(b) (Product of DFs) The function $H : \mathbb{R} \to [0, 1]$ defined by $H(x) := F(x)G(x), \forall x \in \mathbb{R}$ has the relevant properties and hence is a DF. In particular, F^2 is a DF, if F is so.

In fact, a general DF can be written as a convex combination of discrete DFs and some special continuous DFs. We do not discuss such results in this course.

Remark 1.402. In practice, given a known RV X, many times we need to find out the distribution of $h(X)$ for some function $h : \mathbb{R} \to \mathbb{R}$ or even, simply, compute the expectations of the form $\mathbb{E}h(X)$. As already discussed earlier in the course, we can theoretically (i.e., in principle) compute $\mathbb{E}h(X)$ as $\int_{-\infty}^{\infty} h(x)f_X(x)\,dx$, when X is a continuous RV with p.d.f. f_X, for example. However, in practice, it may happen that this integral does not have a closed form expression – which makes it challenging to evaluate. The problem becomes more intractable when we look at similar problems where X is a random vector and the joint/marginal distributions need to be considered. In such situations, as an alternative, we try to find 'good' approximations for the quantities of interest, where the approximation terms are easier to compute than the original expression. This motivation leads to the various notions for convergence of RVs. If some quantity of interest involving an RV X, say $\mathbb{E}X$, is difficult to compute, then we find an appropriate 'approximating' sequence of RVs $\{X_n\}_n$ for X and use the values $\mathbb{E}X_n$ as an approximation for $\mathbb{E}X$.

Remark 1.403. Given a random sample $X_1, X_2, \cdots, X_n$ from $N(\mu, \sigma^2)$ distribution, consider the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Here, we have written $\bar{X}_n$, instead of just $\bar{X}$, to highlight the dependence of the sample mean on the sample size n. Recall that $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. The behaviour of $\bar{X}_n$ for large n is of interest. This is also another motivation for us to study the convergence of sequences of RVs.

We now discuss concepts for convergence of sequences of RVs.

Definition 1.404 (Convergence in r-th mean)**.** Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $r \geq 1$. If $\mathbb{E}|X|^r < \infty, \mathbb{E}|X_n|^r < \infty, \forall n$ and if

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

then we say that the sequence $\{X_n\}_n$ converges to X in r-th mean.

Note 1.405. (a) If a sequence $\{X_n\}_n$ converges to X in r-th mean for some $r \geq 1$, then we have

$$\lim_{n\to\infty} \mathbb{E}|X_n|^r = \mathbb{E}|X|^r,$$

and

$$\lim_{n\to\infty} \mathbb{E}X_n^r = \mathbb{E}X^r,$$

i.e., we have the convergence of the r-th moments.

(b) The sequence $\{X_n\}_n$ converges to X in r-th mean if and only if the sequence $\{X_n - X\}_n$ converges to 0 in r-th mean.

Remark 1.406. Even though we have defined the r-th order moments for $0 < r < 1$, for technical reasons we do not consider the convergence in r-th mean in this case. The details are beyond the scope of this course. In what follows, whenever we consider the convergence in r-th mean, we assume $r \geq 1$.

Definition 1.407 (Convergence in Probability)**.** Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If for all $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0,$$

then we say that the sequence $\{X_n\}_n$ converges to X in probability and write $X_n \xrightarrow[n\to\infty]{P} X$.

Note 1.408. (a) Suppose that a sequence $\{X_n\}_n$ converges to X in probability. Now, for all $\epsilon > 0$, note that

$$\mathbb{P}(|X_n - X| \geq 2\epsilon) \leq \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(|X_n - X| \geq \epsilon).$$

Convergence in probability is equivalent to the fact that

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

(b) The sequence $\{X_n\}_n$ converges to X in probability if and only if the sequence $\{X_n - X\}_n$ converges to 0 in probability.

