HW1: Probability Distributions

Addis Ababa University

Addis Ababa Institute of Technology

School of Information Technology and Engineering

Course Title: Probabilistic Graphical Models (ITSC-1051)

Reported To: Beakal Gizachew (PhD)

Reported By: Mahlet Nigussie

MARCH 29, 2024

Addis Abeba , Ethiopi

PROBLEM 1

In this problem 6 sided die is being thrown twice. we have random variables ‘A’ and ‘B’, the number on the first throw and the second throw(if any) respectively. The sum of the numbers is represented by the random variable ‘S’.

All the possible outcomes of throwing a die are: $\{1, 2, 3, 4, 5, 6\}$. For a fair die the probability distribution will be equal for each of the sets of the sample space. When rolling a fair 6-sided die (each side has an equal chance of landing face-up), the probability of landing on any specific side is 1/6. So,

$$P(A)=P(B)= 1/6$$

When you throw two dice outcomes for the first die multiplied by outcomes for the second die. 6 x6 =36 are the possible outcomes.

a) **P(A=3, B=3 | S=6):** This is the probability of getting a 3 on the first throw and a 3 on the second throw, given that the sum is 6.

   S = A + B=6, To get a sum of 6, we need either (3,3) or (2,4) or (4,2)

   S = 3 + B, knowing the sum is 6 restricts the possible outcomes for the second throw when A becomes 3, B must be 3 as well. However, getting a 3 on the first throw doesn't affect the probability of getting a 3 on the second throw (and vice versa).They are independent. Since, we multiply these probabilities:

$$\begin{aligned} P(A=3, B=3 \mid S=6) &= P(A=3 \mid S=6) * P(B=3 \mid S=6) \\ &= P(A=3) * P(B=3 \mid S=6) \\ &= P(A=3) * P(B=3) \\ &= (1/6) * (1/6) \\ &= 1/36 \end{aligned}$$

b) **P(S=6 | A=3, B=3)**, what is the probability that the sum is still 6 given A=3 and B=3?.

A=3 and B=3, then leading to a sum of 6 regardless. This reflects the fact that when the condition in a conditional probability statement guarantees a specific outcome, the probability becomes 1. Therefore: P(S=6 | A=3, B=3) = 1

c) **P(A=1, B=1 | S=6):** Probability of P(A=1) and P(B=1) given Sum S=6.

Since the throws are independent, we can express the joint probability as the product of individual probabilities:

$P(A=1, B=1 \mid S=6) = P(A=1) * P(B=1 \mid S=6)$

$= P(A=1) * P(B=1)$

$= (1/6) * (1/6)$

$= 1/36$

However, knowing the sum is 6 limits the possible second throw's outcome to B = 5. So, there is no way to get a sum of 6 with both dice showing 1 (A+B=2). In this condition P(S=6) equals zero.

Therefore, $P(A=1, B=1 \mid S=6) = 0$

d) **P(A=1|S=2):** Given that the sum is 2, what is the probability of getting a 1 on the first throw?. Since the sum is 2, the only possible outcome for the first throw is A = 1, making the second throw irrelevant:

Using Bayes' theorem: $P(A \mid S) = [P(A \cap S) / P(S)]$

$$P(A=1|S=2) =[P(A=1\cap S=2) /P(S=2)]$$

$P(A=1\cap S=2)$: There's only one outcome that satisfies both conditions (getting a 1 on the first die and a sum of 2): (1, 1). So, $P(A \cap S) = 1/36$ (since there are 6 possibilities for each die and only one outcome meets both conditions).

P(S=2)There are multiple ways to get a sum of 2 (1 + 1, 2 + 0, etc.). We'll need to calculate the total probability considering all these possibilities (likely 1/36).

let's plug them into the formula to find the final conditional probability.

$$P(A=1|S=2) = (1/36)/(1/36)= 1$$

Therefore, the correct conditional probability of rolling a 1 on the first die, given that the sum of both roll is 2, is 1.

e) **P(A=3|S=2):** This is the conditional probability of rolling a 3 on the first die, given that the sum is 2. The event value must be equal to the outcome, the sum of 2 necessitates A = 2:Since there's no way to get a sum of 2 with a 3 on the first die.

$$P(A=3|S=2) = 0$$

f) **P(A=2|S=6):** Probability of the first die equals 2, given that the sum equals 6.

