all values of } k \text{ satisfying} \left| t - \frac{k}{n} \right| > \delta(\epsilon).$$

Clearly, $S_1 \cap S_2 = 0$. Then

$$\begin{aligned} |x(t) - E_n(t)| &\leqq \sum_{k\in S_1} \left| x(t) - x\left(\frac{k}{n}\right)\right| \binom{n}{k} t^k (1-t)^{n-k} \\ &\quad + \sum_{k\in S_2} \left| x(t) - x\left(\frac{k}{n}\right)\right| \binom{n}{k} t^k (1-t)^{n-k}. \end{aligned}$$

But $|x(t)-x(k/n)|\leqq\epsilon$ for k in S_1, and hence

$$|x(t) - E_n(t)| \leqq \epsilon \sum_{k\in S_1} \binom{n}{k} t^k (1-t)^{n-k} + \sum_{k\in S_2} \left| x(t) - x\left(\frac{k}{n}\right)\right| \binom{n}{k} t^k (1-t)^{n-k}.$$

Now, since a continuous function on a closed interval is bounded, we have $|x(t)| \leqq M$ for $0 \leqq t \leqq 1$ and $|x(t)-x(k/n)| \leqq 2M$. Moreover,

$$\sum_{k\in S_1} \binom{n}{k} t^k(1-t)^{n-k} \leqq \sum_{k=0}^{n} \binom{n}{k} t^k(1-t)^{n-k} = 1,$$

and we obtain

$$|x(t) - E_n(t)| \leqq \epsilon + 2M \sum_{k\in S_2} \binom{n}{k} t^k(1-t)^{n-k}. \tag{1}$$

5. The Tchebyshev inequality and the completion of the proof. Recalling that k is in S_2 if and only if $|t-(k/n)| > \delta(\epsilon)$, or $|k-nt| > n\delta(\epsilon)$, we see that

$$\sum_{k\in S_2} \binom{n}{k} t^k(1-t)^{n-k}$$

is exactly the probability that k, the number of "heads" turning up in one play of n tosses, satisfies the inequality $|k-nt| > n\delta(\epsilon)$. We also notice that nt is, on the average, the number of "heads" we would expect in one play.

Now, if a random variable k has average value a, and if b^2 is the average value of $(k-a)^2$, then the Tchebyshev inequality of elementary probability theory states that

$$Pr[|k-a| > c] \leqq \frac{b^2}{c^2} \cdot \tag{2}$$

Applying this to our case, we obtain

$$\sum_{k\in S_2} \binom{n}{k} t^k(1-t)^{n-k} \leqq \frac{b^2}{n^2\delta(\epsilon)^2} \cdot \tag{3}$$

The value of b^2, i.e., the average of $(k-nt)^2$, is by definition

$$b^2 = \sum_{k=0}^{n} (k-nt)^2 \times [\text{Probability of exactly } k \text{ "heads"}]$$

$$= \sum_{k=0}^{n} (k-nt)^2 \binom{n}{k} t^k(1-t)^{n-k}.$$

We will prove in section 6 that this sum has the value $nt(1-t)$.

We obtain from (3) with $b^2 = nt(1-t)$

$$\sum_{k\in S_2} \binom{n}{k} t^k(1-t)^{n-k} \leqq \frac{t(1-t)}{n\delta(\epsilon)^2} \leqq \frac{1}{4n\delta(\epsilon)^2} \tag{4}$$

since $t(1-t) \leqq \frac{1}{4}$ for $0 \leqq t \leqq 1$. If we substitute this into (1) we obtain

$$|x(t) - E_n(t)| \leqq \epsilon + \frac{M}{2n\delta(\epsilon)^2} \tag{5}$$

which can obviously be made smaller than any $\epsilon^* > 0$, by choosing $\epsilon = \epsilon^*/2$

and then n so large that $M/2n\delta(\epsilon)^2<\epsilon^*/2$. This completes the proof of the approximation theorem.

6. The evaluation of b^2 and a direct proof of (3). If we differentiate the identity $1=\sum_{k=0}^{n}\binom{n}{k}t^k(1-t)^{n-k}$ we obtain after some formal manipulations

$$nt = \sum_{k=0}^{n} k\binom{n}{k}t^k(1-t)^{n-k}$$

and if we differentiate again we obtain

$$n(n-1)t^2 = \sum_{k=0}^{n} k(k-1)\binom{n}{k}t^k(1-t)^{n-k}.$$

Combination of these three identities yields

$$nt(1-t) = \sum_{k=0}^{n}(k-nt)^2\binom{n}{k}t^k(1-t)^{n-k} = b^2. \tag{6}$$

A discussion of the Tchebyshev inequality which is general enough for our purposes is contained in [2]. However, a direct proof of (3) can be easily obtained from (6) as follows: if we sum in (6) over only those values of k which are in S_2, we obtain

$$\sum_{k\in S_2}(k-nt)^2\binom{n}{k}t^k(1-t)^{n-k} \leqq nt(1-t).$$

Now, for all k in S_2, we have $|k-nt|>n\delta(\epsilon)$ and $(k-nt)^2>n^2\delta(\epsilon)^2$. Hence the left member of the above inequality is greater than

$$n^2\delta(\epsilon)^2\sum_{k\in S_2}\binom{n}{k}t^k(1-t)^{n-k}$$

and we obtain

$$n^2\delta(\epsilon)^2\sum_{k\in S_2}\binom{n}{k}t^k(1-t)^{n-k} \leqq nt(1-t),$$

from which (3) follows.

References

1. S. Bernstein, Demonstration du théorème de Weierstrass, fondée sur le calcul des probabilités, Commun. Soc. Math. Kharkow(2), 13(1912–13) 1–2.

2. J. Kemeny, H. Mirkil, J. L. Snell, G. L. Thompson, Finite mathematical structures, Prentice-Hall, Englewood Cliffs, N. J., 1959.