Wednesday December 2nd 2020

Lecture 22

Consider rv's X and Y with finite means and variance, $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ and let $W = (X - cY)^2$ where c is a real constant. Note: W is nonnegative

$$\Rightarrow E[W] \geq 0 \Rightarrow E[X^2 - 2cXY + c^2Y^2] \geq 0 \quad \text{choose } c = \frac{E[XY]}{E[Y^2]} \in \mathbb{R}$$

$$\Rightarrow E[X^2] - 2c\,E[XY] + c^2 E[Y^2] \geq 0 \Rightarrow E[X^2] - 2\frac{E[XY]}{E[Y^2]}E[XY] + \frac{[XY]^2}{E[Y^2]^2}E[Y^2] \geq 0$$

multiply by $E[Y^2]$ $\Rightarrow$ $E[X^2]E[Y^2] - 2E[XY]^2 + E[XY] \geq 0 \Rightarrow E[XY]^2 \leq E[X^2]E[Y^2]$

$$\Rightarrow |E[XY]| \leq \sqrt{E[X^2]E[Y^2]} \overset{\text{if } X,Y \text{ is nonnegative}}{\Longrightarrow} E[XY] \leq \sqrt{E[X^2]E[Y^2]}$$

These are relatively famous, they're called the Cauchy-Schwartz inequalities. We will use it to prove a basic fact useful in Statistics.

$$\text{Cov}[X,Y] = E[XY] - E[X]E[Y]$$

$$\text{Corr}[X,Y] = \frac{\text{Cov}[X,Y]}{SD[X]\,SD[Y]}$$ "This unitless metric called the "Correlation between X and Y".

Let $Z_X = \frac{X - \mu_X}{\sigma_X}$ and $Z_Y = \frac{Y - \mu_Y}{\sigma_Y} \Rightarrow E[Z_X] = E[Z_Y]$

$$SD[Z_X] = SD[Z_Y] = E[Z_X^2] = E[Z_Y^2]$$

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

$$\text{Corr}[X,Y] = \frac{E[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y} = \frac{E[(\sigma_X Z_X + \mu_X)(\sigma_Y Z_Y + \mu_Y)] - \mu_X\mu_Y}{\sigma_X\sigma_Y}$$

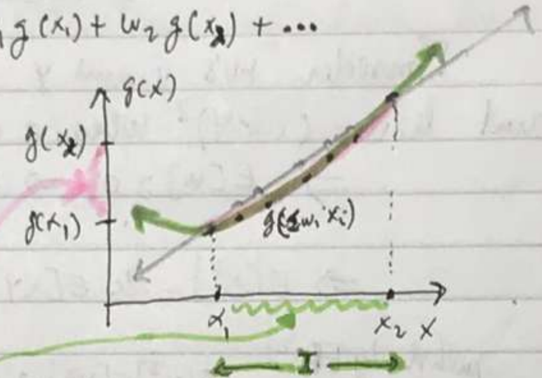
$$= \frac{\cancel{\sigma_X\sigma_Y}E[Z_X Z_Y] + \sigma_X\mu_Y \cancel{E[Z_X]}^{0} + \sigma_Y\mu_X \cancel{E[Z_Y]}^{0} + \cancel{\mu_X\mu_Y} - \cancel{\mu_X\mu_Y}}{\cancel{\sigma_X\sigma_Y}} = E[Z_X Z_Y] \in [-1,1]$$

Def: g is a "convex function" on an interval I (a subset of reals) if for all $x_1, x_2, \dots, \in I$ and all $w_1, w_2, \dots \in (0,1)$ s.t. $\sum w_i = 1$ AKA the "weights",

$$g(w_1 x_1 + w_2 x_2 + \dots) \le w_1 g(x_1) + w_2 g(x_2) + \dots$$

In sum notation,

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



$w_1 + w_2 = 1$

$\{x_1 w_1 + x_2 w_2 : w_1 + w_2 = 1, w_1 > 0, w_2 > 0\}$

$\{w_1 g(x_1) + w_2 g(x_2) : \dots\dots\}$

Let g be convex function and X be a discrete r.v. If discrete, we know $Supp[X] = \{x_1, x_2, \dots\}$ and sum $p(x_i) = 1$ the PMF.
Thus, we can call the PMF values, the weights i.e $W_i := p(x_i)$.

$$E[X] = \sum x_i p(x_i) = \sum w_i x_i$$

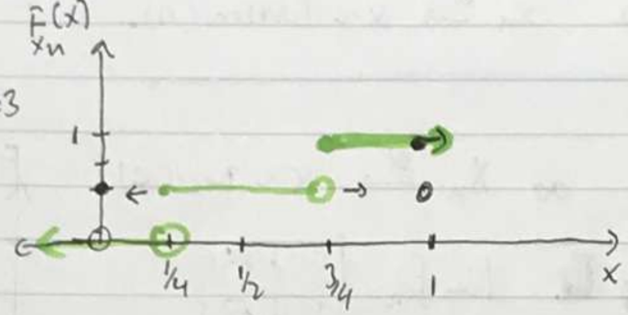
$$g(E[X]) \le \sum w_i g(x_i) = \sum g(x_i) p(x_i) = E[g(x)]$$ Jensen's Inequality.

