6 Strong law of large numbers

There is a zoo of strong laws of large numbers, each of which varies in the exact assumptions it makes on the underlying sequence of random variables.

Theorem 6.1 (Strong law of large numbers)**.** Let $X_1, X_2, \ldots$ be integrable real random variables i.i.d. Then

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

converges almost surely to $\mathbf{E}[X_1]$.

Proof. First, let us note that $X_1 - \mathbf{E}[X_1], X_2 - \mathbf{E}[X_2], \ldots$ are i.i.d. centered. Then, without loss of generality, we may assume that $\mathbf{E}[X_1] = 0$. Second, let $Y_n := X_n \mathbf{1}_{(|X_n| \leq n)}$ for every $n \in \mathbb{N}$ and let $h : \mathbb{R} \to \mathbb{R}$, $h(x) = |x|$. Since,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}(X_n \neq Y_n) &= \sum_{n=1}^{\infty} \mathbf{P}(|X_n| > n) \\
&= \sum_{n=1}^{\infty} \mathbf{P}_{X_1}(h > n) \\
&= \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \mathbf{P}_{X_1}(m + 1 \geq h > m) \\
&= \sum_{m=1}^{\infty} \sum_{n=1}^{m} \mathbf{P}_{X_1}(m + 1 \geq h > m) \\
&= \sum_{m=1}^{\infty} m \mathbf{P}_{X_1}(m + 1 \geq h > m) \\
&\leq \int_{\mathbb{R}} h d\mathbf{P}_{X_1},
\end{aligned}$$

by the Borel-Cantelli lemma,

$$X_n = Y_n \text{ for } n \text{ sufficiently large}$$

happens almost surely. And it therefore suffices to show that

$$\frac{Y_1 + \cdots + Y_n}{n} \xrightarrow{\text{a.s.}} 0. \tag{6.1}$$

Finally, let us note that $Y_1, Y_2, \ldots$ are

(i) independent and

(ii) square integrable, and

(iii) $\sum_{n=1}^{\infty} n^{-2} \mathbf{Var}[Y_n] < \infty$ and

(iv) $\mathbf{E}[Y_n] \to 0$.

Then, by lemma 6.2, proposition 6.1 follows. Indeed, we shall prove every assertion above:

(i) $\sigma(Y_n) \subset \sigma(X_n)$.

(ii) Since,

$$\begin{aligned}
\sum_{n=1}^{\infty} n^{-2}\mathbf{E}[Y_n^2] &= \sum_{n=1}^{\infty} n^{-2} \int_{\mathbb{R}} x^2 \mathbf{1}_{(h\leq n)}(x)\mathbf{P}_{X_1}(dx) \\
&\leq \sum_{n=1}^{\infty}\sum_{m=1}^{n} n^{-2} \int_{(m-1<h\leq m)} x^2\mathbf{P}_{X_1}(dx) \\
&\leq \sum_{n=1}^{\infty}\sum_{m=1}^{n} mn^{-2} \int_{(m-1<h\leq m)} |x|\mathbf{P}_{X_1}(dx) \\
&= \sum_{m=1}^{\infty}\sum_{n=m}^{\infty} mn^{-2} \int_{(m-1<h\leq m)} |x|\mathbf{P}_{X_1}(dx) \\
&= \sum_{m=1}^{\infty}\left(m\sum_{n=m}^{\infty} n^{-2}\right) \int_{(m-1<h\leq m)} |x|\mathbf{P}_{X_1}(dx) \\
&\leq \sum_{m=1}^{\infty} 2 \int_{(m-1<h\leq m)} |x|\mathbf{P}_{X_1}(dx) \\
&\leq 2\int_{\mathbb{R}} |x|\mathbf{P}_{X_1}(dx),
\end{aligned}$$

$$\sum_{n=1}^{\infty} n^{-2}\mathbf{E}[Y_n^2] < \infty. \tag{6.2}$$

(iii) This follows from proposition 6.2.

