5. Modes of Convergence

5.1. **Some important inequalities.** We start this section by proving a result known as Markov inequality.

Lemma 5.1 (Markov inequality). *If X is a non-negative random variable whose expected value exists, then for all $a > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof. Observe that, since X is non-negative

$$\mathbb{E}[X] = \mathbb{E}\Big[X\mathbf{1}_{\{X\geq a\}} + X\mathbf{1}_{\{X<a\}}\Big] \geq \mathbb{E}[X\mathbf{1}_{\{X\geq a\}}] \geq a\mathbb{P}(X \geq a).$$

Hence the result follows. □

Corollary 5.2. *If X is a random variable such that $\mathbb{E}[|X|] < +\infty$, then for all $a > 0$*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

Example 5.1. *A coin is weighted so that its probability of landing on heads is 20%. Suppose the coin is flipped 20 times. We want to find a bound for the probability it lands on heads at least 16 times. Let X be the number of times the coin lands on heads. Then $X \sim B(20, \frac{1}{5})$. We use Markov inequality to find the required bound:*

$$\mathbb{P}(X \geq 16) \leq \frac{\mathbb{E}(X)}{16} = \frac{4}{16} = \frac{1}{4}.$$

The actual probability that this happen is

$$\mathbb{P}(X \geq 16) = \sum_{k=16}^{20} \binom{20}{k} (0.2)^k (0.8)^{20-k} \approx 1.38 \times 10^{-8}\,.$$

Lemma 5.3 (Chebyschev's inequality). *Let Y be an integrable random variable such that* $\mathrm{Var}(Y) < +\infty$. *Then for any $\varepsilon > 0$*

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2}.$$

Proof. To get the result, take $X = |Y - \mathbb{E}(Y)|^2$ and $a = \varepsilon^2$ in Markov inequality. □

Example 5.2. *Is there any random variable X for which*

$$\mathbb{P}\big(\mu - 3\sigma \leq X \leq \mu + 3\sigma\big) = \frac{1}{2},$$

where $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$.
Solution: Observe that

$$\mathbb{P}\big(\mu - 3\sigma \leq X \leq \mu + 3\sigma\big) = \mathbb{P}\big(|X - \mu| \leq 3\sigma\big) = 1 - \mathbb{P}\big(|X - \mathbb{E}(X)| > 3\sigma\big).$$

By Chebyschev's inequality, we get that

$$\mathbb{P}\big(|X - \mathbb{E}(X)| > 3\sigma\big) \leq \mathbb{P}\big(|X - \mathbb{E}(X)| \geq 3\sigma\big) \leq \frac{\sigma^2}{9\sigma^2},$$

and hence

$$\mathbb{P}\big(\mu - 3\sigma \leq X \leq \mu + 3\sigma\big) \geq 1 - \frac{1}{9} = \frac{8}{9}.$$

Since $\frac{1}{2} < \frac{8}{9}$, there exists NO random variable X satisfying the given condition.

In principle Chebyshev's inequality asks about distance from the mean in either direction, it can still be used to give a bound on how often a random variable can take large values, and will usually give much better bounds than Markov's inequality. For example consider Example 5.1. Markov's inequality gives a bound of $\frac{1}{4}$. Using Chebyshev's inequality, we see that

$$\mathbb{P}(X \geq 16) = \mathbb{P}(X - 4 \geq 12) \leq \mathbb{P}(|X - 4| \geq 12) \leq \frac{\mathrm{Var}(X)}{144} = \frac{1}{45}\,.$$

Lemma 5.4 (One-sided Chebyschev inequality)**.** *Let X be a random variable with mean 0 and variance $\sigma^2 < +\infty$. Then for any $a > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}\,.$$

Proof. For any $b \geq 0$, we see that $X \geq a$ is equivalent to $X + b \geq a + b$. Hence by Markov's inequality, we have

$$\mathbb{P}(X \geq a) = \mathbb{P}(X + b \geq a + b) \leq \mathbb{P}((X + b)^2 \geq (a + b)^2) \leq \frac{\mathbb{E}[(X + b)^2]}{(a + b)^2} = \frac{\sigma^2 + b^2}{(a + b)^2}$$

$$\implies \mathbb{P}(X \geq a) \leq \min_{b \geq 0} \frac{\sigma^2 + b^2}{(a + b)^2} = \frac{\sigma^2}{\sigma^2 + a^2}\,.$$

□

One can use one-sided Chebyschev inequality to arrive at the following corollary.

Corollary 5.5. *If $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$, then for any $a > 0$,*

$$\mathbb{P}(X \geq \mu + a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$
$$\mathbb{P}(X \leq \mu - a) \leq \frac{\sigma^2}{\sigma^2 + a^2}\,.$$

Example 5.3. *Let X be a Poisson random variable with mean 20. Show that one-sided Cheby-shev inequality gives better upper bound on $\mathbb{P}(X \geq 26)$ compare to Markov and Chebyschev inequalities. Indeed, by Markov inequality, we have*

$$p = \mathbb{P}(X \geq 26) \leq \frac{\mathbb{E}[X]}{26} = \frac{10}{13}\,.$$

By Chebyschev inequality, we get

$$p = \mathbb{P}(X - 20 \geq 6) \leq \mathbb{P}(|X - 20| \geq 6) \leq \frac{\mathrm{Var}(X)}{36} = \frac{10}{18}\,.$$

