Lecture 10 Conditional Expectation

1. Conditional Distribution Prob. Space $(\Omega, \mathcal{F}, P)$

$B \in \mathcal{F}$ $P(B) > 0$

Cond Prob. of $A \in \mathcal{F}$ given B is $P(A|B) = \frac{P(A \cap B)}{P(B)}$

X, Y r.v defined on prob. space $(\Omega, \mathcal{F}, P)$

Suppose, $y \in \mathbb{R}$ s.t. $F_Y(y) > 0$

$A_x := X^{-1}(-\infty, x]$, $B_y := Y^{-1}(-\infty, y]$

$$P(A_x | B_y) = \frac{P(A_x \cap B_y)}{P(B_y)} = \frac{F_{XY}(x,y)}{F_Y(y)}$$

- We understand cond. prob. of one event given another event. (non trivial event)

1-1 Conditioning on simple r.v

$Y: \Omega \to \underbrace{\mathcal{Y}}_{\text{finite}} \subseteq \mathbb{R}$

$E_y := Y^{-1}\{y\} \in \mathcal{F}$ for $y \in \mathcal{Y}$

$(E_y : y \in \mathcal{Y})$ partition of Ω.

$P_Y : \mathcal{Y} \to [0,1]$
$y \mapsto P(E_y)$ $\quad P_Y(y) > 0$

$X: \Omega \to \mathbb{R}$ r.v on $(\Omega, \mathcal{F}, P)$

$F_{X|E_y} : \mathbb{R} \to [0,1]$

$$x \mapsto P(\{X \le x\} | E_y) = \frac{P(\{X \le x\} \cap E_y)}{P_Y(y)}$$

- $F_{X|E_y}$ is a distribution because it follows all the rules of a distribution. (conditional distribution of X given E_y)

Defn: [Cond. dist. of X given Y]

$\longrightarrow$ space of all functions

$F_{X|Y} : \Omega \to [0,1] \subseteq \mathbb{R}$ (measurable fn. of r.v Y)

wait

$\omega \mapsto F_{X|\{Y=Y(\omega)\}}$

Recall $\left[Y = \sum_{y \in \mathcal{Y}} y \, 1_{E_y} \right]$ on all events, specific event

$$\boxed{F_{X|Y} = \sum_{y \in \mathcal{Y}} F_{X|E_y} 1_{E_y}} \quad *$$

Ex:— $N \sim \mathcal{N}(0, \sigma^2)$

$X \in \{-1, 1\}$ PMF $(1-p, p)$ $p \in (0,1)$

$Y = X + N$

$$F_{Y|X} = F_{Y|\{X=-1\}} 1_{\{X=-1\}} + F_{Y|\{X=1\}} 1_{\{X=1\}}$$

$$F_{Y|\{X=m\}} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} e^{-\left(\frac{(t-m)^2}{2\sigma^2}\right)} dt.$$

1.2 Conditional densities

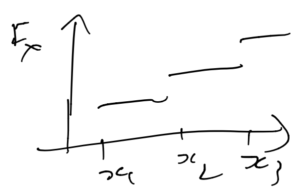
$$f_{X|\{Y=y\}}(x) := \frac{f_{XY}(x,y)}{f_Y(y)} \quad \forall x \in \mathbb{R}$$

Ex. $f_{X|\{Y=y\}}$ is a density function

Defn: $f_{X|Y} : \omega \mapsto f_{X|\{Y=Y(\omega)\}}$

$\mathbb{E}X = \int_{\mathbb{R}} x \, dF_X(x)$ Conditional Expectation (in terms of cond. dist.)

$X = \sum_x x 1_{E_x}$ $\quad \mathbb{E} X = \int_{\mathbb{R}} t \, dF_X(t)$



$$\mathbb{E}[X] = x_1 P_X(x_1) + x_2 P_X(x_2) + x_3 P_X(x_3)$$

2-1 Simple r.v

$$\mathbb{E}[X|E_y] := \int_{x \in \mathbb{R}} x \, dF_{X|E_y}(x)$$

$$= \int_{x \in \mathbb{R}} x \, d_x \frac{P(\{X \le x\} \cap E_y)}{P_Y(y)}$$

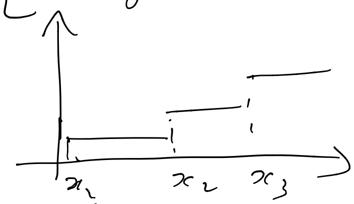
$$= \frac{1}{P_Y(y)} \int_{x \in \mathbb{R}, \{Y=y\}} x \; dF_{XY}(x,t)$$

$$= \frac{1}{P_Y(y)} \int_{(x,y) \in \mathbb{R}\times\mathbb{R}} x \, \mathbb{1}_{E_y} \; dF_{X,Y}(x,t)$$

$$= \frac{1}{P_Y(y)} \int_{(x,y) \in \mathbb{R}\times\mathbb{R}} x \, \mathbb{1}_{E_y} \, dF_{X,Y}(x,t)$$

$$= \frac{\mathbb{E}[X \mathbb{1}_{E_y}]}{P_Y(y)}$$

$$\left[P(E_y) = \mathbb{E}[\mathbb{1}_{E_y}] = \int_{t \in \mathbb{R}} \mathbb{1}_{E_y} \, dF_Y(t) = \int_{t=y} dF_Y(t) \right.$$



$$P(E_{x_1}) = P_X(x_1)$$

$$\int_{x \in \mathbb{R}} \mathbb{1}_{E_{x_1}} \, dF_X(x) = \int_{x \in \mathbb{R}} \mathbb{1}_{\{x = x_1\}} \, dF_X(x)$$

$$\int_{x=x_1} dF_X(x) = F_X(x_1) - F_X(x_1 -) = P_X(x_1)$$

$$\mathbb{E}[X \mid E_y] = \frac{\mathbb{E}[X \mid E_y]}{P_Y(y)}$$

Definition: $\mathbb{E}[X \mid Y] : \Omega \longrightarrow \mathbb{R}$ is a r.v

$$\omega \longmapsto \int_{x \in \mathbb{R}} x \, dF_{X \mid \{Y = Y(\omega)\}}$$

