Math 170E: Winter 2023

Lecture 25, Wed 15th Mar

Weak law of large numbers, the MGF technique, and limiting MGFs

Last time:

- Let $X_1, \ldots, X_n$ be i.i.d. and define $S_n = \sum_{j=1}^{n} X_j$ and $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$
- If $\mathbb{E}[X_j] = \mu$ for each j, then
$$\mathbb{E}[S_n] = n\mu \quad \text{and} \quad \mathbb{E}[\overline{X}_n] = \mu$$
- If $\text{var}(X_j) = \sigma^2$ for each j, then
$$\text{var}(S_n) = n\sigma^2 \quad \text{and} \quad \text{var}(\overline{X}_n) = \frac{\sigma^2}{n}$$
- Given a sequence of random variables $X_1, X_2, \ldots,$ we say they converge in probability to a random variable X if, for any $\varepsilon > 0$,
$$\mathbb{P}(|X - X_n| \geq \varepsilon) \to 0 \quad \text{as } n \to \infty$$
- **Weak Law of Large Numbers**: Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with finite mean μ.
Then,

$\mathbb{P}(|\overline{X}_n - \mathbb{E}[\overline{X}_n]| \geq \varepsilon)$

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad \text{in probability as } n \to \infty.$$

Proposition 5.14: (Generalised Markov's inequality)

Let X be a non-negative random variable. Then, given $\lambda > 0$ and integer $k \geq 1$, we have

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X^k]}{\lambda^k}.$$

Proposition 5.15: (Chebyshev's inequality)

Let X be a random variable with mean μ and variance σ^2. Then, given $\lambda > 0$, we have

$$\mathbb{P}(\underbrace{|X - \mu| \geq \lambda}_{=\{X-\mu \geq \lambda\} \cup \{X-\mu \leq -\lambda\}}) \leq \frac{\sigma^2}{\lambda^2}.$$

Proof:

Set $Y = |X-\mu|$ (non-negative).

So by Gen. Markov with $k=2$:

$$\mathbb{P}(|X-\mu| \geq \lambda) = \mathbb{P}(Y \geq \lambda) \leq \frac{\mathbb{E}[Y^2]}{\lambda^2} = \frac{\overset{\sigma^2}{\mathrm{Var}(X)}}{\lambda^2}$$

Example 7:

$\mathbb{E}[X] = 5$, $\mathrm{Var}(X) = 10 \times \frac{1}{2} \times \frac{1}{2} = 5/2$.

- Let $X \sim \text{Binomial}(10, \frac{1}{2})$
- What estimate does Chebyshev's inequality give for $\mathbb{P}(X \geq 6)$?

$$\mathbb{P}(X \geq 6) = \mathbb{P}(X - 5 \geq 6 - 5)$$
$$= \mathbb{P}(X - 5 \geq 1)$$
$$\leq \mathbb{P}(X-5 \geq 1) + \boxed{\mathbb{P}(X - 5 \leq -1)}$$
$$= \mathbb{P}(|X-5| \geq 1) \underset{\text{Chebyshev}}{\leq} \frac{\mathrm{Var}(X)}{1^2} = \frac{5}{2}$$

If could we that $\mathbb{P}(X \geq 6) = \mathbb{P}(X \leq 4) = \mathbb{P}(X - 5 \leq -1)$

$$\hookrightarrow \mathbb{P}(X \geq 6) = \frac{1}{2}\left[\mathbb{P}(X \geq 6) + \mathbb{P}(X \leq 4)\right]$$
$$= \frac{1}{2}\mathbb{P}(|X-5| \geq 1) \leq 5/4.$$

<u>Chebyshev is useful:</u> – σ^2 is small

– λ is big.

- $\mathbb{P}(X \geq 8) = \mathbb{P}(X-5 \geq 3) \leq \mathbb{P}(|X-5| \geq 3)$

$$\leq \frac{\left(\frac{5}{2}\right)}{9} = \frac{5}{18} \approx 0.28.$$

Gen. Markov with $k=2$

$$\mathbb{P}(X \geq 8) \leq \frac{\mathbb{E}[X^2]}{8^2} = \frac{55}{128} \approx 0.43$$

Tao's blog: Strong law.

