Modes of Convergence

Limit Theorems and Convergence of r.v.s

- When we work with complex random systems, or stochastic processes, we are often interested in the limiting behavior of such processes, i.e., we want to say

$$\boxed{\lim_{n\to\infty} X_n = X}$$, where X_n, X are r.v.

It turns out however that there are multiple ways to define such a notion, with different properties and applications of each. The strongest (but least useful) is:

1) Point-wise Convergence - A sequence of r.v. $(X_n; n \geq 1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ converges pointwise to r.v. X on $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \forall\, \omega \in \Omega$$

- Staying with X_n and X on the same space, we have 3 more modes

2) Almost-Sure Convergence - A sequence of r.v. $(X_n : n \geq 1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ converges almost-surely to r.v. X in $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\mathbb{P}\left[\lim_{n\to\infty} X_n = X\right] = 1 \qquad \left(\text{notation: } X_n \xrightarrow{as} X\right)$$

3) Convergence in Probability - A sequence of r.v. $(X_n : n \geq 1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ converges in probability to r.v. X in $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\forall\, \varepsilon > 0, \ \lim_{n\to\infty} \mathbb{P}\left[|X - X_n| > \varepsilon\right] = 0 \qquad \left(\text{notation: } X_n \xrightarrow{P} X\right)$$

4) Converges in ℓ_p - A sequence of r.v. $(X_n : n \geq 1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ converges to r.v. X in $(\Omega, \mathcal{F}, \mathbb{P})$ in ℓ_p for $p \geq 1$ if $\left(\text{notation: } X_n \xrightarrow{\ell_p} X\right)$

$$\lim_{n\to\infty} \|X_n - X\|_p = 0, \text{ where } \|X_n - X\|_p = \left(\mathbb{E}\left[(X_n - X)^p\right]\right)^{1/p}$$

- The operator (i.e., function acting on functions) $\|\cdot\|_p$ is called the l_p-norm, and is a way to measure 'distance' between objects, in this case, between r.v.s. There are two particular values of p which we are usually interested in

Convergence in Mean $(p=1)$ - $\lim_{n\to\infty} \mathbb{E}[|X_n - X|] = 0$

Convergence in Mean-Square $(p=2)$ - $\lim_{n\to\infty} \left(\mathbb{E}[(X_n - X)^2]\right)^{1/2} = 0$ (and also $\mathbb{E}[X_n^2] < \infty$ $\forall n$)

- All of the above were for X_n and X on the same $(\Omega, \mathcal{F}, \mathbb{P})$. The final mode of convergence is special in that it does not even require this!

5) Convergence in Distribution (or Weak Convergence) - A sequence of r.v. X_n converges to a r.v. X in distribution if

$\lim_{n\to\infty} F_n(t) = F(t)$ $\forall\, t \in \mathbb{R}$ at which $F(t)$ is continuous (notation: $X_n \xrightarrow{d} X$)

- Why so many notions? In a way, this reflects the richness of probability, in that it combines an underlying set Ω, a probability function on sets in the σ-field, functions $X(\omega)$ on Ω (r.v.s), distribution functions of these r.v. and their properties (expectation, variance, etc.). Importantly, they are related as

$$(X_n \xrightarrow{a.s.} X) \Rightarrow (X_n \xrightarrow{P} X) \Rightarrow (X_n \xrightarrow{d} X)$$

$p > q \geq 1$

$$(X_n \xrightarrow{l_p} X) \Rightarrow (X_n \xrightarrow{l_q} X) \Rightarrow (X_n \xrightarrow{l_1} X) \Rightarrow (X_n \xrightarrow{P} X)$$

(implication diagram of the modes of convergence)

