YZV231E - Prob. Stat. for Data Science

Nazrin Abdinli, 150220925 29/11/2023

Question 1

```python
import numpy as np
import random
import matplotlib.pyplot as plt
from scipy.stats import binom, uniform, bernoulli, randint, geom
%matplotlib inline


#Q1 - PART1 SOLUTION
def calculate_expectation(numbers, probs):
    """
    Calculates the expectation with given number and probabilities
    Returns the expectation
    Args:
        numbers: 1D list # values
        probs: 1D list # probabilities
    Return:
        result: float
    """
    result = None
    #calculating the expectation using the formula (x * p) for each value and its probability
    result = sum(x * p for x, p in zip(numbers, probs))
    return result

def calculate_variance(numbers, probs):
    """
    Calculates the variance with given number and probabilities
    Returns the variance
    Args:
        numbers: 1D list # values
        probs: 1D list # probabilities
    Return:
        result: float
    """
```

```
    result = None

    #calculating the expected value using the previously defined function
    expected_value = calculate_expectation(numbers, probs)

    #calculating the variance using the formula ((x - E[x])^2 * p) for each value, its probability, and the expected value
    result = sum((x - expected_value) ** 2 * p for x, p in zip(numbers, probs))
    return result


die = [i for i in range(1, 7)]      #list of the numbers on the faces of the die
die_probs = [1/6 for i in range(6)]      #list of the probabilities for each number in the die

#calculating the expected value and variance for the die using the defined functions
die_expexted_value = calculate_expectation(die, die_probs)
die_variance_value = calculate_variance(die, die_probs)
print(f"Expected value for die: {die_expexted_value:.3f}")
print(f"Variance for die: {die_variance_value:.3f}")
```

Figure 1: Expected Value and Variance for Die

```python
#Q1 - PART2 SOLUTION
def roll_die(n):
    """
    roll die n times
    Returns records of experiments
    Args:
        n: int # number of experiment
    Return:
        array: np.array with size of n
    """

    array = []
    #generating 'n' random numbers within the range of a six-sided die (1 to 6) and record each roll
    array = [random.randint(1, 6) for N in range(n)]
    return np.array(array)

ns = [1,5,10,50,100,500,1000,5000,10000, 50000, 100000] # number of rolls

#performing die rolls for different specified numbers in 'ns'
for n in ns:
    rolls = roll_die(n) # rolling the dice for n times
    # calculate the mean of the rolls
    mean = rolls.mean()
    print(f"Average of {n} die rolls: {mean:.3f}")
```

```
Average of 1 die rolls: 4.000
Average of 5 die rolls: 4.400
Average of 10 die rolls: 3.100
Average of 50 die rolls: 4.280
Average of 100 die rolls: 3.640
Average of 500 die rolls: 3.468
Average of 1000 die rolls: 3.537
Average of 5000 die rolls: 3.498
Average of 10000 die rolls: 3.490
Average of 50000 die rolls: 3.488
Average of 100000 die rolls: 3.498
```

Figure 2: Average for Different Number of Die Rolls

Question 2

```
#Q2 - PART1 SOLUTION
n = 100 # number of trials
probs = [0.2, 0.5, 0.7] # probabilities for each experiments' successful
x = np.arange(0,101,1) # range
plt.figure(figsize=(8,4))
print(x)

#calculation of Probability Mass Function
for prob in probs:
    PMF = binom.pmf(x, n, prob)
    plt.plot(x, PMF, marker='o', linestyle='-', markersize=5)

plt.legend(["$p=0.2$","$p=0.5$","$p=0.7$"],fontsize=15)
plt.title("Probability Mass Function: $\\binom{n}{k}\, p^k (1-p)^{n-k}$\n",fontsize=20)
plt.xlabel("Number of successful trials ($k$)",fontsize=15)
plt.ylabel("$P[k;n,p]$",fontsize=15)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.show()
```

```
[  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53
  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71
  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89
  90  91  92  93  94  95  96  97  98  99 100]
```