Theorem 5.12: (The Weak Law of Large Numbers)

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with finite mean μ. Then,

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad \text{in probability as } n \to \infty.$$

Proof: (Assuming $\sigma^2 = \mathrm{var}(X_j)$ is finite)

Let $\varepsilon > 0$. Want to show that

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mathbb{E}[\overline{X}_n]| \geq \varepsilon) = 0.$$

Recall: $\mathbb{E}[\overline{X}_n] = \mu$, $\mathrm{var}(\overline{X}_n) = \sigma^2/n$.

So by chebyshev inequality,

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \longrightarrow 0 \text{ as } n \to \infty$$

The MGF technique

Goals:

- How to use the MGF to identify the distribution of an r.v.
- How point-wise convergence of a sequence of MGFs determines the distribution of the limit of r.v.s

Proposition 5.13: Let $\underbrace{X_1, X_2, \ldots, X_n}$ be a sequence of independent random variables and let $a_1, a_2, \ldots, a_n \in \mathbb{R}$. Then the random variable

$$Y = \sum_{j=1}^{n} a_j X_j$$

has MGF

$$M_Y(t) = \prod_{j=1}^{n} M_{X_j}(a_j t),$$

whenever it is well-defined.

$$M_Y(t) = \mathbb{E}[\exp(tY)] = \mathbb{E}\left[\exp\left(t \sum_{j=1}^{n} a_j X_j\right)\right],$$

$$= \mathbb{E}\left[\exp\left(\sum_{j=1}^{n} t a_j X_j\right)\right]$$

$$= \mathbb{E}\left[\prod_{j=1}^{n} \exp(t a_j X_j)\right],$$

indep.

$$= \prod_{j=1}^{n} \underbrace{\mathbb{E}[\exp(t a_j X_j)]} = \prod_{j=1}^{n} M_{X_j}(a_j t)$$

Proposition 5.14: (Uniqueness via the MGF) Let X and Y be random variables with MGFs $M_X(t)$ and $M_Y(t)$. Suppose that for some $h > 0$ and all $t \in (-h, h)$, we have

$$M_X(t) = M_Y(t).$$

Then, X and Y are identically distributed.

$\leadsto$ "Same MGF" $\Rightarrow$ "same distribution"!

Proof: Beyond this class :)

If X is discrete & takes values in $\mathbb{N} = S_X$, then

$$\underbrace{\mathbb{P}(X=n)}_{\text{PMF of } X} = \frac{1}{n!} \frac{d^n}{dt^n} \underbrace{M_X(\log t)}_{\text{If we knew } M_X} \Big|_{t=0}.$$

PMF of X $\Longleftarrow$ If we knew M_X

$$\mathbb{P}(Y \le y) = \mathbb{P}\left(\sum_{j=1}^{n} a_j X_j \le y\right)$$

Example 7:

- Let $X_1, X_2, \ldots, X_n$ be independent r.v.s so that $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$
- Let $a_1, a_2, \ldots, a_n \in \mathbb{R}$.
- What is the distribution of $Y = \sum_{j=1}^n a_j X_j$?

$M_{X_j}(t) = \exp(\mu_j t + \frac{1}{2}\sigma_j^2 t^2)$, $j = 1, \ldots, n$, for all $t \in \mathbb{R}$.

As $\{X_j\}_{j=1}^n$ are indep, by our previous Prop$^{\underline{n}}$ 5.13,

$$M_Y(t) = \prod_{j=1}^n M_{X_j}(a_j t)$$

$$= \prod_{j=1}^n \exp\left(\mu_j a_j t + \frac{1}{2}\sigma_j^2 a_j^2 t^2\right).$$

$$= \exp\left(\left(\sum_{j=1}^n \mu_j a_j\right) t + \frac{1}{2}\left(\sum_{j=1}^n \sigma_j^2 a_j^2\right) t^2\right).$$

So by Theorem 5.14,

$$Y \sim N\left(\sum_{j=1}^{n} \mu_j a_j, \sum_{j=1}^{n} \sigma_j^2 a_j^2\right)$$

$\hookrightarrow$ any linear combination of indep. Normal r.v.s is again a normal r.v.