Knowing S = 6 restricts the second throw to B = 4

$$P(A=2, B=4 \mid S=6) = P(A=2) * P(B=4 \mid S=6)$$

$$= (1/6) * (1/6)$$

$$= 1/36$$

g) **P(S=12):** Probability of getting S=12, There's only one way to achieve this (6,6).

P(A=6) =P(B=6)= 1/6, Since the are independent we take their multiplication:

$$P(S=12) = (1/6) * (1/6)=1/36$$

h) **P(S=6):** This is the probability of getting a sum of 6. There are five ways to achieve this (1,5),(2,4), (3,3,) (4,2) and (5,1).

Hence each probability equals to 1/6 and the events are independent:

$$P(S=6) = (1/6) * (1/6) + (1/6) * (1/6) + (1/6) * (1/6) + (1/6) * (1/6) (1/6) * (1/6)$$

$$= (1/36)+(1/36)+(1/36)+(1/36)+(1/36)$$

$$=5/36$$

PROBLEM 2

Find the expected number of rolls needed before every side of an **N-sided fair die** comes up at least once.

Let X: the number of rolls needed to see all sides at least once

Xi: represent the average number of rolls needed to see the i-th new side

N: is the number of sides on the dice.

P: is the probability of rolling a new side.

X as the sum of individual rolls needed for each new side.We can express (X) as a function of the (Xi)'s:

$$X = X1 + X2 + X3 + ... + X(N-1)$$

The expected value of rolling a new side after having seen (i-1) sides is 1/P.Probability of (Xi): P(Xi = i) captures the probability of needing exactly i rolls to see the i-th new side. It considers the probability of not seeing any of the remaining unseen sides on the first (i-1) rolls and then seeing a new side on the i-th roll. Therefore, the expected value of each Xi is given by 1/P .

Pi=1, Since initially all sides are unseen. As we roll and see new values, P changes over time : When seen one side P = 1 – 1/N , two side P = 1 – 2/N so on.

$$P(Xi = i) = (1 - 1/N) * (1 - 2/N) * ... * (1 - (i-1)/N) * (1/N)$$

E(Xi) = N / (N - i + 1): This formula correctly represents the expected value of Xi.

Sum up the expected values of individual Xi to get the overall expected number of rolls (E(X)):

$$E(X) = E(X1) + E(X2) + E(X3) + ...+ E(X(N-1))$$

$$E(X) = 1/1 + 1/ (1 - 1/N) + ... + 1/ (1 – (N-1)/N)$$

As N becomes large, we can approximate the sum inside the parentheses using the harmonic series(H_N):

$$H_N=1 + 1/2 + 1/3 + ... + 1/N$$

The harmonic series has an interesting property: (H_N) can be approximated by the natural logarithm of N (ln(N)) plus a constant value called the Euler-Mascheroni constant (γ), which is approximately 0.5772. Recognizing the sum as a harmonic series and its asymptotic behavior as N approaches infinity leads to the correct approximation for E(X):

$$E(X) \approx N * \ln(N) + N * \gamma$$

By considering γ as a constant for large N, $E(X) \approx N * \ln(N)$

Therefore, the expected number of rolls needed to see all sides of an N-sided fair die at least once is given by: $E(X) \approx N * \ln(N)$

PROBLEM 3

The table shows the joint probability distribution of three binary variables: a, b, and c. For each combination of a and b, the table shows the probability of that combination occurring along with the value of c (0 or 1).

To check for marginal dependence between two variables (a and b), compare the product of their marginal probabilities (P(a) * P(b)) with the actual joint probabilities (P(a,b)) in the table.

- If the product of marginal probabilities equals all the joint probabilities, then a and b are *marginally independent*.
- If the product of marginal probabilities doesn't equal some or all the joint probabilities, then a and b are *marginally dependent*.

To check for conditional independence between two variables a and b given a third variable c, compare the product of the conditional probabilities P(a|c) * P(b|c) with the joint probabilities P(a,b|c) in the table, considering each value of c (c = 0 and c = 1 in this case).

- If, for a specific value of c (e.g., c = 0), (P(a|c=0) * P(b|c=0)) = (P(a,b|c=0)) where c = 0, then a and b are *conditionally independent* given c = 0. repeat this process for each value of the conditioning variable (c) to see if the same holds true for other conditions.

