c) $\pi_j = 0$ for all transient states, $\pi_j > 0$ $\forall$ recurrent states

<u>example</u>

$$\begin{array}{c} \text{up to date} \\ \text{behind} \end{array} \begin{bmatrix} .8 & .2 \\ .6 & .4 \end{bmatrix}$$

$$[\pi_1 \ \pi_2] = [\pi_1 \ \pi_2]\begin{bmatrix} .8 & .2 \\ .6 & .4 \end{bmatrix} = \left.\begin{array}{l} \pi_1 = .8\pi_1 + .6\pi_2 \\ \pi_2 = .2\pi_1 + .4\pi_2 \end{array}\right\} \pi_1 = 3\pi_2$$

$\sum \pi = 1 \Rightarrow \pi_1 + \pi_2 = 1$ $\boxed{\pi_2 = .25, \ \pi_1 = .75}$ 75% chance of being up to date in the steady state

· long term frequency & steady state prob

$\pi_j = \lim_{n\to\infty} \frac{V_{ij}(n)}{n}$, $V_{ij}(n)$: expected # of visits to state j after n steps from i

→ fraction of time spending at state j

· deterministic seq. convergence

② LLN : $S_N = \frac{X_1 + \dots + X_N}{N}$, $\lim_{n\to\infty} S_n = \mu$ for iid X_i

let $Y_i = 1\{X_i = j\}$: indicator that X_i in state j

$N_j(n) = Y_1 + \dots + Y_n$: # of visits to state j

$\frac{N_j(n)}{n}$ = frac of time in state j $\xrightarrow{a.s} \pi_j$ via ergodic thm

③ Big thm : convergence in distribution

$$\lim_{n\to\infty} P^n(i,j) = \pi_j$$

Absorption

Consider a recurrent state K as "absorbing" if $P_{KK} = 1$, $P_{Kj} = 0$ $\forall j \neq K$

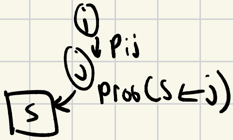
define $a_i = P(X_n$ eventually goes to absorbing state $S \mid X_0 = i)$

methodology : first step eqs

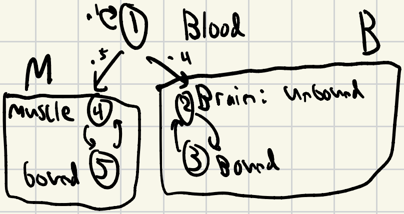
$a_S = 1$

$a_i = 0$ for all absorbing $i \neq S$ } all recurrent state is

$a_i = \sum_{j=1}^{m} p_{ij} a_j$ for all transient i



Example



$\mathbb{P}(X_n \text{ in } M \mid X_0 = i) = a_i$

$a_M = 1$

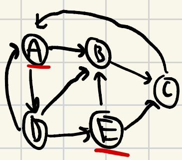
$a_B = 0$ given in brain, prob reaching muscle $= 0$

$a_1 = \sum_{j=1}^{m} P_{i,j} a_j = P_{1,M} \cdot a_M + P_{1,B} a_B + P_{1,1} \cdot a_1 = .5(1) + .4(0) + .1(a_1) \Rightarrow a_1 = 5/9$

$b = 4/9$

$\pi = [0 \quad b \cdot \pi_{brain} \quad a_1 \cdot \pi_{muscle}]$

Hitting times



$X_0 = A$

$\mathbb{E}(T_E \mid X_0 = A)$, T_E : # of timesteps to get to E

Method: first step equations

$\beta(A) = \mathbb{E}(T_E \mid X_0 = A)$, expected # of steps to E given $X_0 = A$

$\beta(i) := \mathbb{E}(\quad \mid X_0 = i)$

Brute Force: $\sum_{i=1}^{\infty} i \cdot P_{AE}(i) \leftarrow \mathbb{E}(\#$ of steps from A to E$)$

↑ # of steps

FSE

$\beta(A) = 1 + \sum_{i \in B,D} P_{A,i} \mathbb{E}(T_E \mid X_0 = i) = 1 + P(A,D) \cdot \mathbb{E}(T_E \mid X_0 = D) + P(B,D) \mathbb{E}(T_E \mid X_0 = B)$

$= 1 + P(A,D)\beta(D) + P(A,B)\beta(B)$

$\beta(B) = 1 + P(B,C)\beta(C)$

$\beta(C) = 1 + P(C,A)\beta(A)$

$\beta(D) = 1 + P(D,A)\beta(A) + P(D,E)\beta(E) + P(D,B)\beta(B)$

$\beta(E) = 0$

Probability of hitting one state before another given a start?

