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.

2a)

$P(A, B) = P(A) + P(B) - (A \cup B)$

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

$P(A, B) = P(A) + P(B) - (1 - P(A' \cap B'))$

$P(A, B) = 0.2 + 0.6 - (1 - 0.3) = 0.1$

2b)

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

Where $P(A') = 1 - P(A) = 1 - 0.2 = 0.8$

and $P(B \cap A') = P(B) - P(A \cap B)$

We know that $P(A \cap B) = 0.1$ from $2a)$

Thus $P(B \cap A') = P(B) - P(A \cap B) = 0.6 - 0.1 = 0.5$

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

Back to top

#### 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) (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?

3a) $P(Virus|Positive) = \frac{P(Positive|Virus) \cdot P(Virus)}{P(Positive|Virus) \cdot P(Virus) + P(Positive|NoVirus) \cdot P(NoVirus)}$

The ability to detect a positive case is 90%. Ability to detect negative case is 92. However we need a (Positive|No Virus) which will be 1 - 0.92 = 0.08. We see that 2% of the population has virus. And that 98% don't have a virus.

We plug these values in

$P(Virus|Positive) = \frac{0.9 \cdot 0.02}{0.9 \cdot 0.02 + 0.98 * 0.08}$

3b)

Prior is what the population has virus before the test result. Thus the Prior is 2%

The likelihood is the probability of getting the positive test result that the person actually has the virus. Thus the likelihood is 90%

Back to top

### 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 (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 $P(n,k)$.

First we see that a deck has 52 cards, 13 ranks from A to K. We need to choose one of the 13 ranks for the three cards.

After choosing a rank, we need to choose 4 out of that 3 for "three of a kind" For the remaining 2 cards, we choose one card from each rank out of the 12 to choose from 2 times. When we choose those cards, we need to choose 4 out of 1 for just one of a kind. It is not 13 and it is 12 because then we may get "4 of a kind." Finally we divide the above with $(\binom{52}{5})$

Thus we can contruct a formula,

$\frac{\binom{13}{1}\cdot\binom{4}{3}\cdot\binom{12}{2}\cdot\binom{4}{1}\cdot\binom{4}{1}}{\binom{52}{5}}$

**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.

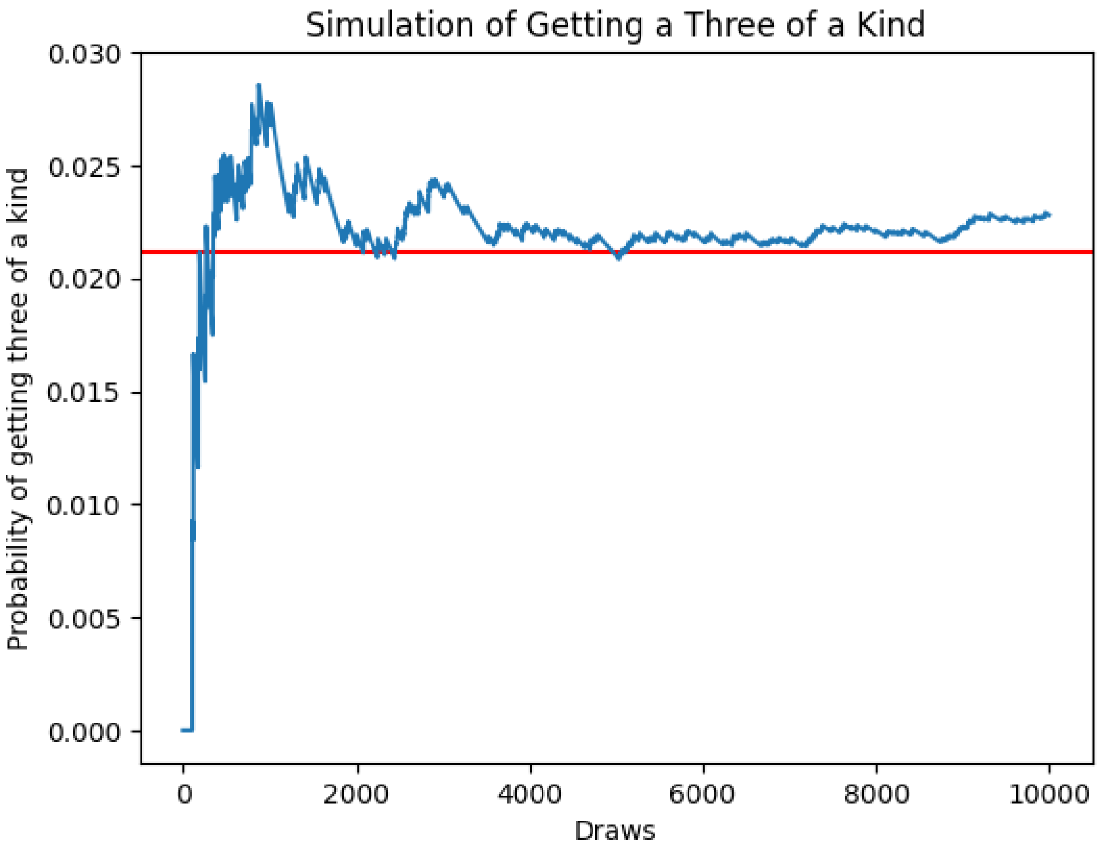
Step 2: Let's visualize how this simulation converges to the theoretical probability. In class, we plotted a running estimate of the probability of an event as a function of the number of trials in our simulation. Write code that completes 10,000 random draws of 5 cards each, but this time outputs a 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 [334]:
```
p=[]
def sim(count):
    n = 0
    for i in range(count):
        randomHand = np.random.shuffle(cards)
        randomHand = cards[:5]
        if three_kind(randomHand) == True:
           n += 1

        into_p = n/(i+1)
        p.append(into_p)
    result = n/count
    return result

sim(10000)
#4biii).Write your code for Step 1 above this line
```