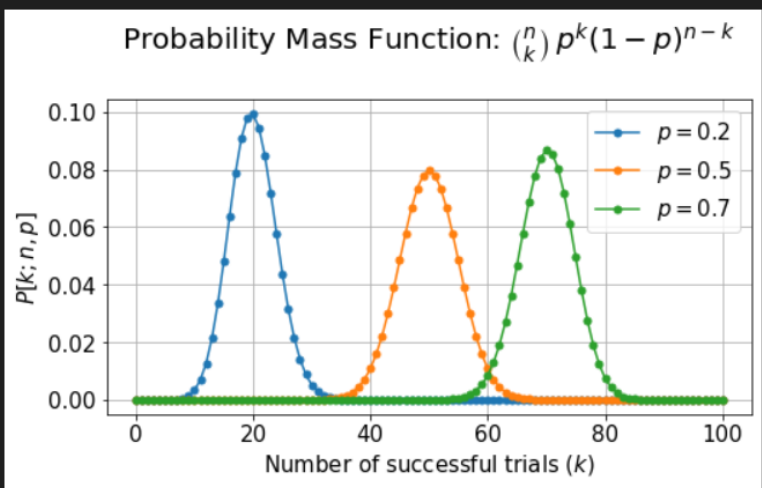

Figure 3: Visualized PMF for Binomial Experiment

```
#Q2 - PART2 SOLUTION
n = 100 # number of trials
probs = [0.2, 0.5, 0.7] # probabilities for each experiments' successful
x = np.arange(0,101,1) # range
plt.figure(figsize=(8,4))

#calculation of Cumulative Distribution Function
for prob in probs:
    CDF = binom.cdf(x, n, prob)
    plt.plot(x, CDF, marker='o', linestyle='-', markersize=5)

plt.legend(["$p=0.2$","$p=0.5$","$p=0.7$"],fontsize=15)
plt.title("Cumulative Distribution Function:",fontsize=20)
plt.xlabel("Number of successful trials ($k$)",fontsize=15)
plt.ylabel("$F[k;n,p]$",fontsize=15)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.show()
```

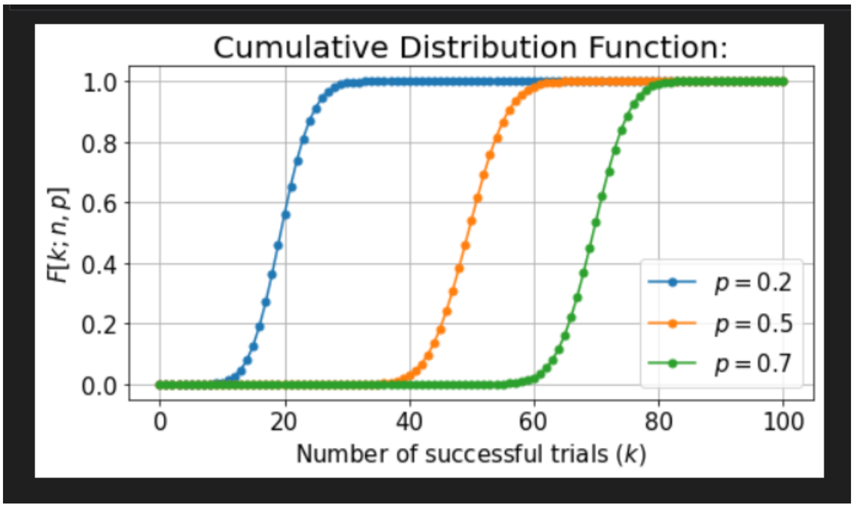

Figure 4: Visualized CDF for Binomial Experiment

Question 3

```
#Q3 - PART1 SOLUTION
probs = [0.2, 0.5, 0.8] # probabilities for each experiments' successful
x = np.arange(0,26,1) # range
plt.figure(figsize=(8,4))
print(x)

#calculation of Probability Mass Function for the geometric distribution
for prob in probs:
    PMF = (1 - prob) ** (x - 1) * prob
    plt.plot(x, PMF, marker='o')

plt.legend(["$p=0.2$","$p=0.5$","$p=0.8$"],fontsize=15)
plt.title("Probability Mass Function: $ (1-p)^{k-1} p$\n",fontsize=20)
plt.xlabel("Success at $k$'th trial ($k$)",fontsize=15)
plt.ylabel("$P[k;p]$",fontsize=15)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.show()
```

```
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
 24 25]
```

