Chapter 3 Review : Conditional Probability & Conditional Expectation

- Discrete Case

def: Let X & Y be discrete RVs. Then the conditional probability mass function (PMF) of X given Y=y is

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

$\forall y$ s.t. $P(Y=y) > 0$.

def: The conditional expectation of X given Y=y is

$$\boxed{E[X \mid Y=y] = \sum_x x\, P(X=x \mid Y=y)}$$

Note: If X is independent of Y, then

$$P(X=x \mid Y=y) = P(X=x)$$

(same is true for CDF & expectation)

- Continuous Case

Recall: Marginal PDF $\quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$

def: If X & Y have joint PDF $f_{XY}(x,y)$, then the conditional PDF of X given Y is

$$\boxed{f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}} \quad \forall y \text{ s.t. } f_Y(y) > 0.$$

def: The conditional expectation of X given $Y=y$ is

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

Computing Expectations by Conditioning (& Variances - skipped)

$E[X|Y]$ is a function of the RV Y whose value at $Y=y$ is $E[X|Y=y]$.

Note: $E[X|Y]$ is itself a random variable.

For all RVs X & Y,

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

To compute $E[X]$, we take a weighted avg. of cond. exp. values (weighted by $P(Y=y)$)

$$\Rightarrow E[X] = \begin{cases} \sum_y E[X|Y=y]\, P(Y=y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} E[X|Y=y]\, f_Y(y)\, dy & \text{if } Y \text{ is continuous} \end{cases}$$

SKIP (Also, $Var(X) = E[Var(X|Y)] + Var(E[X|Y])$)

Example 1: Mean of a geometric distribution (Ex. 3.11 in book)

Flip a coin ($P(\text{Heads}) = p$) until the 1st head appears. What is the expected number of flips required?

Soln: Let N = # of flips required & let

$$Y = \begin{cases} 1 & \text{if 1st flip is heads} \\ 0 & \text{if 1st flip is tails.} \end{cases} \Rightarrow \begin{array}{l} P(Y=1) = p \\ P(Y=0) = 1-p \end{array}$$

Then
$$E[N] = E[N \mid Y=1]\, P(Y=1) + E[N \mid Y=0]\, P(Y=0)$$
$$= p\, E[N \mid Y=1] + (1-p)\, E[N \mid Y=0]. \quad *$$

But $\underbrace{E[N \mid Y=1] = 1}_{\text{why? only takes 1 flip}}$ and $E[N \mid Y=0] = 1 + E[N]$.

1st flip tails so this adds 1 to mean of N

successive flips are indep. so after 1st tail, mean is $E[N]$

Substitute into $*$:

$$E[N] = p(1) + (1-p)(1 + E[N])$$
$$= \cancel{p} + 1 + E[N] - \cancel{p} - p\,E[N]$$

$$\Rightarrow p\,E[N] = 1$$

$$\Rightarrow \boxed{E[N] = 1/p}$$

... Conditioning on 1st event is a useful technique!

[See also Ex 8 in notes PDF → defines compound RV]

Computing Probabilities by Conditioning ← Useful technique!

$$P(X=x) = \sum_{y} P(X=x \mid Y=y)\, P(Y=y) \qquad \text{– discrete case}$$

$$f_X(x) = \int_{-\infty}^{\infty} \underbrace{f_{X|Y}(x|y)}_{\text{cond. PDF}}\ \underbrace{f_Y(y)}_{\text{marginal PDF}}\, dy \qquad \text{– continuous case}$$

Chapter 4: Markov Chains

§4.1: Introduction

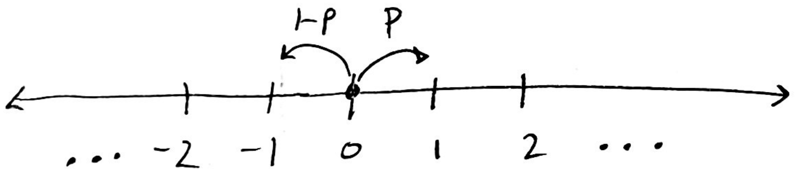
Recall def of stochastic process: a collection of RVs $\{X(t) : t \in T\}$ where t is time, T is an index set ($T \subseteq \mathbb{R}$), each $X(t)$ takes on values in the state space $S \subseteq \mathbb{R}$.

i.e. $X(t)$ is the state of the process at time t.

