CHAPTER 2: LAW OF LARGE NUMBERS

September 7, 2021

Contents

1 Independence

1.1 Definition

Definition. Events $A_1, \ldots, A_n$ are **independent** if

$$P(\cap_{i\in I} A_i) = \prod_{i\in I} P(A_i), \quad \forall\, I \subset \{1, \ldots, n\},$$

or equivalently,

$$P(B_1 \cdots B_n) = P(B_1) \cdots P(B_n), \quad \text{where } B_i = A_i \text{ or } \Omega.$$

Definition. Collections of sets $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are **independent** if

$$P(A_1 \cdots A_n) = P(A_1) \cdots P(A_n), \quad \text{where } A_i \in \mathcal{A}_i \text{ or } A_i = \Omega.$$

Definition. Random variables $X_i, \ldots, X_n$ are **independent** if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent.

Definition. σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots,$ are **pairwise independent** if $\mathcal{F}_i$ is independent of $\mathcal{F}_j$ if $i \neq j$.

Definition. An infinite collection of $\mathcal{F}$'s is **independent** if every finite subcollection is.

1.2 Properties

Theorem. If $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent, and each $\mathcal{A}_i$ is a π-system, then $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ are independent.

Proof. It suffices to show that

$$P(B_1 \cdots B_n) = P(B_1) \cdots P(B_n), \quad \forall\, B_i \in \sigma(\mathcal{A}_i).$$

Fix $B_i \in \mathcal{A}_i$ or $B_i = \Omega$ for $i = 2, \ldots, n$. Define

$$\mathcal{L}_1 := \{B_1 \in \sigma(\mathcal{A}_1) : P(B_1 \cdots B_n) = P(B_1) \cdots P(B_n)\}.$$

Note that $\mathcal{L}_1$ contains the π-system $\mathcal{A}_1$ by assumption. Moreover, it can be easily verified that $\mathcal{L}_1$ is a λ-system (details omitted). Therefore, by the π-λ theorem, $\sigma(\mathcal{A}_1) \subset \mathcal{L}_1$. We have proved that $\sigma(\mathcal{A}_1), \mathcal{A}_2, \ldots, \mathcal{A}_n$ are independent. Repeating the above argument to $\mathcal{A}_2, \ldots, \mathcal{A}_n$ completes the proof. $\square$

Theorem. Random variables $X_1, \ldots, X_n$ are independent if and only if

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i), \quad \forall x_i \in (-\infty, \infty].$$

Proof. Note that $\sigma(X_i) = \sigma(\{\{X_i \le x_i\} : x_i \in (-\infty, \infty]\})$ and $\{\{X_i \le x_i\} : x_i \in (-\infty, \infty]\}$ is a π-system. The equation implies the n π-systems are independent. Therefore, the theorem follows from the previous theorem. $\square$

Exercise. Suppose $(X_1, \dots, X_n)$ has density $f(x_1, \dots, x_n)$ and $f(x_1, \dots, x_n) = g_1(x_1) \dots g_n(x_n)$, $g_i \ge 0$ measurable. Then, $X_1, \dots, X_n$ are independent. [Hint: use the previous theorem and Fubini's theorem 1.7.2]

Exercise. Suppose $X_1, \dots, X_n$ are discrete random variables. Then, $X_1, \dots, X_n$ are independent if and only if

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i), \quad \text{for all possible values } x_i.$$

Theorem. If $X_1, \dots, X_n$ are independent and $I \subset \{1, 2, \dots, n\}$, then $g(X_i, i \in I)$ and $h(X_j, j \in I^c)$ are independent for measurable functions g and h.

Proof. We prove $\sigma(X_i : i \in I)$ and $\sigma(X_j : j \in I^c)$ are independent. The conclusion then follows automatically. Note that

$$\sigma(X_i : i \in I) = \sigma(\cap_{i \in I} A_i : A_i \in \sigma(X_i)),$$

and

$$\sigma(X_j : j \in I^c) = \sigma(\cap_{j \in I^c} B_j : B_j \in \sigma(X_j)).$$

The generating sets are the right-hand sides are π-systems. These two π-systems are independent by the condition. Hence the result follows. $\square$

Fact. If $\mathcal{A}$ and $\mathcal{A}$ itself are independent, then $P(A) = 0$ or 1 for any $A \in \mathcal{A}$.

Proof.

$$P(A) = P(A \cap A) = P(A) \cdot P(A).$$

$\square$

Theorem 2.5.1. (Kolmogorov's 0-1 Law). Suppose $X_1, X_2, \dots$ are independent. Let

$$\mathcal{T} = \cap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$$

be the **tail σ-field**. Then $P(A) = 0$ or 1 for any $A \in \mathcal{T}$.

Proof. Step 1: $\sigma(X_{k+1}, X_{k+2}, \dots)$ can be written as $\sigma(\cup_{j=1}^{\infty} \sigma(X_{k+1}, \dots, X_{k+j}))$. The union inside of $\sigma(\dots)$ is a π-system. Moreover, $\sigma(X_1, \dots, X_k)$ is independent of this π-system by the condition. Therefore, $\sigma(X_1, \dots, X_k)$ is independent of $\sigma(X_{k+1}, X_{k+2}, \dots)$.