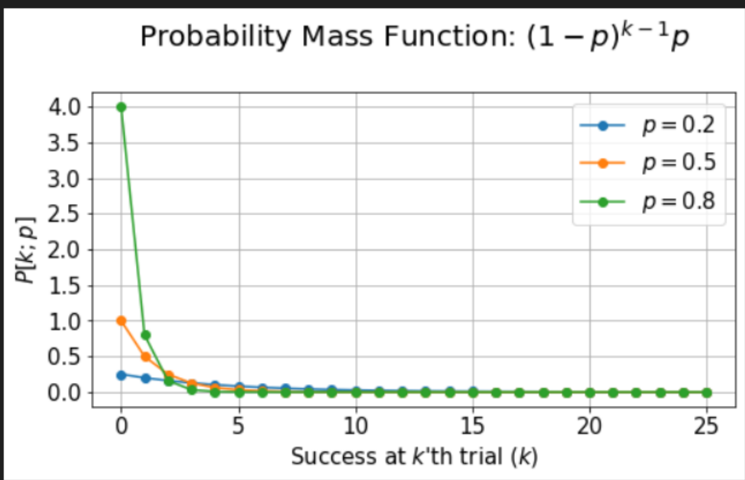

Figure 5: Visualized PMF for Geometric Experiment

```python
#Q3 - PART2 SOLUTION
probs = [0.2, 0.5, 0.8] # probabilities for each experiments' successful
x = np.arange(0,26,1) # range
plt.figure(figsize=(8,4))

#calculation of Cumulative Distribution Function for the geometric distribution
for prob in probs:
    CDF = 1 - (1 - prob) ** x
    plt.plot(x, CDF, marker='o')

plt.legend(["$p=0.2$","$p=0.5$","$p=0.8$"],fontsize=15)
plt.title("Cumulative Distribution Function:",fontsize=20)
plt.xlabel("Success at $k$'th trial ($k$)",fontsize=15)
plt.ylabel("$F[k;p]$",fontsize=15)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.grid(True)
plt.show()
```

