HW-04 MANUAL SOLUTIONS

0.0.1 Problem 2

Let A and B be events in a sample space Ω.
Suppose that the probability that A occurs is 0.2, the probability that B occurs is 0.6, and the probability
that neither A nor B occur is 0.3 (that is, the joint probability that A doesn't occur AND B doesn't occur
is 0.3).
Determine if you have enough information to answer the following questions. If so, write your solution showing all steps. If not, explain what additional information you would need.

2a) (2 pts) What is P (A, B)?

We can describe what we're given with the following table:

	A	A'	T
B	**P(B,A) =**	**P(B,A')=**	**P(B) = 0.6**
B'	**P(B',A)=**	**P(B',A')= 0.3**	
T	**P(A) = 0.2**		

Using the law of total probability, we can fill in the remaining blanks in the table as follows:

	A	A'	T
B	**P(B,A) =0.1**	**P(B,A')=0.5**	**P(B) = 0.6**
B'	**P(B',A)= 0.1**	**P(B',A')= 0.3**	**P(B') = .4**
T	**P(A) = 0.2**	**P(A') = 0.8**	**1**

Thus 2a). $P(A,B) = 0.1$

2b) (2 pts) What is $P(B \mid A')$?

$P(B|A') = \frac{P(B,A')}{P(A')} = \frac{0.5}{0.8} = \frac{5}{8}$

Hint: Use the information provided to create a joint probability table for A and B.
Write up your full solution to both questions in the SAME box below using LaTeX (not code).
Show all steps fully justifying your answers.

0.0.2 Problem 3

The accuracy of a diagnostic test is often described using the following terms:

- Test Sensitivity: Ability to detect a positive case (i.e. probability that the test is positive given that the person actually has the virus).
- Test Specificity: Ability to determine a negative case (i.e. the probability that a person tests negative given that they don't have the virus)).

Suppose a diagnostic test for a virus is reported to have 90% sensitivity and 92%

specificity. Suppose 2% of the population has the virus in question.

Answer the following questions all in ONE cell below using LaTeX. Show all steps.

3a) (6 pts). If a person is chosen at random from the population and the diagnostic test indicates that they have the virus, what is the conditional probability that they do, in fact, have the virus? Write up your full solution using LaTeX.

Let V be the occasion of having the virus and suppose T is the event of getting a positive test result.

According to the question description then, we know $P(V) = 0.02$.

This means we also know that $P(V') = 0.98$.

Furthermore, the test sensitivity tells us $P(T|V) = 0.9$, i.e. the probability of a positive test

result given that the person actually has the virus.

The test specificty implies $P(T'|V') = 0.92$, the probability of a negative test if the person doesn't have the virus.

Thus $P(T|V') = 1-P(T'|V') = 1-0.92 = 0.08$

We seek $P(V|T)$, i.e. the probability of having the virus, given that we have a positive test result.

By Bayes' Theorem $\displaystyle{P(V|T) = \frac{P(T|V)\cdot P(V)}{P(T|V)\cdot P(V)+P(T|V^C)\cdot P(V^C)} = \frac{(0.9)(0.02)}{[(0.9)(0.02)+(0.08)(0.98)]} = \frac{45}{241} \approx 18.7\%}$

3b) (1 pt). Terminology: What is the **prior** and what is the **likelihood** in this scenario?

Prior - 0.02
Likelihood - 0.9

0.0.3 Problem 4: Poker!

A common example for discrete counting and probability questions are poker hands. Consider using a standard 52-card playing deck, with card ranks [A,2,3,4,5,6,7,8,9,10,J,Q,K] across the standard 4 suits: [C,D,H,S].

Part 4A (4 pts)

Suppose we draw 5 cards at random from the deck without replacement.
In Poker, “Three of a Kind” is defined as a hand that contains three cards of one rank and two cards of two other ranks. Notice that in this definition, a Full House (a hand that contains three cards of one rank and two cards of another rank) is NOT classified as “three of a kind”.
What is the probability of drawing 5 cards (without replacement) that are “three of a kind?”
Typeset your work using LaTeX below. Show work justifying all steps. You may leave your answer in terms of a ratio of products, but you should simplify away any combinatorial notation such as P(n, k).

OPTION 1: Unordered solution:

Idea: Choose one card face to be the 3-of-a-kind. There are $\binom{13}{1}$ such ways. Of those 3 faces, choose their 3 suits. There are $\binom{4}{3}$ such ways. Then choose two *different* card face to be the 4th and 5th cards. There are $\binom{12}{2}$ such faces. Then choose a suits for the 2 unmatched faces. Each has 4 such suits.