Step 2: Write $\sigma(X_1, X_2, \dots) = \sigma(\cup_{j=1}^{\infty} \sigma(X_1, \dots, X_j))$ and argue similarly that $\sigma(X_1, X_2, \dots)$ is independent of $\mathcal{T}$.

Step 3: By definition, $\mathcal{T} \subset \sigma(X_1, X_2, \dots)$.

Therefore $\mathcal{T}$ is independent of itself. The theorem then follows from the previous Fact. $\square$

Examples of tail events. If $X_1, X_2, \dots$ are independent, then

$$\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} \quad \text{exists}\} \in \mathcal{T},$$

$$\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = a\} \in \mathcal{T}.$$

Theorem. If X and Y are independent random variables with $E|X|, E|Y| < \infty$ or $X, Y \geq 0$, then

$$E(XY) = E(X)E(Y). \tag{1.1}$$

Proof. Case 1. $X = 1_A, Y = 1_B$. In this case, X and Y are independent implies $P(AB) = P(A)P(B)$. Eq. (1.1) then follows from $E(XY) = P(AB)$ and $E(X)E(Y) = P(A)P(B)$.

Case 2. $X = \sum_{i=1}^n a_i 1_{A_i}$, $Y = \sum_{j=1}^m b_j 1_{B_j}$. Assume without loss of generality that a_i's are distinct and b_j's are distinct. In this case, X and Y are independent implies $P(A_i B_j) = P(A_i)P(B_j)$ for any i, j. Then we have

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j P(A_i B_j) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j P(A_i)P(B_j) = E(X)E(Y).$$

Case 3. $X, Y \geq 0$. Choose a sequence of simple random variables $X_n \uparrow X$ and $\sigma(X_n) \subset \sigma(X)$. ($X_M$ can be chosen as $X_M^{(l)}$ as in Chapter 1, Section 3.2.) Choose also a sequence of simple random variables $Y_n \uparrow X$ and $\sigma(Y_n) \subset \sigma(X)$. From Case 2, we have $E(X_n Y_n) = E(X_n)E(Y_n)$. Then take the limit as $n \to \infty$ to obtain (1.1).

Case 4. For the case $E|X|, E|Y| < \infty$, write $X = X^+ - X^-, Y = Y^+ - Y^-$ and use the previous case. $\square$

Corollary. If $X, Y \in \mathbb{R}^d$ are independent. Let g, h be two measurable functions from $\mathbb{R}^d$ to $\mathbb{R}$. Suppose $g, h \geq 0$ or $E|g(X)|, E|h(Y)| < \infty$. Then

$$E[g(X)h(Y)] = E[g(X)]E[g(Y)].$$

2 Law of Large Numbers

2.1 Weak Law of Large Numbers

Definition. The **covariance** of two square-integrable random variables X, Y is defined as

$$Cov(X,Y) := E[(X - EX)(Y - EX)] = E[XY] - [EX][EY].$$

We say X and Y are **uncorrelated** if $Cov(X,Y) = 0$. Note that if X is independent of Y, then $Cov(X,Y) = 0$. The reverse statement is not true, unless X and Y both follow normal distribution.

Definition. The **correlation** of two square-integrable random variables X, Y is defined as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Here $0/0$ is understood to be 0.

Note that $Corr(aX + b, cY + d) = Corr(X,Y) \cdot sign(ac)$.

Definition. We say $(Y_n)_{n=1}^{\infty}$ **converges in probability** to Y, if $\forall \varepsilon > 0$

$$P(|Y_n - Y| > \varepsilon) \to 0, \quad \text{as } n \to \infty.$$

Definition. For $p > 0$, we say $(Y_n)_{n=1}^{\infty}$ **converges in** L_p to Y if $E|Y_n|^p, E|Y| < \infty$ for all n and

$$E|Y_n - Y|^p \to 0, \quad \text{as } n \to \infty.$$

Note that the above two definitions require $(Y_n)_{n=1}^{\infty}$ and Y to be defined on the same probability space.

Fact. If $Y_n \to Y$ in L_p for some $p > 0$, then $Y_n \to Y$ in probability.
Proof. By Markov's inequality,

$$P(|Y_n - Y| > \varepsilon) \le \frac{E|Y_n - Y|^p}{\varepsilon^p} \to 0.$$

□

Theorem (WWLN with finite second moment). If $X_1, X_2, \ldots$ are uncorrelated, and $E(X_i) = \mu_i, \mathrm{Var}(X_i) \le c > \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then,

$$\frac{S_n - \sum_{i=1}^n \mu_i}{n} \to 0 \quad \text{in } L_2 \text{ and in probability.}$$

Proof.

$$E(\frac{S_n - \sum_{i=1}^n \mu_i}{n})^2 = \mathrm{Var}(\frac{S_n}{n}) = \frac{\sum_{i=1}^n \mathrm{Var}(X_i)}{n^2} \le \frac{c}{n} \to 0.$$