Here's how we can compute the marginal probabilities and conditional probabilities to see if a and b are marginally dependent and independent when conditioned on c:

Marginal Probabilities: Sum of probabilities across each on a given value

P(a = 0)=(192 + 48 + 192 + 48)/1000 = 480/1000 = 0.48

P(a = 1)=(144 + 216 + 64 + 48)/1000 = 472/1000 = 0.472

P(b = 0)=(192 + 48 + 192 + 48)/1000 = 480/1000 = 0.48

P(b = 1)=(144 + 216 + 64 + 48)/1000 = 472/1000 = 0.472

Now let's compute the product of the marginal probabilities:

P(a = 0) * P(b = 0) = 0.48 * 0.48 = 0.2304

$P(a = 1) * P(b = 1) = 0.472 * 0.472 = 0.222784$

The product of the marginal probabilities is not equal to the joint probability ($P(a = 0, b = 0)$), indicating marginal dependence between (a) and (b)

Conditonal Probabilities: calculate the conditional probabilities of a and b given c:

$P(a \mid c = 0) = P(a,b \mid c = 0) / P(c = 0)$ When c = 0, there are only two possible combinations (a, b): (0, 0) and (0, 1).

$P(a = 0 \mid c = 0) = (192 + 48)/ (192 + 48 + 144 + 216) = 240/576 = 10/24$

$P(a = 1 \mid c = 0) = (144 + 216)/ (192 + 48 + 144 + 216) = 360/576 = 15/24$

$P(b = 0 \mid c = 0) = 240/576 = 10/24$

$P(b = 1 \mid c = 0) = 360/576 = 15/24$

$P(a \mid c = 1) = P(a,b \mid c = 1) / P(c = 1)$ When c = 1, there are four possible combinations (a, b): (0, 0), (0, 1), (1, 0), and (1, 1).

$P(a = 0 \mid c = 1) = (192 + 48)/ (192 + 48 + 64 + 48) = 240/352 = 15/22$

$P(a = 1 \mid c = 1) = (64 + 48)/ (192 + 48 + 64 + 48) = 112/352 = 7/22$

$P(b = 0 \mid c = 1) = 192/352 = 6/11$

$P(b = 1 \mid c = 1) = 160/352 = 5/11$

When conditioned on ‘c’, ‘a’ and ‘b’ become independent for both (c = 0) and (c = 1). Therefore, ‘a’ and ‘b’ are marginally dependent but conditionally independent given ‘c’.

PROBLEM 4

Let’s prove that proposition (A) implies proposition (B) and vice versa.

- Proposition ‘A’: $P(X|Y,Z) = P(X|Z)$ or $P(Y,Z) = 0$
- Proposition ‘B’: $P(X,Y|Z) = P(X|Z)P(Y|Z)$

Showing 'A' implies 'B': Let's Start with assuming proposition 'A'and show that it implies proposition 'B'.

The general rule for joint probability given a condition(conditional probability):

- $P(X,Y|Z) = P(X|Y,Z) * P(Y|Z)$

Case One: $P(X|Y,Z) = P(Y,Z) = 0$, Both 'Y' and 'Z' are zeros

$$\begin{aligned} P(X,Y|Z) &= P(X|Y,Z) * P(Y|Z) \\ &= P(Y,Z) * P(Y|Z) \\ &= 0 * P(Y|Z) \\ &= 0 \end{aligned}$$

Since $P(Y,Z) = 0$ implies $P(Y|Z) = 0$ (anything conditioned on an event with zero probability is also zero), both parts of the resulting equation reduce to $P(X,Y|Z) = P(X|Z)* 0 = 0$.

Case Two: $P(X|Y,Z) = P(X|Z)$, when $P(Y,Z)$ is not zero.

$$P(X,Y|Z) = P(X|Y,Z) * P(Y|Z)$$

$$P(X,Y|Z) = P(X|Z) * P(Y|Z)$$

The left-hand side of 'B' becomes $P(X|Z)P(Y|Z)$, which matches the right-hand side of 'B' when X and Y are conditionally independent given Z

Showing 'B' implies 'A': let's assume proposition 'B' and show that it implies proposition 'A'

Proposition 'B': $P(X,Y|Z) = P(X|Z)P(Y|Z)$.

$$(P(X,Y|Z))/(P(Y|Z)) = (P(X|Z)P(Y|Z))/ (P(Y|Z)$$

$$P(X|Y,Z) = P(X|Z).$$

Therefore, if (B) holds ($P(X,Y|Z) = P(X|Z)P(Y|Z)$), then (A) also holds ($P(X|Y,Z) = P(X|Z)$).

