Lec 22 Math 621 12/4/19

1

$X \sim \text{Exp}(1) \Rightarrow M_X(t) = \frac{1}{1-t} \mathbb{1}_{t<1}$

Chernoff

$\Rightarrow P(X \geq a) \leq \min_{t>0} \left\{ e^{-ta} \frac{1}{1-t} \text{ if } t<1 \right\} = \min_{t \in (0,1)} \Big\{ \underbrace{\frac{e^{-ta}}{1-t}}_{g(t)} \Big\} = \frac{e^{-\left(1-\frac{1}{a}\right)a}}{1-\left(1-\frac{1}{a}\right)} = a e^{-a+1} = \frac{ea}{e^a}$

Find min $g(t)$. Set $g'(t)=0$ and solve for $t(a)$.

$g'(t) = \frac{(1-t)(-a)e^{-ta} - e^{-ta}(-1)}{(1-t)^2} = \frac{a(t-1)e^{-ta} + e^{-ta}}{(1-t)^2} \overset{\text{set}}{=} 0 \Rightarrow (at-a+1)e^{-ta} = 0$

$\Rightarrow at - a + 1 = 0 \Rightarrow t^* = \frac{a-1}{a} = 1 - \frac{1}{a} \in (0,1)$ if $a>1$ so it is valid

---

Consider r.v.'s X, Y with finite $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$

let $W = (X - cY)^2$ for some constant $c \in \mathbb{R}$

Note that W is non-negative $\Rightarrow E(W) \geq 0$

$\Rightarrow E\left[(X-cY)^2\right] \geq 0$

$\Rightarrow E\left[X^2 - 2cXY + c^2Y^2\right] \geq 0$

$\Rightarrow E(X^2) - 2cE(XY) + c^2E(Y^2) \geq 0$

let $c = \frac{E(XY)}{E(Y^2)} \in \mathbb{R}$

$\Rightarrow E(X^2) - 2\frac{E(XY)}{E(Y^2)}E(XY) + \frac{E(XY)^2}{E(Y^2)^{\cancel{2}}}\cancel{E(Y^2)} \geq 0$

$\Rightarrow E(X^2)E(Y^2) - 2E(XY)^2 + E(XY)^2 \geq 0$

$\Rightarrow E(X^2)E(Y^2) - E(XY)^2 \geq 0$

$\Rightarrow E(XY)^2 \leq E(X^2)E(Y^2)$

$\Leftrightarrow |E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

Cauchy-Schwarz Inequality

If X,Y non-negative

$\Rightarrow E(XY) \leq \sqrt{E(X^2)E(Y^2)}$

Recall $Cov(X,Y) := E(XY) - \mu_X \mu_Y$ measures "linear dependence" between X,Y. Measured in units of "X times Y"

let $Corr(X,Y) := \frac{Cov(X,Y)}{SE(X)\,SE(Y)}$ measures same thing and it's unitless

WTS $Corr(X,Y) \in [-1,1]$ $\forall$ r.v.'s X,Y.

Intuition: if $Y = cX$ for $c \neq 0$ the following is true:

$$Corr[X,cX] = \frac{Cov(X,cX)}{SE(X)\,SE(cX)} = \frac{c\,Cov(X,X)}{SE(X)\,|c|\,SE(X)} = \frac{c}{|c|}\frac{\sigma_X^2}{\sigma_X\sigma_X} = \begin{cases} 1 & \text{if } c>0 \\ -1 & \text{if } c<0 \end{cases}$$

this is the most "extreme" linear dependence. So it makes sense that $\forall X,Y$ that $Corr(X,Y) \in [-1,1]$.

Pf: let $Z_X = \frac{X-\mu_X}{\sigma_X}$, $Z_Y = \frac{Y-\mu_Y}{\sigma_Y}$ $\Rightarrow E(Z_X) = E(Z_Y) = 0$, $SE(Z_X) = SE(Z_Y) = 1$ $\Rightarrow E(Z_X^2) = E(Z_Y^2) = 1$

by Cauchy-Schwarz...

$$|E[Z_X Z_Y]| \leq \sqrt{E(Z_X^2)\,E(Z_Y^2)} = 1 \Rightarrow E[Z_X Z_Y] \in [-1,1]$$

$$Corr(X,Y) = \frac{E(XY) - \mu_X\mu_Y}{\sigma_X\sigma_Y} = \frac{E\big[\overbrace{(\sigma_X Z_X + \mu_X)(\sigma_Y Z_Y + \mu_Y)}^{\sigma_X\sigma_Y Z_X Z_Y + \mu_X\sigma_Y Z_Y + \mu_Y\sigma_X Z_X + \mu_X\mu_Y}\big] - \mu_X\mu_Y}{\sigma_X\sigma_Y} = \frac{\cancel{\sigma_X\sigma_Y} E[Z_X Z_Y] + \cancel{\mu_X\mu_Y} - \cancel{\mu_X\mu_Y}}{\cancel{\sigma_X\sigma_Y}} \in [-1,1].$$

