CS 188 Homework 1-1: Probability

Question 1 (2 points):

Consider the following 3-sided dice with the given side values. Assume the dice are all fair and all rolls are independent.

A: 2, 2, 5
B: 1, 4, 4
C: 3, 3, 3

a. What is the expected value of each die?

$$\begin{aligned} E[A] &= \frac{1}{3} * 2 + \frac{1}{3} * 5 = 3 \\ E[B] &= \frac{1}{3} * 1 + \frac{2}{3} * 4 = 3 \\ E[C] &= 1 * 3 = 3 \end{aligned}$$

b. Consider the indicator function better(X,Y) which has value 1 if X¿Y and value -1 if X¡Y. What are the expected values of better(A, B), better(B, C), better(C, A)? Why are these sometimes called non-transitive dice?

$$\begin{aligned} E[better(A, B)] &= \frac{5}{9} * 1 + \frac{4}{9} * (-1) = \frac{1}{9} \\ E[better(B, C)] &= \frac{6}{9} * 1 + \frac{3}{9} * (-1) = \frac{1}{3} \\ E[better(C, A)] &= \frac{6}{9} * 1 + \frac{3}{9} * (-1) = \frac{1}{3} \end{aligned}$$

Transitivity says that if you have a relation R then for all a,b,c: aRb and bRc implies aRc. These are sometimes called non-transitive dice because even though $E[better(A, B)] = \frac{1}{9}$ and $E[better(B, C)] = \frac{1}{3}$, $E[better(A, C)] = -E[better(C, A)] = -\frac{1}{3}$. This basically says that A is likely to beat B since the expected value is positive, and B is likely to beat C since the expected value is positive, but A is unlikely to beat C since its expected value is negative. Because of this they are sometimes called non-transitive dice.

Question 2 (2 points):

Assume that a joint distribution over two variables, $X = x, \neg x$ and $Y = y, \neg y$ is known to have the marginal distributions $P(x) = P(\neg x) = P(y) = P(\neg y)$. Give joint distributions satisfying these marginals for each of these conditions:

Since $P(x) = P(\neg x) = P(y) = P(\neg y)$, then we know $P(x) = P(\neg x) = P(y) = P(\neg y) = 0.5$.

a. X and Y are independent

X	Y	$P(X,Y)$
x	y	.25
x	$\neg y$	.25
$\neg x$	y	.25
$\neg x$	$\neg y$	.25

b. Observing $Y = y$ increases the belief in $X = x$, i.e. $P(x|y) > P(x)$
There are a range of possible answers for this question. Here is one possible solution:

X	Y	$P(X,Y)$
x	y	.3
x	$\neg y$	.2
$\neg x$	y	.2
$\neg x$	$\neg y$	.3

For example, $P(x|y) = \frac{P(x,y)}{P(y)} = .3/.5 = .6 > P(x) = 0.5$.
c. Observing $Y = y$ decreases the belief in $X = x$, i.e. $P(x|y) < P(x)$
There are also a range of possible answers for this question. Here is one possible solution:

X	Y	$P(X,Y)$
x	y	.2
x	$\neg y$	.3
$\neg x$	y	.3
$\neg x$	$\neg y$	.2

For example, $P(x|y) = \frac{P(x,y)}{P(y)} = .2/.5 = .4 < P(x) = 0.5$.

Question 3 (2 points):

On a day when an assignment is due ($A = a$), the newsgroup tends to be busy ($B = b$), and the computer lab tends to be full ($C = c$). Consider the following conditional probability tables for the domain, where $A = a, \neg a$, $B = b, \neg b$, $C = c, \neg c$.

$P(A) =$

A	P
a	.2
$\neg a$	.8

$P(B|A) =$

B	A	P
b	a	.9
$\neg b$	a	.1
b	$\neg a$	.4
$\neg b$	$\neg a$	.6

$P(C|A) =$

C	A	P
c	a	.7
$\neg c$	a	.3
c	$\neg a$	.5
$\neg c$	$\neg a$	.5

a. Construct the joint distribution out of these conditional probabilities tables assuming B and C are independent given A.
$P(A,B,C) = P(B|A) * P(C|A) * P(A)$.

