36-752 Advanced Probability Overview	Spring 2018

9. Almost Sure Convergence and Strong Law of Large Numbers

Instructor: Alessandro Rinaldo

Associated reading: Sec 6.1 and 6.2 of Ash and Doléans-Dade; Sec 2.3–2.5 of Durrett.

Overview

Let $\{X_i : i \geq 1\}$ be i.i.d random variables with $-\infty < \mathrm{E}X_1 < \infty$. WLLN says that the partial average $(X_1 + X_2 + ... + X_n)/n$ converges to $\mathrm{E}X_1$ in probability. In fact, one can prove a stronger result: $(X_1 + X_2 + ... + X_n)/n$ converges to μ almost surely.

We start with Kolmogorov's 0-1 law and the notion of tail σ-field.

Theorem 1 (Kolmogorov 0-1 law). *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random quantities. Define $\mathcal{T}_n = \sigma(\{X_i : i \geq n\})$ and $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$. Then every event in $\mathcal{T}$ has probability either 0 or 1.*

Proof: Let $\mathcal{U}_n = \sigma(\{X_i : i \leq n\})$, and let $\mathcal{U} = \bigcup_{n=1}^{\infty} \mathcal{U}_n$. Let $A \in \mathcal{U}$ and $B \in \mathcal{T}$. There exists n such that $A \in \mathcal{U}_n$. Because $B \in \mathcal{T}_{n+1}$, it follows that A and B are independent. So $\mathcal{U}$ and $\mathcal{T}$ are independent. It follows from Proposition 19 of Lecture Notes Set 4 that $\sigma(\mathcal{U}) = \sigma(\{X_n\}_{n=1}^{\infty})$ and $\mathcal{T}$ are independent. Since $\mathcal{T} \subseteq \sigma(\mathcal{U})$, it follows that $\mathcal{T}$ is independent of itself, hence for all $B \in \mathcal{T}$, $\Pr(B) \in \{0, 1\}$ because $P(B) = P(B \cap B) = P(B)P(B)$. ■

Definition 2. *The σ-field $\mathcal{T}$ in Theorem 1 is called the* tail σ-field *of the sequence $\{X_n\}_{n=1}^{\infty}$.*

Now consider the event $A \equiv \{\omega : (X_1 + X_2 + ... + X_n)/n \text{ converges}\}$. Then it is easy to check that $A \in \mathcal{T}$, and hence $P(A) = 0$ or 1 by Kolmogorov's 0-1 law. According to WLLN, we shall conjecture that $P(A) = 1$.

1 Preliminaries and Borel Cantelli Lemmas

Definition 3 (i.o. and ev.). *Let q_n be some statement, true or false for each n. We say q_n happens* infinitely often *or (q_n i.o.) if for all n there is $m \geq n$ such that q_m is true, and (q_n ev.) if there exists n such that for all $m \geq n$, q_m is true. Now consider probability space $(\Omega, \mathcal{F}, P)$ and let q_n depend on $\omega \in \Omega$, giving events*

$$A_n = \{\omega : q_n(\omega) \text{ is true}\}.$$

We now have new events,

$$\{A_n \ i.o.\} = \{\omega : q_n(\omega) \ i.o.\} = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m \equiv \lim \sup_{n\to\infty} A_n \,,$$

and

$$\{A_n \ ev.\} = \{\omega : q_n(\omega) \ ev.\} = \bigcup_{n\geq 1} \bigcap_{m\geq n} A_m \equiv \lim \inf_{n\to\infty} A_n \,.$$

Useful facts.

1. Given a sequence of events A_n, the sequence $(\mathbf{1}_{A_n}(\omega) : n \geq 1)$ can be viewed as a function of $\omega \longmapsto \{0,1\}^{\mathbb{Z}^+}$.
2. $\mathbf{1}_{(A_n \text{ i.o.})} = \limsup_{n\to\infty} \mathbf{1}_{A_n}$ and $\mathbf{1}_{(A_n \text{ ev.})} = \liminf_{n\to\infty} \mathbf{1}_{A_n}$.
3. **(de Morgan)** $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ ev.}\}$ and $\{A_n \text{ ev.}\}^c = \{A_n^c \text{ i.o.}\}$
4. $a_n \to a \Longleftrightarrow \forall \epsilon > 0,\ |a_n - a| < \epsilon$ ev., so

$$\begin{aligned} X_n \stackrel{\text{a.s.}}{\to} X \quad &\Longleftrightarrow \quad \forall \epsilon > 0,\ \Pr(|X_n - X| \leq \epsilon \text{ ev.}) = 1 \\ &\Longleftrightarrow \quad \forall \epsilon > 0,\ \Pr(|X_n - X| > \epsilon \text{ i.o.}) = 0. \end{aligned}$$

   (in the second "$\Leftrightarrow$", showing "$\Rightarrow$" is trivial but "$\Leftarrow$" is less trivial.)

