Week 2

Ω, P, E, $\underset{\sim}{X}$, X

$$\boxed{X \le Y \implies E(X) \le E(Y)}$$

Recall $|a+b| \le |a| + |b|$, $|\underset{\sim}{a}+\underset{\sim}{b}| \le |\underset{\sim}{a}| + |\underset{\sim}{b}|$

$$\implies E(|\underset{\sim}{X}+\underset{\sim}{Y}|) \le E(|\underset{\sim}{X}| + |\underset{\sim}{Y}|) = E(|\underset{\sim}{X}|) + E(|\underset{\sim}{Y}|)$$

Moments

$E(X^k)$ — k th moment $(k = 1, 2, \cdots)$

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = E(X - \underset{E(X)}{\mu})^2$$

$$\sigma = SD(X)$$

$$\underset{\sim}{\mu} = E(\underset{\sim}{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{pmatrix}$$

$$\underbrace{Var(\underset{\sim}{X})}_{\Sigma} = E\Big[\underbrace{(\underset{\sim}{X} - \underset{\sim}{\mu})}_{k\times 1}\underbrace{(\underset{\sim}{X} - \underset{\sim}{\mu})'}_{1\times k}\Big] = \Big\{\overbrace{\underset{\substack{\| \\ E(X_i X_j) - \mu_i \mu_j}}{E[(X_i - \mu_i)(X_j - \mu_j)]}}^{cov(X_i, X_j)}\Big\}_{i=1}^{k}$$

Lemma (i) $E[A\underset{\sim}{X} + \underset{\sim}{b}] = A\, E(\underset{\sim}{X}) + \underset{\sim}{b}$

(ii) $Var(A\underset{\sim}{X} + \underset{\sim}{b}) = A\, Var(\underset{\sim}{X})\, A'$

Pf: Do it

consequence $Var(X_1 + \cdots + X_k)$

$= Var(\underset{\sim}{1}'\underset{\sim}{X}) = \underset{\sim}{1}'\, Var(\underset{\sim}{X})\, \underset{\sim}{1}$

$= \sum_{i,j} cov(X_i, X_j)$

Note ① X's uncorrelated $(cov(X_i, X_j) = 0 \text{ if } i \neq j)$

$\Rightarrow Var(X_1 + \cdots + X_k) = Var(X_1) + \cdots + Var(X_k)$

② $Var(\underset{\sim}{X}) = \underbrace{E(\underset{\sim}{X}\underset{\sim}{X}')}_{\text{product moment matrix}} - \underset{\sim}{\mu}\underset{\sim}{\mu}'$

product moment matrix
i,j th element is $E(X_i X_j)$

③ $\operatorname{Var}(\underset{\sim}{c}'\underset{\sim}{X}) = \underbrace{\underset{\sim}{c}'\Sigma\underset{\sim}{c} \geq 0 \overset{\text{assumption}}{>} 0 \text{ if } \underset{\sim}{c} \neq \underset{\sim}{0}}$

i.e. Σ is <u>positive definite</u>

so that Σ^{-1} exists.

$$\Sigma = TT'$$

$$= Q\underset{\substack{\uparrow \\ \text{diagonal}}}{D}Q'$$

$QQ' = Q'Q = I$ orthogonal

$$= \underbrace{QD^{\frac{1}{2}}Q'}_{\Sigma^{\frac{1}{2}}}\underbrace{QD^{\frac{1}{2}}Q'}_{\Sigma^{\frac{1}{2}}}$$

<u>Basics again</u>

$\underset{\sim}{X} : \Omega \to \mathbb{R}^k$

E on $\{\text{rv's}\}$; P on $\{\text{events}\}$

$X_n \to X$, $X_n \uparrow X$, $X_n \downarrow X$

$A_n \to A$, $A_n \uparrow A$, $A_n \downarrow A$

<u>MCT (Axiom 4)</u> $0 \le X_n \uparrow X \Rightarrow E(X_n) \to E(X)$

<u>Notice</u> $A_n \uparrow A \Rightarrow I_{A_n} \uparrow I_A$

$\Rightarrow E(I_{A_n}) \to E(I_A)$ (MCT)

$\Rightarrow P(A_n) \to P(A)$

If $A_n \downarrow A \Rightarrow A_n^c \uparrow A^c$

$\Rightarrow P(A_n^c) \to P(A^c) \Rightarrow P(A_n) \to P(A)$

<u>Dominated Convergence Theorem (DCT)</u>

$X_n \to X$ & $|X_n| \le W$ with $E(W) < \infty$
then $E(X_n) \to E(X)$.

