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?

\[
E[A] = \frac{1}{3} \times 2 + \frac{1}{3} \times 5 = 3
\]
\[
E[B] = \frac{1}{3} \times 1 + \frac{2}{3} \times 4 = 3
\]
\[
E[C] = 1 \times 3 = 3
\]

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

\[
E[\text{better}(A, B)] = \frac{5}{9} \times 1 + \frac{4}{9} \times (-1) = \frac{1}{9}
\]
\[
E[\text{better}(B, C)] = \frac{6}{9} \times 1 + \frac{3}{9} \times (-1) = \frac{1}{3}
\]
\[
E[\text{better}(C, A)] = \frac{6}{9} \times 1 + \frac{3}{9} \times (-1) = \frac{1}{3}
\]

Transitivity says that if you have a relation \( R \) then for all \( a,b,c \): \( aRb \) and \( b Rc \) implies \( a Rc \). These are sometimes called non-transitive dice because even though \( E[\text{better}(A, B)] = \frac{1}{9} \) and \( E[\text{better}(B, C)] = \frac{1}{3} \), \( E[\text{better}(A, C)] = -E[\text{better}(C, A)] = -\frac{1}{3} \). 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 \( P(x) = P(\neg x) = P(y) = P(\neg y) = 0.5 \).
a. X and Y are independent

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P(X,Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>.25</td>
</tr>
<tr>
<td>X</td>
<td>¬Y</td>
<td>.25</td>
</tr>
<tr>
<td>¬X</td>
<td>Y</td>
<td>.25</td>
</tr>
<tr>
<td>¬X</td>
<td>¬Y</td>
<td>.25</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P(X,Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>.3</td>
</tr>
<tr>
<td>X</td>
<td>¬Y</td>
<td>.2</td>
</tr>
<tr>
<td>¬X</td>
<td>Y</td>
<td>.2</td>
</tr>
<tr>
<td>¬X</td>
<td>¬Y</td>
<td>.3</td>
</tr>
</tbody>
</table>

For example, P(x|y) = \( \frac{P(x,y)}{P(y)} \) = .3/5 = .6 > P(x) = .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:

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P(X,Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>.2</td>
</tr>
<tr>
<td>X</td>
<td>¬Y</td>
<td>.3</td>
</tr>
<tr>
<td>¬X</td>
<td>Y</td>
<td>.3</td>
</tr>
<tr>
<td>¬X</td>
<td>¬Y</td>
<td>.2</td>
</tr>
</tbody>
</table>

For example, P(x|y) = \( \frac{P(x,y)}{P(y)} \) = .2/5 = .4 < P(x) = .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, ¬a, B = b, ¬b, C = c, ¬c.

\[
P(A) =
\begin{align*}
A & \quad P \\
 a & \quad .2 \\
¬a & \quad .8 \\
\end{align*}
\]

\[
P(B|A) =
\begin{align*}
 B & \quad A & \quad P \\
b & \quad a & \quad .9 \\
¬b & \quad a & \quad .1 \\
b & \quad ¬a & \quad .4 \\
¬b & \quad ¬a & \quad .6 \\
\end{align*}
\]

\[
P(C|A) =
\begin{align*}
C & \quad A & \quad P \\
c & \quad a & \quad .7 \\
¬c & \quad a & \quad .3 \\
c & \quad ¬a & \quad .5 \\
¬c & \quad ¬a & \quad .5 \\
\end{align*}
\]

a. Construct the joint distribution out of these conditional probabilities tables assuming B and C are independent given A.

\[
P(A,B,C) = P(B|A) \cdot P(C|A) \cdot P(A).
\]
\[ P(A, B, C) = \]
\[ \begin{array}{ccc}
A & B & C & P \\
a & b & c & .126 \\
a & b & ¬c & .054 \\
a & ¬b & c & .014 \\
a & ¬b & ¬c & .006 \\
¬a & b & c & .16 \\
¬a & b & ¬c & .16 \\
¬a & ¬b & c & .24 \\
¬a & ¬b & ¬c & .24 \\
\end{array} \]

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.

\[ P(B, C) = \sum_A P(A, B, C). \]
\[ P(B, C) = \]
\[ \begin{array}{ccc}
B & C & P \\
b & c & .286 \\
b & ¬c & .214 \\
¬b & c & .254 \\
¬b & ¬c & .246 \\
\end{array} \]

B and C are not absolutely independent in this model. If B and C were independent then we should have \( P(B, C) = P(B) \times P(C) \) for all values of B and C, but this is not true. For example, \( P(b, c) = .286 \), but \( P(b) \times P(c) = .5 \times .54 = .27 \). \( P(b, c) \neq P(b) \times P(c) \). 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 \( P(A|B = b) = P(A, B = b) / P(B = b) \)
\[ P(A = a|B = b) = (.126 + .054) / .5 = .36 \]
\[ P(A = ¬a|B = b) = (.16 + .16) / .5 = .64 \]

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

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

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 \( A = a, ¬a, B = b, ¬b, C = c, ¬c. \)
\[ P(A, B, C) = \]

\[
\begin{array}{ccc|c}
A & B & C & P \\
\hline
a & b & c & .018 \\
a & b & \neg c & .002 \\
a & \neg b & c & .126 \\
a & \neg b & \neg c & .054 \\
\neg a & b & c & .064 \\
\neg a & b & \neg c & .016 \\
\neg a & \neg b & c & .072 \\
\neg a & \neg b & \neg c & .648 \\
\end{array}
\]

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

\[
\begin{array}{ccc|c}
A & B & P(A) \times P(B) \\
\hline
a & b & .02 \\
a & \neg b & .18 \\
\neg a & b & .08 \\
\neg a & \neg b & .72 \\
\end{array}
\]

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

\[ P(A, B) = \sum_c P(A, B, C) \]

\[
\begin{array}{ccc|c}
A & B & P \\\n\hline
a & b & .02 \\
a & \neg b & .18 \\
\neg a & b & .08 \\
\neg a & \neg b & .72 \\
\end{array}
\]

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 \( P(A \mid C = c) \)? Does this change intuitively make sense? Why or why not?

We can calculate \( P(A \mid C = c) \) as \( P(A \mid C = c) = P(A, C = c) / P(C = c) \)

\[
P(A = a \mid C = c) = .144 / .28 = .51 \\
P(A = \neg a \mid C = c) = .136 / .28 = .49 \\
\]

This change makes sense intuitively because 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, \( P(A \mid B = b, C = c) \)? Does it make sense that observing \( B \) should cause this update to \( A \) (called explaining-away)? Why or why not?

We can calculate \( P(A \mid B = b, C = c) \) as \( P(A \mid B = b, C = c) = P(A, B = b, C = c) / P(B = b, C = c) \)

4
\[ P(A = a | B = b, C = c) = .018/.082 = .22 \]
\[ P(A = \neg a | B = b, C = c) = .064/.082 = .78 \]

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.