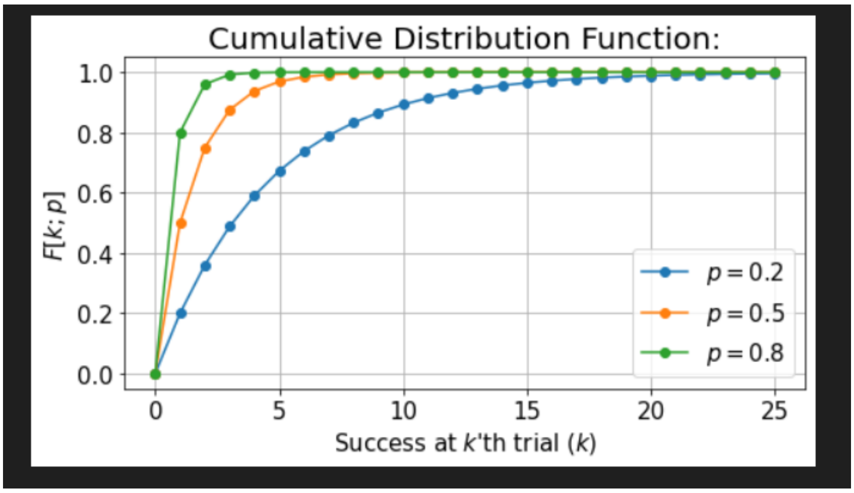

Figure 6: Visualized CDF for Geometric Experiment

Question 4

```python
#Q4 - PART1 SOLUTION
def game_simulation_without_strategy():
    prize_box = random.randint(1, 100)    #selecting a box randomly with a hidden prize (from 1 to 100)

    #simulating 50 attempts by the guest to select a box
    for _ in range(50):
        selected_box = random.randint(1, 100)    #guest selects a box randomly
        if selected_box == prize_box:    #checking if the selected box matches the box with the prize
            return True
    return False

def calculate_prob_without_strategy(trials):
    wins = 0
    #simulating the game for the specified number of trials
    for _ in range(trials):
        if game_simulation_without_strategy():    #checking if the guest wins in each trial
            wins += 1
    prob = wins / trials    #calculating the probability of winning based on the number of wins and total trials
    return prob

# Running 10000 trials for Q3a
trials = 10000
prob_without_strategy = calculate_prob_without_strategy(trials)    #calculating the probability of guests winning the game without any strategy
print("Probability of guests winning the game without strategy:", prob_without_strategy)
```

```
Probability of guests winning the game without strategy: 0.3917
```

Figure 7: Probability of Guests Winning without Strategy

```python
#Q4 - PART2 SOLUTION
def game_simulation_with_strategy(strategy):
    prize_box = random.randint(1, 100)    #selecting a box randomly with a hidden prize (from 1 to 100)

    #iterating through the boxes based on the provided strategy
    for selected_box in strategy:
        if selected_box == prize_box:    #checking if the selected box matches the box with the prize
            return True
    return False

def calculate_prob_with_strategy(trials):
    wins = 0
    for _ in range(trials):
        strategy = random.sample(range(1, 101), 50)    #generating a strategy by randomly selecting 50 unique boxes out of 100
        if game_simulation_with_strategy(strategy):    #checking if the guest wins using the generated strategy
            wins += 1
    prob = wins / trials    #calculate the probability of winning based on the number of wins and total trials
    return prob

# Running 10000 trials for Q3b
prob_with_strategy = calculate_prob_with_strategy(trials)    #calculating the probability of guests winning the game with strategy
print("Probability of guests winning the game with the strategy:", prob_with_strategy)
```

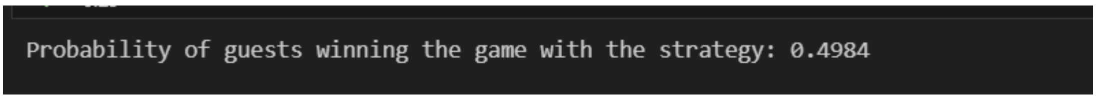

Figure 8: Probability of Guests Winning with Strategy

1 Problem 1 Handwriting

1.1 Subsection A

for $-2 \leqslant x \leqslant 0:$

$$E[X] = \int_{-2}^{0} x\left(\frac{-x}{4}\right)dx = \left[\frac{-x^3}{12}\right]_{-2}^{0} = \frac{-0^3}{12} - \left(-\frac{(-2)^3}{12}\right) = 0 - \left(\frac{8}{12}\right) = -\frac{2}{3}$$

for $0 \leqslant x \leqslant 1:$

$$E[X] = \int_{0}^{1} x x dx = \left[\frac{x^3}{3}\right]_{0}^{1} = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}$$

$$E[X] = (-\frac{2}{3}) + (\frac{1}{3}) = -\frac{1}{3}$$

1.2 Subsection B

for $-2 \leqslant x \leqslant 0:$

$$F_X(x) = \int_{-2}^{x} -\left(\frac{t}{4}dt\right) = \left[\frac{-t^2}{8}\right]_{-2}^{x} = \left(\frac{-x^2}{8}\right) - \left(\frac{-4}{8}\right) = \frac{-x^2}{8} + \frac{1}{2}$$

for $0 \leqslant x \leqslant 1:$

$$F_X(x) = \int_{0}^{x} t dt = \left[\frac{t^2}{2}\right]_{0}^{x} = \frac{x^2}{2} - 0 = \frac{x^2}{2}$$

$$F_X(x) = \begin{cases} -\frac{x^2}{8} + \frac{1}{2}, \text{ if } -2 \leq x \leq 0 \\ \frac{x^2}{2}, \text{ if } 0 \leq x \leq 1 \\ 0, \text{ if } x < -2 \text{ or } x > 1 \end{cases}$$