One-sided Chebyschev inequality gives

$$p = \mathbb{P}(X - 20 \geq 6) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + 36} = \frac{10}{28}\,.$$

Theorem 5.6 (**Weak Law of Large Number**)**.** *Let $\{X_i\}$ be a sequence of iid random variables with finite mean μ and variance σ^2. Then for any $\varepsilon > 0$*

$$\mathbb{P}\big(|\frac{S_n}{n} - \mu| > \varepsilon\big) \leq \frac{\sigma^2}{n\varepsilon^2}\,,$$

where $S_n = \sum_{i=1}^n X_i$. In particular,

$$\lim_{n \to \infty} \mathbb{P}\big(|\frac{S_n}{n} - \mu| > \varepsilon\big) = 0\,.$$

Proof. First inequality follows from Chebyschev's inequality. Sending limit as n tends to infinity in the first inequality, we arrive at the second result. □

We shall discuss various modes of convergence for a given sequence of random variables $\{X_n\}$ defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 5.1 (Convergence in probability). We say that $\{X_n\}$ converges to a random variable X, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in probability if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

We denote it by $X_n \xrightarrow{\mathbb{P}} X$.

Example 5.4. *Let $\{X_n\}$ be a sequence of random variables such that $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$ and $\mathbb{P}(X_n = n) = \frac{1}{n}$. Then $X_n \xrightarrow{\mathbb{P}} 0$. Indeed for any $\varepsilon > 0$,*

$$\mathbb{P}(|X_n| > \varepsilon) = \begin{cases} \frac{1}{n} & \text{if } \varepsilon < n, \\ 0 & \text{if } \varepsilon \ge n\,. \end{cases}$$

Hence, $\lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0$.

Example 5.5. *Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\mathbb{P}(X_n = 1) = \frac{1}{2}$ and $\mathbb{P}(X_n = -1) = \frac{1}{2}$. Then $\frac{1}{n}\sum_{i=1}^n X_i$ converges to 0 in probability. Indeed for any $\varepsilon > 0$, thanks to weak law of large number, we have*

$$\mathbb{P}(|\frac{1}{n}S_n - \mu| > \varepsilon) \le \frac{\mathrm{Var}(X_1)}{n\varepsilon^2}$$

where $\mu = \mathbb{E}(X_1)$. Observe that $\mu = 0$ and $\mathrm{Var}(X_1) = 1$. Hence

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^n X_i| > \varepsilon) \le \frac{1}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

Theorem 5.7. *$X_n \xrightarrow{\mathbb{P}} X$ if and only if $\lim\limits_{n\to\infty} \mathbb{E}\big(\frac{|X_n - X|}{1+|X_n - X|}\big) = 0$.*

Proof. With out loss of generality, take $X = 0$. Thus, we want to show that $X_n \xrightarrow{\mathbb{P}} 0$ if and only if $\lim\limits_{n\to\infty} \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) = 0$.

Suppose $X_n \xrightarrow{\mathbb{P}} 0$. Then given $\varepsilon > 0$, we have $\lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0$. Now,

$$\begin{aligned} &\frac{|X_n|}{1+|X_n|} = \frac{|X_n|}{1+|X_n|}\mathbf{1}_{|X_n|>\varepsilon} + \frac{|X_n|}{1+|X_n|}\mathbf{1}_{|X_n|\le\varepsilon} \le \mathbf{1}_{|X_n|>\varepsilon} + \varepsilon \\ &\implies \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) \le \mathbb{P}(|X_n| > \varepsilon) + \varepsilon \implies \lim_{n\to\infty} \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) \le \varepsilon\,. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have $\lim_{n\to\infty} \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) = 0$.

Conversely, let $\lim_{n\to\infty} \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) = 0$. Observe that the function $f(x) = \frac{x}{1+x}$ is strictly increasing on $[0, \infty)$. Thus,

$$\begin{aligned} &\frac{\varepsilon}{1+\varepsilon}\mathbf{1}_{|X_n|>\varepsilon} \le \frac{|X_n|}{1+|X_n|}\mathbf{1}_{|X_n|>\varepsilon} \le \frac{|X_n|}{1+|X_n|} \\ &\implies \frac{\varepsilon}{1+\varepsilon}\mathbb{P}(|X_n| > \varepsilon) \le \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) \\ &\implies \lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) \le \frac{1+\varepsilon}{\varepsilon} \lim_{n\to\infty} \mathbb{E}\big(\frac{|X_n|}{1+|X_n|}\big) = 0 \implies X_n \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

$\square$

Definition 5.2 (Convergence in r-th mean). Let $X, \{X_n\}$ be random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for $r \in \mathbb{N}$, $\mathbb{E}[|X|^r] < \infty$ and $\mathbb{E}[|X_n|^r] < \infty$ for all n. We say that $\{X_n\}$ converges in the r-th mean to X, denoted by $X_n \xrightarrow{r} X$, if the following holds:

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0.$$

Example 5.6. *Let $\{X_n\}$ be i.i.d. random variables with $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$. Define $Y_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then $Y_n \xrightarrow{2} \mu$. Indeed*

$$\mathbb{E}[|Y_n - \mu|^2] = \mathbb{E}\Big[\Big|\frac{\sum_{i=1}^n X_i - n\mu}{n}\Big|^2\Big] = \frac{1}{n^2}\mathbb{E}[|S_n - \mathbb{E}(S_n)|^2] = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{\sigma^2}{n}$$

where $S_n = \sum_{i=1}^n X_i$. Hence $Y_n \xrightarrow{2} \mu$.

