• Let X be any rv. $Y = e^{tX}$. Note: Y is non-neg.

$$\xrightarrow{\text{Markov}} P(Y \ge c) \le \frac{E[Y]}{c}$$

$$\Rightarrow P(e^{tX} \ge c) \le \frac{E[e^{tX}]}{c} = \frac{M_X(t)}{c}$$

let $c = e^{ta}$

$$\Rightarrow P(e^{tX} \ge e^{ta}) \le \frac{M_X(t)}{e^{ta}}$$

$$\Rightarrow P(tX \ge ta) \le e^{-ta} M_X(t)$$

$$\xrightarrow{t>0} P(X \ge a) \le e^{-ta} M_X(t)$$ valid for all t > 0

because I'm looking for the tightest possible tail bound, I want to minimize the rhs

$$\Rightarrow P(X \ge a) \le \min_{t>0} \{ e^{-ta} M_X(t) \}$$

$$\xrightarrow{t<0} P(X \le a) \le e^{-ta} M_X(t)$$

$$\Rightarrow P(X \le a) \le \min_{t<0} \{ e^{-ta} M_X(t) \}$$

Chernoff's Inequality

Let's derive the Chernoff Bound for X ~ Exp(λ).

$$M_X(t) := E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx$$

$$\lim_{x\to\infty} \frac{1}{e^x} = 0$$

$$= \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)x}\right]_0^\infty = \frac{\lambda}{t-\lambda}\begin{cases} \infty & \text{if } t > \lambda \\ 0 & \text{if } t = \lambda \\ -1 & \text{if } t < \lambda \end{cases} = \frac{\lambda}{\lambda - t} \text{ if } t < \lambda$$

otherwise the MGF doesn't exist

Let λ = 1 and derive the actual bound.

$$P(X \ge a) \le \min_{t>0} \left\{ e^{-ta} \frac{1}{1-t} \text{ if } t<1 \right\} = \min_{t\in(0,1)} \left\{ \frac{\overset{u}{e^{-ta}}}{\underset{v}{1-t}} \right\} \quad // g$$

$$g'(t) = \frac{vu' - uv'}{v^2} = \frac{(1-t)(a e^{-ta}) - e^{-ta}(-1)}{(1-t)^2} = \frac{a(t-1)e^{-ta} + e^{-ta}}{(1-t)^2}$$

$$= \frac{e^{-ta}(a(t-1)+1)}{(1-t)^2} \overset{\text{set}}{=} 0 \Rightarrow \cancel{e^{-ta}}(at - a + 1) = 0$$

$$\Rightarrow at = a - 1 \Rightarrow t_* = \frac{a-1}{a} = 1 - \frac{1}{a} \in (0,1) \text{ if } a > 1$$

$$\Rightarrow P(X \ge a) \le \frac{e^{-(1-\frac{1}{a})a}}{1-(1-\frac{1}{a})} = \frac{e^{-a+1}}{\frac{1}{a}} = \boxed{\frac{ae}{e^a}}$$

Let's compare Markov's, Chebyshev's and Chernoff's bounds for X ~ Exp(1) for different values of a:

a	$P(X \ge a)$	Markov	Chebyshev	Chernoff
2	.1353	0.5	1	0.73576
5	.0067	0.2	0.0635	0.09158
10	.00004	0.1	0.0123	0.00123
⋮				

Consider two rv's X, Y with finite means and variances: $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$

Let W := (X - cY)² for some constant c ∈ ℝ. Note that W is non-negative and thus E[W] ≥ 0.

$$\Rightarrow E[(X-cY)^2] \ge 0$$

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

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

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

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

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

$$\Rightarrow E[XY]^2 \le E[X^2]E[Y^2]$$

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

$$\xRightarrow{\text{If } X,Y \text{ non-neg}} E[XY] \le \sqrt{E[X^2]E[Y^2]}$$

• Recall that Cov[X, Y] := E[XY] - $\mu_X\mu_Y$ has units of X times units Y

Define: Corr[X, Y] := $\frac{Cov[X,Y]}{SD[X]\,SD[Y]}$ is unitless

But is it normalized? Are correlations comparable across different pairs of rv's?

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

$$X = \mu_X + \sigma_X Z_X \qquad Y = \mu_Y + \sigma_Y Z_Y \qquad \Downarrow\ E[Z_X^2] = E[Z_Y^2] = 1$$

Let's use Cauchy-Schwartz on the two Z's:

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

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

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

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

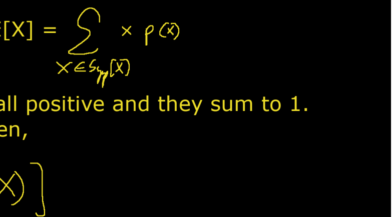
The correlation of any two rv's is between -1 and 1. This means correlation is a standard metric of comparison.

Def: a function g is "convex" on an interval I ⊂ ℝ if for all $\{x_1, x_2, \ldots\} \subset I$ and $\{w_1, w_2, \ldots\}$ s.t. $\Sigma w_i = 1$ and all $w_i > 0$ (i.e. the "weights") then

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

or more compactly

$g(\Sigma w_i x_i) \le \Sigma w_i g(x_i)$



Thm: if g is twice differentiable on I and g''(x) ≥ 0 for all x ∈ I then g is convex on I.

Consider a discrete rv X thus E[X] = $\sum_{x \in Supp[X]} x\, p(x)$

Let p(x) be the w_i's. They're all positive and they sum to 1. Let g be a convex function. Then,

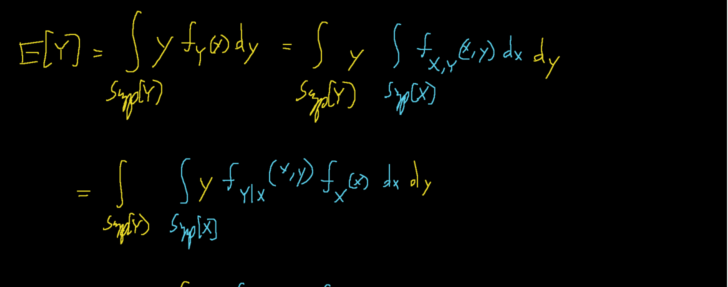
$$g(E[X]) \le E[g(X)]$$

This inequality is also valid for continuous rv's but proof is more involved. Also, if the function is "concave" which means you just flip the ≤ to a ≥ in the definition of convex and the assoicated thm, then,

$$g(E[X]) \ge E[g(X)]$$

Either of these is called "Jensen's Inequality".

Consider $\overset{\text{Cont.}}{\text{rv's}}$ X, Y and a joint density $f_{X,Y}(x,y)$



$$E[Y] = \int_{Supp[Y]} y\, f_Y(y)\, dy = \int_{Supp[Y]} y \int_{Supp[X]} f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{Supp[Y]} \int_{Supp[X]} y\, f_{Y|X}(x,y)\, f_X(x)\, dx\, dy$$

$$= \int_{Supp[X]} \int_{Supp[Y]} y\, f_{Y|X}(x,y)\, f_X(x)\, dy\, dx$$

$$= \int_{Supp[X]} f_X(x) \underbrace{\int_{Supp[Y]} y\, f_{Y|X}(x,y)\, dy}\, dx$$

$$= \int_{Supp[X]} f_X(x)\, E_Y[Y|X=x]\, dx$$

$$= E_X[E_Y[Y|X=x]] = E_Y[Y]$$

The law of iterated expectation.