NOV-25-2020

lecture - 21

Another Application of Multivariate Characteristic function :-

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

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

$$\xRightarrow{P_1/P_8} X_1 \sim f_{X_1}(x)$$

$$f_{X_1}(x) = \int\cdots\int f_{X_1,X_2\cdots X_n}(x_1, u_1, u_2 \cdots u_{n-1})\, du_1 \cdots\cdots du_{n-1}$$

Example :- $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$, $X_1 \sim f_{X_1}(x) = ?$

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

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

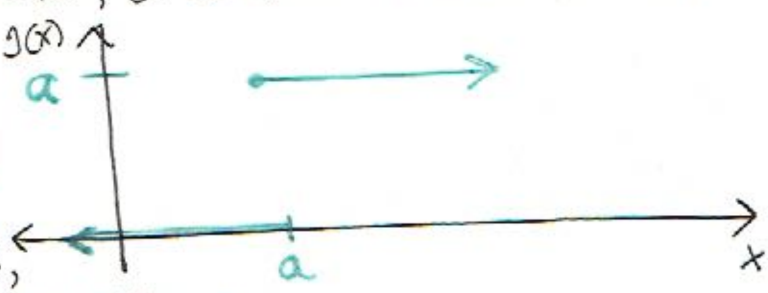
$$= e^{it\mu_1 - \frac{t^2\sigma^2}{2}}$$

$$= \phi_{X_1}(t) \xRightarrow{P_1} X_1 \sim N(\mu_1, \sigma_1^2)$$

HW#3

Pure Math Stuff :-

Assume X is a rv with non negative 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 this, $a\mathbb{1}_{x\geq a} \leq x$? YES!

Two cases,

Case 1: If, $x<a$, $a\mathbb{1}_{x\geq a} = a(0) = 0 \leq x$, becoz $Supp[X] \geq 0$.

Case 2: If, $x \geq a$, $a\mathbb{1}_{x\geq a} = a(1) = a \leq x$, +ve by case assumption

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

let's take the expectation of both Sides:

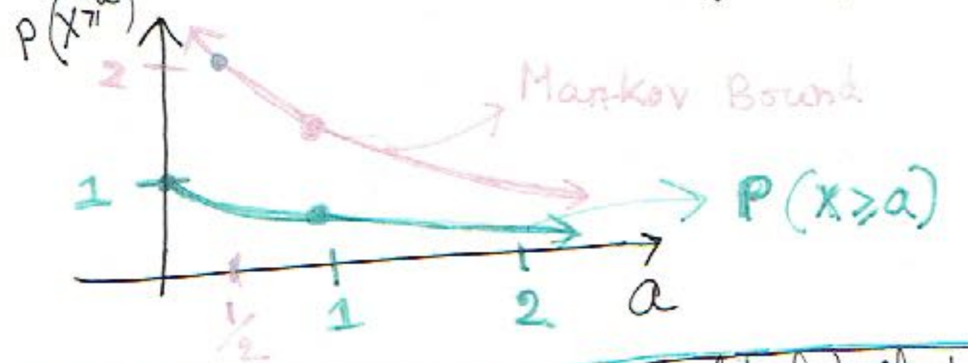
$$E\left[a\mathbb{1}_{X\geq a}\right] \leq E[X] \quad \mathbb{1}_{X\geq a} \sim \begin{cases} 1 & \text{wp } P(X\geq a) \\ 0 & \text{wp } o/t \end{cases}$$

$$\Rightarrow aE\left[\mathbb{1}_{X\geq a}\right] \leq \mu \Rightarrow aP(X\geq a) \leq \mu \qquad = \text{Bern}\left(P(X\geq a)\right)$$

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

For Example,

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



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 = 1/2	1	0.73576
5	0.0067	0.2 = 1/5	0.0635	0.09158
10	0.00004	0.1 = 1/10	0.0123	0.00123

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

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

* let, h be a monotonically increasing function, $y = h(x)$

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

* let, X be continious function in addition to non negative.

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

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

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

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

* let, X be any r.v $\Rightarrow |X|$ is a non negative r.v.

$$P(|X| \geqslant a) \leqslant \frac{E[|X|]}{a}$$

* let, X be any r.v with finite Variance, σ^2

let, $Y = (X-\mu)^2 \Rightarrow Y$ is a non-negative r.v

$$P(Y \geqslant b) \leqslant \frac{E[Y]}{b} \Rightarrow P((X-\mu)^2 \geqslant b) \leqslant \frac{E[(X-\mu)^2]}{b}$$ Variance

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

$$\Rightarrow P(|X-\mu| \geqslant a) \leqslant \frac{\sigma^2}{a^2}$$ → Chebyshev's Inequality

let's manipulate this to get it into a more "user-freindly" form

Assume X - is a non-negative.

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

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

$$= P(X \geqslant a+\mu) + P(X \leqslant \mu - a) \overset{a \geqslant \mu}{=} P(X \geqslant \mu + a) + \cancel{P(X \leqslant -ve}$$

let, $b = \mu + a$

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

let, X be an random variable, let $Y = e^{tX} \Rightarrow Y$ is a non-negative random Variable for all t.

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

Moment generating function for X, $M_X(t)$

$$\Rightarrow P(e^{tX} \geqslant b) \leqslant \frac{M_X(t)}{b} \overset{\text{let } b = e^{ta}}{\Longrightarrow} P(e^{tX} \geqslant e^{ta}) \leqslant e^{-ta} M_X(t)$$

$$\Rightarrow P(tX \geqslant ta) \leqslant e^{-ta} M_X(t)$$

If, $t > 0$

$$P(X \geqslant a) \leqslant e^{-ta} M_X(t)$$

$$P(X \geqslant a) \leqslant \min\{e^{-ta} M_X(t)\}$$

If $t < 0$

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

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

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 $X \sim Exp$ (lamda),

we first need to find mgf for the Exponential r.v:

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

$$= \frac{\lambda}{t-\lambda}\begin{cases} \infty - 1, & t>\lambda \\ 0-1, & \text{if } t \le \lambda \end{cases} = \frac{\lambda}{\lambda - t},$$

only for $t<\lambda$.

If, $t>\lambda$, the mgf doesn't exists.
This is why you shouldn't be using them! chf's always exists! So they are awsome...

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

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

$$\Rightarrow P(X>a) \le \min_{t\in(0,1)}\left\{\overbrace{e^{-ta}\frac{1}{1-t}}^{h(t)}\right\} \longrightarrow = e^{-\left(1-\frac{1}{a}\right)a}\frac{1}{1-\left(1-\frac{1}{a}\right)} = \boxed{\frac{ae}{e^a}}$$

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

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

$$\text{So, } \frac{e^{-ta}(ta-a+1)}{(1-t)^2} = 0$$

$$\Rightarrow ta - a + 1 = 0$$

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

But the chernoff is practically useless. why? Bcoz it requires the mgf. To get the mgf, you need to know 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 with a 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.