$\alpha(A) = \mathbb{P}(T_C < T_E \mid X_0 = A)$

$$\alpha(A) = \sum_{B,D} \mathbb{P}(T_C < E \mid X_0 = i) = P(A,B)\cdot P(T_C < T_E \mid X_0 = B) + P(A,D)\cdot P(T_C < T_E \mid X_0 = D)$$

Lecture 18

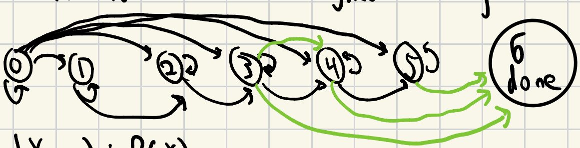
• Countably infinite movies, rating $\in (0,1,2,3,4,5)$, want 2 movies combined rating > 7.5

a) What is approximate MC?

b) $\mathbb{E}$(# of movies until done?)

define X_n to be the highest rating seen thus far

X_n :



$\mathbb{E}(T_6 \mid X = x) : \beta(x)$

$\beta(6) = 0$

$\beta(5) = P_{5,5}\beta(5) + P_{5,6}\beta(6) + 1$

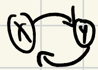
$= 3/6\, \beta(5) + 3/6\, \beta(6) + 1 = \frac{1}{2}\beta(5) + 1$, $\frac{1}{2}\beta(5) = 1$, $\beta(5) = 2$

$\beta(4) = P_{4,4}\beta(4) + P_{4,6}(\beta(6)) + 1$

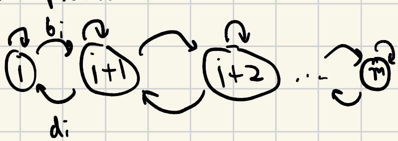
$= 4/6\, \beta(4) + 2/6\, \beta(6) + 1 \implies \beta(4) = 3$

$\mathbb{E}$(# of movies till done) $= \mathbb{P}(X_0 = 0)\beta(0) \cdots \mathbb{P}(X_0 = 5)\beta(5)$

$\boxed{\text{Reversibility}}$ → $\pi(x)P(x,y) = \pi(y)P(y,x)$



Birth death process



$b_i = \mathbb{P}(X_{n+1} = i+1 \mid X_n = i)$

$d_i = \mathbb{P}(X_{n+1} = i-1 \mid X_n = i)$

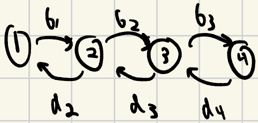
• expected # of transitions $i \to i+1$ = expected # from $i+1 \to i$

• Steady state probabilities π : $\pi_i b_i = \pi_{i+1} d_{i+1}$

$$\pi_i = \frac{b_0 \cdot b_1 \cdots b_{i-1}}{d_1 d_2 \cdots d_i}$$

CTMC

• $\pi_i q_{ij} = \pi_j q_{ji}$



$\pi_2 b_2 = \pi_3 d_3$

$\pi_3 b_3 = \pi_4 d_4$

$\pi_3 = \frac{\pi_4 d_4}{b_3}$

Continuous time Markov chains

Pandemic

0 Susceptible } arrival w/ rate λ

1 infected

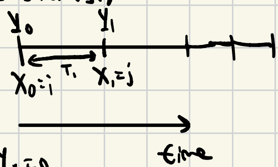
2 recovered } rate μ

Queing system : Customers in line X_n: # ppl in line

arrivals : λ

Served Customers: μ

X_n : State of nth transition



Y_n : time of nth transition, $Y_0 = 0$

T_n : time b/w $(n-1)$ and nth transition

time to next transition $\sim Exp(v_i)$

P_{ij} prob $X_n = j \mid X_{n-1} = i$

X_ns: $X_0, X_1, \ldots, X_n \leftarrow$ DTMC w/ matrix P

"embedded Chain"

