Back to top

### 0.0.1 Problem 2

Let $A$ and $B$ be events in a sample space $\Omega$.

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.

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

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

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 2a Solution

The formula to calculate $P(A, B)$, otherwise known as $P(A \cap B)$ is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \rightarrow \quad P(A \cap B) = P(A) + P(B) - P(A \cup B).$$

Furthermore, we can define the intersection of two probabilities to be

$$P(A \cup B) = 1 - P(A' \cap B').$$

In this context, $P(A' \cap B') = 0.3$. Using the above formulae we can then say

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = P(A) + P(B) - 1 + P(A' \cap B') = 0.2 + 0.6 - 1 + 0.3 = 1.1 - 1 = 0.1.$$

### 0.0.3 2b Solution

Mathematically this is

$$P(B|A') = \frac{P(A' \cap B)}{P(A')}$$

but in order to solve this, we need to calculate $P(A' \cap B)$. This is easiest answered by making a table of the probabilities:

| | A | A' | Accumulative |
|---|---|---|---|
| B | 0.1 | 0.5 | 0.6 |
| B' | 0.1 | 0.3 | 0.4 |
| Accumulative | 0.2 | 0.8 | 1 |

From this, we can see that $P(A' \cap B) = 0.5$. Using this with the other values we can then say:

$$P(B|A') = \frac{1/2}{4/5} = \frac{5}{8}.$$

Back to top

### 0.0.4 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) (5 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.

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

### 0.0.5 3a Solution

To answer this question, we are essentially attempting to solve

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

In the above, the terms are:

- $P(A|B)$: The probability of having the virus given a positive test result.
- $P(B|A)$: The probability of testing negative given that they don't have the virus.
- $P(A)$: The prevalence of the virus.
- $P(B)$: The probability of testing positive.

$P(B)$ is calculated by summing the false positive results with the true negative results, namely

$$P(B) = P(B|A)P(A) + P(B|A')P(A').$$

Here:

- $P(B|A')$: The probability of a false negative.
- $P(A')$: The probability of not having the virus.

Stitching this together we then have

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')} = \frac{(0.90)(0.02)}{(0.90)(0.02) + (0.08)(0.98)} \approx= 0.1867 = \frac{45}{241}.$$

### 0.0.6 3b Solution

In this scenario, the **prior** is referring to the prevalence of the virus, or in English, how many people have the virus: 2%. The **likelihood** is referring to the tests sensitivity, or the likelihood of a true positive test: 0.90%.

---

Back to top

### 0.0.7 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 (3 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". https://en.wikipedia.org/wiki/List_of_poker_hands#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 combinatoric notation such as $\binom{n}{k}$ or $P(n, k)$.

To answer this question, we need to meticulously think about how the cards are being chosen. We are selecting five cards at random from a deck of cards, namely $\binom{52}{5}$. We then need to choose one rank from the 13 ranks for the three of kind part of a full house. Mathematically this is $\binom{13}{1}$. We then need to choose three cards from the chosen rank, so $\binom{4}{3}$ (4 because there are four cards from this rank). We then need to select two cards from the other possible ranks, so this is $\binom{12}{2}$ (12 because we are excluding the previous rank from the last chosen rank for the. three of a kind part). For these last two cards, we need to choose a card from the another two ranks. As the problem statement says, we don't need them to be the same. This is then $\binom{4}{1}^2$. Putting this together we then get

$$
\begin{aligned}
P &= \frac{\binom{13}{1} \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot \binom{4}{1}^2}{\binom{52}{5}} \\
&= \frac{\frac{13!}{1!(13-1)!} \cdot \frac{4!}{3!(4-3)!} \cdot \frac{12!}{2!(12-2)!} \cdot \left(\frac{4!}{1!(4-1)!}\right)^2}{\frac{52!}{5!(52-5)!}} \\
&= \frac{13 \cdot 4 \cdot 66 \cdot 16}{2598960} = \frac{54912}{2595960}
\end{aligned}
$$