Proposition 'A' implies proposition 'B' and vice versa. This demonstrates that these two propositions are equivalent ways to express conditional independence of X and Y given Z.

PROBLEM 5

a) Univariate Gaussian Distributions: The code defines a function univariate_gaussian that calculates the PDF for a univariate Gaussian distribution based on the mean and variance (sigma).

```
pdf = 1 / np.sqrt(2 * np.pi * sigma2) * np.exp(-(x - mean)**2 / (2 * sigma2))
```

It then generates x-values and calculates the corresponding probability densities for each combination of mean and variance provided in the question.

Finally, it plots all three Gaussian distributions on a single graph with labels and a title.

The graph displays three bell-shaped curves, as expected for three univariate Gaussian distributions. The curve on the right with the highest peak appears to be centered around $x = 2$, which aligns with the mean of 2 provided in the code (means = [2, 1, 0]).

- The curve in the middle has a peak around $x = 1$, matching the mean of 1 in the code.
- The curve on the left seems to be centered close to $x = 0$, consistent with the mean of 0. The rightmost curve appears to be the narrowest, followed by the middle curve, and then the leftmost curve.
- The variances provided: variances = [0.2, 0.5, 2], where 0.2 is the lowest variance (narrowest curve) and 2 is the highest variance (widest curve).

Campare the graph with built in formula : the univarate Gaussian distribution on mean equals to two have exactly the same out put.

