Weierstrass' Approximation Theorem

Kexing Ying

Abstract

We will provide a proof for Weierstrass' approximation theorem using the law of large numbers.

The *Weierstrass' approximation theorem* is a powerful theorem that showed that algebraic polynomials are dense in set of continuous real-valued functions. We will prove this fact here.

We shall first consider the following lemmas.

Lemma 0.1. *Let $X_1, X_2, \cdots, X_i, \cdots, X_n$ be a sequence of independently and identically distributed random variables following the Bernoulli distribution with parameter x and let $f : [0,1] \to \mathbb{R}$ be a continuous function. Then*

$$\mathbf{E}\left[f\left(\frac{S_n}{n}\right)\right] = B_n(x)$$

where $S_n := \sum_{i=1}^n X_i$, and

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

Proof. As S_n is defined to be the sum of n i.i.d Bernoulli random variables, it follows a binomial distribution with parameter n and x (prove it!). Thus, S_n has a probability mass function,

$$p_{S_n}(k) = \Pr(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Then, by the *law of the unconscious statistician* (cite this), we have

$$\mathbf{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k \in \operatorname{supp}\left(\frac{S_n}{n}\right)} f(k) \Pr\left(\frac{S_n}{n} = k\right),$$

where $\operatorname{supp}\left(\frac{S_n}{n}\right)$ denotes the support of $\frac{S_n}{n}$.

Now as, S_n is the sum of n Bernoulli random variables, S_n can take valued from $1, 2, \cdots, n$, and thus $\operatorname{supp}\left(\frac{S_n}{n}\right) = \{\frac{1}{n}, \frac{2}{n}, \cdots, 1\}$.

Then,

$$\begin{aligned}
\mathbf{E}\left[f\left(\frac{S_n}{n}\right)\right] &= \sum_{k\in\operatorname{supp}\left(\frac{S_n}{n}\right)} f(k)\Pr\left(\frac{S_n}{n}=k\right) \\
&= \sum_{k\in\operatorname{supp}\left(\frac{S_n}{n}\right)} f(k)\Pr\left(S_n=nk\right) \\
&= \sum_{k\in\{\frac{1}{n},\frac{2}{n},\cdots,1\}} f(k)\binom{n}{nk}x^{nk}(1-x)^{n-nk} \\
&= \sum_{i=0}^{n} f\left(\frac{i}{n}\right)\binom{n}{i}x^{i}(1-x)^{n-i} = B_n(x)
\end{aligned}$$

as required.

□

Lemma 0.2. *Let $f:[0,1]\to\mathbb{R}$ be a continuous function. Then for all $\epsilon>0$, there exists some $\delta>0$ such that*

$$\mathbf{E}\left[\mathbf{1}_{(A_n)^c}\left|f(x)-f\left(\frac{S_n}{n}\right)\right|\right]<\epsilon$$

where $\mathbf{1}_{(A_n)^c}$ is the indicator function for $(A_n)^c$ and

$$A_n(\delta)=\left\{\omega:\left|\frac{S_n(\omega)}{n}-x\right|\geq\delta\right\}.$$

Proof. Let us fix $\epsilon>0$. As f is continuous on a compact interval $[0,1]$, f is uniformally continuous (cite this), thus, there exists some $\delta>0$ such that, for all $x,y\in[0,1]$, if $|x-y|<\delta$, then $|f(x)-f(y)|<\epsilon$.

Then we see, that, if $\omega\in A_n(\delta)^c$, we have $\mathbf{1}_{A_n(\delta)^c}=1$ and as

$$\omega\in A_n(\delta)^c=\left\{\omega:\neg\left|\frac{S_n(\omega)}{n}-x\right|\geq\delta\right\}=\left\{\omega:\left|\frac{S_n(\omega)}{n}-x\right|<\delta\right\},$$

i.e. $\left|x-\frac{S_n}{n}\right|<\delta$, and thus, by the construction of δ, $\left|f(x)-f\left(\frac{S_n}{n}\right)\right|<\epsilon$ and hence,

$$\mathbf{1}_{A_n(\delta)^c}\left|x-\frac{S_n}{n}\right|=\left|x-\frac{S_n}{n}\right|<\epsilon.$$