Application $A_n \to A$

$$\Rightarrow I_{A_n} \to I_A$$

$$\Rightarrow E(I_{A_n}) \to E(I_A) \qquad \text{(DCT)}$$

$$\Rightarrow P(A_n) \to P(A)$$

$\Omega = \{\omega\}$

$\mathcal{E}$

eg $\Omega = \mathbb{R} = \{\overset{\#}{\omega}\}$

Our rv's are f'ns from $\mathbb{R}$ to $\mathbb{R}$

graph of an indicator rv

Suppose $\exists\, f \rightarrow f(\omega) \geq 0, \ \forall \omega$

you have seen many

& $\sum_{\omega \in \mathbb{R}} f(\omega) = 1$

Set

$$E(g) = \sum_{\omega \in \mathbb{R}} g(\omega) f(\omega)$$

Note $g : \mathbb{R} \to \mathbb{R}$

$\rightarrow$ defines a discrete dist'n

The cts version would have

$$f(\omega) \geq 0$$

$$\int_{\mathbb{R}} f(\omega)\, d\omega = 1$$

& define

$$E(g) = \int_{\mathbb{R}} g(\omega) f(\omega)\, d\omega$$

Suppose we have a dist'n on Ω

Let $X : \Omega \to \mathbb{R}$ ← fix X

$B \subset \mathbb{R}$ $\qquad \{\omega : X(\omega) \in B\}$

then set

$$P_X(B) = P(\overbrace{X^{-1}(B)})$$

$(\mathbb{R}, \text{"subsets" of } \mathbb{R}, P_X)$ is a probability space generated by X.

Depending on X one can get simple or complicated dist'ns on $\mathbb{R}$. If

$$E[h(X)] = \int_{\mathbb{R}} h(x) \overbrace{f(x)}^{\geq 0 \,\&\, \text{area } 1} dx, \text{"}\forall\text{" } h

then X is an (absolutely) cts rv with pdf f. If $\int$ were replaced by $\sum_{\mathbb{R}}$ then $X^{\mathbb{R}}$ would be a discrete rv.

<u>Boole's Inequality</u>

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$$

<u>Proof</u>: clearly

$$I\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} I(A_i)$$

$$\Rightarrow E\left[I\left(\bigcup_{i=1}^{\infty} A_i\right)\right] \le E\left[\sum_{i=1}^{\infty} I(A_i)\right]$$

$$= \sum_{i=1}^{\infty} E[I(A_i)]$$

$$\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i) \quad \text{qed}$$

Application

Let $X \geq 0$. Suppose $E(X) = 0$. Then

$$X = 0 \quad \underbrace{\text{wp1}}_{\text{almost surely (a s)}}$$

$\{ X \overset{wp1}{=} 0, \; X \overset{as}{=} 0 \}$

Note $X \overset{as}{=} Y$, $X \overset{wp1}{=} Y$

If $EX^2, EY^2 < \infty$ + $E(Y-X)^2 = 0$

then we say $X \overset{ms}{=} Y$, mean square

Sol'n We want to show $P(X=0) = 1$.

Look at

$$P(X > 0) = P\left(\bigcup_{k=1}^{\infty} \{ X \geq \tfrac{1}{k} \} \right)$$

$$\overset{\text{Boole}}{\leq} \sum_{k=1}^{\infty} P\left(X \geq \tfrac{1}{k} \right)$$

$$\overset{\text{Markov}}{\leq} \sum_{k=1}^{\infty} E(X) / \left(\tfrac{1}{k} \right) = 0$$

$$\Rightarrow P(X=0) = 1$$

<u>qed</u>

<u>Consequence</u> $\mathrm{Var}(Y) = 0$

$$\Rightarrow Y \overset{as}{=} \mu$$

<u>Markov's Inequality</u> Let $c > 0$. Then

$$P(|X| \geq c) \leq E(|X|)/c$$

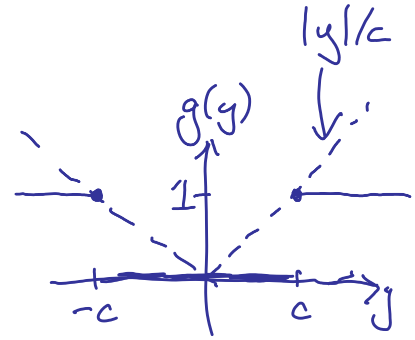
<u>Proof:</u>

$$P(|X| \geq c) = E[I(|X| \geq c)]$$

$$= E[g(X)],$$

where

$$g(y) = 1, \quad |y| \geq c$$
$$= 0, \quad |y| < c$$



Clearly

$$g(X) \le \frac{|X|}{c}$$

& so $E[g(X)] \le E(|X|)/c$

$\Longrightarrow$ $P(|X| \ge c) \le E(|X|)/c$

qed

$$\boxed{\underbrace{P(|X-\mu| \ge k\sigma)} \le \frac{1}{k^2}}$$

$$P[(X-\mu)^2 \ge k^2\sigma^2] \le \frac{\overbrace{E(X-\mu)^2}^{\sigma^2}}{k^2\sigma^2} = \frac{1}{k^2}$$

<u>Application</u> (prediction)

predict Y using $\underset{\sim}{X}$

Let $\hat{Y}$ denote the predictor.

