Question 1

Questions below are related to the following table. Let $A, B, C \in \{0, 1\}$ be three binary random variables with the following joint probability distribution:

a	b	c	f(a,b,c)
0	0	0	0.01
0	0	1	0.07
0	1	0	0.02
0	1	1	0.10
1	0	0	0.02
1	0	1	0.30
1	1	0	0.04
1	1	1	0.44

By direct calculation, compute the marginal $P(A, B)$.

(Recall that $P(A, B)$ is represented by 4 numbers: $P(A = 0, B = 0)$, $P(A = 0, B = 1)$, $P(A = 1, B = 0)$, $P(A = 1, B = 1)$.)

The marginal pmf of two discrete random variables A and B is the sum of the joint pmf $f(A, B, C)$ over all possible values of C for each combination of A and B. This is given by:

$$f_{A,B}(a, b) = \sum_{\forall c} f(a, b, c)$$

Since the table provides the probability of each joint event of A, B, and C, we can directly calculate the marginal pmf of A and B by summing the joint pmf over all values of C.

Calculation of $P(A, B)$:

For $A = 0, B = 0$:

$$f_{A,B}(0, 0) = f(0, 0, 0) + f(0, 0, 1) = 0.01 + 0.07 = 0.08$$

For $A = 0, B = 1$:

$$f_{A,B}(0, 1) = f(0, 1, 0) + f(0, 1, 1) = 0.02 + 0.10 = 0.12$$

For $A = 1, B = 0$:

$$f_{A,B}(1,0) = f(1,0,0) + f(1,0,1) = 0.02 + 0.30 = 0.32$$

For $A = 1, B = 1$:

$$f_{A,B}(1,1) = f(1,1,0) + f(1,1,1) = 0.04 + 0.44 = 0.48$$

Conclusion:

Therefore, the marginal pmf of A and B is:

$$f_{A,B}(a,b) = \begin{cases} 0.08 & \text{if } a = 0, b = 0 \\ 0.12 & \text{if } a = 0, b = 1 \\ 0.32 & \text{if } a = 1, b = 0 \\ 0.48 & \text{if } a = 1, b = 1 \end{cases}$$

Question 2

> By direct calculation compute the marginals $P(A)$ and $P(B)$.

To find the marginal pmf of A, we sum the joint pmf $f(A, B, C)$ over all possible values of B and C for each value of A. Similarly, to find the marginal pmf of B, we sum the joint pmf $f(A, B, C)$ over all possible values of A and C for each value of B.

Calculation of $P(A)$:

For $A = 0$:

$$f_A(0) = f(0,0,0)+f(0,0,1)+f(0,1,0)+f(0,1,1) = 0.01+0.07+0.02+0.10 = 0.20$$

For $A = 1$:

$$f_A(1) = f(1,0,0)+f(1,0,1)+f(1,1,0)+f(1,1,1) = 0.02+0.30+0.04+0.44 = 0.80$$

Calculation of $P(B)$:

For $B = 0$:

$$f_B(0) = f(0,0,0)+f(1,0,0)+f(1,0,1)+f(0,0,1) = 0.01+0.02+0.30+0.07 = 0.40$$

For $B = 1$:

$$f_B(1) = f(0,1,0)+f(1,1,0)+f(1,1,1)+f(0,1,1) = 0.02+0.04+0.44+0.10 = 0.60$$

Conclusion:

Therefore, the marginal pmf of A and B is:

$$f_A(a) = \begin{cases} 0.20 & \text{if } a = 0 \\ 0.80 & \text{if } a = 1 \end{cases}$$

$$f_B(b) = \begin{cases} 0.40 & \text{if } b = 0 \\ 0.60 & \text{if } b = 1 \end{cases}$$

Question 3

Are the random variables A and B independent? Why or why not?

Two random variables A and B are independent if and only if their joint pmf is the product of their marginal pmfs for all possible values of A and B. In other words, A and B are independent if:

$$f_{A,B}(a,b) = f_A(a) \cdot f_B(b)$$

for all possible values of A and B. This definition implies that the occurrence of one event does not affect the occurrence of the other event.