$$E[X \mid Y] = \sum_{y \in \mathcal{Y}} \underbrace{\mathbb{E}[X \mid E_y]}_{\text{Scalar}} \mathbb{1}_{E_y} = \sum_{y \in \mathcal{Y}} \frac{\mathbb{E}[X \mathbb{1}_{E_y}]}{P_Y(y)} \mathbb{1}_{E_y}$$

$$\left[\text{Recall } Y = \sum_{y \in \mathcal{Y}} y \, \mathbb{1}_{E_y} \right]$$

$$\mathbb{E}[X \mid Y] : \quad Y \xrightarrow{g(Y)} \mathbb{E}[X \mid Y]$$

$$\qquad\qquad y \xrightarrow{g(y)} \mathbb{E}[X \mid E_y]$$

- $E[X \mid Y]$ is a r.v
- $E[X \mid Y]$ is a function of Y
- $\mathbb{E}[\mathbb{E}[X \mid Y] \mathbb{1}_{E_y}] = \mathbb{E}[X \mathbb{1}_{E_y}]$

Lemma: $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}X$

Proof from notes or Law of Total Expectation

Ex: $X \in \{1, 2, 3 \dots 6\}$ with $P\{x=i\} = \frac{1}{6}$, $i \in [6]$

$Y = 1_{\{x \le 3\}}$ $\quad \mathbb{E}[X|Y] = \begin{cases} \mathbb{E}(X|Y=1] = 2 = \frac{\frac{1+2+3}{6}}{\frac{1}{2}} = 2 & \text{w.p } \frac{1}{2} \\ \mathbb{E}(X|Y=0] = \frac{\frac{4+5+6}{6}}{\frac{1}{2}} = 5 & \text{w.p } \frac{1}{2} \end{cases}$

$\mathbb{E}[\mathbb{E}[\cancel{X}|Y]] = 3.5$

$\mathbb{E}X = 3.5$

$g(Y) = Y \longmapsto \mathbb{E}[X|Y]$

$0 \longmapsto 5$

$1 \longmapsto 2$

2.2- Continuous r.v

$\mathbb{E}[X|Y] : \omega \longmapsto \int_{x \in \mathbb{R}} x f_{X|Y=Y(\omega)}(x)\, dx$

Lemma: $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}X$

3- Conditional expectation (Formal)

i- $F_{X|B}(x) = \frac{P(\{X \le x\} \cap B)}{P(B)}$ $\quad F_{X|B} := 0$ if $P(B) = 0$

ii- $E[X|B] = \int_{x \in \mathbb{R}} x\, dF_{X|B}(x)$ $\quad E(X|B] = 0$ if $P(B) = 0$

iii- $\mathbb{E}[X|Y]$ is a measurable function of Y $\forall B \in \sigma(Y)$

$\mathbb{E}[\mathbb{E}[X|Y] 1_B) = \mathbb{E}[X 1_B]$

Defn: X is a r.v on $(\Omega, \mathcal{F}, P)$, $E|X| < \infty$.

$Y := \mathbb{E}(X|\mathcal{G})$, $\mathcal{G} \subset \mathcal{F}$ is a r.v on the $(\Omega, \mathcal{F}, P)$ s.t

i- Y is $\mathcal{G}$ measurable

ii- $\mathbb{E}[X 1_A] = \mathbb{E}[Y 1_A]$ $\forall A \in \mathcal{G}$

iii- $\mathbb{E}|Y| < \infty$

$[\mathbb{E}[X|Z]] = [\mathbb{E}[X|\sigma(Z)]]$

Ex: Simple $Y: \Omega \to \mathcal{Y}$ $E_y: Y^{-1}\{y\} \in \mathcal{F}$

$\sigma(Y) = (\bigcup_{y \in I} E_y : I \subseteq \mathcal{Y})$

$Z := \mathbb{E}[X|Y]$ is $\sigma(Y)$ measurable.

i. implies $\mathbb{E}[X|Y] = \sum_{y \in \mathcal{Y}} \alpha_y 1_{E_y}$

ii. implies $\mathbb{E}[X 1_{E_{y_0}}] = \mathbb{E}[\mathbb{E}[X|Y] 1_{E_{y_0}}]$

$= \mathbb{E}[\alpha_{y_0} 1_{E_{y_0}}] = \alpha_{y_0} P_Y(y_0)$

$\left[\mathbb{E}[X|Y] 1_{E_{y_0}} = \sum_{y \in \mathcal{Y}} \alpha_y 1_{E_y} 1_{E_{y_0}} = \alpha_{y_0} 1_{E_{y_0}}\right]$

$\alpha_{y_0} = \dfrac{\mathbb{E}[X 1_{E_{y_0}}]}{P_Y(y_0)}$

After plugging it in, $\left[\mathbb{E}[X|Y] = \sum_{y \in \mathcal{Y}} \dfrac{\mathbb{E}[X 1_{E_y}]}{P_Y(y)} 1_{E_y}\right]$

Lecture-11 Conditional Expectation (General)

Defn: Probability Space $(\Omega, \mathcal{F}, P)$ & event space $\mathcal{G} \subseteq \mathcal{F}$

A r.v $X: \Omega \to \mathbb{R}$ s.t $\mathbb{E}|X| < \infty$.

Then $Y := \mathbb{E}[X|\mathcal{G}]$ is a r.v

i. $\sigma(Y) \subseteq \mathcal{G}$

ii. $\mathbb{E}[Y 1_A] = \mathbb{E}[X 1_A] \; \forall A \in \mathcal{G}$

iii. $\mathbb{E}|Y| < \infty$

Conditional expectation given a r.v Z

$\mathbb{E}[X|Z] = \mathbb{E}[X|\sigma(Z)]$

$(\mathbb{E}[X|E_y] \Rightarrow \mathbb{E}[X|E_y^c$ is also known$]$

$\mathbb{E}[X|1_{E_y}]$, $\mathbb{E}[X|\sigma(1_{E_y})]$

$\mathbb{E}[X|E_y] = \int_{\mathbb{R}} x \, dF_{X|E_y}$

$\mathbb{E}[X|Z] = \sum_{z \in Z} \mathbb{E}[X|E_z] 1_{E_z}$ where $E_z = \{Z = z\}$

$\mathbb{E}[X|X] = X$ ← Finest σ-algebra for X

$\mathbb{E}[X|\{\phi, \Omega\}] = \mathbb{E}X$ ← Coarsest σ-algebra for X.

