Here we give an alternative approach to Section 8.7 - the Weierstrass Approximation Theorem. It comes from my favorite advanced calculus book - *Undergraduate Analysis* by Serge Lang. Taylor’s formula gave us good polynomial approximations to certain but not all infinitely differentiable functions on a finite interval, as we saw in Section 8.6, with the function $f(x) = e^{-1/x}$, for $x > 0$, and $f(x) = 0$, for $x \leq 0$. Here we assume far less about our function $f : [a, b] \rightarrow \mathbb{R}$. That is, we assume only that $f(x)$ is continuous. It may have no derivatives.

**The Weierstrass Approximation Theorem.** If $f : [a, b] \rightarrow \mathbb{R}$ is continuous, $\forall \; \epsilon > 0$, $\exists$ a polynomial $p(x)$ such that

$$|f(x) - p(x)| < \epsilon, \text{ for all } x \in [a, b].$$

For Lang’s proof of this theorem, it is easiest to use improper integrals, even though we will assume our functions are zero outside of $[a, b]$. 
Defn. \[ \int_{-\infty}^{\infty} f(t) dt = \lim_{A \to -\infty} \left( \lim_{B \to \infty} \int_{A}^{B} f(t) dt \right). \]

Next we define an operation on functions \( f, g : \mathbb{R} \to \mathbb{R} \) called convolution or splat.

Defn. \((f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x - t) dt. \) Here we assume that \( f, g \) are bounded, piecewise continuous on every finite interval and that \( \int_{-\infty}^{\infty} |g(t)| dt < \infty. \) For our purposes we may assume that either \( f(x) = 0 \) or \( g(x) = 0, \text{ when } x \text{ is outside a finite interval } [a, b]. \)

**Theorem. Properties of Convolution.** Assume that the functions being convolved satisfy the hypothesis of the definition.

1) \((f + g) * h = f * h + g * h \)

2) If \( c \) is a real number, \( c(f * g) = (cf) * g \)
3) \( f \ast g = g \ast f \)

4) \( f \ast (g \ast h) = (f \ast g) \ast h. \)

5) If \( g(x) \) is a polynomial and \( f(x) = 0 \) for \( x \notin [a, b] \) then \( (f \ast g)(x) \) is a polynomial.

**Proof of 3).** Assume \( f(x) = 0 \) for \( x \notin [a, b] \) if you don’t want to think about improper integrals.

\[
(f \ast g)(x) = \int_{-\infty}^{\infty} f(t)g(x - t)dt.
\]

Now substitute \( u = x - t \). This gives \( t = x - u \) and the minus sign in \( du = -dt \) cancels with the switch in the direction of integration. Thus

\[
(f \ast g)(x) = \int_{-\infty}^{\infty} f(x - u)g(u)du = (g \ast f)(x).
\]
Proof of 5). By properties 1 and 2, it suffices to look at \( g(x) = x^n \). We will use the binomial theorem and facts about integrals (like the integral of a sum is the sum of the integrals).

\[
(f \ast g)(x) = \int_a^b f(t)g(x - t)dt \\
= \int_a^b f(t)(x - t)^n dt \\
= \sum_{k=0}^n \binom{n}{k} x^k \int_a^b f(t)(-t)^{n-k} dt.
\]

This is a polynomial.

\[\square\]

Optional Exercises for Homework # 6.

1) Prove properties 1 and 2 of convolution.
2) Prove property 4 of convolution.

It is common in engineering and physics courses to refer to a "function" known as the Dirac delta function $\delta(x)$. It is supposed to have the values $\delta(x) = 0$, for all $x \neq 0$, and $\delta(0) = \infty$. To a mathematician, this is not a legal definition of a function. We have 2 legal ways to define $\delta(x)$. One way is to approximate $\delta(x)$ by sequences of functions called "delta sequences" and the other is to definite $\delta(x)$ by what it does to other functions. This makes it a generalized function or distribution - a subject originating with Laurent Schwartz around 1950.

For $\varphi : \mathbb{R} \to \mathbb{R}$ (assumed infinitely differentiable and such that $\varphi(x) = 0$, when $|x| > L$), we can define $\delta$ by

$$\int_{-\infty}^{\infty} \varphi(t)\delta(t)dt = \varphi(0).$$

However, here we take the approach of replacing $\delta$ by an approximating sequence of functions $\{K_n\}_{n \geq 1}$. 
Defn. A Dirac sequence of functions $K_n : \mathbb{R} \to \mathbb{R}$, $n = 1, 2, 3, \ldots$; satisfies

**Dirac 1.** $K_n(x) \geq 0$, $\forall x \in \mathbb{R}$, $\forall n \in \mathbb{Z}^+$.

**Dirac 2.** $K_n$ is continuous and $\int_{-\infty}^{\infty} K_n(t)dt = 1$, $\forall n \in \mathbb{Z}^+$.

**Dirac 3.** (area is more and more concentrated at 0 as $n \to \infty$)

$\forall \epsilon > 0$, $\forall \delta > 0$, $\exists N \in \mathbb{Z}^+$, such that $n \geq N$ implies

$$\int_{-\delta}^{\delta} K_n(t)dt + \int_{\delta}^{\infty} K_n(t)dt < \epsilon.$$