Theorem 5.8. *The following holds:*

i) $X_n \xrightarrow{r} X \implies X_n \xrightarrow{\mathbb{P}} X$ for any $r \geq 1$.

ii) Let f be a given continuous function. If $X_n \xrightarrow{\mathbb{P}} X$, then $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Proof. Proof of (i) follows from Markov's inequality. Indeed, for any given $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^r]}{\varepsilon^r} \\ &\implies \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^r} \lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0. \end{aligned}$$

Proof of (ii): For any $k > 0$, we see that

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|f(X_n) - f(X)| > \varepsilon,\ |X| \leq k\} \cup \{|X| > k\}.$$

Since f is continuous, it is uniformly continuous on any bounded interval. Therefore, for any given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ if $|x - y| \leq \delta$ for x and y in $[-k, k]$. This means that

$$\{|f(X_n) - f(X)| > \varepsilon,\ |X| \leq k\} \subset \{|X_n - X| > \delta,\ |X| \leq k\} \subset \{|X_n - X| > \delta\}.$$

Thus we have

$$\begin{aligned} &\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|X_n - X| > \delta\} \cup \{|X| > k\} \\ &\implies \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \delta) + \mathbb{P}(|X| > k). \end{aligned}$$

Since $X_n \xrightarrow{\mathbb{P}} X$ and $\lim_{k\to\infty} \mathbb{P}(|X| > k) = 0$, we obtain that $\lim_{n\to\infty} \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) = 0$. This completes the proof. □

In general, convergence in probability does not imply convergence in r-th mean. To see it, consider the following example.

Example 5.7. *Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ and $\mathbb{P}(dx) = dx$. Let $X_n = n\mathbf{1}_{(0,1/n)}$. Then $X_n \xrightarrow{\mathbb{P}} 0$ but $X_n \stackrel{r}{\nrightarrow} 0$ for all $r \geq 1$. To show this, observe that*

$$\mathbb{P}(|X_n| > \varepsilon) \leq \frac{1}{n} \implies \lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0 \quad i.e.,\ X_n \xrightarrow{\mathbb{P}} 0.$$

On the other hand, for $r \geq 1$

$$\mathbb{E}[|X_n|^r] = \int_0^{\frac{1}{n}} n^r\, dx = n^{r-1} \nrightarrow 0 \quad \text{as } n \to \infty.$$

Definition 5.3 (**Almost sure convergence).** Let $X, \{X_n\}$ be random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $\{X_n\}$ converges to X almost surely or **with probability** 1 if the following holds:

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

We denote it by $X_n \overset{a.s}{\to} X$.

Example 5.8. *Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ and $\mathbb{P}(dx) = dx$. Define*

$$X_n(\omega) = \begin{cases} 1 & \textit{if} \quad \omega \in (0, 1-\frac{1}{n}) \\ n, & \textit{otherwise.} \end{cases}$$

It is easy to check that if $\omega = 0$ or $\omega = 1$, then $\lim_{n\to\infty} X_n(\omega) = \infty$. For any $\omega \in (0,1)$, we can find $n_0 \in \mathbb{N}$ such that $\omega \in (0, 1-\frac{1}{n})$ for any $n \ge n_0$. As a consequence, $X_n(\omega) = 1$ for any $n \ge n_0$. In other words, for $\omega \in (0,1)$, $\lim_{n\to\infty} X_n(\omega) = 1$. Define $X(\omega) = 1$ for all $\omega \in [0,1]$. Then

$$\mathbb{P}(\omega \in [0,1] : \ \{X_n(\omega)\} \textit{ does not converge to } X(\omega)) = \mathbb{P}(\{0,1\}) = 0 \implies X_n \overset{a.s.}{\to} 1.$$

Sufficient condition for almost sure convergence: Let $\{A_n\}$ be a sequence of events in $\mathcal{F}$. Define

$$\limsup_n A_n = \cap_{n=1}^{\infty} \big(\cup_{m\ge n} A_m \big) = \lim_{n\to\infty} \big(\cup_{m\ge n} A_m \big).$$

This can be interpreted probabilistically as

$$\limsup_n A_n = \text{“ } A_n \text{ occurs infinitely often”}.$$

We denote this as

$$\{A_n \text{ i.o.}\} = \limsup_n A_n.$$

Theorem 5.9 (**Borel-Cantelli lemma).** *Let $\{A_n\}$ be a sequence of events in $(\Omega, \mathcal{F}, \mathbb{P})$.*

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

ii) If A_n are mutually independent events, and if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\mathbb{P}(A_n \text{ i.o.}) = 1.$$

Remark 5.1. For mutually independent events A_n, since $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ is either finite or infinite, the event $\{A_n \text{ i.o.}\}$ has probability either 0 or 1. This is sometimes called *zero-one* law.

As a consequence of Borel-Cantelli lemma, we have the following proposition.

