AMAT 362 - PROBABILITY for STATISTICS

LECTURE 23 : [CONDITIONAL DISTRIBUTIONS for DISCRETE RVs + Intro to BAYESIAN INFERENCE]

+ Law of Iterated Expectations
$E[E[X|Y]] = E[X]$

Q: Suppose I roll a 6-sided fair die
$\Rightarrow X \sim \text{Unif}\{1, \ldots, 6\}$

Now, based on outcome of the roll flip a fair coin that # of times. Denote the # of heads by Y.

$$P(y \text{ heads} \mid x \text{ on the die roll}) = \binom{x}{y}\underbrace{\left(\tfrac{1}{2}\right)^y\left(\tfrac{1}{2}\right)^{x-y}}_{=\left(\frac{1}{2}\right)^x = 2^{-x}}$$

CONDITIONAL DISTRIBUTION

AKA $\boxed{P(Y=y \mid X=x) = \binom{x}{y} 2^{-x}}$

FOR Y GIVEN X

Q: What is the (unconditional) PMF for Y?

A: $$P(Y=y) = \sum_x P(Y=y \mid X=x)\,P(X=x)$$

LAW OF TOTAL PROBABILITY / RULE OF AVERAGED CONDIT.

Know that $X \sim \text{Unif}\{1, \dots, 6\}$

$$P(Y=y) = \sum_x \underbrace{P(Y=y \mid X=x)}_{=\binom{x}{y}2^{-x}} \underbrace{P(X=x)}_{=1/6}$$

$$= \frac{1}{6} \sum_{x=1}^{6} \binom{x}{y} \frac{1}{2^x}$$

$$P(Y=0) = \underset{\uparrow\ P(X=1)}{\frac{1}{6}} \cdot \underset{\uparrow\ P(\text{no heads} \mid X=1)}{\frac{1}{2}} + \frac{1}{6}\left(\frac{1}{2}\right)^2 + \frac{1}{6}\left(\frac{1}{2}\right)^3 + \dots + \frac{1}{6}\cdot\left(\frac{1}{2}\right)^6$$

$$\boxed{P(Y=0) = \frac{1}{6} \cdot \frac{63}{64}}$$

IN GENERAL ... From Pitman 6.1

TABLE 1. Probability $P(Y=y)$ of getting y heads.

y	0	1	2	3	4	5	6
$P(Y=y)$	$\frac{63}{384}$	$\frac{120}{384}$	$\frac{99}{384}$	$\frac{64}{384}$	$\frac{29}{384}$	$\frac{8}{384}$	$\frac{1}{384}$

This essentially marginalization of the joint PMF $P(X=x, Y=y)$

Joint PMF for $P(\underset{X}{\text{die roll}} = x, \underset{Y}{\#\text{ of heads}} = y)$

TABLE 2. Joint distribution table for (X, Y).

		Possible values x for X						Marginal distn. of Y
		1	2	3	4	5	6	
Possible values y for Y	0	$\frac{1}{6}\frac{1}{2}$	$\frac{1}{6}\frac{1}{4}$	$\frac{1}{6}\frac{1}{8}$	$\frac{1}{6}\frac{1}{16}$	$\frac{1}{6}\frac{1}{32}$	$\frac{1}{6}\frac{1}{64}$	$\frac{63}{384}$ ← $P(Y=0)$
	1	$\frac{1}{6}\frac{1}{2}$	$\frac{1}{6}\frac{2}{4}$	$\frac{1}{6}\frac{3}{8}$	$\frac{1}{6}\frac{4}{16}$	$\frac{1}{6}\frac{5}{32}$	$\frac{1}{6}\frac{6}{64}$	$\frac{120}{384}$
	2	0	$\frac{1}{6}\frac{1}{4}$	$\frac{1}{6}\frac{3}{8}$	$\frac{1}{6}\frac{6}{16}$	$\frac{1}{6}\frac{10}{32}$	$\frac{1}{6}\frac{15}{64}$	$\frac{99}{384}$
	3	0	0	$\frac{1}{6}\frac{1}{8}$	$\frac{1}{6}\frac{4}{16}$	$\frac{1}{6}\frac{10}{32}$	$\frac{1}{6}\frac{20}{64}$	$\frac{64}{384}$
	4	0	0	0	$\frac{1}{6}\frac{1}{16}$	$\frac{1}{6}\frac{5}{32}$	$\frac{1}{6}\frac{15}{64}$	$\frac{29}{384}$
	5	0	0	0	0	$\frac{1}{6}\frac{1}{32}$	$\frac{1}{6}\frac{6}{64}$	$\frac{8}{384}$
	6	0	0	0	0	0	$\frac{1}{6}\frac{1}{64}$	$\frac{1}{384}$
	Marginal distn. of X	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