$\mathbb{E}[X|\mathcal{F}] = X$; $\mathbb{E}[X|\sigma(Z)]$ → lies somewhere in between.

4. Properties of conditional $\mathbb{E}$

Let X, Y be two r.v on $(\Omega, \mathcal{F}, P)$, s.t. $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$

Two event subspaces $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$.

1. Linearity: $\mathbb{E}[\underbrace{\alpha X + \beta Y | \mathcal{G}}_{\text{r.v } Z}] = \alpha \underbrace{\mathbb{E}[X|\mathcal{G}]}_{\text{r.v } Z_1} + \beta \underbrace{\mathbb{E}[Y|\mathcal{G}]}_{\text{r.v } Z_2}$

Proof: We have to show $Z = Z_1 + Z_2$

i – $\sigma(Z_1) = \mathcal{G}$, $\sigma(Z_2) \subseteq \mathcal{G}$

This implies $\sigma(\alpha Z_1 + \beta Z_2) \subseteq \mathcal{G}$

ii – Let $A \in \mathcal{G}$

$\mathbb{E}[Z_1 1_A] = \mathbb{E}[X 1_A]$ by definition

$\mathbb{E}[\mathbb{E}[X|\mathcal{G}] 1_A] = \mathbb{E}[X 1_A]$

$\mathbb{E}[Z_2 1_A] = \mathbb{E}[Y 1_A]$ by definition

$\mathbb{E}[(\alpha Z_1 + \beta Z_2) 1_A] = \alpha \mathbb{E}[Z_1 1_A] + \beta \mathbb{E}[Z_2 1_A]$
linearity ↗

$\mathbb{E}[Z 1_A] = \alpha \mathbb{E}[X 1_A] + \beta \mathbb{E}[Y 1_A]$
def. of cond. expectation
$= \mathbb{E}[\alpha X 1_A + \beta Y 1_A]$

iii) Similarly for iii so proved.

$\mathbb{E}[Z 1_A] = \int_{y \in A} g(y)\, dF_Y(y) = \int_{x \in A} x\, dF_X(A)$

$\therefore \mathbb{E}[g(y) 1_A] = \mathbb{E}[X 1_A]$

2. Monotonicity: If $X < Y$ a.s. then $\underbrace{\mathbb{E}[X|\mathcal{G}]}_{Z_1} \le \underbrace{\mathbb{E}[Y|\mathcal{G}]}_{Z_2}$ a.s.

Proof: We have to show that $Z_1 \le Z_2$

Let $\epsilon > 0$, $A_\epsilon := \{Z_1 - Z_2 > \epsilon\} \in \mathcal{G}$.

$\mathbb{E}[(Z_1 - Z_2) 1_{A_\epsilon}] = \mathbb{E}[(X - Y) 1_{A_\epsilon}] \le 0$

$\mathbb{E}[(Z_1 - Z_2) 1_{\{Z_1 - Z_2 > \epsilon\}}] > \mathbb{E}[\epsilon 1_{\{Z_1 - Z_2 > \epsilon\}}]$

$> \epsilon P(A_\epsilon)$

3. If $\sigma(X) \subseteq \mathcal{G}$ and $\mathbb{E}|X| < \infty$, then $\mathbb{E}[X|\mathcal{G}] = X$ a.s.

Proof: – We have to show that

i – $\sigma(X) \subseteq \mathcal{G}$.

ii – $\mathbb{E}[X 1_A] = \mathbb{E}[X 1_A]$ for any $A \in \mathcal{G}$

iii – $\mathbb{E}|X| < \infty$

If $\sigma(X) \not\subseteq \mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] \ne X$

4. If $\sigma(Y) \subseteq \mathcal{G}$, & $\mathbb{E}|XY| < \infty$, then $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$ a.s

5. If $\mathbb{E}|X^2| < \infty$, then $\xi^* = \mathbb{E}[X|\mathcal{G}]$ minimises $\mathbb{E}(X-\xi)^2$
$\forall$ $\sigma(\xi^2) < \infty$ & $\mathbb{E}|\xi^2| < \infty$

$$\left[\underset{\substack{\xi \in L^2 \\ \sigma(\xi) \subseteq \mathcal{G}}}{\arg\min} \ \mathbb{E}(X-\xi)^2 = \mathbb{E}[X|\mathcal{G}] \right]$$

6. Tower Property: $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$
$\mathbb{E}\{\mathbb{E}(X|\mathcal{G})|\mathcal{H}\} = \mathbb{E}[X|\mathcal{H}]$ a.s.

7. If $\mathcal{H}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$ then $\mathbb{E}[X|\sigma(\mathcal{G},\mathcal{H})]$
$= \mathbb{E}[X|\mathcal{G}]$ a.s.