Proposition 5.10. *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \varepsilon) < +\infty$ for any $\varepsilon > 0$, then $X_n \overset{a.s.}{\to} 0$.*

Proof. Fix $\varepsilon > 0$. Let $A_n = \{|X_n| > \varepsilon\}$. Then $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty$, and hence by Borel-Cantelli lemma, $\mathbb{P}(A_n \text{ i.o.}) = 0$. Now

$$\big(\limsup_n A_n\big)^c = \{\omega : \exists\, n_0(\omega) \, \text{such that} \, |X_n(\omega)| \le \varepsilon \quad \forall\, n \ge n_0(\omega)\} := B_\varepsilon.$$

Thus, $\mathbb{P}(B_\varepsilon) = 1$. Let $B = \cap_{r=1}^{\infty} B_{\frac{1}{r}} \implies B^c = \cup_{r=1}^{\infty} B_{\frac{1}{r}}^c$. Moreover, since $\mathbb{P}(B_{\varepsilon=\frac{1}{r}}) = 1$, we have $\mathbb{P}(B_{\frac{1}{r}}^c) = 0$. Observe that

$$\{\omega : \lim_{n\to\infty} |X_n(\omega)| = 0\} = \cap_{r=1}^{\infty} B_{\frac{1}{r}}.$$

Again, $\mathbb{P}(B^c) \le \sum_{r=1}^{\infty} \mathbb{P}(B_{\frac{1}{r}}^c) = 0$, and hence $\mathbb{P}(B) = 1$. In other words,

$$\mathbb{P}\big(\{\omega : \lim_{n\to\infty} |X_n(\omega)| = 0\}\big) = 1, \quad \text{i.e.,} \ X_n \overset{a.s.}{\to} 0.$$

□

Example 5.9. *Let $\{X_n\}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(X_n = 1) = \frac{1}{2}$ and $\mathbb{P}(X_n = -1) = \frac{1}{2}$. Let $S_n = \sum_{i=1}^{n} X_i$. Then $\frac{1}{n^2} S_{n^2} \stackrel{a.s.}{\to} 0$. To show the result, we use Proposition 5.10. Note that*

$$\mathbb{P}(\frac{1}{n^2}|S_{n^2}| > \varepsilon) \le \frac{\mathbb{E}[|S_{n^2}|^2]}{n^4\varepsilon^2} \le \frac{1}{n^2\varepsilon^2} \implies \sum_{n=1}^{\infty} \mathbb{P}(\frac{1}{n^2}|S_{n^2}| > \varepsilon) < \infty \implies \frac{1}{n^2} S_{n^2} \stackrel{a.s.}{\to} 0.$$

Let us consider the following example.

Example 5.10. *Let $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and $\mathbb{P}(dx) = dx$. Define*

$$X_n = \mathbf{1}_{[\frac{j}{2^k}, \frac{j+1}{2^k}]}, \quad n = 2^k + j,\ j \in \{0, \dots, 2^k - 1\}, \quad k = \lfloor \log_2(n) \rfloor.$$

Note that, for each positive integer n, there exist integers j and k(uniquely determined) such that

$$n = 2^k + j, \ ,\ j \in \{0, \dots, 2^k - 1\}, \quad k = \lfloor \log_2(n) \rfloor.$$

(for $n = 1$, $k = j = 0$, and for $n = 5$, $k = 2, j = 1$ and so on). Let $A_n = \{X_n > 0\}$. Then, clearly $\mathbb{P}(A_n) \to 0$. Consequently, $X_n \stackrel{\mathbb{P}}{\to} 0$ but $X_n(\omega) \nrightarrow 0$ for all $\omega \in \Omega$.

Theorem 5.11. *The followings hold.*

i) If $X_n \stackrel{a.s}{\to} X$, then $X_n \stackrel{\mathbb{P}}{\to} X$.
ii) If $X_n \stackrel{\mathbb{P}}{\to} X$, then there exists a subsequence X_{n_k} of X_n such that $X_{n_k} \stackrel{a.s.}{\to} X$.
iii) If $X_n \stackrel{a.s}{\to} X$, then for any continuous function f, $f(X_n) \stackrel{a.s}{\to} f(X)$.

Proof. Proof of i): For any $\varepsilon > 0$, define $A_n^\varepsilon = \{|X_n - X| > \varepsilon\}$ and $B_m^\varepsilon = \cup_{n=m}^{\infty} A_n^\varepsilon$. Since $X_n \stackrel{a.s}{\to} X$, $\mathbb{P}(\cap_m B_m^\varepsilon) = 0$. Note that $\{B_m^\varepsilon\}$ are nested and decreasing sequence of events. Hence from the continuity of probability measure $\mathbb{P}$, we have

$$\lim_{m\to\infty} \mathbb{P}(B_m^\varepsilon) = \mathbb{P}(\cap_m B_m^\varepsilon) = 0.$$

Since $A_m^\varepsilon \subset B_m^\varepsilon$, we have $\mathbb{P}(A_m^\varepsilon) \le \mathbb{P}(B_m^\varepsilon)$. This implies that $\lim_{m\to\infty} \mathbb{P}(A_m^\varepsilon) = 0$. In other words, $X_n \stackrel{\mathbb{P}}{\to} X$.