$X_j \sim N(0,1)$ $\quad Y = \sum_{j=1}^{n} a_j X_j$

$\hookrightarrow$ $Y \sim N\left(0, \sum_{j=1}^{n} a_j^2\right).$

Proposition 5.14:

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with common MGF $M(t)$. Then

$$M_{S_n}(t) = \left[M(t)\right]^n,$$

$$M_{\overline{X}_n}(t) = \left[M\left(\tfrac{t}{n}\right)\right]^n$$

Proof:

ident. dist. $M_{X_j}(t) = M_{X_1}(t)$ for all $j = 1, \ldots, n$.

$a_j = 1$

$S_n = \sum_{j=1}^{n} X_j$

$$M_{S_n}(t) = \prod_{j=1}^{n} M_{X_j}(t) = \prod_{j=1}^{n} M_{X_1}(t) = \left[M_{X_1}(t)\right]^n$$

$a_j = 1/n$.

$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$

$$\hookrightarrow M_{\overline{X}_n}(t) = \prod_{j=1}^{n} M_{X_1}(t/n) = \left[M_{X_1}(t/n)\right]^n.$$

Example 8:

- Let $0 < p < 1$ and $X_1, X_2, \ldots, X_n$ be an i.i.d sequence of Bernoulli(p) random variables
- What is the distribution of S_n?

$$S_n = \sum_{j=1}^{n} X_j . \quad [\sim Bin(n,p)]$$

$$M_{X_j}(t) = 1-p+pe^t, \text{ for all } j=1,\ldots,n, \ t\in\mathbb{R}.$$

$$\hookrightarrow M_{S_n}(t) = [M_{X_1}(t)]^n = \underbrace{(1-p+pe^t)^n}_{\text{This is the MGF of a } Bin(n,p) \text{ r.v.}}.$$

So by Theorem 5.14,

$$S_n \sim Bin(n,p).$$

Example 9:

- Let $\theta > 0$ and $X_1, X_2, \ldots, X_n$ be an i.i.d sequence of Exponential(θ) random variables

- What is the distribution of S_n?

Recall that if $X \sim Exp(\theta)$, then $M_X(t) = \frac{1}{1-\theta t}$, $t < 1/\theta$.

So

$$M_{S_n}(t) = [M_{X_1}(t)]^n = \underbrace{\frac{1}{(1-\theta t)^n}}_{\text{MGF of Gamma}(n,\theta)}, \quad t < 1/\theta$$

$\Rightarrow S_n \sim Gamma(n,\theta)$, by Theorem 5.14.

$$X_1, X_2, \ldots \rightsquigarrow \boxed{M_{X_n}(t) \to M(t)}$$

$$\boxed{X_n \to X}$$

We want to take $n \to \infty$: Limiting MGFs

Convergence in distribution:

- Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables and X be another random variable

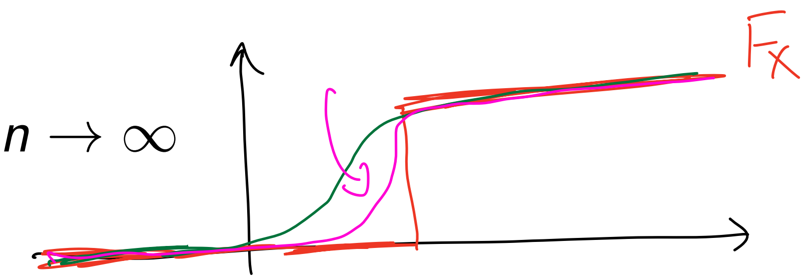
- We say that

$$X_n \to X \quad \text{in distributuon: as} \quad n \to \infty$$

if the CDF

'' $\mathbb{P}(X_n \le x)$.

$$F_{X_n}(x) \to F_X(x) \quad \text{as } n \to \infty$$



for all $x \in \mathbb{R}$ where F_X is continuous at x.

In this course, I am never going to ask you to prove a sequence of r.v.s. "Converge in distribution" using only this definition. I do expect you to be able to prove convergence in distribution using Theorem 5-15 below and I will ask about this.

