EE126

LECTURE 1

(base outcomes)
Sample space (Ω): set of all outcomes
Collectively exhaustive (CE) mutually exclusive (ME)

Events: subsets of Ω

Prob space (Ω, F, P) is a math construct to model experiments.

Ω: set of all possible outcomes

(power set)
F: set of events where each set contains 0 or more base outcomes

P: probability of events

$P: F \mapsto [0,1]$

Properties of P:

(1) $P(\emptyset) = 0$ (2) $P(\Omega) = 1$

(3) $P(A_1 \cup A_2 \cup A_3 \cdots) = P(A_1) + P(A_2) + \cdots$
for disjoint (ME) events $A_1, A_2 \cdots$

AXIOMS

$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ finite additivity

(1) $P(A^c) = 1 - P(A)$

(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Discrete Prob: $P(A) = \sum_{\omega \in A} P(\omega)$

Uniform Sample Space: $P(A) = \frac{|A|}{|\Omega|}$

Union Bound

$$P(A_1 \cup A_2 \cup \cdots A_n) \leq \sum_{i=1}^{n} P(A_i)$$

Inclusion-Exclusion Principle

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{\substack{I \subseteq \{1,\dots,n\} \\ |I| = k}} P\left(\bigcap_{i\in I} A_i\right)\right) = \sum P(A_i) - \sum_{\{i,j\}} P(A_i \cap A_j) + \cdots \pm P\left(\bigcap_{i=1}^{n} A_i\right)$$

Conditional Probability

$P(A|B) = \frac{P(A\cap B)}{P(B)}$ $P(B) \neq 0$

Product Rule

$$P(A_1 \cap A_2 \cap \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots$$

Total Probability Rule

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots P(A_n \cap B)$$
$$= \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

Bayes' Theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum P(A_i)P(B|A_i)}$$

LECTURE 2

Birthday Paradox

$$P(A^c) = \frac{365 \times 364 \cdots 365-(n-1)}{365^n} = 1\left(1-\frac{1}{k}\right)\left(1-\frac{2}{k}\right)\cdots\left(1-\frac{n-1}{k}\right)$$

$\boxed{\text{Taylor's approx: } e^x \approx 1+x}$

$$e^{-\frac{1}{k}} \cdot e^{-\frac{2}{k}} \cdot e^{-\frac{3}{k}} \cdots e^{-\frac{n-1}{k}} = e^{-\frac{1}{k}(1+2+\cdots n-1)}$$

$$= e^{-\frac{1}{k}\frac{n(n-1)}{2}} \approx e^{-\frac{n^2}{2k}}$$

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - e^{-\frac{n^2}{2k}}$$

$k = 365 \quad n = 23 \quad \mathbb{P}(\text{at least a pair}) \approx 50.4\%$

$n = 50 \implies 97\%$

INDEPENDENCE

$$\boxed{\mathbb{P}[A|B] = \mathbb{P}[A]} \qquad \boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]}$$

Disjoint $\not\equiv$ Independence

A, B are disjoint $\Leftrightarrow \mathbb{P}[A \cap B] = 0 \Rightarrow \mathbb{P}[A] = 0$ or $\mathbb{P}[B] = 0$

Conditional Independence

$$\boxed{\mathbb{P}[A \cap B | C] = \mathbb{P}(A|C)\,\mathbb{P}[B|C]}$$

(1) Dep. events can be cond. indep.
(2) Indep. events can be cond. dep.

e.g. HH TT — Pick one at random, flip twice

H_i: ith flip is H

$\mathbb{P}[H_2|H_1] = 1 \neq \mathbb{P}[H_2] = \frac{1}{2}$

$\mathbb{P}[H_1 \cap H_2 | A] = \mathbb{P}[H_1|A] \cdot \mathbb{P}[H_2|A]$

Cond. indep.

A: pick HH coin.

Independence of collection of events

$$\boxed{\mathbb{P}\left[\bigcap_{i\in S} A_i\right] = \prod_{i\in S} \mathbb{P}[A_i]}$$

"Pairwise" indep $\not\Rightarrow$ Joint indep of 3 or more events

Ex. Roll 2 6-sided dice, $\mathbb{P}[6 \text{ before } 7]$?

E: 6 before 7 $\quad$ S: first roll is 6

T: first roll is 7

Condition on first roll

$$\mathbb{P}[E] = \underbrace{\mathbb{P}[E \mid S]}_{1}\underbrace{\mathbb{P}[S]}_{5/36} + \underbrace{\mathbb{P}[E \mid T]}_{0}\mathbb{P}[T] + \underbrace{\mathbb{P}[E \mid (S\cup T)^c]}_{\mathbb{P}[E]}\underbrace{\mathbb{P}[(S\cup T)^c]}_{\frac{25}{36}}$$

$$\mathbb{P}[E] = \frac{5}{36} + \frac{25}{36}\mathbb{P}[E] \Rightarrow \boxed{\mathbb{P}[E] = \frac{5}{11}}$$

Random Variables

RV : $\Omega \mapsto \mathbb{R}$

$\mathbb{P}_X(x) = \mathbb{P}(X = x)$ $\quad \sum_x p_X(x) = 1$

Functions of RVs are RVs

$Y = g(X)$ $\quad \boxed{\mathbb{P}_Y(y) = \sum_{\{x \mid g(x) = y\}} \mathbb{P}_X(x)}$

e.g. $Y = |X|$; X is uniform $[-2, 2]$ $p = \frac{1}{5}$

Expectation

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\,\mathbb{P}[X = x] = \sum x\, p_X(x)$$

$$\boxed{\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)}$$

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ $\quad \mathbb{E}[g(x)] = \sum_x g(x)\, p_X(x)$

$\mathbb{E}[cX] = c\,\mathbb{E}[X]$

Tail Sum Formula

$$\boxed{\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}[X \geq k]}$$

Pf

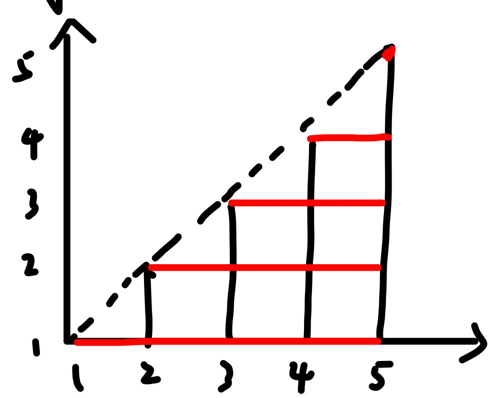
$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i\, \mathbb{P}[X=i]$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{i} 1\, \mathbb{P}[X=i]$$

$$= \sum_{j=1}^{\infty} \underbrace{\sum_{i=j}^{\infty} \mathbb{P}[X=i]}$$

$$= \sum_{j=1}^{\infty} \mathbb{P}[X \geq j] \qquad \blacksquare$$

$$\sum_{j=1}^{\infty} \mathbb{P}[X \geq j]$$



Variance

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

Popular RV's

① **Uniform**

$$p_X(k) = \frac{1}{n} \qquad \mathbb{E}[X] = \frac{n+1}{2} \qquad Var(X) = \frac{n^2-1}{12}$$

---

② **Bernoulli**

$$p_X(k) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \end{cases}$$

$$\mathbb{E}[X] = p \qquad Var(X) = p(1-p)$$

---

③ **Indicator RV**

$$X = \{1\}_A = \begin{cases} 1 & \text{if } A \text{ True} \\ 0 & \text{else} \end{cases}$$

$$\mathbb{E}[X] = \sum_x x P[X=x] = P(A)$$

---

④ **Binomial**

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \binom{n}{k} p^k (1-p)^{n-k} = np \qquad Var(X) = np(1-p)$$

---

⑤ **Geometric**

$$p_X(k) = (1-p)^{k-1} p$$

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

$$\text{CDF: } P(X \leq k) = \sum_{i=1}^{k} (1-p)^{i-1} p = 1 - (1-p)^k$$

$$\Downarrow$$

$$P(X > k) = 1 - (1 - (1-p)^k) = (1-p)^k$$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

LECTURE 4

Poisson (λ)

Limit of Bin(n,p) $n \longrightarrow \infty$ $p \longrightarrow 0$ $np = \lambda$

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad n >> k$$

$$= \underbrace{\frac{n!}{k!(n-k)!}} p^k (\underbrace{1-p})^{n-k} \qquad 1-p \approx e^{-p}$$

$$= \frac{n^k}{k!} p^k e^{-pn}$$

$$= \frac{(np)^k e^{-pn}}{k!}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!} \qquad \sum_{k=0}^{\infty} P(X=k) = 1$$

$$\sum_{k\geq 0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

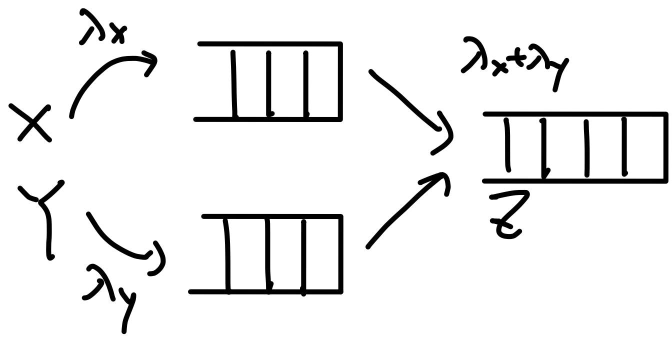
$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} = \boxed{\lambda}$$

$$\boxed{\mathrm{Var}(X) = \lambda}$$

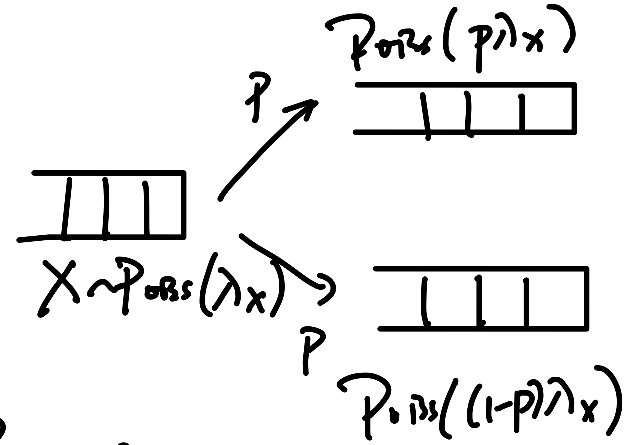
Thm: $X \sim \text{Poiss}(\lambda_x)$, $Y \sim \text{Poiss}(\lambda_y)$

X, Y indep. $\Rightarrow X+Y \sim \text{Poisson}(\lambda_x + \lambda_y)$

MERGING



SPLITTING



Poisson Recursion

$$\boxed{\mathbb{E}[X^n] = \lambda \mathbb{E}\left[(X+1)^{n-1}\right]}$$

Proof:

$$\mathbb{E}[X^n] = \sum_{x=0}^{\infty} x^n p_X(x)$$

$$= \sum_{x=0}^{\infty} x^n \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \lambda \sum_{x=0}^{\infty} x^{n-1} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}$$

$$= \lambda \sum_{k=-1}^{\infty} (k+1)^{n-1} \frac{\lambda^k e^{-\lambda}}{k!}$$

$$= \lambda \mathbb{E}\left[(X+1)^{n-1}\right] \quad \blacksquare$$

Thm: X, Y indep $\overset{*}{\Longrightarrow}$ $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

Proof:

$$\mathbb{E}[XY] = \sum_x \sum_y xy\, p_{XY}(x,y)$$

$$= \sum_x \sum_y xy\, p_X(x)\, p_Y(y)$$

$$= \sum_x x\, p_X(x) \sum_y y\, p_Y(y)$$

$$= \mathbb{E}[X]\,\mathbb{E}[Y]$$

■

X, Y indep. $\Longrightarrow$ $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Proof:

WLOG, consider $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$.

$$\mathrm{Var}(X) = \mathbb{E}[X^2] = \mathbb{E}[(X_1+X_2)^2]$$

$$= \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + 2\cancel{\mathbb{E}[X_1X_2]} \quad 0$$

$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

■

Note:
Can use to prove $\mathrm{Var}(X)$ of $\mathrm{Bin}(n,p)$.

$$\boxed{Cov(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]}$$

$X = \sum_i X_i$, each X_i indep. $\Rightarrow Var(X) = \sum_i Var(X_i)$

* X_1, X_2 not indep. $\Rightarrow$

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2\underbrace{\left(\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]\right)}_{Cov(X_1, X_2)}$$

$$Cov(X_1, X_2) = \mathbb{E}\left[\left(X - \mathbb{E}[X_1]\right)\left(X_2 - \mathbb{E}[X_2]\right)\right]$$

$$\boxed{\rho = \frac{cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}} \qquad -1 \leq \rho \leq 1$$

Problem 21 p.123

St Peterburg — Toss fair coin until H.

n tosses $\Rightarrow$ get \$$2^n$

Q) How much to pay to play this game?

W = amount

$$\mathbb{E}[W] = \sum_{k=0}^{\infty} \left(\frac{1}{2^k}\right) 2^k = \infty$$

Conditioning on RVs

$$p_{X|Y}(k|n) = P(X=k \mid Y=n)$$

$$p_{X|Y}(k|n) \geq 0 \qquad \sum_x p_{X|Y}(x|y) = 1$$

Memoryless Property of Geo(p)

$$\boxed{\mathbb{P}(X > s+t \mid X > s) = \mathbb{P}(X > t)}$$

Proof:

$$\mathbb{P}(X > s+t \mid X > s) = \frac{\mathbb{P}(X > s+t \cap X > s)}{\mathbb{P}(X > s)}$$

$$= \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)}$$

$$= \frac{(1-p)^{s+t}}{(1-p)^s}$$

$$= (1-p)^t$$

$$= \mathbb{P}(X > t)$$

$\blacksquare$

$$\boxed{\mathbb{E}[g(X) \mid X > 1] = \mathbb{E}[g(1+X)]}$$

Proof:

$$\mathbb{E}[g(X) \mid X > 1] = \sum_{k=1}^{\infty} g(k)\,\mathbb{P}(X = k \mid X > 1)$$

$$= \sum_{k=1}^{\infty} g(k)\,\mathbb{P}(X = k-1) \quad \text{(Memoryless)}$$

$$= \sum_{n=1}^{\infty} g(1+n)\,\mathbb{P}(X = n) \quad (n = k-1)$$

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

$\blacksquare$

Cool Way to calculate $\mathbb{E}[X] = \frac{1}{p}$ for Geo(p)

$$\mathbb{E}[X] = \mathbb{E}[X \mid X=1]\,P(X=1) + \mathbb{E}[X \mid X>1]\,P(X>1)$$

$$= 1 \cdot p + \underbrace{\left(1 + \mathbb{E}[X]\right)}_{\text{Memoryless}} \cdot (1-p)$$

$$\boxed{\mathbb{E}[X] = \frac{1}{p}}$$

Conditional Expectation

$$\boxed{\mathbb{E}[X \mid A] = \sum_{x} x\, p_{X|A}(x)}$$

Total Expectation Theorem

$$\boxed{\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{E}[X \mid A_i]}$$

$$\boxed{\mathbb{E}[X \mid B] = \sum_{i=1}^{n} \mathbb{P}(A_i \mid B)\,\mathbb{E}[X \mid A_i \cap B]}$$

$$\boxed{\mathbb{E}[X] = \sum_{y} p_Y(y)\,\mathbb{E}[X \mid Y=y]}$$

Law of Iterated Expectation

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

Proof:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_y \mathbb{E}[X|Y=y]\,\mathbb{P}[Y=y] \\
&= \sum_y \sum_x x\,\underbrace{\mathbb{P}[X|Y=y]\,\mathbb{P}[Y=y]}_{\mathbb{P}[X=x,\,Y=y]} \\
&= \sum_x x \sum_y \mathbb{P}(X=x, Y=y) \\
&= \sum_x x\,\mathbb{P}[X=x] \\
&= \mathbb{E}[X]
\end{aligned}$$