Proposition 1.409. *Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If the sequence $\{X_n\}_n$ converges to X in r-th mean for some $r \geq 1$, then $X_n \xrightarrow[n\to\infty]{P} X$.*

Proof. By Markov's inequality (Corollary 1.255), we have

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \epsilon^{-r} \, \mathbb{E}|X_n - X|^r.$$

Since $\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0$, we have the result. □

Corollary 1.410. *Let $\{X_n\}_n$ be a sequence of RVs with finite second moments. If $\lim_n \mathbb{E}X_n = \mu$ and $\lim_n Var(X_n) = 0$, then $\{X_n\}_n$ converges to μ in 2nd mean and in particular, in probability.*

Proof. We have $\mathbb{E}|X_n - \mu|^2 = \mathbb{E}\left[(X_n - \mu_n) + (\mu_n - \mu)\right]^2 = \mathbb{E}(X_n - \mu_n)^2 + (\mu_n - \mu)^2 = Var(X_n) + (\mu_n - \mu)^2$. By our hypothesis, $\lim_n \mathbb{E}|X_n - \mu|^2 = 0$. Hence, $\{X_n\}_n$ converges to μ in 2nd mean. By Proposition 1.409, the sequence also converges in probability. □

Example 1.411. Let $X_1, X_2, \cdots$ be i.i.d. $Uniform(0, \theta)$ RVs, for some $\theta > 0$. The sequence $\{X_n\}_n$ being i.i.d. means that the collection $\{X_n : n \geq 1\}$ is mutually independent and that all the RVs have the same law/distribution. Here, the common p.d.f. and the common DF are given by

$$f(x) = \begin{cases} \frac{1}{\theta}, \text{ if } x \in (0, \theta), \\ 0, \text{ otherwise} \end{cases}, \quad F(x) = \begin{cases} 0, \text{ if } x < 0, \\ \frac{x}{\theta}, \text{ if } 0 \leq x < \theta, \\ 1, \text{ if } x \geq \theta. \end{cases}$$

Consider $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$. Using Proposition 1.344, we have the marginal p.d.f. of $X_{(n)}$ is given by

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

Then,

$$\mathbb{E}X_{(n)} = \int_0^\theta x \, \frac{n}{\theta^n} x^{n-1} \, dx = \frac{n}{n+1}\theta, \quad \mathbb{E}X_{(n)}^2 = \int_0^\theta x^2 \, \frac{n}{\theta^n} x^{n-1} \, dx = \frac{n}{n+2}\theta^2$$

and

$$Var(X_{(n)}) = \theta^2 \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right] = \theta^2 \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2} = \theta^2 \frac{n}{(n+2)(n+1)^2}.$$

Now, $\lim_n \mathbb{E}X_{(n)} = \theta$ and $\lim_n Var(X_{(n)}) = 0$. Hence, by Corollary 1.410, $\{X_{(n)}\}_n$ converges in 2nd mean to θ and also in probability.

Remark 1.412 (Convergence in probability does not imply convergence in r-th mean)*.* Consider a sequence of discrete RVs $\{X_n\}_n$ with $X_n \sim Bernoulli(\frac{1}{n}), \forall n$. Consider $Y_n := nX_n, \forall n$. Then Y_n's are also discrete with the p.m.f.s given by

$$f_{Y_n}(y) = \begin{cases} 1 - \frac{1}{n}, \text{ if } y = 0, \\ \frac{1}{n}, \text{ if } y = n, \\ 0, \text{ otherwise.} \end{cases}$$

For all $\epsilon > 0$, we have $\mathbb{P}(|Y_n| \geq \epsilon) = \frac{1}{n} \xrightarrow{n\to\infty} 0$ and hence $Y_n \xrightarrow[n\to\infty]{P} 0$. But, for any $r > 1$, $\mathbb{E}|Y_n|^r = n^{r-1}, \forall n$. Here, $\{Y_n\}_n$ does not converge to 0 in r-th mean.

Example 1.413. $X_1, X_2, \cdots$ be i.i.d. RVs following $N(\mu, \sigma^2)$ distribution. Recall that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Then $\lim_n \mathbb{E}\bar{X}_n = \lim_n \mu = \mu$ and $\lim_n Var(\bar{X}_n) = \lim_n \frac{\sigma^2}{n} = 0$. By Corollary 1.410, $\{\bar{X}_n\}_n$ converges in 2nd mean to μ and also in probability.

The above example leads to the following result.

Theorem 1.414 (Weak Law of Large Numbers (WLLN))**.** *Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1$ exists. Then, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{P} \mathbb{E}X_1$.*

Remark 1.415*.* We only discuss the proof of Theorem 1.414, when $\mathbb{E}X_1^2$ exists. The proof of the theorem when $\mathbb{E}X_1^2$ does not exist is beyond the scope of this course. However, we shall use this theorem in its full generality.

