Strong Law of Large Numbers

We'll prove the strong law of large numbers in 4 steps. We state the theorem below.

Strong Law of Large Numbers. *Let $X_1, X_2, ...$ be iid random variables and assume $\mathbb{E}[|X_1|] < \infty$. Let $S_n := X_1 + ... + X_n$. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}[X_1].$$

Step 1 (Kolmogorov's inequality).

Theorem. Let $X_1, ..., X_n$ be mutually independent random variables. Suppose $\sigma_i^2 := \mathbb{E}[X_i^2] < \infty$ and $\mathbb{E}[X_i] = 0$. Let $\lambda > 0$. Then

$$\mathbb{P}\left(\max_{1 \le i \le n} |X_1 + ... + X_i| \ge \lambda\right) \le \frac{\sum_{i=1}^n \sigma_i^2}{\lambda^2}.$$

Proof. Note that this is a stronger version of Chebyshev's inequality. It bounds the probability of the absolute value of the maximum partial sum exceeding a given value. Let $S_k := \sum_{i=1}^k X_i$. And $A := \{\max_{1 \le k \le n} |S_k| \ge \lambda\} = \bigcup_{k=1}^n \{|S_k| \ge \lambda\}$ i.e. the event that the maximum partial sum excedes λ. Now let $A_k := \{\max_{1 \le i \le k-1} |S_i| < \lambda, |S_k| \ge \lambda\}$ i.e. the event that the largest partial sum over indices less than or equal to k exceeds λ no sooner than the k partial. Then $A = \bigcup_{k=1}^n A_k$. Also note that $A_k \cap A_\ell = \emptyset$ for $k \ne \ell$. So

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A_k) = \mathbb{E}[\mathbb{1}_{\{\omega \in A_k\}}] \le \sum_{k=1}^n \mathbb{E}\left[\frac{S_k}{\lambda^2} \cdot \mathbb{1}_{\{\omega \in A_k\}}\right] = \frac{1}{\lambda^2} \sum_{k=1}^n \mathbb{E}[S_k^2 \cdot \mathbb{1}_{\{\omega \in A_k\}}].$$

Now just consider the summation. We have

$$\begin{aligned}
\sum_{k=1}^n \mathbb{E}[S_k^2 \cdot \mathbb{1}_{\{\omega \in A_k\}}] &\le \sum_{k=1}^n \left[\mathbb{E}[S_k^2 \cdot \mathbb{1}_{\{\omega \in A_k\}}] + \mathbb{E}[(S_n - S_k)^2 \cdot \mathbb{1}_{\{\omega \in A_k\}}]\right] && \text{(second term is non-negative)} \\
&= \sum_{k=1}^n \mathbb{E}[S_n^2] \cdot \mathbb{1}_{\{\omega \in A_k\}} && (**) \\
&= \mathbb{E}[S_n^2 \sum_{i=1}^n \mathbb{1}_{\{\omega \in A_k\}}] \\
&= \mathbb{E}[S_n^2 \cdot \mathbb{1}_{\{\omega \in A\}}] \\
&\le \mathbb{E}[S_n^2] && \text{(monotonicity of integral)} \\
&= \sum_{i=1}^n \sigma_i^2 && \text{(independence and mean zero).}
\end{aligned}$$

Now we explain (**) since it's not immediately obvious. We focus on the two terms in the sum. We have

$$\begin{aligned}
\mathbb{E}[S_k^2 \cdot \mathbb{1}_{\{\omega \in A_k\}}] + \mathbb{E}[(S_n - S_k)^2 \cdot \mathbb{1}_{\{\omega \in A_k\}}] &= \mathbb{E}\left[[S_k^2 + (S_n - S_k)^2] \cdot \mathbb{1}_{\{\omega \in A_k\}}\right] && \text{(linearity of expectation)} \\
&= \mathbb{E}[S_n^2 + 2S_k(S_k - S_n)] && \text{(expanding everything)} \\
&= \mathbb{E}[S_n^2] + \mathbb{E}[2S_k(S_k - S_n)] && \text{(linearity of expectation)} \\
&= \mathbb{E}[S_n^2] + \mathbb{E}[2S_k]\mathbb{E}[S_k - S_n] && (S_k \text{ and } (S_k - S_n) \text{ independent}) \\
&= \mathbb{E}[S_n^2] && \text{(each partial has expectation 0)}
\end{aligned}$$

Putting everything together, we have

$$\mathbb{P}\left(\max_{1 \le i \le n} |X_1 + ... + X_i| \ge \lambda\right) \le \frac{\sum_{i=1}^n \sigma_i^2}{\lambda^2}.$$

□

Step 2 (Infinite sum of finite, zero-mean random variables exists and is finite).

Theorem. Let $X_1, X_2, ...$ be mutually independent random variables, and assume $\mathbb{E}[X_i] = 0$, and $\sigma_i^2 := \mathbb{E}[X_i^2] < \infty$, and $\sum_{i=1}^n \sigma_i^2 < \infty$. Then $\lim_{n\to\infty} \sum_{i=1}^n X_i$ exists and is finite almost surely.