□

Theorem. If $(S_n)_{n=1}^{\infty}$ is a sequence of random variables with $\sigma_n^2 = \mathrm{Var}(S_n)$ and $\sigma_n^2/b_n^2 \to 0$, then

$$\frac{S_n - E(S_n)}{b_n} \to 0 \quad \text{in } L_2 \text{ and in probability.}$$

Example 2.2.3 Coupon Collector's Problem. Let $n \geq 1$ be an integer and $X_1, X_2, \dots$ be independent and uniformly distributed on $\{1, \dots, n\}$. Define for $k = 0, 1, \dots, n$

$$\tau_k^n = \inf\{m \geq 0 : |\{X_1, \dots, X_m\}| = k\},$$

where $|\cdot|$ denotes the size of a set (the number of distinct elements in a set). Let $T_n = \tau_n^n$. Then we have

$$\frac{T_n}{n \log n} \to 1 \quad \text{in probability.}$$

Proof. Let $Y_k = \tau_k^n - \tau_{k-1}^n$, $k = 1, 2, \dots, n$. Then Y_k follows a geometric distribution with parameter $\frac{n-k+1}{n}$ and

$$EY_k = \frac{n}{n-k+1}, \qquad \mathrm{Var}(Y_k) = \frac{k-1}{n} \frac{n^2}{(n-k+1)^2}.$$

Moreover, $Y_1, \dots, Y_n$ are independent and $T_n = \sum_{k=1}^n Y_k$. Therefore,

$$E(T_n) = n(1 + \frac{1}{2} + \cdots + \frac{1}{n}) \sim n \log n$$

and

$$\mathrm{Var}(T_n) = \sum_{k=1}^n \mathrm{Var}(Y_k) \leq \sum_{k=1}^n \frac{n^2}{(n-k+1)^2} \leq Cn^2$$

for some absolute constant C. Choosing $b_n = n \log n$ in the above theorem, we obtain

$$\frac{T_n - n \log n}{n \log n} \to 0 \quad \text{in probability.}$$

□

The proofs of above results are based on computation of the first and second moments. In the remainder of this subsection, we consider the case where $S_n = \sum_{i=1}^n X_i$ and X_i may not have finite moments. In this case, we use the technique of TRUNCATION .

Theorem 2.2.6 (Weak LLN for Triangular Arrays.) Assume for each $n \geq 1$, $\{X_{n,k} : 1 \leq k \leq n\}$ are independent. Let $b_n > 0$ and

$$\bar{X}_{n,k} = X_{n,k} 1_{\{|X_{n,k}| \leq b_n\}}.$$

If

$$(i) \quad \sum_{k=1}^{n} P(|X_{n,k}| > b_n) \to 0 \text{ and}$$

$$(ii) \quad b_n^{-2} \sum_{k=1}^{n} E[(\bar{X}_{n,k})^2] \to 0,$$

then, with $S_n = \sum_{k=1}^n X_{n,k}$ and $a_n = \sum_{k=1}^n E\bar{X}_{n,k}$,

$$\frac{S_n - a_n}{b_n} \to 0 \quad \text{in probability.}$$

Proof. Let $\bar{S}_n = \sum_{k=1}^n \bar{X}_{n,k}$. We have

$$P(|\frac{S_n - a_n}{b_n}| > \varepsilon) \leq \sum_{k=1}^{n} P(|X_{n,k}| > b_n) + P(|\frac{\bar{S}_n - a_n}{b_n}| > \varepsilon).$$

The first term tends to 0 by condition (i). The second term tends to 0 by condition Chebyshev's inequality and condition (ii). □

Theorem 2.2.7 (WLLN without moment assumption). Let $X_1, X_2, \dots$ be i.i.d. with

$$xP(|X_1| > x) \to 0, \quad \text{as } x \to \infty.$$

Let $S_n = \sum_{i=1}^n X_i$, $\mu_n = E(X_1 1_{\{|X_1| \leq n\}})$. Then

$$\frac{S_n}{n} - \mu_n \to 0 \quad \text{in probability.}$$

Proof. In the previous theorem, let $X_{n,k} = X_k$ and $b_n = n$. Then $\bar{X}_{n,k} = X_k 1_{\{|X_k| \leq n\}}$. We have

$$\sum_{k=1}^{n} P(|X_{n,k}| > b_n) = \sum_{k=1}^{n} P(|X_k| > n) = nP(|X_1| > n) \to 0.$$

We also have

$$b_n^{-2} \sum_{k=1}^{n} E(\bar{X}_{n,k})^2 = n^{-2} \sum_{k=1}^{n} E[(X_k 1_{\{|X_k| \leq n\}})^2] = n^{-1} E(\bar{X}_1^2 1_{\{|X_1| \leq n\}}).$$

From Lemma 2.2.8 of the textbook (which follows form Fubini's theorem), we have

$$n^{-1} E(\bar{X}_1^2 1_{\{|X_1| \leq n\}}) = n^{-1} \int_0^n 2y P(|X_1| > y) dy \to 0,$$

where the last convergence follows by the condition $xP(|X_1| > x) \to 0$ and ignoring an arbitrarily large initial portion in the "average" $n^{-1}\int_0^n$. This verifies the conditions (i) and (ii) of the previous theorem; hence the result follows. $\square$

