Weak Law of Large Numbers

Theorem (WLLN). *If* $\{X_1, ..., X_n\}$ *are iid with* $E|X_i| < \infty$ *and then* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \to_p E(X_i)$.

Proof. Without loss of generality, we can set $E(X_i) = 0$ (by recentering X_i on its expectation). We need to show that for all $\delta > 0$ and $\eta > 0$ there is some $\overline{n} < \infty$ so that for all $n \geq \overline{n}$, $P\left(\left|\overline{X}_n\right| > \delta\right) \leq \eta$. Fix δ and η. Set $\varepsilon = \delta\eta/3$. Pick $C < \infty$ large enough so that

$$E\left(|X|\,1\left(|X| > C\right)\right) \leq \varepsilon \tag{1}$$

(where $1(\cdot)$ is the indicator function) which is possible since $E|X| < \infty$. Then set

$$\overline{n} \geq 4C^2/\varepsilon^2. \tag{2}$$

Define the random vectors

$$\begin{aligned} W_i &= X_i 1\left(|X_i| \leq C\right) - E\left(X_i 1\left(|X_i| \leq C\right)\right) \\ Z_i &= X_i 1\left(|X_i| > C\right) - E\left(X_i 1\left(|X_i| > C\right)\right). \end{aligned}$$

Since X_i is iid, W_i and Z_i are also.

By Jensen's inequality and (1),

$$\left|E\left(X_i 1\left(|X_i| > C\right)\right)\right| \leq E\left(|X_i|\,1\left(|X_i| > C\right)\right) \leq \varepsilon.$$

By the triangle inequality and (1),

$$E\left|\overline{Z}_n\right| \leq E|Z_i| \leq E|X_i|\,1\left(|X_i| > C\right) + \left|E\left(X_i 1\left(|X_i| > C\right)\right)\right| \leq 2\varepsilon.$$

Note that $|W_i| \leq 2C$. Thus (crudely) $EW_i^2 \leq 4C^2$. Since the W_i are iid and mean zero,

$$E\overline{W}_n^2 = \frac{EW_i^2}{n} \leq \frac{4C^2}{n} \leq \varepsilon^2$$

the final inequality holding for $n \geq \overline{n}$ by (2). Thus by Jensen's inequality

$$\left(E\left|\overline{W}_n\right|\right)^2 \leq E\overline{W}_n^2 \leq \varepsilon^2.$$

Finally, by Markov's inequality, the fact that $\overline{X}_n = \overline{W}_n + \overline{Z}_n$, the triangle inequality, and these two bounds,

$$P\left(\left|\overline{X}_n\right| > \delta\right) \leq \frac{E\left|\overline{X}_n\right|}{\delta} \leq \frac{E\left|\overline{W}_n\right| + E\left|\overline{Z}_n\right|}{\delta} \leq \frac{3\varepsilon}{\delta} = \eta,$$

the equality by the definition of ε. We have shown that for any $\delta > 0$ and $\eta > 0$ there is some $\overline{n} < \infty$ so that for all $n \geq \overline{n}$, $P\left(\left|\overline{X}_n\right| > \delta\right) \leq \eta$, as needed. ■

Strong Law of Large Numbers

Theorem (SLLN). *If $\{X_1, ..., X_n\}$ are iid with $E|X_i| < \infty$ and $EX_i = \mu$ then $\overline{X}_n \to_{a.s.} \mu$ as $n \to \infty$.*

Classical proofs of strong laws are based on convergence results from analysis. Two powerful results are known as the Toeplitz Lemma and the Kronecker Lemma.

A **Toeplitz array** $\{a_{ni}\}$ satisfies the following three characteristics:
(i) For all $n \geq 1$, $\sum_{i=1}^{\infty} |a_{ni}| \leq c < \infty$
(ii) As $n \to \infty$, $\sum_{i=1}^{\infty} a_{ni} \to 1$
(iii) For all $i \geq 1$, as $n \to \infty$, $a_{ni} \to 0$

An example of a Toeplitz array is $a_{ni} = 1/n$ if $i \leq n$, else $a_{ni} = 0$.