$\hat{Y} = g(\underset{\sim}{X})$, where g is <u>real valued</u>

$E(\hat{Y}-Y)^2$ — <u>MSE</u>

Stick to linear f'ns $A\underset{\sim}{X}$. Assume means are zero (if not subtract them).
For this case we use

$$\hat{Y} = \underset{\sim}{a}'\underset{\sim}{X}$$

Now

$$\underbrace{E(\hat{Y}-Y)^2}_{MSE(\underset{\sim}{a})} = E(\underset{\sim}{a}'\underset{\sim}{X} - Y)^2$$

$$= E\left(Y^2 + \underset{\sim}{a}'\underset{\sim}{X}\underset{\sim}{X}'\underset{\sim}{a} - 2Y\underset{\sim}{a}'\underset{\sim}{X}\right)$$

$$= E(Y^2) + \underset{\sim}{a}'\underbrace{E(\underset{\sim}{X}\underset{\sim}{X}')}_{\text{product moment matrix}}\underset{\sim}{a} - 2\underset{\sim}{a}'\underbrace{E(Y\underset{\sim}{X})}_{\begin{pmatrix} E(YX_1) \\ E(YX_2) \\ \vdots \end{pmatrix}}$$

$MSE(\underset{\sim}{a})$

Look at

$$\frac{\partial\, MSE(\underset{\sim}{a})}{\partial \underset{\sim}{a}'} = \underset{\sim}{0}'$$

+ you get an optimal $\underset{\sim}{a}$

Back to calculus

$g : \mathbb{R}^n \to \mathbb{R}^m$

$g'(\underset{\sim}{x})$

matrix

$$y = g(\underset{\sim}{x})$$

$$\boxed{\frac{dy}{d(x_1, x_2, \ldots, x_n)}} = \frac{dy}{d\underset{\sim}{x}'}$$

$$= \begin{pmatrix} \frac{dy}{dx_1} & \frac{dy}{dx_2} & \cdots & \frac{dy}{dx_n} \end{pmatrix}$$

$$y = \underset{\sim}{x}' \underset{\sim}{d}$$

$$\frac{dy}{d\underset{\sim}{x}'} = \underset{\sim}{d}'$$

$$\underbrace{y = \underset{\sim}{x}' B \underset{\sim}{x}}$$

[illegible]

$$\frac{\partial}{\partial \underset{\sim}{x}'} \underset{\sim}{b}'\underset{\sim}{x} = \underset{\sim}{b}'$$

$$\frac{\partial}{\partial \underset{\sim}{x}'} \underset{\sim}{x}'A\underset{\sim}{x} = 2\underset{\sim}{x}'A$$

$$\frac{\partial}{\partial \underset{\sim}{x}'} A\underset{\sim}{x} = A$$

used in linear prediction problem

$$\boxed{\Sigma_{\underset{\sim}{X}\underset{\sim}{X}} \underset{\sim}{a} = \Sigma_{Y\underset{\sim}{X}} \qquad (*)}$$

$$\left.\begin{aligned} X &\le |X| \\ -X &\le |X| \end{aligned}\right\} \Longrightarrow \left.\begin{aligned} E(X) &\le E(|X|) \\ -E(X) &\le E(|X|) \end{aligned}\right\} \Longrightarrow |E(X)| \le E(|X|)$$

$$Y \ge 0 \qquad 0 < k_1 \le k_2$$

$$E(Y^{k_2}) < \infty \Longrightarrow E(Y^{k_1}) < \infty$$

$$E(Y^{k_1}) = E\Big[Y^{k_1}\big(\underbrace{I(0 \le Y < 1) + I(Y \ge 1)}_{1}\big)\Big]$$

$$= E\big[Y^{k_1} I(0 \le Y < 1)\big] + E\big[Y^{k_1} I(Y \ge 1)\big]$$

$$\le 1 + E\big[Y^{k_2} I(Y \ge 1)\big]$$

$$\le 1 + E(Y^{k_2}) < \infty$$

Look at

$$\sum_{k=1}^{\infty} X_k = \lim_{n\to\infty} S_n \quad \leftarrow X_1 + \cdots + X_n$$