■

CONTINUOUS PROB.

$$\mathbb{P}(X \in B) = \int_B f_X(x)\,dx \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

$$\boxed{\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,dx}$$

$$\mathbb{P}(X < a) = \mathbb{P}(X \le a)$$

$$\mathbb{P}(X \in [x, x+\delta]) \approx f_X(x)\,\delta$$

C.D.F

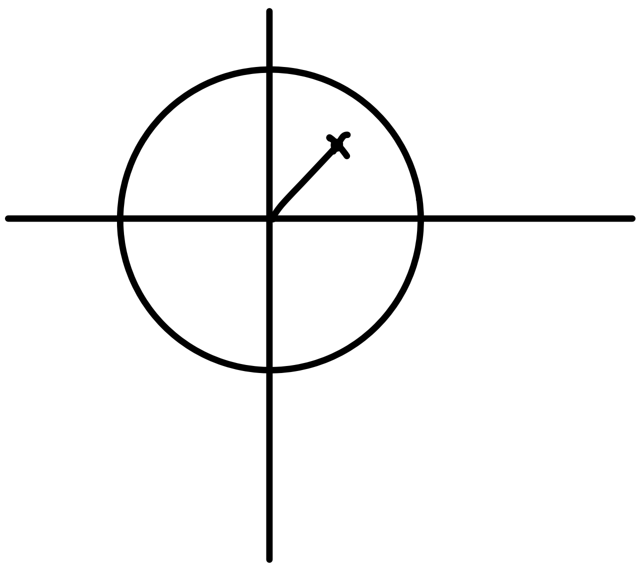
$$\boxed{F_X(x) = \mathbb{P}(X \le x)}$$

(i) $F_X(\infty) = 1$; $F_X(-\infty) = 0$

(ii) $\mathbb{P}(X = k) = F_X(k) - F_X(k-1)$

(iii) $f_X(x) = \dfrac{dF_X(x)}{dx}$

Random Dart Throw



Y: dist. from (0,0)

Find pdf, cdf of Y.

Sol:

$$\mathbb{P}(Y \le y) = \frac{\pi (y)^2}{\pi (1)^2} = \boxed{y^2} \quad y \in (0,1)$$

$$\frac{dF_y(y)}{dy} = \boxed{2y}$$

EXPECTATION (CONTINUOUS)

$$\boxed{\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx}$$

$$\boxed{\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx}$$

$$\boxed{\mathbb{E}[g(X,Y) \mid Y=y] = \int g(x,y) f_{X|Y}(x|y)\,dx}$$

$$\boxed{\mathbb{E}[g(X,Y)] = \int \mathbb{E}[g(X,Y) \mid Y=y] f_Y(y)\,dy}$$

X, Y independent $\iff F_{XY}(x,y) = F_X(x) F_Y(y)$

Popular Continuous Distribution

① Uniform $X \sim U[a,b]$

$$f_X(x) = \frac{1}{b-a}$$

$$\mathbb{E}[X] = \int_a^b x \cdot \frac{1}{b-a}\, dx = \boxed{\frac{a+b}{2}}$$

$$\mathrm{Var}(X) = \boxed{\frac{(b-a)^2}{12}}$$

---

② Exponential $X \sim \mathrm{Expo}(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x} \; ; \; x > 0$$

$$\mathbb{P}(X \le x) = F_X(x) = 1 - e^{-\lambda x} \; ; \; x > 0$$

$$\mathbb{P}(X > x) = e^{-\lambda x} \; ; \; x > 0$$

$$\mathbb{E}[X] = \boxed{\frac{1}{\lambda}} \qquad \mathrm{Var}(X) = \boxed{\frac{1}{\lambda^2}}$$

MEMORYLESS PROP. of Expo(λ)

$$\mathbb{P}(X > t+s \mid X > t) = \mathbb{P}(X > s) \qquad s,t > 0$$

Proof:

$$\mathbb{P}(X > t+s \mid X > t) = \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > t)}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}}$$

$$= e^{-\lambda s}$$

$$= \mathbb{P}(X > s)$$

$\blacksquare$

NORMAL DISTRIBUTION

$$X \sim N(\mu, \sigma^2)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \qquad F_X(x) = \Phi(x)$$

$$\left(\int_{-\infty}^{\infty} f_X(x)\,dx\right)^2 = 1$$

$$\left(\int_{-\infty}^{\infty} f_X(x)\,dx\right)^2 = \left(\int_{-\infty}^{\infty} f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} f_Y(x)\,dy\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x) f_Y(y)\,dx\,dy$$

$x^2+y^2=r^2$

$dy\,dx = r\,dr\,d\theta$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right)\cdot\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\right)dy\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\,dy\,dx$$

$$= \int_{0}^{2\pi}\int_{0}^{\infty} \frac{1}{2\pi} e^{-\frac{r^2}{2}}\,r\,dr\,d\theta$$

$$= \int_{0}^{\infty} e^{-\frac{r^2}{2}}\,r\,dr$$

$u = \frac{r^2}{2}$

$du = r\,dr$

$$= \int_{0}^{\infty} e^{-u}\,du$$

$$= -e^{-u}\Big|_{0}^{\infty} = 1$$

$\blacksquare$

Sum of ⫫ Normal RVs is Normal

$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ $Z = X + Y$

$$Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Sum of dependent Normal ≠ Normal

$$X \sim N(0,1)$$

$$Y = \begin{cases} X & p: \frac{1}{2} \\ -X & p: \frac{1}{2} \end{cases}$$

$Z = X + Y$ not Normal.

$X \sim N(\mu, \sigma^2)$, $Y = aX \Rightarrow Y \sim N(a \cdot \mu, a^2 \cdot \sigma^2)$

$$X \sim N(\mu, \sigma^2)$$

$$Z = \frac{X - \mu}{\sigma} \qquad Z \sim N(0,1)$$

Male height: $N(70,5)$ $\qquad$ $\mathbb{P}(M>F)$?

Female ... : $N(64,4)$

<u>Sol</u>:

$$\mathbb{P}(X>Y) = \mathbb{P}(X+(-Y)>0)$$
$$= \mathbb{P}(N(70,5)+(-N(64,4))>0)$$
$$= \mathbb{P}(N(70,5)+N(-64,4)>0)$$
$$= \mathbb{P}(N(6,9)>0)$$
$$= \mathbb{P}(N(0,9)>-6)$$
$$= \mathbb{P}\left(N(0,1)>-\frac{6}{3}\right)$$
$$= \mathbb{P}(N(0,1)<2)$$
$$= \boxed{\Phi(2) \approx 0.977}$$

$$\frac{X-\mu}{\sigma} > \frac{x-\mu}{\sigma}$$

JOINT PDFs

Discrete PMF: $p_{X,Y}(x,y)$

$$\boxed{\mathbb{P}((X,Y)\in A) = \sum_{(x,y)\in A} p_{X,Y}(x,y)}$$

Continuous PDF: $f_{X,Y}(x,y)$

$$\boxed{\mathbb{P}((X,Y)\in A) = \int_A f_{X,Y}(x,y)\,dx\,dy}$$

$$\boxed{F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y)}$$

$$\boxed{f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)}$$

MARGINAL PDF

Discrete:

$$\boxed{p_X(x) = \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y)}$$

Continuous:

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy}$$

CONDITIONAL PDF

Discrete:

$$\boxed{p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}}$$

Continuous:

$$\boxed{f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}}$$

INDEPENDENCE

$$f_{X,Y}(x,y) = f_X(x)\,f_Y(y) \qquad\qquad f_{X|Y}(x|y) = f_X(x)$$

BAYES RULE

Discrete:

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)\,p_X(x)}{p_Y(y)}$$

$$= \boxed{\frac{p_{Y|X}(y|x)\,p_X(x)}{\sum_{x' \in X} p_{Y|X}(y|x')\,p_X(x')}}$$

Continuous:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\,f_X(x)}{f_Y(y)}$$

$$\boxed{f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\,f_X(x)}{\int_{-\infty}^{\infty} f_X(t)\,f_{Y|X}(y|t)\,dt}}$$

CONDITIONAL EXPECTATION

Discrete:

$$\boxed{\mathbb{E}[Y \mid X = x] = \sum_{y \in \mathcal{Y}} y \cdot p_{Y|X}(y \mid x)}$$

Continuous:

$$\boxed{\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\, dy}$$

Change of Variables.

$$Y = g(X)$$

$$\mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \mathbb{P}(X \in g^{-1}(y))$$

$X \sim U[0,1] \quad Y = 2X$

FALSE!

$$f_Y(y) = \mathbb{P}(\overset{=0}{Y = y}) = \mathbb{P}(\overset{=0}{2X = y}) = f_X(\tfrac{y}{2})$$

$$\int_0^2 f_X(\tfrac{y}{2})\,dy = 2 \neq 1$$

Use CDF instead.

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(2X \le y) = F_X(\tfrac{y}{2})$$

$$f_Y(y) = \frac{d}{dy} F_X(\tfrac{y}{2}) = \frac{1}{2} \cdot f_X(\tfrac{y}{2})$$

Note PDF does not satisfy $f_Y(y) = \mathbb{P}(Y = y)$

$$\begin{aligned} f_Y(y)\,dy &= \mathbb{P}(y \le Y \le y + dy) \\ &= \mathbb{P}(y \le 2X \le y + dy) \\ &= \mathbb{P}\left(\frac{y}{2} \le X \le \frac{y}{2} + \frac{dy}{2}\right) \\ &= f_X(\tfrac{y}{2}) \cdot \frac{dy}{2} \\ &= \frac{1}{2} f_X(\tfrac{y}{2}) \end{aligned}$$

X continuous $\qquad Y = g(X)$

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y)$$

g monotone increasing:

$$\mathbb{P}(g(X) \le y) = \mathbb{P}(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$

$$\mathbb{P}(g(X) \le y)$$

LAW OF TOTAL VARIANCE

$$\boxed{Var(X) = \mathbb{E}[Var(X|Y)] + Var[\mathbb{E}[X|Y]]}$$

E.g. Biased coin n tosses X: # of H's

$Y = P[H] \sim U[0,1]$ $\qquad \mathbb{E}[X] = \frac{n}{2}$

$$Var(X) = Var[\mathbb{E}[X|Y]] + \mathbb{E}[Var[X|Y]]$$

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

$$= Var(nY) + \mathbb{E}[nY(1-Y)]$$

$$= n^2 Var(Y) + n\mathbb{E}[Y] - n\mathbb{E}[Y^2]$$

$$= n^2 \cdot \frac{1}{12} + \frac{n}{2} - \frac{n}{3}$$

$$= \boxed{\frac{n^2}{12} + \frac{n}{6}}$$

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

ORDER STATISTICS

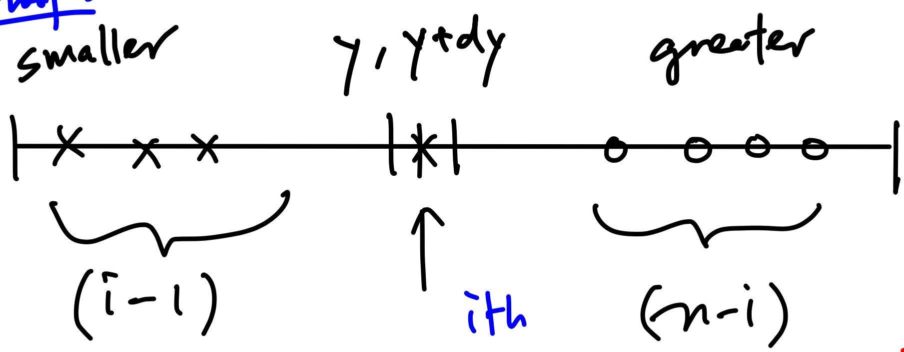
$\mathbb{P}[X_i = X_j] = 0 \quad \forall i,j$

$X^{(i)} = (x^{(i)})$: ith order statistics

Marginal PDF of ith.

$$\boxed{f_{X^{(i)}}(y) = \frac{n!}{(i-1)!\,(n-i)!}\left(F_X(y)\right)^{i-1}\left(1-F_X(y)\right)^{n-i} f_X(y)}$$

Sketch of Proof:



$$\underbrace{\mathbb{P}(X^{(i)} \in \{y, y+dy\})}_{f_{X^{(i)}}(y)\,dy} = \overset{(1)}{\mathbb{P}[(i-1) \text{ elts} < y]} \cdot \overset{(2)}{\mathbb{P}[i \in [y, y+dy]]} \cdot \overset{(3)}{\mathbb{P}[(n-i) \text{ elts} > y+dy]} \cdot \overset{(4)}{(\#\text{ arrangements})}$$

(1): $[F_X(y)]^{i-1}$

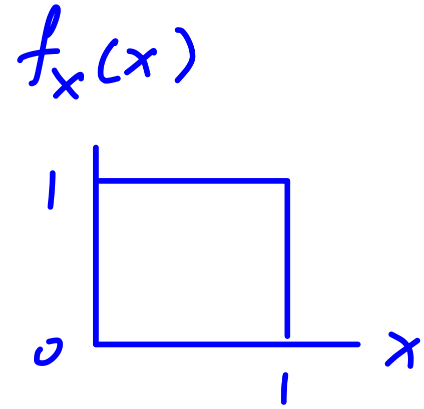
(2): $f_Y(y)\,dy$

(3): $[1-F_X(y)]^{n-i}$

(4): $n\binom{n-1}{i-1}$

Special Case.

$X \sim U[0,1]$



$$\boxed{f_{X^{(i)}}(y) = \frac{n!}{(i-1)(n-i)!}\, y^{i-1}(1-y)^{n-i} \qquad 0<y<1}$$

CONVOLUTION

$Z = X + Y$ continuous, independent.

$$f_Z(z) = \int_x f_{X,Z}(x,z)\,dx = \int_x f_X(x) f_{Z|X}(z|x)\,dx \quad (1)$$

$$F_{Z|X}(z|x) = \mathbb{P}[X+Y \le z \mid X = x]$$

$$= \mathbb{P}[Y \le z - x \mid X = x]$$

$$= \mathbb{P}[Y \le z - x] \quad \text{(independence)}$$

$$= F_Y(z-x)$$

$$f_{Z|X}(z|x) = f_Y(z-x) \quad (2)$$

$$\boxed{f_Z(z) = \int_x f_X(x) f_Y(z-x)\,dx}$$

$$f_Z(z) = (f_X * f_Y)(z)$$

* Overlapping area

Discrete case.

$$Z = X + Y$$

$$\mathbb{P}[Z = n] = \sum_{k} \mathbb{P}(X = k)\,\mathbb{P}(Y = n-k)$$

MOMENT GENERATING FUNCTIONS (MGF's)

$$e^{sX} = 1 + sX + \frac{s^2X^2}{2!} + \cdots$$

MGF (transform)

$$\boxed{M_X(s) = \mathbb{E}[e^{sX}] = 1 + s\,\mathbb{E}[X] + \frac{s^2}{2!}\mathbb{E}[X^2] + \cdots}$$

$$M_X(0) = 1$$

$$\frac{d}{ds}M_X(s) = \mathbb{E}[X] + s\,\mathbb{E}[X^2] + \frac{s^2}{2!}\mathbb{E}[X^3] + \cdots$$

$$M_X'(0) = \mathbb{E}[X]$$

$$\vdots$$

$$\boxed{\left.\frac{d^n M_X(s)}{ds^n}\right|_{s=0} = \mathbb{E}[X^n]}$$

1) Find moments.