[In particular, if X is indep. of $\mathcal{H}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$]

Proof:

Lecture 17

1- Transforms for distribution functions

$X: \Omega \to \mathbb{R}$ r.v defined on $(\Omega, \mathcal{F}, P)$ with dist F_X

Ex. $g_\theta : \mathbb{R} \to \mathbb{R}_+$
$x \mapsto e^{\theta x}$ is Borel measurable $\forall$ $\theta \in \mathbb{R}$

This implies that $g_\theta(X)$ is a r.v

$h_\theta : \mathbb{R} \to \mathbb{C}$
$x \mapsto e^{j\theta x}$ is Borel measurable. $j^2 = -1$, $e^{j\theta x} = \cos\theta x + j\sin\theta x$

This implies $h_\theta x$ is also a r.v

Remark: If $\mathbb{E}|X|^N < \infty$ for some $N \in \mathbb{N}$, then $\mathbb{E}|X|^k < \infty$ $\forall$ $k \in [N]$

$$|X|^k = |X|^k 1_{\{|X| \leq 1\}} + |X|^k 1_{\{|X| > 1\}}$$
$$\leq 1_{\{|X| \leq 1\}} + |X|^N 1_{\{|X| > 1\}}$$
$$\leq 1 + |X|^N$$

1.1- Characteristic function

For a r.v $X: \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, P)$

$\Phi_X : \mathbb{R} \to \mathbb{C}$
$u \mapsto \mathbb{E}[e^{juX}]$

Remarks: 1- $\Phi_X(u) = \mathbb{E}[\cos uX] + j\,\mathbb{E}[\sin uX]$

2- If X discrete, then

$$\Phi_X(u) = \sum_{x \in X} e^{jux} P_X(u)$$

3. If X is continuous, then
$$\Phi_X(u) = \int_{x \in \mathbb{R}} e^{jux} f_X(x)\, dx$$

4. $\Phi_X(u)$ is always finite, since
$\mathbb{E}\,|e^{jux}| = 1$

Ex. [Gaussian r.v.] If $X \sim (\mu, \sigma^2)$ Gaussian, then

$$\Phi_X(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x \in \mathbb{R}} e^{jux} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \Rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x \in \mathbb{R}} \exp\left[\frac{-(x^2+\mu^2-2\mu x-2jux\sigma^2)}{2\sigma^2}\right] dx$$

$$-\left[\frac{(x-\mu-ju\sigma^2)^2 + (\mu+ju\sigma^2)^2 - \mu^2}{2\sigma^2}\right]$$

$$= e^{-\frac{u^2\sigma^2}{2} + ju\mu} \times \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{x \in \mathbb{R}} e^{-\left[\frac{(x-\mu-ju\sigma^2)^2}{2\sigma^2}\right]} dx}_{=1}$$

$$= \exp\left(-\frac{u^2\sigma^2}{2} + ju\mu\right)$$

$|\Phi_X(u)| = e^{-\frac{u^2\sigma^2}{2}}$ $\sim$ Gaussian with mean 0 and variance $\frac{1}{\sigma^2}$

Thm: If $\mathbb{E}|X|^N < \infty$ for some $N \in \mathbb{N}$, then $|\Phi_X^{(k)}(u)| < \infty$ & it is continuous func. of u $\forall\, k \in [N]$

Further $\Phi_X^{(k)}(0) = j^k\, \mathbb{E}[X^k]$ $\forall\, k \in [N]$

Thm: Two r.v have the same prob. dist. iff they have the same characteristic function.

1.2 Moment generating function

$M_X : \mathbb{R} \to \mathbb{R}_+$
$t \mapsto \mathbb{E}[e^{tX}]$, for all $t \in \mathbb{R}$ where $\mathbb{E}[e^{tX}]$ is finite

Lemma: $\Rightarrow M_X(t) = 1 + \sum_{n \in \mathbb{N}} \frac{t^n}{n!} \mathbb{E}[X^n]$

1.3 Prob. gen. Function

Convenient for integer-valued r.v $\Psi_X(z) = \mathbb{E}[z^X]$, $z \in \mathbb{C}$

Lemma: $\Psi_X(z)$ is always finite for $|z| \le 1$

Proof: $|z| \le 1 \Rightarrow |z|^X \le 1 \Rightarrow \mathbb{E}\,|z|^X \le 1$

Thm: $\Psi_X^{(k)}(1) = \mathbb{E}[X(X-1) \ldots (X-k+1)]$

Proof: $\Psi_X(z) = \mathbb{E}[z^X]$

$\Psi_X^{(1)}(z) = \mathbb{E}[X z^{X-1}]$ for X integer

$\Psi_X(1) = \mathbb{E}[X]$

$\Psi_X^{(k)}(z) = \mathbb{E}[X(X-1)(X-2) \ldots (X-k+1) z^{X-k}]$

$\Psi_X^{(k)}(1) = \mathbb{E}[X(X-1)(X-2) \ldots (X-k+1)]$

$\mathbb{E}(X^2) = \Psi_X^{(2)}(1) + \Psi_X^{(1)}(1)$

Do Gaussian Random Vectors from Notes

Lecture 13 Almost Sure Convergence

Consider Prob. space $(\Omega, \mathcal{F}, P)$

A seq. of r.v $X: \Omega \to \mathbb{R}^{\mathbb{N}}$

$\omega \mapsto (X_1(\omega), X_2(\omega), \ldots)$

Recall that $X_i: \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable function

Interested: Convergence of $X_n(\omega) \to ?$

"Defn: A statement holds "almost surely" (a.s) if there exists an event called the "exception set" $N \in \mathcal{F}$, s.t $P(N) = 0$ and the statement holds for all $\omega \notin N$

Ex: [Almost sure equality] X, Y r.v on $(\Omega, \mathcal{F}, P)$

$X = Y$ a.s if $\exists$ an exception set N, i- $N = \{\omega \in \Omega: X(\omega) \neq Y(\omega)\} \in \mathcal{F}$

ii- $P(N) = 0$

Defn: $X: \Omega \to \mathbb{R}^{\mathbb{N}}$ $\quad X_n = \pi_n \circ X$ is the projection in the n_{th} dimension for each $n \in \mathbb{N}$

$\lim_n X_n$ exists a.s if $\exists$ an exception set

(i) $N = \{\omega \in \Omega: \limsup X_n(\omega) \neq \liminf X_n(\omega)\}$

(ii) $P(N) = 0$

Let $X_\infty(\omega) = \begin{cases} \lim_n X_n(\omega) & \omega \notin N \\ 0 & \omega \notin N \end{cases}$ seq. X_n converges a.s to X_∞.

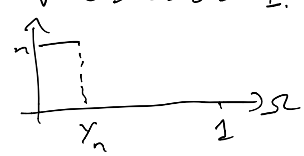
$\lim_n X_n = X_\infty$ a.s.

