Admin
- HW 5 due Oct 23RD 11:59 PM (GRADED FOR COMPLETION)
- FINAL EXAM Oct 22ND 1:10 - 3:40

Recap
Goal: learn about/prove The Law of Large Numbers (LLN) ✓
Central Limit Theorem (CLT)

Recall: (Weak Law of Large numbers) If $X_1, \dots, X_n$ are iid with finite mean $\mathbb{E}[X]$, Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_i]$$

$\xrightarrow{P}$ means convergence in probability, which, in This case, means That

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > t\right) \to 0 \text{ as } n \to \infty \text{ for all } t > 0$$

We proved This using Chebyshev's inequality, which said That if Y has finite $\mathbb{E}[X]$ and finite var(X), Then

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq t) \leq \frac{\text{var}(X)}{t^2}$$

Assuming $\text{var}(X_i) < \infty$ are plugging in $\frac{1}{n}\sum X_i$, we found That
$\mathbb{E}[\frac{1}{n}\sum X_i] = \mathbb{E}[X_i]$ and $\text{var}(\frac{1}{n}\sum X_i) = \frac{\text{var}(X)}{n} \to 0 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_i]$

Convergence in probability: How to prove in general?
Want to prove That $Y_n \xrightarrow{P} c$ (for Y_n r.v. and a constant, c)
Specifically, we need to show That $\forall t > 0$, $\mathbb{P}(|Y_n - c| > t) \to 0$ as $n \to \infty$

① Directly from The def, using a probability inequality

$$\mathbb{P}(|Y_n - c| > t) \leq \text{something} \to 0$$

We can obtain upper bounds using probability inequality for some r.v. W
- Chebyshev (requires $\text{var}(W) < \infty$ for all n (or all n large enough))
- Markov's inequality: for any $W \geq 0$ (W non-negative) we have That

$$\mathbb{P}(W \geq t) \leq \frac{\mathbb{E}[W]}{t}$$ (we could apply with $W = |Y_n - c|$)

Markov ⇒ Chebyshev
Proof: W has mean 0 and $\text{var}(W) < \infty$ (You can always define $W = Y - \mathbb{E}[Y]$)

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq t) = \mathbb{P}(|W| \geq t) = \mathbb{P}(W^2 \geq t^2) \leq \frac{\mathbb{E}[W^2]}{t^2} = \frac{\text{var}(W)}{t}$$

Markov's inequality

$\text{var}(W) = \mathbb{E}[W^2] - \mathbb{E}[W]^2 = \mathbb{E}[W^2]$

- Chernoff's inequality for $W \geq 0$

$$\mathbb{P}(W > t) \leq e^{-st} M_X(s) \quad \text{for some } s \in \mathbb{R}$$

(e^{-st}: dependence on t is exponential; $M_X(s)$: moment generating function)

Proof: in lecture notes (good exercise in Markov's ineq.)

② (Comes out of the first strategy) Show that $\mathbb{E}[Y_n] \to c$ and $\text{var}(Y_n) \to 0$ (as $n \to \infty$). Chebyshev $\Rightarrow Y_n \xrightarrow{P} c$

③ Continuous mapping theorem: If $W_n \xrightarrow{P} a$ and $Y_n = g(W_n)$ where g is continuous, then $Y_n = g(W_n) \xrightarrow{P} g(a)$

LLN $\Rightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mathbb{E}[X]$

$$\bar{X}^2 \xrightarrow{P} \mathbb{E}[X]^2$$

Careful: $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \mathbb{E}[X_i^2] \neq \mathbb{E}[X_i]^2$

We're using that if $\mathbb{E}[g(X_i)] < \infty$ then $\frac{1}{n}\sum_{i=1}^n g(X_i) \xrightarrow{P} \mathbb{E}[g(X_i)]$

(Also by LLN)