Theorem 2.2.9 (WLLN with finite first moment). Let $X_1, X_2, \ldots$ be i.i.d. with $E|X_1| < \infty$. Let $S_n = \sum_{i=1}^n X_i, \mu = EX_1$. Then

$$\frac{S_n}{n} \to \mu \quad \text{in probability.}$$

Proof. Note that by the dominated convergence theorem (DCT),

$$xP(|X_1| > x) \leq E(|X_1| 1_{\{|X_1|>x\}}) \to 0.$$

Again by DCT,

$$E(X_1 1_{\{|X_1|\leq n\}}) \to \mu.$$

Therefore, this theorem follows from the previous theorem. $\square$

2.2 Strong Law of Large Numbers

Definition. We say a sequence of random variables Y_n converges to Y **almost surely (a.s.)** if

$$P(w : \lim_{n\to\infty} Y_n(w) = Y(w)) = 1.$$

To give a way of proving a.s. convergence, we will have an equivalent statement below. We first need a definition.

Definition. Let $A_1, A_2, \ldots$ be a sequence of events. The event that they **occur infinitely often** is defined to be (the four expressions involved are equivalent)

$$\begin{aligned}
\{A_n \ i.o.\} :=& \{w \in \Omega : w \text{ is in infinitely many } A_i\text{'s}\} \\
=& \cap_{n=1}^\infty \cup_{k=n}^\infty A_k \\
=& \limsup A_n \\
=& \{w : \exists \text{ a subsequence } n_k,\ w \in \cap_k A_{n_k}\}.
\end{aligned}$$

Fact. $Y_n \to Y$ a.s. if and only if $\forall\, \varepsilon > 0$,

$$P(|Y_n - Y| > \varepsilon \ i.o.) = 0.$$

Proof. The right-hand side is equivalent to $\forall\, \varepsilon > 0$ (may not be the same ε as above),

$$P(\limsup_{n\to\infty} |Y_n - Y| \leq \varepsilon) = 1.$$

This implies, by tbe continuity from above property of probability measures,

$$P(\limsup_{n\to\infty} |Y_n - Y| = 0) = \lim_{\varepsilon \downarrow 0} P(\limsup_{n\to\infty} |Y_n - Y| \le \varepsilon) = 1.$$

This is equivalent to a.s. convergence by definition. $\square$

The following result is extremely important to study the event of infinitly often.

Theorem (Borel-Cantelli Lemmas). We have

(i) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \ i.o.) = 0$;

(ii) If A_n's are independent, then $\sum_{n=1}^{\infty} P(A_n) = \infty$ implies $P(A_n \ i.o.) = 1$.

Proof. (i) We have

$$P(A_n \ i.o.) = P(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k) = \lim_{n\to\infty} P(\cup_{k=n}^{\infty} A_k) = 0,$$

where the last equality follows form the condition.

(ii) We have

$$\begin{aligned}
P(A_n \ i.o.) &= 1 - P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k^c) \\
&= 1 - \lim_{n\to\infty} P(\cap_{k=n}^{\infty} A_k^c) \\
&= 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} P(A_k^c) && \text{(by independence)} \\
&= 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} (1 - P(A_k)) \\
&\ge 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} e^{-P(A_k)} && \text{(from } 1 + x \le e^x) \\
&= 1 - \lim_{n\to\infty} e^{-\sum_{k=n}^{\infty} P(A_k)} \\
&= 1. && \text{(by condition)}
\end{aligned}$$

$\square$

A simple application of the B-C lemma (i) leads to the second statement (a very useful fact) of the following theorem.

Theorem. If $Y_n \to Y$ a.s., then $Y_n \to Y$ in probability. If $Y_n \to Y$ in probability, then there exists a subsequence n_k such that $Y_{n_k} \to Y$ a.s.

Proof. The first statement follows directly by definition. To prove the second statement, we choose a subsequence $n_1 < n_2 < n_3 < \dots$ such that

$$P(|Y_{n_k} - Y| > \frac{1}{2^k}) \le \frac{1}{2^k}.$$

Such a subsequence exists because of the condition that $Y_n \to Y$ in probability. For this subsequence, we have

$$\sum_{k=1}^{\infty} P(|Y_{n_k} - Y| > \frac{1}{2^k}) < \infty.$$

By B-C lemma (i),

$$P(|Y_{n_k} - Y| > \frac{1}{2^k} \ i.o.) = 0.$$

The implies $P(|Y_{n_k} - Y| > \varepsilon \ i.o.) = 0$ for any $\varepsilon > 0$; hence almost sure convergence. $\square$

Next, we state a theorem that generalize B-C lemma (ii). The proof uses the **subsequence method.**

Theorem 2.3.8. If $A_1, A_2, \ldots$ are pairwise independent, and $\sum_{i=1}^n P(A_i) = \infty$. Then, as $n \to \infty$,

$$\frac{\sum_{i=1}^n 1_{A_i}}{\sum_{i=1}^n P(A_i)} \to 1 \quad a.s.$$

Proof. Let $S_n = \sum_{i=1}^n 1_{A_i}, b_n = \sum_{i=1}^n P(A_i)$. We have, for $\varepsilon > 0$

$$\begin{aligned} P(|\frac{S_n}{b_n} - 1| > \varepsilon) \leq & \frac{\operatorname{Var}(S_n/b_n)}{\varepsilon^2} = \frac{\sum_{i=1}^n P(A_i)(1-P(A_i))}{\varepsilon^2 b_n^2} \\ \leq & \frac{\sum_{i=1}^n P(A_i)}{\varepsilon^2 (\sum_{i=1}^n P(A_i))^2} = \frac{1}{\varepsilon b_n}. \end{aligned}$$

