We'll continue with two more inequalities. Consider rv's X and Y with finite means and variances, $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ and let W = (X - cY)^2 for some constant c. Note that W is non-negative by construction.

$$\Rightarrow E[W] \ge 0 \Rightarrow E[X^2 - 2cXY + c^2Y^2] \ge 0$$

$$\Rightarrow E[X^2] - 2c\,E[XY] + c^2 E[Y^2] \ge 0 \qquad \text{let } c = \frac{E[XY]}{E[Y^2]} \in \mathbb{R}$$

$$\Rightarrow E[X^2] - 2\frac{E[XY]}{E[Y^2]}E[XY] + \frac{E[XY]^2}{E[Y^2]^{\cancel{2}}}\cancel{E[Y^2]} \ge 0$$

multiply by $E[Y^2]$

$$\Longrightarrow E[X^2]E[Y^2] - \cancel{2}E[XY]^2 + \cancel{E[XY]^2} \ge 0$$

$$\Rightarrow E[X^2]E[Y^2] - E[XY]^2 \ge 0$$

$$\Rightarrow E[XY]^2 \le E[X^2]E[Y^2] \Rightarrow |E[XY]| \le \sqrt{E[X^2]E[Y^2]}$$

if X, Y non-neg. $\Longrightarrow E[XY] \le \sqrt{E[X^2]E[Y^2]}$

these are called the "Cauchy-Schwartz Inequalities" and they're famous.

Recall $Cov[X,Y] := E[XY] - E[X]E[Y]$

$$Corr[X,Y] := \frac{Cov[X,Y]}{SD[X]SD[Y]} \in [-1,1]$$

Correlation of X and Y, a unitless metric. We now prove that its range is always -1 to 1:

let $Z_X = \frac{X-\mu_X}{\sigma_X}$ and $Z_Y = \frac{Y-\mu_Y}{\sigma_Y} \Rightarrow E[Z_X] = E[Z_Y] = 0$, $SD[Z_X] = SD[Z_Y] = 1 \Rightarrow E[Z_X^2] = E[Z_Y^2] = 1$

Cauchy-Schwarz

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

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

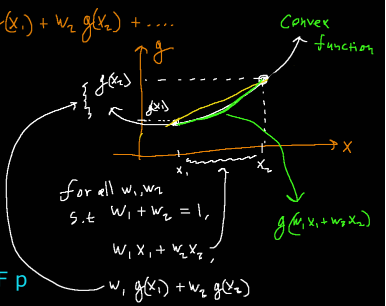
$$= \frac{\cancel{\sigma_X}\cancel{\sigma_Y}E[Z_X Z_Y] + \mu_X\sigma_Y \cancelto{0}{E[Z_Y]} + \mu_Y\sigma_X \cancelto{0}{E[Z_X]} + \cancel{\mu_X\mu_Y} - \cancel{\mu_X\mu_Y}}{\cancel{\sigma_X}\cancel{\sigma_Y}}$$

$$= E[Z_X Z_Y] \in [-1,1]$$

Definition: g is a "convex function" on and interval I, a connected subset of the reals if for all $\{x_1, x_2, \dots\} \subset I$ and for all $w_1, w_2, \dots \in (0,1)$ s.t. the sum of the w_i's = 1,

$$g(w_1x_1 + w_2x_2 + \dots) \le w_1 g(x_1) + w_2 g(x_2) + \dots$$

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



Thm: if g is twice differentiable and g''(x) >= 0 for all x in I, then g is convex on I.

Consider a discrete rv X with PMF p

$$E[X] = \sum_{x \in Supp[X]} p(x)x \quad \text{and} \quad Supp[X] = \{x_1, x_2, \dots\}, \text{ let } w_i := p(x_i) \Rightarrow \sum w_i = 1$$

and a convex function g, then just using the definition of convexity, we get the following inequality:

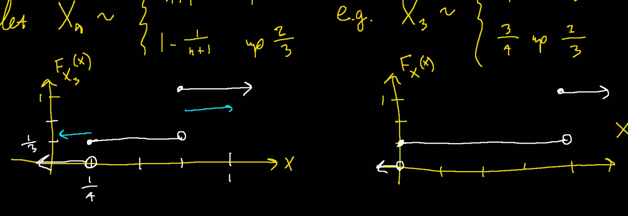
$$E[g(X)] = \sum \underbrace{p(x_i)}_{w_i} g(x_i) \Rightarrow g(E[X]) \le E[g(X)]$$

"Jensen's Inequality"

Types of convergence of random variables. We begin with reviewing "convergence in distribution". Consider a sequence of rv's X_1, X_2, ... denoted X_n:

$X_n \xrightarrow{d} X$ is defined as $\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \;\forall x$

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



$X \sim \begin{cases} 0 & \text{wp } \frac{1}{3} \\ 1 & \text{wp } \frac{2}{3} \end{cases}$

It seems plausible that PMF convergence and CDF convergence are equivalent. Thm: $Supp[X_n] \subset \mathbb{Z}$ and $Supp[X] \subset \mathbb{Z}$ then they are equivalent.

Pf: CDF convergence => PMF convergence (for discrete sequences)

$$p_{X_n}(x) = F_{X_n}\left(x + \tfrac{1}{2}\right) - F_{X_n}\left(x - \tfrac{1}{2}\right) \quad \forall x \in \mathbb{Z}$$

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

Pf: PMF convergence => CDF convergence

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

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

$X_n \sim Binom(n, \frac{\lambda}{n})$ Show $X_n \xrightarrow{d} X \sim Poisson(\lambda)$ HW

let $c \in \mathbb{R}$ and let $X_n \xrightarrow{d} c$ be defined as $X_n \xrightarrow{d} X \sim Deg(c)$

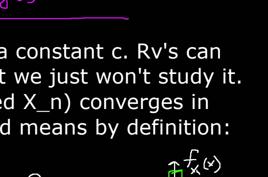
which means by definition of conv. in distr, $\forall x \lim F_{X_n}(x) = \begin{cases} 1 & \text{if } x \ge c \\ 0 & \text{o/w} \end{cases}$

For continuous rv's is PDF convergence equivalent to CDF convergence unconditionally? No. PDF convergence => CDF converg. Here's a counterexample to the other direction:

$$X_n \sim U\left(0, \tfrac{1}{n}\right) = n\mathbb{1}_{x \in [0, \frac{1}{n}]}$$

$\lim f_{X_n}(x) = \infty$ no PDF

HW show $\lim F_{X_n}(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0 \end{cases} \Rightarrow X \sim Deg(0)$

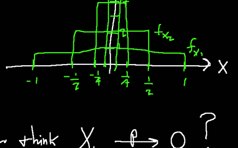


Convergence in Probability. But only to a constant c. Rv's can converge in probability to other rv's, but we just won't study it. A sequence of rv's X_1, X_2, ... (denoted X_n) converges in probability to c is denoted $X_n \xrightarrow{p} c$ and means by definition:

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(|X_n - c| \ge \varepsilon) = 0$$

equivalently,

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



Consider $X_n \sim U\left(-\frac{1}{n}, \frac{1}{n}\right) = \frac{n}{2}\mathbb{1}_{x \in [-\frac{1}{n}, \frac{1}{n}]}$. Do you think $X_n \xrightarrow{p} 0$?