Proof of ii): To prove ii), we will use Borel-Cantelli lemma. Since $X_n \stackrel{\mathbb{P}}{\to} X$, we can choose a subsequence X_{n_k} such that $\mathbb{P}(|X_{n_k} - X| > \frac{1}{k}) \le \frac{1}{2^k}$. Let $A_k := \{|X_{n_k} - X| > \frac{1}{k}\}$. Then $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < +\infty$. Hence, by Borel-Cantelli lemma $\mathbb{P}(A_k \text{ i.o.}) = 0$. This implies that

$$\mathbb{P}(\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c) = 1 \implies \mathbb{P}\big(\{\omega \in \Omega : \exists\, n_0 :\ \forall k \ge n_0,\ |X_{n_k} - X| \le \frac{1}{k}\}\big) = 1$$
$$\implies X_{n_k} \stackrel{a.s.}{\to} X.$$

Proof of iii): Let $N = \{\omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}$. Then $\mathbb{P}(N) = 0$. If $\omega \notin N$, then by the continuity property of f, we have

$$\lim_{n\to\infty} f(X_n(\omega)) = f(\lim_{n\to\infty} X_n(\omega)) = f(X(\omega))\,.$$

This is true for any $\omega \notin N$ and $\mathbb{P}(N) = 0$. Hence $f(X_n) \stackrel{a.s}{\to} f(X)$. □

Definition 5.4 (Convergence in distribution). Let $X, X_1, X_2, \ldots$ be real-valued random variables with distribution functions $F_X, F_{X_1}, F_{X_2}, \ldots$ respectively. We say that (X_n) converges to X is distribution, denoted by $X_n \overset{d}{\to} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{for all continuity points } x \text{ of } F_X.$$

Remark 5.2. In the above definition, the random variables $X, \{X_n\}$ need not be defined on the same probability space.

Example 5.11. *Let $X_n = \frac{1}{n}$ and $X = 0$. Then*

$$F_{X_n}(x) = \mathbb{P}(X_n \le x) = \begin{cases} 1 & \text{if } x \ge \frac{1}{n} \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \quad F_X(x) = \begin{cases} 1 & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Observe that 0 is the only discontinuity point of F_X, and $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for $x \ne 0$. Thus, $X_n \overset{d}{\to} 0$.

Example 5.12. *Let X be a real-valued random variable with distribution function F. Define $X_n = X + \frac{1}{n}$. Then*

$$\begin{aligned} &F_{X_n}(x) = \mathbb{P}(X + \frac{1}{n} \le x) = F(x - \frac{1}{n}) \\ &\implies \lim_{n\to\infty} F_{X_n}(x) = F(x-) = F(x) \quad \text{for continuity point } x \text{ of } F. \end{aligned}$$

This implies that $X_n \overset{d}{\to} X$.

Theorem 5.12. *$X_n \overset{\mathbb{P}}{\to} X$ implies that $X_n \overset{d}{\to} X$.*

Proof. Let $\varepsilon > 0$. Since $F_{X_n}(t) = \mathbb{P}(X_n \le t)$, we have

$$\begin{aligned} F_{X_n}(t) &= \mathbb{P}(X_n \le t, |X_n - X| > \varepsilon) + P(X_n \le t, |X_n - X| \le \varepsilon) \\ &\le \mathbb{P}(|X_n - X| > \varepsilon) + \mathbb{P}(X_n \le t, |X_n - X| \le \varepsilon) \\ &\le \mathbb{P}(|X_n - X| > \varepsilon) + \mathbb{P}(X \le t + \varepsilon) \\ &\le \mathbb{P}(|X_n - X| > \varepsilon) + F_X(t + \varepsilon), \end{aligned}$$

and

$$\begin{aligned} F_X(t - \varepsilon) = \mathbb{P}(X \le t - \varepsilon) &= \mathbb{P}(X \le t - \varepsilon, |X_n - X| > \varepsilon) + \mathbb{P}(X \le t - \varepsilon, |X_n - X| \le \varepsilon) \\ &\le \mathbb{P}(|X_n - X| > \varepsilon) + \mathbb{P}(X \le t - \varepsilon, |X_n - X| \le \varepsilon) \\ &\le \mathbb{P}(|X_n - X| > \varepsilon) + \mathbb{P}(X_n \le t) \\ &\le \mathbb{P}(|X_n - X| > \varepsilon) + F_{X_n}(t)\,. \end{aligned}$$

Thus, since $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$, we obtain from the above inequalities

$$F_X(t - \varepsilon) \le \liminf_{n\to\infty} F_{X_n}(t) \le \limsup_{n\to\infty} F_{X_n}(t) \le F_X(t + \varepsilon).$$

Thet t be the continuity point of F. Then sending $\varepsilon \to 0$ in the above inequality, we get

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t), \quad \text{i.e., } X_n \overset{d}{\to} X.$$

□

Converse of this theorem is NOT true in general.

Example 5.13. *Let $X \sim \mathcal{N}(0,1)$. Define $X_n = -X$ for $n = 1, 2, 3, \ldots$. Then $X_n \sim \mathcal{N}(0,1)$ and hence $X_n \overset{d}{\to} X$. But*

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|2X| > \varepsilon) = \mathbb{P}(|X| > \frac{\varepsilon}{2}) \ne 0 \implies X_n \overset{\mathbb{P}}{\nrightarrow} X.$$

Theorem 5.13 (**Continuity theorem**). *Let $X, \{X_n\}$ be random variables having the characteristic function $\phi_X, \{\phi_{X_n}\}$ respectively. Then the followings are equivalent.*

i) $X_n \xrightarrow{d} X$.
ii) $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ *for all bounded Lipschitz continuous function.*
iii) $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ *for all* $t \in \mathbb{R}$.