Compare this to the number of ways to pick any 5 cards from the deck. There are $\binom{52}{5}$ such ways. Combined, this is:

$$P(\text{3 of a kind})= \frac{\text{3 of a kind}}{\text{All hands}}= \frac{\binom{13}{1}\cdot\binom{4}{3}\cdot \binom{12}{2}\cdot4\cdot4}{\binom{52}{5}}= \frac{13\cdot 4 \cdot12\cdot11\cdot4\cdot4 \cdot 5!}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48\cdot 2}= \frac{88}{4165}\approx 0.02113$$

OPTION 2: ORDERED SOLUTION:

A 2nd way of approaching this problem is to start with calculating the probability for a specific example of 3 of a kind hand. Then count the number of different ways this could occur and add up the probability for the different ways. For example consider the following full house: 3, 3, 3, 4, 5 (note since 3 of a kind is determined by the numbers, here we'll just specify that).

$P(3, 3, 3, 4, 5) = P(3)P(3|\text{already drew 3})P(3|\text{already drew two 3 s})P(4|3, 3, 3)P(5|3, 3, 3, 4) = \frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{4}{49}\frac{4}{48}$ (by the multiplication rule)

Now let's count the number of orders this could occur in (for example 3, 4, 5, 3, 3 etc).

We can reorder these 5 cards in $\binom{5}{3}\binom{2}{1}$ ways (number of ways to choose the 3 spaces for the repeat number and number of ways to choose the unique numbers).

Now we need to count the number of ways we could choose the 3-peat number and the 2 unique numbers. There are 13 ways to choose the 3-peat number. Then there are 12 numbers left and we need to choose 2 of them. This gives us: $\binom{13}{1}\binom{12}{2}$

Thus there are $\binom{5}{3}\binom{2}{1}\binom{13}{1}\binom{12}{2}$ ways to draw a 3 of a kind. And the probability of one specific 3 of a kind is $\frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{4}{49}\frac{4}{48}$

Thus $P(\text{3 of a kind}) = \binom{5}{3}\binom{2}{1}\binom{13}{1}\binom{12}{2}(\frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{4}{49}\frac{4}{48}) = 88/4165

4biii)(4 pts).

Write code in the space below that completes the following steps:

Step 1: Write a function to simulate 10,000 random draws from cards of 5 cards each, and check if each draw is Three of a Kind. The function should return the overall proportion of random hands (out of the 10,000) in which Three of a Kind was observed. If you have coded your simulation correctly, your answer to this part should be very close to your theoretical answer from Part 4A.

Step 2: Let’s visualize how this simulation converges to the theoretical probability. In class, we plotted a running estimate of the probability of an event as a function of the number of trials in our simulation. Write code that completes 10,000 random draws of 5 cards each, but this time outputs a plot of a running estimate of the proportion of hands that are Three of a Kind as a function of the number of trials (from 1 to 10,000) in your simulation. **Include a red horizontal line on your plot with the theoretical probability that you calculated in part 4A.** Be sure to include a title on your plot and be sure to label both your axes on the plot.

In [ ]: ...
```
def card_sim(num_trials):
    FH=[]

    for ii in range(num_trials):
        hand=np.random.choice(cards, replace=False, size=5)
        FH.append(three_kind(hand))
    return np.sum(FH)/num_trials

card_sim(10000)



#4biii).Write your code for Step 1 above this line
```

In [ ]:
```
hand=np.random.choice(cards, replace=False, size=5)

hand
```

In [ ]: ...
```
def plot_estimates(num_trials):
    FH=[]
    running_prob = []
    for ii in range(num_trials):
        hand=np.random.choice(cards, replace=False, size=5)
        FH.append(three_kind(hand))
        running_prob.append(np.sum(FH)/(ii+1))
    return running_prob

p=plot_estimates(10000)


fig, ax = plt.subplots(figsize=(12,6))
ax.plot(p)

ax.grid(True, alpha=0.25)
ax.set_axisbelow(True)
ax.set_xlabel("number of trials", fontsize=16)
ax.set_ylabel("Estimate of prob of three of kind", fontsize=16)
```

```
plt.axhline(y = 0.021, color = 'r', linestyle = '-')
  # 4biii) Write your code for Step 2 above this line
```

0.0.4 Problem 5

To play a game, you have a bag containing 28 fair four-sided dice, with faces {1,2,3,4}. This bag also contains 9 fair six-sided dice (faces {1,2,3,4,5,6}) and 3 fair twenty-sided dice (faces {1,2,3,4,...,19,20}).

Call these 3 classes of die "Four", "Six" and "Twenty" (or D4, D6, and D20, for short). You grab one die at random from the box.

Work the following problems by hand and write up your full solution using LaTeX unless otherwise stated (but don't be afraid to simulate to check your result!).

Part 5A (3 pts): You grab one die at random from the box and roll it one time. What is the probability of the event R5, that you roll a 5? Explain your reasoning mathematically (using LaTeX).

Solution:

$P(R5)=P(R5|D4)P(D4)+P(R5|D6)P(D6)+P(R5|D20)P(D20)$

=0+(1/6 9/40)+(1/20 3/40)=9/240+3/800

=33/800

Part 5B (3 pts): Suppose you roll a 5. Given this information, what is the probability that the die you chose from the box is a Six-sided die? Write up your full solution using LaTeX. Show all steps.