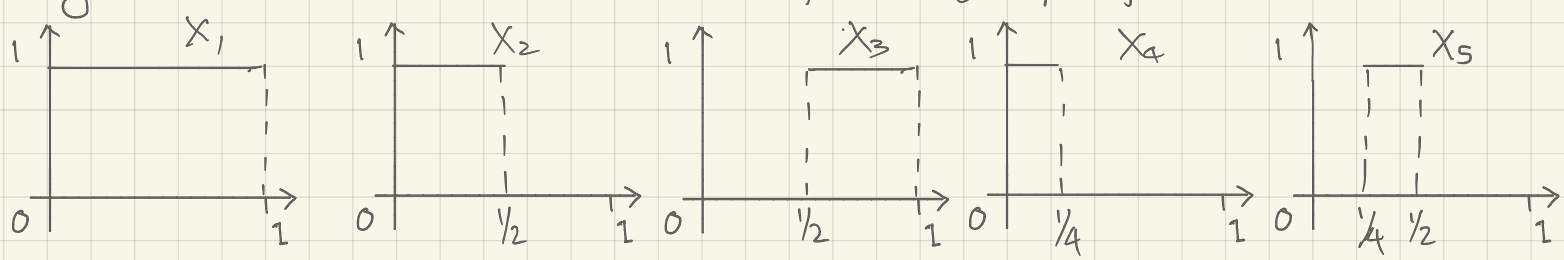
We now build some intuition behind each definition

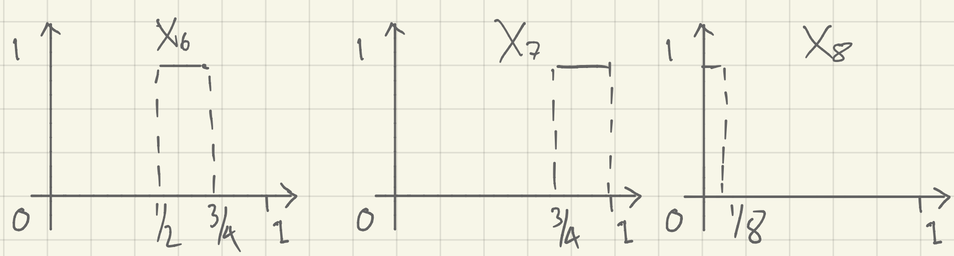
Convergence in almost-sure vs in probability

- These are concerned with the probability of events under r.v s $(X_n; n \geq 1)$ and X defined on a common space $(\Omega, \mathcal{F}, \mathbb{P})$. They differ in the 'relative position' of the $\mathbb{P}$ and lim operators

$$\underbrace{\lim_{n\to\infty} \mathbb{P}[X_n \text{ `='} X] = 0}_{X_n \xrightarrow{P} X} \quad \text{vs} \quad \underbrace{\mathbb{P}[\lim_{n\to\infty} X_n = X] = 0}_{X_n \xrightarrow{a.s} X}$$

Eg - Let $\{X_n; n \geq 1\}$ be the following seq^n of r.v on Uniform$[0,1]$





and in general for $n = 2^k + j$,

$$X_n(\omega) = \begin{cases} 1 & \text{if } \omega \in (j2^{-k}, (j+1)2^{-k}) \\ 0 & \text{ow} \end{cases}$$

- To check if $X_n \xrightarrow{a.s} X$ for some X, fix any ω and consider the sequence $(X_1(\omega), X_2(\omega) \ldots)$. Observe that $\lim_{n\to\infty} X_n(\omega)$ does not exist!
$\Rightarrow$ X_n does not converge a.s. to any r.v.
- However, note also that $\mathbb{P}[X_n > 0] = 1/2^{\lceil \log_2 n \rceil} \xrightarrow[n\to\infty]{} 0$
$\Rightarrow \lim_{n\to\infty} \mathbb{P}[|X_n - 0| > \varepsilon] = 0 \ \forall \varepsilon \Rightarrow X_n \xrightarrow{P} 0$