Example 5.14. *Let $\{X_n\}$ be a sequence of Poisson random variables with parameter $\lambda_n = n$. Define $Z_n = \frac{X_n - n}{\sqrt{n}}$. Then*

$$Z_n \xrightarrow{d} Z, \quad \text{where } \mathcal{L}(Z) = \mathcal{N}(0,1).$$

Solution: *To see this, we use Levy's continuity theorem. Let $\Phi_{Z_n} : \mathbb{R} \to \mathbb{C}$ be the characteristic function of Z_n. Then we have*

$$\Phi_{Z_n}(u) = \mathbb{E}\big[e^{iuZ_n}\big] = e^{-iu\sqrt{n}}\mathbb{E}\big[e^{i\frac{u}{\sqrt{n}}X_n}\big] = e^{-iu\sqrt{n}}e^{n(e^{\frac{iu}{\sqrt{n}}}-1)}.$$

Using Taylor's series expansion, we have

$$e^{\frac{iu}{\sqrt{n}}} - 1 = \frac{iu}{\sqrt{n}} - \frac{u^2}{2n} - \frac{iu^3}{6n^{\frac{3}{2}}} + \dots$$

and hence we get

$$\Phi_{Z_n}(u) = e^{-iu\sqrt{n}}e^{n(\frac{iu}{\sqrt{n}} - \frac{u^2}{2n} - \frac{iu^3}{6n^{\frac{3}{2}}} + \dots)} = e^{-iu\sqrt{n} + -iu\sqrt{n}}e^{-\frac{u^2}{2}}e^{-\frac{h(u,n)}{\sqrt{n}}}$$

where $h(u,n)$ stays bounded in n for each u and hence $\lim_{n\to\infty}\frac{h(u,n)}{\sqrt{n}} = 0$. Therefore, we have

$$\lim_{n\to\infty}\Phi_{Z_n}(u) = \lim_{n\to\infty} e^{-iu\sqrt{n} + -iu\sqrt{n}}e^{-\frac{u^2}{2}}e^{-\frac{h(u,n)}{\sqrt{n}}} = e^{-\frac{u^2}{2}}.$$

Since $e^{-\frac{u^2}{2}}$ is the characteristic function of $\mathcal{N}(0,1)$, we conclude that $Z_n \xrightarrow{d} Z$.

Theorem 5.14 (**Strong law of large number (SLLN)**). *Let $\{X_i\}$ be a sequence of i.i.d. random variables with finite mean μ and variance σ^2. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu, \quad \text{where } S_n = \sum_{i=1}^n X_i.$$

The special case of 4-th order moment, above theorem is refered as **Borel's SLLN**. Before going to prove the theorem, let us consider some examples.

Example 5.15. *1) Let $\{X_n\}$ be a sequence of i.i.d. random variables that are bounded, i.e., there exists $C < \infty$ such that $\mathbb{P}(|X_1| \le C) = 1$. Then $\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}(X_1)$.*
2) Let $\{X_n\}$ be a sequence of i.i.d. Bernoulli(p) random variables. Then $\mu = p$ and $\sigma^2 = p(1-p)$. Hence by SLLN theorem, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i = p \quad \text{with probability } 1.$$

To prove the theorem, we need following lemma.

Lemma 5.15. *Let $\{X_i\}$ be a sequence of random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

i) If $\{X_n\}$ are positive, then

$$\mathbb{E}\Big(\sum_{n=1}^\infty X_n\Big) = \sum_{n=1}^\infty \mathbb{E}[X_n]. \tag{5.1}$$

ii) If $\sum_{n=1}^\infty \mathbb{E}[|X_n|] < \infty$, then $\sum_{i=1}^\infty X_i$ converges almost surely and (5.1) holds as well.

Proof of Strong law of large number: With out loss of generality we can assume that $\mu = 0$ (otherwise, if $\mu \neq 0$, then set $\tilde{X}_i = X_i - \mu$, and work with $\tilde{X}_i$). Set $Y_n = \frac{S_n}{n}$. Observe that, thanks to independent property,

$$\mathbb{E}[Y_n] = 0, \quad \mathbb{E}[Y_n^2] = \frac{1}{n^2} \sum_{1 \le j,k \le n} \mathbb{E}(X_j X_k) = \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{E}[X_j^2] = \frac{\sigma^2}{n}.$$