Toeplitz Lemma. *If $\{a_{ni}\}$ is a Toeplitz array and x_n is a real sequence such that $x_n \to x$ as $n \to \infty$ then as $n \to \infty$*

$$y_n = \sum_{i=1}^{\infty} a_{ni} x_i \to x.$$

Proof: Using property (ii) WLOG assume $x = 0$. Fix $\varepsilon > 0$ and pick N so that $|x_i| \leq \varepsilon/2c$ for all $i \geq N$. Then by property (i)

$$\begin{aligned} |y_n| &\leq \sum_{i=1}^{N} |a_{ni}| \, |x_i| + \sum_{i=N+1}^{\infty} |a_{ni}| \, |x_i| \\ &\leq \sum_{i=1}^{N} |a_{ni}| \, |x_i| + \varepsilon/2 \\ &\leq \varepsilon \end{aligned}$$

the final inequality holding for n sufficiently large by property (iii). ■

Kronecker Lemma. If b_n is an increasing real sequence with $b_n \to \infty$, and x_n is a real sequence such that $\sum_{i=1}^{\infty} x_i$ exists (that is, $\sum_{i=1}^{n} x_i$ converges to a finite limit as $n \to \infty$), then

$$\frac{1}{b_n} \sum_{i=1}^{n} b_i x_i \to 0.$$

Proof. Let $s_n = \sum_{i=1}^{n} x_i$ and define $s_0 = 0$ and $b_0 = 0$. Now

$$\begin{aligned} \sum_{i=1}^{n} b_i x_i &= \sum_{i=1}^{n} b_i s_i - \sum_{i=1}^{n} b_i s_{i-1} \\ &= \sum_{i=1}^{n} b_i s_i - \sum_{i=1}^{n} b_{i-1} s_{i-1} - \sum_{i=1}^{n} (b_i - b_{i-1}) \, s_{i-1} \\ &= b_n s_n - \sum_{i=1}^{n} (b_i - b_{i-1}) \, s_{i-1}. \end{aligned}$$

Thus

$$\frac{1}{b_n}\sum_{i=1}^{n} b_i x_i = s_n - \sum_{i=1}^{n} a_{ni} s_{i-1} \tag{3}$$

where

$$a_{ni} = \frac{b_i - b_{i-1}}{b_n}$$

and we define $a_{ni} = 0$ for $i > n$. Note that $|a_{ni}| \leq 1$, $\sum_{i=1}^{\infty} a_{ni} = 1$, and $a_{ni} \to 0$ as $n \to \infty$, so a_{ni} is a Toeplitz array. Since $s_n \to x$ then $\sum_{i=1}^{n} a_{ni} s_{i-1} \to x$ by the Toeplitz Lemma and (3) converges to $x - x = 0$. ■

We also need a strengthening of Markov's inequality.

Kolmogorov's Inequality. *Assume $U_1, ..., U_n$ are independent (but not necessarily iid) with $EU_i = 0$. Set $S_j = \sum_{i=1}^{j} U_i$. Then for any $\lambda > 0$*

$$P\left(\max_{1\leq i\leq n} |S_i| > \lambda\right) \leq \frac{ES_n^2}{\lambda^2} = \frac{1}{\lambda^2}\sum_{i=1}^{n} EU_i^2. \tag{4}$$

Proof:. Define

$$I_{i-1} = \left\{|S_i| > \lambda; \max_{j<i} |S_j| \leq \lambda\right\},$$

the event that the sequence $|S_j|$ first exceeds λ at $j = i$. Since these events are disjoint,

$$P\left(\max_{1\leq i\leq n} |S_i| > \lambda\right) = P\left(\bigcup_{i=1}^{n} I_{i-1}\right) = \sum_{i=1}^{n} P\left(I_{i-1}\right) \leq \sum_{i=1}^{n} P\left(I_{i-1}\left|S_i\right| > \lambda\right) \leq \lambda^{-2}\sum_{i=1}^{n} E\left(I_{i-1}S_i^2\right). \tag{5}$$

