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 where c is a real constant. Note: W is nonnegative.

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

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

$$\xrightarrow[\text{by } E[Y^2]]{\text{multiply}} E[X^2]E[Y^2] - \cancel{2}E[XY]^2 + \cancel{E[XY]} \ge 0 \Rightarrow E[XY]^2 \le E[X^2]E[Y^2]$$

$$\Rightarrow |E[XY]| \le \sqrt{E[X^2]E[Y^2]} \xrightarrow[\text{non-neg}]{\text{if } X,Y} E[XY] \le \sqrt{E[X^2]E[Y^2]}$$

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

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

$$Corr[X,Y] := \frac{Cov[X,Y]}{SD[X]\,SD[Y]}$$ this unitless metric is called the "correlation between X and Y"

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

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

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

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

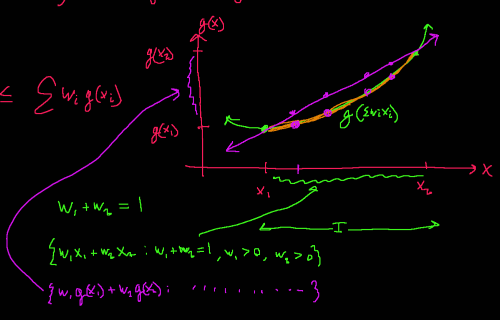
Def: g is a "convex function" on an interval I (a subset of reals) if for all $x_1, x_2, \ldots \in I$ and all $w_1, w_2, \ldots \in (0,1)$ s.t. $\sum w_i = 1$ AKA the "weights",

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

In sum notation,

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

Thm: g is convex on I if $\forall x \in I\ g''(x) > 0$.



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

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

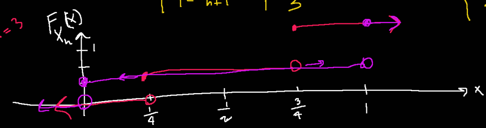
$$g(E[X]) \le \sum w_i g(x_i) = \sum g(x_i)\,p(x_i) = E[g(X)]$$ Jensen's Inequality

Convergence of rv's. We will study three different types.

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

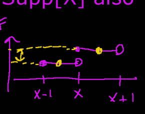
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)\ \forall x$$

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



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

Conjecture: PMF convergence and CDF convergence 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 Z, the integers and let Supp[X] also be a subset of Z, the integers. Let's prove it.



Pf: CDF convergence implies PMF convergence:

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

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

Pf: PMF convergence implies CDF convergence:

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

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

HW $X_n \sim Bin\left(n, \frac{\lambda}{n}\right)$, $\lambda > 0$ Prove: $X_n \xrightarrow{d} X \sim Poisson(\lambda)$.

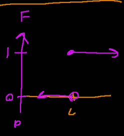
How about for continuous rv's? Is PDF convergence equivalent to CDF convergence? Not always. PDF convergence always implies CDF convergence but not vice versa. Here's a counterexample:

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

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

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

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

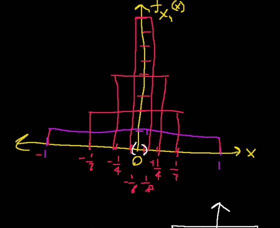


Convergence in probability to a constant.

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

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

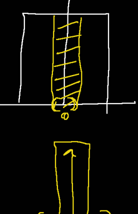
$$X_n \sim U\left(-\tfrac{1}{n}, \tfrac{1}{n}\right) = \frac{n}{2}\mathbb{1}_{x \in \left[-\frac{1}{n}, \frac{1}{n}\right]}$$



$\varepsilon = 0.0001$

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

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



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

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