Thus, $\lim_{n\to\infty} \mathbb{E}[Y_n^2] = 0$, and hence along a subsequence, Y_n converges to 0 almost surely. But we need to show that *original* sequence converges to 0 with probability 1. To do so, we proceed as follows. Since $\mathbb{E}[Y_n^2] = \frac{\sigma^2}{n}$, we see that $\sum_{n=1}^{\infty} \mathbb{E}[Y_{n^2}^2] = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < +\infty$ and hence by Lemma 5.15, ii), $\sum_{n=1}^{\infty} Y_{n^2}^2$ converges almost surely. Thus,

$$\lim_{n\to\infty} Y_{n^2} = 0 \quad \text{with probability } 1. \tag{5.2}$$

Let $n \in \mathbb{N}$. Then there exists $m(n) \in \mathbb{N}$ such that $(m(n))^2 \le n < (m(n)+1)^2$. Now

$$\begin{aligned}
Y_n - \frac{(m(n))^2}{n} Y_{(m(n))^2} &= \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{(m(n))^2}{n} \Big(\frac{1}{(m(n))^2} \sum_{i=1}^{(m(n))^2} X_i \\
&= \frac{1}{n} \sum_{i=1+(m(n))^2}^{n} X_i \\
\implies \mathbb{E}\Big[\Big(Y_n - \frac{(m(n))^2}{n} Y_{(m(n))^2}\Big)^2\Big] &= \frac{1}{n^2} \sum_{i=1+(m(n))^2}^{n} \mathbb{E}[X_i^2] \\
&= \frac{n-(m(n))^2}{n^2}\sigma^2 \le \frac{2m(n)+1}{n^2}\sigma^2 \quad (\because n < (m(n)+1)^2) \\
&\le \frac{2\sqrt{n}+1}{n^2}\sigma^2 \le \frac{3\sigma^2}{n^{\frac{3}{2}}} \quad (\because m(n) \le \sqrt{n}) \\
\implies \sum_{n=1}^{\infty} \mathbb{E}\Big[\Big(Y_n - \frac{(m(n))^2}{n} Y_{(m(n))^2}\Big)^2\Big] &\le \sum_{n=1}^{\infty} \frac{3\sigma^2}{n^{\frac{3}{2}}} < +\infty .
\end{aligned}$$

Thus, again by Lemma 5.15, ii), we conclude that

$$\lim_{n\to\infty} Y_n - \frac{(m(n))^2}{n} Y_{(m(n))^2} = 0 \quad \text{with probability } 1. \tag{5.3}$$

Obsere that $\lim_{n\to\infty} \frac{(m(n))^2}{n} = 1$. Thus, in view of (5.2) and (5.3), we conclude that

$$\lim_{n\to\infty} Y_n = 0 \quad \text{with probability } 1.$$

This completes the proof.

Theorem 5.16 (Kolmogorov's strong law of large numbers). *Let $\{X_n\}$ be a sequence of i.i.d. random variables and $\mu \in \mathbb{R}$. Then*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu \quad \textit{a.s. if and only if } \mathbb{E}[X_n] = \mu.$$

In this case, the convergence also holds in L^1.

Example 5.16 (Monte Carlo approximation). *Let f be a measurable function in $[0,1]$ such that $\int_0^1 |f(x)|\,dx < \infty$. Let $\alpha = \int_0^1 f(x)\,dx$. In general we cannot obtain a closed form expression* *for α and need to estimate it. Let $\{U_j\}$ be a sequence of independent uniformly random variables on $[0,1]$. Then by Theorem 5.16,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} f(U_j) = \mathbb{E}[f(U_j)] = \int_0^1 f(x)\,dx$$

a.s. and in L^2. Thus, to get an approximation of $\int_0^1 f(x)\,dx$, we need to simulate the uniform random variables U_j (by using a random number generator).

Theorem 5.17 **(Central limit theorem).** *Let $\{X_n\}$ be a sequence of i.i.d. random variables with finite mean μ and variance σ^2 with $0<\sigma^2<+\infty$. Let $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then Y_n converges in distribution to Y , where $\mathcal{L}(Y) = \mathcal{N}(0,1)$.*

Proof. With out loss of generality, we assume that $\mu = 0$. Let Φ, Φ_{Y_n} be the characteristic function of X_j and Y_n respectively. Since $\{X_j\}$ are i.i.d., we have

$$\begin{aligned}\Phi_{Y_n}(u) &= \mathbb{E}[e^{iuY_n}] = \mathbb{E}[e^{iu\frac{S_n}{\sigma\sqrt{n}}}] = \mathbb{E}[e^{iu\frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}}] \\ &= \mathbb{E}\Big[\prod_{i=1}^{n} e^{iu\frac{X_i}{\sigma\sqrt{n}}}\Big] = \prod_{i=1}^{n}\mathbb{E}\big[e^{iu\frac{X_i}{\sigma\sqrt{n}}}\big] = \big(\Phi(\frac{u}{\sigma\sqrt{n}})\big)^n.\end{aligned}$$

Since $\mathbb{E}[|X_j|^2] < +\infty$, the function Φ has two continuous derivatives. In particular,

$$\Phi'(u) = i\mathbb{E}[X_j e^{iuX_j}], \quad \Phi''(u) = -\mathbb{E}[X_j^2 e^{iuX_j}] \implies \Phi'(0)=0,\ \Phi''(0) = -\sigma^2\,.$$

Expanding Φ in a Taylor expansion about $u=0$, we have

$$\Phi(u) = 1 - \frac{\sigma^2 u^2}{2} + h(u)u^2, \quad \text{where } h(u)\to 0 \text{ as } u\to 0.$$

Thus, we get

$$\begin{aligned}&\Phi_{Y_n}(u) = e^{n\log(\Phi(\frac{u}{\sigma\sqrt{n}}))} = e^{n\log(1-\frac{u^2}{2n}+\frac{u^2}{n\sigma^2}h(\frac{u}{\sigma\sqrt{n}}))} \\ &\implies \lim_{n\to\infty}\Phi_{Y_n}(u) = e^{-\frac{u^2}{2}} = \Phi_Y(u) \quad \text{(by L'Hôpital rule)}.\end{aligned}$$

Hence by Levy's continuity theorem, we conclude that Y_n converges in distribution to Y with $\mathcal{L}(Y) = \mathcal{N}(0,1)$. □

Remark 5.3. If $\sigma^2 = 0$, then $X_j = \mu$ a.s. for all j, and hence $\frac{S_n}{n} = \mu$ a.s.