Choose a subsequence $n_k = \min\{m : b_m \geq k^2\}$ (it is possible because of the condition $\sum_{i=1}^n P(A_i) = \infty$). Then

$$\sum_{k=1}^{\infty} P(|\frac{S_{n_k}}{b_{n_k}} - 1| > \varepsilon) < \infty,$$

and by B-C lemma (i),

$$\frac{S_{n_k}}{b_{n_k}} \to 1 \quad a.s.$$

For any sufficiently large integer m, there exists k such that $n_{k-1} \leq m \leq n_k$, and

$$\frac{S_{n_{k-1}}}{b_{n_{k-1}}} \frac{b_{n_{k-1}}}{b_{n_k}} = \frac{S_{n_{k-1}}}{b_{n_k}} \leq \frac{S_m}{b_m} \leq \frac{S_{n_k}}{b_{n_{k-1}}} = \frac{S_{n_k}}{b_{n_k}} \frac{b_{n_k}}{b_{n_{k-1}}}.$$

Because both sides tend to 1 a.s., S_m/m in the middle must tend to 1 a.s.. $\square$

We now turn to Strong Law of Large Numbers (SLLN). As a warm-up, we first prove an easier version.

Theorem 2.3.5. Let $X_1, X_2, \dots$ be i.i.d. with $EX_i = \mu, E(X_i^4) < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n}{n} \to \mu \quad a.s.$$

Proof. WLOG, assume $\mu = 0$ (otherwise, consider $X_i' = X_i - \mu$). We have

$$P(|\frac{S_n}{n}| > \varepsilon) \le \frac{E[S_n^4]}{\varepsilon^4 n^4} \le \frac{\sum_{i,,j=1}^n E[X_i^2 X_j^2]}{\varepsilon^4 n^4} \le \frac{C}{\varepsilon^4 n^2},$$

which sums up to a finite number over n. The theorem follows by applying B-C lemma (i) and by the definition of a.s. convergence. □

Theorem 2.4.1 (SLLN). Let $X_1, X_2, \dots$ be i.i.d. with $E|X_i| < \infty$, $EX_i = \mu$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n}{n} \to \mu \quad a.s.$$

Proof. WLOG, assume $X_i \ge 0$ (otherwise, consider $X_i = X_i^+ - X_i^-$). The basic idea of the proof is to use truncation, a second moment calculation and the subsequence method.

[Step 1: Truncation.] Let $Y_k = X_k 1_{\{|X_k| \le k\}}$, $T_n = \sum_{i=1}^n Y_i$. Because

$$\sum_{k=1}^\infty P(|X_k| > k) \le E|X_1| < \infty,$$

we have by B-C lemma (i)

$$P(X_k \ne Y_k \ i.o.) = 0.$$

Therefore, to prove the theorem, if suffices to show $\frac{T_n}{n} \to \mu \ a.s.$.

[Step 2: 2nd moment calculation.] Fix $\alpha > 1$. Choose $k(n) = \lfloor \alpha^n \rfloor$ (this means the integer part, and it is obviously $\ge \alpha^n/2$). We have, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{n=1}^\infty P\left(|\frac{T_{k(n)} - ET_{k(n)}}{k(n)}| > \varepsilon\right) \\
\le&\varepsilon^{-2} \sum_{n=1}^\infty \frac{\mathrm{Var}(T_{k(n)})}{(k(n))^2} = \varepsilon^{-2} \sum_{n=1}^\infty \frac{1}{(k(n))^2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) \\
=&\varepsilon^{-2} \sum_{m=1}^\infty \mathrm{Var}(Y_m) \sum_{n:k(n)\ge m} (k(n))^{-2} && \text{(Fubini)} \\
\le&\varepsilon^{-2} \sum_{m=1}^\infty \mathrm{Var}(Y_m) 4 \sum_{n:\alpha^n \ge m} \alpha^{-2n} \\
\le&4(1-\alpha^{-2})^{-1} \varepsilon^{-2} \sum_{m=1}^\infty \frac{\mathrm{Var}(Y_m)}{m^2}.
\end{aligned}$$

Note that

$$\begin{aligned}\sum_{m=1}^{\infty}\frac{\text{Var}(Y_m)}{m^2}\leq\sum_{m=1}^{\infty}\frac{EY_m^2}{m^2}&=\sum_{m=1}^{\infty}\frac{\int_0^m 2yP(|Y_m|>y)dy}{m^2}\\ &\leq\sum_{m=1}^{\infty}\frac{\int_0^\infty 2y1_{\{y<m\}}P(|X_1|>y)dy}{m^2}\\ &=\int_0^\infty\{\sum_{m:m>y}\frac{1}{m^2}\}2yP(|X_1|>y)dy \qquad \text{(Fubini)}\\ &\leq C\int_0^\infty P(|X_1|>y)dy=CE|X_1|<\infty.\end{aligned}$$