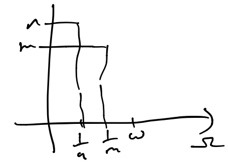
Ex. [A.s convergence $\not\Rightarrow$ everywhere convergence]

$([0,1], \mathcal{B}([0,1]), \lambda)$ s.t $\lambda([a,b]) = b - a \quad \forall\ 0 \le a \le b \le 1$.

For each $n \in \mathbb{N}$, $X_n = \Omega \to \{0, 1\}$

$\omega \mapsto n\, \mathbb{1}_{[0, \frac{1}{n})}(\omega)$





Let $N = \{0\}$ then $\lambda(N) = 0$.

$N^c = \{\omega \in [0,1]: \lim_n X_n(\omega) = 0\} = (0, 1]$

Let $\omega \notin N$, $\exists\, m = \lceil \frac{1}{\omega_0} \rceil \in \mathbb{N}$ s.t $\forall n \ge m$ $\frac{1}{n} \le \frac{1}{m} \le \omega_0$

$X_n(\omega) = 0$

$\lim_n X_n(\omega) = 0$

2- Convergence in Probability

Defn: $X: \Omega \to \mathbb{R}^{\mathbb{N}}$ converges in Prob. to a r.v.

$X_\infty: \Omega \to \mathbb{R}$ if, for any $\epsilon > 0$

$\lim_n P\{\omega \in \Omega: |X_n(\omega) - X_\infty(\omega)| > \epsilon\} = 0$

Remark: A.s. convergence: For almost all outcomes ω, the diff. $|X_n - X_\infty|$ gets small and stays small.

Convergence in Probability: Probability of the difference $|X_n - X_\infty|$ being non-trivial is small.

Ex: [Conv. in Prob. but not a.s] $([0,1], \mathcal{B}([0,1]), \lambda)$

$\forall\ 0 \le a \le b \le 1 \qquad \lambda([a,b]) = b - a$

$S_k := \sum_{i=1}^{k} i \quad (1, 3, 6, 10, 15, \ldots)$ $\qquad I_1 = \{1\}$; $I_2 = \{2, 3\}$, $I_3 = \{4, 5, 6\}$

$I_k := \{S_{k-1} + 1, \ldots, S_{k+1}\}$ $\qquad I_4 = \{7, 8, 9, 10\}$

$(I_k : k \in \mathbb{N})$ partition the natural numbers.
For each $n \in \mathbb{N}$, $n \in I_k$ for some $k \in \mathbb{N}$
$n = S_{k-1} + i$ uniquely, for $i \in [k]$
$X_n := \mathbb{1}_{\left[\frac{i-1}{k}, \frac{i}{k}\right]}$; For any $\omega \in [0,1]$ $X_n(\omega) = 1$ for infinitely many, s.t. $\frac{i-1}{k} \leq \omega \leq \frac{i}{k}$

$\limsup_n X_n(\omega) = 1$ $\quad \lim_n X_n(\omega) \neq 0$ (Check rest from notes)

3. Borel - Cantelli Lemma

Lemma [Infinitely often and almost all] $= (A_n \in \mathcal{F} : n \in \mathbb{N})$

a. For some sel. seq. $(k_n : n \in \mathbb{N})$ dep. on ω,

$$\limsup_n A_n = \left\{ \omega \in \Omega : \sum_{n \in \mathbb{N}} \mathbb{1}_{A_n}(\omega) = \infty \right\}$$
$$= \left\{ \omega \in \Omega : \omega \in A_{k_n}, n \in \mathbb{N} \right\}$$
$$= \{ A_n \text{ infinitely often} \}$$

b. For a finite $n_0(\omega) \in \mathbb{N}$ dep. on ω,

$$\liminf_n A_n = \left\{ \omega \in \Omega : \omega \in A_n \ \forall\, n \geq n_0(\omega) \right\}$$
$$= \left\{ \omega \in \Omega : \sum_{n \in \mathbb{N}} \mathbb{1}_{A_n^c}(\omega) < \infty \right\}$$
$$= \{ A_n \ \forall \text{ but finitely many } n \}.$$

Proof. From Notes.

Lecture 22

Communicating classes: Let $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ be a time homogeneous DTMC with transition matrix P.

Defn: For $x, y \in \mathcal{X}$, y is accessible from x if there exists $n \in \mathbb{Z}_+$ s.t. $P_{xy}^{(n)} > 0$, and is denoted by $x \to y$.

If $x \to y$ and $y \to x$ then x and y communicate and is denoted by $x \leftrightarrow y$

(i) a set of states that communicate are called communicating class.

(ii) A communicating class $\mathcal{C}$ is closed

(iii) A communicating class $\mathcal{C}$ is "open" if $\exists$ one edge that leaves this class.

$$\exists \ (x, y) \in \mathcal{C} \times \mathcal{C}^c \text{ s.t. } P_{xy} > 0$$

Prop: Communication is an equivalence relation

[A relation $R \subseteq \mathcal{X} \times \mathcal{X}$ and is denoted by xRy for $(x,y) \in R$

1. For eg. $f : \mathcal{X} \to \mathcal{X}$ is a relation, given by

$$\{ (x, f(x)) : x \in \mathcal{X} \}$$

2. Comm. class is also a relation

$$\mathcal{C}(x) = \{ (x, y) : y \leftrightarrow x, y \in \mathcal{X} \} \ ]$$

Eq. relation : Symmetry $(x,y) \in R$ & $(y,x) \in R$

Transitivity: If $(x,y) \in R$ & $(y,z) \in R$ then $(x,z) \in R$

Reflexivity $(x,x) \in R$

Remark: The communicating relation partitions state space $\mathcal{X}$ into equivalence classes.

Defn: Each equivalent class created by this commutative relation is called a "communicating class".
A property of states is said to a "class property" if for each comm. class either all states in the class have the property or none do.

