13. Week 13

Remark 13.1. The WLLN suggests that for large sample sizes, the sample mean based on a random sample from a given distribution (also referred to as a population) is close to the expectation/mean of the distribution, in the sense of convergence in probability. In practice, this principle can be used to find an approximate value of the expectation of a distribution.

Example 13.2. Let $\{X_n\}_n$ be i.i.d. RVs with the common distribution $Bernoulli(p)$ for some $p > 0$. Here, we may visualize X_n's as a sequence of coin tosses with probability of success (obtaining head) as p. By the WLLN, $\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1 = p$, i.e. for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - p| \geq \epsilon) = 0$. This supports the intuitive notion that by tossing a coin, with unknown p, a large number of times we can make an educated guess about the value of p.

Example 13.3. Continuing with the discussion of the previous example, we can justify the working methodology of assigning probabilities by a relative frequency approach. Suppose we repeat a random experiment n times and observe whether an event E occurs or not in each trial. For $i = 1, 2, \cdots, n$, we consider an RV X_i to be 1 if E occurs and 0 otherwise. As discussed earlier in Remark 7.20, $X_i \sim Bernoulli(p)$, where $p = \mathbb{P}(E)$. If p is unknown, then by the WLLN we have $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbb{P}} p$, i.e., the observed relative frequency $\frac{1}{n}\sum_{i=1}^n X_i$ in first n trials approximates p in probability, for large n.

Theorem 13.4 (Continuous Mapping Theorem for convergence in Probability)**.** *Let $\{X_n\}_n$ be a sequence of RVs converging to X in probability. Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then $\{h(X_n)\}_n$ converges to $h(X)$ in probability.*

Note 13.5. In general, continuous mapping theorem is not true for convergence in r-th mean. Construction of such an example is left as an exercise in the problem set 10.

The proof of the next result is not part of the course.

Theorem 13.6 (Algebraic Operations with Convergence in Probability)**.** *Let $\{X_n\}_n$ and $\{Y_n\}$ be sequences of RVs such that $X_n \xrightarrow[n\to\infty]{\mathbb{P}} x$ and $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} y$ for some $x, y \in \mathbb{R}$. Let $\{a_n\}_n$ and $\{b_n\}_n$ be sequences in $\mathbb{R}$ converging to $a, b \in \mathbb{R}$ respectively. Then the following statements hold.*

(a) $X_n + Y_n \xrightarrow[n\to\infty]{\mathbb{P}} x + y$.
(b) $X_n - Y_n \xrightarrow[n\to\infty]{\mathbb{P}} x - y$.
(c) $X_n Y_n \xrightarrow[n\to\infty]{\mathbb{P}} xy$.
(d) $\frac{X_n}{Y_n} \xrightarrow[n\to\infty]{\mathbb{P}} \frac{x}{y}$, *provided* $y \neq 0$.
(e) $a_n X_n + b_n \xrightarrow[n\to\infty]{\mathbb{P}} ax + b$.

Remark 13.7. Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists. Consider the sample variance $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}(\bar{X}_n)^2$. By the assumption, the RVs X_i^2 are i.i.d. with finite expectation $\mathbb{E}X_1^2$, and hence by the WLLN (Theorem 12.24)

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1^2.$$

Again by WLLN $\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1$ and by Theorem 13.4 applied to the function $h(x) = x^2, \forall x \in \mathbb{R}$, we have

$$(\bar{X}_n)^2 \xrightarrow[n\to\infty]{\mathbb{P}} (\mathbb{E}X_1)^2.$$

Since $\frac{n}{n-1} \xrightarrow{n\to\infty} 1$, using Theorem 13.6, we have $S_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} Var(X_1)$. By Theorem 13.4, we have $S_n \xrightarrow[n\to\infty]{\mathbb{P}} \sqrt{Var(X_1)}$.

Definition 13.8 (Almost Sure Convergence)**.** Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If there exists $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$ such that for all $\omega \in A$, we have $\lim_n X_n(\omega) = X(\omega)$, then we say that the sequence $\{X_n\}_n$ converges almost surely (a.s., in short) to X and write $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X$.