Exercise 4. $X_n \stackrel{\text{a.s.}}{\to} 0 \Longleftrightarrow \sup_{k\geq n} |X_k| \stackrel{P}{\to} 0$.

Next we present a basic tool in the study of almost sure convergence.

Theorem 5 (First Borel-Cantelli lemma). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then $\mu(\limsup_{n\to\infty} A_n) = 0$ or equivalently, $\mu(A_n \ i.o.) = 0$.*

Proof: Let $B_i = \bigcup_{n=i}^{\infty} A_n$. Then $\{B_i\}_{i=1}^{\infty}$ is a decreasing sequence of sets, each of which has finite measure, so by continuity of measure we have

$$\lim_{i\to\infty} \mu(B_i) = \mu\left(\lim_{i\to\infty} B_i\right) = \mu\left(\bigcap_{i=1}^{\infty} B_i\right) = \mu\left(\limsup_{n\to\infty} A_n\right).$$

Since $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, it follows that $\lim_{i\to\infty} \sum_{n=i}^{\infty} \mu(A_n) = 0$. Since $\mu(B_i) \leq \sum_{n=i}^{\infty} \mu(A_n)$, $\lim_{i\to\infty} \mu(B_i) = 0$, and the result follows. ■

Theorem 6 (Second Borel-Cantelli lemma). *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $\sum_{n=1}^{\infty} P(A_n) = \infty$ and if $\{A_n\}_{n=1}^{\infty}$ are mutually independent, then $P(\limsup_{n\to\infty} A_n) = 1$ or equivalently, $P(A_n \ i.o.) = 1$.*

Proof: Let $B = \limsup_{n\to\infty} A_n$. We shall prove that $P(B^C) = 0$. Let $C_i = \bigcap_{n=i}^{\infty} A_n^C$. Then $B^C = \bigcup_{i=1}^{\infty} C_i$. So, we shall prove that $P(C_i) = 0$ for all i. Now, for each i and $k > i$,

$$P(C_i) = P\left(\bigcap_{n=i}^{\infty} A_n^C\right) \le P\left(\bigcap_{n=i}^{k} A_n^C\right) = \prod_{n=i}^{k}[1 - P(A_n)].$$

Use the fact that $\log(1-x) \le -x$ for all $0 \le x \le 1$ to see that, for every $k > i$,

$$\log[P(C_i)] \le \sum_{n=i}^{k} \log[1 - P(A_n)] \le -\sum_{n=i}^{k} P(A_n).$$

Since this is true for all $k > i$, it follows that $\log[P(C_i)] \le -\sum_{n=i}^{\infty} P(A_n) = -\infty$. Hence, $P(C_i) = 0$ for all i. ■

Now we use the Borel-Cantelli Lemma to prove some results in Lecture Notes Set 5.

Theorem (Lemma 25 of Lecture Notes Set 5). *If $X_n \xrightarrow{P} X$, then there is a subsequence $\{X_{n_k}\}_{k=1}^{\infty}$ such that $X_{n_k} \xrightarrow{\text{a.s.}} X$.*

Proof: Let n_k be large enough so that $n_k > n_{k-1}$ and $\Pr(d(X_{n_k}, X) > 1/2^k) < 1/2^k$. Because $\sum_{k=1}^{\infty} \Pr(d(X_{n_k}, X) > 1/2^k) < \infty$, we know that $\Pr(d(X_{n_k}, X) > 1/2^k \text{ i.o.}) = 0$. Let $A = \{d(X_{n_k}, X) > 1/2^k \text{ i.o.}\}$. Then $\Pr(A^C) = 1$ and $\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega)$ for every $\omega \in A^C$. ■

The next application of Borel-Cantelli lemma shows that $L^P(\Omega, \mathcal{F}, \mu)$ is complete.