Solution:

We use Bayes' Theorem:

$P(D6|R5)=P(R5|D6)P(D6)$ / $P(R5)$=(1/6·9/40)/(33/800)=(3/80)/(33/800)

=10/11

Part 5C (3 pts): Are the events R5 and D6 independent? Write up your full solution using LaTeX. Show all steps. Justify your answer using the mathematical definition of independence.

Solution:

No they are not independent. We know this because $P(R5) \neq P(R5|D6)$

$P(R5) \approx 0.044167$ (from Part A)

$P(R5|D6) = 16 \approx 0.167$ (from part of Part B)

NB: We could also have checked either:

$P(R5 \cap D6) = ? P(R5)P(D6)$

, or

$P(D6|R5) = ? P(D6)$

0.0.5 Problem 6

Suppose you roll two fair six-sided dice. Let C be the event that the two rolls are close to one another in value, in the sense that they're either equal or differ by only 1.

Part 6A (4 pts): Compute P (C) by hand. Show all steps using LaTeX.

Solution:

$P(C) = P(\text{same}) + P(\text{differ by one})$

We can start with the more straightforward part, $P(\text{same})$
.

There are a total of 36 possible outcomes (not necessarily distinct sums of the dice), and each is equally likely with probability 1/36.
Six of these 36 outcomes correspond to rolling the same number of both dice (1-1, 2-2, and so on).
Thus, the probability of rolling the same on both dice is $P(\text{same}) = 6 \cdot 1/36 = 6/36$

Now it is time for the slightly tougher part, calculating $P(\text{differ by one})$

There are five ways for the two dice to come up differing by one: 1-2, 2-3, 3-4, 4-5 and 5-6
Each of these five ways has two possible "orientations": 1-2 and 2-1, for example
This gives 10 distinct outcomes of the total 36, where the dice differ by one
Thus, the probability of rolling two numbers that differ by one is $P(\text{differ by one}) = 10 \cdot 1/36 = 10/36$

So we have $P(C) = 6/36 + 10/36 = 16/36 \approx 0.444$

Part 6B (3 pts): Write a simulation to run 10,000 trials of rolling a pair of dice and estimate the value of P(C) you calculated in Part A. Your estimate should agree with the exact calculation you did in Part A. If it doesn't, try increasing the number of trials in your simulation.

In [ ]: ...

```
def dice_sim(num_trials=1000):

    roll1 = np.random.choice(np.array([1,2,3,4,5,6]), size=num_trials)
    roll2 = np.random.choice(np.array([1,2,3,4,5,6]), size=num_trials)

    running_prob = np.array([np.sum(np.abs(roll1[1:ii+1]-roll2[1:ii+1])<=1)/(ii+1) for ii in range(num_trials-1)])

    return running_prob

p = dice_sim(num_trials=100000)
print("P(C) \u2248 {}".format(p[-1]))
 #Your code above this line
```

Part 6C (3 pts): In class we plotted a running estimate of the probability of an event as a function of the number of trials in our simulation. Write code to run 5 independent simulations of 50,000 trials each to estimate �� (��) and plot their running estimate curves on the same set of axes. **Hint**: This is a lot of computation, so try to leverage Numpy and **list comprehensions** as much as possible so that your code doesn't run forever.

Include a red horizontal line on your plot with the theoretical probability that you calculated in part 6A. Be sure to include a title on your plot and be sure to label both your axes on the plot.

In [ ]: ...

```
def dice_sim(num_trials=1000):

    roll1 = np.random.choice(np.array([1,2,3,4,5,6]), size=num_trials)
    roll2 = np.random.choice(np.array([1,2,3,4,5,6]), size=num_trials)

    running_prob = np.array([np.sum(np.abs(roll1[1:ii+1]-roll2[1:ii+1])<=1)/(ii+1) for ii in range(num_trials-1)])

    return running_prob

def plot_estimates(num_sims=10):

    fig, ax = plt.subplots(figsize=(12,6))
```

```
    for ii in range(num_sims):
        p = dice_sim(50000)
        ax.plot(p)

    ax.grid(True, alpha=0.25)
    ax.set_axisbelow(True)
    ax.set_xlabel("number of trials", fontsize=16)
    ax.set_ylabel("Estimate of P(C)", fontsize=16)
    ax.set_ylim([.34,.54])

plot_estimates()
 # Your code above this line
```

Part 6D (1 pt): Describe the behavior of the running estimates as the number of trials increases.

i). What value(s) are they converging to?
ii). How many trials does it take until they appear to converge?

Solution:

The different estimates take about 10,000 iterations to start to converge on 44%. But as the number of trials increases, the ten estimates start to blend together, although we can see that even by 50,000 trials, the ten estimates are not yet indistinguishable from one another.