Remark 13.9. It can be shown that almost sure convergence implies convergence in probability.

Remark 13.10 (a.s. convergence and convergence in r-th mean are not comparable). We show by examples that a.s. convergence does not imply convergence in r-th mean and vice versa. First, we consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ given by $\Omega = [0, 1]$ and $\mathbb{P}((a, b]) = b - a, \forall\ 0 \le a < b \le 1$. It can be shown that $\mathbb{P}$ is a probability function on $\mathcal{F}$, where $\mathcal{F}$ is a large collection of subsets of $\Omega = [0, 1]$ including all intervals.

(a) Fix $\alpha > 1$ and $r \geq 1$. Now, consider the RVs $X_n, n = 1, 2, \cdots$ defined by

$$X_n(\omega) := \begin{cases} \alpha^n, \text{ if } \omega \in [0, \frac{1}{n}], \\ 0, \text{ otherwise.} \end{cases}$$

Then, X_n is a discrete RV with $\mathbb{P}(X_n = \alpha^n) = \frac{1}{n}$ and $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$. In particular, $\lim_n \mathbb{E}|X_n|^r = \lim_n \alpha^{rn} = \infty$. So, X_n's do not convergence to 0 in the r-th mean. However, for all $\omega \in (0, 1]$, we have $\lim_n X_n(\omega) = 0$, which implies $X_n \xrightarrow[n\to\infty]{\text{a.s.}} 0$.

(b) For natural numbers n and $1 \leq m \leq n$, we construct RVs X_{nm} as follows

$$X_{nm}(\omega) = 1_{(\frac{m-1}{n}, \frac{m}{n}]}(\omega), \forall \omega \in \Omega = [0, 1].$$

We have $X_{nm} \sim Bernoulli(\frac{1}{n})$. This forms a triangular array of RVs as follows

$$\begin{array}{lll} X_{11} & & \\ X_{21} & X_{22} & \\ X_{31} & X_{32} & X_{33} \\ \cdots & & \end{array}$$

If we read the RVs, first left to right along a row and then move to the left-most entry of the next row, we end up with the sequence

$$X_{11} \quad X_{21} \quad X_{22} \quad X_{31} \quad X_{32} \quad X_{33} \quad \cdots$$

Note that for each $\omega \in (0, 1]$, the sequence $\{X_{nm}(\omega)\}_{nm}$ has infinitely many 0's and infintely many 1's. Therefore, the sequence $\{X_{nm}(\omega)\}_{nm}$ does not convergence for all $\omega \in (0, 1]$, and hence the sequence $\{X_{nm}\}_{nm}$ of RVs does not convergence a.s.. But, for any $r \geq 1$, $\mathbb{E}|X_{nm}|^r = \frac{1}{n}$ and consequently $\{X_{nm}\}_{nm}$ converges to 0 in r-th mean.

Remark 13.11 (Convergence in probability does not imply a.s. convergence). Note that in both of the examples mentioned in Remark 13.10, the sequence of RVs converge in probability to 0. In particular, the second example shows that convergence in probability does not imply a.s. convergence.

Note 13.12. There is a stronger version of the WLLN, called the Strong Law of Large Numbers, in terms of a.s. convergence. We shall not discuss this in the course.

Note 13.13. In the discussion involving convergence of RVs, we have seen the following notions of convergence, viz. convergence in r-th mean, convergence in probability and almost sure convergence. Now, given a sequence of RVs $\{X_n\}_n$, the law/distribution of each X_n is determined by the corresponding DFs F_{X_n}. It is, therefore, reasonable to consider the problem of the convergence of the DFs.

Remark 13.14 (Pointwise limit of DFs need not be a DF)*.* We show by examples that the pointwise limit of DFs need not be a DF.