To check if A and B are independent, we compare the joint pmf $f_{A,B}(a,b)$ with the product of the marginal pmfs $f_A(a) \cdot f_B(b)$ for all possible values of A and B:

For $A = 0, B = 0$:

$$f_{A,B}(0,0) = 0.08 = 0.20 \cdot 0.40 = 0.08$$

For $A = 0, B = 1$:

$$f_{A,B}(0,1) = 0.12 = 0.20 \cdot 0.60 = 0.12$$

For $A = 1, B = 0$:

$$f_{A,B}(1,0) = 0.32 = 0.80 \cdot 0.40 = 0.32$$

For $A = 1, B = 1$:

$$f_{A,B}(1,1) = 0.48 = 0.80 \cdot 0.60 = 0.48$$

Since for all possible values of A and B, $f_{A,B}(a,b) = f_A(a) \cdot f_B(b)$, the random variables A and B are **independent**.

Question 4

Compute the conditional distribution $P(A, B|C)$.

Note that this includes computing $P(A, B|C = 0)$ as well as $P(A, B|C = 1)$, each of which is represented by 4 numbers (in total 8 numbers).

A conditional probability for two random variables X and Y is given by:

$$P(X|Y) = \frac{P(X, Y)}{P(Y)}$$

where $P(X, Y)$ is the joint probability of X and Y, and $P(Y)$ is the marginal probability of Y.

Extending this to three random variables A, B, and C, the conditional probability $P(A, B|C)$ is given by:

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

where $P(A, B, C)$ is the joint probability of A, B, and C, and $P(C)$ is the marginal probability of C.

Calculate the marginal probability of C For $C = 0$:

$$f_C(0) = f(0, 0, 0)+f(1, 0, 0)+f(1, 1, 0)+f(0, 1, 0) = 0.01+0.02+0.04+0.02 = 0.09$$

Since C is a binary random variable, $P(C = 1) = 1 - P(C = 0) = 1 - 0.09 = 0.91$.

Calculate the conditional probability of $A, B|C$ For $A = 0, B = 0, C = 0$:

$$f(0, 0, 0) = 0.01 \implies f(0, 0|0) = \frac{f(0, 0, 0)}{f_C(0)} = \frac{0.01}{0.09} = 0.1111$$

For $A = 0, B = 0, C = 1$:

$$f(0,0,1) = 0.07 \implies f(0,0|1) = \frac{f(0,0,1)}{f_C(1)} = \frac{0.07}{0.91} = 0.0769$$

For $A = 0, B = 1, C = 0$:

$$f(0,1,0) = 0.02 \implies f(0,1|0) = \frac{f(0,1,0)}{f_C(0)} = \frac{0.02}{0.09} = 0.2222$$

For $A = 0, B = 1, C = 1$:

$$f(0,1,1) = 0.10 \implies f(0,1|1) = \frac{f(0,1,1)}{f_C(1)} = \frac{0.10}{0.91} = 0.1099$$

For $A = 1, B = 0, C = 0$:

$$f(1,0,0) = 0.02 \implies f(1,0|0) = \frac{f(1,0,0)}{f_C(0)} = \frac{0.02}{0.09} = 0.2222$$

For $A = 1, B = 0, C = 1$:

$$f(1,0,1) = 0.30 \implies f(1,0|1) = \frac{f(1,0,1)}{f_C(1)} = \frac{0.30}{0.91} = 0.3297$$

For $A = 1, B = 1, C = 0$:

$$f(1,1,0) = 0.04 \implies f(1,1|0) = \frac{f(1,1,0)}{f_C(0)} = \frac{0.04}{0.09} = 0.4444$$

For $A = 1, B = 1, C = 1$:

$$f(1,1,1) = 0.44 \implies f(1,1|1) = \frac{f(1,1,1)}{f_C(1)} = \frac{0.44}{0.91} = 0.4835$$

Conclusion:

Therefore, the conditional pmf of A and B given C is:

$$f_{A,B|C}(a,b|c) = \begin{cases} 0.1111 & \text{if } a=0, b=0, c=0 \\ 0.0769 & \text{if } a=0, b=0, c=1 \\ 0.2222 & \text{if } a=0, b=1, c=0 \\ 0.1099 & \text{if } a=0, b=1, c=1 \\ 0.2222 & \text{if } a=1, b=0, c=0 \\ 0.3297 & \text{if } a=1, b=0, c=1 \\ 0.4444 & \text{if } a=1, b=1, c=0 \\ 0.4835 & \text{if } a=1, b=1, c=1 \end{cases}$$

Question 5

Let

$$f(x,y) = \begin{cases} c(x+y^2) & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $P(X \le \frac{1}{2} | Y = \frac{1}{2})$.

For a continuous random variable, the probability of an event is given by the integral of the joint PDF over the region corresponding to the event. In this case, we need to find the probability that X is less than or equal to $\frac{1}{2}$ given that $Y = \frac{1}{2}$.