Y_n : Can be PP if v_i same regardless of i

Event $A = \{T_1 = t_1, T_2 = t_2 \cdots T_n = t_n, X_0 = i_0, \ldots, X_n = i\}$

a) $\text{Prob}(X_{n+1} = j, T_{n+1} \geq t \mid A) = \mathbb{P}(X_{n+1} = j, T_{n+1} \geq t \mid X_n = i)$ by markov property

$= \mathbb{P}(X_{n+1} = j \mid X_n = i) \cdot \mathbb{P}(T_{n+1} \geq t \mid X_n = i)$ and Memorylessness

$= P_{ij}(1 - (1 - e^{-v_i t}))$

$= P_{ij} e^{-v_i t}$

b) expected time of next transition

Property of exp.

$$\mathbb{E}(T_{n+1} \mid X_n = i) = \int_0^\infty \tau \cdot v_i e^{-v_i \tau} \, d\tau = \frac{1}{v_i}$$

v_i : avg # of transitions out of state i per unit time spent at i

$q_{ij} = v_i \cdot p_{ij}$, average # of transitions from $i \to j$ per unit time spent at i

Note 1 : if q_{ij} is known $\to$ get v_i via $\sum_{j=1}^{m} q_{ij} = \sum_{j=1}^{m} v_i p_{ij} = v_i \sum p_{ij} = v_i$

Note 2 q_{ij} s $\to v_i$, can find p_{ij}

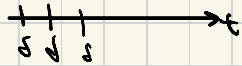
Note 3 $i \to i+1$ Self arrivals dont affect interarrival $T_n \to T_{n+1}$

$X_n = i, X_{n+1} = i$ so $p_{ii} = 0$, $q_{ii} = 0$

Steady State behavior

1) describe CTMC as DTMC

$Z_n = X(n\delta)$ at time $n\delta$ State of $X = Z_n$



- δ is small enough for 1 transition per δ
- Z_n is itself a MC
- $\overline{p_{ij}}$ are Z_n's transition probabilities
- if $Z_n = i \to Z_{n+1} = j$, $P(Z_{n+1} = j \mid Z_n = i) = p_{ij} \cdot P(T_{n+1} \leq \delta)$

$$= p_{ij}(1 - e^{-v_i \delta}) \quad \text{for small } \delta$$
$$\approx p_{ij} v_i \delta$$

- define transition on Z : $\overline{p_{ij}} = p_{ij} v_i \delta = q_{ij} \delta$

$$\overline{p_{ii}} = 1 - \sum_{j \neq i} \overline{p_{ij}} = 1 - \delta \sum_{j \neq i} q_{ij}$$

Steady State Convergence for CTMC

$P(X(t) = i) \to P(Z_n = i)$

if $\pi_j = \lim_{n \to \infty} P(Z_n = j \mid Z_0 = i)$, Same as DTMC

- Single recurrent class
- balance eq for Z_n , $\pi = \pi \overline{P}$

$$\pi_j = \sum \pi_k \overline{p_{kj}} = \pi_j \overline{p_{jj}} + \sum_{k \neq j} \pi_k \overline{p_{kj}} \; \cdots$$

$$\boxed{\pi_j \sum_{k \neq j} q_{jk} = \sum_{k \neq j} \pi_k q_{kj}}$$

Taxi problem

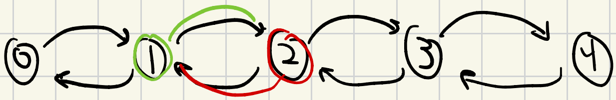
Taxis arrive 2 per min | • if someone is waiting Taxis pickup
People arrive 1 per min | • People only wait if queue size < 4