One can weaken slightly the hypotheses of Theorem 5.17. Indeed, we have the following Central limit theorem.

Theorem 5.18. *Let $\{X_n\}$ be independent but not necessarily identicallu distributed. Let $\mathbb{E}[X_n] = 0$ for all n, ane let $\sigma_n^2 = \mathrm{Var}(X_n)$. Assume that*

$$\sup_n \mathbb{E}[|X_n|^{2+\varepsilon}] < +\infty \textit{ for some } \varepsilon>0, \quad \sum_{n=1}^{\infty}\sigma_n^2 = \infty.$$

Then $\frac{S_n}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$ converges in distribution to Y with $\mathcal{L}(Y) = \mathcal{N}(0,1)$.

Example 5.17. *Let $\{X_n\}$ be a sequence of i.i.d random variables such that $\mathbb{P}(X_n = 1) = \frac{1}{2}$ and $\mathbb{P}(X_n = 0) = \frac{1}{2}$. Then $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{4}$. Hence by central limit theorem, $Y_n = \frac{2S_n - n}{\sqrt{n}}$ converges in distribution to Y with $\mathcal{L}(Y) = \mathcal{N}(0,1)$.*

Example 5.18. *Let $X \sim B(n,p)$. For any given $0 \le \alpha \le 1$, we want to find n such that $\mathbb{P}(X > \frac{n}{2}) \le 1-\alpha$. We can thaink of X as a sum of n i.i.d. random variable X_i such that $X_i \sim B(1,p)$. Hence by central limit theorem, for large n,*

$$\mathbb{P}(X - np \le x\sqrt{np(1-p)}) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du.$$

Choose x such that $np + x\sqrt{np(1-p)} = \frac{n}{2}$. This implies that $x = \frac{\sqrt{n}}{2}\frac{1-2p}{\sqrt{p(1-p)}}$. Thus,

$$\mathbb{P}(X > \frac{n}{2}) = 1 - \mathbb{P}(X \le \frac{n}{2}) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\sqrt{n}}{2}\frac{1-2p}{\sqrt{p(1-p)}}} e^{-\frac{u^2}{2}}\,du\,.$$

Therefore, we need to choose n such that

$$\begin{aligned}\alpha &\le \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\sqrt{n}}{2}\frac{1-2p}{\sqrt{p(1-p)}}} e^{-\frac{u^2}{2}}\,du \implies \sqrt{n} \ge \frac{2\sqrt{p(1-p)}}{1-2p}\Phi^{-1}(\alpha)\\ &\implies n \ge \frac{4p(1-p)}{(1-2p)^2}\big(\Phi^{-1}(\alpha)\big)^2.\end{aligned}$$

Example 5.19. *Let $\{X_i\}$ be sequence of i.i.d. random variables with* $\exp(1)$ *distributed. Let $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$. How large should n be such that*

$$\mathbb{P}(0.9 \le \bar{X} \le 1.1) \ge 0.95?$$

Since X_i's are $\exp(1)$ *distributed, $\mu = \mathbb{E}[X_i] = 1$ and $\sigma^2 = \mathrm{Var}(X_i) = 1$. Let $Y = \sum_{i=1}^n X_i$. Then by central limit theorem, $\frac{Y-n}{\sqrt{n}}$ is approximately $\mathcal{N}(0,1)$. Now*

$$\begin{aligned}\mathbb{P}(0.9 \le \bar{X} \le 1.1) &= \mathbb{P}((0.9)n \le Y \le (1.1)n) = \mathbb{P}(\frac{(0.9)n-n}{\sqrt{n}} \le \frac{Y-n}{\sqrt{n}} \le \frac{(1.1)n-n}{\sqrt{n}})\\ &= \mathbb{P}\big(-(0.1)\sqrt{n} \le \frac{Y-n}{\sqrt{n}} \le (0.1)\sqrt{n}\big) = \Phi((0.1)\sqrt{n}) - \Phi(-(0.1)\sqrt{n})\\ &= 2\Phi((0.1)\sqrt{n}) - 1 \quad (\therefore \Phi(-x) = 1-\Phi(x))\end{aligned}$$

Hence we need to find n such that

$$\begin{aligned}&2\Phi((0.1)\sqrt{n}) - 1 \ge 0.95 \implies \Phi((0.1)\sqrt{n}) \ge 0.975 \implies (0.1)\sqrt{n} \ge \Phi^{-1}(0.975) = 1.96\\ &\implies \sqrt{n} \ge 19.6 \implies n \ge 384.16 \implies n \ge 385 \quad (\therefore n \in \mathbb{N}).\end{aligned}$$