Out[334]: 0.0228

In [335]:
```
fig, ax = plt.subplots()
ax.axhline(y=54912/2598960, color = "red")
ax.plot(p)
plt.title('Simulation of Getting a Three of a Kind')
plt.xlabel('Draws')
plt.ylabel('Probability of getting three of a kind')
# 4biii) Write your code for Step 2 above this line
```

Out[335]: Text(0, 0.5, 'Probability of getting three of a kind')


Simulation of Getting a Three of a Kind
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Probability of getting three of a kind
0
2000
4000
6000
8000
10000
Draws

Back to top

### 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 $D_4$, $D_6$, 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).

We see that grabbing the die is independent, and the number we get from the die is dependent. This is the die we pick has different probabilties when we role it.

To calculate the probability of $P(R_5)$ we can use the multiplication rule:

First we can exclude $(D_4)$ as there is no 5 in it. Thus the probability would be 0. When we grab a die and the probability for $(D_6) = \frac{9}{40}$ If we role a die for $(D_6)$ the probability is $\frac{1}{6}$ Thus the total probability for $(D_6)$ is $\frac{1}{6} \cdot \frac{9}{40} = \frac{9}{240}$

When we grab a die and the probability for $(D_{20}) = \frac{3}{40}$ If we role a die for $(D_{20})$ the probability is $\frac{1}{20}$ Thus the total probability for $(D_{20})$ is $\frac{1}{20} \cdot \frac{3}{40} = \frac{3}{800}$

We can then add these two results to find that the probability of rolling the event $R_5$ is $D_6 + D_{20} = \frac{9}{240} + \frac{3}{800} = \frac{33}{800}$

**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.

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

Where $P(A) = \frac{9}{240}$

$P(B) = \frac{33}{800}$

$P(B|A) = \frac{1}{6}$

When we plug all the values in we get

$P(A|B) = \frac{\frac{9}{240} \cdot \frac{1}{6}}{\frac{33}{800}} = \frac{2400}{2640} = \frac{10}{11}$

**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**.

**We see that to prove independency that it needs to be** We know for $P(D_6)$ which is probability of rolling the die, is $\frac{9}{40}$.

We calculated $P(D_6|R_5) = \frac{10}{11}$

Thus from the first step, we see that $P(D_6|R_5) \neq P(R_5)$.

Thus we know these events are dependent of each other.

Back to top

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

First we roll the first die. When we throw the second, we have the 6 possibilies where both dice roll the same. (1,1), (2,2), (3,3), (4,4), (5,5), (6,6). The other could differ by 1. (1,2), (2,1), (2,3),(3,2), (3,4),(4,3), (4,5), (5,4), (5,6),(6,5). We see the total count to be 16 different possibilites. Now we divide this over by the total probability which is 36.