- A more useful way to think of this is via the set of **Bad Events** $B_n(\varepsilon) = \{\omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}$ and the tail set of bad events $B_n^\infty(\varepsilon) = \{\omega \mid |X_k(\omega) - X(\omega)| > \varepsilon \ \forall k \geq n\}$
  - Now by defn, $X_n \xrightarrow{P} X$ if $\lim_{n\to\infty} \mathbb{P}[B_n(\varepsilon)] = 0$
  - On the other hand, let $C = \{\omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ then by defn $X_n \xrightarrow{a.s} X$ if $\mathbb{P}[C] = 1$
  - Now note that
    - $B_n(\varepsilon) \supseteq B_n^\infty(\varepsilon)$
    - $B_1^\infty(\varepsilon) \subseteq B_2^\infty(\varepsilon) \subseteq \ldots$

    (by sequential continuity) $\Rightarrow \lim_{n\to\infty} \mathbb{P}[B_n^\infty(\varepsilon)] = \mathbb{P}\left[\bigcup_{n=1}^{\infty} B_n^\infty(\varepsilon)\right]$
    - $C \subseteq \bigcup_{n=1}^{\infty} B_n^\infty(\varepsilon) \Rightarrow \mathbb{P}[C] \leq \mathbb{P}\left[\bigcup_{n=1}^{\infty} B_n^\infty(\varepsilon)\right]$

    $\Rightarrow$ If $X_n \xrightarrow{a.s} X$, then $\mathbb{P}\left[\bigcup_{n=1}^{\infty} B_n^\infty(\varepsilon)\right] = 1$
  - Also since $\mathbb{P}[B_n(\varepsilon)] \geq \mathbb{P}[B_n^\infty(\varepsilon)] \ \forall n$

    $\Rightarrow \lim_{n\to\infty} \mathbb{P}[B_n(\varepsilon)] \geq \lim_{n\to\infty} \mathbb{P}[B_n^\infty(\varepsilon)] = 1$

    Thus $X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{P} X$

- Thinking about bad sets also allows us to get a partial converse. First we need an additional defn.

<u>Def</u> - Given a sequence of events $(A_n; n \geq 1)$, the event **A_n occurs infinitely often** (or $\{A_n \text{ i.o.}\}$) is defined as

$$\{A_n \text{ i.o}\} = \{\omega \mid \omega \in A_n \text{ for infinitely many } n\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_n$$

- Lemma (Borel-Cantelli Lemmas). Let $(A_n, n \geq 1)$ be a sequence of events. Then

i) $\sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty \Rightarrow \mathbb{P}[A_n \text{ i.o.}] = 0$

(less useful 'converse') ii) If A_n are independent and $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty \Rightarrow \mathbb{P}[A_n \text{ i.o.}] = 1$

Pf. Note that $\bigcup_{k=n}^{\infty} A_k \supseteq \bigcup_{k=n+1}^{\infty} A_k \supseteq \bigcup_{k=n+2}^{\infty} A_k \ldots$

$$\Rightarrow \mathbb{P}\left[\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right] = \lim_{n\to\infty} \mathbb{P}\left[\bigcup_{k=n}^{\infty} A_k\right] \quad \text{(sequential continuity)}$$

Also $\mathbb{P}\left[\bigcup_{k=n}^{\infty} A_k\right] \leq \sum_{k=n}^{\infty} \mathbb{P}[A_k]$ (union bound)

and since $\sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty \Rightarrow \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}[A_k] = 0$

$$\Rightarrow \mathbb{P}[A_n \text{ i.o}] = \mathbb{P}\left[\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right] \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}[A_n] =$$

- For the converse, w.l.o.g assume $\mathbb{P}[A_n] > 0 \; \forall n \geq 1$

Then $\prod_{k=n}^{\infty} (1 - \mathbb{P}[A_k]) \leq \prod_{k=n}^{\infty} e^{-\mathbb{P}[A_k]} = e^{-\sum_{k=n}^{\infty} \mathbb{P}[A_k]} = 0$ by defn