Want to know: $\mathbb{E}(\text{wait time} \mid \text{joined the queue})$
Assume process is in the steady state
if nobody in queue → $\mathbb{E}(T) = \frac{1}{\lambda} = \frac{1}{2}$ min

if 1 person ahead → $\mathbb{E}(T$ of 2nd arrival$)$ →
We know $\mathbb{E}(T \mid N)$ where N ppl in queue

$$\mathbb{E}(T) = \mathbb{E}(\mathbb{E}(T \mid N)) = p_0 \mathbb{E}(T \mid N=0) + \dots + p_3 \mathbb{E}(T \mid N=3)$$

X_n: how many people in queue : $\{0,1,2,3,4\}$



Birth death

π : Steady state prob of being in each state

$$\pi_i q_{i,i+1} = \pi_{i+1} q_{i+1,i}$$

person arrival rate

$$\pi_i \cdot 1 = \pi_{i+1} \cdot 2 \longrightarrow$$

π_0
π_1
π_2
π_3
π_4
$\sum \pi_i = 1$

Lecture 19

CTMCs : DTMCs + transition time

DTMCs $\rightarrow p_{ij}$

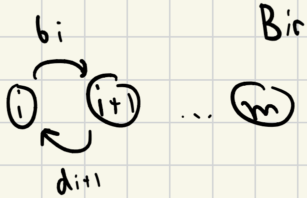
CTMCs $\rightarrow q_{ij}$: rate $- \underbrace{v_i}_{\text{rate of leaving } i} p_{ij}$

$p_{ij} \rightarrow P(X_{n+1} | X_n = i)$

$\pi \rightarrow$ steady state dist

$\pi_j \sum_{i \neq j} q_{ji} = \sum_{k \neq j} \pi_k q_{kj} \rightarrow$ balance equations

Birth death process



Taxi problem $\rightarrow$ CTMC found π

Statistical inference:

Prob: axioms $\rightarrow$ dist of R.Vs $\rightarrow$ Computations based on axioms/dist

Static: noisy data $\rightarrow$ compute qty. of interest.

Bayesian vs. frequentists

Bayesian: unkown variable treated as RV has a dist
$\mu \sim N$ "prior"

frequentists: best estimate of your parameter given the data
$\mu \leftarrow$ data

Let data $= X$

qty of interest $= \theta$

$\mathbb{P}(\text{data} | \theta)$ max this when deciding θ by choosing θ, define CI around it.

Bayesian approach

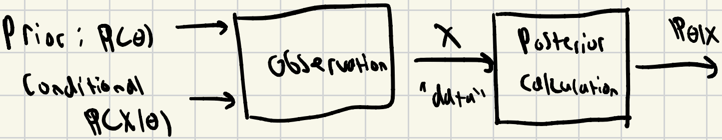
$$\mathbb{P}(\text{data} | \theta) \cdot \mathbb{P}(\theta) = \mathbb{P}(\text{data}, \theta) \rightarrow \underbrace{\mathbb{P}(\theta | \text{data})}_{\uparrow \text{ maximize}} : \frac{\mathbb{P}(\text{data}, \theta)}{\mathbb{P}(\text{data})}$$

"Posterior" "Conditional" "prior"

$$\mathbb{P}(\theta \mid \text{data}) = \frac{\mathbb{P}(\text{data} \mid \theta) \cdot \mathbb{P}(\theta)}{\mathbb{P}(\text{data})}$$

Bayesian: Choose θ to max

frequentist: Choose θ to max



Example: Normal dist

data: $x_1, \dots, x_n \rightarrow$ iid $\sim N(\mu, \sigma^2)$ Want: μ from our data

$\mu \rightarrow$ prior

$N(\mu_0, \sigma_0^2)$ Specific prior

Want: Posterior, estimate μ given X

$\mathbb{P}(\mu \mid X)$

Prior: $f_\mu = C_1 e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$

Conditional: $f_{X \mid \mu} = C_2 e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$

Want $\mathbb{P}(\mu \mid X) = f_{\mu \mid X} = \dfrac{f_\mu \cdot f_{X \mid \mu}}{f_X} =$