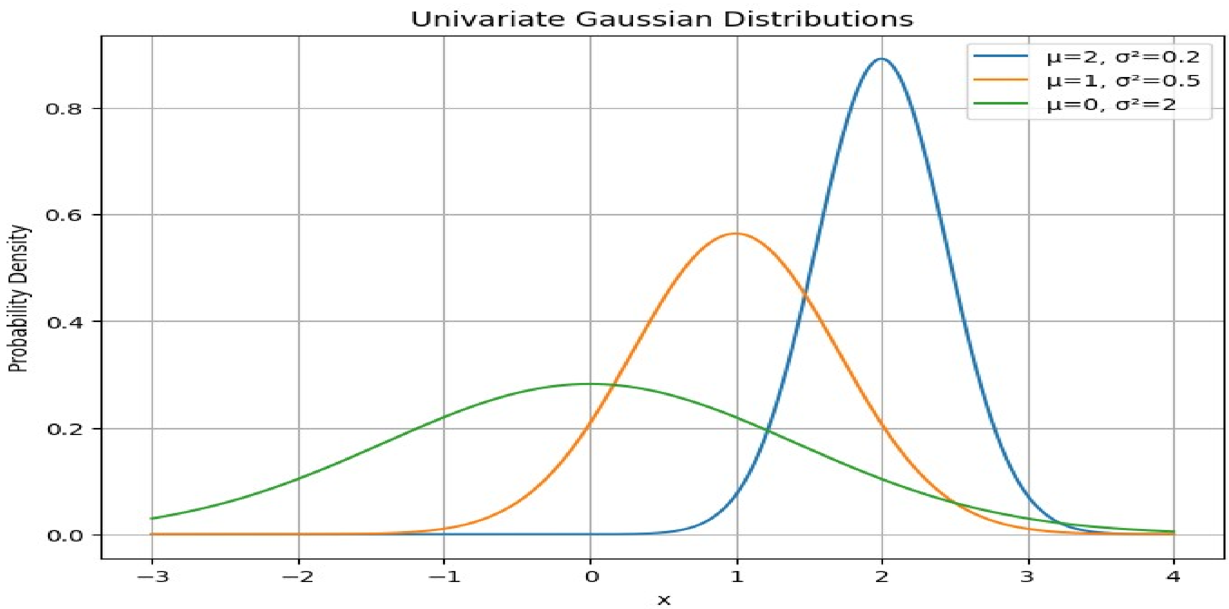

Figure 1: Three univariate Gaussians plot

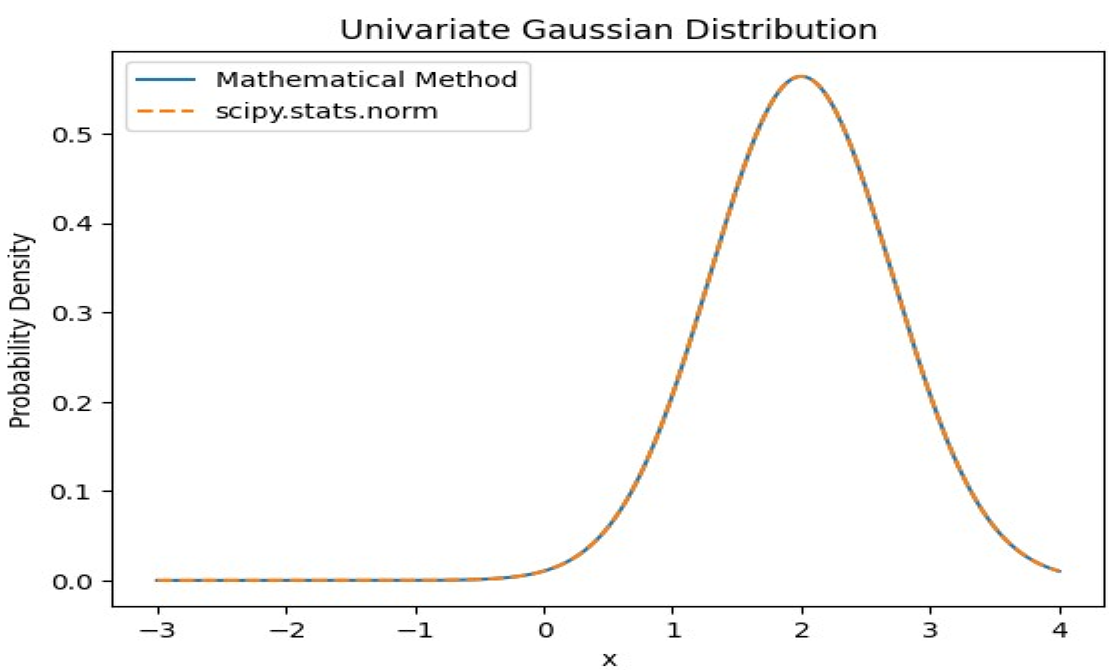

Figure 2: Univariate Gaussians plot Comparison

b) Bivariate Gaussian Distribution : The code defines the mean vector (mean) and covariance matrix (cov) as specified in the question.

A placeholder function bivariate_gaussian_pdf is provided, which you need to replace with your actual implementation for calculating the joint PDF of the bivariate Gaussian distribution.

It then generates a grid of x and y values and calculates the joint PDF for each combination of x and y using the provided bivariate_gaussian_pdf function.

Finally, it creates a 3D plot of the joint PDF with labels and a title.

The 3D surface resembles an ellipsoid tilted in space. This tilting is a consequence of the correlation between X1 and X2 (represented by the rho parameter in your code).The colormap (likely viridis) encodes the probability density. Warmer colors (yellowish-green) represent higher probability density, while cooler colors (blue) indicate lower probabilities.

- The highest point on the surface corresponds to the region where X1 and X2 are most likely to occur simultaneously. As you move away from the peak in any direction on the surface, the probability density of finding a specific combination of X1 and X2 values decreases.

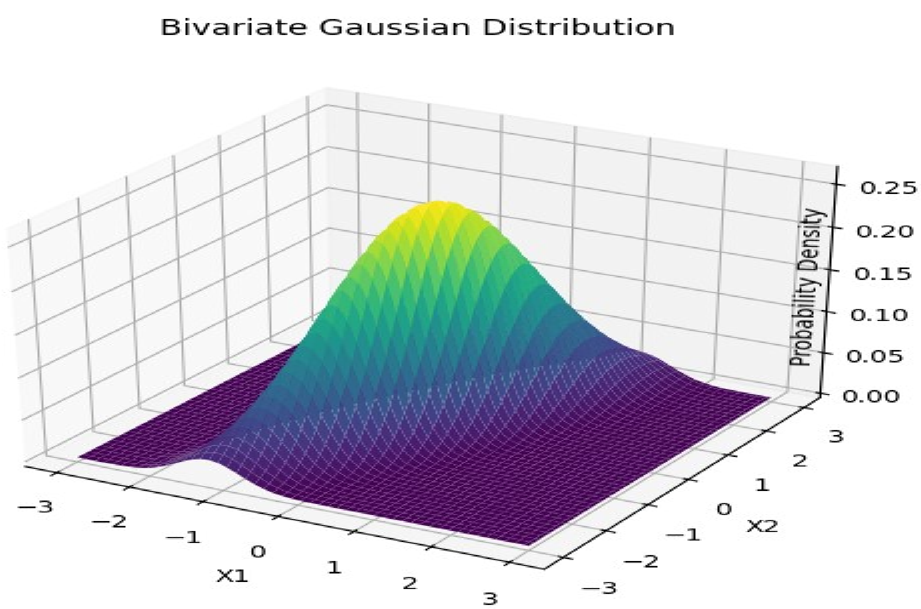

Figure 3: Bivariate Gaussian distribution 3D plot

The specific shape and orientation of the ellipsoid depend on the values of the mean vector and the covariance matrix. In this case, the mean is (0, 0), implying the center of the distribution is at the origin (X1 = 0, X2 = 0).