$P(A, B, C) =$

A	B	C	P
a	b	c	.126
a	b	$\neg c$	.054
a	$\neg b$	c	.014
a	$\neg b$	$\neg c$	.006
$\neg a$	b	c	.16
$\neg a$	b	$\neg c$	.16
$\neg a$	$\neg b$	c	.24
$\neg a$	$\neg b$	$\neg c$	.24

b. What is the marginal distribution $P(B, C)$? Are these two variables absolutely independent in this model? Justify your answer using the actual probabilities, not your intuitions.

$P(B, C) = \sum_A P(A, B, C)$.

$P(B, C) =$

B	C	P
b	c	.286
b	$\neg c$	.214
$\neg b$	c	.254
$\neg b$	$\neg c$	.246

B and C are not absolutely independent in this model. If B and C were independent then we should have $P(B, C) = P(B) * P(C)$ for all values of B and C, but this is not true. For example, $P(b, c) = .286$, but $P(b) * P(c) = .5 * .54 = .27$. $P(b, c) \neq P(b) * P(c)$. Therefore they cannot be independent.

c. What is the posterior distribution over A given that $B = b$, $P(A|B = b)$? What is the posterior distribution over A given that $C = c$, $P(A|C = c)$? What about $P(A|B = b, C = c)$? Explain the pattern among these posteriors and why it holds.

We can calculate $P(A|B = b)$ as $P(A|B = b) = P(A, B = b)/P(B = b)$
$P(A = a|B = b) = (.126 + .054)/.5 = .36$
$P(A = \neg a|B = b) = (.16 + .16)/.5 = .64$

We can calculate $P(A|C = c)$ as $P(A|C = c) = P(A, C = c)/P(C = c)$
$P(A = a|C = c) = (.126 + .014)/.54 = .26$
$P(A = \neg a|C = c) = (.16 + .24)/.54 = .74$

We can calculate $P(A|B = b, C = c)$ as $P(A|B = b, C = c) = P(A, B = b, C = c)/P(B = b, C = c)$
$P(A = a|B = b, C = c) = (.126)/.286 = .44$
$P(A = \neg a|B = b, C = c) = (.16)/.286 = .56$

The pattern is that the probability of an assignment being due, A is increased by observing either that the newsgroup is busy ($B = b$) or that the lab is full ($C = c$). If both are observed, the probability of an assignment being due goes up even more.

Question 4 (2 points):

Sometimes, there is traffic (cars) on the freeway ($C = c$). This could either be because of a ball game ($B = b$) or because of an accident ($A = a$). Consider the following joint probability table for the domain, where $A = a, \neg a$, $B = b, \neg b$, $C = c, \neg c$.

$P(A, B, C) =$

A	B	C	P
a	b	c	.018
a	b	$\neg c$	.002
a	$\neg b$	c	.126
a	$\neg b$	$\neg c$	.054
$\neg a$	b	c	.064
$\neg a$	b	$\neg c$	.016
$\neg a$	$\neg b$	c	.072
$\neg a$	$\neg b$	$\neg c$	.648

a. What is the distribution $P(A, B)$? Are A and B independent in this model given no evidence? Justify your answer using the actual probabilities, not your intuitions.

$P(A, B) = \sum_c P(A, B, C)$

A	B	P
a	b	.02
a	$\neg b$	.18
$\neg a$	b	.08
$\neg a$	$\neg b$	.72

If we calculate $P(A, B)$ as $P(A) * P(B)$, where $P(a) = .2, P(\neg a) = .8, P(b) = .1, P(\neg b) = .9$ we get the following table:

A	B	$P(A) * P(B)$
a	b	.02
a	$\neg b$	.18
$\neg a$	b	.08
$\neg a$	$\neg b$	.72

For every value of A and B, we have that $P(A, B) = P(A) * P(B)$. Therefore, A and B are independent.

b. What is the marginal distribution over A given no evidence?