$f_X \rightarrow \underbrace{\int f_{X \mid \mu}(x, \mu') \cdot f_\mu(\mu')}_{\text{total prob thm}} \, d\mu'$

Can ignore f_X, $g_{\mu \mid X} = f_\mu \cdot f_{X \mid \mu}$

- normal prior + normal conditional $\rightarrow$ normal posterior "Conjugate priors"
- for a given conditional $\mathbb{P}(X \mid \theta)$, the posterior $\mathbb{P}(\theta \mid X)$ is in the same family of distributions as the prior.

ex: Condition $\sim N(\mu, \sigma^2)$ Prior $\sim N(\mu_0, \sigma_0^2)$ Posterior $\sim N(\dots, \dots)$

<u>Conjugate Priors</u>:

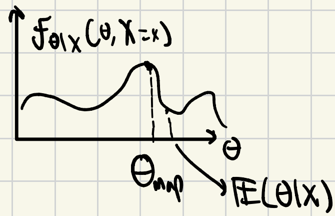
Conditioned $P(X \mid \theta)$ is bernoulli/binomial, prior: $\text{Beta}(\alpha, \beta) \rightarrow$ Posterior is also $\text{Beta}(\hat{\alpha}, \hat{\beta})$

Point estimation

$\hat{\theta}$ is an estimate of θ, $\hat{\Theta}$ is RV $= g(X)$ $\quad \hat{\theta} = g(X=x)$

Estimator: $\hat{\Theta} = \mathbb{E}(\Theta | X=x)$, least mean square estimate

Estimator: MAP estimator



Estimator: Maximum likelihood estimate ← frequentist

$\mathbb{P}(\text{data} | \theta)$ ← likelihood, choosing θ s.t. $\mathbb{P}(\text{data}|\theta)$ maxxed

$$\underbrace{\mathbb{P}(\theta|\text{data})}_{\text{MAP}} = \frac{\overbrace{\mathbb{P}(\text{data}|\theta)}^{\text{MLE}} \cdot \mathbb{P}(\theta)}{c}$$

MLE = MAP if $\mathbb{P}(\theta)$ is uninformative

Example: 100,000 ppl in Berkeley

- 500 eat ice cream
  - 50, $X=1$
  - 450, $X=0$
- 99,500 don't
  - 600, $X=1$
  - 98,900, $X=0$

Ice Cream

$\theta = 1$ if ate

$\theta = 0$ else

Sunburn

$X=1$ yes

$X=0$ no

· given $X=0$ what is MAP and MLE estimate of θ?

MLE: $\mathbb{P}(X=1 | \theta=0)$ and $\mathbb{P}(X=1 | \theta=1)$

$\Rightarrow \frac{600}{99500} = .006$ and $\frac{50}{500} = .1$ likelihood is max when $\theta=1$

MAP: Posterior $\mathbb{P}(\theta|X) \rightarrow \dfrac{\mathbb{P}(X=1|\theta=0) \cdot \mathbb{P}(\theta=0)}{c} = \frac{600}{99500} \cdot \frac{99500}{100000} \approx .006$

$\rightarrow \mathbb{P}(\theta=1|X=1) = \dfrac{\mathbb{P}(X=1|\theta=1) \cdot \mathbb{P}(\theta=1)}{c} = \frac{50}{500} \cdot \frac{500}{100{,}000} = .0005$

Hypothesis testing

- $\hat{\Theta}$ takes on one of m values $\theta_1, \dots, \theta_m$ m is small int
- Computing $\mathbb{P}(\Theta = \theta_i \mid X = x)$ posterior apply MAP rule

Example

- 2 biased coins; p_1, p_2
- Choose coin at random
- decide whether we picked up coin 1 or 2 based on single flip
- $\Theta = 1$: hypothesis coin 1 was chosen
  $\Theta = 2$
- $X \to 0$ tails
  $X \to 1$ heads