The first inquality holds since $I_{i-1} = 1$ implies $I_{i-1}|S_i| > \lambda$, and the last inequality is Markov's. Let $\tilde{U}_i = (U_1, ..., U_i)$ and note that

$$E\left(S_n^2 \mid \tilde{U}_i\right) = E\left(S_i^2 \mid \tilde{U}_i\right) + 2E\left(S_i(S_n - S_i) \mid \tilde{U}_i\right) + E\left((S_n - S_i)^2 \mid \tilde{U}_i\right) = S_i^2 + E(S_n - S_i)^2 \geq S_i^2$$

so using iterated expectations,

$$E\left(I_{i-1}S_i^2\right) \leq E\left(I_{i-1}E\left(S_n^2 \mid \tilde{U}_i\right)\right) = E\left(E\left(I_{i-1}S_n^2 \mid \tilde{U}_i\right)\right) = E\left(I_{i-1}S_n^2\right). \tag{6}$$

Together, (5) and (6) show that

$$\lambda^2 P\left(\max_{1\leq i\leq n} |S_i| > \lambda\right) \leq \sum_{i=1}^{n} E\left(I_{i-1}S_n^2\right) = E\left(\left(\sum_{i=1}^{n} I_{i-1}\right) S_n^2\right) \leq ES_n^2.$$

■

Given the Kronecker Lemma and Kolmogorov's inequality, it is straightforward to establish the SLLN if $Var(X) < \infty$.

Almost Sure Convergence Theorem. If

$$\sum_{i=1}^{\infty} \frac{Var(X_i)}{i^2} < \infty \tag{7}$$

then $\overline{X}_n \to 0$ almost surely.

Before we prove this theorem, we state the following implication.
Kolmogorov SLLN. If X_i is iid and $Var(X_i) < \infty$ then $\overline{X}_n \to 0$ almost surely.

Proof of Kolmogorov SLLN.

$$\sum_{i=1}^{\infty} \frac{Var(X_i)}{i^2} = Var(X_i) \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

since $\sum_{i=1}^{\infty} \frac{1}{i^2} = \pi^2/6 < \infty$. Then by the almost sure convergence theorem, $\overline{X}_n \to 0$ almost surely.

Proof of Almost Sure Convergence Theorem. WLOG assume $EX_i = 0$. Let $U_i = i^{-1} X_i$. The Kronecker Lemma implies that if $S_n = \sum_{i=1}^{n} U_i$ converges to a finite random limit as $n \to \infty$, then $\overline{X}_n \to 0$. (To see this, set $x_i = U_i$ and $b_i = i$.) We now show that S_n converges almost surely as $n \to \infty$, so $\overline{X}_n \to 0$ almost surely.

One characterization of convergence is that S_n converges iff $S_{m+k} - S_m \to 0$ as $m, k \to \infty$. In other words, S_n converges if for all $\varepsilon > 0$, there is a sufficiently large $\overline{m} < \infty$ such that for all $m \geq \overline{m}$, $|S_{m+k} - S_m| < \varepsilon$ for all $k \geq 1$. But for all $\varepsilon > 0$ and $m < \infty$

$$\begin{aligned}
P\left(\text{for some } k \geq 1,\ \ \{|S_{m+k} - S_m| \geq \varepsilon\}\right) &= \lim_{n\to\infty} P\left(\max_{1 \leq k \leq n} |S_{m+k} - S_m| \geq \varepsilon\right) \\
&= \lim_{n\to\infty} P\left(\max_{1 \leq k \leq n} \left| \sum_{i=m+1}^{m+k} U_i \right| \geq \varepsilon\right) \\
&\leq \frac{1}{\varepsilon^2} \lim_{n\to\infty} \sum_{i=m+1}^{m+n} EU_i^2 \qquad (8) \\
&= \frac{1}{\varepsilon^2} \sum_{i=m+1}^{\infty} \frac{Var(X_i)}{i^2}. \qquad (9)
\end{aligned}$$

Under (7), (9) tends to 0 as $m \to \infty$, as required. Note that inequality (8) is Kolmogorov's inequality (4). ■

For a proof of the SLLN without assuming the variance is finite, we need another intermediate result.