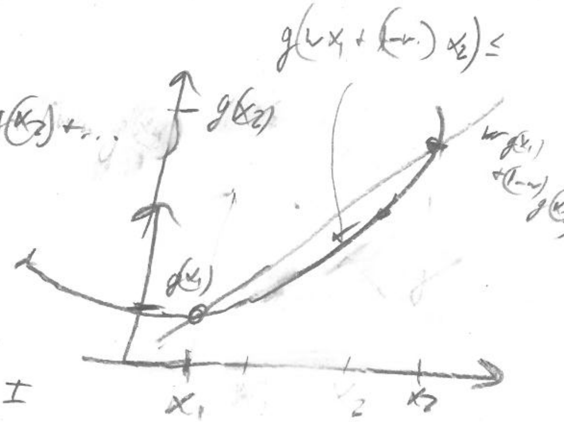
<u>Def</u>: g is a "convex function" on an interval $I \subset \mathbb{R}$

If $\forall \{x_1, x_2, \ldots\} \subset I$ and $\forall \{w_1, w_2, \ldots\}$ s.t. $\sum w_i = 1$, $\forall w_i \geq 0$, i.e. the w's are the "weights", that implies

$$g(w_1 x_1 + w_2 x_2 + \ldots) \leq w_1 g(x_1) + w_2 g(x_2) + \ldots$$

$$\Longleftrightarrow$$

$$g\left(\sum w_i x_i\right) \leq \sum w_i g(x_i)$$



<u>Thm</u>: if g is twice differentiable on I, ~~then g is convex on I~~

~~if~~ $g''(x) \geq 0 \quad \forall x \in I \Longrightarrow g$ is convex on I.

Consider a discrete r.v. X ~~...~~

and $p(x_1), \ldots, p(x_n) \Rightarrow E(X) = \sum_{x \in \mathrm{supp}\, X} x\, p(x)$. Consider $p(x)$ the "weights"

$$\Longrightarrow g(E(X)) \leq E(g(X))$$

Proof for X continuous is slightly more involved but it holds.

Note: if g is "concave" i.e. $g\left(\sum w_i x_i\right) \geq \sum w_i g(x_i)$

$$\Rightarrow g(E(X)) \geq E(g(X))$$

"Jensen's Inequalities"

e.g. From Math 241 we know $E[X^2] = \sigma^2 + \mu^2$

let $g(X) = X^2$

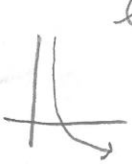
$$g(E(X)) \le E[g(X)] \Rightarrow \mu^2 \le E[X^2]$$

We can derive many inequalities with this one. Let's do one as an example.

let $a,b > 0$, $p,q > 0$ but $\frac{1}{p} + \frac{1}{q} = 1$

Consider $X \sim \begin{cases} a^p & wp\ \frac{1}{p} \\ b^q & wp\ \frac{1}{q} \end{cases} \Rightarrow E[X] = \frac{a^p}{p} + \frac{b^q}{q}$

let $g(t) = -\ln(t)$ convex! $\Rightarrow g(X) \sim \begin{cases} -p\ln(a) & wp\ \frac{1}{p} \\ -q\ln(b) & wp\ \frac{1}{q} \end{cases}$



$$\Rightarrow E[g(X)] = \frac{-p\ln(a)}{p} + \frac{-q\ln(b)}{q} = -\ln(ab)$$

Jensen's

$$\Rightarrow -\ln\left(\frac{a^p}{p} + \frac{b^q}{q}\right) \le -\ln(ab) \Rightarrow ab \le \frac{a^p}{p} + \frac{b^q}{q} \quad \text{"Young's Inequality"}$$

let $a = X$, $b = Y \Rightarrow XY \le \frac{X^p}{p} + \frac{Y^q}{q} \Rightarrow E(XY) \le \frac{E(X^p)}{p} + \frac{E(Y^q)}{q}$