Definition 7 (Cauchy sequence). *Let E be a metric space with metric d. A sequence $\{x_n\}_{n=1}^{\infty}$ in E is a* Cauchy sequence *if, for every $\epsilon > 0$ there exists N such that $d(x_n, x_m) < \epsilon$ for all $m, n \ge N$. The metric space E is* complete *if every Cauchy sequence in E converges to an element of E.*

Proposition 8. *If $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in a metric space and if a subsequence converges to x, the whole sequence converges to x.*

Lemma 9 (Completeness of L^P spaces). *Each Cauchy sequence in L^p converges.*

Proof: Let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $L^p(\Omega, \mathcal{F}, \mu)$. Let $\{n_k\}_{k=1}^{\infty}$ be a sequence of integers such that $\|f_{n_k} - f_{n_{k+1}}\|_p < 1/3^k$ for all k. For finite p, apply the Markov inequality to $|f_{n_k} - f_{n_{k+1}}|^p$ to get

$$\mu\left(|f_{n_k} - f_{n_{k+1}}| > \frac{1}{2^k}\right) < 2^{pk}\|f_{n_k} - f_{n_{k+1}}\|_p^p \le \left(\frac{2}{3}\right)^{pk}.$$

Since $\sum_{k=1}^{\infty} \mu(|f_{n_k} - f_{n_{k+1}}| > 1/2^k) < \infty$, it follows from Theorem 5 that

$$\mu\left(|f_{n_k} - f_{n_{k+1}}| > \frac{1}{2^k} \quad \text{i.o.}\right) = 0.$$

For $p = \infty$, we have $\mu(|f_{n_k} - f_{n_{k+1}}| > 1/3^k) = 0$, for all k, hence

$$\mu\left(|f_{n_k} - f_{n_{k+1}}| > \frac{1}{3^k} \text{ i.o.}\right) = 0.$$

In either case, it follows that, a.e. $[\mu]$ $\sum_{k=1}^{\infty} |f_{n_k}(\omega) - f_{n_{k+1}}(\omega)| < \infty$, hence $\{f_{n_k}\}_{k=1}^{\infty}$ converges a.e. $[\mu]$ to some limit, call it f. To see that f is the L^p limit of $\{f_{n_k}\}$, use Fatou's lemma and repeated applications of the triangle inequality to see that

$$\|f\|_p \leq \liminf_{k\to\infty} \|f_{n_k}\|_p \leq \left(\|f_{n_1}\|_p + \lim_{k\to\infty} \sum_{m=1}^{k} \|f_{n_m} - f_{n_{m+1}}\|_p\right) < \infty.$$

Also,

$$\|f - f_{n_k}\|_p \leq \sum_{m=k}^{\infty} \|f_{n_m} - f_{n_{m+1}}\|_p < \frac{2}{3^k}.$$

Proposition 8 then says that the whole sequence converges to f in L^p. ■

2 Sums of independent random variables

The proof of strong law of large numbers requires a series of results about sums of independent random variables. These are also interesting classical results.

Theorem 10 (Kolmogorov's maximal inequality). *Let $\{X_k\}_{k=1}^{n}$ be a finite collection of independent random variables with finite variance and mean 0. Define $S_k = \sum_{i=1}^{k} X_i$ for all k. Then*

$$\Pr\left(\max_{1\leq k\leq n} |S_k| \geq \epsilon\right) \leq \frac{\mathrm{Var}(S_n)}{\epsilon^2}.$$

Proof: For $n = 1$, the result is just Chebyshev's inequality. So assume that $n > 1$ for the rest of the proof. Let A_k be the event that $|S_k| \geq \epsilon$ but $|S_j| < \epsilon$ for $j < k$. Then $\{A_k\}_{k=1}^{n}$ are disjoint and

$$\left\{\max_{1\leq k\leq n} |S_k| \geq \epsilon\right\} = \bigcup_{k=1}^{n} A_k. \tag{1}$$

It follows that

$$
\begin{aligned}
\mathrm{E}(S_n^2) &\geq \sum_{k=1}^{n} \int_{A_k} S_n^2 dP \\
&= \sum_{k=1}^{n} \int_{A_k} \left[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right] dP \\
&\geq \sum_{k=1}^{n} \int_{A_k} [S_k^2 + 2S_k(S_n - S_k)] dP \\
&= \sum_{k=1}^{n} \int_{A_k} S_k^2 dP \\
&\geq \epsilon^2 \sum_{k=1}^{n} \Pr(A_k) \\
&= \epsilon^2 \Pr\left(\max_{1\leq k\leq n} |S_k| \geq \epsilon\right),
\end{aligned}
$$

where the first two inequalities and the first equality are obvious. The second inequality follows from the fact that $I_{A_k} S_k$ is independent of $(S_n - S_k)$ which has mean 0. The third inequality follows since $S_k^2 \geq \epsilon^2$ on A_k, and the third equality follows from Equation (1). ■

The reason that this theorem works is that whenever the maximum $|S_k|$ is large, it most likely is $|S_n|$ that is large.