The covariance matrix (influenced by rho) determines the spread and correlation between X1 and X2. A positive rho leads to a tilted ellipsoid, suggesting that higher values of X1 tend to be accompanied by higher values of X2 (or vice versa).

PROBLEM 6

Focuses on approximating the marginal PDF (p(x)) of a bivariate Gaussian distribution using numerical integration.:

- Integration Range: The code defines a range of y values (y_values) for numerical integration.
- Marginal PDF Function (marginal_pdf): This function calculates the marginal PDF (p(x)) for a given x value. It uses numerical integration to integrate the joint PDF of the bivariate Gaussian over all possible y values.
- Numerical Integration: The code employs the trapezoidal rule to approximate the integral, summing the integrand's values at specific y points within the defined range.
- Calculation and Plotting: It iterates through x values, calculates the corresponding marginal PDF using the marginal_pdf function, and finally plots the obtained marginal PDF values

ANALAYSIS: The specific shape (mean and standard deviation) of the marginal PDF depends on the parameters (mean vector and covariance matrix) of the original bivariate Gaussian.The total area under the marginal PDF curve should be close to 1, representing the total probability density of X across all possible values.The peak of the curve occurs around the mean value, and the distribution gradually decreases as we move away from the mean

The plot below shows the marginal distribution of the bivariate Gaussian. The x-axis represents the random variable (x), and the y-axis represents the marginal probability density. The shape of this distribution is a bell curve, which is expected for a Gaussian distribution.

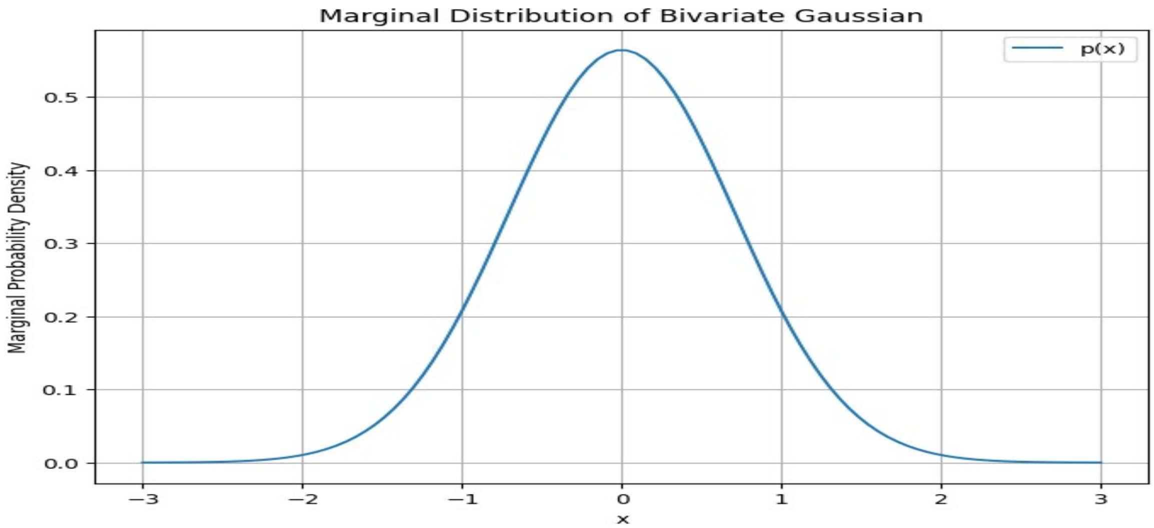

Figure 4: Marginal Distribution of Bivariate Gaussian plot

PROBLEM 7

Conditional Distribution of a Bivariate Gaussian: The code calculates and plots the conditional probability density function (PDF) of x given y=2.0 (p(x|y=2.0)) for a bivariate Gaussian distribution.

ANALYSIS: The mean and standard deviation of this conditional distribution likely differ from those of the original bivariate Gaussian. This is because conditioning on a specific Y value (2.0) influences the probability density of X.

