Back to top

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.

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)

$P(A, B) = P(A \cap B) = 0.1$

	A	A'	Total
B	**.1**	.5	.6 (given)
B'	.1	.3 (given)	.4
Total	.2 (given)	.8	1

2b)

$P(B|A') = \frac{P(B \cap A')}{P(A')} = \frac{.5}{.8} = 0.625$

	A	A'	Total
B	.1	**.5**	.6 (given)
B'	.1	.3 (given)	.4
Total	.2 (given)	**.8**	1

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)

Bayes' Rule

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

$P(A)$ = probability the person has the disease (the prior) = 0.02

$P(B)$ = probability of a positive (sum of the true positives and false negatives probabilities) $= 0.02 * 0.9 + .98 * .08$

$P(B|A)$ = the probability the test is a true positive (test sensitivity) = 0.9

$P(A|B) = \frac{0.9*0.02}{0.02*0.9+.98*.08}$

3b)

The prior is that 2% of the population has the virus, and the liklihood is the probability that they really do have the virus.

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 $P(n,k)$.

The total possible ways to choose 5 cards out of 52 is

$C(52,5) = \frac{52!}{5!(52-5)!} = \frac{52!}{5!*47!} = \frac{52*51*50*49*48}{2*3*4*5} = 2598960$

When we pull out the first rank, there are

13

different possibilities for what rank that can be chosen. Then there are only 12 different ranks left and we have to choose 2 of them for the other two ranks. Therefore the total number of ways to choose those ranks is

$C(12,2) = \frac{12!}{2!(12-2)!} = \frac{12!}{2!*10!} = \frac{12*11}{2} = 66$

Now we have to account for the suits of the cards. To find all the ways the suits for the three same ranks can be chosen, we calculate

$C(4,3) = \frac{4!}{3!(4-3)!} = \frac{4!}{3!} = 4$

And to find all the ways the suits for one of the other different ranks can be chosen, we calculate

$C(4,1) = \frac{4!}{1(4-1)!} = \frac{4!}{3!} = 4$

And the same can be done for the last different rank.

Thus, the probability of drawing 5 cards that are three of a kind is

$\frac{13*66*4*4*4}{2598960} = \frac{54912}{2598960} = 0.021128$

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.

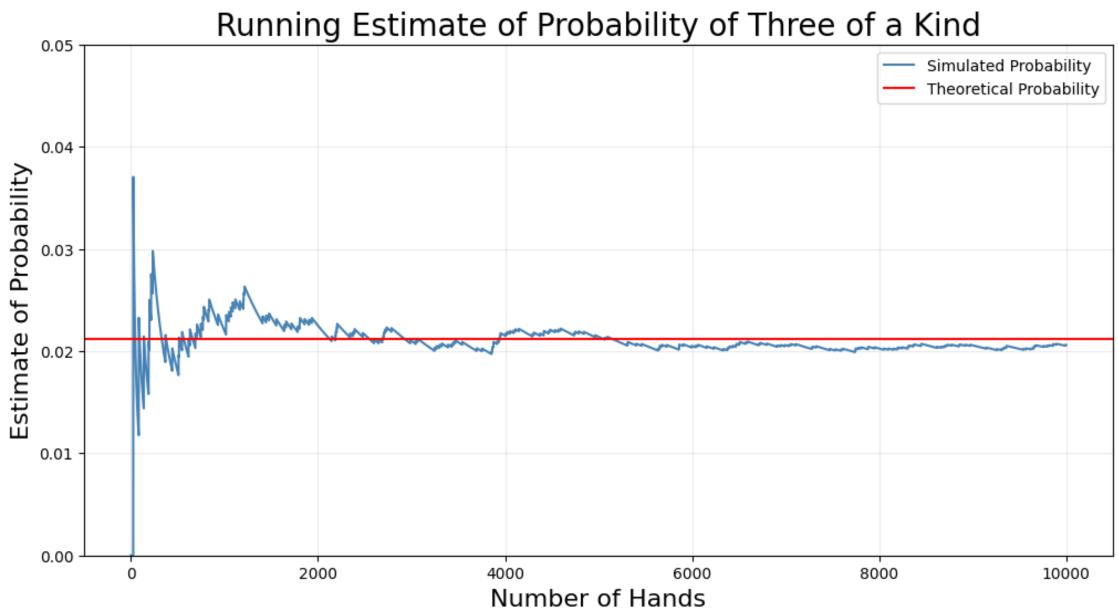
In [145]:
```
def draw_five():
    np.random.shuffle(cards)
    first_five = cards[:5]
    return three_kind(first_five)

def probability_three_kind(num_samples=10000):
    threes = np.array([draw_five() for ii in range(num_samples)])
    return threes.sum()/num_samples

probability_three_kind()

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

Out[145]: 0.0203

In [155]:
```
def plot_estimates(n):