let $a = \frac{X}{(E(X^p))^{\frac{1}{p}}}$, $b = \frac{Y}{E(Y^q)^{\frac{1}{q}}} \Rightarrow \frac{E[XY]}{E(X^p)^{\frac{1}{p}} E(Y^q)^{\frac{1}{q}}} \le \frac{\frac{\cancel{E(X^p)}}{\cancel{((E(X^p))^{\frac{1}{p}})^p}}}{p} + \frac{\frac{\cancel{E(Y^q)}}{\cancel{((E(Y^q))^{\frac{1}{q}})^q}}}{q} = \frac{1}{p} + \frac{1}{q} = 1$

$$\Rightarrow E(XY) \le E(X^p)^{\frac{1}{p}} E(Y^q)^{\frac{1}{q}} \quad \text{Holder's Inequality}$$

let $0<r<s$, $p=\frac{s}{r}$, $q=\frac{p}{p-1}=\frac{\frac{s}{r}}{\frac{s}{r}-1}=\frac{s}{s-r}$ $\Rightarrow \frac{1}{p}+\frac{1}{q}=\frac{r}{s}+\frac{s-r}{s}=1$

let $X=|Y|^r$, $Y=1$ both non-negative... By Holder's,

$$E[|Y|^r] \le \left(E[(|Y|^r)^{\frac{s}{r}}]\right)^{\frac{r}{s}} \cancel{E[1^{\frac{s}{s-r}}]^{\frac{s-r}{s}}} \to 1$$

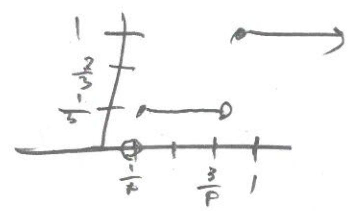
$\Rightarrow E[|Y|^r] \le E[|Y|^s]^{\frac{r}{s}}$ $\Rightarrow$ $E[|Y|^s]$ finite $\Rightarrow$ $E[|Y|^r]$ finite

this was a "taste" of inequalities. Lots more!

Convergence of r.v.'s

$X_n \xrightarrow{d} X$ means $\lim_{n\to\infty} F_n(x) = F(x)\ \forall x$ CDF convergence

Consider $X_n \sim \begin{cases} \frac{1}{n+1} & wp\ \frac{1}{3} \\ 1-\frac{1}{n+1} & wp\ \frac{2}{3} \end{cases}$ e.g. $X_3 \sim \begin{cases} \frac{1}{4} & wp\ \frac{1}{3} \\ \frac{3}{4} & wp\ \frac{2}{3} \end{cases}$



$X_n \to X \sim \begin{cases} 0 & wp\ \frac{1}{3} \\ 1 & wp\ \frac{2}{3} \end{cases}$ but does PMF convergence $\Rightarrow$ CDF converge?

Thm $Supp[X_n] \subseteq \mathbb{Z}$ and $Supp(X) \subseteq \mathbb{Z}$ then PMF convergence $\Leftrightarrow$ CDF convergence

Proof $\Leftarrow$

Note: $\forall x \in \mathbb{Z}$ $p_{X_n}(x) = F_{X_n}(x+\frac{1}{2}) - F_{X_n}(x-\frac{1}{2})$ why? there is no support outside of $\mathbb{Z}$
$\Rightarrow P(X \in (x, x+\frac{1}{2}]) = 0\ \forall x$

$$\lim_{n\to\infty} p_{X_n}(x) = \lim_{n\to\infty}\left(F_{X_n}(x+\tfrac{1}{2}) - F_{X_n}(x-\tfrac{1}{2})\right) = \lim_{n\to\infty} F_{X_n}(x+\tfrac{1}{2}) - \lim_{n\to\infty} F_{X_n}(x-\tfrac{1}{2})$$

$$= F_X(x+\tfrac{1}{2}) - F_X(x-\tfrac{1}{2}) = p_X(x)\ \checkmark$$

Proof $\Rightarrow$

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} P(X_n \le x) = \lim_{n\to\infty} \sum_{y=-\infty}^{x} p_{X_n}(x) \overset{\text{real analysis}}{=} \sum_{y=-\infty}^{x} \lim_{n\to\infty} p_{X_n}(x)$$

$$= \sum_{y=-\infty}^{x} p_X(x) = P(X \le x) = F_X(x)$$

~~try this then. Since $p_{X_n}(x) = p_X(x) = \mathrm{Bern}(\frac{2}{3}) \Rightarrow X_n \xrightarrow{d} \mathrm{Bern}(\frac{2}{3})$~~

Let $X_n \sim \mathrm{Binom}(n, \frac{\lambda}{n})$, $X_n \xrightarrow{d} \mathrm{Poisson}(\lambda)$

Consider $X_n \xrightarrow{d} c$ s.t. $c \in \mathbb{R}$. AKA $\mathrm{Deg}(c)$.

$$\forall x \lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 1 & \text{if } x \ge c \\ 0 & \text{if } x < c \end{cases} \quad \text{(HW)}.$$

Do we get the same with cont r.v.'s? PDF convergence $\Leftrightarrow$ CDF convergence?