(a) Let $X_n \sim Uniform(-n, n)\, \forall n = 1, 2, \cdots$. Here,

$$F_{X_n}(x) = \begin{cases} 0, \text{ if } x > -n, \\ \frac{x+n}{2n}, \text{ if } -n \leq x < n, \\ 1, \text{ if } x \geq n. \end{cases}$$

Then, the pointwise limit exists and is given by $\lim_n F_{X_n}(x) = \frac{1}{2}, \forall x$. However, the pointwise limit function, say, $F(x) = \frac{1}{2}, \forall x$ is not a DF.

(b) Consider the sequence $\{X_n\}_n$ with X_n degenerate at $\frac{1}{n}$. Then,

$$F_{X_n}(x) = \begin{cases} 0, \text{ if } x < \frac{1}{n}, \\ 1, \text{ if } x \geq \frac{1}{n}. \end{cases}$$

Then, the pointwise limit function exists and is given by

$$F(x) := \lim_n F_{X_n}(x) = \begin{cases} 0, \text{ if } x \leq 0, \\ 1, \text{ if } x > 0. \end{cases}$$

Since F is not right continuous at 0, it is not a DF. However, we may change the value of F at 0 and obtain the following DF $\widetilde{F}$ given by

$$\widetilde{F}(x) := \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0. \end{cases}$$

and $\widetilde{F}$ matches with $\lim_n F_{X_n}$ except at the point of discontinuity of $\widetilde{F}$. Note that $\widetilde{F}$ is the DF of the degenerate RV at 0.

Motivated by the above examples, we now consider the following notion of convergence of RVs.

Definition 13.15 (Convergence in Law/Distribution)**.** Let X be an RV defined on a Probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with DF F. For each n, let X_n be an RV defined on (possibly different) probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ with DF F_n. Let D_F denote the point of discontinuities of F. We say the sequence $\{X_n\}_n$ converges in law/distribution to X, denoted by $X_n \xrightarrow[n\to\infty]{d} X$, if

$$F_n(x) \xrightarrow{n\to\infty} F(x), \forall x \in D_F^c.$$

Example 13.16. Consider the sequence $\{X_n\}_n$ with X_n degenerate at $\frac{1}{n}$ and let X be an RV degenerate at 0. Then, as discussed in Remark 13.14, we have $X_n \xrightarrow[n\to\infty]{d} X$.

Note 13.17.
(a) Recall that the set D_F of discontinuities of a DF F, if it is non-empty, is either finite or countably infinite. If $X_n \xrightarrow[n\to\infty]{d} X$, then as per the definition, we must have $F_n(x) \xrightarrow{n\to\infty} F(x)$ everywhere, except possibly at a countable number of points.
(b) If X is a continuous RV, then $D_F = \emptyset$. If $X_n \xrightarrow[n\to\infty]{d} X$, then as per the definition $F_n(x) \xrightarrow{n\to\infty} F(x), \forall x \in \mathbb{R}$.
(c) Even if the RVs $X, X_1, X_2, \cdots$ are defined on different probability spaces, we can consider the notion of convergence in law/distribution. However, to consider the notion of convergence in r-th mean or in probability, we must have the RVs defined on the same probability space.

We state the following result without proof. The details of the proof are not part of the course.

Proposition 13.18. *Let the RVs $X, X_1, X_2, \cdots$ be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ (or converges in the r-th mean for some $r \geq 1$), then $X_n \xrightarrow[n\to\infty]{d} X$.*

Corollary 13.19. *Let the RVs $X, X_1, X_2, \cdots$ be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If the sequence $\{X_n\}_n$ converges in r-th mean (or a.s.) to X, then $X_n \xrightarrow[n\to\infty]{d} X$.*

A special case of the above result requires more attention.