2) Convolution operation $\longleftrightarrow$ multiplication

3) Prove CLT.

$$\boxed{M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx}$$

Properties:

(1) $M_X(0) = 1$

(2) $X > 0$, $M_X(-\infty) = 0$; $X < 0$, $M_X(+\infty) = 0$.

* (3) If $Y = aX + b$

$$M_Y(s) = \mathbb{E}[e^{s(aX+b)}]$$
$$= e^{sb}\,\mathbb{E}[e^{asX}]$$
$$= e^{sb} M_X(as)$$

$$\boxed{M_Y(s) = e^{sb} M_X(as)}$$

Normal MGF

* $$\boxed{M_X(s) = e^{s^2/2}}$$ $X \sim N(0,1)$

If $Y \sim N(\mu, \sigma^2)$, $Y = \sigma X + \mu$

$$\boxed{M_Y(s) = e^{s\mu} M_X(\sigma s)}$$

$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$

$$M_{X+Y}(s) = \mathrm{MGF}(N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2))$$

$Z = X + Y$, X, Y independent

$$M_Z(s) = \mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX} \cdot e^{sY}]$$
$$= \mathbb{E}[e^{sX}] \cdot \mathbb{E}[e^{sY}]$$
$$= M_X(s) \cdot M_Y(s)$$

$$\boxed{M_Z(s) = M_X(s) \cdot M_Y(s)}$$

Convolution of densities $\longleftrightarrow$ product of MGFs.

E.g. $X \sim N(0,1)$, $Y \sim N(0,1)$

$$Z = X + Y \Rightarrow M_Z(s) = M_X(s) \cdot M_Y(s)$$
$$= e^{s^2/2} \cdot e^{s^2/2}$$
$$= e^{s^2}$$
$$= \mathrm{MGF}(N(0,2))$$

LIMITING BEHAVIOR of R.V.s

$$M_n = \frac{\sum_i X_i}{n} \quad \text{sample mean.}$$

① $$\mathbb{E}[M_n] = \frac{n\,\mathbb{E}[X_i]}{n} = \mathbb{E}[X_i] = \mu.$$

② $$\mathrm{Var}(M_n) = \frac{1}{n^2}\Big[\sum_i \mathrm{Var}(X_i)\Big] = \frac{\mathrm{Var}(X_i)}{n}$$

As $n \to \infty$, $\mathbb{E}[M_n] = \mathbb{E}[X_i] = \mu.$

$$\mathrm{Var}(M_n) = 0$$

Markov Bound

$$\boxed{\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}}$$

Proof:

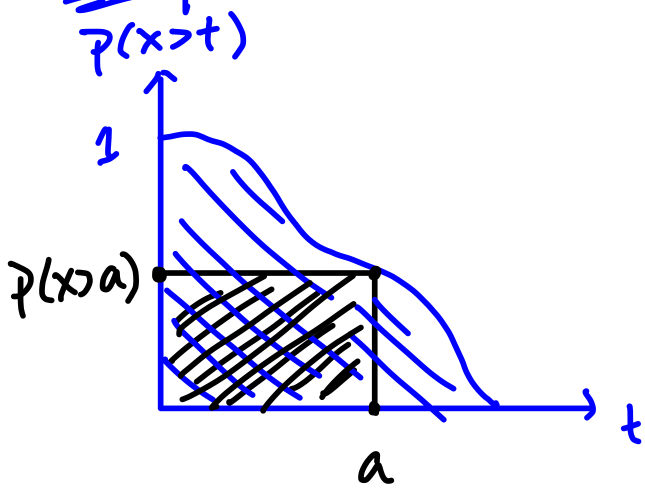
$$\{1\}_{X>a} \leq \frac{X}{a}$$

$$\mathbb{E}\big[\{1\}_{X>a}\big] \leq \frac{\mathbb{E}[X]}{a}$$

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}.$$

■

Alt. proof:



Tail Sum: $X \geq 0$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} P(X>t)\,dt.$$

$$\text{Area} = P(X>a)\cdot a \leq \mathbb{E}[X] \quad \blacksquare$$

Chebyshev Bound

$$\boxed{P(|X-\mathbb{E}[X]| \geq c] \leq \frac{\mathrm{Var}(X)}{c^2}} \quad \forall c>0$$

$$\boxed{P(|X-\mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}}$$

Proof:

$$P(|X-\mathbb{E}[X]| \geq c] = P(|X-\mathbb{E}[X]|^2 \geq c^2)$$

$$\leq \frac{\mathbb{E}[|X-\mathbb{E}[X]|^2]}{c^2}$$

$$= \frac{\mathrm{Var}(X)}{c^2} \quad \blacksquare$$

Chernoff Bound

$X = e^{sY} \Rightarrow P(e^{sY} \geq a) \leq \dfrac{\mathbb{E}[e^{sY}]}{a}$

$a = e^{sb} \Rightarrow s > 0, \; P(Y > b) = P(e^{sY} > e^{sb})$

$$\leq \frac{\mathbb{E}[e^{sY}]}{e^{sb}}$$

$s > 0$

$$\boxed{\begin{aligned} P(Y \geq b) &\leq e^{-sb} M_Y(s) \\ &\leq \min_{s>0} e^{-sb} M_Y(s) \end{aligned}}$$

$s < 0$

$$\boxed{P(Y \leq b) \leq \min_{s<0} e^{-sb} M_Y(s)}$$

WEAK LAW OF LARGE NUMBERS

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (X_i \text{ i.i.d})$$

$$\mathbb{E}[M_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\mathbb{E}[X_i]\,n = \mu.$$

$$Var(M_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

As $n \to \infty$ $\quad \mathbb{E}[M_n] = \mu \quad Var(M_n) = 0$

$$\boxed{P\left(\left|\frac{X_1 + \cdots X_n}{n} - \mu\right| \geq \epsilon\right) \longrightarrow 0 \qquad n \to \infty}$$

$$P(|M_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon} \longrightarrow 0 \qquad n \longrightarrow \infty$$

$$\Downarrow$$

$$\lim_{n\to\infty} P(|M_n - \mu| \geq \epsilon) = 0$$

$$P(|M_n - \mu| \geq \epsilon) \leq \delta \qquad n > n_0(\epsilon, \delta)$$

SLLN

$$\boxed{P\left(\lim_{n\to\infty} M_n = \mu\right) = 1}$$

CENTRAL LIMIT THEOREM

$$\boxed{\lim_{n\to\infty} P(\hat{S}_n \le x) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt}$$

Proof

$$S_n = \sum_{i=1}^{n} X_i \Rightarrow \mathbb{E}[S_n] = n\mu \qquad Var(S_n) = n\sigma^2$$

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathbb{E}[M_n] = \mu \qquad Var(M_n) = \frac{\sigma^2}{n}$$

$$\hat{S}_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \Rightarrow \mathbb{E}[\hat{S}_n] = 0 \qquad Var(\hat{S}_n) = 1$$

$$Z_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \qquad \mathbb{E}[X_i] = 0, \quad Var(X_i) = 1$$

$$\text{If } Y \sim N(0,1) \Rightarrow M_Y(s) = e^{s^2/2}$$

$$\ln M_Y(s) = s^2/2 \qquad (1)$$

$$M_{Z_n}(s) = \mathbb{E}[e^{sZ_n}] = \mathbb{E}\left[e^{\frac{s}{\sqrt{n}}\cdot\sum_{i=1}^{n} X_i}\right]$$

$$= \mathbb{E}\left[e^{\left(\frac{s}{\sqrt{n}}\right)\cdot X_1} e^{\left(\frac{s}{\sqrt{n}}\right)\cdot X_2} \cdots e^{\left(\frac{s}{\sqrt{n}}\right)\cdot X_n}\right]$$

$$= \mathbb{E}\left[e^{\left(\frac{s}{\sqrt{n}}\right)\cdot X_1}\right]^n$$

$$= \left[M_X\left(\frac{s}{\sqrt{n}}\right)\right]^n \qquad (2)$$

$M_X(0) = 1$

$M_X'(0) = \mathbb{E}[X] = 0$

$M_X''(0) = \mathbb{E}[X^2] = 1$

$$\lim_{n\to\infty} \ln M_{Z_n}(s) = \lim_{n\to\infty}\left[n \ln M_X\left(\frac{s}{\sqrt{n}}\right)\right] = \lim_{n\to\infty}\left[\frac{\ln M_X\left(\frac{s}{\sqrt{n}}\right)}{\frac{1}{n}}\right]$$

$$= \lim_{y\to 0} \frac{\ln[M_X(sy)]}{y^2} = \lim_{y\to 0} \frac{s M_X'(sy)}{2y\,\underbrace{\cancel{M_X(sy)}}_{1}} = \lim_{y\to 0} \frac{s^2 \overbrace{\cancel{M_X''(sy)}}^{1}}{2} = \frac{s^2}{2}$$

■

INFORMATION THEORY

10.2 Information Theory

This entire field was born with Claude Shannon's 1948 paper *A mathematical theory of communication*, which was actually rejected from the publishing journal Shannon sent it to for not being rigorous enough. The reviewer of the paper remarked 30 years later that, "One of my biggest regrets was rejecting that paper."

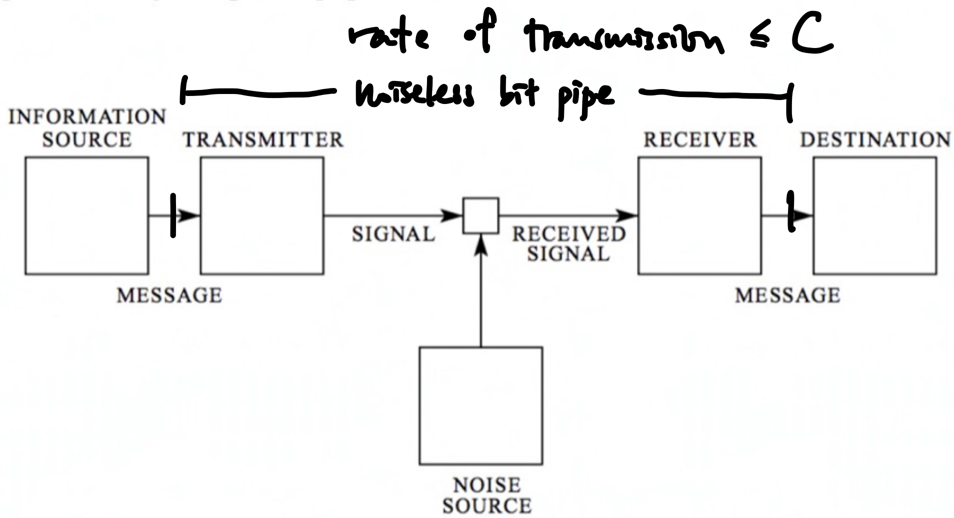
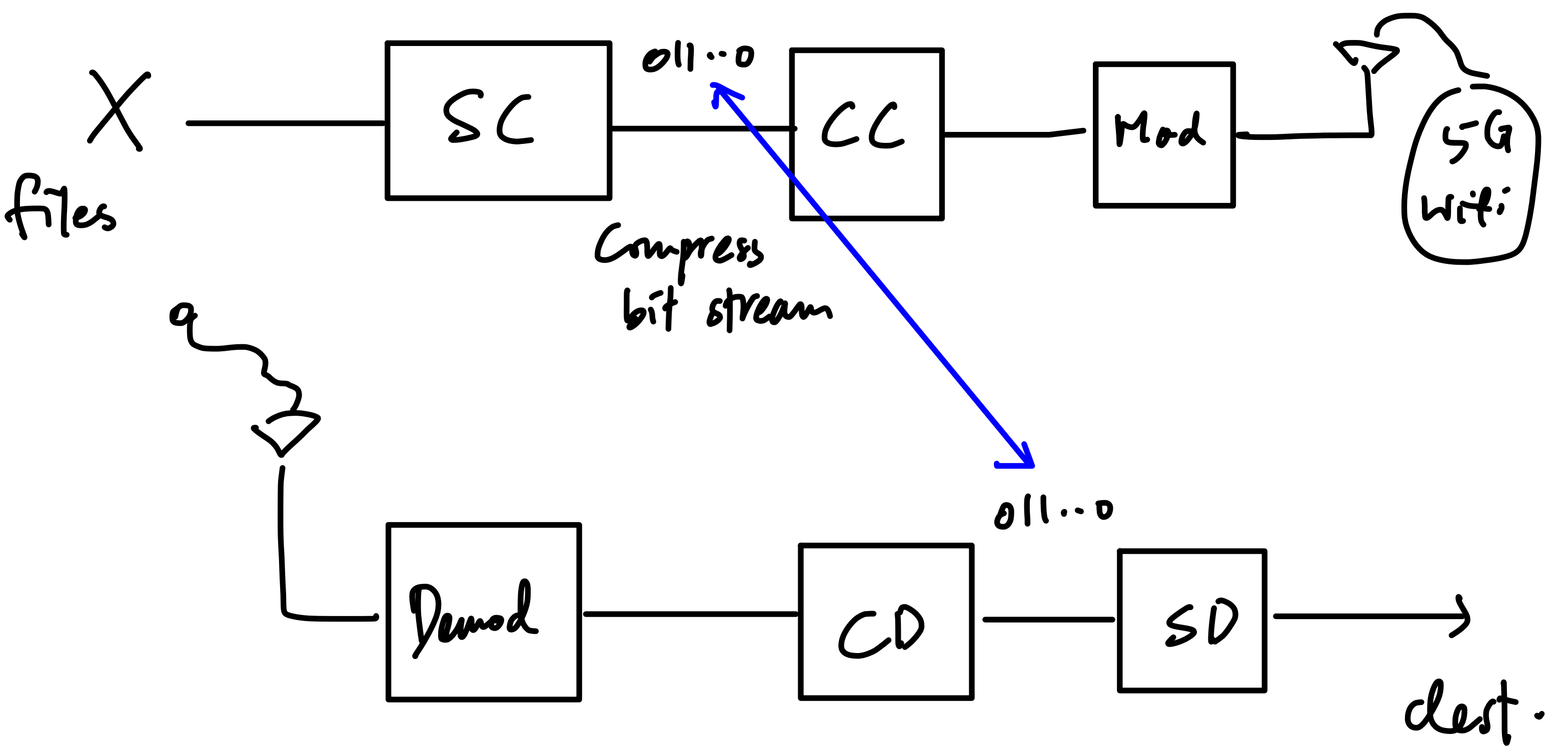


Fig. 1—Schematic diagram of a general communication system.