Proof of WLLN (Theorem 1.414) (assuming $\mathbb{E}X_1^2 < \infty$). Observe that $\mathbb{E}\bar{X}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n}n\mathbb{E}X_1 = \mathbb{E}X_1$ and, using independence of X_i's we have

$$Var(\bar{X}_n) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}Var(X_1) \xrightarrow{n\to\infty} 0.$$

By Corollary 1.410, the result follows. □

Remark 1.416. 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 1.417. 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, $\bar{X}_n \xrightarrow[n\to\infty]{P} \mathbb{E}X_1 = p$, 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 1.418. 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 1.228, $X_i \sim Bernoulli(p)$, where $p = \mathbb{P}(E)$. If p is unknown, then by the WLLN we have $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{P} p$, 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.

Note 1.419. There is a stronger version of the WLLN, called the Strong Law of Large Numbers, which we do not discuss in the course.

Theorem 1.420 (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 1.421. 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 1.422 (Algebraic Operations with Convergence in Probability)**.** *Let $\{X_n\}_n$ and $\{Y_n\}$ be sequences of RVs such that $X_n \xrightarrow[n\to\infty]{P} x$ and $Y_n \xrightarrow[n\to\infty]{P} y$ 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) $X_n + Y_n \xrightarrow[n\to\infty]{P} x + y$.
(b) $X_n - Y_n \xrightarrow[n\to\infty]{P} x - y$.
(c) $X_n Y_n \xrightarrow[n\to\infty]{P} xy$.
(d) $\frac{X_n}{Y_n} \xrightarrow[n\to\infty]{P} \frac{x}{y}$, *provided* $y \neq 0$.
(e) $a_n X_n + b_n \xrightarrow[n\to\infty]{P} ax + b$.

Remark 1.423. 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 1.414)

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

Again by WLLN $\bar{X}_n \xrightarrow[n\to\infty]{P} \mathbb{E}X_1$ and by Theorem 1.420 applied to the function $h(x) = x^2, \forall x \in \mathbb{R}$, we have

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

Since $\frac{n}{n-1} \xrightarrow{n\to\infty} 1$, using Theorem 1.422, we have $S_n^2 \xrightarrow[n\to\infty]{P} Var(X_1)$. By Theorem 1.420, we have $S_n \xrightarrow[n\to\infty]{P} \sqrt{Var(X_1)}$.

Note 1.424. In the discussion involving convergence of RVs, we have seen two notions of convergence, viz. convergence in r-th mean and convergence in probability. 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 1.425 (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 1.426 (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 1.427. 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 1.425, we have $X_n \xrightarrow[n\to\infty]{d} X$.

Note 1.428. (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 1.429. *Let the RVs $X, X_1, X_2, \cdots$ be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n \xrightarrow[n\to\infty]{P} X$ (or converges in the r-th mean for some $r \geq 1$), then $X_n \xrightarrow[n\to\infty]{d} X$.*

A special case of the above result requires more attention.

Proposition 1.430. *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 $X_n \xrightarrow[n\to\infty]{P} c$ 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 $\epsilon > 0$,

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

This proves the sufficiency part. The necessity part follows from Proposition 1.429. $\square$

Example 1.431. 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 10.

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

Theorem 1.432. *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* $X_n \xrightarrow[n\to\infty]{d} X$.

Example 1.433. Consider the discrete RVs X_n with the p.m.f.s and DFs given by

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

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 1.434. 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 1.435 (Continuous Mapping Theorem for Convergence in Distribution)**.** *Let $X_n \xrightarrow[n\to\infty]{d} X$ and $Y_n \xrightarrow[n\to\infty]{P} c$ 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 $h(X_n, Y_n) \xrightarrow[n\to\infty]{d} h(X, c)$.

Notation 1.436. For any RV X, we treat $0 \times X$ as an RV degenerate at 0.

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

Theorem 1.437 (Slutky's Theorem)**.** *Let* $X_n \xrightarrow[n\to\infty]{d} X$ *and* $Y_n \xrightarrow[n\to\infty]{P} c$ *for some* $c \in \mathbb{R}$*. Then* $X_n + Y_n \xrightarrow[n\to\infty]{d} X + c$ *and* $X_nY_n \xrightarrow[n\to\infty]{d} cX$*.*

Note 1.438. We now look at an example of convergence in distribution which is quite useful in practice.