Also since A_k are $\perp\!\!\!\perp \Rightarrow \mathbb{P}\left[\bigcap_{k=n}^{\infty} \bar{A}_k\right] = \prod_{k=n}^{\infty} (1 - \mathbb{P}[A_k])$

$$= 0$$

$$\Rightarrow \mathbb{P}\left[\bigcap_{k=n}^{\infty} \bar{A}_k\right] = 1 - \mathbb{P}\left[\bigcap_{k=n}^{\infty} \bar{A}_k\right] = 1$$

$$\Rightarrow \mathbb{P}[A_n \text{ i.o}] = \mathbb{E}\left[\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right] = 1$$

∎

- Now returning to $X_n \xrightarrow{a.s} X$ vs $X_n \xrightarrow{a.s} X$

$\boxed{\text{Thm}}$ i) $X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{P} X$

ii) $X_n \xrightarrow{P} X$ $\left(\text{ie. } \lim_{n\to\infty} \mathbb{P}[B_n(\varepsilon)] = 0 \ \forall \varepsilon > 0\right)$ and $\sum_{n=1}^{\infty} \mathbb{P}[B_n(\varepsilon)] < \infty \ \forall \varepsilon > 0$

$\Rightarrow X_n \xrightarrow{a.s} X$

Convergence in probability vs l_p

Recall bad events
$B_n(\varepsilon) = \{\omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}$

- While $X_n \xrightarrow{P} X$ implies that $\mathbb{P}[B_n(\varepsilon)]$ is small, it does not say anything about $|X_n(\omega) - X(\omega)|$ for $\omega \in B_n(\varepsilon)$. This extra 'control' is ensured by l_p convergence.

- l_p norm $\|Y\|_p \triangleq \left(\mathbb{E}[|Y|^p]\right)^{1/p}$ is a norm on r.v. for $p \geq 1$

$\Rightarrow$ 3 properties
i) $\|aY\|_p = a\|Y\|_p$ ii) $\|Y\|_p = 0 \Rightarrow Y = 0$ a.s
iii) $\|Y+Z\|_p \leq \|Y\|_p + \|Z\|_p$ (triangle inequality)

Eg - Consider $(X_n ; n \geq 0)$ where $X_n = \begin{cases} a_n & \text{for } \omega \in [0, 1/n] \\ 0 & \text{ow} \end{cases}$

- For any a_n, we have $\mathbb{P}[B_n(\varepsilon)] = 1/n \searrow 0 \ \forall \varepsilon > 0$
- If $a_n \searrow 0$, then $\mathbb{P}[\lim_{n\to\infty} X_n(\omega) = 0] = 1 \Rightarrow X_n \xrightarrow{a.s} X$
(Note - $\sum_{n=1}^{\infty} \mathbb{P}[B_n(\varepsilon)] = \infty$ but B_n not iid $\Rightarrow$ can't use Borel Cantelli)
- $\left(\mathbb{E}[(X_n - 0)^2]\right)^{1/2} = \frac{a_n}{\sqrt{n}} \Rightarrow$ for $X_n \xrightarrow{m.s} X$, we need $\lim_{n\to\infty} \frac{a_n}{\sqrt{n}} = 0$

Thm - Convergence in $\mathbb{P}$ and ℓ_p are related as follows

i) If $r > s \geq 1$, then $X_n \xrightarrow{\ell_r} X \Rightarrow X_n \xrightarrow{\ell_s} X$

ii) If $X_n \xrightarrow{\ell_1} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$

iii) If $X_n \xrightarrow{\mathbb{P}} X$ <u>and</u> $\mathbb{P}[X_n < k] = 1 \ \forall n$ for some k then $X_n \xrightarrow{\ell_r} X$ for all $r \geq 1$

We first need 2 inequalities, which on their own are perhaps more useful!