$P(A) = \sum_B P(A, B)$

$P(A = a) = .02 + .18 = .2$
$P(A = \neg a) = .08 + .72 = .8$

c. How does this change if we observe that C=c; what is the posterior distribution P(A — C=c)? Does this change intuitively make sense? Why or why not?

We can calculate $P(A|C = c)$ as $P(A|C = c) = P(A, C = c)/P(C = c)$
$P(A = a|C = c) = .144/.28 = .51$
$P(A = \neg a|C = c) = .136/.28 = .49$

This change makes sense intuitively becuase if there is traffic (C=c) then the probability of an accident (A=a) should go up.

d. What is the conditional distribution over A if we then learn there is a ball game, $P(A|B = b, C = c)$? Does it make sense that observing B should cause this update to A (called explaining-away)? Why or why not?

We can calculate $P(A|B = b, C = c)$ as $P(A|B = b, C = c) = P(A, B = b, C = c)/P(B = b, C = c)$

$P(A = a|B = b, C = c) = .018/.082 = .22$
$P(A = \neg a|B = b, C = c) = .064/.082 = .78$

It makes sense that observing that there is a ballgame (B=b) and traffic (C=c) should make the probability of an accident (A=a) go down because the ballgame can explain the traffic and there is less need for the accident to be causing the traffic (the ballgame explains away the presence of traffic).

Question 5 (2 points):

Often we need to carry out reasoning over some pair of variables X, Y conditioned on the value of other variable E.

a. Using the definitions of conditional probabilities, prove the conditionalized version of the product rule: $P(x, y|e) = P(x|y, e)P(y|e)$

1.) Starting with the left hand side, $P(x, y|e) = \frac{P(x,y,e)}{P(e)}$
2.) We also know, $P(x, y, e) = P(x|y, e) * P(y, e)$
3.) Substituting 2 into 1 we get, $P(x, y|e) = \frac{P(x|y,e)P(y,e)}{P(e)}$
4.) We also know, $P(y, e) = P(y|e)P(e)$
5.) Substituting 4 into 3 we get, $P(x, y|e) = \frac{P(x|y,e)P(y|e)P(e)}{P(e)} = P(x|y, e)P(y|e)$.

b. Prove the conditionalized version of Bayes' rule: $P(y|x, e) = P(x|y, e)P(y|e)/P(x|e)$

1.) Starting with the right hand side, $\frac{P(x|y,e)P(y|e)}{P(x|e)} = \frac{P(x,y|e)}{P(x|e)}$ by part a.
2.) We also know, $P(x|e) = \frac{P(x,y|e)}{P(y|x,e)}$ by part a.
3.) Substituting 2 into 1 we get, $\frac{P(x|y,e)P(y|e)}{P(x|e)} = P(x, y|e)\frac{P(y|x,e)}{P(x,y|e)} = P(y|x, e)$.

Question 6 (2 points):

Suppose we wish to calculate $P(C = c|A = a, B = b)$.

a. If we have no conditional independence information, which of the following sets of tables are sufficient to calculate $P(C = c|A = a, B = b)$?

1. $P(A, B), P(C), P(A|C), P(B|C)$
2. $P(A, B), P(C), P(A, B|C)$
3. $P(A, B, C)$
4. $P(C), P(A|C), P(B|C)$
5. $P(C|A, B), P(A)$

If we have no conditional independence information then 2, 3 and 5 are sufficient to calculate $P(C = c|A = a, B = b)$, but 1 and 4 are not.

b. Which are sufficient if we know that A and B are conditionally independent given C?

If we know that A and B are conditionally independent given C, then we can calculate $P(C = c|A = a, B = b)$ from any of the 5 sets of tables.