and simplifying the last part of the solution we then get

$$P = \frac{88}{4165} \approx 0.0211.$$

**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.

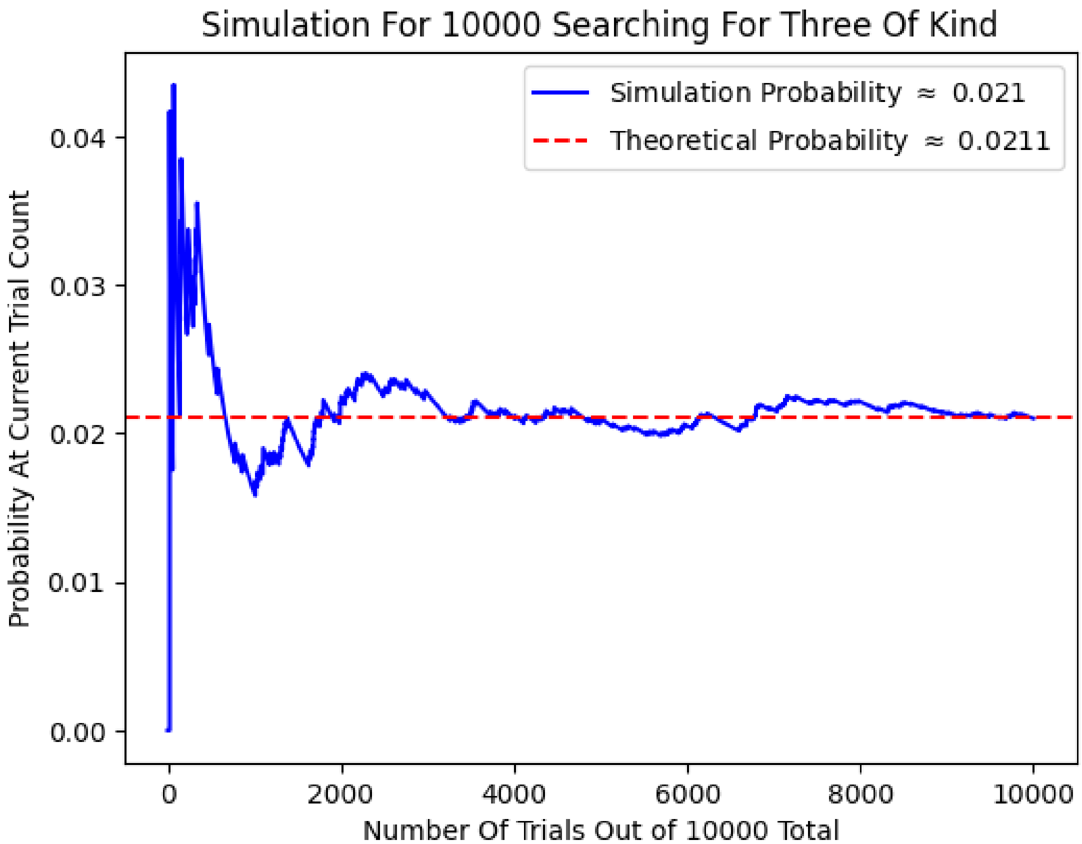
In [23]:
```
def threeOfKind(trials):
    success = 0
    successArr = np.zeros(trials)
    probArr = np.zeros(trials)
    trialsArr = np.arange(1, trials + 1)
    for i in range(trials):
        testHand = np.random.choice(cards, size = 5, replace = False)
        threeOfKind = three_kind(testHand)
        if (threeOfKind):
            success += 1
        successArr[i] = success
        probArr[i] = successArr[i] / trialsArr[i]
    return (success / trials), successArr, probArr, trialsArr

print(f"Calculated Probability: {threeOfKind(10000)[0]}")
#4biii).Write your code for Step 1 above this line
```

```
Calculated Probability: 0.0221
```

In [24]:
```
trials = 10000
threeOfKindSim = threeOfKind(trials)
plt.plot(threeOfKindSim[3], threeOfKindSim[2], color='b', label=f'Simulation Probability $\\ap
plt.axhline(y=(88 / 4165), color='r', linestyle='--', label=f'Theoretical Probability $\\appro
plt.xlabel(f"Number Of Trials Out of {trials} Total")
plt.ylabel(f"Probability At Current Trial Count")
plt.title(f"Simulation For {trials} Searching For Three Of Kind")
plt.legend()
plt.show()

# 4biii) Write your code for Step 2 above this line
```