Theorem 1.439 (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 1.432). We have for all $t \in \mathbb{R}$

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

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. □

Remark 1.440. In Theorem 1.439, 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 1.351). 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 1.441. If $X \sim Binomial(1000, 0.003)$, 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 $Y \sim Poisson(1000 \times 0.003) = Poisson(3)$. Here, $\mathbb{P}(Y = 5) = e^{-3}\frac{3^5}{5!}$ is comparatively easier to compute.

Remark 1.442 (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 1.414), we have $\bar{X}_n \xrightarrow[n\to\infty]{P} \mathbb{E}X_1$ 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 1.443. For i.i.d. RVs, recall from the proof of WLLN (Theorem 1.414), 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 1.444 (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 1.445 (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 1.444). 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 $\lim_{t\to 0} R(t) = 0$.

Then, using the i.i.d. nature of $\frac{X_i - \mathbb{E}X_1}{\sigma}, i = 1, 2, \cdots$, we have

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

$$\xrightarrow{n\to\infty} \exp\left(\frac{t^2}{2}\right) = M_Z(t)$$

for all t in a neighbourhood of 0. Using Theorem 1.432, we conclude the proof. $\square$

Remark 1.446. 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 1.447. 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 1.448 (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 1.442, 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 1.444), 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 Slutky's theorem (Theorem 1.437),

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

where Y denotes an RV degenerate at 0. By Proposition 1.430, we have $\bar{X}_n - \mathbb{E}X_1 \xrightarrow[n\to\infty]{P} 0$. Finally, by Theorem 1.422, we conclude $\bar{X}_n \xrightarrow[n\to\infty]{P} \mathbb{E}X_1$, 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.

Note 1.449. 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 1.450. 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 1.423 and Theorem 1.422, we have

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

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

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

Note 1.451. Recall from Theorem 1.420 that if $X_n \xrightarrow[n\to\infty]{P} X$, then for any continuous function $h : \mathbb{R} \to \mathbb{R}$, we have $h(X_n) \xrightarrow[n\to\infty]{P} h(X)$. 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 1.452 (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 1.453. *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 1.454. Let $X_1, X_2, \cdots$ be i.i.d. $Uniform(0,\theta)$ RVs, for some $\theta > 0$. Recall from Example 1.411 that $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\} \xrightarrow[n\to\infty]{P} \theta$. 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 $Y \sim \mathit{Exponential}(\theta)$. Since the DF F_Y of Y is continuous everywhere, from the above computation we conclude that $Y_n = n(\theta - X_{(n)}) \xrightarrow[n\to\infty]{d} Y \sim \mathit{Exponential}(\theta)$. The sequence $\{X_{(n)}\}_n$ another example where Delta method can be applied.

2. To add

14.31 Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution with the parameters μ and σ^2 being unknown. Recall from Remark 1.375 that $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$, where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Then, $\mathbb{E}_{\mu,\sigma^2}\frac{(n-1)S_n^2}{\sigma^2} = (n-1)$ or $\mathbb{E}_{\mu,\sigma^2} S_n^2 = \sigma^2$, for all possible values of μ and σ^2.

14.39 Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution, with the parameters μ and σ^2 being unknown. Consider the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Recall from Remark 1.369

that $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. By an application of the Chebyshev's inequality (Corollary 1.257), for all $\epsilon > 0$, we have

$$\mathbb{P}_{\mu,\sigma^2}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{1}{\epsilon^2} Var(\bar{X}_n) = \frac{1}{n\epsilon^2}\sigma^2$$

and hence $\lim_{n\to\infty} \mathbb{P}_{\mu,\sigma^2}(|\bar{X}_n - \mu| \geq \epsilon) = 0$ (also see Proposition 1.409). Therefore, $\{\bar{X}_n\}_n$ is consistent for μ.

15.10 We shall use the following pivotal quantities in our discussion. Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

(a) Suppose that μ is unknown but σ^2 is known. If the estimand is $h(\mu) = \mu$, then we consider the pivotal quantity $g(X_1, X_2, \cdots, X_n; \mu) = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$ (see Remark 1.369).

(b) Suppose that μ and σ^2 both are unknown. If the estimand is $h(\mu, \sigma^2) = \mu$, then we consider the pivotal quantity $g(X_1, X_2, \cdots, X_n; \mu) = \sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1}$ (see Note 1.378).