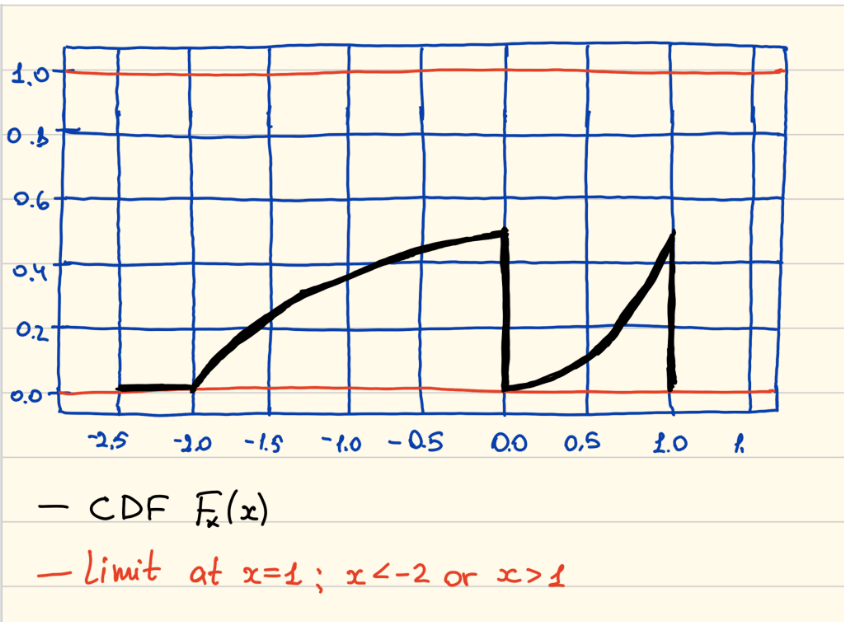

Figure 9: CDF of X

1.3 Subsection C

$P(X > 0)$: For finding $P(X > 0)$, we can use cdf of the random variable X and then find the complement of the event $X \leq 0$

$$P(X > 0) = 1 - P(X \leq 0)$$

For $X \leq 0$:

$$F_X(0) = -\frac{0^2}{8} + \frac{1}{2} = \frac{1}{2} \rightarrow \text{ so, } P(X > 0) = 1 - \frac{1}{2} = \frac{1}{2}$$

In this case, the probability that the random variable x is greater than 0 is $\frac{1}{2}$

1.4 Subsection D

$$f_{X|A}(x \mid A) = \frac{f_X(x)}{P(X > 0)}$$

We need to compute the conditional pdf $f_{X|A}(x \mid A)$ for $X > 0$. For $0 \leqslant X \leqslant 1$, the pdf of X given event $A(X > 0)$ will be :

$$f_{X|A}(x|A) = \frac{f_X(x)}{P(X > 0)} = \frac{x}{P(A)} = \frac{x}{\frac{1}{2}} = 2x$$

In this case, the conditional pdf $f_{X|A}(x \mid A)$ for $X > 0$ is $2x$ within the interval $0 \leq x \leq 1$.