Convergence of r.v's. We will study three different types.

First, let's review "Convergence in distribution". We say a sequence of r.v's $X_1, X_2 \dots$ denote X_n converges in distribution to X denote $X_n \xrightarrow{d} X$ means by definition that the limiting CDF is X's CDF:

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x) \quad \forall x$$

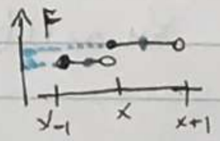
Conside $X_n \sim \begin{cases} \frac{1}{n+1} & w.p\ \frac{1}{3} \\ 1-\frac{1}{n+1} & w.p\ \frac{2}{3} \end{cases}$ e.g $X_3 \sim \begin{cases} \frac{1}{4} & w.p\ \frac{1}{3} \\ \frac{3}{4} & w.p\ \frac{2}{3} \end{cases}$



$\Rightarrow X \sim \begin{cases} 0 & w.p\ \frac{1}{3} \\ 1 & w.p\ \frac{2}{3} \end{cases}$

Conjecture: PMF Convergence and CDF Converence are equivalent. This is not true in general. But here's a situation where it is true: If Supp $[X_n]$ be a subset of $\mathbb{Z}$, the integers and let Supp $[X]$ also be a subset of $\mathbb{Z}$, the integers. Let's prove it.

Pf: CDF Convergence implies PMF convergence:



$$P_{X_n}(x) = F_{X_n}(x+\tfrac{1}{2}) - F_{X_n}(x-\tfrac{1}{2})$$

$$\lim P_{X_n}(x) = \lim F_{X_n}(x+\tfrac{1}{2}) - \lim F_{X_n}(x-\tfrac{1}{2}) = F_X(x+\tfrac{1}{2}) - F_X(x-\tfrac{1}{2}) = P_X(x)$$

Pf: PMF Convergence implies CDF Convergence:

$$F_{X_n}(x) := P(X_n \le x) = \sum_{y=-\infty}^{x} P_{X_n}(y)$$

$$\lim F_{X_n}(x) = \lim \sum_{y=-\infty}^{x} P_{X_n}(y) = \sum_{y=-\infty} \lim P_{X_n}(y) = \sum_{y=-\infty}^{x} P_X(y) = P(X \le x) = F_X(x)$$

How about for continuous rv's? Is PDF Convergence equivalent to CDF Convergence? Not always. PDF Convergence always implies CDF Convergence but not vice versa. Here's a Counter example:

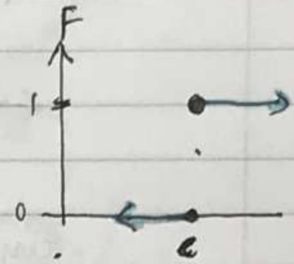
$$X_n \sim \overset{\text{uniform}}{U}\left(-\tfrac{1}{n}, \tfrac{1}{n}\right) \overset{\text{PDF}}{=} \frac{n}{2} \mathbb{1}_x\left[-\tfrac{1}{n}, \tfrac{1}{n}\right] = f_{X_n}(x)$$

$\lim f_{X_n}(x) = \infty$! Not a PDF! HW: $X_n \xrightarrow{d} 0$

HW $X_n \sim \text{Bin}(n, \frac{\lambda}{n}),\ \lambda > 0 \quad X_n \xrightarrow{d} X \sim \text{Poisson}(\lambda)$.

Define $X_n \xrightarrow{d} c,\ c \in \mathbb{R}$ as $X_n \xrightarrow{d} X \sim \text{Deg}(c)$

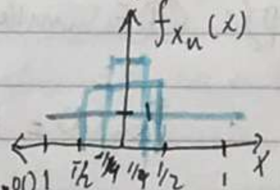
$$\lim F_{X_n} = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \geq 1 \end{cases}$$



Convergence in Probability to a Constant. For a sequence of rv's $X_1, X_2, \ldots$ denoted X_n, X_n converges in probability to a constant c, $X_n \xrightarrow{P} c$ is defined to be:

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

$$X_n \sim U(-\tfrac{1}{n}, \tfrac{1}{n}) = \frac{n}{2} \mathbb{1}_{x \in [-\frac{1}{n}, \frac{1}{n}]}$$



$\varepsilon = 0.0001 \quad n = 100 \quad X_n \sim U(-.01, .01)$

$$P(|X_n - 0| \leq 0.0001)$$
$$= P(X_n \in [-0.0001, 0.0001])$$
$$= \frac{2}{100} \cdot \frac{2}{1000} \neq 1$$

$\uparrow \quad n = 1000 \quad X_n \sim U(-.0001, .0001)$

$$P(X_n \in [-.0001, .0001]) = 1$$