Figure 3: Shannon's proposed communication framework (Shannon 1948).



Separation Theorem

1) Compression (Source Coding) Theorem

N i.i.d R.V's each having entropy

$H(X)$ compressed into no more than

$N(H(X)+\epsilon)$ bits $\forall \epsilon > 0$ as $N \to \infty$.

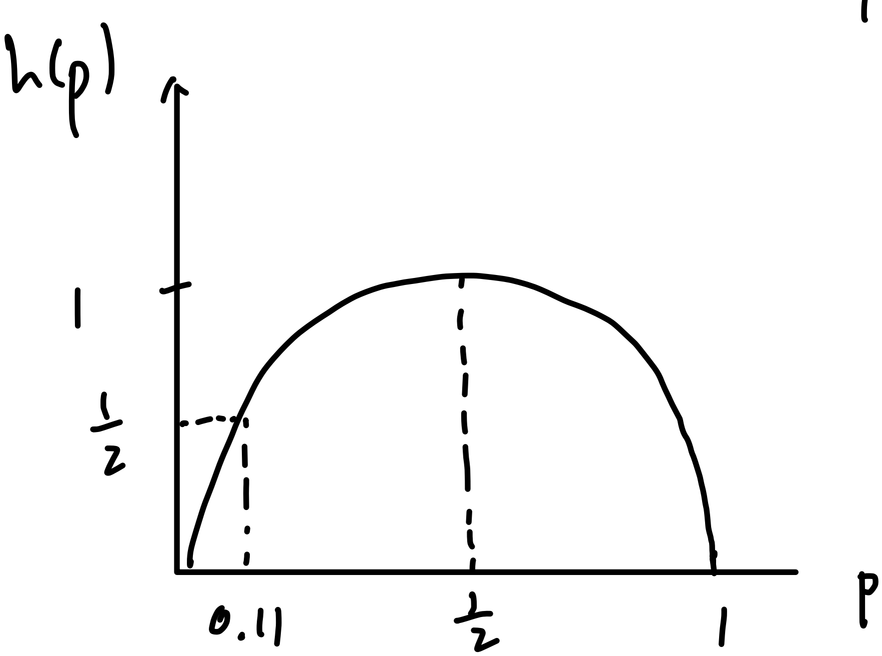
Compression to fewer than $N(H(X))$ bits is impossible without loss of info.

Entropy

$$\boxed{H(X) = \mathbb{E}\left[\log \frac{1}{p(X)}\right] = \sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)}}$$

$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} = h(p)$$

$X \sim \text{Bern}(p)$



$$H(X,Y) = \mathbb{E}\left[\log \frac{1}{p(x,y)}\right]$$

If X,Y indep.

$$H(X,Y) = \sum_x \sum_y p(x,y) \log \frac{1}{p(x,y)}$$

$$= \sum_x \sum_y p(x)p(y)\left[\log \frac{1}{p(x)} + \log \frac{1}{p(y)}\right]$$

$$= \left(\sum_y p(y) \sum_x p(x) \log \frac{1}{p(x)}\right) +$$

$$\left(\sum_x p(x) \sum_y p(y) \log \frac{1}{p(y)}\right)$$

$$= H(X) + H(Y)$$

X,Y indep. $\Rightarrow$ $H(X,Y) = H(X) + H(Y|X)$

$= H(Y) + H(X|Y)$

AEP (Asymptotic Equipartition Property)

S : np heads & $n(1-p)$ tails

$$P(S) = p^{np}(1-p)^{n(1-p)}$$

$$= 2^{np\log p} \cdot 2^{n(1-p)\log(1-p)}$$

$$= 2^{n(\underbrace{p\log p + (1-p)\log(1-p)}_{-h(p)})}$$

$$= 2^{-nh(p)}$$

Total # of such sequence:

Stirling

$$\binom{n}{np} = \frac{n!}{(np)!(n-np)!} \approx 2^{nh(p)} \qquad n! = \left(\frac{n}{e}\right)^n$$

If $p = 0.11$, $2^{nh(p)} = 2^{n/2}$

E.g. $n = 1000$, $p = 0.11$ Total # $= 2^{1000}$

# of typical seq $= 2^{500}$

Thm: (AEP) $X_1 \ldots X_n$ i.i.d $\sim p(x)$.

$$-\frac{1}{n} \log p(x_1, x_2 \ldots x_n) \longrightarrow H(X) \text{ in prob.}$$

Proof:

Functions of RV = RV

If X_i's are i.i.d, so are $\log p(X_i)$.

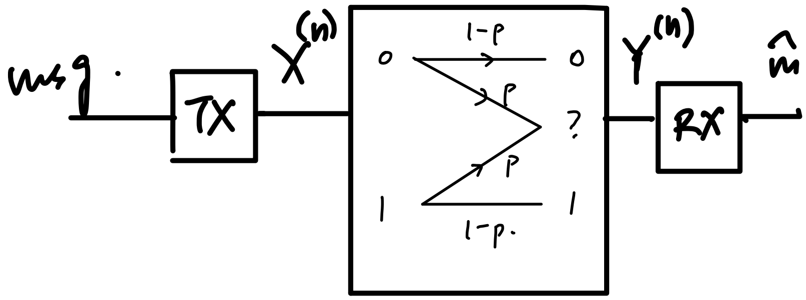
WLLN

$$-\frac{1}{n} \log p(X_1, \cdots, X_n) = -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i)$$

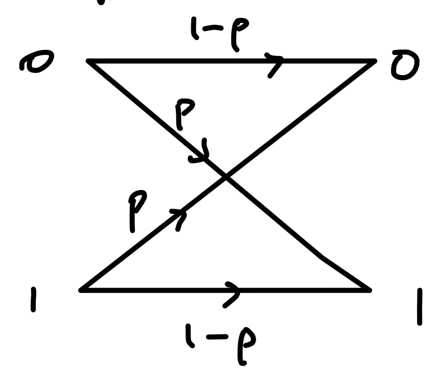
$$\longrightarrow -\mathbb{E}[\log p(X)] \text{ in prob.}$$

$$= H(X).$$

Channel Capacity.

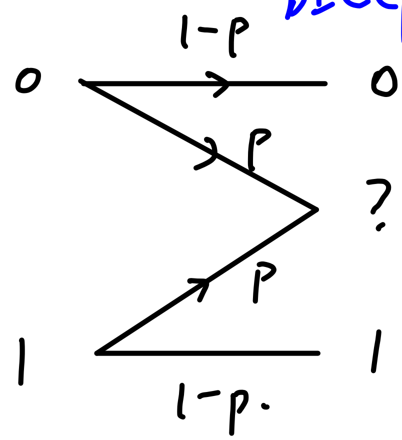


BSC(p)
Binary Symm. Channel



Find capacity of BEC(p) channel

BEC(p)



<u>Capacity</u>: Max rate of reliable communication

$$\text{Rate} = \frac{L_n}{n} = R \text{ bits/ch. use}$$

$$\boxed{R = \frac{L_n}{n}}$$

$$m \xrightarrow[E]{f_n} X^{(n)} \xrightarrow{\text{Channel}} Y^{(n)} \xrightarrow[D]{g_n} \hat{m}$$

$$\boxed{P_e^{(n)} = \max_{(\text{all } m)} P(\hat{m} \neq m)}$$

R is <u>achievable</u> if $\forall n$ (block-length) that is "long enough", $\exists\ f_n(\cdot), g_n(\cdot)$ s.t. $P_e^{(n)} \to 0$ as $n \to \infty$. The largest achievable R = Capacity.

$$\boxed{C_{BEC(p)} = 1-p} \quad \frac{\text{bits}}{\text{channel use}}$$

By SLLN, as $n \to \infty$ $P(np \text{ bits get erased}) \to 1$

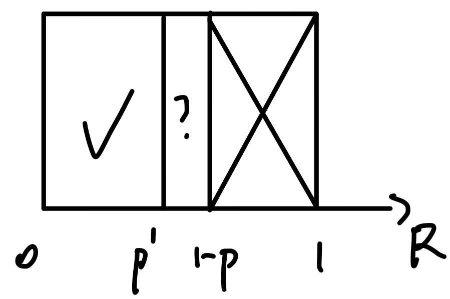
$$C_{BEC(p)} \leq 1-p \ \text{bits/channel use}$$

<u>Proof</u>: (1) Converse (2) Achievability.

(1)

Tx: 011 0110 ⋯

Rx: 01* 01 * ⋯



max rate $\leq 1-p$ bits / ch. use even with genie.

(2) $R = 1-p-\epsilon$

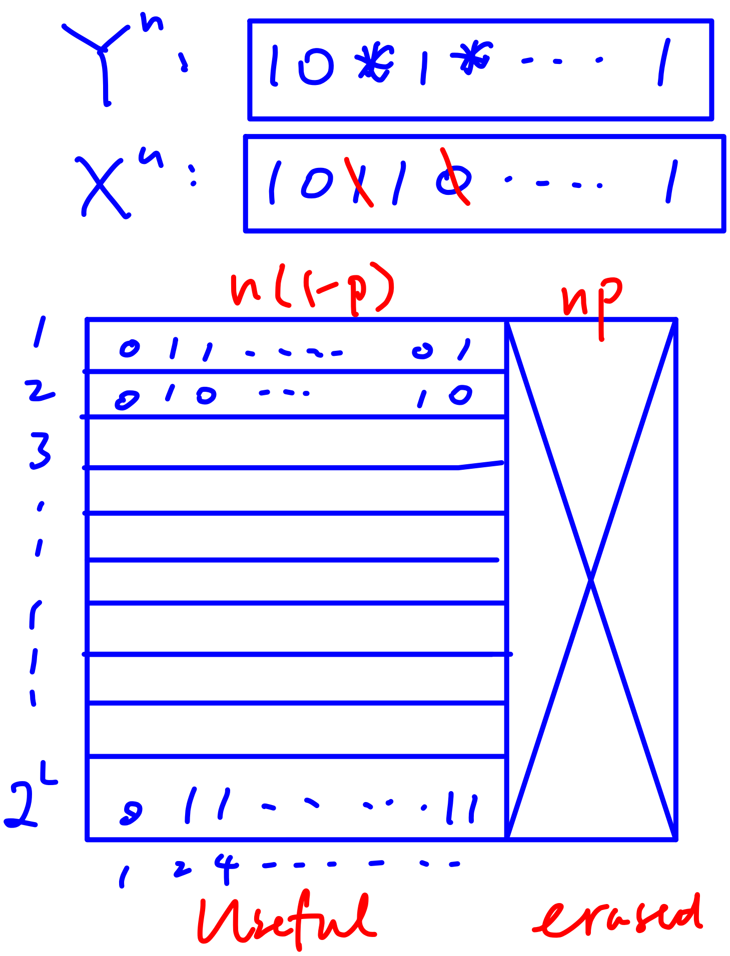
By SLLN, $P(\text{erase } np \text{ bits}) \xrightarrow{n\to\infty} 1$

msg #	1	2	3	⋯	⋯		n
1	0	1	0	1	---	1	0
2	0	1	1	0		0	1
3	1	0	1	0			
⋮							
i	1	0	1	1	0		1
⋮							
2^L	0	1	1	0	--	1	1

Bernoulli($\frac{1}{2}$)

codeword of the i-th msg.

Assume WLOG that last np bits are erased.



Decoding Rule

If c_j' is the only entry in $\mathcal{C}'$ matching $Y^{(n)}$ exactly, declare $\hat{m} = j$.

else:

declare FAILURE.

P_e Analysis

Assume WLOG that msg #1 was sent.

$$P(\text{error}) = P(c_j' \text{ is not unique})$$

$$= P\left(\bigcup_{i=2}^{2^L} \{c_i = c_1\}\right) \le \sum_{i\ne 1}^{2^L} \frac{1}{2^{n(1-p)}}$$

$R = \frac{Ln}{n}$

$$= 2^{Ln} \cdot 2^{-n(1-p)} = 2^{nR}\, 2^{-n(1-p)}$$

$$= 2^{n(R-(1-p))}$$

$P_e \xrightarrow{n\to\infty} 0$

$$\boxed{P(\text{error}) \le 2^{n(R-(1-p))}}$$

$\Rightarrow R-(1-p) < 0$

$\Rightarrow \boxed{R < 1-p}$

$$\text{Make } R = 1 - p - \epsilon \Rightarrow \boxed{P_e \le 2^{-n\epsilon}} \xrightarrow{n\to\infty} 0$$

Ex. $n = 10{,}000$ $\quad p = 0.5$ $\quad \epsilon = 0.01$

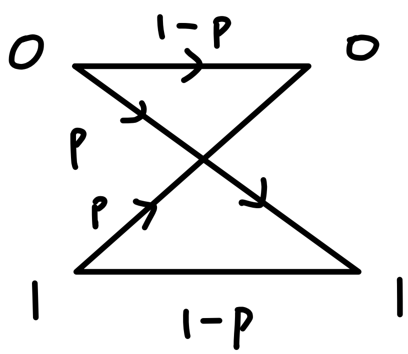
$$C_{BEC(\frac{1}{2})} = \frac{1}{2} \text{ bit/ch. use} \Rightarrow C = 5{,}000 \text{ bits.}$$

$$L = \overset{n}{10{,}000}\,\overset{R}{\left(1 - \underbrace{0.5}_{p} - \underbrace{0.01}_{\epsilon}\right)}$$

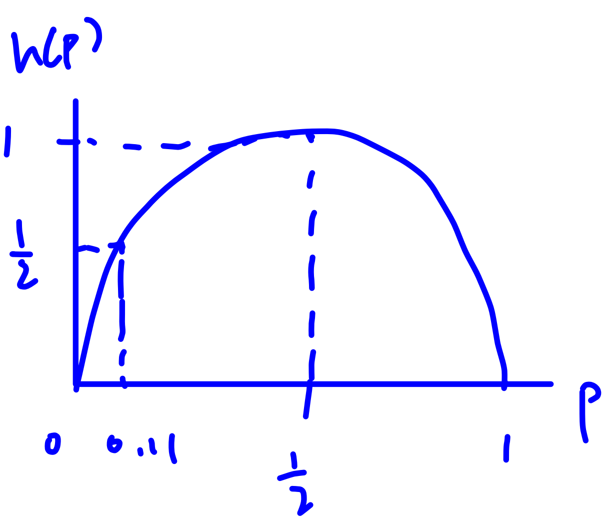
$$= 4{,}900 \Rightarrow \boxed{L = 4{,}900}$$

$$P_e \le 2^{-n\epsilon} = 2^{-100} \approx 0$$

BSC:

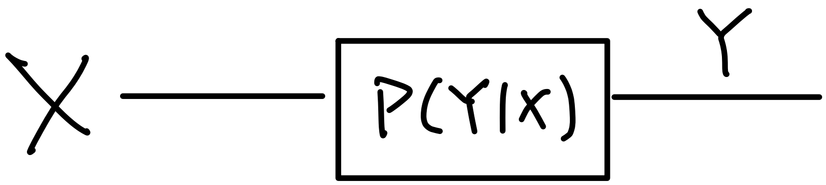


$$C_{BSC(p)} = 1 - h(p) \ \frac{\text{bits}}{\text{ch. use}}$$



$$p = \frac{1}{2} \Rightarrow C = 0.$$

General Discrete Memoryless Channel (DMC)