(iv) Let $f_n : \mathbb{R} \to \mathbb{R}$, $f(x) = x\mathbf{1}_{(h\leq n)}(x)$, and $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$. Since, $f_n \to f$, by the dominated convergence theorem, $\mathbf{E}[Y_n] = \int_{\mathbb{R}} f_n d\mathbf{P}_{X_1} \to \int_{\mathbb{R}} f d\mathbf{P}_{X_1} = 0$. □

Lemma 6.2. Let $X_1, X_2, \ldots$ be square integrable and independent real random variables with $\sum_{n=1}^{\infty} n^{-2}\mathbf{Var}[X_n] < \infty$ and

$$\frac{\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n]}{n} \to a \in \mathbb{R}.$$

Then

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{a.s.}} a.$$

Proof. Let $S_n := \sum_{k=1}^{n}(X_k - \mathbf{E}[X_k])$. Fix $\epsilon > 0$. For every $k \in \mathbb{N}$, let A_k be the event where

$$n^{-1}|S_n| \geq \epsilon \quad \text{for some } n \text{ with } 2^{k-1} \leq n < 2^k.$$

Then on A_k we have

$$|S_n| \geq \epsilon 2^{k-1} \quad \text{for some } n < 2^k,$$

so by Kolmogorov's inequality,

$$\mathbf{P}(A_k) \leq (\epsilon 2^{k-1})^2 \sum_{n=1}^{2^k} \mathbf{Var}[X_n].$$

Therefore,

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbf{P}(A_k) &\leq \frac{4}{\epsilon^2} \sum_{k=1}^{\infty} \sum_{n=1}^{2^k} 2^{-2k} \mathbf{Var}[X_n] \\
&= \frac{4}{\epsilon^2} \sum_{n=1}^{\infty} \sum_{k=\log_2 n}^{\infty} 2^{-2k} \mathbf{Var}[X_n] \\
&\leq \frac{8}{\epsilon^2} \sum_{n=1}^{\infty} n^{-2} \mathbf{Var}[X_n],
\end{aligned}$$

so

$$\mathbf{P}\left(\limsup_{k \to \infty} A_k \right) = 0$$

by the Borel-Cantelli lemma. But $\limsup A_k$ is precisely the set where

$$n^{-1}|S_n| \geq \epsilon \quad \text{for infinitely many } n,$$

so

$$\mathbf{P}\left(\limsup_{n \to \infty} n^{-1}|S_n| < \epsilon \right) = 1.$$

Letting $\epsilon \to 0$ through a countable sequence of values, we have that

$$\frac{X_1 + \cdots + X_n}{n} - \frac{\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n]}{n} \xrightarrow{\text{a.s.}} 0,$$

and the conclusion follows. $\square$

Remark 6.3. The core of the weak law of large numbers is Chebyshev's inequality. Here we present a stronger inequality that claims the same bound but now for the maximum over all partial sums until a fixed time.

Lemma 6.4 (Kolmogorov's inequality)**.** Let $X_1, \ldots, X_n$ be square integrable, independent and centered real random variables; let $S_k = X_1 + \cdots + X_k$ for $k = 1, \ldots, n$. Then, for any $\epsilon > 0$,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon \right) \leq \epsilon^{-2} \mathbf{Var}[S_n]. \tag{6.3}$$

Proof. First, fix $\epsilon > 0$. We decompose the probability space according to the first time τ at which the partial sums exceed the value ϵ. Hence, let

$$\tau := \min\{k = 1, \ldots, n \; ; \; |S_k| \geq \epsilon\}$$

and $A_k := (\tau = k)$ for $k = 1, \ldots, n$. Further, let

$$A := \left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon \right) = \biguplus_{k=1}^{n} A_k.$$

The random variables $S_n - S_k$ and $S_k \mathbf{1}_{A_k}$ are $\sigma(X_{k+1}, \ldots, X_n)$ and $\sigma(X_1, \ldots, X_k)$ measurable, and thus

$$\mathbf{E}[(S_n - S_k) S_k \mathbf{1}_{A_k}] = \mathbf{E}[S_n - S_k] \mathbf{E}[S_k \mathbf{1}_{A_k}] = 0.$$

