1 Prerequisites

Random variables, independence, expectation, Kolmogorov’s maximal inequality, Borel-Cantelli lemmas

2 Summary

Two classic theorems concerning the almost sure convergence of a sum of independent random variables are stated and proved. The $L^2$ convergence theorem assumes the random variables are independent and in $L^2$ to arrive at convergence both almost surely and in $L^2$. The Strong Law of Large Numbers assumes the variables are IID with finite mean only, and shows the average converges almost surely.

3 Strong Law of Large Numbers

**Theorem 1 (Basic $L^2$ Convergence Theorem)** Let $X_1, X_2, \ldots$ be independent random variables with $E(X_i) = 0$ and $E(X_i^2) = \sigma_i^2 < \infty$, $i = 1, 2, \ldots$, and $S_n = X_1 + X_2 + \cdots + X_n$. If $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$, then $S_n$ converges a.s. and in $L^2$ to some $S_\infty$ with $E(S_\infty^2) = \sum_{i=1}^{\infty} \sigma_i^2$.

*Remark:* The proof uses Kolmogorov’s maximal inequality. Thus, the conclusion is valid for a martingale $\{S_n\}$ with $E[X_{n+1} f(X_1, \ldots, X_n)] = 0$ for all bounded measurable $f : \mathbb{R}^n \to \mathbb{R}$. 
**Proof:** First note that $L^2$ convergence and existence of $S_\infty$ is implied by the orthogonality of the $X_i$'s: since $E(X_i X_j) = 0$ for $i \neq j$,

$$
E(S_n^2) = \sum_{i=1}^{n} \sigma_i^2, \text{ and}$$

$$
E((S_n - S_m)^2) = \sum_{i=m+1}^{n} \sigma_i^2 \to 0 \text{ a.s. } m, n \to \infty,
$$

so $S_n$ is Cauchy in $L^2$. Since $L^2$ is complete, there is a unique $S_\infty$ (up to a.s. equivalence) such that $S_n \to S_\infty$ in $L^2$.

Turning to a.s. convergence, show the sequence $(S_n)$ is a.s. Cauchy. The limit of $S_n$ then exists a.s. by completeness of the set of real numbers. The same argument applies more generally to martingale differences $X_i$. Note that this method gives $S_\infty$ more explicitly, and does not appeal to completeness of $L^2$.

Recall that $S_n$ is Cauchy a.s. means $M_n := \sup_{p,q \geq n} |S_p - S_q| \to 0$ a.s. Note that $0 \leq M_n(\omega) \downarrow$ implies that $M_n(\omega)$ converges to a limit in $[0, \infty)$. So, if $P(M_n > \epsilon) \to 0$ for all $\epsilon > 0$, then $M_n \downarrow 0$ a.s.

Let $M_n^* := \sup_{p \geq n} |S_p - S_n|$. By the triangle inequality,

$$
|S_p - S_q| \leq |S_p - S_n| + |S_q - S_n| \Rightarrow M_n^* \leq M_n \leq 2M_n^*,
$$

so it is sufficient to show that $M_n^* \to 0$ in probability.

For all $\epsilon > 0$,

$$
P\left(\sup_{p \geq n} |S_p - S_n| > \epsilon \right) = \lim_{N \to \infty} P\left(\max_{n \leq p \leq N} |S_p - S_n| > \epsilon \right)
\leq \lim_{N \to \infty} \sum_{i=n+1}^{N} \frac{\sigma_i^2}{\epsilon^2} = \sum_{i=n+1}^{\infty} \frac{\sigma_i^2}{\epsilon^2},
$$

where we applied Kolmogorov’s inequality in the second step. Since $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$,

$$
\lim_{n \to \infty} P\left(\sup_{p \leq n} |S_p - S_n| > \epsilon \right) = 0
$$

**Remark:** Just orthogonality rather than independence of the $X_i$s is not enough to get an a.s. limit. Counterexamples are hard. According to classical results of Rademacher-Menchoff, for orthogonal $X_i$ the condition $\sum_i (\log^2 i) \sigma_i^2 < \infty$ is enough for a.s. convergence of $S_n$, whereas if $b_i \uparrow$ with $b_i = o(\log^2 i)$, there exist orthogonal $X_i$ such that $\sum_i b_i \sigma_i^2 < \infty$ and $S_n$ diverges almost surely.
Suppose the $X_i$ are IID with finite mean only. We conclude that the average converges almost surely to the mean. To prove this, use Kronecker’s Lemma, which is stated below, but proved in Section 1.8 of Durrett.

**Lemma 1 (Kronecker’s Lemma)** Suppose $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n$ converges then

$$a_n^{-1} \sum_{m=1}^{n} x_m \to 0$$

**Theorem 2 (Strong Law of Large Numbers)** Let $X, X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}(|X|) < \infty$, $S_n = X_1 + \ldots + X_n$. Then $S_n/n \to \mathbb{E}(X)$ a.s. as $n \to \infty$.