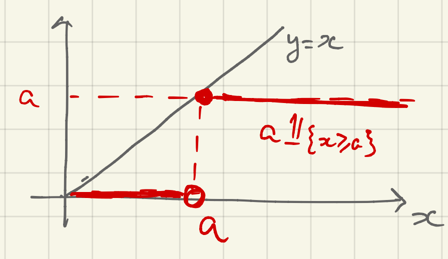
- (Markov's Inequality) For any <u>non-negative</u> r.v Z, and any $a > 0$

$$\mathbb{P}[Z \geq a] \leq \mathbb{E}[Z]/a$$

<u>Pf</u> - Observe that $(a \mathbb{1}_{\{x \geq a\}}) \geq x \ \forall x \geq 0$

$\Rightarrow \mathbb{E}[Z] \leq \mathbb{E}[a \cdot \mathbb{1}_{\{Z \geq a\}}]$

$= a \mathbb{P}[Z \geq a$



- (Jensen's Inequality) Given any r.v Z and fn f

i) If f is convex $\Rightarrow \mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$

ii) If f is concave $\Rightarrow \mathbb{E}[f(x)] \leq f(\mathbb{E}[x])$

(We will see this in more detail in the assignment)

- <u>Proposition</u> - If $p > q \geq 1$, then $\|X\|_p \geq \|X\|_q$

<u>Pf</u> - For $x \geq 0$, let $f(x) = x^{p/q} \Rightarrow f''(x) = \frac{p}{q}\left(\frac{p}{q}-1\right) x^{p/q-2} \geq 0$

for all $p > q \Rightarrow f$ is convex

Also given any r.v. X, let $Y = X^q$.

By Jensen's Inequality we have $f(\mathbb{E}[Y]) \leq \mathbb{E}[f(Y)]$

$\Rightarrow \left(\mathbb{E}[X^q]\right)^{p/q} \leq \mathbb{E}\left[(X^q)^{p/q}\right] = \mathbb{E}[X^p]$

$\Rightarrow \quad \|X\|_q \leq \|X\|_p$

- <u>Pf of (i) in theorem</u>

$r > s \Rightarrow \mathbb{E}\left[|X_n - X|^r\right]^{1/r} \geq \mathbb{E}\left[|X_n - X|^s\right]^{1/s}$

Also $X_n \xrightarrow{\ell_r} X \Rightarrow \lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right]^{1/r} = 0$

$\Rightarrow \lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^s\right]^{1/s} = 0 \Rightarrow X_n \xrightarrow{\ell_s} X$

- <u>Pf of (ii) in theorem</u>

$X_n \xrightarrow{\ell_1} X \Rightarrow \lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0$

By Markov's Inequality, $\mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \frac{\mathbb{E}\left[|X_n - X|\right]}{\varepsilon} \quad \forall \varepsilon > 0 \;\; \forall n$

$\Rightarrow$ for any $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \lim_{n\to\infty} \frac{\mathbb{E}\left[|X_n - X|\right]}{\varepsilon} = 0$

$\Rightarrow X_n \xrightarrow{P} X$

- Pf of (iii) in theorem

$X_n \xrightarrow{P} X$ and $\mathbb{P}[|X_n| \leq k] = 1 \Rightarrow \mathbb{P}[|X| \leq k] = 1$ (prove this!)

Now for any $r \geq 1$,

$$\mathbb{E}[|X_n - X|^r] = \mathbb{E}[|X_n-X|^r \mid |X_n-X| < \varepsilon] \overbrace{\mathbb{P}[|X_n-X| < \varepsilon]}^{\leq 1}$$
$$+ \mathbb{E}[\underbrace{|X_n-X|^r}_{\leq (2k)^r} \mid |X_n-X| \geq \varepsilon] \mathbb{P}[|X_n-X| \geq \varepsilon]$$
$$\leq \varepsilon^r + (2k)^r \mathbb{P}[|X_n-X| \geq \varepsilon]$$

$\because X_n \xrightarrow{P} X$, $\lim_{n\to\infty} \mathbb{P}[|X_n-X| \geq \varepsilon] = 0$ for any ε. Finally we can take $\varepsilon \searrow 0$ to get $\mathbb{E}[|X_n-X|^r] \searrow 0$

$\Rightarrow X_n \xrightarrow{r} X$

Note - The above style of proof is very typical and important - it will show up repeatedly in this course, starting from next week!