From the definition of conditional probability for continuous random variables, we have:

$$P(X \le \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{P(X \le \frac{1}{2}, Y = \frac{1}{2})}{P(Y = \frac{1}{2})}$$

where $P(X \le \frac{1}{2}, Y = \frac{1}{2})$ is the joint PDF of X and Y, and $P(Y = \frac{1}{2})$ is the marginal PDF of Y at $y = \frac{1}{2}$.

Calculate the marginal PDF of Y

The marginal PDF of Y is obtained by integrating the joint PDF over all possible values of X (in this case, from 0 to 1):

$$f_Y(y) = \int_0^1 c(x+y^2)dx = c\left[\frac{x^2}{2} + y^2 x\right]_0^1 = c\left(\frac{1}{2} + y^2\right)$$

For $Y = \frac{1}{2}$:

$$f_Y\left(\frac{1}{2}\right) = c\left(\frac{1}{2} + \left(\frac{1}{2}\right)^2\right) = c\left(\frac{1}{2} + \frac{1}{4}\right) = c\left(\frac{3}{4}\right)$$

Calculate $P(X \leq \frac{1}{2}, Y = \frac{1}{2})$

Next, calculate the joint probability $P(X \leq \frac{1}{2}, Y = \frac{1}{2})$ by integrating the joint PDF over the region defined by $0 \leq x \leq \frac{1}{2}$ and $y = \frac{1}{2}$:

$$P(X \leq \frac{1}{2}, Y = \frac{1}{2}) = \int_0^{\frac{1}{2}} c(x + \left(\frac{1}{2}\right)^2) dx = c \left[\frac{x^2}{2} + \frac{1}{4}x\right]_0^{\frac{1}{2}} = c\left(\frac{1}{8} + \frac{1}{8}\right) = c\left(\frac{1}{4}\right)$$

Calculate the conditional probability

Now we can calculate the conditional probability:

$$P(X \leq \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{P(X \leq \frac{1}{2}, Y = \frac{1}{2})}{P(Y = \frac{1}{2})} = \frac{c\left(\frac{1}{4}\right)}{c\left(\frac{3}{4}\right)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Conclusion:

The conditional probability is:

$$P(X \leq \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{1}{3}$$

Question 6

> Assume there are 100 boys and 150 girls in a classroom. We are going to select 70 students at random from the classroom without replacement. Let X denote the number of boys that are selected and let Y denote the number of girls that are selected. Find the expectation of X −Y .

From the properties of expectation, we know that the expectation of the difference of two random variables is the difference of their expectations. Therefore, we can find the expectation of $X - Y$ by finding the expectations of X and Y separately and then taking their difference.

Since we are sampling without replacement, we know to use the hypergeometric distribution given by:

$$P(X = x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$$

In this case:

- $A = 100$
- $B = 150$
- $n = 70$

Calculate the expectation of X

The expectation of a hypergeometric distribution is given by:

$$E(X) = \sum_{x=0}^{n} x \cdot P(X = x)$$

Substitute the values into the formula:

$$E(X) = \sum_{x=0}^{70} x \cdot \frac{\binom{100}{x}\binom{150}{70-x}}{\binom{250}{70}} = 28$$

Calculate the expectation of Y

Similarly, the expectation of Y is:

$$E(Y) = \sum_{y=0}^{70} y \cdot \frac{\binom{150}{y}\binom{100}{70-y}}{\binom{250}{70}} = 42$$

Calculate the expectation of $X - Y$

Finally, the expectation of $X - Y$ is the difference of the expectations of X and Y:

$$E(X - Y) = E(X) - E(Y) = 28 - 42 = -14$$

Question 7

Assume X, Y are two random variables that have a negative correlation. Determine the relationship between $Var(X+Y)$ and $Var(X-Y)$ (i.e. find out whether the former is larger or smaller than the latter).