**Examples.**

1) Gauss Kernel: $G_n(x) = \sqrt{\frac{n}{2\pi}}e^{-nx^2/2}$
Here we use Mathematica to plot $n = 1, 10, 20$:

Plot[{Exp[-x^2/(2*1)]/Sqrt[2*Pi*1],
      Exp[-x^2/(2*.1)]/Sqrt[2*Pi*.1],
      Exp[-x^2/(2*.05)]/Sqrt[2*Pi*.05]},
     {x,-10,10},PlotRange->All]

You can see that the larger the $n$ the closer the function to a spike at the origin. In physics and engineering, the delta function is thought of as an impulse.
Actually the Gauss kernel is usually defined as a continuous kernel $G_t(x)$, as $t \to 0^+$. Here I replaced $t$ with $1/n$.

2) **Landau Kernel.**

\[
L_n(t) = \begin{cases} 
\frac{1}{c_n}(1 - t^2)^n, & \text{if } |t| \leq 1, \\
0, & \text{if } |t| > 1.
\end{cases}
\]

Here the constant $c_n$ is chosen to make Dirac 2 true. That is:

\[
c_n = \frac{1}{\int_{-1}^{1} (1 - t^2)^n dt}.
\]

**Theorem.** $L_n$ is a Dirac Sequence.

**Proof.**
Dirac 1 and 2 are clear.

To prove Dirac 3, we first

**Claim.** \( c_n \geq \frac{2}{n+1} \).

**Proof of Claim.**

\[
\frac{c_n}{2} = \int_{0}^{1} (1 - t^2)^n dt
\]

\[
= \int_{0}^{1} (1 - t)^n (1 + t)^n dt
\]

\[
\geq \int_{0}^{1} (1 - t)^n dt = \frac{1}{n + 1}.
\]

Now we can finish the proof of Dirac 3. It suffices, since
\[ L_n(-x) = L_n(x), \text{ to look at} \]
\[ \int_{\delta}^{1} L_n(t) \, dt = \int_{\delta}^{1} \frac{1}{c_n} (1 - t^2)^n \, dt \]
\[ \leq \int_{\delta}^{n+\frac{1}{2}} \frac{n+1}{2} (1 - \delta^2)^n \, dt \]
\[ = \frac{n+1}{2} (1 - \delta^2)^n (1 - \delta). \]

The last expression \( \frac{n+1}{2} (1 - \delta^2)^n (1 - \delta) \) approaches 0 as \( n \to \infty \).

\[ \square \]

The following theorem will imply the Weierstrass approximation theorem.

**Theorem.** Suppose that \( f \) is continuous on \([a, b]\). Let \( K_n \) be a Dirac sequence. Then \( \forall \ \epsilon > 0, \ \exists N \) such that \( n \geq N, \) implies \( |(K_n * f)(x) - f(x)| < \epsilon, \) for
all \( x \in [a, b] \). Here we assume \( f \) and \( K_m \) vanish when 
\( x \not\in [a, b] \).

Proof.

Using the definition of convolution plus Dirac 1 and 2, we see that

\[
\left| (K_n * f)(x) - f(x) \right| \\
= \left| \int_{-\infty}^{\infty} K_n(t) f(x-t) dt - \int_{-\infty}^{\infty} f(x) K_n(t) dt \right| \\
= \left| \int_{-\infty}^{\infty} (f(x-t) - f(x)) K_n(t) dt \right| \\
\leq \int_{-\infty}^{\infty} |f(x-t) - f(x)| K_n(t) dt.
\]
We split the last integral into 2 parts

\[
\int_{-\infty}^{\infty} |f(x - t) - f(x)| \, K_n(t) \, dt = \int_{|t| \leq \delta} |f(x - t) - f(x)| \, K_n(t) \, dt + \int_{|t| > \delta} |f(x - t) - f(x)| \, K_n(t) \, dt
\]

By uniform continuity of \( f \), \( \forall \epsilon > 0, \exists \delta > 0 \), such that \( |f(x - t) - f(x)| < \epsilon \). Use this \( \delta \) and Dirac 2 to see that the first integral is less than

\[
\epsilon \int_{|t| \leq \delta} K_n(t) \, dt < \epsilon \int_{-\infty}^{\infty} K_n(t) \, dt = \epsilon.
\]

Then use Dirac 3 to see that if \( M \) is the bound on \( |f| \) on our interval \([a, b]\), the second integral is \( < 2M\epsilon \).
The Weierstrass theorem is a corollary of the preceding theorem, once we adjust our interval \([0,1]\) for the Landau kernel to an arbitrary interval. Our text does this argument on page 226.