Then

$$\begin{aligned}
\mathbf{Var}[S_n] &= \mathbf{E}[S_n^2] \\
&\geq \mathbf{E}\left[\sum_{k=1}^{n} S_n^2 \mathbf{1}_{A_k} \right] \\
&= \sum_{k=1}^{n} \mathbf{E}[S_n^2 \mathbf{1}_{A_k}] \\
&= \sum_{k=1}^{n} \mathbf{E}[((S_n - S_k)^2 + 2(S_n - S_k) S_k + S_k^2) \mathbf{1}_{A_k}] \\
&= \sum_{k=1}^{n} \mathbf{E}[(S_n - S_k)^2 \mathbf{1}_{A_k}] + \sum_{k=1}^{n} \mathbf{E}[S_k^2) \mathbf{1}_{A_k}] \\
&\geq \sum_{k=1}^{n} \mathbf{E}[S_k^2 \mathbf{1}_{A_k}] \\
&\geq \sum_{k=1}^{n} \mathbf{E}[\epsilon^2 \mathbf{1}_{A_k}] \\
&= \epsilon^2 \mathbf{P}(A),
\end{aligned}$$

and inequality 6.3 follows. □

Example 6.5 (Monte Carlo Integration)**.** ...

Definition 6.6 (Empirical distribution function)**.** Let $X_1, X_2, \ldots$ be real random variables. The map $F_n : \mathbb{R} \to [0, 1]$, $x \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{]-\infty, x]}(X_i)$ is called the **empirical distribution function** of $X_1, X_2, \ldots$. □

Theorem 6.7 (Glivenko–Cantelli)**.** Let $X_1, X_2, \ldots$ be i.i.d. real random variables with distribution function F, and let F_n, $n \in \mathbb{N}$, be the empirical distribution functions. Then

$$\limsup_{n \to +\infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \quad \text{a.s.}$$

Proof. Exercise. □

Example 6.8 (Shannon's theorem)**.** ...

Exercise 6.1. Let $X_1, X_2, \ldots$ be i.i.d. real random variables with

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{a.s.}} Y$$

for some random variable Y. Show that $X_1 \in \mathcal{L}^1(\mathbf{P})$ and $Y = \mathbf{E}[X_1]$ a.s. (Hint: first show that

$$\mathbf{P}(|X_n| > n \text{ for infinitely many } n) = 0 \quad \Leftrightarrow \quad X_1 \in \mathcal{L}^1(\mathbf{P}).)$$

Exercise 6.2. Let E be a finite set and let p be a probability vector on E. Show that the entropy $H(p)$ is minimal (in fact, zero) if $p = \delta_e$ for some $e \in E$. It is maximal (in fact, $log(\#E)$) if p is the uniform distribution on E.

Exercise 6.3. Let $X_1, X_2, \ldots$ be independent and centered real random variables with $\sum_{n=1}^{\infty} \mathbf{Var}[X_n] < \infty$. Prove that $(X_1 + \cdots + X_n)$ converges almost surely. (hint: apply Kolmogorov's inequality to show that the partial sums are Cauchy almost surely.)

Exercise 6.4. If the plus and minus signs in $\sum_{n=1}^{\infty} \pm n^{-1}$ are determined by successive tosses of a fair coin, prove that the resulting series converges almost surely.

Exercise 6.5. Let $X_1, X_2, \ldots$ be real random variables i.i.d. that are not integrable. Prove that

$$\limsup_{n\to\infty} n^{-1}|X_1 + \cdots + X_n| \xrightarrow{\text{a.s.}} \infty.$$

(Hint: show that $\sum_{n=1}^{\infty} \mathbf{P}(|X_n| > n) = \infty$ and apply the Borel-Cantelli lemma.)

Exercise 6.6. A collection or "population" of N objects (such as mice, grains of sand, etc.) may be considered as a sample space in which each object has probability N^{-1}. Let X be a random variable on this space (a numerical characteristic of the objects such as mass, diameter, etc.) with mean m and variance v. In statistics one is interested in determining m and v by taking a sequence of random samples from the population and measuring X for each sample, thus obtaining a sequence (X_n) of numbers that are values of independent random variables with the same distribution as X. The nth sample mean is $M_n = n^{-1} \sum_{i=1}^{n} X_n$ and the nth sample variance is $V_n = (n-1)^{-1} \sum_{i=1}^{n} (X_i - M_i)^2$.

(i) Show that $\mathbf{E}[M_n] = m$, $\mathbf{E}[V_n] = v$, and $M_n \xrightarrow{\text{a.s.}} m$ and $V_n \xrightarrow{\text{a.s.}} v$.

(ii) Can you see why one uses $(n-1)^{-1}$ instead of $n - 1$ in the definition of V_n?