Lemma. $E|X| < \infty$ iff

$$\sum_{i=1}^{\infty} P(|X| > i) < \infty. \tag{10}$$

Proof. Let $Y = |X|$. By expansion

$$\begin{aligned} EY &= \sum_{i=1}^{\infty} E\left(|X| 1\left(i-1 < Y \leq i\right)\right) \\ &\leq \sum_{i=1}^{\infty} iE\left(1\left(i-1 < Y \leq i\right)\right) \\ &= \sum_{i=1}^{\infty} iP\left(i-1 < Y \leq i\right) \\ &= \sum_{i=1}^{\infty} iP\left(Y > i-1\right) - \sum_{i=1}^{\infty} iP\left(Y > i\right) \\ &= \sum_{i=1}^{\infty} \left(i+1\right) P\left(Y > i\right) - \sum_{i=0}^{\infty} iP\left(Y > i\right) \\ &= \sum_{i=1}^{\infty} P\left(Y > i\right) \end{aligned}$$

Thus $\sum_{i=1}^{\infty} P\left(Y > i\right) < \infty$ implies $EY < \infty$. The converse can be shown similarly. ■

General Proof of SLLN. WLOG assume $EX_i = 0$. By the previous Lemma, $E|X_i| < \infty$ and X_i identically distributed implies

$$\sum_{i=1}^{\infty} P(|X_i| > i) < \infty. \tag{11}$$

The Borel-Cantelli Lemma states that (11) implies that $P\left(\{|X_i| > i\} \text{ infinitely often}\right) = 0$. This means that $\overline{X}_n \to 0$ almost surely iff

$$\frac{1}{n} \sum_{i=1}^{n} X_i 1\left(|X_i| \leq i\right) \to 0$$

almost surely, which occurs iff

$$\frac{1}{n} \sum_{i=1}^{n} \left[X_i 1\left(|X_i| \leq i\right) - E\left(X_i 1\left(|X_i| \leq i\right)\right)\right] \to 0$$

almost surely, since $EX_i = 0$ and identically distributed implies $E\left(X_i 1\left(|X_i| \leq i\right)\right) \to 0$ as $i \to \infty$, and an application of the Toeplitz Lemma yields

$$\frac{1}{n} \sum_{i=1}^{n} E\left(X_i 1\left(|X_i| \leq i\right)\right) \to 0.$$

By the almost sure convergence theorem, it is therefore sufficient to show that

$$\sum_{i=1}^{\infty} \frac{Var(X_i 1\left(|X_i| \leq i\right))}{i^2} \leq \sum_{i=1}^{\infty} \frac{E(X_i^2 1\left(|X_i| \leq i\right))}{i^2} < \infty.$$

Let

$$A_j = \sum_{i=j}^{\infty} \frac{1}{i^2} \leq \frac{2}{j}.$$

The inequality holds since for $j = 1$, $A_1 = \pi^2/6 < 2$, and for $j \geq 2$, by comparing A_j to the sum of rectangles beneath the curve x^{-2},

$$\sum_{i=j}^{\infty} \frac{1}{i^2} \leq \int_{j-1}^{\infty} x^{-2} dx = \frac{1}{j-1} \leq \frac{2}{j}.$$

Then by expanding and changing the order of summation

$$\begin{aligned}
\sum_{i=1}^{\infty} \frac{E(X_i^2 1\left(|X_i| \leq i\right))}{i^2} &= \sum_{i=1}^{\infty} \sum_{j=1}^{i} \frac{E(X_i^2 1\left(j-1 < |X_i| \leq j\right))}{i^2} \\
&= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{E(X_i^2 1\left(j-1 < |X_i| \leq j\right))}{i^2} \\
&= \sum_{j=1}^{\infty} E(X_i^2 1\left(j-1 < |X_i| \leq j\right)) A_j \\
&\leq 2 \sum_{j=1}^{\infty} \frac{E(X_i^2 1\left(j-1 < |X_i| \leq j\right))}{j} \\
&\leq 2 \sum_{j=1}^{\infty} E(|X_i| \, 1\left(j-1 < |X_i| \leq j\right)) \\
&= 2E\,|X| < \infty
\end{aligned}$$

which is what we wanted to show.