2 Problem 2 Handwriting

2.1 Subsection A

$$f_P(p) = \begin{cases} \frac{1}{0.7-0.3} = \frac{1}{0.4}, \text{ for } 0.3 \leqslant p \leqslant 0.7 \\ 0, \text{ otherwise} \end{cases}$$

$$P(H) = \int_{0.3}^{0.7} p f_P(p) dp = \int_{0.3}^{0.7} \frac{1}{0.4} p dp = \left.\frac{p^2}{0.8}\right|_{0.3}^{0.7} =$$
$$= \frac{0.49}{0.8} - \frac{0.09}{0.8} = \frac{0.4}{0.8} = \frac{1}{2}$$

answer is $\frac{1}{2}$

2.2 Subsection B

Bayes' theorem states:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

given that $P(H) = 0.5$, because the coin is fair $P(E \mid H)$: the probability of observing 7 heads in 10 lasses given that the coin has a bias $X = a$ for heads

$$P(E \mid H) = \begin{pmatrix} 10 \\ 7 \end{pmatrix} \cdot a^7 \cdot (1 - a)^3$$

$P(E)$: the marginal probability of observing 7 heads in to tosses and can be found by integrating all possible values of X within the interval $(0.3, 0.7)$

$$P(E) = \int_{0.3}^{0.7} P(E \mid H = a) \cdot f(a) da$$

Where $f(a)$ is the probability density function (PDF) of the uniform distribution within the interval $(0.3, 0.7)$.

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} = \frac{\begin{pmatrix} 10 \\ 7 \end{pmatrix} \cdot a^7 \cdot (1-a)^3 \cdot \frac{1}{2}}{\int_{a.3}^{0.7} P(E \mid H = a) \cdot f(a) da}$$

3 Problem 3 Handwriting

3.1 Subsection A

Each guest selects a box randomly out of 100 boxes. The probability of a guest not finding the prize in a box is $99/100$, as there is only 1 box with the prize out of 100 boxes

The probability that none of the guests find the prize is $(99/100)$ raised to the power of 50 (prob. of not finding the price for each guest multiplied together for all 50 guests)

Prob. of the guests winning the game is the complement of the prob. that they lose. So:
prob. of guests winning $= 1$ - prob. of guests losing
prob. of guests winning $= 1 - \left(\frac{99}{100}\right)^{50} \approx 0.395$
answer is 0.395 .

3.2 Subsection B

Each guest selects a unique number from 1 to 100. That's why the probability that all guests miss the prize in a single try is:

$$\frac{90}{100} \cdot \frac{98}{99} \cdot \frac{97}{98} \cdot \ldots \cdot \frac{50}{51} = \frac{50}{100} = \frac{1}{2}$$

$1 - \frac{1}{2} = \frac{1}{2} \rightarrow$ the probability that guests win the game with the certain strategy

3.3 Subsection C

In the first strategy, if first player finds the prize, he/she takes it and the rest of the players gets nothing. If the lost player finds the prize, he/she takes it and previous players don't get anything. If any player except the first and the last one finds the prize, both the first and the last players get nothing. So the expected value is 1 for the player who finds the prize and is 0 for the rest of the players who doesn't find the prize.

prob. of finding the prize for a tingle player: $\rightarrow \frac{1}{100}$
value of finding the prise for a single player: $\longrightarrow 100$
Expected value for a single player: $\longrightarrow \frac{1}{100} \cdot 100 = 1$

In the second strategy, it doesn't matter who finds the price, all players have the same expectation value.
prob. of finding the prize for a single player: $\rightarrow \frac{1}{100}$
value of finding the prize: $\longrightarrow \frac{100}{50} = 2$
Expected value for a single player: $\frac{1}{100} \cdot 2 = 0.02$

4 Problem 4 Handwriting

4.1 Subsection A

The expected number of accidents for a portion distribution is equal to its parameter λ.
for highway $1: \lambda_1 = 0.3$
for highway 2: $\lambda_2 = 0.5$
for highway 3: $\lambda_3 = 0.7$
Expected number of accidents that will happen on any of there high rays can be found by adding the expected number of accidents for each highway:

$$\text{Expected number of daily accidents on any highway} = \lambda_1 + \lambda_2 + \lambda_3 = 0.3 + 0.5 + 0.7 = 1.5$$