(c) Suppose that σ^2 is unknown but μ is known. If the estimand is $h(\sigma^2) = \sigma^2$, then we consider the pivotal quantity $g(X_1, X_2, \cdots, X_n; \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2_n$ (see Note 1.372).

(d) Suppose that μ and σ^2 both are unknown. If the estimand is $h(\sigma^2) = \sigma^2$, then we consider the pivotal quantity $g(X_1, X_2, \cdots, X_n; \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi^2_{n-1}$ (see Remark 1.375).

16.1 (pivotal quantity – pooled sample mean)

17.11 If $Z \sim N(0,1)$, it can be checked that $\mathbb{P}(|Z| \leq 3) \approx 0.997$ and $\mathbb{P}(|Z| \leq 6) \approx 0.9997$. More generally, for $X \sim N(\mu, \sigma^2)$, we have $\mathbb{P}(|X - \mu| \leq 3\sigma) \approx 0.997$ and $\mathbb{P}(|X - \mu| \leq 6\sigma) \approx 0.9997$. This shows that the values of a normal RV is quite concentrated near its mean.

3. To add from practice problems

Set 9 question 4 (sample mean and sample variance)

Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution. Consider the sample mean $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ and sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Show that $\bar{X}$ and S_n^2 are independent.

Look at the joint MGF of $(X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X}, \bar{X})$ given by

$$M(t_1, t_2, \cdots, t_n, t_{n+1}) = \mathbb{E}\exp\left(\sum_{j=1}^n t_j(X_j - \bar{X}) + t_{n+1}\bar{X}\right), \forall (t_1, t_2, \cdots, t_n, t_{n+1}) \in \mathbb{R}^{n+1}$$

$$= \mathbb{E}\exp\left(\sum_{j=1}^{n} s_j X_j\right),$$

where $s_j = t_j + \frac{t_{n+1} - \sum_{i=1}^{n} t_i}{n}$. Using the independence of X_j's, we have

$$\begin{aligned}
M(t_1, t_2, \cdots, t_n, t_{n+1}) &= \prod_{j=1}^{n} \mathbb{E}\exp(s_j X_j) \\
&= \prod_{j=1}^{n} \exp\left(\mu s_j + \frac{1}{2}\sigma^2 s_j^2\right) \\
&= \exp\left(\mu \sum_{j=1}^{n} s_j + \frac{1}{2}\sigma^2 \sum_{j=1}^{n} s_j^2\right) \\
&= \exp\left(\mu t_{n+1} + \frac{1}{2}\sigma^2 \frac{t_{n+1}^2}{n}\right) \exp\left(\frac{1}{2}\sigma^2 \sum_{j=1}^{n}\left(t_j - \frac{\sum_{i=1}^{n} t_i}{n}\right)^2\right) \\
&= M_{\bar{X}}(t_{n+1}) M_{X_1-\bar{X}, X_2-\bar{X}, \cdots, X_n-\bar{X}}(t_1, t_2, \cdots, t_n).
\end{aligned}$$

Here, we use the observation that

$$\begin{aligned}
M_{X_1-\bar{X}, X_2-\bar{X}, \cdots, X_n-\bar{X}}(t_1, t_2, \cdots, t_n) &= M_{X_1-\bar{X}, X_2-\bar{X}, \cdots, X_n-\bar{X}, \bar{X}}(t_1, t_2, \cdots, t_n, 0) \\
&= \exp\left(\frac{1}{2}\sigma^2 \sum_{j=1}^{n}\left(t_j - \frac{\sum_{i=1}^{n} t_i}{n}\right)^2\right)
\end{aligned}$$

and

$$M_{\bar{X}}(t_{n+1}) = M_{X_1-\bar{X}, X_2-\bar{X}, \cdots, X_n-\bar{X}, \bar{X}}(0, 0, \cdots, 0, t_{n+1}) = \exp\left(\mu t_{n+1} + \frac{1}{2}\sigma^2 \frac{t_{n+1}^2}{n}\right).$$

Therefore, $(X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X})$ and $\bar{X}$ are independent. Consequently, the sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $\bar{X}$ are independent.

4. To add: additional content

Computation of DF for constant RV and for RV with two values (immediately after the definition)