Proposition 13.20. *Let $\{X_n\}_n$ be a sequence of RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $c \in \mathbb{R}$. Then $X_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ if and only if $X_n \xrightarrow[n\to\infty]{d} c$.*

Proof. Here, the constant c is being treated as a degenerate RV at c. The DF for this RV is given by

$$F(x) := \begin{cases} 0, & \text{if } x < c, \\ 1, & \text{if } x \geq c \end{cases}$$

with the only point of discontinuity at c. Now, $X_n \xrightarrow[n\to\infty]{d} c$ implies

$$\lim_n F_n(x) = F(x), \forall x \neq c,$$

where F_n denotes the DF of X_n. Observe that, for any $\epsilon > 0$,

$$\begin{aligned} \lim_n \mathbb{P}(|X_n - c| > \epsilon) &= \lim_n \mathbb{P}(X_n > c + \epsilon) + \lim_n \mathbb{P}(X_n < c - \epsilon) \\ &\leq \lim_n [1 - \mathbb{P}(X_n \leq c + \epsilon)] + \lim_n \mathbb{P}(X_n \leq c - \epsilon) \\ &= \lim_n [1 - F_n(c + \epsilon)] + \lim_n F_n(c - \epsilon) \\ &= [1 - F(c + \epsilon)] + F(c - \epsilon) \\ &= 0. \end{aligned}$$

This proves the sufficiency part. The necessity part follows from Proposition 13.18. □

Example 13.21. If a sequence of RVs converges in law/distribution, they need not converge in probability. Construction of such an example is left as an exercise in the problem set 13.

We now state some sufficient conditions which imply the convergence in law/distribution. The proofs are not included in the course.

Theorem 13.22. *Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space.*

(a) If these RVs are taking values in the set of non-negative integers (in particular, the RVs are discrete) and if the corresponding p.m.fs converge pointwise, i.e. $\lim_n f_{X_n}(x) = f(x), \forall x \in \{0, 1, 2, \cdots\}$*, then* $X_n \xrightarrow[n\to\infty]{d} X$.

(b) If all the RVs are continuous with the corresponding p.d.fs $f_X, f_{X_1}, f_{X_2}, \cdots$ *and if* $\lim_n f_{X_n}(x) = f(x), \forall x \in \mathbb{R}$*, then* $X_n \xrightarrow[n\to\infty]{d} X$.

(c) If these RVs have the MGFs $M, M_1, M_2, \cdots$ *existing on* $(-h, h)$ *for some* $h > 0$ *and if* $\lim_n M_n(t) = M(t), \forall t \in (-h, h)$*, then* $X_n \xrightarrow[n\to\infty]{d} X$.

(d) (Lévy's Continuity Theorem) Let $\Phi, \Phi_1, \Phi_2, \cdots$ *denote the Characteristic functions of* $X, X_1, X_2, \cdots$*, respectively. If* $\lim_n \Phi_n(t) = \Phi(t), \forall t \in \mathbb{R}$*, then* $X_n \xrightarrow[n\to\infty]{d} X$.

Example 13.23. Consider the discrete RVs X_n with the p.m.f.s and DFs given by

$$f_{X_n}(x) = \begin{cases} \frac{1}{2}, \text{ if } x \in \{\frac{1}{2n}, \frac{1}{n}\}, \\ 0, \text{ otherwise} \end{cases} \quad , \quad F_{X_n}(x) = \begin{cases} 0, \text{ if } x < \frac{1}{2n}, \\ \frac{1}{2}, \text{ if } \frac{1}{2n} \le x < \frac{1}{n}, \\ 1, \text{ if } x \ge \frac{1}{n}. \end{cases}$$

Since

$$\lim_n F_{X_n}(x) = \begin{cases} 0, \text{ if } x < 0, \\ 1, \text{ if } x > 0 \end{cases}$$

equals the DF F of the degenerate RV at 0, except at the point of discontinuity 0 of F, we have $X_n \xrightarrow[n\to\infty]{d} 0$. However, $\lim_n f_{X_n}(0) = 0 \neq 1 = f_X(0)$. Here, the pointwise convergence of the p.m.f.s do not hold.