A consequence of Kolmogorov's maximal inequality is the basic L^2 convergence theorem.

Theorem 11 (Basic L^2 Convergence Theorem). *Let X_1 X_2, … be independent random variables with $\mathrm{E}(X_i) = 0$ and $\mathrm{E}(X_i^2) = \sigma_i^2 < \infty$, $i = 1, 2, \ldots$, 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_∞ with $\mathrm{E}(S_\infty^2) = \sum_{i=1}^{\infty} \sigma_i^2$.*

Recall: The conclusion has been proved in the completeness of L^p for $p = 2$. Here we give a different argument for a.s. convergence using Kolmogorov's maximal inequality.

Proof: We say that S_n is Cauchy a.s. if $M_n := \sup_{p,q\geq n} |S_p - S_q| \to 0$ a.s. In light of Exercise 4, if $\Pr(M_n > \epsilon) \to 0$ 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 $M_n^* \xrightarrow{P} 0$.

For all $\epsilon > 0$,

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

where we used continuity of measure in the first step and applied Kolmogorov's inequality in the second step. Since $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$,

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

■

Remark: Later in this class we shall see that the conclusion is valid for a martingale $\{S_n\}$ with $\mathrm{E}[X_{n+1} f(X_1, \ldots, X_n)] = 0$ for all bounded measurable $f : \mathbb{R}^n \to \mathbb{R}$.

A consequence of the basic L^2 theorem is the following interesting theorem about sums of independent random variables. It gives necessary and sufficient conditions for convergence of S_n. For each $c > 0$ and each n, let $X_n^{(c)}(\omega) = X_n(\omega) I_{[0,c]}(|X_n(\omega)|)$. We will prove only the sufficiency part of the result. The necessity proof involves martingale theory and will be given later.

Theorem 12 (Three-series theorem). *Suppose that $\{X_n\}_{n=1}^{\infty}$ are independent. For each $c > 0$, consider the following three series:*

$$\sum_{n=1}^{\infty} \Pr(|X_n| > c), \quad \sum_{n=1}^{\infty} \mathrm{E}(X_n^{(c)}), \quad \sum_{n=1}^{\infty} \mathrm{Var}(X_n^{(c)}). \tag{2}$$

A necessary condition for S_n to converge a.s. is that all three series are finite for all $c > 0$. A sufficient condition is that all three series converge for some $c > 0$.

Proof: First, define some notation. For each $c > 0$ and each n, define

$$S_n^{(c)} = \sum_{k=1}^{n} X_k^{(c)},$$
$$M_n^{(c)} = \sum_{k=1}^{n} \mathrm{E}(X_k^{(c)}),$$
$$s_n^{(c)} = \sqrt{\sum_{k=1}^{n} \mathrm{Var}(X_k^{(c)})}.$$

For sufficiency, assume that all three series converge for some $c > 0$. Because the second and third series in Equation (2) converge, Theorem 11 says that $S_n^{(c)}$ converges a.s. We know that $\Pr(X_n \neq X_n^{(c)}) = \Pr(|X_n| > c)$. Since the first series in Equation (2) converges, the first Borel-Cantelli lemma says that $\Pr(X_n \neq X_n^{(c)} \text{ i.o.}) = 0$. Hence, for almost all ω, there exists $N(\omega)$ such that $S_n(\omega) - S_n^{(c)}(\omega)$ is the same for all $n \geq N(\omega)$. Hence $S_n(\omega)$ converges for almost all ω. ■

Example 13. *Let X_n have a uniform distribution on the interval $[a_n, b_n]$. A necessary condition for convergence of S_n is that $\sum_{n=1}^{\infty}(b_n - a_n)^2 < \infty$ (the third series). Another necessary condition is that $\sum_{n=1}^{\infty}(a_n + b_n)$ converge (the second series). It follows that a_n and b_n must both converge to 0 so that the first series also converges for all $c > 0$. That the two conditions above are sufficient for the convergence of S_n follows from Theorem 11.*