Proof. Let $S_n := X_1 + ... + X_n$. We will prove that the sequence $\{S_n\}_{n\geq 1}$ is a Cauchy sequence with probability 1. Define the event $A_{N,r} := \left\{\exists i, j \geq N : |S_i - S_j| \geq \frac{1}{r}\right\}$ Then event that $\{S_n\}_{n\geq 1}$ is not Cauchy is

$$\bigcup_{r=1}^{\infty} \bigcap_{N=1}^{\infty} A_{N,r}.$$

Note that $A_{N,r}$ is increasing in r and decreasing in N, so we can use sequential continuity of measure. We have

$$\begin{aligned} \mathbb{P}\left(\{S_n\}_{n\geq 1} \text{ is not Cauchy}\right) &= \mathbb{P}\left(\bigcup_{r=1}^{\infty} \bigcap_{N=1}^{\infty} A_{N,r}\right) \\ &= \lim_{r\to\infty} \lim_{N\to\infty} A_{N,r}. \end{aligned}$$

Now we show for every $r \geq 1$ we have $\lim_{N\to\infty} A_{N,r} = 0$. Fix $r \geq 1$, and define $B_{N,r} := \left\{\exists i \geq N : |S_i - S_N| \geq \frac{1}{2r}\right\}$. Then

$$\begin{aligned} \mathbb{P}\left(\left\{\exists i \geq N : |S_i - S_N| \geq \frac{1}{2r}\right\}\right) &= \mathbb{P}\left(\bigcup_{N=N'}^{\infty} \left\{\exists N \leq i \leq N' : |S_i - S_N| \geq \frac{1}{2r}\right\}\right) \\ &= \lim_{N'\to\infty} \mathbb{P}\left(\max_{N\leq i\leq N'} |X_{N+1} + X_{N+2} + ... + X_{N'}| \geq \frac{1}{2r}\right) \\ &\leq \lim_{N'\to\infty} 4r^2 \sum_{i=N}^{N'} \sigma_i^2 \\ &= 4r^2 \sum_{i=N}^{\infty} \sigma_i^2. \end{aligned}$$

The second line follows from Kolmogorov's inequality shown in Step 1. Hence

$$\lim_{N\to\infty} \mathbb{P}(B_{N,r}) \leq \lim_{N\to\infty} 4r^2 \sum_{i=N}^{\infty} \sigma_i^2 = 0$$

because the sum of the variances converges by assumption, and therefore its tail sums go to 0. To conclude, note that $A_{N,r} \subseteq B_{N,r}$ by the triangle inequality. By monotonicity, we have $\lim_{N\to\infty} \mathbb{P}(A_{N,r}) = 0$. Equivalently,

$$\mathbb{P}\left(\{S_n\}_{n\geq 1} \text{ is Cauchy}\right) = 1.$$

$\square$

Step 3 (SLLN with special condition)

First we state a useful lemma that will help us prove the strong law of large numbers when we have summable variances. Then we will use it to prove a version of the strong law of large numbers with a special case.

Lemma. Suppose $\{a_n\}_{n\geq 1}$ is a sequence such that $\sum_{k=1}^{\infty} \frac{a_k}{k} < \infty$. Then $\frac{1}{n}\sum_{i=1}^{\infty} a_k \xrightarrow{n\longrightarrow\infty} 0$.

SLLN with summable variances. Let $X_1, X_2, ...$ be mutually independent random variables, and define $\mu_i := \mathbb{E}[X_i], \sigma_i^2 := \mathrm{Var}(X_i) < \infty$, and $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{\infty}(X_i - \mu_i) \xrightarrow{\text{a.s.}} 0.$$

Proof. Define $Y_i := \frac{X_i - \mu_i}{i}$. Then $\mathbb{E}[Y_i] = 0$, and $\mathrm{Var}(Y_i) = \frac{\sigma_i^2}{i^2}$. By assumption, we have $\sum_{i=1}^{\infty} \mathrm{Var}(Y_i) = \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty$. By Step 2, $\sum_{i=1}^{\infty} Y_i < \infty$ almost surely. Thus $\sum_{i=1}^{\infty} Y_i = \sum_{i=1}^{\infty} \frac{X_i - \mu_i}{i} < \infty$. Hence, by Kronecker's Lemma, it follows that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{\infty}(X_i - \mu_i) = 0.$$

□

Step 4 (General case of SLLN)

Now we can prove SLLN in the more general case where we do not require the variances to be summable as they are in Step 3.