Example 13.24. Let $X, X_1, X_2, \cdots$ be independent RVs with $X \sim N(0, 1)$ and $X_n \sim N(0, 1 + \frac{1}{n})$. Looking at the MGFs we have

$$\lim_n M_{X_n}(t) = \lim_n \exp\left(\frac{1}{2}\left(1 + \frac{1}{n}\right)t^2\right) = \exp\left(\frac{1}{2}t^2\right) = M_X(t), \forall t \in \mathbb{R}.$$

Therefore, $X_n \xrightarrow[n\to\infty]{d} X$. However, using the independence of $X, X_1, X_2, \cdots$, we have $X_n - X \sim N(0, 2+\frac{1}{n})$ and an argument similar to above shows that $X_n - X \xrightarrow[n\to\infty]{d} Z$, where $Z \sim N(0,2)$. Here, $X_n - X$ does not converge to the degenerate RV at 0.

The proof of the next result is not included in the course.

Theorem 13.25 (Continuous Mapping Theorem for Convergence in Distribution)**.** *Let $X_n \xrightarrow[n\to\infty]{d} X$ and $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ for some $c \in \mathbb{R}$.*

(a) Let $h : \mathbb{R} \to \mathbb{R}$ be a function continuous on the support S_X of X. Then $h(X_n) \xrightarrow[n\to\infty]{d} h(X)$.
(b) Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a function continuous on the set $\{(x,y) \in \mathbb{R}^2 : x \in S_X, y = c\}$. Then $h(X_n, Y_n) \xrightarrow[n\to\infty]{d} h(X, c)$.

Remark 13.26. The above Continuous Mapping Theorem implies the following result related to the Lévy's Continuity Theorem. Let $\Phi, \Phi_1, \Phi_2, \cdots$ denote the Characteristic functions of $X, X_1, X_2, \cdots$, respectively. If $X_n \xrightarrow[n\to\infty]{d} X$, then $\lim_n \Phi_n(t) = \Phi(t), \forall t \in \mathbb{R}$.

A special case of the above theorem is useful in practice.

Theorem 13.27 (Slutsky's Theorem)**.** *Let $X_n \xrightarrow[n\to\infty]{d} X$ and $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ for some $c \in \mathbb{R}$. Then $X_n + Y_n \xrightarrow[n\to\infty]{d} X + c$ and $X_n Y_n \xrightarrow[n\to\infty]{d} cX$.*

Notation 13.28. For any RV X, we treat $0 \times X$ as an RV degenerate at 0.

Note 13.29. We now look at an example of convergence in distribution which is quite useful in practice.

Theorem 13.30 (Poisson approximation to Binomial Distribution)**.** *Let $X_n \sim Binomial(n, p_n), n = 1, 2, \cdots$ where $p_n \in (0,1), \forall n$ and*

$$\lim_n np_n = \lambda > 0.$$

Then $X_n \xrightarrow[n\to\infty]{d} X$ with $X \sim Poisson(\lambda)$ with $\mathbb{P}(X_n = k) \xrightarrow{n\to\infty} \mathbb{P}(X = k)$ for all $k = 0, 1, 2, \cdots$.

Proof. We prove the stated convergence using MGFs (see Theorem 13.22). We have for all $t \in \mathbb{R}$

$$\lim_n M_{X_n}(t) = \lim_n (1 - p_n + p_n e^t)^n = \lim_n \left(1 + \frac{np_n(e^t - 1)}{n}\right)^n = \exp(\lambda(e^t - 1)) = M_X(t).$$

Hence, we have the convergence in law/distribution.

We may check the same using the p.m.fs. For fixed $k = 0, 1, 2, \cdots$ and for all $n \geq k$, we have

$$\begin{aligned}\lim_n \mathbb{P}(X_n = k) &= \lim_n \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{1}{k!} \lim_n \frac{n(n-1)\cdots(n-k+1)}{n^k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k} \\ &= \frac{1}{k!} \lambda^k e^{-\lambda} = \mathbb{P}(X = k).\end{aligned}$$

This completes the proof. $\square$

Remark 13.31. In Theorem 13.30, from the assumption $\lim_n np_n = \lambda > 0$, we have the probability of success p_n is 'small' for large n. We may therefore treat X_n as the number of successes of a 'rare' event in n trials of a random experiment with probability of success p_n (see Remark 10.22). Here, we have kept $\mathbb{E}X_n = np_n$ close to $\lambda > 0$. So the number n of trials are increases, but the probability of success is decreases with n.