- Markov's Inequality can also be used to give stronger bounds

- (Chebyshev's Inequality) - For any r.v. X, and $t \geq 0$

$$\mathbb{P}[|X - \mathbb{E}[X]| > \varepsilon] \leq \frac{Var(X)}{\varepsilon^2}$$

Pf - $\mathbb{P}[|X-\mathbb{E}[X]| > \varepsilon] = \mathbb{P}[|X-\mathbb{E}[X]|^2 > \varepsilon^2]$

$$\leq \frac{\mathbb{E}[(X-\mathbb{E}X)^2]}{\varepsilon^2} \quad \text{(By Markov's)}$$

$\blacksquare$

The Law of Large Numbers

- The most famous applications of a.s and p convergence in probability!

- **Thm** (Weak Law of Large Numbers) - Let $\{X_i ; i \geq 1\}$ be an iid sequence of r.v.s. Then $\frac{1}{n}\sum_{i=1}^{n} X_n \xrightarrow{P} \mathbb{E}[X]$

If $X \perp\!\!\!\perp Y$
$Var(X+Y) = Var(X)+Var(Y)$

Pf - Let $S_n = \sum_{i=1}^{n} X_i \Rightarrow \mathbb{E}[S_n] = \sum_{i=1}^{n}\mathbb{E}[X_i] = n\,\mathbb{E}[X]$

$Var(S_n) = \sum_{i=1}^{n} Var(X_i) = n\,Var(X)$

Now $\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > \varepsilon\right] = \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X])\right| > \varepsilon\right]$

$= \mathbb{E}\left[\left|\frac{S_n}{n} - \mathbb{E}[X]\right| > \varepsilon\right]$

$Var(aX) = a^2 Var(X)$

$\leq \frac{Var(S_n)}{n^2\varepsilon^2} = \frac{Var(X)}{n\varepsilon^2}$

$\Rightarrow \lim_{n\to\infty} \mathbb{P}\left[\left|S_n - \mathbb{E}[X]\right| > \varepsilon\right] = 0 \Rightarrow S_n \xrightarrow{P} \mathbb{E}[X]$

- Now we want to convert this to a.s. We will do this via Borel-Cantelli to get the result assuming $\mathbb{E}[X^4] = m_4 < \infty$.

<u>Note</u>: this is a more strict condition than we need for the SLLN - we will see a more general version when we study Martingales

Thm (Borel's Strong Law of Large Numbers) - Let $(X_i; i \geq n)$ be iid r.v with $\mathbb{E}[X_i] = \mu$, $Var(X_i) = \sigma^2$ and $\mathbb{E}[(X_i-\mu)^4] = m_4 < \infty$. Then

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} \mathbb{E}[X]$$

Pf - $\mathbb{P}\left[\left|\frac{S_n}{n}-\mu\right| > \varepsilon\right] = \mathbb{P}\left[\left|\frac{S_n}{n}-\mu\right|^4 > \varepsilon^4\right]$

$$\leq \mathbb{E}\left[\left(\frac{S_n}{n}-\mu\right)^4\right] \cdot \frac{1}{\varepsilon^4}$$

$$= \frac{\mathbb{E}\left[\left(\sum_{i=1}^{n}(X_i-\mu)\right)^4\right]}{n^4\,\varepsilon^4}$$

Let $Y_i = X_i - \mu \Rightarrow \mathbb{E}[Y_i] = 0,\ Var(Y_i) = \sigma^2$

also $Y_i \perp\!\!\!\perp Y_j \Rightarrow \mathbb{E}[Y_i Y_j^3] = \mathbb{E}[Y_i Y_j Y_k Y_l] = 0 \quad \forall i,j,k,l$ distinct

$$\Rightarrow \mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4\right] = \sum_{(i,j,k,l)} \mathbb{E}[Y_i Y_j Y_k Y_l]$$

$$= n\,\mathbb{E}[Y_i^4] + 3n(n-1)\,\mathbb{E}[Y_i^2 Y_j^2]$$

$$= n\,m_4 + 3n(n-1)\,\sigma^4$$

$$\Rightarrow \mathbb{P}\Big[\underbrace{\left|\frac{S_n}{n}-\mu\right| > \varepsilon}_{B_n(\varepsilon)}\Big] \leq \frac{m_4}{n^3\varepsilon^4} + \frac{3\sigma^4}{n^2\varepsilon^4}$$