Strong Law of Large Numbers without variance condition. Assume $X_1, X_2, ...$ are iid random variables with $\mathbb{E}[|X_1|] < \infty$ and $\mathbb{E}[X_1] = 0$. Let $S_n := X_1 + ... + X_n$. Then

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}[X_1].$$

Proof. We use truncation to prove the statement. Define $Y_k := X_k \cdot \mathbb{1}_{\{|X_k|\leq k\}}$ and $Z_k := X_k \cdot \mathbb{1}_{\{|X_k|>k\}}$. Then $X_k = Y_k + Z_k$. So

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} Y_i + \frac{1}{n}\sum_{i=1}^{n} Z_i.$$

We want to show that both sums on the right tend to zero as n tends to ∞. We start with the term summing Z_i. Our goal is to show that only finitely many Z_k are non-zero. This way, the entire term tends to 0. We use Borel-Cantelli. Define $A_k := \{Z_k \neq 0\}$. Then

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbb{P}(A_k) &= \sum_{k=1}^{\infty} \mathbb{P}(X_k > k) + \mathbb{P}(X_k < -k) \\
&= \sum_{k=1}^{\infty} F(-k) + (1 - F(k)) \\
&\leq \sum_{k=1}^{\infty} \left[\int_{-k}^{-(k-1)} F(y)\, dy + \int_{k}^{k+1} (1 - F(y))\, dy\right] && \text{(monotonicity of integral)} \\
&= \int_{-\infty}^{0} F(y)\, dy + \int_{0}^{\infty} (1 - F(y))\, dy \\
&= -\int_{-\infty}^{0} x dF(x) + \int_{0}^{\infty} x dF(x) && \text{(integration by parts)} \\
&= \mathbb{E}[|X_1|] < \infty && \text{(by assumption).}
\end{aligned}$$

Hence, by Borel-Cantelli, finitely many Z_k are non-zero almost surely, and thus $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n Z_i = 0$ almost surely. Now we can show that $\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{\text{a.s.}} 0$. This is equivalent to showing

$$\left[\frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) + \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i]\right] \xrightarrow{\text{a.s.}} 0.$$

We want to show that $\sum_{i=1}^\infty \frac{Y_i}{i^2} < \infty$ so that by Step 3 the first term in the sum above goes to 0. Define

$$a_n := \int_{n-1}^n x dF(x) - \int_{-n}^{-n+1} x dF(x).$$

As shown above, $\sum_{n=1}^\infty a_n = \mathbb{E}[|X_1|] < \infty$. We have

$$\int_{n-1}^n x^2 dF(x) + \int_{-n}^{-n+1} x^2 dF(x) \leq \int_{n-1}^n nx dF(x) - \int_{-n}^{-n+1} nx dF(x) = n a_n.$$

It follows that

$$\begin{aligned}
\sum_{k=1}^\infty \frac{\mathrm{Var}(Y_k)}{k^2} &\leq \sum_{k=1}^\infty \frac{\mathbb{E}[Y_k^2]}{k^2} \\
&= \sum_{k=1}^\infty \frac{1}{k^2}\int_{-k}^k x^2 dF(x) \\
&= \sum_{k=1}^\infty \frac{1}{k^2} \cdot \sum_{\ell=1}^k \left[\int_{\ell-1}^{\ell=1} x^2 dF(x) + \int_{-\ell}^{-\ell+1} x^2 dF(x)\right] \\
&\leq \sum_{k=1}^\infty \frac{1}{k^2} \cdot \sum_{\ell=1}^k \ell a_\ell && \text{(how we've defined } a_n\text{)} \\
&= \sum_{\ell=1}^\infty \ell a_\ell \sum_{k=\ell}^\infty \frac{1}{k^2} && \text{(interchanging sums)} \\
&\leq \sum_{\ell=1}^\infty \ell a_\ell \frac{c}{\ell} && \text{(each tail sum proportional to } \tfrac{1}{\ell}\text{)} \\
&= c \sum_{\ell=1}^\infty a_\ell \\
&= c\, \mathbb{E}[|X_1|] \\
&< \infty && \text{(by assumption).}
\end{aligned}$$

Above, c is some constant bounding the tail sums from above. Thus, $\frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \xrightarrow{\text{a.s.}} 0$ as desired. Now we need to take care of the $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i]$ term. This average should converge to 0 almost surely because Y_i tends to X_1 as $i \to \infty$ and $\mathbb{E}[X_1] = 0$. Formally, we have

$$\begin{aligned}
\lim_{k\to\infty} \mathbb{E}[Y_k] &= \lim_{k\to\infty} \int_{-k}^k x \; dF(x) \\
&= \lim_{k\to\infty} \int_{-\infty}^\infty x \; \mathbb{1}_{\{|X_k| \leq k\}} \; dF(x) \\
&= \int_{-\infty}^\infty \lim_{k\to\infty} x \; \mathbb{1}_{\{|X_k| \leq k\}} \; dF(x) && \text{(Dominated Convergence Theorem)} \\
&= \int_{-\infty}^\infty x \; dF(x) \\
&= \mathbb{E}[X_1] \\
&= 0.
\end{aligned}$$

It can be shown that for a sequence $\{a_n\}_{n\geq 1}$, if $a_n \longrightarrow 0$, then $\lim_{n\to\infty} \frac{1}{n}\sum_{n=1}^\infty a_n = 0$. It follows that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i] \xrightarrow{n\to\infty} 0 \text{ almost surely.}$$

$\square$