Therefore,

$$\sum_{n=1}^{\infty}P\left(|\frac{T_{k(n)}-ET_{k(n)}}{k(n)}|>\varepsilon\right)<\infty,$$

and by B-C lemma (i),

$$\frac{T_{k(n)}-ET_{k(n)}}{k(n)}\to 0 \quad a.s.$$

Note also that, because $EY_m=EX_1 1_{\{|X_1|\leq m\}}\to\mu$ as $m\to\infty$, we have

$$\frac{ET_{k(n)}}{k(n)}=\frac{\sum_{m=1}^{k(n)}EY_m}{k(n)}\to\mu.$$

Therefore,

$$\frac{T_{k(n)}}{k(n)}\to\mu \quad a.s.$$

[Step 3: Subsequence method.] For $k(n)\leq m<k(n+1)$, we have

$$\frac{k(n)}{k(n+1)}\frac{T_{k(n)}}{k(n)}\leq\frac{T_m}{m}\leq\frac{k(n+1)}{k(n)}\frac{T_{k(n+1)}}{k(n+1)}.$$

The left-hand side tends to μ/α and the right-hand side tends to $\mu\alpha$. Therefore, for any $\alpha>1$,

$$\limsup_{m\to\infty}\frac{T_m}{m}\leq\mu\alpha, \quad \liminf_{m\to\infty}\frac{T_m}{m}\geq\mu/\alpha.$$

The theorem follows by letting $\alpha\downarrow 1$. □

Remark. Let $X_1,X_2,\ldots$ be i.i.d. and $S_n=\sum_{i=1}^n X_i$. If $\frac{S_n}{n}\to\mu$ $a.s.$ for some finite number μ, then we must have $E|X_i|<\infty$. This means the condition of the above SLLN is necessary.

Proof of the Remark. Consider $\frac{X_n}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n-1}\frac{n-1}{n}$. This tends to 0 a.s. because of the condition $\frac{S_n}{n} \to \mu$ *a.s.*. Therefore, $P(|\frac{X_n}{n}| > 1\ i.o.) = 0$ and by the B-C lemma (ii),

$$\sum_{n=1}^{\infty} P(|\frac{X_n}{n}| > 1) < \infty.$$

Therefore, from the i.i.d. assumption,

$$E|X_1| \leq 1 + \sum_{n=1}^{\infty} P(|X_1| > n) < \infty.$$

$\square$

Next result concerns the case of infinite mean.

Theorem. Let $X_1, X_2, \ldots$ be i.i.d. and $S_n = \sum_{i=1}^n X_i$. If $E(X_i^+) = \infty$ and $E(X_i^-) < \infty$, then

$$\frac{S_n}{n} \to \infty \quad a.s.$$

Proof. Define $X_i^M = X_i \wedge M$. Then $E|X_i^M| < \infty$. By SLLN, for $S_n^M := \sum_{i=1}^n X_i^M$, we have

$$\frac{S_n^M}{n} \to EX_i^M \quad a.s.$$

Therefore,

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \lim_{n\to\infty} \frac{S_n^M}{n} = EX_i^M \to \infty, \text{ as } M \to \infty.$$

$\square$

Finally, we give an application of the SLLN.

Example. Let $X_1, X_2, \ldots$ be i.i.d. from a population distribution $F(x)$. Define the **Empirical Distribution Function**

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leq x\}}.$$

By SLLN, we have $F_n(x) \to F(x)$ a.s. for any $x \in \mathbb{R}$.

In fact, we have the following stronger result, which follows from the above point-wise convergence and monotonicity and boundedness of distribution functions. We omit the proof.

Theorem 2.4.7 (The Glivenko-Cantelli Theorem). Under the setting of the above example, we have

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0 \quad a.s.$$

3 Convergence of Random Series

3.1 Kolmogorov's Maximal Inequality

Theorem 2.5.2 (Komogorov's Maximal Inequality). Let $X_1, X_2, \dots$ be independent with $EX_i = 0, E(X_i^2) < \infty$ for all i. Let $S_n = \sum_{i=1}^n X_i$. Then for any $x > 0$, we have

$$P(\max_{1\le k\le n} |S_k| \ge x) \le \frac{E(S_n^2)}{x^2}.$$

Proof. Let

$$A_k := \{|S_i| < x \text{ for } i < k, |S_k| \ge x\}.$$

These A_k's are disjoint and

$$\{\max_{1\le k\le n} |S_k| \ge x\} = \cup_{k=1}^n P(A_k).$$

Therefore,

$$P(\max_{1\le k\le n} |S_k| \ge x) = \sum_{k=1}^n P(A_k) \le \sum_{k=1}^n \frac{E(S_k^2 1_{A_k})}{x^2}.$$

We also have

$$\begin{aligned} &E(S_n^2 1_{A_k}) = E[(S_k + S_n - S_k)^2 1_{A_k}] \\ =&E(S_k^2 1_{A_k}) + 2E[S_k 1_{A_k}(S_n - S_k)] + E[(S_n - S_k)^2 1_{A_k}] \\ \ge&E(S_k^2 1_{A_k}), \end{aligned}$$

where we used, by the independence and mean zero assumption,

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

Therefore,

$$P(\max_{1\le k\le n} |S_k| \ge x) \le \frac{E(S_n^2 1_{A_k})}{x^2} \le \frac{E(S_n^2)}{x^2}.$$

$\square$

Kolmogorov's maximal inequality can be used to obtain the following result giving a sufficient condition for the convergence of random series.