(OR denoted X_t)

Example 1: Random Walk on $S = \mathbb{Z}$, $T = \{0, 1, 2, \ldots\}$

for some $0 < p < 1$



possible realization of the process

	time	state		
$X(0) = 0$				
$X(1) = 1$	moved	R	by 1 step	
$X(2) = 0$	"	L	"	
$X(3) = -1$	"	L	"	
$X(4) = -2$	"	L	"	
$X(5) = -1$	"	R	"	
$\vdots$				

$$P(X(t+1) = j \mid X(t) = i) = ?$$

* This process only relies on the current state $X(t)$ to decide what the next state $X(t+1)$ will be.
* Past history doesn't matter! ∴ Markov Property

def: Consider a stochastic process $\{X_n : n=0,1,2,\dots\}$. The possible values for X_n are the states of the system. If $X_n = i$, then the process is said to be in state i at time n. This version is a discrete-time stoch. process since the time index n belongs to a countable set $\{0,1,2,\dots\}$.

def: A Markov chain is a discrete-time stochastic process $\{X_n : n=0,1,2,\dots\}$ whose state space is finite or countable (& whose index set is $T=\{0,1,2,\dots\}$), and has the Markov property:

$$\boxed{P(X_{n+1}=j \mid X_0=i_0, X_1=i_1, \dots, X_{n-1}=i_{n-1}, X_n=i) = P(X_{n+1}=j \mid X_n=i)}$$

$\forall$ time points $i_0, i_1, \dots, i_{n-1}, i, j$.

* Only depends on current state i, future state j, and time n.

Example: Random walk on $S=\mathbb{Z}$ & $T=\{0,1,\dots\}$ is a Markov chain

$P(X_{n+1}=i+1 \mid X_n=i) = p$ ← only can move R by 1

$P(X_{n+1}=i-1 \mid X_n=i) = 1-p$ ← move L by 1 def: The movement of the MC (Markov chain) from 1 state to another is called a transition. ... move in ONE STEP

def: Transition probability : $P(X_{n+1}=j \mid X_n=i) = P_{ij}$ (denoted)

$\Rightarrow$ probability of moving from state i (at current time n) to state j (in the next time step, $n+1$)

* In many examples, the transition prob. does NOT depend on time n $\rightarrow$ this is a stationary transition prob.

$$\boxed{P(X_{n+1}=j \mid X_n=i) = P(X_1=j \mid X_0=i)}$$

Arrange these P_{ij}'s in a matrix :

def: $\mathbb{P}$ - matrix of 1-step transition probabilities

Let $S=\{0,1,2,\dots\}$ - state space

$$\begin{array}{c|c} & \begin{array}{cccccc} 0 & 1 & 2 & \cdots & j & \cdots \end{array} \leftarrow \text{future states} \\ \mathbb{P} = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ i \\ \vdots \end{array} & \begin{bmatrix} P_{00} & P_{01} & P_{12} & \cdots & P_{0j} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots & P_{1j} & \\ P_{20} & P_{21} & P_{22} & & \vdots & \\ & \vdots & \vdots & & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \cdots & P_{ij} & \cdots \\ & \vdots & & & \vdots & \end{bmatrix} \end{array}$$

current states

$P_{ij} \geq 0 \ \forall i,j$, $\sum_{j=0}^{\infty} P_{ij} = 1$ for $i=0,1,\dots$ $\leftarrow$ Rows Σ to 1 ✓

Example 2: Forecasting the Weather

Suppose that the chance of rain tomorrow depends only on whether or not it is raining today:

- If it rains today, then it will rain tomorrow w/prob. α
- If it doesn't rain today, then it will rain tomorrow w/prob. β

Assume process is defined as
- → state 0 when it rains
- → state 1 when it doesn't rain

$$\begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \leftarrow \text{tomorrow (future state)} \\ P = \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} \end{array}$$

today (current state)

since rows sum to 1

$P(A \cap B \cap C) = P(C \mid A \cap B) P(B \mid A) P(A)$

Q. Compute $P(\underset{\text{rains}}{X_0 = 0}, \underset{\text{doesn't rain}}{X_1 = 1}, \underset{\text{rains}}{X_2 = 0})$ if $P(X_0 = 0) = 0.8$.

$$P(X_0 = 0, X_1 = 1, X_2 = 0) = P(X_2 = 0 \mid X_1 = 1, X_0 = 0) \cdot \underbrace{P(X_1 = 1, X_0 = 0)}_{= P(X_1 = 1 \mid X_0 = 0) \cdot P(X_0 = 0)}$$

by Markov Property

$$= \underbrace{P(X_2 = 0 \mid X_1 = 1)}_{= P_{10} = \beta} \cdot \underbrace{P(X_1 = 1 \mid X_0 = 0)}_{P_{01} = 1-\alpha} \cdot \underbrace{P(X_0 = 0)}_{0.8}$$

assuming this is stationary transition prob.

$$= \boxed{0.8\beta(1-\alpha)}$$