So the expected number of daily accidents on any of these highways is 1.5

4.2 Subsection B

The probability of no accidents occuring on a particular road that follows a Poison distribution with parameter λ is given by the PMF of the Poisson distribution:

$$P(X = 0) = \frac{e^{-\lambda}\lambda^{e}}{0!} = e^{-\lambda}$$

for highway 1: P (no accidents) $= e^{-0.3}$
for highway 2:P(no accidents) $= e^{-0.5}$
for highway $3 : P$ (no accidents) $= e^{-0.7}$

$$P(\text{ no accident on any road }) = e^{-0.3} \cdot e^{-0.5} \cdot e^{-0.7} = e^{-1,5} \approx 0.223$$

$$P\left(\begin{array}{l}\text{at least one accident)}\\ \text{on any road}\end{array}\right) = 1 - P(\text{ no accident on any road })$$

$$P\left(\begin{array}{l}\text{at bast one accident)}\\ \text{on any road}\end{array}\right) = 1 - 0.223 = 0.777$$

answer is 0.777

5 Problem 5 Handwriting

5.1 Subsection A

The area under the curve is:

$$A_1 = 2 \cdot a - 2a$$
$$A_2 = \frac{a \cdot 2}{2} = a$$

The total area under the curve should be equal to 1. $\rightarrow$ in this case: $3a = 1$

$$a = \frac{1}{3}$$

answer is $a = \frac{1}{3}$

5.2 Subsection B

$$f_X(x) = \frac{ax}{2} + a$$
$$E[X] = \int_0^2 x f_X(x) dx = \int_0^2 x \cdot \left(\frac{a}{2}x + a\right) dx =$$
$$= \int_0^2 \left(\frac{ax^2}{2} + ax\right) dx = \left[\frac{ax^3}{6} + \frac{ax^2}{2}\right]_0^2 =$$
$$= \left(\frac{8a}{6} + \frac{4a}{2}\right) - 0 = \frac{4a}{3} + \frac{4a}{2} = \frac{8a + 12a}{6} = \frac{20a}{6} =$$
$$= \frac{10}{3}a$$

$E[X] = \frac{10}{3}a \Rightarrow$ we round $a = \frac{1}{3}$ in a).

$$\frac{10}{3} + \frac{1}{3} = \frac{10}{9} \quad E[x] = \frac{10}{9}$$

answer is $E[X] = \frac{10}{9}$

5.3 Subsection C

$$f_X(x) = \frac{a}{2}x + a$$

$$\begin{aligned}
E\left[X^2\right] &= \int_0^2 x^2 \cdot f_X(x)dx \\
E\left[X^2\right] &= \int_0^2 x^2 \cdot \left(\frac{a}{2}x + a\right)dx = \int_0^2 \left(\frac{ax^3}{2} + ax^2\right)dx = \\
&= \left[\frac{ax^4}{8} + \frac{ax^3}{3}\right]_0^2 = \left(\frac{16a}{8} + \frac{8a}{3}\right) - 0 = 2a + \frac{8a}{3} = \\
&= \frac{6a}{3} + \frac{8a}{3} = \frac{14a}{3} \to a = \frac{1}{3} \to \frac{14}{3} \cdot \frac{1}{3} = \frac{14}{9}
\end{aligned}$$

we found $E[X] = \frac{10}{9}$

$$\begin{aligned}
\sigma_x^2 &= E\left[X^2\right] - (E[X])^2 = \frac{14}{9} - \frac{100}{81} = \frac{126}{81} - \frac{100}{81} = \frac{26}{81} \\
\sigma_x^2 &= \frac{26}{81}
\end{aligned}$$

5.4 Subsection D

$f_{X|A}(x \mid A) = \frac{f_X(x)}{P(A)}$

The prob. of the event $A(i.e., X > 1)$ is the area under the curve of the PDF from $x = 1$ to $x = 2$

$$\left.\begin{array}{l}\text{the area under}\\ \text{the curve of the}\\ \text{PDF from } x=1\\ \text{to } x=2\end{array}\right\} \frac{\frac{3a}{2}+2a}{2} \cdot 1 = \frac{7a}{2} \cdot \frac{1}{2} = \frac{7a}{4}$$