Variance of $X+Y$ is given by:

$$Var(X+Y) = E[(X+Y)^2] - E[X+Y]^2 = Var(X) + Var(Y) + 2Cov(X,Y)$$

Variance of $X-Y$ is given by:

$$Var(X-Y) = E[(X-Y)^2] - E[X-Y]^2 = Var(X) + Var(Y) - 2Cov(X,Y)$$

Covariance is a measure of the relationship between two random variables. If two random variables have a negative correlation, it means that as one variable increases, the other variable decreases. It is given by:

$$Cov(X,Y) = E[XY] - E[X]E[Y] = \sigma_X \sigma_Y \rho_{X,Y}$$

The correlation of X and Y is given by:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Since the product of the sigmas cannot be negative, if the correlation is negative $\implies$ the covariance is negative:

$$Cov(X,Y) < 0$$

Since $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ and $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$, when $Cov(X,Y) < 0$, the variance of $X-Y$ will increase by $2 \cdot |Cov(X,Y)|$ whereas the variance of $X+Y$ reduces by $2 \cdot |Cov(X,Y)|$ compared to the variance of X and Y.

Thus, the relationship between $Var(X+Y)$ and $Var(X-Y)$ is:

$$\rho_{X,Y} < 0 \implies Cov(X,Y) < 0 \implies Var(X-Y) > Var(X+Y)$$

Question 8

Questions below are related to random variable $X \sim N(3, 16)$. You need to use scipy.stats. Paste your relevant code for each question separately.

Find $P(X > -2)

$X \sim N(3, 16)$ denotes that X has a normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 16$. To find $P(X > -2)$, we need to calculate the cumulative distribution function (CDF) of the normal distribution at -2.

```
from scipy.stats import norm

# Parameters of the normal distribution
mu = 3
sigma = 4

# Calculate the probability P(X > -2)
p_x_gt_minus_2 = 1 - norm.cdf(-2, loc=mu, scale=sigma)

p_x_gt_minus_2
# >>> 0.8943502263331446
```

Question 9

Find x such that $P(X > x) = 0.05$

To find the value of x such that $P(X > x) = 0.05$, we need to find the 95th percentile of the normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 16$.

```python
from scipy.stats import norm

# Parameters of the normal distribution
mu = 3
sigma = 4

# Calculate the value of x such that P(X > x) = 0.05
x = norm.ppf(0.95, loc=mu, scale=sigma)

x
# >>> 9.57941450780589
```

Question 10

Find $P(0 \leq X \leq 4)$

To find $P(0 \leq X \leq 4)$, we need to calculate the cumulative distribution function (CDF) of the normal distribution at 0 and 4 and then take the difference.

```
from scipy.stats import norm

# Parameters of the normal distribution
mu = 3
sigma = 4

# Calculate the probability P(0 <= X <= 4)
p_x_between_0_and_4 = norm.cdf(4, loc=mu, scale=sigma) - norm.cdf(0, loc=mu, scale=sigma)

p_x_between_0_and_4
# >>> 0.3720789733060555
```

Question 11

In continuous probability, we often need to solve messy integrals. For example, in this class we might need to use integrals to evaluate the probability of an event under a cumulative distribution function (CDF). Rather than solve this by hand, we can approximate it using discrete intervals. This problem will explore discrete approximation of integrals using a Gaussian model. Recall that the probability density function of a Gaussian random variable $X \sim N(\mu, \sigma 2)$ is,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In the questions below, we will use Python to form a discrete approximation of this continuous distribution, and evaluate associated probabilities.

Form a discrete approximation of the Normal PDF with mean $\mu = 70$ and standard deviation $\sigma = 2$. To do this, create an array x of evenly spaced values in the range $[68, 76]$ at increments of 2 excluding 76 (this array will include 68 and 74). The function `numpy.arange` might be helpful.

Create an array p containing values of the PDF at each location x. Plot the result as a bar chart (use `matplotlib.pyplot.bar`). In the same figure, overlay a PDF curve (use `matplotlib.pyplot.plot`) at more finely spaced intervals (e.g. 0.01). Paste your code.

```
import numpy as np
import matplotlib.pyplot as plt

# Parameters of the normal distribution
mu = 70
sigma = 2

# Create an array of evenly spaced values in the range [68, 76] at increments of 2 excluding
x = np.arange(68, 76, 2)

# Calculate the PDF values at each location x
p = 1 / np.sqrt(2 * np.pi * sigma**2) * np.exp(-0.5 * ((x - mu) / sigma)**2)

# Create a finer array of x values for plotting the PDF curve
```

```python
x_fine = np.linspace(68, 76, 1000)

# Calculate the PDF values at each fine location x
p_fine = 1 / np.sqrt(2 * np.pi * sigma**2) * np.exp(-0.5 * ((x_fine - mu) / sigma)**2)

# Plot the discrete approximation of the PDF as a bar chart
plt.bar(x, p, width=1.5, color='b', alpha=0.5, label='Discrete PDF')

# Plot the PDF curve
plt.plot(x_fine, p_fine, color='r', label='PDF Curve')

plt.xlabel('x')
plt.ylabel('f(x)')
plt.title('Discrete Approximation of Normal PDF')
plt.legend()
plt.show()
```