On the other hand, if $\omega\notin A_n(\delta)^c$, we have $\mathbf{1}_{A_n(\delta)^c}=0$, so

$$\mathbf{1}_{A_n(\delta)^c}\left|x-\frac{S_n}{n}\right|=0.$$

Now, as the events of $\omega \in A_n(\delta)^c$ and the event $\omega \notin A_n(\delta)^c$ partitions the sample space by the *law of excluded middle* (cite this), by the *total law of expectation*, we have,

$$\begin{aligned}
\mathbf{E}\left[\mathbf{1}_{(A_n)^c}\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] &= \mathbf{E}\left[\left(\mathbf{1}_{(A_n)^c}\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right) \mid \omega \in A_n(\delta)^c\right] \Pr(\omega \in A_n(\delta)^c) \\
&+ \mathbf{E}\left[\left(\mathbf{1}_{(A_n)^c}\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right) \mid \omega \notin A_n(\delta)^c\right] \Pr(\omega \notin A_n(\delta)^c) \\
&= \mathbf{E}\left[\left(\mathbf{1}_{(A_n)^c}\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right) \mid \omega \in A_n(\delta)^c\right] \Pr(\omega \in A_n(\delta)^c) \\
&< \epsilon \Pr(\omega \in A_n(\delta)^c) \leq \epsilon
\end{aligned}$$

as required. □

Now armed with the above two lemmas, we can finally prove the *Weierstrass' approximation theorem.*

Theorem 1 (Weierstrass' approximation theorem)**.** *Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Then for all $\epsilon > 0$, there exists a polynomial P_n such that*

$$\sup_{x \in [0,1]} |f(x) - P_n(x)| < \epsilon.$$

Proof. Fix $\epsilon > 0$, then by lemma 0.2, there is some $\delta > 0$ such that

$$\begin{aligned}
\frac{1}{2}\epsilon &> \mathbf{E}\left[\mathbf{1}_{A_n(\delta)^c}\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \\
&= \mathbf{E}\left[\left(\mathbf{1}_{(A_n)^c}\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right) \mid A_n(\delta)^c\right] \Pr(A_n(\delta)^c),
\end{aligned}$$

where $A_n(\delta)$ is defined the same way as lemma 0.2.

By considering the *weak law of large numbers*, we have $P(A_n(\delta)) \to 0$ as $n \to \infty$. So there exists some $N \in \mathbb{N}$, such that for all $n \geq N$, $|P(A_n(\delta))| < \frac{1}{2}$.

Now, consider by triangle inequality, for any discrete random variable Y,

$$\mathbf{E}(|Y|) = \sum |y| \Pr(Y = y) \geq \left|\sum y \Pr(Y = y)\right| = |\mathbf{E}(Y)|,$$

where the second equality is due to the law of the unconscious statistician.

Thus, with this, we can establish the following inequality,

$$\begin{aligned}
&\mathbf{E}\left[\left(\mathbf{1}_{(A_N)^c}\left|f(x) - f\left(\frac{S_N}{N}\right)\right|\right) \mid (A_N)^c\right] \Pr((A_N)^c) \\
&= \mathbf{E}\left[\left|f(x) - f\left(\frac{S_N}{N}\right)\right|\right] \Pr((A_N)^c) \\
&\geq \left|\mathbf{E}\left[f(x) - f\left(\frac{S_N}{N}\right)\right]\right| \Pr((A_N)^c) \\
&= \left|f(x) - \mathbf{E}\left[f\left(\frac{S_N}{N}\right)\right]\right| \Pr((A_N)^c).
\end{aligned}$$

However, by lemma 0.1, we have $\mathbf{E}\left[f\left(\frac{S_N}{N}\right)\right] = B_N$, thus,

$$\begin{aligned}\left|f(x) - \mathbf{E}\left[f\left(\frac{S_N}{N}\right)\right]\right| \Pr((A_N)^c) &= \left|f(x) - B_N(x)\right| \Pr((A_N)^c) \\ &= \left|f(x) - B_N(x)\right| (1 - \Pr(A_N)).\end{aligned}$$

Now, as we have choosen N, such that $\Pr(A_N) < \frac{1}{2}$, $1 - \Pr(A_N) > \frac{1}{2}$, so

$$\left|f(x) - B_N(x)\right| (1 - \Pr(A_N)) > \frac{1}{2}\left|f(x) - B_N(x)\right|,$$

and hence,

$$\frac{1}{2}\epsilon > \frac{1}{2}\left|f(x) - B_N(x)\right| \implies \epsilon > \left|f(x) - B_N(x)\right|.$$

As B_N is a polynomial, we have found polynomial that is at most ϵ distance away from f, as required. $\square$