1.1 Irreducibility & Periodicity

Defn: A MC with a single class is called an "irreducible" MC. Let
$\mathcal{J}(x) = \{ n \in \mathbb{N} : P_{xx}^{(n)} > 0 \}$, $x \in \mathcal{X}$

The period of any state $x \in \mathcal{X}$ is defined as $d(x) := \gcd \mathcal{J}(x)$
We define $d(x) := \infty$ if $\mathcal{J}(x) = \phi$, i.e., $P_{xx}^{(n)} = 0 \ \forall n \in \mathbb{N}$

A state is called "aperiodic" if the period $d(x) = 1$.

$$P_{xx}^{(n)} = P_x \{ X_n = x \} = \{P^n\}_{xx} \neq \{P^{n+1}\}_{xx}$$

Prop. If $x \leftrightarrow y$ then $d(x) = d(y)$. The periodicity is a class property.
Proof: Since $\left.\begin{matrix} x \to y \\ y \leftarrow y \end{matrix}\right\}$ then $\exists\ n \in \mathbb{Z}_+$ $P_{xy}^{(n)} > 0$

Suppose $s \in \mathcal{J}(x)$ then $P_{xx}^{(s)} > 0$.
Then, $P_{yy}^{(m+s+n)} \geq P_{yx}^{(m)} P_{xx}^{(s)} P_{xy}^{(n)} > 0$ $\quad P_{yy}^{(m+n)} > 0$
Since $d(y) \mid m+s+n$ and $d(y) \mid m+n$, this implies that
$d(y) \mid s, \quad \forall s \in \mathcal{J}(x)$
$d(y) \mid d(x)$
$\blacksquare$

For an irreducible chain, the period of the chain is defined as the common period of states.
An irreducible MC is called "aperiodic" if MC single comm class is aperiodic.
If the finite transition matrix P is aperiodic & is irreducible, then $\exists$ an integer r_0 s.t. $P_{xy}^{(r)} > 0$
$\forall x, y \in \mathcal{X}$ & $r \geq r_0$

1.2 Transient & recurrence states

Prop. Transience & recurrence are class properties.

Proof: Let x be a recurrent state and $x \leftrightarrow y$
$\exists\ m, n \in \mathbb{Z}_+$ s.t. $P_{xy}^{(m)} P_{yx}^{(n)} > 0$

Since x is recurrent, $\sum_{n \in \mathbb{N}} P_{xx}^{(n)} = \infty$

$$\sum_{s \in \mathbb{N}} P_{yy}^{(s)} \geq \sum_{r} P_{yy}^{(n+r+m)} \geq \sum_{r \in \mathbb{N}} P_{yx}^{(n)} \left[P_{xx}^{(r)} \right] P_{xy}^{(m)} = \infty$$

Hence y is recurrent

Let x be a transient state, i.e. $\sum_{r \in \mathbb{N}} P_{xx}^{(r)} < \infty$

$$P_{xy}^{(m)} P_{yy}^{(r)} P_{yx}^{(n)} \le P_{xx}^{(m+r+n)}$$

$$\sum_{r \in \mathbb{N}} P_{yy}^{(r)} \le \left[\sum_{r \in \mathbb{N}} \frac{P_{xx}^{(m+r+n)}}{P_{xy}^{(m)} P_{yx}^{(n)}} \right] < \infty \quad \boxed{\text{E}}$$

Cor: If y is recurrent, then any state x s.t. $y \to x$ then $x \leftrightarrow y$ and $f_{xy} = 1$.

Proof: Let $y \in \mathcal{X}$ be recurrent, and $x \in \mathcal{X}$ s.t. $y \longmapsto x$

We will show that $f_{xy} = 1$

Then $f_{xy}^{(n)} > 0$ for some $n \in \mathbb{N}$.

Lem: Let $x, y \in \mathcal{X}$ in the same commun. class, and the state y is recurrent.

Then $\displaystyle \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P_{xy}^{(k)} = \frac{1}{\tau_{yy}}$ $\quad \displaystyle \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_x N_y(n)$

$$= \frac{1}{\mathbb{E}_y \tau_{yy}^{(1)}}$$

Thm: All states in a commun. class are one of the foll. types

(i) All transient (ii) All null recurrent (iii) All +ve recurrent

Proof: It suffices to show that $x, y \in \mathcal{C}$ and y is null recurrent then x is null recurrent.

$$\exists\ r, s \in \mathbb{N} \text{ s.t. } P_{yx}^{(r)} P_{xy}^{(s)} > 0$$

$$\sum_{l=1}^{n-r-s} P_{yy}^{(r+l+s)} \ge \sum_{l=1}^{n-r-s} P_{yx}^{(r)} P_{xx}^{(l)} P_{xy}^{(s)} > 0$$

$$\frac{1}{n} \sum_{k=1}^{n} P_{yy}^{(k)} \ge \frac{1}{n} \sum_{k=r+s+1}^{n} P_{yy}^{(k)} \ge \left(\frac{n-r-s}{n}\right) \left[\frac{1}{n-r-s} \sum_{l=1}^{n-r-s} P_{xx}^{(l)} \right] P_{yx}^{(r)} P_{xy}^{(s)}$$

$$0 = \limsup_{n\to\infty} \sum_{k=1}^{n} P_{yy}^{(k)} \ge \limsup_{n\to\infty} \left[\frac{1}{n} \sum_{l=1}^{n} P_{xx}^{(l)} \right] \ge 0$$

$\boxed{\text{E}}$

Lecture 23: DTMC: Invariant Distribution

1. Invariant Distribution

Let $X: \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ be a time homogeneous MC with transition probability matrix $P: \mathcal{X} \times \mathcal{X} \to [0,1]$

Defn: For any countable set $\mathcal{X}$, we define set of pmf over $\mathcal{X}$ as $\left\{ \mathcal{M}(\mathcal{X}) = \nu \in [0,1]^{\mathcal{X}} : \sum_{x \in \mathcal{X}} \nu_x = 1 \right\}$