# Keep a "running estimate" of the probability of getting a heads as num_trials gets larger:

    running_prob = []  #create an empty list - we will be appending probabilities to this lis
    numerator = 0;
    for ii in range(n):
        if draw_five():
            numerator+=1
        denom=ii+1
        running_prob.append(numerator/denom)
         # A growing sequence with ratio of 'H' being counted per times coin flipped.
        # Suppose we got T, H, T, ...
        # running_prob[0]=0/1   running_prob[1]=1/2   running_prob[2]=1/3   etc
        # running_prob = [0, 0.5, 0.3333..., etc]
```

```
    return running_prob


# Run code for num trials
num_trials=10000

p=plot_estimates(num_trials)

print("the probability of threes is approximately {:.3f}".format(p[num_trials-1]))



# Plot running estimate of probability of getting heads as num_trials gets larger:
fig, ax = plt.subplots(figsize=(12,6))

# plot the terms in p
ax.plot(p, color='steelblue', label='Simulated Probability')

#Plot the theoretical probability
plt.axhline(y=0.021128, color = 'r', label='Theoretical Probability')

# put labels on the axes and give the graph a title.
ax.set_title("Running Estimate of Probability of Three of a Kind", fontsize=20)
ax.set_xlabel("Number of Hands", fontsize=16)
ax.set_ylabel("Estimate of Probability", fontsize=16)
# fix the y-axis to be between 0 and 1:
ax.set_ylim(0,.05)
# include a legend:
ax.legend()
# put a faded grid behind the graphic
ax.grid(True, alpha=0.25)

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

```
the probability of threes is approximately 0.021
```



Running Estimate of Probability of Three of a Kind
Simulated Probability
Theoretical Probability
Estimate of Probability
0.05
0.04
0.03
0.02
0.01
0.00
0
2000
4000
6000
8000
10000
Number of Hands

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).

To calculate the probability of rolling a 5, we must calculate

$P(R_5 \cap D_4) + P(R_5 \cap D_6) + P(R_5 \cap D_{20})$

Since $P(A \cap B) = P(A|B) * P(B)$ and the total amount of die is $28 + 9 + 3 = 40$

$P(R_5|D_4) = 0$

$P(R_5|D_6) = \frac{1}{6}$ and $P(D_6) = \frac{9}{40}$

$P(R_5|D_{20}) = \frac{1}{20}$ and $P(D_{20}) = \frac{3}{40}$

Then we can calculate

$0 + (\frac{1}{6})(\frac{9}{40}) + (\frac{1}{20})(\frac{3}{40}) = \frac{9}{240} + \frac{3}{800} = \frac{33}{800}$

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.

Since we want to calculate the probability of the die being six-sided given we rolled a 5, we will use Bayes' Rule

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

Our case is

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

From **Part 5A** we can plug in these values

$P(D_6|R_5) = \frac{(\frac{1}{6})(\frac{9}{40})}{\frac{33}{800}} = \frac{7200}{7920} = \frac{10}{11}$

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**.