Simulation For 10000 Searching For Three Of Kind
Simulation Probability ≈ 0.021
Theoretical Probability ≈ 0.0211
Probability At Current Trial Count
0.04
0.03
0.02
0.01
0.00
0
2000
4000
6000
8000
10000
Number Of Trials Out of 10000 Total

Back to top

#### 0.0.8 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 $D_4$, $D_6$, and $D_{20}$, 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 $R_5$, that you roll a 5? Explain your reasoning mathematically (using LaTeX).

There are a total of 12 di in this scenario that actually have a 5 on them. There are a total of 28 di in this example that do not have a 5 on them. The only chance that we get a 5 from this scenario is if we first pick a di that actually has a 5 on it, and then roll a 5 from there. So, we need to sum the probabilities of obtaining a 5 on each di, namely

$$
\begin{aligned}
R_5 &= (R_{D_4})_5 + (R_{D_6})_5 + (R_{D_{20}})_5 && (1)\\
&= \frac{28}{40} \cdot \frac{0}{4} + \frac{9}{40} \cdot \frac{1}{6} + \frac{3}{40} \cdot \frac{1}{20} && (2)\\
&= 0 + \frac{9}{240} + \frac{3}{800} = \frac{33}{800} && (3)
\end{aligned}
$$

finally, we have

$$\frac{33}{800} \approx 0.0413.$$

**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.

For this problem we are essentially calculating

$$P(D_6|R_5) = \frac{P(R_5 \cap D_6)}{P(R_5)}.$$

For the first part $P(R_5 \cap D_6)$, this is the probability of selecting a 5 from a six sided di, so $\frac{9}{40} \cdot \frac{1}{6} = \frac{9}{240}$. $P(R_5)$ is the total probability of selecting 5 out of all the di. Stitching this together we then have

$$P(D_6|R_5) = \frac{9/240}{33/800} = \frac{10}{11}$$

giving us the final probability of

$$P(D_6|R_5) = \frac{10}{11} \approx 0.9091.$$

**Part 5C (2 pts):** Are the events $R_5$ and $D_6$ independent? Write up your full solution using LaTeX. Show all steps. Justify your answer **using the mathematical definition of independence**.

Two events are considered to be independent if the following statement is found to be true

$$P(A \cap B) = P(A) \cdot P(B).$$

In this context, $P(A)$ refers to the probability of rolling a 5 and $P(B)$ corresponds to the probability of selecting a six sided di. From the previous part we know

$$P(R_5 \cap D_6) = \frac{10}{11}.$$

Furthermore, we can show that

$$P(R_5) \cdot P(D_6) = \frac{33}{800} \cdot \frac{9}{40} = \frac{297}{32000}.$$

Comparing these two we can see

$$\left(P(R_5 \cap D_6) = \frac{10}{11}\right) \neq \left(P(R_5) \cdot P(D_6) = \frac{297}{32000}\right).$$

Thus, we can conclude that these events are dependent.

Back to top

#### 0.0.9 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 (3 pts)**: Compute $P(C)$ by hand. Show all steps using LaTeX.

In essence, there are going to be a total of 36 possible combinations of rolled di, we just need to count the close values:

- (1,1), (1,2)
- (2,1), (2,2), (2,3)
- (3,2), (3,3), (3,4)
- (4,3), (4,4), (4,5)
- (5,4), (5,5), (5,6)
- (6,5), (6,6)