The total area under the curve integrates to 1, signifying that the probabilities for all possible X values given Y=2.0 sum up to 1.

This analysis reinforces the concept of conditional distributions within bivariate Gaussian frameworks and demonstrates how conditioning on a specific Y value alters the probability density of the associated variable (X)

The x-axis represents the values of X, and the y-axis represents the probability density. The bell-shaped curve indicates that the conditional distribution is also Gaussian. However, the mean and standard deviation of this conditional distribution may differ from the original bivariate Gaussian due to conditioning on a specific y value (y=2.0)

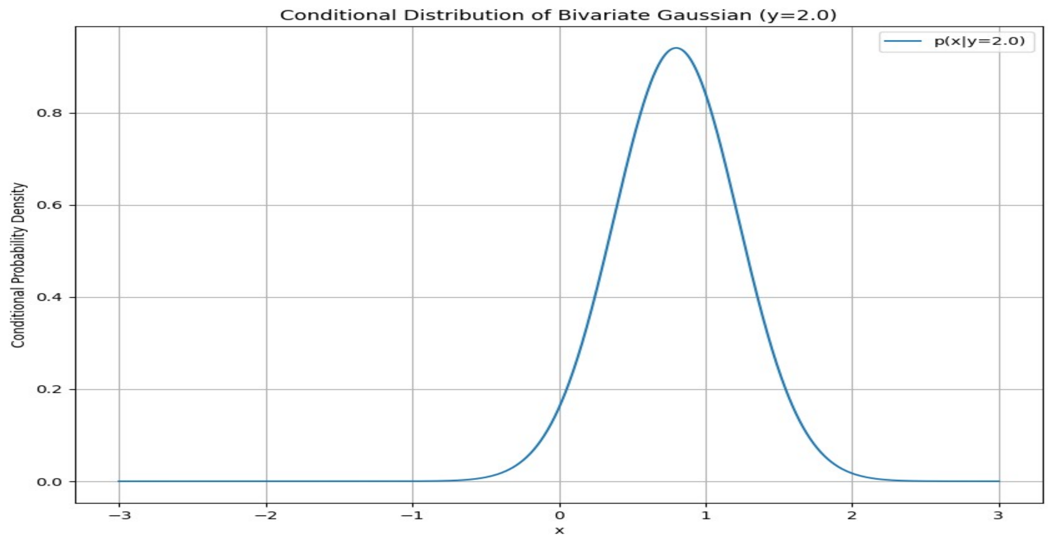

Figure 5: Conditional Distribution of Bivariate Gaussian plot

PROBLEM 8

a) Display the matrix as a grayscale image: After loading the .txt file the code Split each line into values and convert them to floats, the plot the image