(Borel-Cantelli)

Now since $\sum_{n=1}^{\infty} B_n(\varepsilon) < \infty \ \forall \varepsilon \Rightarrow \mathbb{P}[B_n(\varepsilon)\ i.o.] = 0$

$$\Rightarrow \frac{S_n}{n} \xrightarrow{a.s} \mu$$

Weak Convergence

- Unlike all the previous notions of convergence, convergence in distribution does not need X_n, X to be on the same $(\Omega, \mathcal{F}, \mathbb{P})$.
- Even otherwise, the idea is somewhat counterintuitive...

Eg - Let $X \sim \text{Bernoulli}(1/2)$, and $X_1, X_2, \ldots$ be identical r.v. given by $X_n = X$ for all n.

- X_n are not independent, but clearly $X_n \xrightarrow{d} X$ (and indeed, in all modes of conv!)
- Now let $Y = 1 - X$. $\because$ X and Y have the same distribution

$\Rightarrow X_n \xrightarrow{d} Y$. Note though that $|X_n - Y| = 1 \ \forall n$!

- Another aspect to get used to is that $X_n \xrightarrow{d} X$ only requires $\lim_{n\to\infty} F_n(t) = F(t)$ at continuity points of $F(\cdot)$

Eg - Let X be any r.v., and $X_n = X + 1/n$

$$\Rightarrow F_n(t) = \mathbb{P}[X_n \leq t] = \mathbb{P}[X \leq t - 1/n] = F(t - 1/n)$$

Thus $\lim_{t\to\infty} F(t - 1/n) = F(t)$, but only at points where F is continuous — this is because we defined F in a way that it is RCLL (continuous from the right, but only having a limit from the left). However, we do not want this arbitrary convention to make us decide such an example is not converging in distribution (it would if we assumed LCRL...)

- So if convergence in distribution is 'weak', why do we care. Should we not always strive for $X_n \xrightarrow{a.s} X$?

Not so fast...

Thm (Skorohod Representation Theorem) Given r.v.s $(X_n; n \geq 1)$ and X, with distributions $(F_n; n \geq 0)$ and F, s.t $X_n \xrightarrow{d} X$ (i.e. $F_n(t) \to F(t)$). Then $\exists$ probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and r.v.s $(Y_n; n \geq 1)$ and Y on $(\Omega, \mathcal{F}, \mathbb{P})$ s.t the following are true

i) $Y_n \sim F_n \ \forall n$, $Y \sim F$

ii) $Y_n \xrightarrow{a.s} Y$

- This is a somewhat magical theorem, and one of the first examples you will see of a 'probabilistic way of thinking'. Essentially, it takes a setting, moves it to another space using 'probability magic', and then get a very different property!
- The proof though, is 'elementary' - it constructs $(\Omega, \mathcal{F}, \mathbb{P})$, Y_n, Y in a 'natural' way, and then carefully make sure all definitions work.

Proof - First, we choose $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}(0,1)$ (i.e., the Borel σ-algebra on $(0,1)$), and $\mathbb{P}$ as the Lebesgue measure (i.e., the 'usual' notion of length).

- Now we define Y_n, Y in a 'natural' way

$$Y_n(\omega) = \inf_x \{ \omega \in (0,1) \mid \omega \leq F_n(x) \}$$

$$Y(\omega) = \inf_x \{ \omega \in (0,1) \mid \omega \leq F(x) \}$$

This is the natural notion of the inverse fn of F_n, F

- Note that by definition, we have shown (i)!