Example 14. *Let*

$$\Pr(X_n = x) = \begin{cases} \frac{1}{2n^2} & \text{if } x = n \text{ or } x = -n, \\ \frac{1}{2} - \frac{1}{2n^2} & \text{if } x = -1/n \text{ or } x = 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathrm{E}(X_n) = 0$ and $\mathrm{Var}(X_n) = 1 + 1/n^2 - 1/n^4$. So Theorem 11 does not imply that S_n converges a.s. However, for $c > 0$, $\mathrm{E}(X_n^{(c)}) = 0$ and $\mathrm{Var}(X_n^{(c)})$ eventually equals $1/n^2 - 1/n^4$ while $\Pr(|X_n| > c)$ eventually equals $1/n^2$, so the three-series theorem does imply that S_n converges a.s.

3 Strong Law of Large Numbers

We now prove the strong law of large numbers. We first need to recall some results in elementary analysis.

Lemma 15 (Kronecker's lemma). *If let $\{x_n : n \geq 1\}$ and $\{a_n : n \geq 1\}$ be sequences of real numbers, such that $0 < a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n < \infty$, then $(\sum_{i=1}^{n} x_i)/a_n \to 0$.*

Observation. Let $X_1, X_2, \ldots$ be independent with mean 0 and $S_n = X_1 + X_2 + \cdots + X_n$. If $\sum_{n=1}^{\infty} \mathrm{E}(X_n^2)/a_n^2 < \infty$, then by the basic L^2 convergence theorem $\sum_{n=1}^{\infty} X_n/a_n$ converges a.s., hence $S_n/a_n \to 0$ a.s. by Kronecker's lemma.

Example 16. *Let $X_1, X_2, \ldots$ be i.i.d., $\mathrm{E}(X_i) = 0$, and $\mathrm{E}(X_i^2) = \sigma^2 < \infty$.*
Take $a_n = n$:

$$\sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty \Rightarrow \frac{S_n}{n} \overset{a.s.}{\to} 0.$$

Now take $a_n = n^{\frac{1}{2}+\epsilon}$, $\epsilon > 0$*:*

$$\sum_{n=1}^{\infty} \frac{\sigma^2}{n^{1+2\epsilon}} < \infty \;\Rightarrow\; \frac{S_n}{n^{\frac{1}{2}+\epsilon}} \overset{a.s.}{\to} 0.$$

Theorem 17 (Kolmogorov's Law of Large Numbers). *Let* $X_1, X_2, \ldots$ *be i.i.d. with* $\mathrm{E}(|X_i|) < \infty$*,* $S_n = X_1 + \ldots + X_n$*. Then* $S_n/n \to \mathrm{E}(X)$ *a.s. as* $n \to \infty$*.*

Note that the theorem is true with just pairwise independence instead of the full independence assumed here. The theorem also has an important generalization to stationary sequences.

Proof: Without loss of generality, assume $\mathrm{E}(X_1) = 0$.
Consider truncated variables

$$\widehat{X}_n := X_n \mathbf{1}(|X_n| \le n).$$

Observe that

$$\Pr(X_n = \widehat{X}_n \text{ ev.}) = 1.$$

To see this, check

$$\Pr(X_n \neq \widehat{X}_n \; i.o.) = \Pr(|X_n| > n \; i.o.)$$

and use Borel-Cantelli lemma by observing

$$\sum_{n=1}^{\infty} \Pr(|X_n| > n\ ) = \sum_{n=1}^{\infty} \Pr(|X| > n\ ) \le \int_{[0,\infty)} \Pr(|X| > t) dt = \mathrm{E}|X| < \infty\,.$$