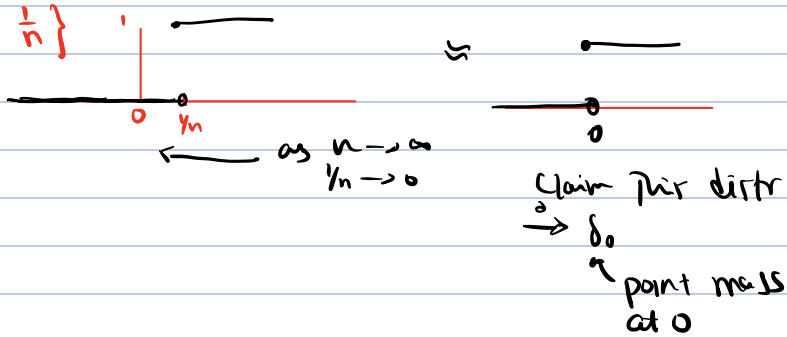
Central Limit Theorem

Before we define this, we need a notion of convergence that's called convergence in distribution

DEF: Let $X_1, X_2, \ldots$ be a collection of r.v.s each with cdf $F_n(t) = \mathbb{P}(X_n \leq t)$. We say "$X_n$ converges in distribution to X" (where X is a r.v.) if

$F_n(t) \to F(t) \equiv \mathbb{P}(X \leq t)$ as $n \to \infty$ for all t such that F is continuous at t

↳ e.g. If $F_n(t) = \mathbb{1}\{t \geq \frac{1}{n}\}$



THM: (Central Limit Theorem) If $X_1, \ldots, X_n$ are iid r.v.s with $\text{var}(X_i) < \infty$ then

$$\frac{\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right)}{\sqrt{\frac{\text{var}(X_i)}{n}}} \xrightarrow{d} N(0,1) \qquad n \to \infty$$

Implication: if "n is large" ($n \geq 30$) then $\frac{\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right)}{\sqrt{\frac{\text{var}(X_i)}{n}}} \approx N(0,1)$

used in hypothesis testing — z-test

Proof: Idea is that mgfs characterize distributions. This is true even in the limit

↳ **mgf method** (TTM) Suppose $X_1, X_2, \ldots$ have mgfs $M_n(t)$ ($M_i(t) = \mathbb{E}[e^{tX_i}]$)
we say $X_n \xrightarrow{d} X$ (where X has mgf $M(t)$) if the following hold
(1) $M_n(t)$, $M(t)$ exist and are finite for all t in a neighborhood of 0
(2) $M_n(t) \to M(t)$ $\forall$ t in a neighborhood of 0

" $X_n \overset{d}{=} X$ if $M_n(t) \overset{\cdot}{=} M(t)$ $\forall$ t in a nbd of 0 "

We will use mgf method and show that $M_{z_n}(t) \to M_z(t)$ $\Rightarrow z_n \xrightarrow{d} z$

$$z_n = \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right)}{\sqrt{\frac{\mathrm{var}(X_i)}{n}}}$$

$z \sim N(0,1)$

Calculate $M_{z_n}(t)$. We will assume that $\mathbb{E}[X_i] = 0$ and $\mathrm{var}(X_i) = 1$

$$z_n = \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right)}{\sqrt{\frac{\mathrm{var}(X_i)}{n}}} = \frac{\frac{1}{n}\sum_{i=1}^n X_i}{\sqrt{\frac{1}{n}}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$$

$$M_{z_n}(t) = M_{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i}(t) = \mathbb{E}\left[e^{\sum_{i=1}^n \frac{X_i}{\sqrt{n}}t}\right] \overset{\text{independence}}{=} \prod_{i=1}^n \mathbb{E}\left[e^{X_i\left(\frac{t}{\sqrt{n}}\right)}\right] = \prod_{i=1}^n M_{X_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= M_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n$$

(identically distributed)

$M_{z_n}(t) = M_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n \longrightarrow ?$ as $n \to \infty$?