Question 12

The bar chart above is a discrete approximation of the continuous PDF. We will use it to approximate $P(68 < X \leq 76)$. Recall that $X \sim N(\mu, \sigma 2)$, so

$$ P(68 < X ≤76) = ∫_{68}^{76} f(x) dx

We will approximate this integral using a Riemann sum. Let N be the number of grid points in your array x. The spacing between grid points is Δx and let the ith point of array p be p_i. The Riemann sum approximation is,

$$\sum_{i=1}^{N} p_i \Delta x$$

Find the value of the approximation to $P(68 < X \leq 76)$. Paste your code.

```
import numpy as np

# Parameters of the normal distribution
mu = 70
sigma = 2

# Create an array of evenly spaced values in the range [68, 76] at increments of 2 excludin
x = np.arange(68, 76, 2)

# Calculate the PDF values at each location x
p = 1 / np.sqrt(2 * np.pi * sigma**2) * np.exp(-0.5 * ((x - mu) / sigma)**2)

# Calculate the spacing between grid points
delta_x = x[1] - x[0]

# Calculate the Riemann sum approximation of the integral
riemann_sum = np.sum(p * delta_x)

riemann_sum
# >>> 0.9368746959529074
```

Question 13

Now, reduce the spacing $\Delta x = 0.01$ and recompute the discrete approximation of $P(68 < X \leq 76)$. Paste your code and argue: How do the two approximations compare? What is the practical downside of smaller spacing?

```
import numpy as np

# Parameters of the normal distribution
mu = 70
sigma = 2

# Create an array of evenly spaced values in the range [68, 76] at increments of 0.01 exclu
x = np.arange(68, 76, 0.01)

# Calculate the PDF values at each location x
p = 1 / np.sqrt(2 * np.pi * sigma**2) * np.exp(-0.5 * ((x - mu) / sigma)**2)

# Calculate the spacing between grid points
delta_x = x[1] - x[0]

# Calculate the Riemann sum approximation of the integral
riemann_sum = np.sum(p * delta_x)

riemann_sum
# >>> 0.8405881634221108
```

The second approximation being about 11% lower than the first approximation — which seems like it would be meaningful in most contexts in which we are interested in the exact value of the probability of an event.

The practical downside of using smaller spacing ($\Delta x = 0.01$) is that it requires more computation and memory resources. The finer grid results in a larger number of grid points and a higher computational cost for calculating the Riemann sum.

Question 14

Repeat the steps above to show the distribution over the range $[20, 120]$ and compute $P(20 \leq X < 120)$. What is the value? This interval should contain almost all of the probability in this distribution, i.e. the event is almost certain. Paste your code.

```
import numpy as np

# Parameters of the normal distribution
mu = 70
sigma = 2

# Create an array of evenly spaced values in the range [20, 120] at increments of 0.01 excl
x = np.arange(20, 120, 0.01)
x_fine = np.linspace(20, 120, 1000)

# Calculate the PDF values at each location x
p = 1 / np.sqrt(2 * np.pi * sigma**2) * np.exp(-0.5 * ((x - mu) / sigma)**2)
p_fine = 1 / np.sqrt(2 * np.pi * sigma**2) * np.exp(-0.5 * ((x_fine - mu) / sigma)**2)

# Calculate the spacing between grid points
delta_x = x[1] - x[0]
delta_x_fine = x_fine[1] - x_fine[0]

# Plot the discrete approximation of the PDF as a bar chart
plt.bar(x, p, width=0.5, color='b', alpha=0.5, label='Discrete PDF')

# Plot the PDF curve
plt.plot(x_fine, p_fine, color='r', label='PDF Curve')

plt.xlabel('x')
plt.ylabel('f(x)')
plt.title('Discrete Approximation of Normal PDF')
plt.legend()
plt.show()

# Calculate the Riemann sum approximation of the integral
riemann_sum = np.sum(p * delta_x)
riemann_sum_fine = np.sum(p_fine * delta_x_fine)

riemann_sum, riemann_sum_fine
# >>> (1.0, 1.0)
```

The value of $P(20 \leq X < 120)$ is given as 1.0 for both degrees of approximation. This indicates that the event is almost certain to occur within the range $[20, 120]$ for the given normal distribution.