To check if the events are independent, we will check if $P(A \cap B) = P(A)P(B|A) = P(A)P(B)$

Our case is $P(R_5 \cap D_6)$

From **5A** and **5B** we can plug in the values.

$P(R_5)P(D_6|R_5) = (\frac{33}{800})(\frac{10}{11}) = \frac{330}{8800}$

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

Since $P(A)P(B|A) \neq P(A)P(B)$ these events are dependent.

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.

The total different combinations is $6 * 6 = 36$

Then I wrote out all the different combinations of being close:

C = {(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)}

Since $|C| = 16$

$P(C) = \frac{|C|}{Total} = \frac{16}{36}$

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 [164]:
```
def roll_die():
    die1= [1,2,3,4,5,6]
    die2= [1,2,3,4,5,6]

    roll1=np.random.choice(die1,1)
    roll2=np.random.choice(die2,1)

    rolls = []
    rolls.append(roll1)
    rolls.append(roll2)

    return rolls

def check_close(die):
    die1=die[0]
    die2=die[1]

    dif = abs(die1-die2)

    if dif == 1 or dif == 0:
        return True

    return False

def probability_close(num_samples=10000):
    close=np.array([check_close(roll_die()) for ii in range(num_samples)])
    return close.sum()/num_samples

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

Out[164]: 0.4405

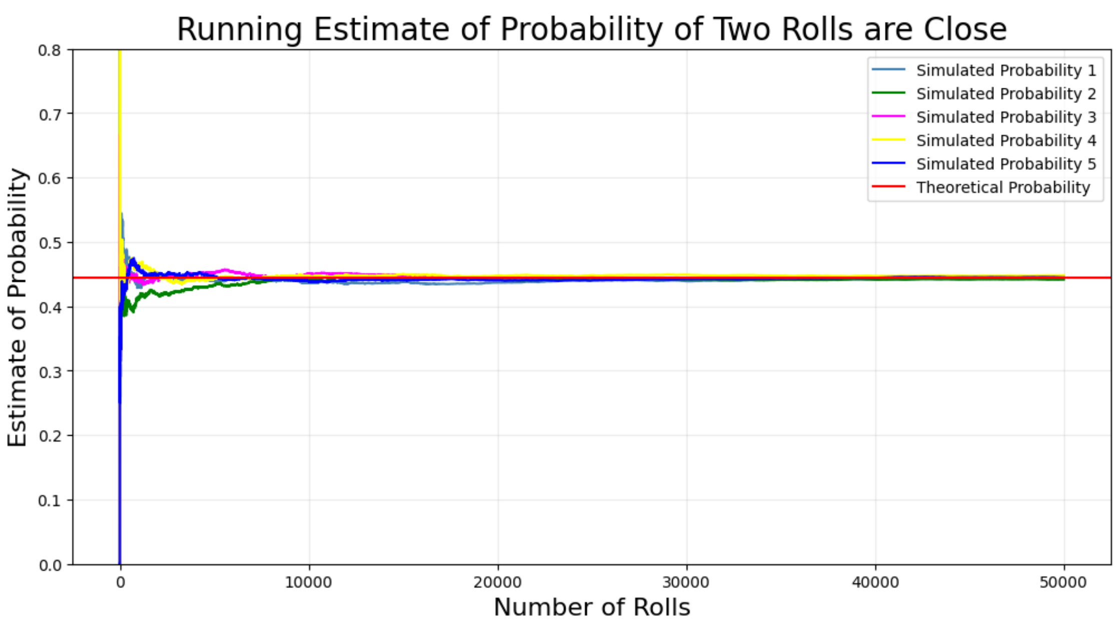
Part 6C (3 pts): In class we plotted a running estimate of the probability of an event as a function of the number of trials in our simulation. Write code to run 5 independent simulations of 50,000 trials each to estimate $P(C)$ and plot their running estimate curves on the same set of axes. **Hint**: 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 [165]:
```
def plot_estimates(n):