Recall (from 2.5 weeks ago ...) — Taylor of e^{tX}

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[1 + tX + \frac{1}{2}(tX)^2 + \cdots\right]$$
$$= 1 + t\mathbb{E}[X] + \frac{1}{2}t^2\mathbb{E}[X^2] + \cdots$$
$$\Rightarrow \frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = \mathbb{E}[X^n]$$

$$M_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n = \mathbb{E}\left[e^{\frac{t}{\sqrt{n}}X_1}\right]^n \approx \mathbb{E}\left[1 + \frac{t}{\sqrt{n}}X_1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}X_1\right)^2\right]^n$$
$$= \left(1 + \frac{t}{\sqrt{n}}\mathbb{E}[X_1] + \frac{t^2}{2n}\mathbb{E}[X_1^2]\right)^n$$
$$= \left(1 + \frac{t}{\sqrt{n}}(0) + \frac{t^2}{2n}(1)\right)^n \quad \text{Assumption that } \mathbb{E}[X_1] = 0,\ \mathbb{E}[X_1^2] = \mathrm{var}(X_1) = 1$$
$$= \left(1 + \frac{t^2}{2n}\right)^n \longrightarrow e^{t^2/2} \quad (\text{L'Hopital's rule})$$

Done!

$\to M_Z(t)$

(because $M_{N(\mu,\sigma^2)}(t) = e^{\mu t + \sigma^2 t^2/2}$)

$\mu = 0 \quad \sigma^2 = 1$

We introduced $\xrightarrow{P}$ and $\xrightarrow{d}$

What is the relationship?

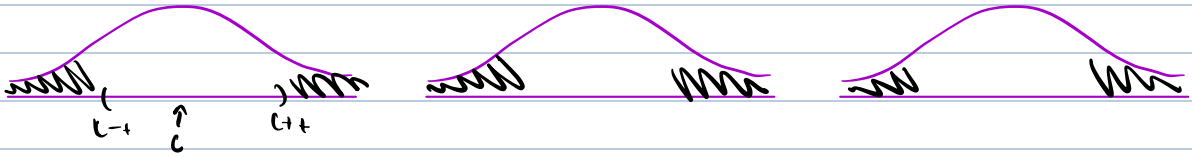
① $\xrightarrow{P}$ is stronger than $\xrightarrow{d}$ (i.e., $\xrightarrow{P} \Rightarrow \xrightarrow{d}$ but the reverse does not hold)

exercise

e.g., Suppose $X \sim N(0,1)$. Then $X \overset{d}{=} -X$

This implies that the sequence $X, -X, X, -X, \ldots$ converges in distribution to X. But this sequence does not converge in probability to anything

Distr of X | Distr of $-X$ | Limiting distr.



$$\mathbb{P}(|X - c| > \varepsilon) \not\to 0$$

② What is the implication of $\xrightarrow{d}$ in CLT?

(simulation)

$Z_n = \sqrt{n}(\bar{X}_n - 0.2) \xrightarrow{d} N(0, 0.2(0.8))$

(equivalent to $\frac{\sqrt{n}(\bar{X} - 0.2)}{\sqrt{0.2(0.8)}} \xrightarrow{d} N(0,1)$)

$\xrightarrow{d}$ means that $\mathbb{P}(\sqrt{n}(\bar{X}_n - 0.2) \le t) \to \mathbb{P}(N(\ ) \le t)$

Does not mean that

$\mathbb{P}(\bar{X} \le t) = \mathbb{P}(\sqrt{n}\bar{X} \le \sqrt{n}t) \xrightarrow{\text{"CLT"}} \mathbb{P}(Z \le \sqrt{n}t)$

Not true by CLT (Large deviation thm)

Misc. Topics

① Delta method: Suppose you know $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

Delta method: $\sqrt{n}(g(\bar{X}) - g(\mu)) \to N(0, g'(\mu)^2 \sigma^2)$

② Other prob. inequalities (Markov, Chebyshev, Chernoff : "tail inequalities")

- Cauchy Schwarz
- Jensen's inequality: $g(\mathbb{E}[X]) \underset{\geq}{\leq} \mathbb{E}[g(X)]$ if g convex / concave

e.g., $\mathbb{E}[\frac{1}{X}] \neq \frac{1}{\mathbb{E}[X]}$

$\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$

⋮

③ $\xrightarrow{d}$ also has a continuous mapping Thm

if $X_n \xrightarrow{d} X$

$g(X_n) \xrightarrow{d} g(X)$ g cont