Example 13.32. If $X \sim \mathit{Binomial}(1000, 0.003)$, then the exact value of

$$\mathbb{P}(X = 5) = \binom{1000}{5} (0.003)^5 (0.997)^{995}$$

is hard to compute. Instead, we can approximate the value by $\mathbb{P}(Y = 5)$ where $Y \sim \mathit{Poisson}(1000 \times 0.003) = \mathit{Poisson}(3)$. Here, $\mathbb{P}(Y = 5) = e^{-3} \frac{3^5}{5!}$ is comparatively easier to compute.

Remark 13.33 (A question about the rate of convergence). We have seen three types of convergences of RVs and their examples. However, in these examples, we can ask how 'fast' does the convergences occur? For example, by the WLLN (Theorem 12.24), we have $\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1$ for any i.i.d. sequence of RVs $X_1, X_2, \cdots$ with finite expectation. How 'fast' does the 'error term' $\bar{X}_n - \mathbb{E}X_1$ go to 0? In other words, how 'small' is $\bar{X}_n - \mathbb{E}X_1$ for 'large' n? If we can show, '$n^\alpha \left(\bar{X}_n - \mathbb{E}X_1\right) \xrightarrow{n\to\infty} c$' for some $c \in \mathbb{R}, \alpha > 0$, then for large n, we may say $\bar{X}_n - \mathbb{E}X_1$ is close to $\frac{c}{n^\alpha}$ - which gives an idea about the magnitude. This, however, is only an idea and not a concrete result. In fact, in this description, it is more likely to have an RV in the place of c above, with a clear notion of convergence for the 'error term', again in terms of some notion of convergence of RVs. It is to be noted that a convergence result with a 'rate of convergence' is stronger than another convergence result without any clear 'rate of convergence'. We shall come back to this discussion later.

Note 13.34. For i.i.d. RVs, recall from the proof of WLLN (Theorem 12.24), that $Var(\bar{X}_n) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}Var(X_1)$. Provided $Var(X_1) > 0$, we have $\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sqrt{Var(X_1)}}$ is an RV with mean 0 and variance 1.

Theorem 13.35 (Lindeberg-Lévy Central Limit Theorem (CLT))**.** *Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists and $Var(X_1) = \sigma^2 > 0$. Then,*

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$.

Remark 13.36 (Restatements of the CLT). Under the assumptions of the CLT above, we can restate the conclusion in various useful ways. Note that the DF Φ of $Z \sim N(0,1)$ is continuous everywhere on $\mathbb{R}$.

(a) $\lim_{n\to\infty} \mathbb{P}(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq x) = \Phi(x), \forall x \in \mathbb{R}$.
(b) For all $a < b$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(a < \sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq b) \\
&= \lim_{n\to\infty} \mathbb{P}(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq b) - \lim_{n\to\infty} \mathbb{P}(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq a) \\
&= \Phi(b) - \Phi(a) \\
&= \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt
\end{aligned}$$

(c) Writing $Y_n = X_1 + X_2 + \cdots + X_n$, for all $a < b$ we have

$$\lim_{n\to\infty} \mathbb{P}(a < \frac{Y_n - n\mathbb{E}X_1}{\sigma\sqrt{n}} \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

Proof of CLT (Theorem 13.35). We find the limit of the MGFs of

$$Z_n := \sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - \mathbb{E}X_1}{\sigma}.$$

