$\phi_{\vec{X}}(\vec{t}) := E\left[e^{i\vec{t}^T\vec{X}}\right]$ for any vector rv X of dimension n.

Consider: $\phi_{\vec{X}}\left(\begin{bmatrix} t \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) = E\left[e^{i[t\ 0 \cdots 0]\vec{X}}\right] = E\left[e^{itX_1}\right] = \phi_{X_1}(t)$

$$\stackrel{P1/P8}{\Longrightarrow} X_1 \sim f_{X_1}(x)$$

$$f_{X_1}(x) = \int \cdots \int f_{X_1, X_2, \ldots X_n}(x, u_1, u_2, \ldots, u_{n-1})\, du_1 \ldots du_{n-1}$$

---

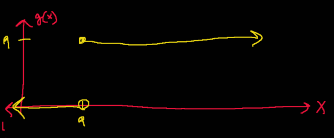
e.g. $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$, $X_1 \sim f_{X_1}(x) = ?$

$$\phi_{\vec{X}}\left(\begin{bmatrix} t \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) = e^{i[t\ 0 \cdots 0]\vec{\mu} - \frac{1}{2}[t\ 0 \cdots 0]\Sigma\begin{bmatrix} t \\ 0 \\ \vdots \\ 0 \end{bmatrix}} = e^{it\mu_1 - \frac{t}{2}[t\ 0 \cdots 0]\begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \vdots \\ \sigma_{1n} \end{bmatrix}}$$

$$= e^{it\mu_1 - \frac{t^2\sigma_1^2}{2}} = \phi_{X_1}(t) \stackrel{P1}{\Longrightarrow} X_1 \sim N(\mu_1, \sigma_1^2)$$

Assume X is a rv with nonnegative support i.e. Supp[X] >= 0 and has finite expectation. Let a > 0, a constant. Consider the following function:

$g(x) = a\mathbb{1}_{x \geq a}$



Is $a\mathbb{1}_{x \geq a} \leq x$? Yes! Two cases:

- if $x < a$ $\quad a\mathbb{1}_{x \geq a} = a(0) = 0 \leq x$ because $Supp[X] \geq 0$.
- if $x \geq a$ $\quad a\mathbb{1}_{x \geq a} = a(1) = a \leq x$ true by case assumption.

$$\Rightarrow a\mathbb{1}_{X \geq a} \leq X$$

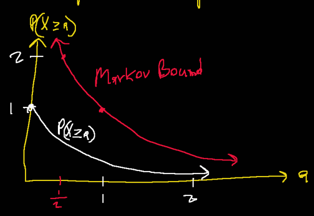
Let's take the expectation of both sides:

$$E[a\mathbb{1}_{X \geq a}] \leq E[X] \qquad \mathbb{1}_{X \geq a} \sim \begin{cases} 1 & \text{w.p. } P(X \geq a) \\ 0 & \text{o/w} \end{cases} = \text{Bern}(P(X \geq a))$$

$$\Rightarrow a\,E[\mathbb{1}_{X \geq a}] \leq \mu \Rightarrow a\,P(X \geq a) \leq \mu$$

$\Rightarrow P(X \geq a) \leq \frac{\mu}{a}$ this is called "Markov's Inequality" and it's very famous AKA: Markov Tail Bound.

For example $X \sim Exp(1) = e^{-x} \Rightarrow P(X \geq a) = 1 - F_X(a) = e^{-a} \Rightarrow \mu = 1$



We see here is that the Markov Bound is very "crude" meaning very approximate, much bigger than the truth.

a	$P(X \geq a)$	Markov Bound	Chebyshev Bound	Chernoff Bound
2	0.1353	0.5	1	0.73576
5	0.0067	0.2	0.0635	0.09158
10	0.00004	0.1	0.0123	0.00123

We will now prove many, many corollaries of the Markov Inequality:

- let $b = a\mu$ $\quad P(X \geq b) \leq \frac{\mu}{b} \Rightarrow P(X \geq a\mu) \leq \frac{1}{a}$

- Let h be a monotonically increasing function, $Y = h(X)$

$$P(Y \geq h(a)) \leq \frac{E[Y]}{h(a)} \Rightarrow P(h(X) \geq h(a)) \leq \frac{E[h(X)]}{h(a)} \Rightarrow P(X \geq a) \leq \frac{E[h(X)]}{h(a)}$$

- Let X be continuous in addition to nonnegative.

let $a = \text{Quantile}[X, p] \stackrel{\cdot}{=} F_X^{-1}(p)$

$$P(X \geq F_X^{-1}(p)) \leq \frac{\mu}{F_X^{-1}(p)} \Rightarrow 1 - \cancel{F_X}(\cancel{F_X^{-1}}(p)) \leq \frac{\mu}{F_X^{-1}(p)}$$

$$\Rightarrow 1 - p \leq \frac{\mu}{F_X^{-1}(p)} \Rightarrow \underbrace{F_X^{-1}(p)}_{Q[X,p]} \leq \frac{\mu}{1-p} \qquad \text{e.g. } Med[X] \leq 2\mu$$

- Let X be any rv => |X| is a nonnegative rv. $P(|X| \geq a) \leq \frac{E[|X|]}{a}$.