$R \leq C$

$$\boxed{C = \max_{p_X(x)} I(X;Y) = \max_{p_X(x)} \left[H(X) - H(X|Y) \right]}$$

design choice

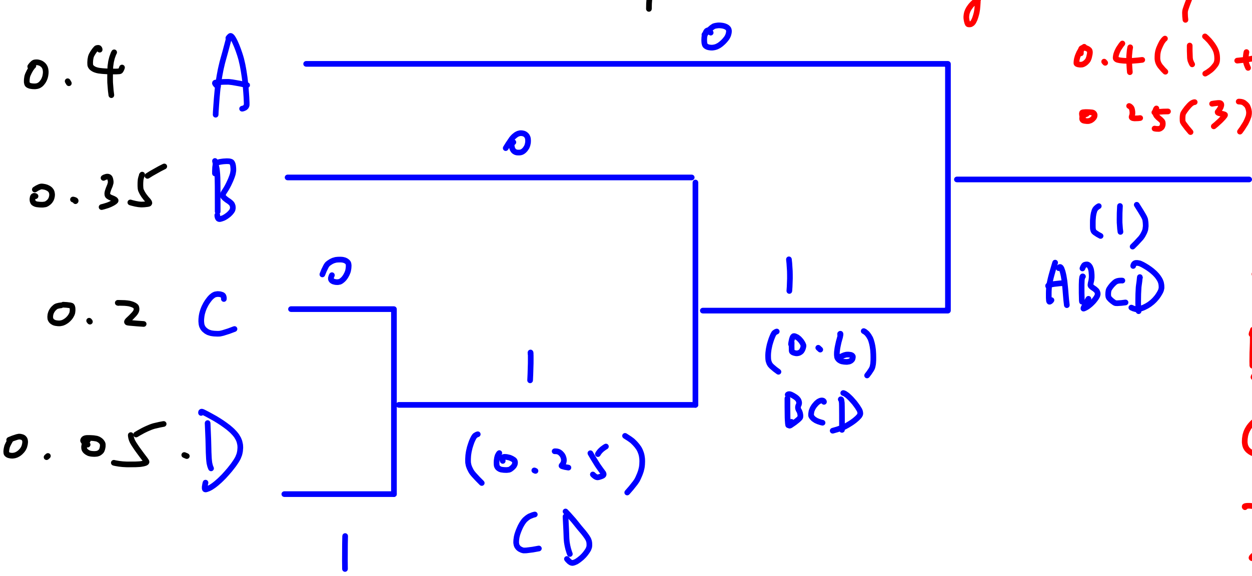
Huffman Coding

$X \in \{A, B, C, D\}$

$P_A = 0.4$
$P_B = 0.35$
$P_C = 0.2$
$P_D = 0.05.$

AABACDAAB

$H(X) = 1.74$ bits/symb.



Avg # bits/symbol = $0.4(1) + 0.35(2) + 0.25(3) = 1.85$ bits/sym.

A : 0
B : 10
C : 110
D : 111

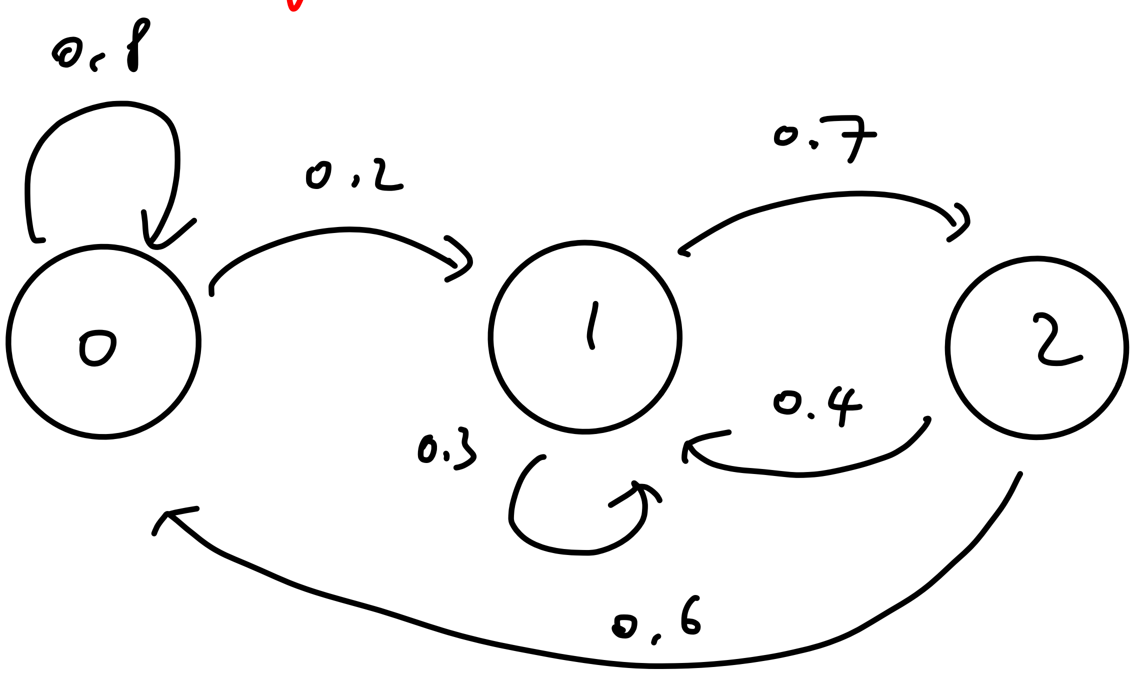
MARKOV CHAIN

$$\boxed{P(X_n = x_n \mid X_{n-1} = x_{n-1}, \cdots X_0 = x_0) = P(X_n = x_n \mid X_{n-1} = x_{n-1})}$$

from $\rightarrow P_{x_{n-1}, x_n} \leftarrow$ to

Does not imply X_n is indep. of $X_{n-2}, \ldots$
but conditionally indep. given X_{n-1}.

Ex.



$$P = \begin{array}{cc} & \text{to} \\ \text{from} & \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.3 & 0.7 \\ 0.6 & 0.4 & 0 \end{bmatrix} \end{array}$$

Row sums = 1

Initial Dist:

$$\pi_0 = [\pi_0(0) \quad \pi_0(1) \quad \pi_0(2)]$$

$$\pi_n = [\pi_n(0) \quad \pi_n(1) \quad \pi_n(2)]$$

Start with $n=1$:

$$\underbrace{P(X_1 = j)}_{\pi_i(j)} = \sum_{i=0}^{2} \underbrace{P(X_1 = j \mid X_0 = i)}_{P_{ij}} \underbrace{P(X_0 = i)}_{\pi_0(i)}$$

$$\pi_i(j) = \sum_{i=0}^{2} P_{ij}\, \pi_0(i)$$

$$\underbrace{[\pi_1(0) \quad \pi_1(1) \quad \pi_2(1)]}_{\pi_1} = \underbrace{[\pi_0(0) \quad \pi_0(1) \quad \pi_0(0)]}_{\pi_0} \underbrace{\begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}}_{P.}$$

$$\pi_1 = \pi_0 P \Rightarrow \pi_2 = \pi_1 P = \pi_0 P^2$$

$$\Rightarrow \quad \boxed{\pi_n = \pi_0 P^n}$$

If $\pi = \pi P$,

then π is called invariant dist. of the MC (stationary)

$\pi = \pi P$ balance equation

Flow in = Flow out at each state at equilibrium.

Let $\pi_0 = [x \ \ y \ \ z]$ ∵ $\pi_0 = \pi_0 P$

State 0: $0.6z = 0.2x$

State 1: $0.2x + 0.4z = 0.7y$

State 2: $0.7y = z$

$x + y + z = 1$ Solve

$\pi_0 = [0.55 \quad 0.26 \quad 0.19]$

Irreducible: go from any state to any state

Aperiodic: $d(i) = \gcd\{ n \geq 1 \mid P^n(i,i) > 0 \}$

$d(i) = d \quad \forall i \in \mathcal{X} \quad d=1$

Self-loop $\Rightarrow d=1$

Invariant dist exists ✓ if # states is finite

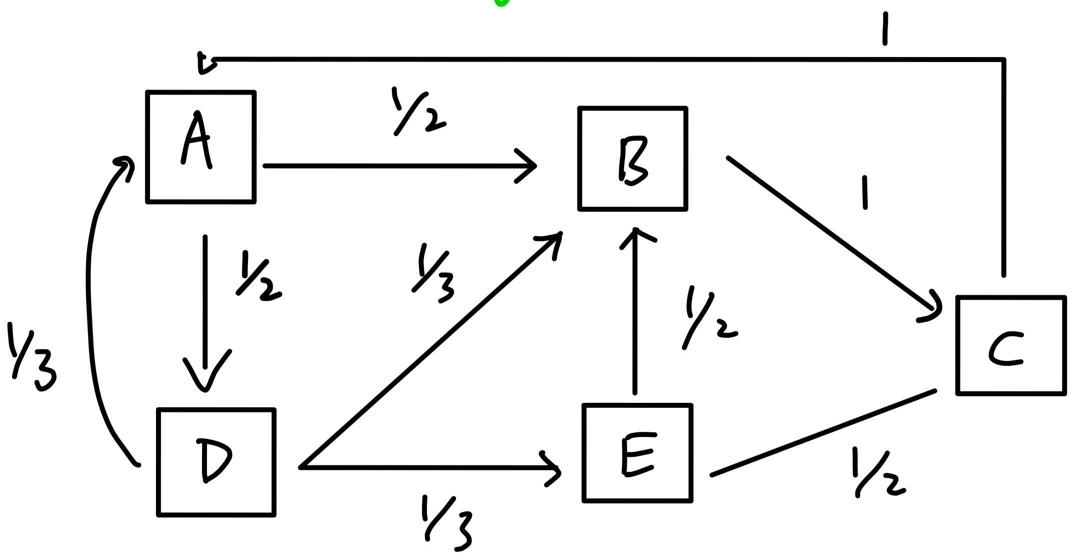
Unique? ✓ if finite, irreducible.

<u>Big Theorem:</u>

(1) If MC is finite and irreducible, it has a unique invariant dist. π and $\pi(i)$ is the long term frac of time that X_n is equal to i.

(2) If also aperiodic, $\lim_{n\to\infty} \pi_n = \pi$

Hitting Time



Starting at A: how many steps to reach E?

(Hitting Time)

$$\beta_E(A) \triangleq \mathbb{E}\left[T_E \mid X_0 = A\right]$$

$$\beta(A) = 1 + \tfrac{1}{2}\beta(B) + \tfrac{1}{2}\beta(D) = 17$$

$$\beta(B) = 1 + \beta(C) = 19$$

$$\beta(C) = 1 + \beta(A) = 18$$

$$\beta(D) = 1 + \tfrac{1}{3}\beta(A) + \tfrac{1}{3}\beta(B) + \tfrac{1}{3}\beta(E) = 13$$

$$\beta(E) = 0$$

First step Eq.

Classification

Transcience: (recurrence) A state i is (recurrent) transient if there is a non-zero (zero) prob. we will never return to i.

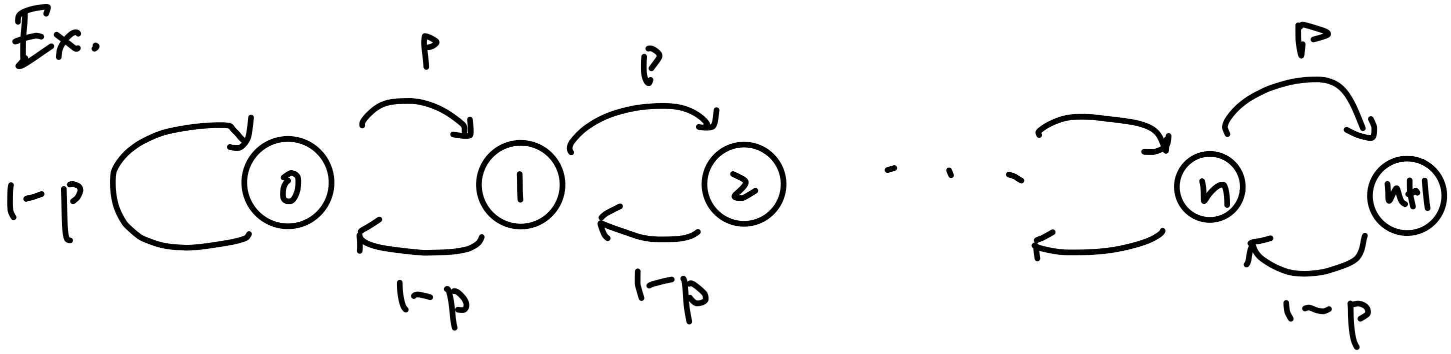
$$P(T_i < \infty \mid X_0 = i) = \begin{cases} 1 & \text{recurrent} \\ <1 & \text{transient} \end{cases}$$

If MC is recurrent:

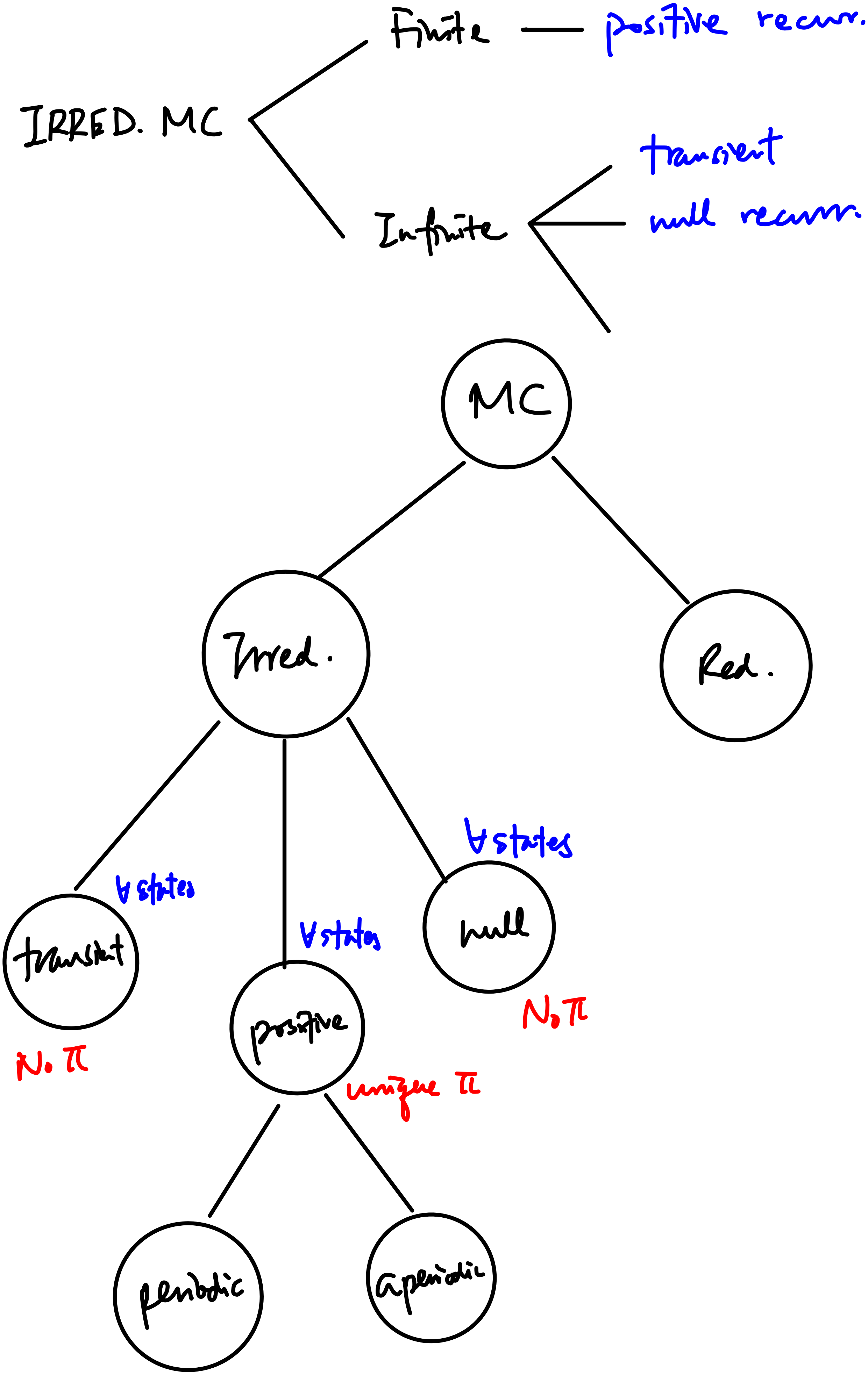
$$\text{if } \begin{cases} \mathbb{E}[T_i \mid X_0 = i] < \infty : & \text{positive recurrent.} \\ \quad \cdots \quad = \infty : & \text{null recurrent.} \end{cases}$$

If MC irred., aperiodic, and positive, then asymptotically stationary.

