# Math 525: Lecture 11

February 8, 2018

## 1 Laws of large numbers

Intuitively, we know that

$$\frac{\text{number of successes in } n \text{ independent trials}}{n} \to \mathbb{P}(\text{success})$$

as $n \to \infty$. This lecture aims to prove this.

Let's first restate the above in terms of random variables. Let $X_i$ be the random variable whose value is 1 if the $i$-th trial is a success and whose value is 0 otherwise. In other words,

$$X_i = I_{\{i\text{-th trial is a success}\}}.$$

Since the trials are independent, we must assume that $X_i$ is independent of $X_j$ for $i \neq j$. Then, the number of successes in the first $n$ trials is simply

$$X_1 + \cdots + X_n$$

Using this notation, our first claim becomes

$$\frac{1}{n}\left(X_1 + \cdots + X_n\right) \to \mathbb{E}[X_1],$$

which is a statement about sums of random variables converging to the mean $\mu = \mathbb{E}[X_1]$ (we are purposely vague as to the type of convergence for the time being). Defining $X_i' = X_i - \mu$ and using the claim above, note that

$$\frac{1}{n}\left(X_1 + \cdots + X_n\right) - \mu = \frac{1}{n}\left(X_1 - \mu + \cdots + X_n - \mu\right) = \frac{1}{n}\left(X_1' + \cdots + X_n'\right) \to 0$$

and hence we can simply focus on the mean-zero case.

**Proposition 1.1** (Weak law of large numbers)**.** *Let $(X_n)_n$ be an i.i.d. sequence of square integrable, mean zero random variables. Then,*

$$\frac{1}{n}\left(X_1 + \cdots + X_n\right) \to 0$$

*in probability.*

For the remainder of this lecture, we will use the notation $S_n = X_1 + \cdots + X_n$.

*Proof.* First, note that

$$\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left[X_1X_1 + X_1X_2 + \cdots + X_1X_n + \cdots + X_nX_n\right].$$

By independence, $\mathbb{E}[X_iX_j] = 0$ whenever $i \neq j$. Therefore, the above becomes

$$\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^2\right] = \frac{1}{n^2}\left(\mathbb{E}\left[X_1^2\right] + \cdots + \mathbb{E}\left[X_n^2\right]\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

where we have used $\sigma^2 = \mathbb{E}[X_1^2]$. By Markov's/Chebyshev's inequality,

$$\mathbb{P}\left\{\frac{1}{n}\left|S_n\right| \geq \epsilon\right\} \leq \frac{1}{\epsilon^2}\frac{\sigma^2}{n} \to 0.$$ □

The above is called a "weak law" since convergence is in probability. There are various "strong laws" which guarantee convergence a.e. Here's one:

**Proposition 1.2** (Cantelli's strong law of large numbers)**.** *Let $(X_n)_n$ be an i.i.d. sequence of mean zero random variables, with $X_1^4$ integrable. Then,*

$$\frac{1}{n}S_n \to 0 \ a.s.$$

*Proof.* Note that

$$\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^4\right] = \frac{1}{n^4}\sum_{1\leq i,j,k,\ell\leq n}\mathbb{E}\left[X_iX_jX_kX_\ell\right].$$

Using the same arguments as in the proof of the weak law of large numbers, note that the only terms that do not have expectation of zero in the sum are of the form $X_i^4$ and $X_i^2X_j^2$ for $i \neq j$. That is,

$$\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^4\right] = \frac{1}{n^4}\sum_{i=1}^n\left(\mathbb{E}\left[X_i^4\right] + \sum_{\substack{j=1\\ j\neq i}}^n\mathbb{E}\left[X_i^2X_j^2\right]\right).$$

Let $\mathbb{E}[X_1^4] = M$. Then,

$$\mathbb{E}\left[X_i^2X_j^2\right] \leq \sqrt{\mathbb{E}\left[X_i^4\right]\mathbb{E}\left[X_j^4\right]} \leq \sqrt{M\cdot M} = M$$

so that

$$\mathbb{E}\left[\left(\frac{1}{n}S_n\right)^4\right] \leq \frac{1}{n^4}\sum_{i=1}^n\left(M + (n-1)M\right) = \frac{M + (n-1)M}{n^3} = \frac{M}{n^2}.$$

Chebyshev's inequality tells us

$$\mathbb{P}\left\{\frac{|S_n|}{n} \geq \lambda\right\} \leq \frac{1}{\lambda^4}\frac{M}{n^2}.$$