Defn: A prob. distribution $\pi \in \mathcal{M}(\mathcal{X})$ is said to be "invariant distribution" of the MC X: if it satisfies the global balance equation

$$\pi = \pi P$$

Defn: When the initial dist. of a MC is $\nu \in \mathcal{M}(\mathcal{X})$ then the cond. prob. is denoted by

$$P_\nu : \mathcal{F} \to [0,1]$$

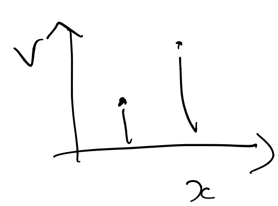
$$A \mapsto P_\nu(A) = \sum_{x \in \mathcal{X}} \nu(x) P_x(A)$$

Ex: $\mathcal{X} = \{0, 1\}$ $\quad \mathcal{M}(\mathcal{X}) = \{ (p, 1-p) : p \in [0,1) \}$

$$= \left\{ \nu \in [0,1)^{\{0,1\}} : \nu(0) + \nu(1) = 1 \right\}$$

$$P_x(A) = P(A(\{X_0 = x\}))$$

$$P_\nu(A) = \sum_{x \in \mathcal{X}} \nu(x) P_x(A)$$



Defn: For a MC $X: \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we denote the dist. of $X_n : \Omega \to \mathcal{X}$ by $\nu_n \in \mathcal{M}(\mathcal{X})$.

That is, $v_n(x) = P\{X_n = x\}$, $\forall x \in X$

$$= P_{v_0}\{X_n = x\}$$

$$= \sum_{y \in X} P(\{X_n = x\} | \{X_0 = y\}) \, P(\{X_0 = y\})$$

$$= \sum_{y \in X} v_0(y) \, P_y\{X_n = x\}$$

$$= P_{v_0}\{X_n = x\}$$

Remark: 1) $v_n(x) = \sum_{z \in X} v_0(z) \underbrace{P_{zx}^{(n)}}_{[P^n]_{zx}}$

Remarks: 1. Global balance equation is a set of $|X|$ equations

$$\pi_y = \sum_{x \in X} \pi_x P_{xy} \quad \forall y \in X$$

$\underline{\pi \in \mathcal{M}(X)}$ s.t. $\pi = \pi P$

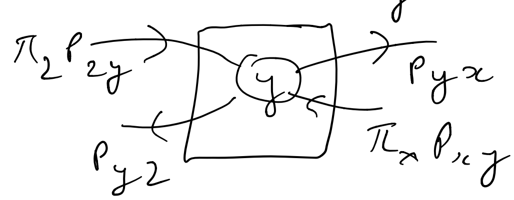
$[\pi_y := \pi(y)]$ $\qquad \pi_y = \sum_{x \in X} \pi_x P_{xy}$

$\pi \in \mathcal{M}(X)$

$\pi : X \longmapsto [0,1]$

$\quad y \longmapsto \pi_y \in [0,1)$

ii)



$$\pi_y(1 - P_{yy})$$

$$\le \pi_y \sum_{x \ne y} P_{yx}$$

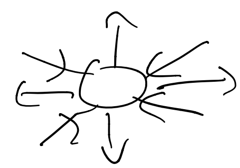
$$= \sum_{x \ne y} \pi_x P_{xy}$$

$$\pi_y = \sum_{x \in X} \pi_x P_{xy} = \pi_y P_{yy} + \sum_{x \neq y} \pi_x P_{xy}$$

$$\pi_y (1 - P_{yy}) = \sum_{x \neq y} \pi_x P_{xy}$$

$$y^n \begin{bmatrix} \\ P_{y1} \quad P_{y2} \end{bmatrix}$$

$$\pi_y \left(\sum_{x \neq y} P_{yx} \right) = \sum_{x \neq y} \pi_x P_{xy}$$



leaving

$$\left[\sum_{x \neq y} \pi_y P_{yx} = \sum_{x \neq y} \pi_x P_{xy} \right]$$

(iii) $\boxed{\pi = \pi P}$ Invariant distribution π is a left eigenvector of P with "largest eigenvalue" 1.

$$P \mathbb{1}^T = \mathbb{1}^T \qquad \frac{1}{|X|} (1 \; 1 \; 1 \; 1 \; 1)^T$$

If P is doubly stochastic then $\mathbb{1}$ is also a left eigenvector $\mathbb{1} P^T = \mathbb{1}$

iv $\quad v_n = v_0 P^n \qquad [\pi = \pi P]$

$$[\pi = \pi P = \pi P^2 \ldots \ldots \pi P^n]$$

$$\pi = \pi P^n \quad \forall n \in \mathbb{N}$$

If, $v_0 = \pi$, then $v_n = \pi$, $\forall n \in \mathbb{Z}_+$

v $\quad$ If $v_0 = \pi$, then MC is stationary.

$$P_\pi \{X_0 = x_0, \ldots X_n = x_n\} = \pi_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1} x_n}$$

$$P_\pi \{X_k = x_0, \ldots X_{k+n} = x_n\} = P_\pi\{X_k = x_0\} P_{x_0 x_1} \ldots P_{x_{n-1} x_n}$$

vi. If the MC is irreducible with $\pi_x > 0$ for some $x \in \mathcal{X}$. Then for any $y \in \mathcal{X}$, $P^{(m)}_{xy} > 0$ for some $m \in \mathbb{N}$ $\quad \pi_y \ge \pi_x P^{(m)}_{xy} > 0$

That is, the entire invariant vector is positive.

vii Any scaled version of π satisfies the gbe.

Therefore, for any $\alpha : \mathcal{X} \to \mathbb{R}_+$ s.t. $\alpha = \alpha P$

and $\|\alpha\|_1 = \sum_{x \in \mathcal{X}} \alpha < \infty$

then $\frac{1}{\|\alpha\|_1} \alpha$ is an invariant dist. of P.