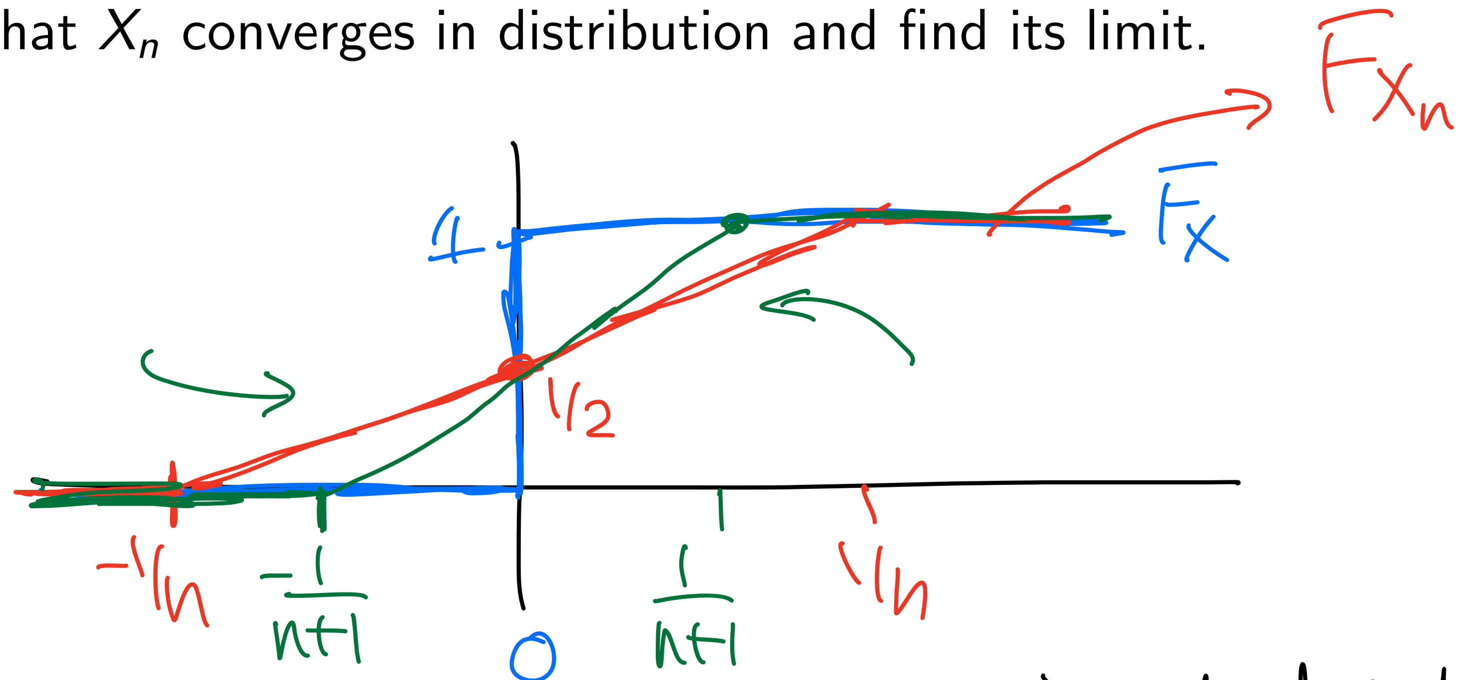
→ The following example proves convergence in distribution for a particular sequence of r.v.s. The full details are here for you to:

(i) be convinced that, even in this simple situation, one needs to argue carefully

(ii) compare with an argument in lecture (below) which uses Theorem 5-15 to prove the same result (i.e. conv. in distribution).

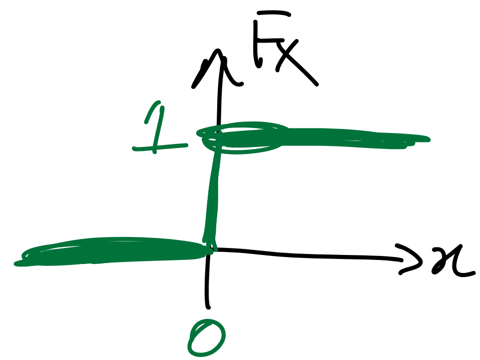
Example 10:

- Let $X_n \sim \text{Uniform}\big(\big(-\frac{1}{n}, \frac{1}{n}\big)\big)$
- Show that X_n converges in distribution and find its limit.