Pick $\lambda = n^{-1/8}$ and define

$$\Lambda_n = \left\{\frac{|S_n|}{n} \geq \frac{1}{n^{1/8}}\right\}$$

so that

$$\sum_n \mathbb{P}(\Lambda_n) \leq \sum_n \frac{M}{n^{3/2}} < \infty.$$

Now, the Borel-Cantelli lemma tells us

$$1 - \mathbb{P}(\liminf_n \Lambda_n^c) = \mathbb{P}(\limsup_n \Lambda_n) = 0.$$

In other words, for all $\omega \in \liminf_n \Lambda_n^c$, we can find $N(\omega)$ such that for all $n \geq N(\omega)$,

$$\frac{|S_n|}{n} < \frac{1}{n^{1/8}}.$$

That is, $S_n/n$ converges to zero a.s. □

## 2 Empirical distribution

Let $(X_n)_n$ be a sequence of i.i.d. random variables and define its *empirical distribution* by

$$F_n(x) = \frac{1}{n}\left|\{j \leq n \colon X_j \leq x\}\right|.$$

In other words, $F_n(x)$ counts exactly how many of the first $n$ random variables are at most $x$. Unlike the distribution function $F$, $F_n(x)$ is itself a random variable. Sometimes, we might write

$$(F_n(x))(\omega)$$

to stress this fact.

**Proposition 2.1.** *Let $(X_n)_n$ be a sequence of i.i.d. random variables with distribution function $F$ and empirical distribution $F_n$. Then,*

$$F_n \to F \text{ a.s.}$$

Actually, $F_n \to F$ in a stronger sense than pointwise, but we will not prove that.

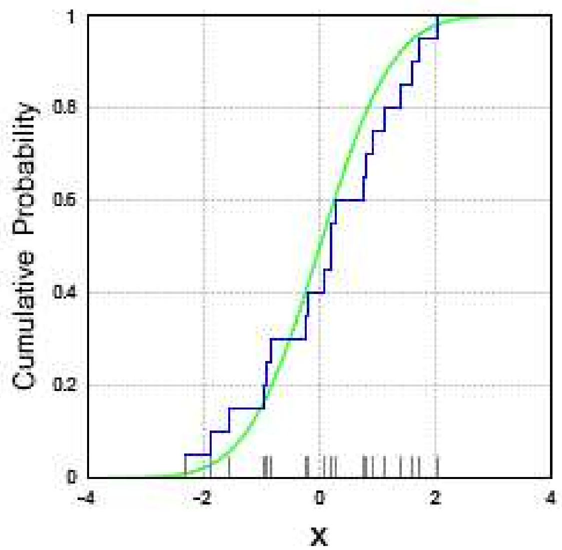


Figure 1: Empirical distribution converging to the *actual* distribution

*Proof.* Let $(U_n)_n$ be a sequence i.i.d. $U[0,1]$ random variables with empirical distribution $F_n^0$. Let $F^0$ be the *actual* distribution function of $U_1$. Now, fix $x \in [0,1]$ and let $Y_i = I_{\{U_i \le x\}}$. Note that $\mathbb{E}[Y_i] = \mathbb{P}(U_i \le x) = x$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = F_n^0(x).$$

Since $\mathbb{E}[Y_i^4]$ is a constant independent of $i$, Cantelli's strong law of large numbers tells us

$$F_n^0(x) \to F^0(x) \text{ a.s.}$$

It remains to extend this to general random variables. We will not do the most general case here and instead focus on the case when $F$ is bijective. Without loss of generality, $X_n = F^{-1}(U_n)$. Therefore,

$$\begin{aligned} F_n(x) = \frac{1}{n}\left|\{j \le n \colon X_j \le x\}\right| = \frac{1}{n}\left|\left\{j \le n \colon F^{-1}(U_j) \le x\right\}\right| = \frac{1}{n}\left|\{j \le n \colon U_j \le F(x)\}\right| \\ = F_n^0(F(x)). \end{aligned}$$

Since $F^0(x) = x$ whenever $x \in [0,1]$,

$$|F(x) - F_n(x)| = \left|F^0(F(x)) - F_n^0(F(x))\right| \to 0 \text{ a.s.},$$

as desired. $\square$

# 3 Bernstein polynomials

In a real analysis course, you might have learned that the set of all polynomials on $[0,1]$, call it $\mathcal{P}[0,1]$, is dense in $C[0,1]$, the set of all continuous real-valued functions from $[0,1]$. This is called the *Stone-Weirestrass theorem*. We now give a constructive proof of this fact, using probability theory!