No. Only PDF convergence $\Rightarrow$ CDF convergence.

For counterexamples, consider $X_n \sim U(0, \frac{1}{n}) = n\,\mathbb{1}_{x \in (0, \frac{1}{n})} = f_{X_n}(x)$

$$\lim f_{X_n}(x) = \begin{cases} \infty & x = 0 \\ 0 & \text{o/t} \end{cases} \quad \text{Not a PDF!}$$

But $X_n \xrightarrow{d} X \sim \mathrm{Deg}(0)$

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ nx & \text{if } x \in [0, \frac{1}{n}) \\ 1 & \text{if } x \ge \frac{1}{n} \end{cases} \xrightarrow{d} F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \ge 0 \end{cases}$$

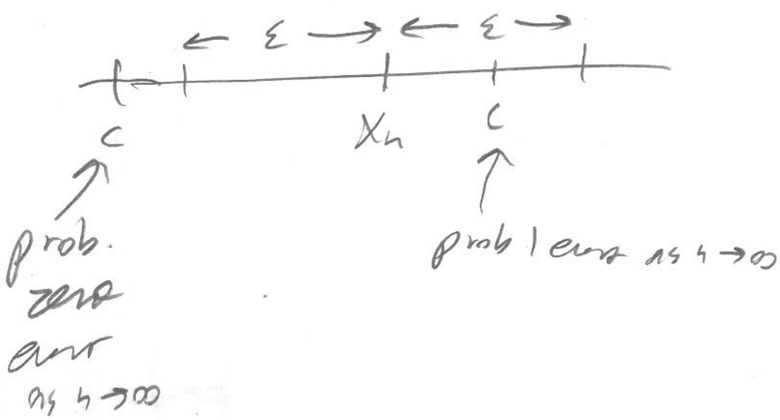
See HW for another

Convergence In Probability: $X_n \xrightarrow{P} c$ read

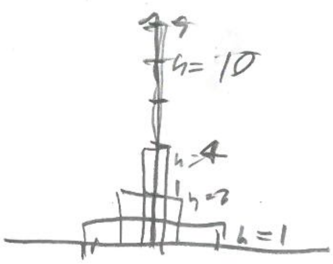
to a constant, we will ~~let conv to arbitrary~~

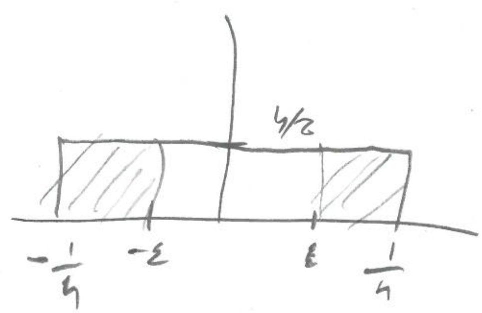
"X_n converges in prob to a constant c" if by definition

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left(|X_n - c| \geq \varepsilon\right) = 0 \iff \lim_{n\to\infty} P(|X_n - c| < \varepsilon) = 1$$



e.g. let $X_n \sim U\left(-\frac{1}{n}, \frac{1}{n}\right) = \frac{n}{2}\mathbb{1}_{x\in\left[-\frac{1}{n},\frac{1}{n}\right]}$





Prove $X_n \xrightarrow{P} 0$

$$\lim_{n\to\infty} P(|X_n - 0| \geq \varepsilon) = \lim P(|X_n| \geq \varepsilon) = \lim \left(P(X_n \leq -\varepsilon) + P(X_n \geq \varepsilon)\right) =$$

$$\lim\left(\left(-\varepsilon - -\frac{1}{n}\right)\frac{n}{2}\mathbb{1}_{-\varepsilon > -\frac{1}{n}} + \left(\frac{1}{n} - \varepsilon\right)\frac{n}{2}\mathbb{1}_{\varepsilon < \frac{1}{n}}\right) = \lim 2\left(\frac{1}{n} - \varepsilon\right)\frac{n}{2}\mathbb{1}_{\varepsilon < \frac{1}{n}}$$

$$= \lim (1 - n\varepsilon)\mathbb{1}_{\varepsilon < \frac{1}{n}} = 0$$

Pick ε, find n s.t. $\varepsilon \geq \frac{1}{n} \Rightarrow 0$