11/25/20 Michael Velez

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

Consider: $\phi_{\vec{x}}\left(\begin{bmatrix} t \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) = E\left[e^{i[t\ 0 \dots, 0]\vec{x}}\right]$

$= E[e^{itX_1}] = \phi_{X_1}(t) \overset{P_1/P_8}{\Rightarrow} X_1 \sim f_{X_1}(x)$

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

[illegible] only special cases

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

$\phi_X\left(\begin{bmatrix} t \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) = e^{i[t\ 0 \dots 0]\vec{\mu} - \frac{1}{2}[t\ 0 \dots 0]\Sigma \begin{bmatrix} t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}$

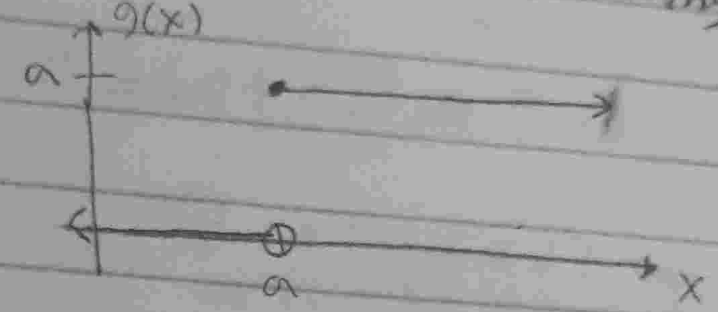
$= e^{it\mu_1 - \frac{t}{2}[t\ 0 \dots 0]\begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \vdots \\ \sigma_{1n} \end{bmatrix}}$

chf of univariate $N$

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

Assume $X$ is a rv with non-negative support $\mathrm{Supp}[X] \geq 0$ 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 $\mathrm{Supp}[X] \geq 0$

- 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:

$$E[a \mathbb{1}_{X \geq a}] \leq E[X]$$

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

$$\mathbb{1}_{X \geq a} \sim \begin{cases} 1 & \text{w.p. } P(X \geq a) \\ 0 & \text{o.w.} \end{cases}$$

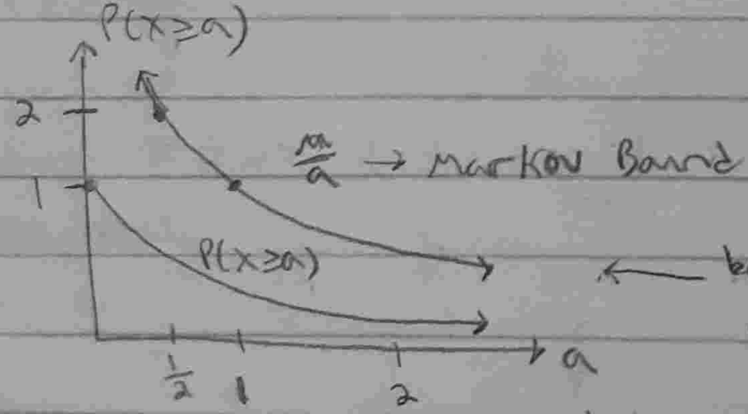
$$= \text{Bern}(P(X \geq a))$$

$$\mu \Rightarrow a P(X \geq a) \leq \mu$$

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

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

$= 1 - F_X(a) = e^{-a} \Rightarrow \mu = 1$




becomes useful only when $a > \mu$

very bad / app.

| a | $P(X \geq a)$ | Markov Bound | Chebyshev | Chernoff |
|---|---|---|---|---|
| 2 | 0.1353 | 0.5 | 1 | 0.73576 |
| 5 | 0.0067 | 0.2 | 0.0635 | 0.09158 |
| 10 | 0.00004 | 0.1 | 0.0123 | 0.000125 |

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

- let $b = a\mu$

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

- 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 non-negative

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

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

$$\Rightarrow 1 - p \leq \frac{\mu}{F_X^{-1}(p)} \Rightarrow \underbrace{F_X^{-1}(p)}_{Q[X,p]} \leq \frac{\mu}{1-p}$$

e.g. $\text{Med}[X] \leq 2\mu$

• Let $X$ be any r.v. $\Rightarrow |X|$ is non-negative r.v.

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}$$

****** "famous one"

• Let $X$ be any r.v. with finite variance, $\sigma^2$. Let $Y = (X-\mu)^2 \Rightarrow Y$ is a non-neg. r.v.

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

def of variance

let $b = a^2$

$$\Rightarrow P((X-\mu)^2 \geq b) \leq \frac{\sigma^2}{b} \Rightarrow P((X-\mu)^2 \geq a^2) \leq \frac{\sigma^2}{a^2}$$

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

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

disjoint

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

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

(for usefulness) if $a \geq \mu$

$a \geq m$

$$P(X \geq \mu + a) + P(X \leq \mu - a) \Rightarrow$$

$$P(X \geq \mu + a) + \cancel{P(X \leq \text{neg.} \#)}^{0}$$

let $b = \mu + a$

$$= P(X \geq b) \leq \frac{\sigma^2}{(b-\mu)^2}$$

· Let $X$ be any r.v.

Let $Y = e^{tX} \Rightarrow Y$ is a non-neg. r.v. for all $t$

$$P(Y \geq b) \leq \frac{E[Y]}{b} \Rightarrow P(e^{tX} \geq b) \leq \frac{\overbrace{E[e^{tX}]}^{\text{MGF for } X}_{M_X(t)}}{b}$$

let $b = e^{ta}$

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

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

if $t > 0$

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

if $t < 0$

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

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

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

$$P(X \leq a) \leq \min_{t < 0} \left\{ e^{-ta} M_X(t) \right\}$$

This is called "Chernoff's Inequality".

Let's calculate it for $X \sim Exp(\lambda)$.
warning it's a lot of work...

First, we need to find mgf for the exponential r.v.

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

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

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

If $t > \lambda$, the mgf does not exist. This is why chf's are better, they always exist.

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

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

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

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

$$= \frac{(t-1)ae^{-ta} + e^{-ta}}{(1-t)^2}$$

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

$$\Rightarrow t_* = \frac{a-1}{a} = 1 - \frac{1}{a}$$

$$h(t_*) = \frac{e^{-a}e}{\frac{1}{a}} = \frac{ae}{e^a}$$

But... the Chernoff bound is practically useless. Why? Because it requires the mgf. To get the mgf, you need the PDF or PMF. If I know the PDF or PMF, then I know analytically or can numerically compute the CDF, so I know the tail exactly or within small numerical error! So it is really only useful when you only have the MGF and not the PDF/PMF.