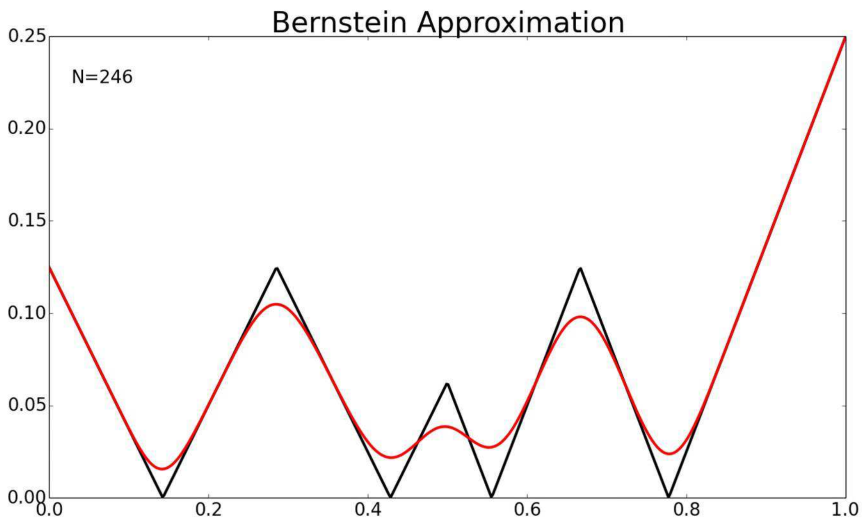


Figure 2: Bernstein polynomials approximating a nondifferentiable function

**Proposition 3.1.** *Let $f$ be continuous on $[0,1]$. Then, there exists a sequence $(p_n)_n$ of polynomials on $[0,1]$ such that $p_n \to f$ uniformly. That is,*

$$\|p_n - f\|_\infty \to 0 \qquad \text{as } n \to \infty.$$

*where $\|g\|_\infty = \sup_x |g(x)|$. Moreover, these polynomials, called Bernstein polynomials, have the form*

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

*Proof.* Let $(X_n)_n$ be a sequence of i.i.d. Bernoulli($p$) random variables, for some $0 \le p \le 1$. As usual, define $S_n = X_1 + \cdots + X_n$. Note that $S_n \sim B(n,p)$, and hence $\mathbb{E}[S_n] = np$ (equivalently, $\mathbb{E}[\frac{1}{n}S_n] = p$). Moreover, $\mathrm{Var}(S_n) = np(1-p)$.

By the strong law of large numbers, $S_n/n \to p$ a.s. Let $\delta > 0$. By Chebyshev's inequality,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - p\right| \ge \delta\right\} \le \frac{1}{\delta^2}\mathbb{E}\left[\left|\frac{1}{n}S_n - p\right|^2\right] = \frac{\mathrm{Var}(\frac{1}{n}S_n)}{\delta^2} = \frac{\mathrm{Var}(S_n)}{n^2\delta^2} = \frac{p\,(1-p)}{n\delta^2} \le \frac{1}{4n\delta^2}.$$

The last step is since

$$\max_{p\in[0,1]} p\,(1-p) = \frac{1}{4}$$

(you can prove that by ordinary calculus: take the first derivative and set it to zero).

Now, let $f \in C[0,1]$ be arbitrary. Since

$$\frac{1}{n}S_n \to p \text{ a.s.}$$

a.s., then by continuity,

$$f\left(\frac{1}{n}S_n\right) \to f(p) \text{ a.s.}$$

By the dominated convergence theorem,

$$\mathbb{E}\left[f\left(\frac{1}{n}S_n\right)\right] \to \mathbb{E}\left[f\left(p\right)\right] = f(p).$$

Let $\epsilon > 0$. Since continuity and uniform continuity are equivalent on compact sets, we can pick $\delta > 0$ such that for $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\left|f\left(\frac{1}{n}S_n\right) - f(p)\right|\right] &\leq \epsilon \mathbb{P}\left\{\left|\frac{1}{n}S_n - p\right| < \delta\right\} + 2\|f\|_\infty \mathbb{P}\left\{\left|\frac{1}{n}S_n - p\right| \geq \delta\right\} \\
&\leq \epsilon + 2\|f\|_\infty \frac{1}{4n\delta^2}.
\end{aligned}$$

Since this bound is independent of $p$, we see $\mathbb{E}[f(S_n/n)] \to f(p)$ uniformly. Now, if we set $p = x$,

$$\mathbb{E}\left[f\left(\frac{1}{n}S_n\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^n \left(1 - x\right)^{n-k}.$$

The right hand side above defines the $n$-th Bernstein polynomial, $p_n(x)$. $\square$