$$\mathbb{P}[Y_n \leq x] = \mathbb{P}[\{\omega \in [0, F_n(x)]\}] = F_n(x)$$

$$\mathbb{P}[Y \leq x] = \mathbb{P}[\{\omega \in [0, F(x)]\}] = F(x)$$

- Finally we want to argue that $\mathbb{P}[Y_n \leq x]$ converges to $\mathbb{P}[Y \leq x]$ for all 'continuity points' of $F(x)$. If F_n, F are absolutely continuous, then this is true by definition! (Essentially $Y_n = F_n^{-1}(U)$, $Y = F^{-1}(U)$)

- Else, for ω pt of continuity and $\varepsilon > 0$, we pick x as a pt of continuity s.t. $Y(\omega) - \varepsilon < x < Y(\omega)$ and $x < Y_n(\omega)$ for large enough $n \Rightarrow \liminf_{n \to \infty} Y_n(\omega) \geq Y(\omega)$ $\forall\, \omega \in \Omega$

- Similarly show $\limsup_{n\to\infty} Y_n(\omega) \leq Y(\omega) \ \forall\, \omega \in \Omega'$
- Combining we get $Y_n(\omega) \longrightarrow Y(\omega)$ for all points ω of continuity of Y.
- Finally we use the following fact $\equiv$ Any monotone non-decreasing fn on a compact set has a countable # of discontinuities

$\Rightarrow \quad Y_n(\omega) \longrightarrow Y(\omega)$ for almost all ω !

<u>Note</u> - The above proof is somewhat technical, and only given for illustration - its ok for this course if you do not get all the continuity details! The result though is super useful, for eg, for the following

---

<u>Thm</u> - Suppose $X_n \xrightarrow{d} X$. Then

i) $g(X_n) \xrightarrow{d} g(X)$ for all continuous fns g

ii) $\mathbb{E}[g(X_n)] \longrightarrow \mathbb{E}[g(X)]$ for all bounded cont fn g

<u>Pf</u> - For (i), consider the $Y_n \xrightarrow{a.s.} Y$ from the Skorohod representation. Then $g(Y_n) \xrightarrow{a.s.} g(Y) \Rightarrow g(X_n) \xrightarrow{d} g(X)$

For (ii), use bounded convergence

Summary

- For $\{X_n; n \geq 0\}$, X on same $(\Omega, \mathcal{F}, \mathbb{P})$
  - $X_n \xrightarrow{a.s} X$ if $\mathbb{P}[\lim_{n\to\infty} X_n = X] = 1$
  - $X_n \xrightarrow{P} X$ if $\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0 \ \forall \varepsilon > 0$
  - $X_n \xrightarrow{l_q} X$ if $\lim_{n\to\infty} \|X_n\|_q = \|X\|_q$, where $\|X\|_q = \left(\mathbb{E}[|X|^q]\right)^{1/q}$
- For any $X_n \sim F_n$, $X \sim F$, $X_n \xrightarrow{d} X$ if $\lim_{n\to\infty} F_n(t) = F(t)$ for all t points of continuity of F
-

$$
\begin{tikzcd}
 & & \text{a.s} \arrow[dr] & & \\
 & & & P \arrow[r] \arrow[ul, bend right, "\text{If } \sum_{n=1}^{\infty} \mathbb{P}[|X_n - X| > \varepsilon] < \infty \ \forall \varepsilon > 0 \ \text{(Borel Cantelli)}"'] \arrow[dl, bend left, "\text{if } |X_n| < k \text{ a.s}"] & d \\
(p>q)\ l_p \arrow[r] & l_q \arrow[r] & l_1 \arrow[ur] & &
\end{tikzcd}
$$

- $X_n \xrightarrow{d} X \Rightarrow Y_n \sim F_n$, $Y \sim F$, $Y_n \xrightarrow{a.s} Y$ (Skorohod representation)

  $X_n \xrightarrow{d} X \Longrightarrow g(X_n) \xrightarrow{d} g(X)$ for continuous g

  $\iff \mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for bounded cont g
- $\forall X \geq 0$, $\mathbb{P}[X \geq a] \leq \mathbb{E}[X]/a$ (Markov's Ineq)
- $\forall X$, if f is convex $\Rightarrow f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$ (Jensen's Ineq)