RULE OF AVERAGED CONDITIONALS

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

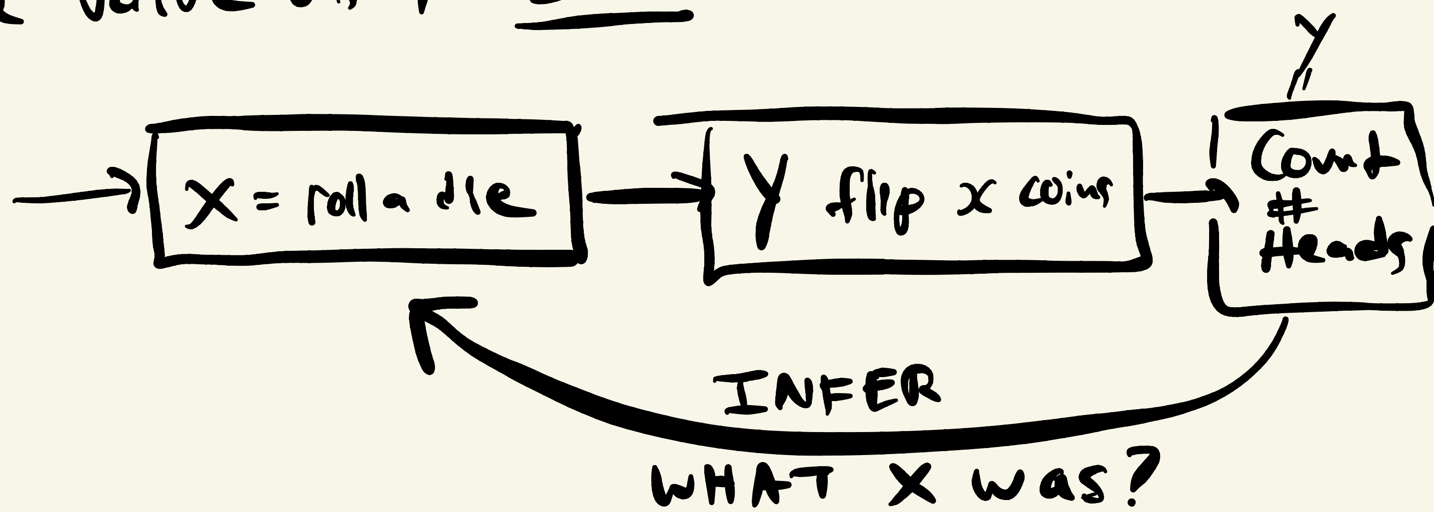
$$= \sum_x \frac{P(Y=y, X=x)}{\cancel{P(X=x)}} \cdot \cancel{P(X=x)}$$

Rule for Marginalization

$$= \sum_x P(Y=y, X=x)$$

EXAMPLE OF BAYES' RULE

Suppose we're told the result of this two step experiment, i.e. we're told the value of Y **BUT** we don't know X



Q: Find $P(X=x \mid Y=y)$

value of roll — # heads observed

A: BAYES RULE

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

Say... $Y=2$ ... $P(X=1 \mid Y=2) = 0$ B/c had 2 heads $\Rightarrow x \geq 2$

$$P(X=2 \mid Y=2) = \frac{\binom{2}{2}\left(\frac{1}{2}\right)^2 \cdot \frac{1}{6}}{\sum_{x=2}^{6} \binom{x}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{x-2} \cdot \frac{1}{6} = P(Y=2) = \frac{99}{384}}$$

$$\boxed{P(X=2 \mid Y=2) = \frac{16}{99}}$$

Repeating this calculation for $x = 3, 4, 5, 6$ conditioned on $Y = 2$

TABLE 3. Conditional distribution of X given $Y = 2$.