# Keep a "running estimate" of the probability of getting a heads as num_trials gets larger:

    running_prob = []  #create an empty list - we will be appending probabilities to this lis
    numerator = 0;
    for ii in range(n):
        die = roll_die()
        if check_close(die):
            numerator+=1
        denom=ii+1
        running_prob.append(numerator/denom)
         # A growing sequence with ratio of 'H' being counted per times coin flipped.
        # Suppose we got T, H, T, ...
        # running_prob[0]=0/1   running_prob[1]=1/2   running_prob[2]=1/3   etc
        # running_prob = [0, 0.5, 0.3333..., etc]

    return running_prob



# Run code for num trials
num_trials=50000

p1=plot_estimates(num_trials)
p2=plot_estimates(num_trials)
p3=plot_estimates(num_trials)
p4=plot_estimates(num_trials)
p5=plot_estimates(num_trials)

print("the probability of threes is approximately {:.3f}".format(p1[num_trials-1]))
print("the probability of threes is approximately {:.3f}".format(p2[num_trials-1]))
print("the probability of threes is approximately {:.3f}".format(p3[num_trials-1]))
print("the probability of threes is approximately {:.3f}".format(p4[num_trials-1]))
print("the probability of threes is approximately {:.3f}".format(p5[num_trials-1]))



# Plot running estimate of probability of getting heads as num_trials gets larger:
fig, ax = plt.subplots(figsize=(12,6))

# plot the terms in p
ax.plot(p1, color='steelblue', label='Simulated Probability 1')
ax.plot(p2, color='green', label='Simulated Probability 2')
```

```python
ax.plot(p3, color='magenta', label='Simulated Probability 3')
ax.plot(p4, color='yellow', label='Simulated Probability 4')
ax.plot(p5, color='blue', label='Simulated Probability 5')

#Plot the theoretical probability
plt.axhline(y=0.444444, color = 'r', label='Theoretical Probability')

# put labels on the axes and give the graph a title.
ax.set_title("Running Estimate of Probability of Two Rolls are Close", fontsize=20)
ax.set_xlabel("Number of Rolls", fontsize=16)
ax.set_ylabel("Estimate of Probability", fontsize=16)
# fix the y-axis to be between 0 and 1:
ax.set_ylim(0,.8)
# include a legend:
ax.legend()
# put a faded grid behind the graphic
ax.grid(True, alpha=0.25)

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

```
the probability of threes is approximately 0.443
the probability of threes is approximately 0.442
the probability of threes is approximately 0.445
the probability of threes is approximately 0.448
the probability of threes is approximately 0.445
```



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?

They are converging to the theoretical probability of 0.444, and it appears to take around 10000 trials to converge.

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.

The total outcomes for rankings is R = {XYZ, XZY, YXZ, YZX, ZXY, ZYX} |R| = 6

Therefore the set of

Event A = {XYZ, XZY, ZXY},|A|=3

Event B = {XYZ, XZY},|B|=2

Event C = {YXZ, ZXY},|C|=2

Event D = {YZX, ZYX},|D|=2

7a)

$P(A)P(B|A) = (\frac{1}{2})(\frac{2}{3}) = \frac{1}{3}$

$P(A)P(B) = (\frac{1}{2})(\frac{1}{3}) = \frac{1}{6}$

Since $P(A)P(B|A) \neq P(A)P(B)$ these events are dependent.

7b)

$P(A)P(C|A) = (\frac{1}{2})(\frac{1}{3}) = \frac{1}{6}$

$P(A)P(C) = (\frac{1}{2})(\frac{1}{3}) = \frac{1}{6}$

Since $P(A)P(B|A) = P(A)P(B)$ these events are independent.

7c)

$P(A)P(D|A) = (\frac{1}{2})(0) = 0$

$P(A)P(D) = (\frac{1}{2})(\frac{1}{3}) = \frac{1}{6}$

Since $P(A)P(B|A) \neq P(A)P(B)$ these events are dependent.