Lecture 21

11-25-2020

Math 621

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

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

$= E[e^{itX_1}] = \phi_{X_1}(t)$

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

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

---

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

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

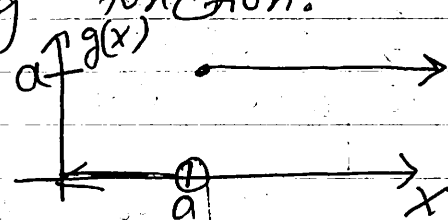
$= e^{it\mu_1 - \frac{t}{2}[t\,0 \cdots 0]\begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \vdots \\ \sigma_{1n} \end{bmatrix}}$

$= e^{it\mu_1 - \frac{t^2\sigma_1^2}{2}}$

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

Assume X is a rv with non-negative support i.e. $\text{Supp}[X] \geq 0$ and has finite expectation. Let $a > 0$, a constant. Consider the following function:

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



Is $a\mathbb{1}_{x\geq a} \leq x$?

Two cases:

- if $x < a$, $a\mathbb{1}_{x\geq a} = a(0) = 0 \leq x$ because $\text{Supp}[X] \geq 0$.

(2)

- if $X \geq a$, $a\mathbb{1}_{X \geq a} = a(1) = a \leq X$ True by Case Assumption

So. Answer is Yes.

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

Let's take the expectation of both sides:

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

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

$\mathbb{1}_{X \geq a} \sim \begin{cases} 1 & \text{wp } P(X \geq a) \\ 0 & \text{otherwise} \end{cases} = Bern(P(X \geq a))$

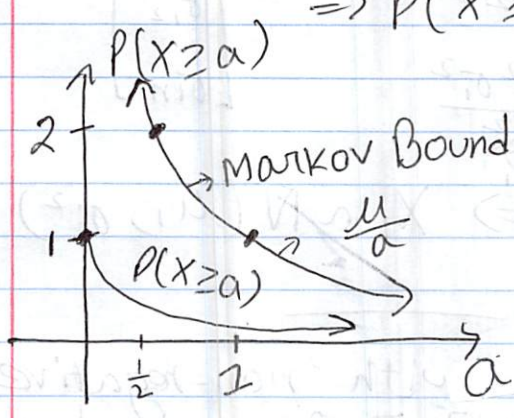
$\Rightarrow a P(X \geq a) \leq \mu$

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

For example: $X \sim Exp(1) = e^{-x}$

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

$\Rightarrow \mu = 1$



The table shows the Markove Bound is very "crude" meaning very approximate, much bigger than the truth.

a	$P(X \geq a)$	Markove Bound	Cheb. Bound	Chernoff bound
2	0.1353	$\frac{1}{2} = 0.5$	1	0.73576
5	0.0067	$\frac{1}{5} = 0.2$	0.0635	0.09158
10	0.00004	$\frac{1}{10} = 0.1$	0.0123	0.00123

We will now prove corollaries of the Markov Inequality:

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

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

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

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

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

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

3) Let X be continuous in addition to non-negative.

Let $a = \text{Quantile}[X, p] = F_X^{-1}(p)$ ← CDF of inverse

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

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

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

e.g.

$\text{Med}\,[X] \leq 2\mu$

← this is true for when X be continuous ... non-negative.

4) Let X be any rv $\Rightarrow |X|$ is a non-negative rv.

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

5) Let X be any rv with finite variance, σ^2.
Let $Y = (X - \mu)^2 \Rightarrow Y$ is a non-negative rv.

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

definition of variance

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

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

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

This is called "Chebyshev's Inequality".

Let's manipulate this:

Assume X is nonnegative:

$P(|X-\mu| \geq a) = P(\underbrace{(X-\mu \geq a) \overset{\text{union}}{\cup} (-(X-\mu) \geq a)}_{\text{disjoint events}})$

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

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

if $a \geq \mu$

$\Rightarrow P(X \geq \mu+a) + \cancel{P(X \leq \text{negative } \#)} \rightarrow 0$

Let $b=\mu+a \Rightarrow b-\mu=a$

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

6) • Let X be an r.v. Let $Y = e^{tX}$

$\Rightarrow Y$ is a non-negative r.v for all t.

$P(Y \geq b) \leq \frac{E[Y]}{b}$

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

moment-generating function for X, $M_X(t)$

$\Rightarrow P(e^{tX} \geq b) \leq \dfrac{M_X(t)}{b}$

Let $b = e^{ta}$

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

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

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?

$\Rightarrow P(X \geq a) \leq \min\limits_{t>0} \{e^{-ta} M_X(t)\}$

$\Rightarrow P(X \leq a) \leq \min\limits_{t<0} \{e^{-ta} M_X(t)\}$

This is called "Chernoff's Inequality".

Let's calculate it for $X \sim Exp(\lambda)$.

first: find mgf for the exponential rv:

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

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

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

$$= \frac{\lambda}{t-\lambda} \quad \text{only for } t < \lambda.$$

If $t > \lambda$, the mgf doesn't exist. This is why you shouldn't be using them!

Chf's always exist!

⑥ $X \sim Exp(1) \Rightarrow M_X(t) = \frac{1}{1-t}$ for $t < 1$.

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

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

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

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

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

set to 0,

$ta - a + t = 0$

$t^* = \frac{a-1}{a} = \boxed{1 - \frac{1}{a}}$

subtitute

$$\therefore \min_{t \in (0,1)} e^{-(1-\frac{1}{a})a} \frac{1}{1-(1-\frac{1}{a})}$$

$$= \frac{e^{-a} e}{1/a} = \frac{ae}{e^a}$$

But .. the Chernoff is sometimes useless.
Why? Because it requires the mgf.
To get 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 within small numerical error!
Only useful when you have mgf and not have PDF or PMF.