Assume outcome is tails $X=0$

$$\mathbb{P}(\Theta=1 \mid X=0) \propto \mathbb{P}(X=0 \mid \Theta=1) \cdot \mathbb{P}(\Theta=1) = (1-p_1)\cdot\frac{1}{2}$$

$$\mathbb{P}(\Theta=2 \mid X=0) \qquad (1-p_2)\cdot\frac{1}{2}$$

$p_1 < p_2 \Rightarrow \theta = 1$

Example:

- Customers arrive $\sim PP(1)$
- $X_1, X_3, \dots$ are sent away   (RUMOR)
- Let $\theta = 1$: odds sent away
  $\theta = 0$: odds not sent

- $Y_1, \dots, Y_n$ where Y_i is time of ith arrival, $T_i = Y_{i+1} - Y_i$
- Derive MAP rule to determine if odds sent away or not
- $\mathbb{P}_{\Theta|T} \propto \mathbb{P}_{T|\Theta} \, \mathbb{P}_\Theta \to$ ignore bc prior is uninformative

$$\mathbb{P}_{T|\theta=0}, \; T_i \sim Exp(1) = \prod_{i=1}^{n} e^{-T_i}$$ prob of observing data under rumor being false

$$\mathbb{P}_{T|\theta=1} \sim \prod \text{Erlang order 2 rate 1} = \prod_{i=1}^{n} t_i e^{-t_i}$$

m users

thinking intervals iid $\exp(\lambda)$

active mode by submitting request

each request served one at a time

Service time iid $\exp(\mu)$

$[0, m]$ possible requests

Lecture 20

Bayesian

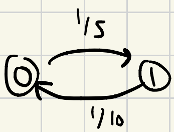
- unknowns are R.V ~ prior
- MAP, max posterior dist

Frequentist

- unknowns ~ parameters estimated from data
- MLE rule

Quiz

Machine $X \begin{cases} 0 \\ 1 \end{cases}$



$\pi_0 \cdot \frac{1}{5} = \pi_1 \cdot \frac{1}{10}$

$\pi_0 = \frac{1}{3}, \ \pi_1 = 2/3$

Balance eqs for C.t.s Markov Chain

Noisy Sensor Y $\quad \mathbb{P}(Y = X) = .9$

$$\mathbb{P}(X=0 \mid Y=1)? = \frac{\mathbb{P}(Y=1 \mid X=0)\cdot \mathbb{P}(X=0)}{\mathbb{P}(Y=1)} = \frac{.1 \cdot 1/3}{.1 \cdot 1/3 + .9 \cdot 2/3}$$

Use MAP to find most likely X given $Y=1$

$$\max_{x \in S} \mathbb{P}(X = x \mid Y=1) \propto \mathbb{P}(Y=1 \mid X=x) \propto \mathbb{P}(Y=1 \mid X=0)\cdot \mathbb{P}(X=0) \text{ vs. } \mathbb{P}(Y=1 \mid X=1)\cdot \mathbb{P}(X=1) \propto 2/3 \cdot .9$$

X = unknown, Y = observations/data, $C(g) = \mathbb{E}\left[C(X, \underbrace{g(Y)}_{\text{estimator}})\right]$

example: $C(X, g(Y)) = C(X, \hat{X}) = |X - \hat{X}|^2$

Minimum MSE

$$C(g) = \mathbb{E}((X-\hat{X})^2) = \mathrm{Var}(X-\hat{X}) + \mathbb{E}(X-\hat{X})^2 = \mathrm{Var}(X) + \underbrace{(\mathbb{E}(X)-\hat{X})^2}_{\text{min this via } \hat{X} = \mathbb{E}(X)}$$

Observation of $Y=y$, $\hat{X}$ given $Y=y$ $= \mathbb{E}(X|Y=y)$ MMSE estimator

MAP $= \max \mathbb{P}(X|y)$
MMSE $= \mathbb{E}(X|Y)$

Example

X — 1 empty, 2 med, 3 full

$Y-1$ happy

$\mathbb{P}(Y=1|X=1) = .1$ | $\mathbb{P}(X=1) = .7$
$\mathbb{P}(Y=1|X=2) = .3$ | $\mathbb{P}(X=2) = .2$
$\mathbb{P}(Y=1|X=3) = .6$ | $\mathbb{P}(X=3) = .1$