Theorem: An irreducible MC with transition matrix P is positive recurrent iff there exists a unique invariant $\pi \in \mathcal{M}(\mathcal{X})$ that satisfies gbe and

$$\pi_x = \frac{1}{\mu_{xx}} > 0 \quad \forall x \in \mathcal{X}.$$

$\mu_{xx} = \mathbb{E}_x \tau_x^{(1)}$

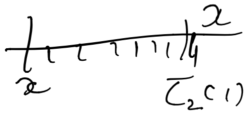
Proof ($\Rightarrow$) We assume that X is irreducible

Let $x_0 = x$

$$N_y(n) = \sum_{k=1}^{n} \mathbb{1}_{\{x_k = y\}}$$

$$\sum_{y \in \mathcal{X}} N_y(n) = n \quad \forall n \in \mathbb{N}$$

$$\begin{cases} N_x(\tau_x^{(1)}) = 1 \\ \sum_{y \in \mathcal{X}} N_y(\tau_x^{(1)}) = \tau_x^{(1)} \end{cases}$$



Existence of $\pi \in \mathcal{M}(\mathcal{X})$:

$$\nu_y := \mathbb{E}_x\left[N_y(\tau_x^{(1)})\right] \text{ for } y \in \mathcal{X}$$

mean # visits to state y
first recurrence to state x.

1. $\nu_y \geq 0$
2. $\nu_x = 1$
3. $\sum_{y \in \mathcal{X}} \nu_y = \mathbb{E}_x \tau_x^{(1)} = \mu_{xx} < \infty$ (pos. recurr.)

Claim: $\nu : \mathcal{X} \mapsto \mathbb{R}_+$ satisfies GBE

Then $\pi : \frac{\nu}{\sum_{y \in \mathcal{X}} \nu_y}$ is an invariant dist. of P

Proof: $\nu_y = \sum_{x \in \mathcal{X}} \nu_x P_{xy}$ (we need to show)

$$\nu_y = \mathbb{E}_x N_y(\tau_x^{(1)}) = \mathbb{E}_x \sum_{n=1}^{\tau_x^{(1)}} \mathbb{1}_{\{X_n = y\}}$$

$$= \mathbb{E}_x \sum_{n \in \mathbb{N}} \mathbb{1}_{\{X_n = y,\, n \leq \tau_x^{(1)}\}}$$

$$\underset{\text{MCT}}{=} \sum_{n \in \mathbb{N}} \underbrace{P_x\{X_n = y,\, n \leq \tau_x^{(1)}\}}_{\lambda_{xy}^{(n)}}$$

$$\lambda_{xy}^{(1)} = P_x\{X_1 = y,\, 1 \leq \tau_x^{(1)}\} = P_x(X_1 = y)$$

$$= P_{xy},\ y \neq x$$

For $n \geq 2$, $\lambda_{xy}^{(n)} = P_x$ (See from notes the proof.)

Cor. An irreducible MC on a finite space $\mathcal{X}$, has a unique st. distribution π.

Defn. An irreducible, aperiodic, pos. recurrent MC is called "ergodic".

Remarks: 1. $\underset{\substack{\downarrow \\ \pi_1}}{C_1}, \ldots\ldots \underset{\substack{\downarrow \\ \pi_m}}{C_n}$ for $\underline{\underline{X}}$

$$\pi := \sum_{i=1}^{m} \alpha_i \pi_i$$

ii If $v_0 = e_x$, i.e. $X_0 = x$

Then $\pi_y = \frac{1}{\mu_{yy}} = \lim_{n\to\infty} \frac{1}{n} \underbrace{\sum_{k=1}^{n} P_{xy}^{(k)}}_{\mathbb{E}_x N_y(n)}$

π_y = limiting avg # visits to state y

(iii) If X is pos. recurr. & aperiodic,

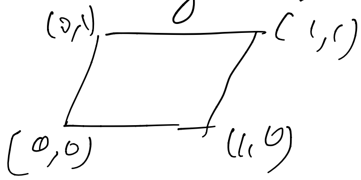
$$\pi_y = \lim_{n\to\infty} P_{xy}^{(n)} = \lim_{n\to\infty} P_x\{X_n = y\}$$

π_y

Lec - 24 : Poisson Processes

1 Simple Point Processes.

Consider the d-dimensional Euclidean space $\mathbb{R}^d$
The collection of Borel measurable subsets $\mathcal{B}(\mathbb{R}^d)$
generated by $B(x) := \{ y \in \mathbb{R}^d : y_i \leq x_i \}$ of $x \in \mathbb{R}^d$



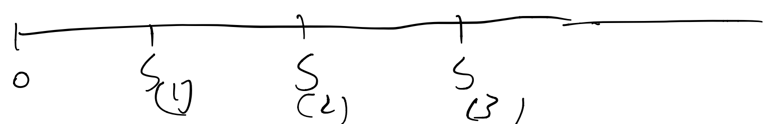
Def: A *simple point processes* is a random countable collection of distinct points $S : \Omega \to (\mathbb{R}^d)^{\mathbb{N}}$ s.t. the distance $\|S_n\| \to \infty$ as $n \to \infty$.

$$\Big[\big(S : \Omega \to (\mathbb{R}^d)^{\mathbb{N}} \qquad \omega \longmapsto (\underset{\substack{\uparrow \\ \mathbb{R}^d}}{S_1(\omega)}, \ldots\ldots \underset{\substack{\uparrow \\ \mathbb{R}^d}}{S_n(\omega)}, \ldots\ldots \big) \Big]$$

Ex. [simple point processes on half-line] $x = \mathbb{R}_+$

$$S : \Omega \to \mathbb{R}_+^{\mathbb{N}}$$

$$\omega \to (\underset{\substack{\uparrow \\ \mathbb{R}^+}}{S_1(\omega)}, \ldots S_n(\omega), \ldots)$$



$S \to \tilde{S} \qquad \tilde{S}_n = S_{(n)} \qquad n^{th}$ order statistics

$S_{(0)} = 0 \qquad S_{(n)} := \Big\{ \inf \big[S_k > S_{(n+1)} : k \in \mathbb{N} \Big\}$

$\lim_{n \to \infty} S_{(n)} = \infty \qquad S(1) < S(2) < \ldots S$

$$\mathcal{B}(\mathbb{R}_+) = \sigma((0, t] : t \in \mathbb{R}_+)$$

Def: For a pt. process $S : \Omega \to x^{\mathbb{N}}$, the no. of pt.s in $A \in \mathcal{B}(x)$ by