The theorem is true with just pairwise independence instead of the full independence assumed here [Durrett, p.55 (7.1)]. The theorem also has an important generalization to stationary sequences (the *ergodic theorem*, [Durrett, p.337 (2.1)]).

**Proof:** *Step 1:* Replace $X_i$ by $\tilde{X}_i = X_i - \mathbb{E}X$ (note $\mathbb{E}X_i = \mathbb{E}X$). Then

$$\frac{\tilde{S}_n}{n} = \frac{S_n}{n} - \mathbb{E}X.$$ 

So it’s enough to consider $\mathbb{E}X = 0$.

*Step 2:* Now assume $\mathbb{E}X = 0$. Introduce truncated variables

$$\tilde{X}_n := 1(|X_n| \leq n)$$

Observe that

$$\mathbb{P}(X_n = \tilde{X}_n \text{ ev.}) = 1.$$ 

To see this, check

$$\mathbb{P}(X_n \neq \tilde{X}_n \text{ i.o.}) = \mathbb{P}(|X_n| > n \text{ i.o.})$$

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) = \mathbb{E} \left( \sum_{n=1}^{\infty} 1(|X| > n) \right) < \infty$$

since

$$\sum_{n=1}^{\infty} 1(|X| > n) = \sum_{1 \leq n < |X|} 1 \leq |X| + 1$$

This argument is similar to the tail sum formula for the expectation of a random variable $X$ with values in $0, 1, 2, \ldots$

$$\mathbb{E}X = \sum_{n=1}^{\infty} n \mathbb{P}(X = n) = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$$
Step 3: Center the truncated variables. Define \( \tilde{X}_n := \tilde{X}_n - \mathbb{E}(\tilde{X}_n) \). Show that

\[
\left( \frac{S_n}{n} \to 0 \right) \overset{a.s.}{\Rightarrow} \left( \frac{\tilde{S}_n}{n} \to 0 \right) \overset{a.s.}{\Rightarrow} \left( \frac{\tilde{S}_n}{n} \to 0 \right),
\]

where \( \tilde{S}_n = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n \) and \( \tilde{S}_n = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n \). Show that \( \mathbb{P}(\tilde{S}_n/n \to 0) = 1 \).

(a) comes from the fact that if \( \omega \in \left\{ \omega : X_n(\omega) = \hat{X}_n(\omega) \ \text{ev.} \right\} \) (which has probability 1), then \( S_n(\omega) - \hat{S}_n(\omega) \) is eventually not dependent on \( n \). So

\[
\frac{S_n(\omega) - \hat{S}_n(\omega)}{n} \to 0 \quad \text{for such } \omega.
\]

(b) comes from

\[
\frac{\hat{S}_n - \tilde{S}_n}{n} = \frac{\mathbb{E}\hat{X}_1 + \mathbb{E}\hat{X}_2 + \cdots + \mathbb{E}\hat{X}_n}{n} \to 0 \quad \text{as } n \to \infty \quad \text{(By analysis and } \mathbb{E}\hat{X}_i \to 0)\
\]

But

\[
\mathbb{E}\hat{X}_n = \mathbb{E}[X_n1(|X_n| \leq n)] = \mathbb{E}[X1(|X| \leq n)] \to EX \quad \text{as } n \to \infty,
\]

as the integrand is dominated by \( |X| \) and note \( \mathbb{E}(|X|) < \infty \).

Use Kronecker’s lemma and the \( L^2 \) convergence theorem to show that

\[
\sum_{n=1}^\infty \frac{\mathbb{E}(\tilde{X}_n^2)}{n^2} < \infty.
\]

\[
\mathbb{E}(\tilde{X}_n^2) = \mathbb{E}\left[\left(\tilde{X}_n - \mathbb{E}(\tilde{X}_n)\right)^2\right] = \mathbb{E}\left[(X1(|X| \leq n) - \mathbb{E}(X1(|X| \leq n)))^2\right] \leq \mathbb{E}(X1(|X| \leq n))^2.
\]

So

\[
\sum_{n=1}^\infty \frac{\mathbb{E}(\tilde{X}_n^2)}{n^2} \leq \sum_{n=1}^\infty \frac{\mathbb{E}X^21(|X| \leq n)}{n^2} = \mathbb{E}\left(X^2\sum_{n=1}^\infty \frac{1(|X| \leq n)}{n^2}\right) \approx \mathbb{E}\left(\frac{X^2}{|X|}\right) = \mathbb{E}(|X|) < \infty.
\]

This came from

\[
\sum_{n=1}^\infty \frac{x^21(|x| \leq n)}{n^2} = x^2 \sum_{n=|x|}^\infty \frac{1}{n^2} \approx x^2 \frac{1}{|x|} = |x|
\]
4 References