Best estimate of X given $Y=1$

MLE: $\mathbb{P}(Y=1|X) \Rightarrow X=3$, $\mathbb{P}(Y=1|X=3) = .6$

MAP: $\max \mathbb{P}(X|Y=1) \propto \mathbb{P}(Y=1|X=1)\cdot\mathbb{P}(X=1) = \boxed{.07} \rightarrow X=1$

$\mathbb{P}(Y=1|X=2)\cdot\mathbb{P}(X=2) = .3\cdot.2 = .06$

$\mathbb{P}(Y=1|X=3)\cdot\mathbb{P}(X=3) = .6\cdot.1 = .06$

MMSE: $\mathbb{E}(X|Y=1) = \sum_x x\cdot\mathbb{P}(X|Y) = \sum_x \frac{x\cdot\mathbb{P}(Y=1|X=x)\cdot\mathbb{P}(X=x)}{\mathbb{P}(Y=1)} = \frac{\cdots}{.19} = 1.95$

Properties of MMSE $\tilde{X} = \hat{X} - X$, $\mathbb{E}(\tilde{X}) = 0$, $\mathbb{E}(\tilde{X}|Y=y) = 0 \ \forall y$

$\text{Cov}(\tilde{X}, \hat{X}) = 0 \Rightarrow \text{Var}(X) = \text{Var}(\hat{X}) + \text{Var}(\tilde{X})$

Can think of as $\mathbb{E}[Y|X]$ as proj onto all funcs of X in $\mathcal{H}$

Linear Least Squares

$C(g) = \mathbb{E}[|X-g(Y)|^2]$, $g(Y) = a+bY$, $C(g) = \mathbb{E}[|X-(a+bY)|^2]$ want a,b

expand and find $\frac{dC}{da}, \frac{dC}{db} = 0$

$C(g) = X^2 + a^2 + b^2Y^2 - 2aX + 2abXY$

$\frac{dC}{da}$ $\mathbb{E}(X^2) + a^2 + b^2\mathbb{E}(Y^2) - 2a\mathbb{E}(X) + 2ab\mathbb{E}(XY)$

$2a - 2\mathbb{E}(X) + 2b\mathbb{E}(Y) = 0$, $\boxed{a = \mathbb{E}(X) - b\mathbb{E}(Y)}$

$\frac{dC}{db} = 2b\mathbb{E}(Y^2) - 2\mathbb{E}(XY) + 2a\mathbb{E}(Y) = 0 \Rightarrow b\,\text{Var}(Y) = \text{Cov}(X,Y)$

$$\boxed{b = \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}}$$

LSE of X: $a+bY = \mathbb{E}(X) - \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}\mathbb{E}(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}\cdot Y = \boxed{\mathbb{E}(X) + \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}(Y-\mathbb{E}(Y))}$

Example

Prob of heads of biased coin $'p' \to X \sim \text{Uniform}(0,1)$
n tosses $\to Y$ heads, $\mathbb{P}(Y=k) \sim \text{Binom}(X, n)$ given X
Derive llse for X given Y heads in n tosses

$$\hat{X} = \underbrace{\mathbb{E}(X)}_{\frac{1}{2}} + \frac{\text{cov}(X,Y)}{\text{Var}(Y)}(Y - \mathbb{E}(Y))$$

$\mathbb{E}(Y) \to \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(nX) = \frac{n}{2}$

$\text{Var}(Y) \to = \text{Var}(\mathbb{E}(Y|X)) + \mathbb{E}(\text{Var}(Y|X)) = \frac{n^2}{12} + \mathbb{E}(nX(1-X)) = \frac{n^2}{12} + \frac{n}{2} - \frac{n}{3}$

$\text{cov}(X,Y) \to \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Lec 21