Assume

$$\sum_{k=1}^{\infty} |X_k| < \infty \text{ or } E\left(\sum_{k=1}^{\infty} |X_k|\right) < \infty \implies \sum_{k=1}^{\infty} X_k \quad \text{exists}$$

$$|S_n| \leq |X_1| + \cdots + |X_n|$$
$$\leq \sum_{k=1}^{\infty} |X_k|$$

Look at

$$E\left(\sum_{k=1}^{\infty} X_k\right) = E\left(\lim_{n\to\infty} S_n\right) \qquad \sum_{k=1}^{\infty} E(X_k)$$
$$\overset{DCT}{=} \lim_{n\to\infty} E(S_n) = \lim_{n\to\infty} \sum_{k=1}^{n} E(X_k)$$

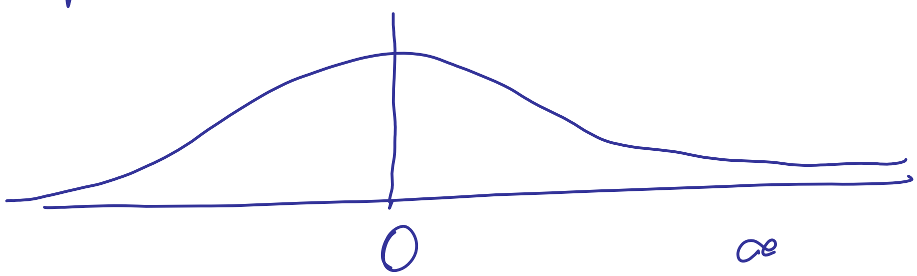
<u>Recall</u> If $|Y_n| \leq W$ & $E(W) < \infty$ &

<u>DCT</u> $Y_n \longrightarrow Y$

then $E(Y_n) \to E(Y)$ $\left(\lim_{n\to\infty} E(Y_n) = E(\lim_{n\to\infty} Y_n)\right)$

Note: A Cauchy rv X has pdf

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad \forall x$$

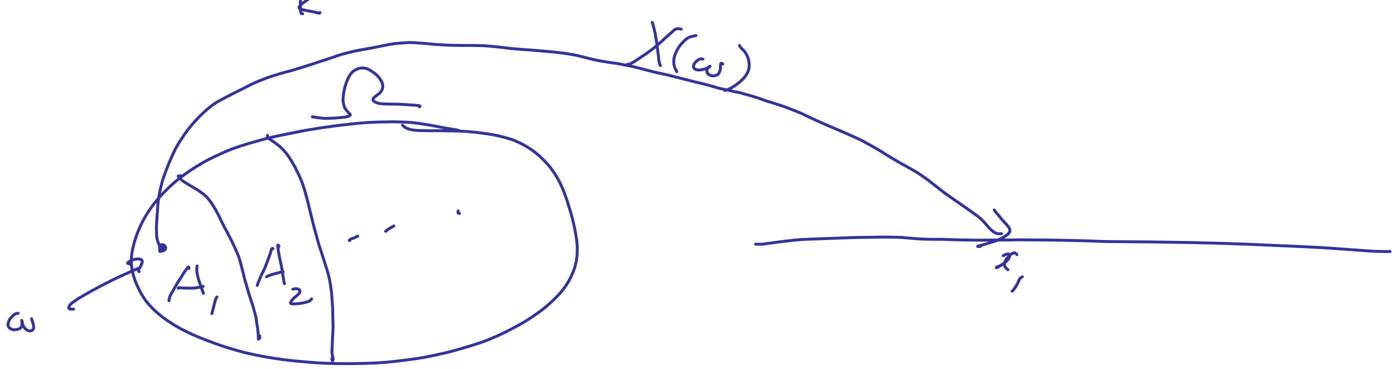


$$E(|X|) = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}\,dx = 2\int_0^{\infty} \frac{x}{\pi(1+x^2)}\,dx = \infty$$

discrete rv: range is countable

X - discrete, range $= \{x_1, x_2, \dots\}$

Let $A_k = \{X = x_k\} = \{\omega : X(\omega) = x_k\}$



The A_k partition Ω.

Let $g: \mathbb{R} \to \mathbb{R}$. Then

$g(X) = g \circ X$ is a discrete r.v.

$$X = \sum_{k=1}^{\infty} x_k I(A_k)$$

$$g(X) = \sum_{k=1}^{\infty} g(x_k) I(A_k)$$

~~If~~ we can interchange E & $\sum_{k=1}^{\infty}$ then

$$E[g(X)] = \sum_{k=1}^{\infty} E[g(x_k) I(A_k)]$$

$$= \sum_{k=1}^{\infty} g(x_k) \; E[I(A_k)]$$

$$= \sum_{k=1}^{\infty} g(x_k) \; P(A_k)$$

$$= \sum_{k=1}^{\infty} g(x_k) \underbrace{P(X = x_k)}_{f(x_k)}$$

$$\boxed{f(x) = P(X=x) \quad - \text{Pf}}$$