Theorem 2.5.3. Let $X_1, X_2, \dots$ be a sequence of independent random variables with $EX_i = 0$ for all i. If $\sum_{i=1}^\infty E(X_i^2) < \infty$, then $\sum_{i=1}^\infty X_i$ converge almost surely (recall this means $\sum_{i=1}^n X_i(w)$ converges a.s.).

Proof. Let $S_n = \sum_{i=1}^n X_i$. It suffices to prove S_n is a Cauchy sequence, a.s., that is,

$$\omega_M := \sup_{m,n \geq M} |S_m - S_n| \downarrow 0 \quad a.s. \quad \text{as } M \uparrow \infty.$$

To show this, for any $\varepsilon > 0$, we write

$$\begin{aligned} P(\omega_M > 2\varepsilon \ i.o.) =& P(\omega_M > 2\varepsilon, \forall\ M) \qquad \text{(by monotonicity of } \omega_M\text{)} \\ =& P(\cap_M \{\omega_M > 2\varepsilon\}) = \lim_{M\to\infty} P(\omega_M > 2\varepsilon). \end{aligned}$$

Note that, by definition of ω_M and the union bound,

$$\begin{aligned} P(\omega_M > 2\varepsilon) =& P(\sup_{m,n\geq M} |S_m - S_n| > 2\varepsilon) \\ \leq& P(\sup_{m\geq M} |S_m - S_M| > \varepsilon) + P(\sup_{n\geq M} |S_n - S_M| > \varepsilon) \\ =& 2P(\sup_{m\geq M} |S_m - S_M| > \varepsilon). \end{aligned}$$

Moreover, by the Kolmogorov's maximal inequality,

$$\begin{aligned} P(\sup_{m\geq M} |S_m - S_M| > \varepsilon) = \lim_{N\to\infty} P(\sup_{M\leq m\leq N} |S_m - S_M| > \varepsilon) \\ \leq \limsup_{N\to\infty} \frac{\mathrm{Var}(S_N - S_M)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{i=M+1}^{\infty} E(X_i^2) \to 0. \end{aligned}$$

Combining the above arguments, we conclude that

$$P(\omega_M > 2\varepsilon \ i.o.) = 0, \quad \forall\ \varepsilon > 0,$$

hence $\omega_M \to 0$. □

Example. Let $X_1, X_2, \ldots$ be i.i.d. Rademacher variables, i.e., $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. We have, for $\alpha > 1/2$, $\sum_{i=1}^\infty \frac{X_i}{i^\alpha}$ converges a.s.

3.2 Kolmogorov's Three-series Theorem

In this subsection, we present the most general result to determine whether a random series is convergent or not.

Theorem 2.5.4 (Komogorov's Three-series Theorem). Let $X_1, X_2, \ldots$ be a sequence of independent random variables. Let $A > 0$ be a constant. Define $Y_i = X_i 1_{\{|X_i|\leq A\}}$. If

(i) $\sum_{n=1}^\infty P(|X_n| > A) < \infty$
(ii) $\sum_{n=1}^\infty EY_n$ converges, and
(iii) $\sum_{n=1}^\infty \mathrm{Var}(Y_n) < \infty$,

then $\sum_{n=1}^{\infty} X_n$ converges a.s.

Remark. These three conditions are also necessary conditions, we will prove it in Chapter 3.

Proof. Let $\mu_n = EY_n$. From Theorem 2.5.3 above and condition (iii), we have $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ converges a.s. Combined with condition (ii), we have $\sum_{n=1}^{\infty} Y_n$ converges a.s. From the B-C lemma and condition (i), we have

$$P(X_n \neq Y_n \ i.o.) = 0.$$

Therefore, $\sum_{n=1}^{\infty} X_n$ converges a.s. $\square$

We will use the following result. We omit its proof, which is by analysis.

Theorem 2.5.5 (Kronecker's Lemma). If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converges, then $\frac{\sum_{i=1}^{n} x_n}{a_n} \to 0$.