Counting the above, we have a total of 16 total possible outcomes where this is true. Therefore we can then say the probability $P(C)$ occurring is then

$$P(C) = \frac{16}{36} = 0.\bar{4}.$$

**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.

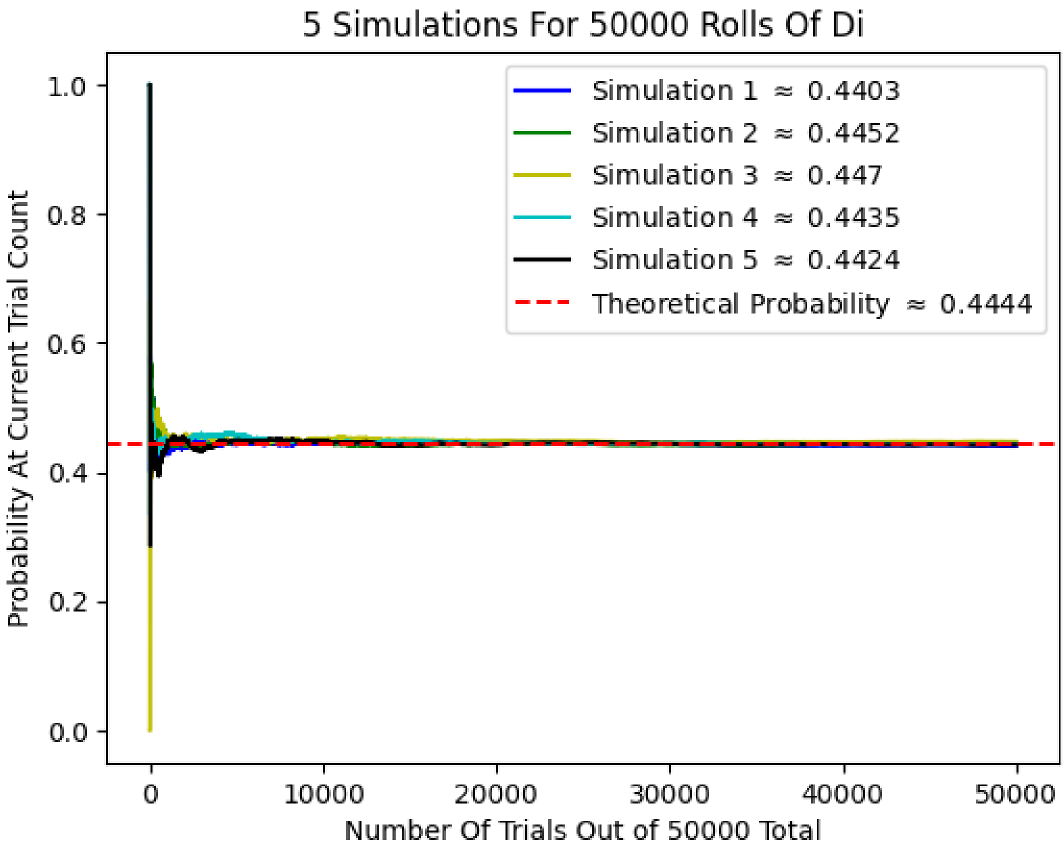
In [33]:
```
# Here I am making the number of trials arbitrary, simply pass in 10,000 to the function to ge
def DiRolls(trials):
    di1 = [1, 2, 3, 4, 5, 6]
    di2 = [1, 2, 3, 4, 5, 6]
    success = 0
    successArr = np.zeros(trials)
    probArr = np.zeros(trials)
    trialsArr = np.arange(1, trials + 1)
    for i in range(trials):
        roll1 = np.random.choice(di1, size=1, replace=True)
        roll2 = np.random.choice(di2, size=1, replace=True)
        val1 = roll1[0]
        val2 = roll2[0]
        if ((val1 == val2) or (np.abs(val1 - val2) == 1)):
            success += 1
        successArr[i] = (success)
        probArr[i] = successArr[i] / trialsArr[i]
    return (success / trials), successArr, probArr, trialsArr
testRun = DiRolls(10000)
print(f"Calculated Probability: {testRun[0]}")
#Your code above this line
```

```
Calculated Probability: 0.441
```