Ex.



$p < \frac{1}{2}$: positive $\quad p > \frac{1}{2}$: transient $\quad p = \frac{1}{2}$: null.


IRRED. MC
Finite
positive recurr.
Infinite
transient
null recurr.
MC
Irred.
Red.
transient
∀ states
No π
positive
∀ states
unique π
null
∀ states
No π
periodic
aperiodic

Reversability

Facts: ① MC runs backward is always a MC.

② If reversible, same MC.

Proof: ①

$$P[X_k = i \mid X_{k+1} = j, X_{k+2} = i_{k+2} \cdots X_n = i_n]$$

$$= \frac{P[X_k = i, X_{k+1} = j, X_{k+2} = i_{k+2} \cdots X_n = i_n]}{P[X_{k+1} = j, X_{k+2} = i_{k+2} \cdots X_n = i_n]}$$

$$= \frac{\pi(i) P_{ij} \cancel{P_{j i_{k+2}}} \cancel{\cdots} \cancel{P_{i_{n-1} i_n}}}{\pi(j) \cancel{P_{j i_{k+2}}} \cancel{\cdots} \cancel{P_{i_{n-2} i_{n-1}}} \cancel{P_{i_{n-1} i_n}}} \quad \left(\text{Markov Property}\right)$$

$$= \frac{\pi(i) P_{ij}}{\pi(j)}$$

$$= P[X_k = i \mid X_{k+1} = j] = \tilde{P}_{ji} = P[\text{backward chain from } j \text{ to } i]$$

$$\tilde{P}_{ji} = P_{ji}$$

time-reversible MC.

Condition for reversibility:

$$\tilde{P}_{ji} = P_{ji} = \frac{\pi(i) P_{ij}}{\pi(j)}$$

$$\Rightarrow \quad \pi(i) P_{ij} = \pi(j) P_{ji} \quad \forall i, j \in \mathcal{X} \quad *$$

Detailed-balance Equation

■

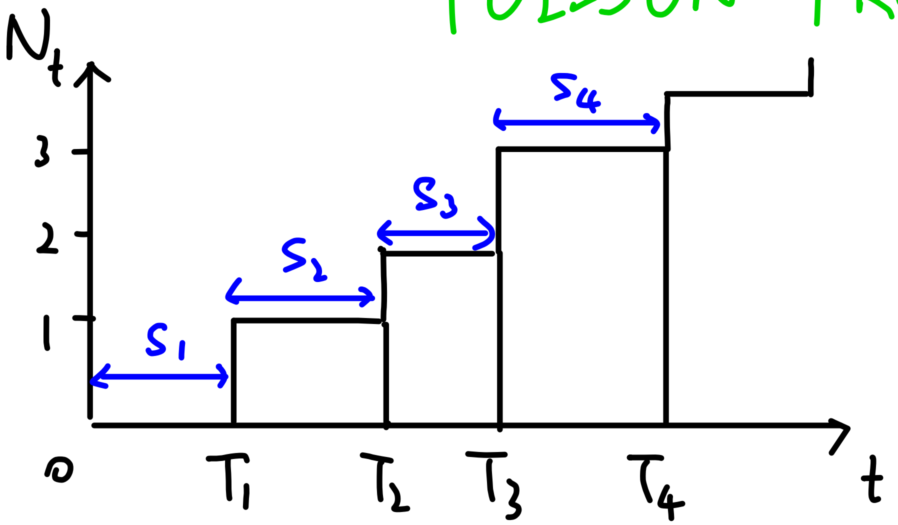
Thm If a MC is reversible, it has **invariant** dist. π

Proof: Show $\pi(j) = \sum_i \pi(i) P_{ij}$ $\qquad \pi = \pi P$.

$$\sum_i \pi(i) P_{ij} = \sum_i \pi(j) P_{ji} = \pi(j) \sum_i P_{ji} = \pi(j)$$

$\blacksquare$

POISSON PROCESSES



$T_1 = S_1$

$T_2 = S_1 + S_2$

$\vdots$

$T_n = \sum_{i=1}^{n} S_i$

- $S_1, S_2 \text{---} S_n$ i.i.d Expo(λ) R.V.s $\lambda > 0$
- T_i's are the CT-arrival times

$$f_{S_i}(t) = \lambda e^{-\lambda t} \;;\quad t > 0$$

$$N_t = \begin{cases} \max\limits_{n \geq 1} \{n \mid T_n \leq t\} & t \geq 0 \\ 0 & \text{if } t < T_1 \end{cases}$$

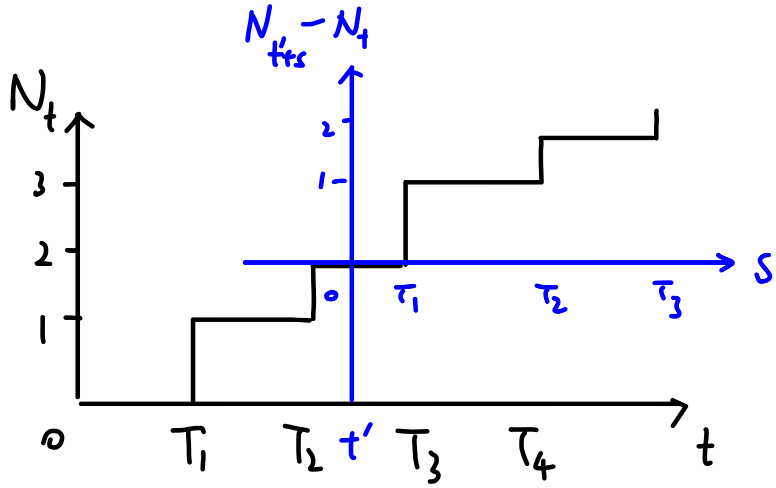
$\lim\limits_{\epsilon \to 0} \frac{o(\epsilon)}{\epsilon} = 0$ e.g. ϵ^2

- $P[\tau \leq t + \epsilon \mid \tau > t] = \underbrace{\lambda\epsilon}_{P(1 \text{ arrival})} + \underline{o(\epsilon)}$

Proof: $P[\tau > t + \epsilon \mid \tau > t] = P[\tau > \epsilon]$

$$= e^{-\lambda\epsilon}$$

$$= \underbrace{1 - \lambda\epsilon + o(\epsilon)}_{P(0 \text{ arrival})} \quad \blacksquare$$



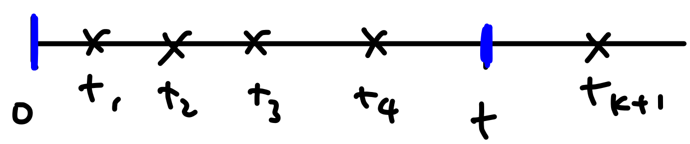
Thm: Poisson Process is memoryless.

- If $N_t \sim PP(\lambda)$, then so is $(N_{t'+s} - N_{t'})$
- $\forall\ 0 \le t_1 < t_2 < \cdots$, $\{N_{t_{n+1}} - N_{t_n}\}$ are indep. and dist. depends only on $(t_{n+1} - t_n)$

Thm: If $N = \{N_t, t \ge 0\}$ is a $PP(\lambda)$, then $N_t = (\#$ of arrivals in $[0,t])$ has a Poisson Dist.

$$\boxed{P[N_t = k] = \frac{e^{-\lambda t}(\lambda t)^k}{k!}}$$

Proof:



Joint Density $P[T_1 \in \{t_1, t_1 + dt_1\}, \cdots T_k \in \{t_k, t_k + dt_k\}, T_{k+1} > t]$

$$= P[S_1 \in \{t_1, t_1 + dt_1\}, S_2 \in \{t_2 - t_1, t_2 - t_1 + dt_2\}, \cdots$$
$$S_k \in \{t_k - t_{k-1}, t_k - t_{k-1} + dt_k\}, S_{k+1} > t - t_k]$$

$$= (\lambda e^{-\lambda t_1} dt_1)(\lambda e^{-\lambda(t_2 - t_1)} dt_2) \cdots (\lambda e^{-\lambda(t_k - t_{k-1})} dt_k)(\lambda e^{-\lambda(t - t_k)})$$

$$= \lambda^k e^{-\lambda t} dt_1 \cdots dt_k$$

* Uniform

$$\boxed{f_{T_1,T_2 \cdots T_k}(t_1, t_2, \cdots t_k) = \lambda^k e^{-\lambda t}}$$

$$N_t(k) = \int_{t_1} \int_{t_2} \cdots \int_{t_k} f_{T_1,T_2,\cdots T_k}(t_1, t_2, \cdots t_k)\, dt_1 dt_2 \cdots dt_k$$

$$= \lambda^k e^{-\lambda t} \underbrace{\int_0^t \int_0^t \cdots \int_0^t}_{S:\ t_1 < t_2 < \cdots t_k} \underbrace{dt_1 dt_2 \cdots dt_k}_{Vol(S) = t^k}$$

(independence)

w/o constraints.

$$= \lambda^k e^{-\lambda t} \frac{t^k}{k!}$$

$$= \frac{(\lambda t)^k e^{-\lambda t}}{k!} \sim \text{Poisson}(\lambda t)$$

■

ERLANG DISTRIBUTION

Interarrival time of $PP(\lambda)$

$$T_k = S_1 + \dots S_k$$

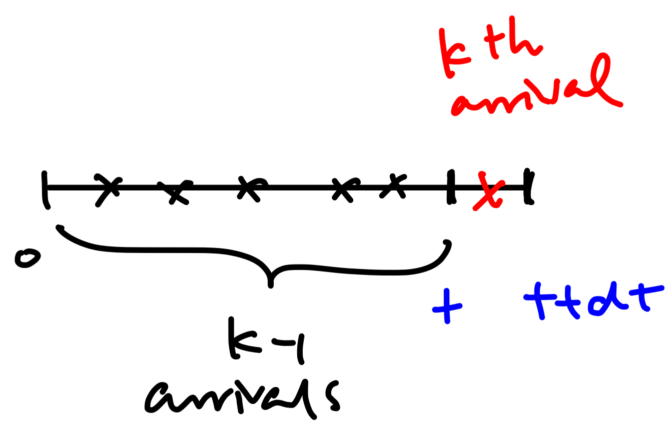
$$\mathbb{E}[T_k] = \sum_{i=1}^{k} \mathbb{E}[S_i] = \frac{k}{\lambda}$$

$$Var(T_k) = k\,Var(T_k) = \frac{k}{\lambda^2}$$

$$f_{T_k}(t)\,dt = P((k-1) \text{ arrs} \in [0,t]) \cdot P(k\text{th} \in [t, t+dt])$$

$$= \frac{e^{-\lambda t}(\lambda t)^{k-1}}{(k-1)!} \cdot \lambda\, dt$$

$$\boxed{f_{T_k}(t) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t}}$$

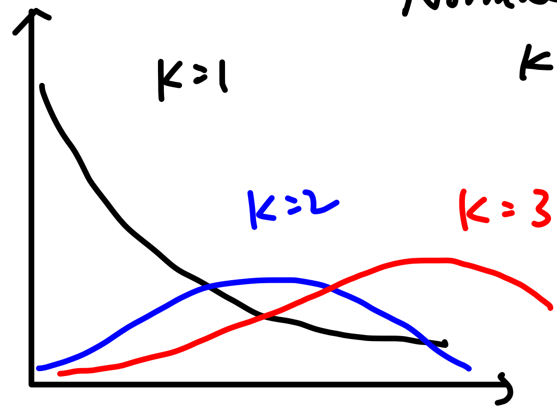


kth order

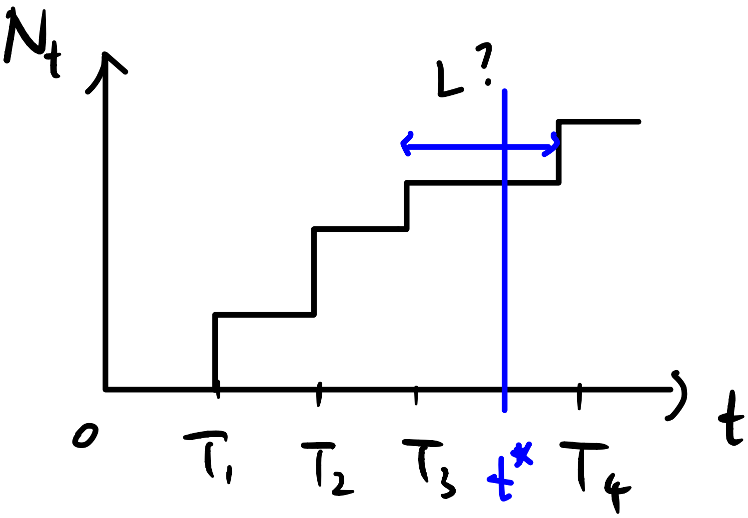
Erlang pdf.

$$T_k \sim \text{Erlang-}k(\lambda)$$

Normal as $k \to \infty$



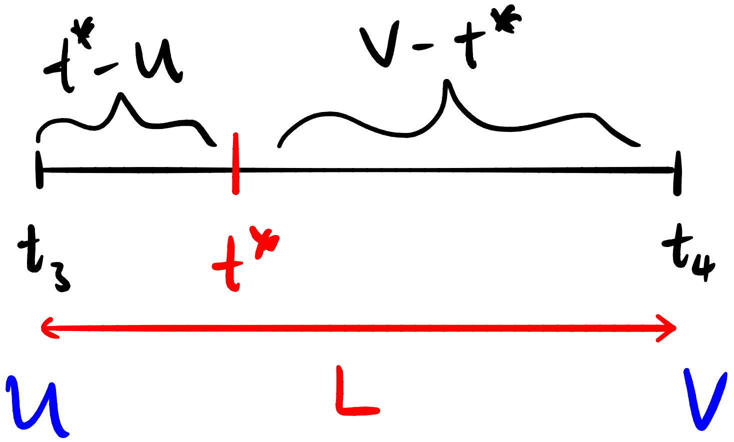
Random Incidence Paradox RIP



$$L \sim \text{Erlang-2}(\lambda)$$

$$\Downarrow$$

$$\mathbb{E}[L] = \frac{2}{\lambda}$$



$$L = \underbrace{(t^* - u)}_{\text{Expo}(\lambda)} + \underbrace{(V - t^*)}_{\text{Expo}(\lambda)}$$

$$\begin{aligned} P(t^* - u > x) &= P(\text{no arrivals} \in [t^* - x, t^*]) \\ &= P(\text{no arrivals in } (0, x)) \\ &= e^{-\lambda x} \end{aligned}$$