We found $a = \frac{1}{3}$ from a).
so, $\frac{7a}{4} \Rightarrow \frac{7}{4} \cdot \frac{1}{3} = \frac{7}{12} \to$ Therefore, $P(A) = \frac{7}{12}$

$$\begin{aligned}
f_X(x) &= \frac{ax}{2} + a \Longleftrightarrow a = \frac{1}{3} \\
f_X(x) &= \frac{x}{6} + \frac{1}{3} \\
\int_1^2 \left(\frac{x}{6} + \frac{1}{3}\right)dx &= \left[\frac{x^2}{12} + \frac{1}{3}x\right]_1^2 = \left(\frac{4}{12} + \frac{2}{3}\right) - \left(\frac{1}{12} + \frac{1}{3}\right) = 1 - \frac{5}{12} = \frac{7}{12} \\
f_{X|A}(x|A) &= \frac{f_X(x)}{P(A)} \to \frac{\frac{7}{12}}{\frac{7}{12}} = 1
\end{aligned}$$

answer is 1.

6 Problem 6 Handwriting

6.1 Subsection A

The curve $y = \ln(x+1)$ intersects the x-axis when $y = 0$

$$0 = \ln(x+1) \to x+1 = e^0 \to x+1 = 1 \to x = 0$$

it intersects the line $x = 1$ when:

$$y = \ln(1+1) = \ln 2$$

Is this case, the limits of integration for x are from 0 to 1 and for y are from 0 to $\ln 2$

$$\int_0^1 \int_0^{\ln 2} cxe^y dydx = 1$$
$$c\int_0^1 \left[x\int_0^{\ln 2} e^y dy\right] dx = 1 \to c\int \left[x\left(e^{\ln 2} - e^0\right)\right] dx = 1$$
$$c\int_0^1 (x(2-1))dx = 1 \to c\int_0^1 xdx = 1$$
$$c\left[\frac{x^2}{2}\right]_0^1 = 1 \to c \cdot \frac{1}{2} = 1 \quad c = 2$$

answer: $c = 2$

6.2 Subsection B

$$f_X(x) = \int_0^{\ln 2} cxe^y dy$$
$$f_X(x) = cx\int_0^{4x} e^y dy - f_X(x) = cx\left[e^y\right]_0^{\ln 2}$$
$$f_X(x) = cx\left(e^{\ln 2} - e^0\right) \to f_x(x) = cx$$
$$f_X(x) = 2x$$

$$f_Y(y) = \mathrm{c}e^y\int_0^1 xdx \to f_Y(y) = \mathrm{c}e^y\left[\frac{x^2}{2}\right]_0^1$$
$$f_Y(y) = ce^y \cdot \frac{1}{2} \to f_Y(y) = \frac{c}{2}e^y$$

From a), c=2

$$f_Y(y) = e^y$$

$f_X(x) = 2x$ for $0 \leqslant x \leqslant 1$
$f_Y(y) = e^y$ for $0 \leqslant y \leqslant \ln 2$

6.3 Subsection C

$$P(X > 0.5 \text{ and } Y < 0.5) = \int_0^{0.5} \int_{0.5}^{1} 2xe^y dxdy$$

$$\int_0^{0.5} e^y \left[x^2\right]_{0.5}^{1} dy = \int_0^{0.5} \left(e^y - 0.25e^y\right) dy =$$

$$= \int_0^{0.5} \left(0.75e^y\right) dy = 0.75 \int_0^{0.5} e^y dy = 0.75 \left[e^y\right]_0^{0.5} =$$

$$= 0.75 \left(e^{0.5} - e^0\right) = 0.75 \left(e^{0.5} - 1\right) \approx 0.75(1.6487 - 1) \approx 0.75.0.6487 \approx 0.4865$$

answer: $P(X > 0.5 \text{ and } Y < 0.5) \approx 0.4865$