**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 $P(C)$ and plot their running estimate curves on the same set of axes. **Hint**: This is a lot of computation, so try to leverage Numpy 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 [34]:
```
def DiSimulation(trials, simulations):
    indSims = [np.array([]) for _ in range(simulations)]
    for i in range(simulations):
        indSims[i] = DiRolls(50000)
    plt.plot(indSims[0][3], indSims[0][2], color='b', label=f'Simulation 1 $\\approx$ {np.round
    plt.plot(indSims[1][3], indSims[1][2], color='g', label=f'Simulation 2 $\\approx$ {np.round
    plt.plot(indSims[2][3], indSims[2][2], color='y', label=f'Simulation 3 $\\approx$ {np.round
    plt.plot(indSims[3][3], indSims[3][2], color='c', label=f'Simulation 4 $\\approx$ {np.round
    plt.plot(indSims[4][3], indSims[4][2], color='black', label=f'Simulation 5 $\\approx$ {np.r
    plt.axhline(y=(16 / 36), color='r', linestyle='--', label=f'Theoretical Probability $\\appr
    plt.xlabel(f"Number Of Trials Out of {trials} Total")
    plt.ylabel(f"Probability At Current Trial Count")
    plt.title(f"{simulations} Simulations For {trials} Rolls Of Di")
    plt.legend()
    plt.show()
DiSimulation(50000, 5)
# Your code above this line
```


5 Simulations For 50000 Rolls Of Di
Simulation 1 ≈ 0.4403
Simulation 2 ≈ 0.4452
Simulation 3 ≈ 0.447
Simulation 4 ≈ 0.4435
Simulation 5 ≈ 0.4424
Theoretical Probability ≈ 0.4444
Probability At Current Trial Count
Number Of Trials Out of 50000 Total

**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?

#### 0.0.10 Solution 6Di

The values are converging to the theoretical value, as expected.

#### 0.0.11 Solution 6Dii

It appears that around 10,000 total trials for each simulation, this is where the values converge. Some occur earlier, but this is where it seems that most are around that value.

Back to top

## 0.1 Problem 7

Three brands of coffee, $X$, $Y$ and $Z$, are to be ranked according to taste by a single judge. Define the following events:

Event A: Brand $X$ is preferred to $Y$

Event B: Brand $X$ is ranked best.

Event C: Brand $X$ is ranked second best.

Event D: Brand $X$ is ranked third best.

If the judge actually has no taste preference and just randomly assigns ranks to the brands, which of the following events are independent and which are dependent? Justify your answers using the mathematical definition of independence. Write up your full solution using LaTeX.

**Part 7a (2 pts)**. Are events $A$ and $B$ independent or dependent?

**Part 7b (2 pts)**. Are events $A$ and $C$ independent or dependent?

**Part 7c (2 pts)**. Are events $A$ and $D$ independent or dependent?

Answer all 3 parts using LaTeX in the ONE cell provided below.

Referring to earlier in this notebook, the mathematical definition of independence is defined to be

$$P(A \cap B) = P(A)P(B).$$

If the LHS of the aforementioned expression is not equal to that of the RHS, then the relationship is defined to be **dependent**.

The coffees can be ranked in the following permutations:

- $X, Y, Z$
- $X, Z, Y$
- $Y, X, Z$

- $Y, Z, X$
- $Z, X, Y$
- $Z, Y, X$

### 0.1.1 Solution 7a

Event A occurs three times of the six possibilities. Event B occurs two times out of the six possibilities. In this context, for $X$ to be preferred over $Y$ and ranked as best, this only occurs two times out of the six possibilities. Mathematically this is

$$\left(P(A \cap B) = \frac{2}{6}\right) \neq \left(P(A)P(B) = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6}\right)$$

and therefore these Events are **dependent**.

### 0.1.2 Solution 7b

Event C occurs two times out of the six possibilities. In this context, for $X$ to be preferred over $Y$ and be ranked second best, this only occurs one time out of the six possibilities. Mathematically this is

$$\left(P(A \cap C) = \frac{1}{6}\right) = \left(P(A)P(C) = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6}\right)$$

and therefore these Events are **independent**.

### 0.1.3 Solution 7c

Event D occurs two times out of the six possibilities. In this context, it is not possible for $X$ to be preferred over $Y$ and ranked third best. Mathematically this is

$$(P(A \cap D) = 0) \neq \left(P(A)P(D) = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6}\right)$$

and therefor these Events are **dependent**.