Continuous-time Markov Chains (CTMC)

Rate Matrix $Q = \{ Q(i,j) \mid i,j \in \mathcal{X} \}$

i) $Q(i,j) \geq 0 \quad \forall i \neq j$

ii) $\sum_j Q(i,j) = 0$ row sum = 0

$$\text{CTMC } (\pi, Q) : \{ X_t, t \geq 0 \} \text{ s.t. } P(X_0 = i) = \pi(i)$$

$$P(X_{t+\epsilon} = j \mid X_t = i, X_u, u<t) = \begin{cases} \epsilon Q(i,j) + o(\epsilon) & i \neq j \\ 1 + \epsilon Q(i,j) + o(\epsilon) & i = j \end{cases}$$

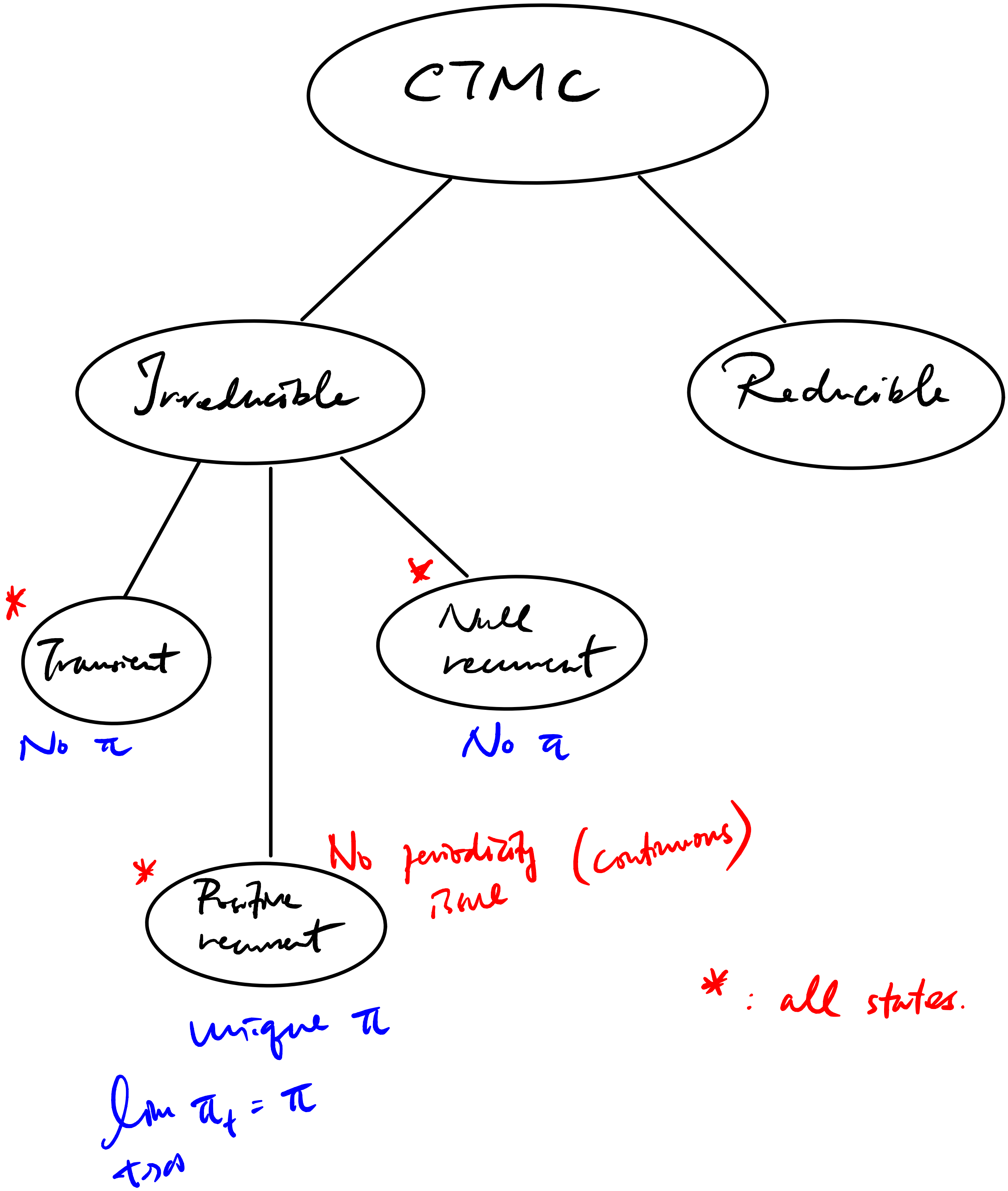
$$\lim_{\epsilon \to 0} \frac{o(\epsilon)}{\epsilon} = 0$$

$$P(X_{t+\epsilon} = j \mid X_t = i, X_u, u<t) = q(i) \epsilon \, T(i,j)$$

$$= \cancel{q(i)} \epsilon \cdot \frac{Q(i,j)}{\cancel{q(i)}}$$

$$= Q(i,j) \epsilon$$

Big Theorem for CTMC



Random Graphs

Erdős Rényi $G(n,p)$

$$\mathbb{P}[G=G_0] = p^m(1-p)^{\binom{n}{2}-m}$$

1) $\mathbb{E}[\#\text{edges}] = \binom{n}{2}p$

2) Let $D = \deg(v)$, $\forall v$, distribution of D?

$$\mathbb{P}[D=d] = \binom{n-1}{d} p^d (1-p)^{n-d-1} \qquad D \sim \text{Bin}(n-1,p)$$

$$\mathbb{E}[D] = (n-1)p$$

3) $n \to \infty$, $p \to 0$, $(n-1)p = \lambda$, $p_D(d)$?

$$D \sim \text{Poisson}(\lambda) \qquad \mathbb{P}[D=d] \approx \frac{\lambda^d e^{-\lambda}}{d!}$$

4) Probability of a node being isolated?

$$q = (1-p)^{n-1} \; *$$

Thm $p(n) = \lambda \frac{\log n}{n}$

a) $\lambda < 1 \Rightarrow \mathbb{P}[G(n,p) \text{ connected}] \to 0 \quad n \to \infty$

b) $\lambda > 1 \Rightarrow \mathbb{P}[G(n,p) \text{ connected}] \to 1 \quad n \to \infty$

Proof: (stronger statement)

a) $\mathbb{P}[\text{no isolated}] \to 0$ as $n \to \infty$

Let X = # isolated nodes in $G(n,p)$

Goal: $\mathbb{E}[X]$

* $p = p(n)$

I_i = node i is isolated

$$\Rightarrow \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[I_i] = n\,\mathbb{E}[I_1] = nq = n(1-p)^{n-1}\ *$$

Note $\ln \mathbb{E}[X] = \ln n + (n-1)\ln(1-p)$

$\approx -p = \lambda \frac{\ln n}{n}$ Taylor

$$\ln \mathbb{E}[X] \sim \ln n - \left(\frac{n-1}{n}\right)\lambda \ln n$$
$$\approx \ln n\,(1-\lambda)$$

$n \to \infty$

$\lambda < 1$

* $\boxed{\mathbb{E}[X] = nq = n^{1-\lambda}}$

Lemma: If $X \geq 0$,

$$\mathbb{P}[X=0] \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}[X])^2}$$

① Proof:

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$= \mathbb{P}[X=0]\,\mathbb{E}[X]^2 + \mathbb{P}[X=1]\left[\mathbb{E}[X]-1\right]^2 + \mathbb{P}[X=2]\left[\mathbb{E}[X]-2\right]^2 + \cdots$$

$$\geq \mathbb{P}[X=0]\,\mathbb{E}[X]^2$$

$\blacksquare$

② Proof:

$$\mathbb{P}[X_n=0] \leq \mathbb{P}[|X_n - \mathbb{E}[X_n]| \geq \mathbb{E}[X_n]]$$

$$\leq \frac{\mathrm{Var}(X_n)}{\mathbb{E}[X_n]^2}$$

Second Moment Method.

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{n} I_i\right)$$

$$= n\,\mathrm{Var}(I_1) + \sum_i \sum_{j \neq i} \mathrm{Cov}(I_i, I_j)$$

$$= n\,\mathrm{Var}(I_1) + n(n-1)\,\mathrm{cov}(I_1, I_2)$$

$$\mathrm{Cov}(I_1, I_2) = \mathbb{E}[I_1 I_2] - \overbrace{\mathbb{E}[I_1]}^{q}\,\overbrace{\mathbb{E}[I_2]}^{q}$$

$$= (1-p)^{n-1}(1-p)^{n-2} - q^2$$

$$= \frac{q^2}{1-p} - q^2$$

$$\mathrm{Var}(X) = nq(1-q) + n(n-1)\Bigg[\underbrace{\frac{q^2}{1-p} - q^2}_{\frac{pq^2}{1-p}}\Bigg]$$

$$\mathbb{P}[X=0] \le \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{nq(1-q) + n(n-1)\frac{pq^2}{1-p}}{n^2 q^2}$$

$$= \frac{1-q}{nq} + \frac{n-1}{n} \cdot \frac{p}{1-p}$$

$$\longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

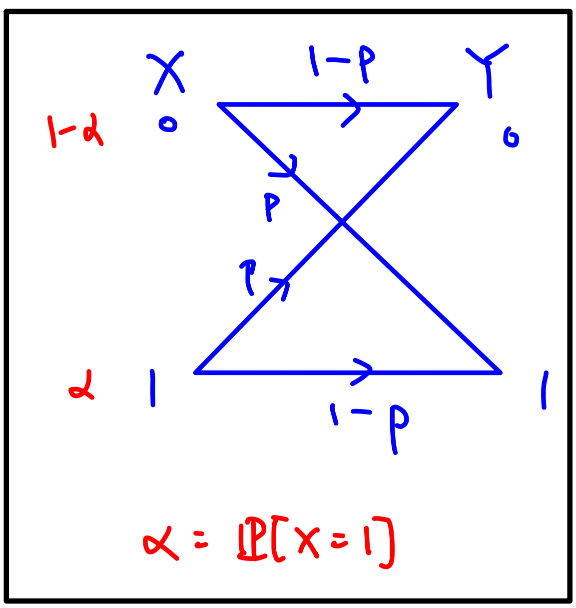
MAP: Max A Posteriori estimate

MLE: Maximum likelyhood estimate

$$\text{MAP} \triangleq \arg\max_i \pi_i = \arg\max_i p_i q_i$$

$$\text{MAP}(X|Y=y) = \max_x P(X=x|Y=y)$$

$$\text{MLE}(X|Y=y) = \max_x P(Y=y|X=x)$$

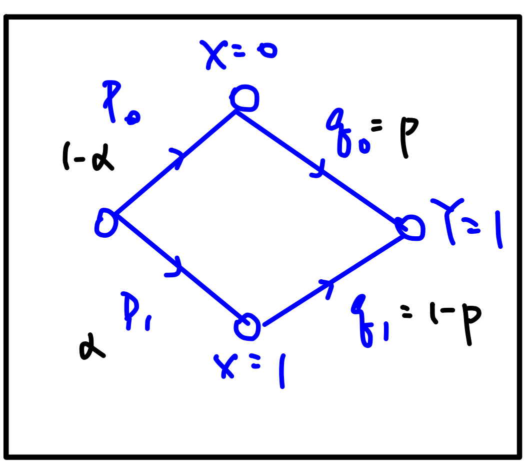


$BSC(p),\ p < \frac{1}{2}$

$$\text{MAP}(X|Y=0) = \{1\}_{p>1-\alpha} = \begin{cases} 0 & p \le 1-\alpha \\ 1 & p > 1-\alpha \end{cases}$$

$$\text{MAP}(X|Y=1) = \begin{cases} 0 & p \ge \alpha \\ 1 & p < \alpha \end{cases}$$

$$\text{MLE}(X|Y) = Y$$



$$\text{MAP}(X|Y=1) = \max_{i\in\{0,1\}} p_i q_i$$

If $\quad p_0 q_0 \underset{\hat{x}_0=0}{\overset{\hat{x}_1=0}{\gtrless}} p_1 q_1, \quad \Rightarrow$

$$(1-\alpha)p \underset{1}{\overset{0}{\gtrless}} \alpha(1-p)$$

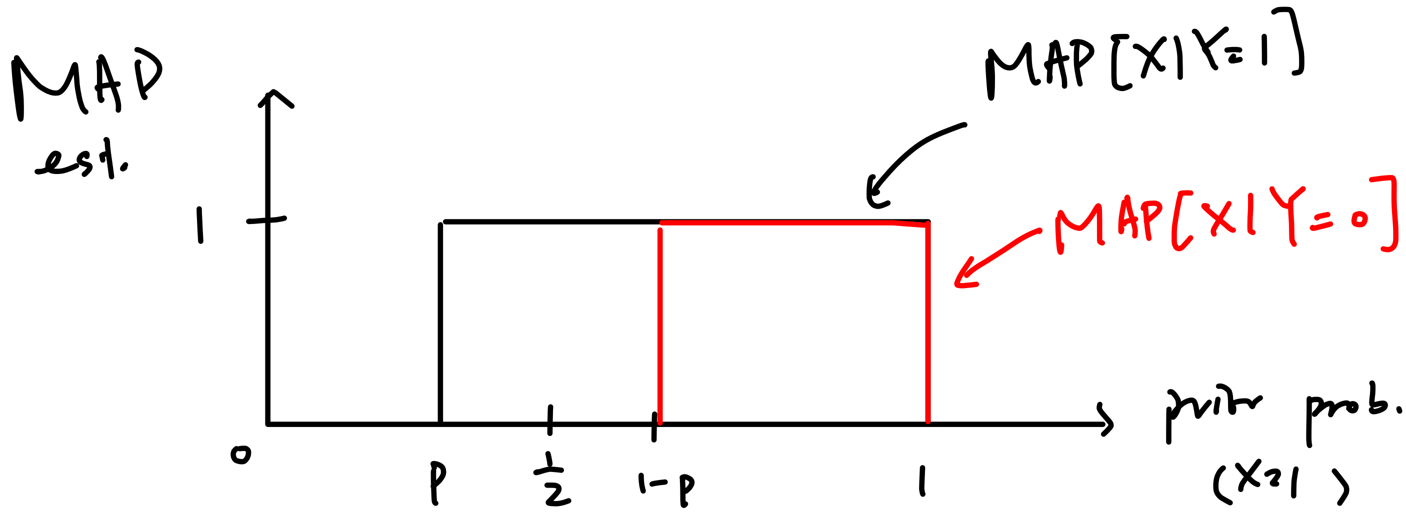
$$p - \cancel{\alpha p} \underset{1}{\overset{0}{\gtrless}} \alpha - \cancel{\alpha p} \Rightarrow p \underset{1}{\overset{0}{\gtrless}} \alpha$$

$$\hat{X}_{MAP}(Y=1) = \begin{cases} 0 & p \ge \alpha \\ 1 & p < \alpha \end{cases}$$

$$\mathrm{MLE}[X|Y] = \mathrm{MAP}[X|Y] \text{ when } \alpha = \tfrac{1}{2}$$

$$\left.\begin{aligned} \mathrm{MLE}[X|Y=1] &= 1 \\ \mathrm{MLE}[X|Y=0] &= 0 \end{aligned}\right\} \quad \text{because } p < \tfrac{1}{2}$$

$$\mathrm{MLE}[X|Y] = Y$$



HYPOTHESIS TESTING

$X \in \{0,1\}$

Probability of correct decision

<u>Goal</u>: $\max P_{CD} = P(\hat{X}=1 \mid X=1)$

$P_{FA} = P(\hat{X}=1 \mid X=0) \leq \beta$

probability of False Alarm

<u>Neyman-Pearson</u> (N-P)

no priors

(i) Observe Y

(ii) Two hypothesis

$H_0: Y \sim f(y|0)$ (Null)