x	1	2	3	4	5	6
$P(X = x \mid Y = 2)$	0	$\frac{16}{99}$	$\frac{24}{99}$	$\frac{24}{99}$	$\frac{20}{99}$	$\frac{15}{99}$

STATISTICAL ASIDE

The principle of Maximum Likelihood (Maximal Likelihood Estimator = MLE) says that you should assume the value of x that makes $Y=2$ maximally likely. i.e. $x=3$ or 4 is the most likely explanation for the data. OCCAM'S RAZOR

$\boxed{\text{MODE}}$

BUT OTHER STATISTICS OPTIMIZE DIFFERENT REWARD SCHEMES, i.e. MEDIAN & MEAN

For this example instead of guessing $x=3$ or 4 given $Y=2$

$$E(X \mid Y=2) = \sum x P(X=x \mid Y=2)$$

could be another choice.

EXPECTATION OF CONDITIONAL RVS

(LAW OF ITERATED EXPECTATIONS)

Q: Suppose we want to know $E(Y)$, but we know Y is conditioned on X, so..

$$\underbrace{E(Y \mid X=x)}_{\substack{\text{Binomial} \\ \text{R.V. w/} \\ n=x \\ p=1/2}} = \sum_y y P(Y=y \mid X=x)$$

$$= np$$

$$= x \cdot \frac{1}{2}$$

$$E(Y) = \sum \underbrace{E(Y \mid X=x)}_{\substack{\text{This is just} \\ \text{a function of} \\ \text{RV } X\text{! so..}}} \cdot P(X=x)$$

$E[g(X)] = \sum g(x) P(X=x)$

$$= \sum_{x=1}^{6} \left(\frac{x}{2}\right) \frac{1}{6}$$

$$= \frac{1}{2} \underbrace{\sum_{x=1}^{6} x \cdot P(X=x)}_{E(X) = \frac{7}{2} = 3.5}$$

$$\boxed{E(Y) = \frac{E(X)}{2} = \frac{7}{4} = 1.75}$$

LAW OF ITERATED EXPECTATION

$$E\left[\underbrace{E[Y|X]}_{\text{\# that is dependent on value } x \text{ of } X \text{ i.e. } g(X)}\right] = E[Y]$$

Pf

$$E[E[Y|X]] = \sum_x E[Y|X=x]\,P(X=x)$$

$$= \sum_x \sum_y y\,P(Y=y|X=x)\,P(X=x)$$

$$= \sum_x \sum_y y\,\frac{P(X=x, Y=y)}{\cancel{P(X=x)}}\,\cancel{P(X=x)}$$

Reorder $\sum \sum$

$$= \sum_y y \underbrace{\sum_x P(X=x, Y=y)}_{\text{MARGINALIZES}}$$

$$= \sum_y y\,P(Y=y)$$

$$= E[Y] \quad \blacksquare$$

Turn this result on its head...

Recall this "simpler" problem, but hard to calculate

$$E[X|Y=2] = \sum_x x\, P(X=x|Y=2)$$

$$= \cdots$$

OK, so what is...

$$E[\underbrace{E[X|Y]}_{\text{function of } Y}] = E[X] = \frac{7}{2}$$

WHAT IS GOING ON HERE?

$$P(Y=n_H | X=N, p=\tfrac{1}{2}) = \binom{N}{n_H} p^{n_H} q^{N-n_H}$$

Now fix $n_H=2$ $p=1/2$ let X vary...

$$P(X=x | n_H=2, p=1/2) = \frac{\left[\binom{x}{2} p^2 q^{x-2}\right] \cdot \overbrace{\tfrac{1}{6}}^{P(X=x)}}{\sum_{x=2}^{6} \binom{x}{2} p^2 q^{x-2} \cdot \frac{1}{6}}$$

Computing total probability of $n_H = 2$

FUNDAMENTAL PARADIGM for BAYESIAN STATS

$$\text{POSTERIOR PROBABILITY} = \frac{\text{Likelihood} \times \text{Prior PROBABILITY}}{P(\text{Evidence})}$$