Now we use Kolmogorov's Three-series Theorem and Kronecker's lemma to give an alternative proof of SLLN. Recall:

Theorem 2.4.1 (SLLN). Let $X_1, X_2, \dots$ be i.i.d. with $E|X_i| < \infty$, $EX_i = \mu$. Let $S_n = \sum_{i=1}^{n} X_i$. Then

$$\frac{S_n}{n} \to \mu \quad a.s.$$

Second Proof. Recall that WLOG, we can assume $X_i \geq 0$. As in the first proof, we let $Y_k = X_k 1_{\{|X_k| \leq k\}}$, $T_n = \sum_{i=1}^{n} Y_i$. We have argued that $P(X_k \neq Y_k \ i.o.) = 0$ and it suffices to show $\frac{T_n}{n} \to \mu \ a.s.$. Recall that we have also argued that $\frac{ET_n}{n} \to \mu$. Therefore, we are left to show

$$\sum_{k=1}^{n} \frac{Y_k - EY_k}{n} \to 0 \quad a.s.$$

By Kronecker's lemma, we only N.T.S.

$$\sum_{k=1}^{n} \frac{Y_k - EY_k}{k} \text{ converges a.s.}$$

Taking $A = 1$ in Kolmogorov's Three-series Theorem, (i) and (ii) therein are automatically satisfied. For (iii), recall that we have verified in the first proof of SLLN that

$$\sum_{k=1}^{\infty} \mathrm{Var}\left(\frac{Y_k - EY_k}{k}\right) \leq \sum_{k=1}^{n} \frac{E(Y_k^2)}{k^2} < \infty.$$

$\square$

3.3 Marcinkiewicz-Zygmund SLLN

Finally, we present the following result.

Theorem 2.5.8 (Marcinkiewicz-Zygmund SLLN). Let $X_1, X_2, \dots$ be i.i.d. with $EX_i = 0, E|X_i|^p < \infty$ for some $1 < p < 2$. Then with $S_n = \sum_{i=1}^n X_i$, we have

$$\frac{S_n}{n^{1/p}} \to 0 \quad a.s..$$

Proof. By Kronecker's lemma, it suffices to prove

$$\sum_{n=1}^{\infty} \frac{X_n}{n^{1/p}} \quad \text{converges a.s.}$$

We choose $A = 1$ in Komogorov's Three-series Theorem, define

$$Y_i = \frac{X_i}{i^{1/p}} 1_{\{|X_i| \le i^{1/p}\}},$$

and verify the three conditions below.

(i) We have

$$\sum_{n=1}^{\infty} P(|\frac{X_n}{n^{1/p}}| > 1) = \sum_{n=1}^{\infty} P(|X_1|^p > n) \le E|X_1|^p < \infty.$$

(ii) We have

$$\begin{aligned}
\sum_{i=1}^{\infty} |EY_i| =& \sum_{i=1}^{\infty} \frac{1}{i^{1/p}} |EX_i 1_{\{|X_i| \le i^{1/p}\}}| \\
=& \sum_{i=1}^{\infty} \frac{1}{i^{1/p}} |EX_i 1_{\{|X_i| > i^{1/p}\}}| && \text{(from } EX_i = 0\text{)} \\
\le& \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \frac{1}{i^{1/p}} E[|X_1| 1_{\{j \le |X_1|^p < j+1\}}] && \text{(from the i.i.d. assumption)} \\
=& \sum_{j=1}^{\infty} \sum_{i=1}^{j} \frac{1}{i^{1/p}} E[|X_1| 1_{\{j \le |X_1|^p < j+1\}}] && \text{(Fubini)} \\
\le& C_p \sum_{j=1}^{\infty} j^{1-\frac{1}{p}} E[|X_1| 1_{\{j \le |X_1|^p < j+1\}}] && \text{(sum of geometric series)} \\
\le& C_p \sum_{j=1}^{\infty} E\left\{ (|X_1|^p)^{1-\frac{1}{p}} [|X_1| 1_{\{j \le |X_1|^p < j+1\}}] \right\} \\
\le& C_p E[|X_1|^p] < \infty.
\end{aligned}$$

(iii) We have, similarly as in verifying (ii),

$$\begin{aligned}
\sum_{i=1}^{\infty} E[Y_i^2] =& \sum_{i=1}^{\infty} \frac{1}{i^{2/p}} E(X_i^2 1_{\{|X_i| \le i^{1/p}\}}) \\
=& \sum_{i=1}^{\infty} \sum_{j=1}^{i} \frac{1}{i^{2/p}} E[X_1^2 1_{\{j-1 < |X_1|^p \le j\}}] \\
=& \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{1}{i^{2/p}} E[X_1^2 1_{\{j-1 < |X_1|^p \le j\}}] \\
\le& C_p \sum_{j=1}^{\infty} j^{1-\frac{2}{p}} E[X_1^2 1_{\{j-1 < |X_1|^p \le j\}}] \\
\le& C_p \sum_{j=1}^{\infty} E\left\{ ((|X_1|+1)^p)^{1-\frac{2}{p}} [X_1^2 1_{\{j \le |X_1|^p < j+1\}}] \right\} \\
\le& C_p E[(|X_1|+1)^p] < \infty.
\end{aligned}$$

$\square$

Exercise. Prove the reverse statement of the above theorem that if $\frac{S_n}{n^{1/p}} \to 0 \quad a.s.$, then $E|X_1|^p < \infty$. [Hint: follow the same proof of the reverse statement of SLLN.]

Exercise. Let $X_1, X_2, \dots$ be i.i.d. Suppose $E|X_i|^p < \infty$ for some $0 < p < 1$, then

$$\frac{S_n}{n^{1/p}} \to 0 \quad a.s..$$

Note that in this case, the expectation of X_1 may not exist.