$H_1: Y \sim f(y|1)$ (Alternate)

<u>Decision Rule</u>:

$r: \mathbb{R} \to \{0,1\}$

<u>Goal</u>

$\min P[r(Y) \mid X=1]$ false negative

$P(r(Y) \mid X=0] \leq \beta$ false positive

error types

	H_0	H_1
$r(Y)=0$	✓	false negative
$r(Y)=1$	false positive	✓

Type-II error

PFA

Type-I error

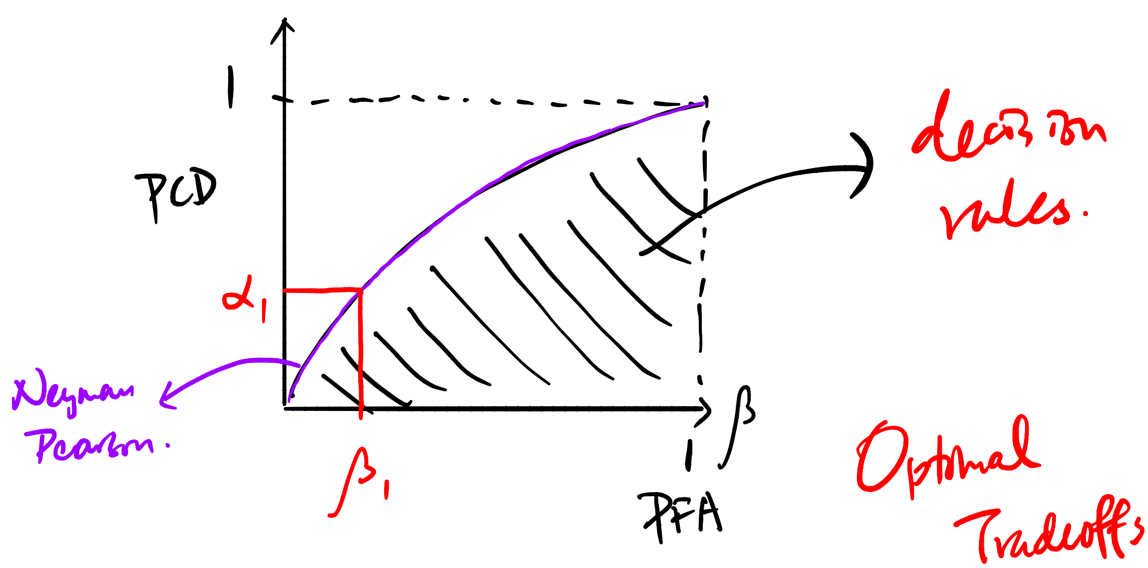
PCD

"power"

<u>Goal</u>: $\max PCD = P(\hat{X}=1 \mid X=1)$

$$PFA = P(\hat{X}=1 \mid X=0) \leq \beta$$

R.O.C (Receiver Operating Characteristic)



$$\mathcal{L}(y) = \frac{f(y|1)}{f(y|0)}$$

<u>N-P Thm</u> OPTIMAL DECISION RULE

$$\hat{X} = r^*(Y) = \begin{cases} 1 & \text{if } \mathcal{L}(Y) > \lambda \\ 0 & \text{if } \mathcal{L}(Y) < \lambda \\ 1 \text{ w.p. } \gamma & \text{if } \mathcal{L}(Y) = \lambda \end{cases}$$

$\lambda > 0$ $\gamma \in [0,1]$ chosen s.t.

$$P[\hat{X} = 1 \mid X = 0] = \beta$$

If $\mathcal{L}(\lambda)$ is large, declare $\hat{X} = 1$

else, declare $\hat{X} = 0$.

Proof: Consider binary hypothesis testing problem.

Goal: Show that any OTHER decision rule (other than N-P) having the same "PFA spec" will not result in a better "PCD spec" i.e.

if $\tilde{X}$ is an alternate decision rule $(N\text{-}P = \hat{X})$

If $P(\tilde{X}=1 \mid X=0) \le \beta$, then $P(\tilde{X}=1 \mid X=1) \le P(\hat{X}=1 \mid X=1)$

Lemma: $(\hat{X}-\tilde{X})(\mathcal{L}(Y)-\lambda) \ge 0$ (1)

Proof: If $\mathcal{L}(Y) > \lambda$, $\hat{X}=1 \Rightarrow \hat{X}-\tilde{X} \ge 0$.

If $\mathcal{L}(Y) < \lambda$, $\hat{X}=0 \Rightarrow \hat{X}-\tilde{X} \le 0$

If $\mathcal{L}(Y) = \lambda$, $0 \ge 0$. ✓

$$\hat{X}\mathcal{L}(Y) - \tilde{X}\mathcal{L}(Y) \ge \lambda\hat{X} - \lambda\tilde{X}$$

$$\mathbb{E}[\hat{X}\mathcal{L}(Y) \mid X=0] - \mathbb{E}[\tilde{X}\mathcal{L}(Y)] \ge \lambda\left[\mathbb{E}[\hat{X} \mid X=0] - \mathbb{E}[\tilde{X} \mid X=0]\right]$$

$$= \lambda\left[\underbrace{P(\hat{X}=1 \mid X=0)}_{\beta} - \underbrace{P(\tilde{X}=1 \mid X=0)}_{\le \beta}\right]$$

$$\ge 0$$

$$\mathbb{E}[\underbrace{\hat{X}}_{g(Y)}\mathcal{L}(Y) \mid X=0] \ge \mathbb{E}[\tilde{X}\mathcal{L}(Y) \mid X=0]$$

$$\int g(y)\mathcal{L}(y) f_{Y|X}(y|0)\,dy \ge \int h(y)\mathcal{L}(y) f_{Y|X}(y|0)\,dy$$

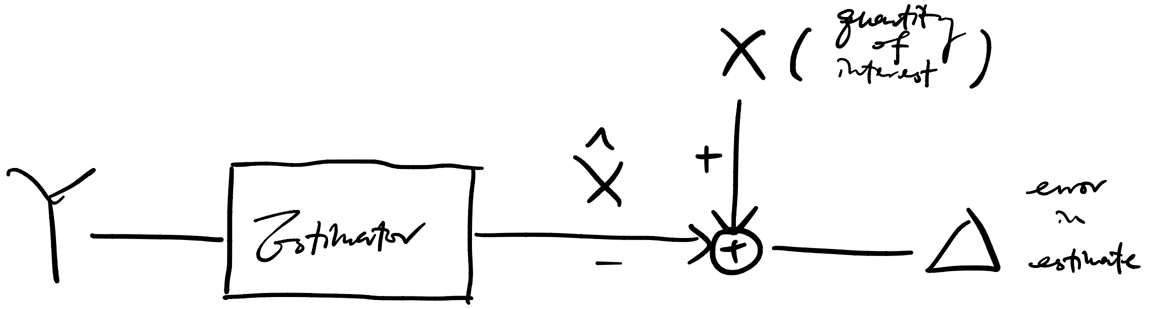
$$\int g(y)\frac{f_{Y|X}(y|1)}{\cancel{f_{Y|X}(y|0)}} \cancel{f_{Y|X}(y|0)}\,dy \ge \int h(y)\frac{f_{Y|X}(y|1)}{\cancel{f_{Y|X}(y|0)}} \cancel{f_{Y|X}(y|0)}\,dy$$

$$\mathbb{E}[\hat{X} \mid X=1] \ge \mathbb{E}[\tilde{X} \mid X=1]$$

$$\mathbb{P}[\hat{X}=1 \mid X=1] \ge \mathbb{P}[\tilde{X}=1 \mid X=1]$$

■

ESTIMATION

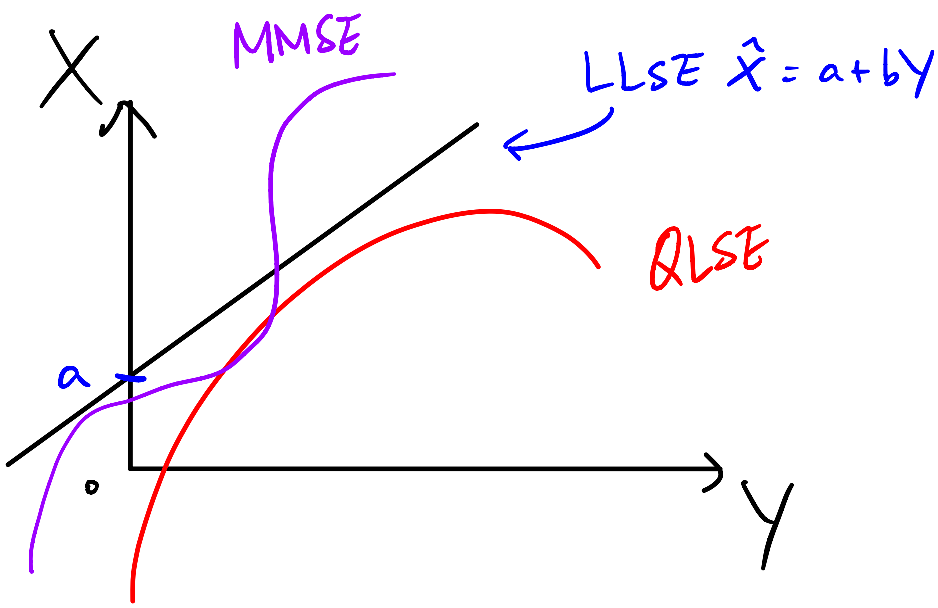


Goal: Est. X from Y as accurate as possible.

Error $= \Delta \Rightarrow$ want $\mathbb{E}[\Delta^2]$ to be small.

LLSE: $\hat{X} = a + bY$

MMSE: minimize $\mathbb{E}\left[(X - \hat{X}(Y))^2\right]$ (no constraints on $\hat{X}$)



$$\min_{a,b} \mathbb{E}\left[(X-(a+bY))^2\right]$$

$$\mathbb{E}\left[(X-a-bY)^2\right] = F(a,b)$$

$$= \mathbb{E}\left[X^2 + a^2 + b^2Y^2 - 2aX - 2bXY + 2abY\right]$$

$$\frac{\partial F}{\partial a} = 2a - 2\mathbb{E}[X] + 2b\mathbb{E}[Y] = 0 \qquad (1)$$

$$\frac{\partial F}{\partial b} = 2b\,\mathbb{E}[Y^2] + 2a\mathbb{E}[Y] - 2\mathbb{E}[XY] = 0 \qquad (2)$$

$$\boxed{\begin{aligned} L[X|Y] &= a + bY \\ &= \mathbb{E}[X] + \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)}\left[Y - \mathbb{E}[Y]\right] \end{aligned}}$$

Properties of $L[X|Y]$:

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

$$\mathbb{E}[\Delta] = \mathbb{E}[X - \hat{X}] = 0$$

(2) $\mathrm{cov}(\Delta, Y) = 0$ (Projection Property)

- unbiased
- uncorrelated.

1) $\mathbb{E}[XY] \iff \langle \vec{X}, \vec{Y} \rangle = |\vec{X}||\vec{Y}|\cos\vartheta$

$0 \iff \theta = \frac{\pi}{2}$

2) $\mathbb{E}[X^2] \iff \langle \vec{X}, \vec{X} \rangle = |\vec{X}|^2$

3) $$\rho = \frac{\mathbb{E}[XY]}{\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}} \iff \frac{\langle \vec{X}, \vec{Y} \rangle}{|\vec{X}||\vec{Y}|} = \cos\vartheta$$

KALMAN FILTER

Kalman Eqs:

(1) $\hat{X}_{n|n} = \hat{X}_{n|n-1} + K_n \underbrace{(Y_n - \hat{X}_{n|n-1})}_{\tilde{Y}_n}$

(a) $\hat{X}_{n|n-1} = a\hat{X}_{n-1|n-1}$

(b) $\tilde{Y}_n = Y_n - \hat{X}_{n|n-1}$

(2)

Kalman Gain: $K_n = \dfrac{\sigma^2_{n|n-1}}{\sigma^2_{n|n-1} + \sigma^2_w}$

(3) $\sigma^2_{n|n-1} = a^2\sigma^2_{n-1|n-1} + \sigma^2_v$

(4) $\sigma^2_{n|n} = (1-K_n)\,\sigma^2_{n|n-1}$

$$\begin{cases} x_n = a x_{n-1} + v_n & (1) \\ y_n = c x_n + w_n & (2) \end{cases}$$

$$\sigma^2_{n|k} := \mathbb{E}\left[(x_n - \hat{x}_{n|k})^2\right]$$

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + k_n \tilde{y}_n$$

$$\tilde{y}_n = y_n - a\hat{x}_{n-1|n-1}$$

$$k_n = \frac{\sigma^2_{n|n-1}}{\sigma^2_{n|n-1} + \sigma^2_w}$$

$$\sigma^2_{n|n-1} = a^2 \sigma^2_{n-1|n-1} + \sigma^2_v$$

$$\sigma^2_{n|n} = \sigma^2_{n|n-1}(1 - k_n)$$

$$\hat{x}_{n|n-1} = a\hat{x}_{n-1|n-1}$$

Discussion Cheatsheet.

$U \sim \min(\text{Expo}(\lambda_1), \text{Expo}(\lambda_2)) = \text{Expo}(\lambda_1 + \lambda_2)$

$V \sim \max(\text{Expo}(\lambda_1), \text{Expo}(\lambda_2))$

$U \quad , \quad V - U$ indep.

(LLSE)

- $\mathbb{E}[\hat{X}-X]=0$ $\quad \mathbb{E}[\Delta]=0$ Unbiased.
- $\mathrm{Cov}(\hat{X}-X, Y)=0$ $\quad \mathbb{E}[\Delta Y]=0$ projection property.

(MMSE)

$$\mathbb{E}[\hat{X}|Y]=\mathbb{E}[X|Y] \iff \mathbb{E}[\hat{X}-X|Y]=0.$$

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1)$$

$$H(X, Y) = H(X) + H(X \mid Y)$$

$$H(X \mid Y) \neq H(Y \mid X)$$