Here, $\frac{X_i - \mathbb{E}X_1}{\sigma}, i = 1, 2, \cdots$ are i.i.d. with mean 0 and variance 1. In particular, $\mathbb{E}\left(\frac{X_i-\mathbb{E}X_1}{\sigma}\right)^2 = 1$. Since we have assumed the existence of the MGFs, we have $M'_{\frac{X_i-\mathbb{E}X_1}{\sigma}}(0) = 0$ and $M''_{\frac{X_i-\mathbb{E}X_1}{\sigma}}(0) = 1$. We also have a Taylor series expansion in a neighbourhood of 0 as

$$M_{\frac{X_i-\mathbb{E}X_1}{\sigma}}(t) = M_{\frac{X_i-\mathbb{E}X_1}{\sigma}}(0) + tM'_{\frac{X_i-\mathbb{E}X_1}{\sigma}}(0) + \frac{t^2}{2}\left(M''_{\frac{X_i-\mathbb{E}X_1}{\sigma}}(0) + R(t)\right) = 1 + \frac{t^2}{2}(1 + R(t))$$

with the remainder term satisfying $\lim_{t\to 0} R(t) = 0$.
Then, using the i.i.d. nature of $\frac{X_i-\mathbb{E}X_1}{\sigma}, i = 1, 2, \cdots$, we have

$$\begin{aligned}
M_{Z_n}(t) &= \mathbb{E}\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - \mathbb{E}X_1}{\sigma}\right) \\
&= \left(M_{\frac{X_1-\mathbb{E}X_1}{\sigma}}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&= \left(1 + \frac{t^2}{2n}\left(1 + R\left(\frac{t}{\sqrt{n}}\right)\right)\right)^n \\
&\xrightarrow{n\to\infty} \exp\left(\frac{t^2}{2}\right) = M_Z(t)
\end{aligned}$$

for all t in a neighbourhood of 0. Using Theorem 13.22, we conclude the proof. $\square$

Remark 13.37. The CLT suggests that for large sample sizes, the normarlized version $\frac{\bar{X}_n - \mathbb{E}X_1}{\sqrt{Var(X_1)}}$ of the sample mean $\bar{X}_n$ based on a random sample from any given distribution is close to the Standard Normal distribution, in the sense of convergence in law/distribution. In practice, this result can be used to obtain estimates for probabilities involving the sample mean.

Note 13.38. In the statement of the CLT, we have taken $\sigma > 0$. If $\sigma = 0$, then note that all the RVs $X_1, X_2, \cdots$ are actually degenerate at some constant $c \in \mathbb{R}$ and $\bar{X}_n = c, \forall n$.

Remark 13.39 (From CLT to WLLN). Our motivation to study CLT type results was to find a 'rate of convergence' for the WLLN. As mentioned in Remark 13.33, a convergence result with a 'rate of convergence' is stronger than another convergence result without any clear 'rate of convergence'. We illustrate this idea by deriving the WLLN from the CLT. Under the assumptions of CLT (Theorem 13.35), we have

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$. Since $\frac{\sigma}{\sqrt{n}} \xrightarrow{n\to\infty} 0$, by Slutsky's theorem (Theorem 13.27),

$$\bar{X}_n - \mathbb{E}X_1 \xrightarrow[n\to\infty]{d} 0 \times Z = Y,$$

where Y denotes an RV degenerate at 0. By Proposition 13.20, we have $\bar{X}_n - \mathbb{E}X_1 \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Finally, by Theorem 13.6, we conclude $\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1$, which is the WLLN. Note that, however, to show this we needed the additional assumption on the second moments, which is not required if we are only interested in the WLLN.

One may ask what happens if we deal with a sequence of independent RVs, which are not necessarily identically distributed.

Theorem 13.40 (Lyapunov CLT)**.** *Let $\{X_n\}_n$ be a sequence of independent RVs with $\mathbb{E}X_n^2 < \infty, \forall n$. Write $c_n^2 := Var(X_1 + X_2 + \cdots + X_n), \forall n$. If*

$$\frac{1}{c_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}|X_k - \mathbb{E}X_k|^{2+\delta} \xrightarrow{n\to\infty} 0$$

for some $\delta > 0$, then

$$\frac{X_1 + X_2 + \cdots + X_n - \mathbb{E}(X_1 + X_2 + \cdots + X_n)}{c_n} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$.

Remark 13.41. Another well-known variant of the Lindeberg-Lévy CLT is the Lindeberg-Feller CLT, which is more general than the Lyapunov CLT. We do not discuss this result in this course.