$\hat{\Theta}_n$ estimator for param θ, func of $Y_1, \ldots, Y_n$ data pts
estimation error $\tilde{\Theta}_n = \hat{\Theta}_n - \theta$
bias: $\mathbb{E}(\tilde{\Theta}_n)$, unbiased estimator: $\mathbb{E}(\tilde{\Theta}_n) = 0$ or $\mathbb{E}(\hat{\Theta}_n) = \theta$
Consistent: if a sequence $\hat{\Theta}_n$ converges in probability to the true value of parameter θ $\forall$ possible θ

MLE

now $Y_1, \ldots, Y_n$

$$p_Y(Y_1, \ldots, Y_n; \theta) = \prod_{i=1}^{n} p_{Y_i}(Y_i; \theta) \leftarrow \text{likelihood}$$

log likelihood: $\sum_{i=1}^{n} \log(p_{Y_i}(Y_i; \theta))$

example:

arrivals $\sim PP(\theta)$
ith arrival Y_i
ith interarrival time $X_i = Y_i - Y_{i-1}$
estimate θ given $X_1, \ldots, X_n$

$$f_X(X; \theta) = \prod_{i=1}^{n} \underbrace{f_{X_i}(X_i; \theta)}_{\text{exp}} = \prod \theta e^{-\theta X_i} \quad \text{What } \theta \text{ maximizes this?}$$

$$= \sum_{i=1}^{n} \log(\theta e^{-\theta X_i}) = \sum_i \log(\theta) + \log(e^{-\theta X_i})$$

$$= n\log(\theta) - \theta \sum X_i = \boxed{n\log(\theta) - \theta Y_n}$$ log likelihood

$d/d\theta$: $\frac{n}{\theta} - Y_n = 0 \Rightarrow \theta = n/Y_n \to$ empirical rate Θ_{MLE}

$\hat{\Theta}_n = n/Y_n \to Y_n/n = \frac{X_1 + \ldots + X_n}{n} \leftarrow$ Sample mean of interarrival times

By WLLN converges to true mean in prob

$\hat{\theta}_n \rightarrow$ in prob to θ (consistent estimator)

Properties of MLE

1) Invariance, if $\hat{\theta}$ is MLE of θ for any bijective $h(\theta)$, MLE $h(\theta) = h(\hat{\theta}_n)$

Confidence intervals

- A "$1-\alpha$" CI satisfies $\mathbb{P}(\hat{\theta}_n^- \leq \theta \leq \hat{\theta}_n^+) \geq 1-\alpha$

- Standard normal dist.

$Y = X_1 + \dots + X_n \qquad X_n \sim \mu, \sigma^2$

$[Y_-, Y_+]$ found via variance

$$M_n = \frac{X_1 + \dots + X_n}{n} \rightarrow Z_n = \frac{M_n - \mu}{\sigma / \sqrt{n}} \rightarrow N(0,1)$$

- Using CLT and std normal to compute C.I of sample mean if we know σ^2 of X_i.
- $Var(X_i\text{'s}) = \hat{S}_n^2$ (estimator of sample var.)
- $\hat{\theta}_n = \frac{X_1 + \dots + X_n}{n}$ estimate of sample mean
- $Var(\hat{\theta}_n) = \sum Var\left(\frac{X_i}{n}\right) = \frac{1}{n^2} \cdot n \cdot \hat{S}_n^2 = \hat{S}_n^2 / n$

if applied CLT $Z_n = \frac{\hat{\theta}_n - \mu}{\sqrt{\hat{S}_n^2 / n}} \sim N(0,1)$

$Z_{\alpha/2} : \Phi(Z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

$$\mathbb{P}(Z_{\alpha/2}^- \leq Z_n \leq Z_{\alpha/2}^+) = 1-\alpha$$

$$= \mathbb{P}\left(-Z_{\alpha/2} \leq \frac{\hat{\theta}_n - \mu}{S_n / \sqrt{n}} \leq Z_{\alpha/2}\right) = 1-\alpha$$

$$= \mathbb{P}\left(\hat{\theta} - \frac{Z_{\alpha/2} \cdot S_n}{\sqrt{n}} \leq \mu \leq \hat{\theta} + \frac{Z_{\alpha/2} S_n}{\sqrt{n}}\right) = 1-\alpha$$ Noisy estimate!

T dist when n is high