- Let X be any rv with finite variance, σ^2. Let Y = (X - mu)^2 => Y is a nonnegative rv.

$$P(Y \geq b) \leq \frac{E[Y]}{b} \Rightarrow P((X-\mu)^2 \geq b) \leq \frac{E[(X-\mu)^2]}{b}$$

← definition of variance

$$\Rightarrow P((X-\mu)^2 \geq b) \leq \frac{\sigma^2}{b} \stackrel{\text{let } b = a^2}{\Longrightarrow} P((X-\mu)^2 \geq a^2) \leq \frac{\sigma^2}{a^2}$$

$\Rightarrow P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$ This is "Chebyshev's Inequality" and it's also very famous.

Let's manipulate this to get it into a more "user-friendly" form. Assume X is nonnegative:

$$P(|X-\mu| \geq a) = P(X - \mu \geq a \cup -(X-\mu) \geq a) = P(X - \mu \geq a) + P(-(X-\mu) \geq a)$$

disjoint events

$$= P(X \geq \mu + a) + P(X \leq \mu - a) \stackrel{a \geq \mu}{=} P(X \geq \mu + a) + \cancelto{0}{P(X \leq \text{negative \#})}$$

$$\stackrel{\text{let } b = \mu + a}{=} P(X \geq b) \leq \frac{\sigma^2}{(b-\mu)^2}$$

- Let X be an rv. Let Y = e^{tX} => Y is a nonnegative rv for all t.

$$P(Y \geq b) \leq \frac{E[Y]}{b} \Rightarrow P(e^{tX} \geq b) \leq \frac{E[e^{tX}]}{b}$$

← the moment generating function for X, M_X(t)

$$\Rightarrow P(e^{tX} \geq b) \leq \frac{M_X(t)}{b} \stackrel{\text{let } b = e^{ta}}{\Longrightarrow} P(e^{tX} \geq e^{ta}) \leq e^{-ta}M_X(t)$$

$$\Rightarrow P(tX \geq ta) \leq e^{-ta}M_X(t)$$

$\Downarrow$ if $t > 0$ $\qquad\qquad$ $\Downarrow$ if $t < 0$

$P(X \geq a) \leq e^{-ta}M_X(t) \qquad P(X \leq a) \leq e^{-ta}M_X(t)$

$\Downarrow \qquad\qquad\qquad \Downarrow$

$$P(X \geq a) \leq \min_{t>0}\{e^{-ta}M_X(t)\} \qquad P(X \leq a) \leq \min_{t<0}\{e^{-ta}M_X(t)\}$$

If these inequalities are valid for all t, why not choose the "best" t to get the "sharpest" (lowest) bound?

This is called "Chernoff's Inequality".

Let's calculate it for X ~ Exp(lambda). Warning: it's a lot of work...

We first need to find the mgf for the exponential rv:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,dx = \lambda\int_0^\infty e^{(t-\lambda)x}\,dx = \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)x}\right]_0^\infty$$

$$= \frac{\lambda}{t-\lambda}\begin{cases} \infty - 1 & \text{if } t > \lambda \\ 0 - 1 & \text{if } t < \lambda \end{cases} = \frac{\lambda}{\lambda - t} \text{ only for } t < \lambda.$$

If t > lambda, the mgf doesn't exist. This is why you shouldn't be using them! Chf's always exist! So they're awesome...

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

$$P(X > a) \leq \min_{t>0}\left\{e^{-ta}\frac{1}{1-t}\right\} \text{ for } t < 1$$

$$\Rightarrow P(X > a) \leq \min_{t \in (0,1)}\left\{\underbrace{e^{-ta}\frac{1}{1-t}}_{h(t)}\right\} = e^{-(1-\frac{1}{a})a}\frac{1}{1-(1-\frac{1}{a})} = \frac{e^{-a}e}{\frac{1}{a}} = \frac{ae}{e^a}$$

$h(t_*)$

$$h'(t) = \frac{(1-t)(-a)e^{-ta} - e^{-ta}(-1)}{(1-t)^2} = \frac{(t-1)a\,e^{-ta} + e^{-ta}}{(1-t)^2}$$

$$= \frac{e^{-ta}(ta - a + 1)}{(1-t)^2} \stackrel{\text{set}}{=} 0 \Rightarrow ta - a + 1 = 0 \Rightarrow t_* = \frac{a-1}{a} = 1 - \frac{1}{a}$$

But... the Chernoff is sometimes useless. Why? Because it requires the mgf. To get the mgf, you need to kniw the PDF or PMF. If I know the PDF or PMF, then I know analytically or can numerically compute the CDF which means I know the tail exactly or within small numerical error! So it really is only useful if you're in a situation where you only have the MGF and not the PDF/PMF.