Note 13.42. As discussed above, using information from higher moments, we have improved results. The CLT stated here can be improved to the Berry-Esseen Theorem using information from 3-rd absolute moments and the WLLN can be improved to Hoeffding's inequality for bounded

RVs. The CLT has a huge literature and many CLT-type results have been proved even in the non-i.i.d. setting. These results are not part of this course.

Remark 13.43. Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists and $Var(X_1) = \sigma^2 > 0$. By the CLT,

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$. From Remark 13.7 and Theorem 13.6, we have

$$\frac{\sigma}{S_n} \xrightarrow[n\to\infty]{\mathbb{P}} 1,$$

where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is the sample variance. By Slutsky's theorem,

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{S_n} \xrightarrow[n\to\infty]{d} Z.$$

Note 13.44. Recall from Theorem 13.4 that if $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, then for any continuous function $h : \mathbb{R} \to \mathbb{R}$, we have $h(X_n) \xrightarrow[n\to\infty]{\mathbb{P}} h(X)$. We can now ask about the rate of convergence. This question leads to a useful result, known as the Delta method. We do not discuss the proof of this result in this course.

Theorem 13.45 (Delta method)**.** *Let $\{X_n\}_n$ be a sequence of RVs such that $n^b(X_n - a) \xrightarrow[n\to\infty]{d} X$ for $a \in \mathbb{R}, b > 0$ and some RV X. Let $g : \mathbb{R} \to \mathbb{R}$ be a function differentiable at a. Then*

$$n^b(g(X_n) - g(a)) \xrightarrow[n\to\infty]{d} g'(a)X.$$

Combining with the CLT, using the Delta method we get the following result often used in practice.

Theorem 13.46. *Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists and $Var(X_1) = \sigma^2 > 0$. Let $g : \mathbb{R} \to \mathbb{R}$ be a function differentiable at $a = \mathbb{E}X_1$ with $g'(a) \neq 0$. Then,*

$$\sqrt{n}\frac{g(\bar{X}_n) - g(\mathbb{E}X_1)}{\sigma} \xrightarrow[n\to\infty]{d} g'(a)Z \sim N(0, (g'(a))^2),$$

where $Z \sim N(0,1)$.

Remark 13.47. Let $X_1, X_2, \cdots$ be i.i.d. $Uniform(0,\theta)$ RVs, for some $\theta > 0$. Recall from Example 12.21 that $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\} \xrightarrow[n\to\infty]{\mathbb{P}} \theta$. We can now ask about the limiting distribution of $(\theta - X_{(n)})$ to understand the rate of convergence. Recall that the p.d.f. of $X_{(n)}$ is given by

$$g_{X_{(n)}}(x) = \begin{cases} \frac{n}{\theta^n} x^{n-1}, \text{ if } x \in (0,\theta), \\ 0, \text{ otherwise.} \end{cases}$$

Look at $Y_n := n(\theta - X_{(n)})$. Then for all $y \in \mathbb{R}$,

$$\begin{aligned} F_{Y_n}(y) &= \mathbb{P}(Y_n \leq y) \\ &= \mathbb{P}\left(X_{(n)} \geq \theta - \frac{y}{n}\right) \\ &= \int_{\theta-\frac{y}{n}}^{\infty} g_{X_{(n)}}(x)\, dx \\ &= \begin{cases} 0, \text{ if } y \leq 0 \\ 1 - \left(1 - \frac{y}{n\theta}\right)^n, \text{ if } 0 < y < n\theta, \\ 1, \text{ if } y > n\theta \end{cases} \\ &\xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } y \leq 0 \\ 1 - \exp\left(-\frac{y}{\theta}\right), \text{ if } y > 0 \end{cases} \\ &= F_Y(y) \end{aligned}$$

where $Y \sim Exponential(\theta)$. Since the DF F_Y of Y is continuous everywhere, from the above computation we conclude that $Y_n = n(\theta - X_{(n)}) \xrightarrow[n\to\infty]{d} Y \sim Exponential(\theta)$. The sequence $\{X_{(n)}\}_n$ another example where Delta method can be applied.