The limit should be a discrete r.v. X which takes values only on $S_X = \{0\}$ & has PMF $P_X(0) = 1$.

This r.v. has CDF $F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$



We need to show that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$

at every x for which F_X is continuous.

i.e. at every non-zero $x \neq 0$! (F_X is discts only at $x=0$).

First need CDF of X_n

$$f_{X_n}(x) = \begin{cases} n/2 & \text{if } -1/n < x < 1/n \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow F_{X_n}(x) = \int_{-\infty}^{x} f_{X_n}(t)\,dt = \begin{cases} 0 & \text{if } x \leq -1/n \\ \frac{nx}{2} + \frac{1}{2} & \text{if } -1/n < x < 1/n \\ 1 & \text{if } x \geq 1/n. \end{cases}$$

(see picture above).

For $x \neq 0$, $F_{X_n}(x) \longrightarrow F_X(x)$.

Why? Suffices to show that

$$g_n(x) := \left(\frac{nx}{2} + \frac{1}{2}\right) \mathbb{1}_{\{-1/n < x < 1/n\}}(x) \quad |x| < 1/n$$

converges to 0 as $n \to +\infty$, for each $x \neq 0$.

We have $|g_n(x)| \leq \frac{|nx+1|}{2} \mathbb{1}_{\{|x|<1/n\}}$.

depends on x

Fix $\varepsilon > 0$ and $|x| > 0$. Then, there exists $N \in \mathbb{N}$, $N = N(x)$, such that $|x| > 1/N$. So for all $n > N$, we have

$$|x| > 1/N > 1/n$$

$$\Rightarrow |g_n(x)| = 0 < \varepsilon.$$

So $g_n(x) \to 0$ if $x \neq 0$.

Note: If $x = 0$, $g_n(x) = 1/2 \to 1/2$ as $n \to \infty$

So we have shown that $F_{X_n}(x) \to F_X(x)$ as $n \to +\infty$ for all $x \neq 0$

$$\Rightarrow X_n \to X \text{ in distribution as } n \to +\infty.$$

Note: $F_{X_n}(0) = 1/2$ for all n.

So $F_{X_n}(0) \not\to F_X(0)$ as $n \to \infty$.

This is not a problem b/c F_X is not continuous at $x = 0$.

THEOREM 5.15

~~Proposition 5.?~~: (Limiting MGF determines the distribution)

Let $X_1, X_2, \ldots$ and X be random variables. Suppose that for some $h > 0$ and all $t \in (-h, h)$, we have

$$M_{X_n}(t) \to M_X(t) \quad \text{as } n \to \infty.$$

Then, $X_n \to X$ in distribution as $n \to \infty$.

Proof: Beyond this class.

- $X_n \sim \text{Uniform}((-1/n, 1/n))$

$X_n \to X$ in distribution where X is discrete $\mathbb{P}(X=0)=1$.

$$M_{X_n}(t) = \mathbb{E}[e^{tX_n}] = \int_{-1/n}^{1/n} e^{tx} \frac{1}{2/n}\, dx.$$

$$= \frac{e^{t/n} - e^{-t/n}}{2t/n}.$$

$$\lim_{x \to 0} \frac{e^{ax} - 1}{ax} = 1. \quad a \neq 0$$

$$= e^{-t/n} \cdot \frac{e^{2t/n}-1}{2t/n}$$

$e^{-t/n} \to e^0 = 1$, $\quad \frac{e^{2t/n}-1}{2t/n} \to 1$

$$\frac{e^{2t/n}-1}{2t/n} = \frac{\frac{2t}{n} + O\left(\left(\frac{2t}{n}\right)^2\right)}{2t/n} = 1 + \underbrace{O\left(\tfrac{2t}{n}\right)}_{\to 0}$$

So $\lim_{n\to\infty} M_{X_n}(t) = 1 = M_X(t)$

$= e^{t\cdot 0}\,\mathbb{P}(X=0)$

So we have by Theorem 5.15,

$X_n \to X$ in distribution as $n \to +\infty$