Now center the truncated variables. Define $\widetilde{X}_n := \widehat{X}_n - \mathrm{E}\Big(\widehat{X}_n\Big)$.
We will show that

$$\left(\frac{S_n}{n} \overset{\text{a.s.}}{\to} 0\right) \underset{\text{(a)}}{\Leftarrow} \left(\frac{\hat{S}_n}{n} \overset{\text{a.s.}}{\to} 0\right) \underset{\text{(b)}}{\Leftarrow} \left(\frac{\tilde{S}_n}{n} \overset{\text{a.s.}}{\to} 0\right),$$

where $\hat{S}_n = \hat{X}_1 + \hat{X}_2 + \cdots + \hat{X}_n$ and $\tilde{S}_n = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n$.
(a) comes from the fact that if $\omega \in \Big\{\omega : X_n = \widehat{X}_n \text{ ev.}\Big\}$ (which has probablity 1), then $S_n(\omega)/n - \widehat{S}_n(\omega)/n \to 0$.
(b) comes from (fact: if $c_n \to 0$ then $(c_1 + \ldots + c_n)/n \to 0$)

$$\frac{\widehat{S}_n}{n} - \frac{\widetilde{S}_n}{n} = \frac{\mathrm{E}\widehat{X}_1 + \mathrm{E}\widehat{X}_2 + \ldots + \mathrm{E}\widehat{X}_n}{n} \to 0 \text{ as } n \to \infty$$

because by DCT we have

$$\mathrm{E}\widehat{X}_n = \mathrm{E}[X_n \mathbf{1}(|X_n| \le n)] = \mathrm{E}[X \mathbf{1}(|X| \le n)] \to 0\,.$$

Now, if we can show that

$$\sum_{n=1}^{\infty} \frac{\mathrm{E}\left(\widetilde{X}_n^2\right)}{n^2} < \infty \,,$$

then the proof can be completed by Kronecker's lemma and the L^2 convergence theorem (see the observation following Lemma 15).

In fact, note that

$$\mathrm{E}\left(\widetilde{X}_n^2\right) = \mathrm{Var}(\hat{X}_n) \leq \mathrm{E}(\hat{X}_n^2) = \mathrm{E}(X^2 \mathbf{1}(|X| \leq n)).$$

So, by some basic manipulation, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\mathrm{E}\left(\widetilde{X}_n^2\right)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{\mathrm{E}X^2 \mathbf{1}\,(|X| \leq n)}{n^2} = \mathrm{E}\left(X^2 \sum_{n=1}^{\infty} \frac{\mathbf{1}\,(|X| \leq n)}{n^2}\right) \\
&\leq \mathrm{E}\left(X^2 \sum_{n=1}^{\infty} \frac{\mathbf{1}\,(|X| \leq n)}{n^2} \mathbf{1}(|X| \leq 2)\right) + \mathrm{E}\left(X^2 \sum_{n=1}^{\infty} \frac{\mathbf{1}\,(|X| \leq n)}{n^2} \mathbf{1}(|X| > 2)\right) \\
&\leq 4 \sum_{n=1}^{\infty} \frac{1}{n^2} + \mathrm{E}\left(X^2 \sum_{n=1}^{\infty} \frac{\mathbf{1}\,(\lfloor |X| \rfloor \leq n)}{n^2} \mathbf{1}(|X| > 2)\right) \\
&\leq \sum_{n=1}^{\infty} \frac{4}{n^2} + \mathrm{E}\left(X^2 \sum_{n=\lfloor |X| \rfloor}^{\infty} \frac{1}{n^2} \mathbf{1}(|X| > 2)\right) \\
&\leq \sum_{n=1}^{\infty} \frac{4}{n^2} + \mathrm{E}\left(X^2 \frac{1}{\lfloor |X| \rfloor - 1} \mathbf{1}(|X| > 2)\right) \\
&\leq \sum_{n=1}^{\infty} \frac{4}{n^2} + \mathrm{E}\left(X^2 \frac{3}{|X|} \mathbf{1}(|X| > 2)\right) \\
&\leq \sum_{n=1}^{\infty} \frac{4}{n^2} + \mathrm{E}(3|X|) < \infty \,.
\end{aligned}$$

■

4 Law of the Iterated Logarithm

Let X_1, X_2, ... be i.i.d. with $\mathrm{E}X_i = 0$, $\mathrm{E}X_i^2 = \sigma^2$, $S_n = X_1 + ... + X_n$. We know

$$\frac{S_n}{n^{\frac{1}{2}+\varepsilon}} \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

For general interest, we state, without proof, the *Law of the Iterated Logarithm*:

$$\limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log n)}} = 1 \text{ a.s.}$$
$$\liminf_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log n)}} = -1 \text{ a.s.}$$

We will show later

$$\frac{S_n}{\sigma n^{\frac{1}{2}}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$