Thus we see the $P(C) = \frac{16}{36} = \frac{4}{9}$

**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 [344]:
```
def sim(n):
    count = 0
    z = []
    for i in range(n):
        r1 = np.random.randint(1,7)
        r2 = np.random.randint(1,7)
        if((r1 == r2) or (r1 - r2) == 1 or (r2 - r1) == 1):
           count += 1
        z.append(count/(i+1))
    result = count/n
    return result,z

sim(10000)[0]

#Your code above this line
```

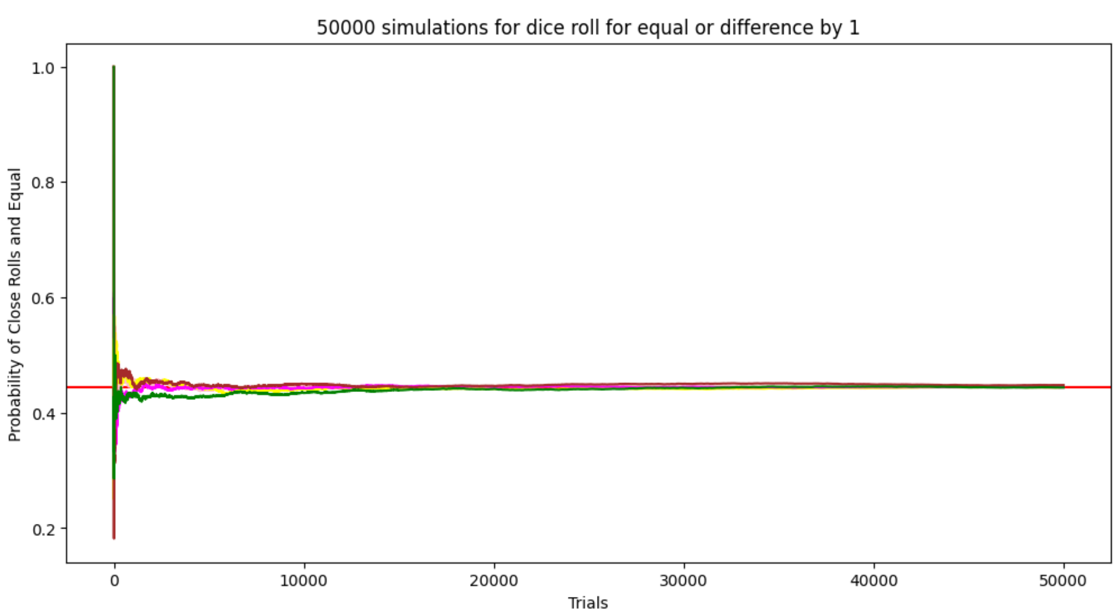
Out[344]:
```
0.4317
```

**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 [345]:
```
fig, ax = plt.subplots(figsize=(12, 6))
ax.axhline(y=16/36, color = "red", label = 'Theoretical Probability')
l1 = sim(50000)[1]
l2 = sim(50000)[1]
l3 = sim(50000)[1]
l4 = sim(50000)[1]
l5 = sim(50000)[1]
ax.plot(l1, color = 'pink')
ax.plot(l2, color = 'yellow')
ax.plot(l3, color = 'Magenta')
ax.plot(l4, color = 'brown')
ax.plot(l5, color = 'green')
plt.title("50000 simulations for dice roll for equal or difference by 1")
plt.xlabel("Trials")
plt.ylabel("Probability of Close Rolls and Equal")
```

Out[345]:
```
Text(0, 0.5, 'Probability of Close Rolls and Equal')
```



**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?

i)

They all converge to 4/9

ii)

It takes about 15000 trials until they seem to converge. This is why we run many simulations in order to remove any uncertainties.

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.

7a)

We see that if by Event B, Brand X is ranked best, and that in Event A Brand X is preffered more than Brand Y, we see that it is dependent. Thus we need to show that $P(A) \neq P(A|B)$ , where $P(A) = \frac{1}{3}$, while $P(A|B) = 1$. Thus we show $P(A) \neq P(A|B)$ and it is dependent.

7b)

The information in with A or C is vague. We know that Brand X is second best, but it doesn't tell anything compared to Brand Y. We can show its independency that $P(C|A) = P(C)$, where $P(C|A) = \frac{1}{3}$ and $P(A) = \frac{1}{3}$. Thus we show $P(C|A) = P(C)$ that it is independent

7c)

We know that Brand X is ranked third or last. This means that there is 0 chance that X is preferred to Y. Thus we see $P(D|A) = 0$ and $P(D) = \frac{1}{3}$, that $P(D|A) \neq P(D)$. Thus we can conclude they are dependent.