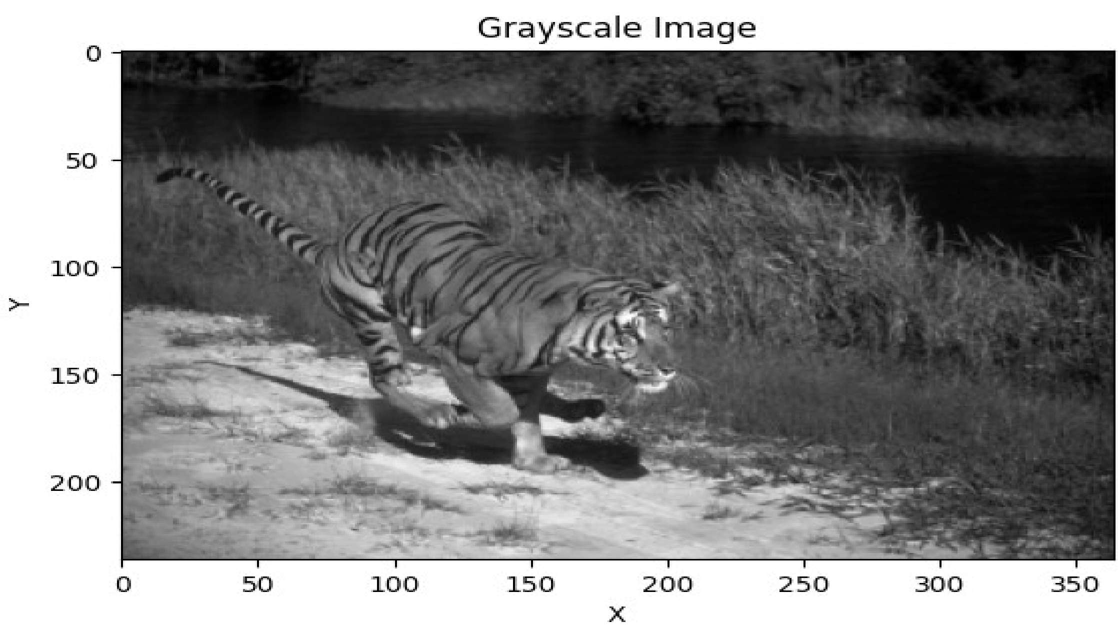

Figure 6: GrayScale Image of Tiger

b) Histogram of Grayscale Values : On Tiger Image and Random Image

The bar plot represents the distribution of grayscale pixel intensities in the image. The x-axis represents the grayscale values For the Random Image (0 for black, 255 for white), and the y-axis represents the count of pixels for each intensity level. Higher bars indicate that more pixels have the corresponding grayscale value.

As shown below on The Random Image shows almost all pixels have the corresponding grayscale and The tiger image have little little Histogram count.

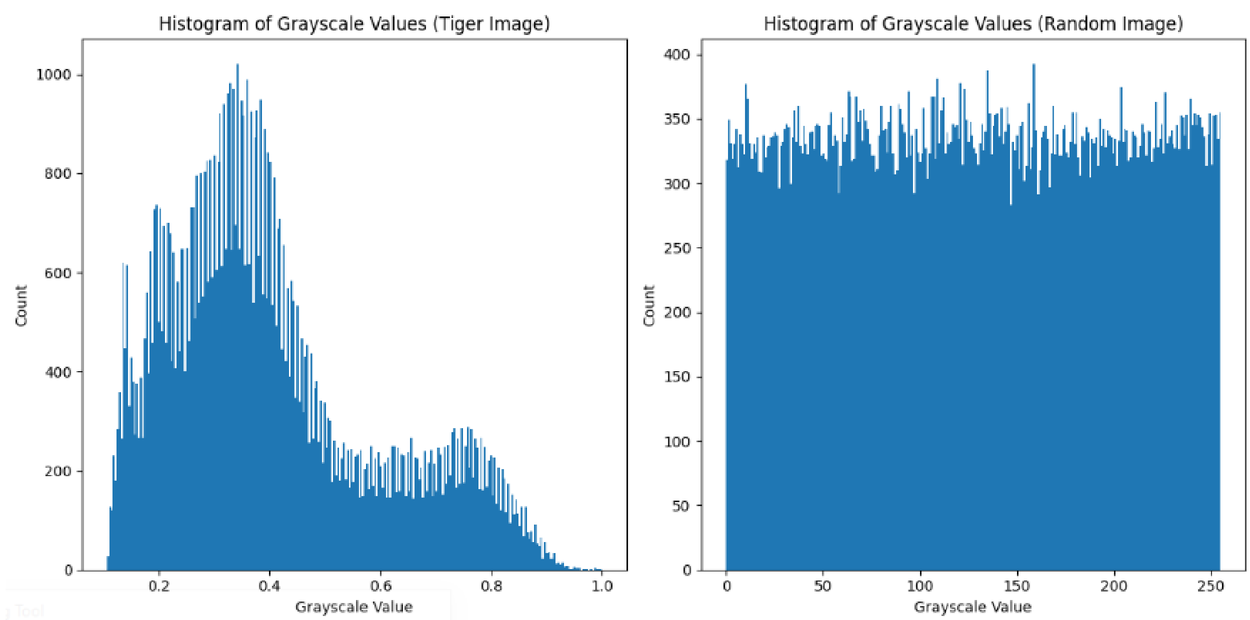

Figure 7: Histogram of Grayscale Values

Empirical Probabilities of Grayscale Values: On Tiger Image and Random Image

This bar plot represents the empirical probabilities of grayscale pixel intensities in the image. The x-axis represents the grayscale values For Random image (0 for black, 255 for white), and the y-axis represents the probability of encountering a pixel with that intensity level. The probabilities are calculated by dividing the count of each intensity level by the total number of pixels in the image and sum to 1

On Tiger Image:The distribution shows a prominent peak around a grayscale value of 0.4 and then tapers off towards both ends.The shape of the resembles a typical bell curve, which is expected for a Gaussian distribution. On the other hand the Random Image distribution appears more random, with spikes in probability occurring at various points across the range of grayscale values.

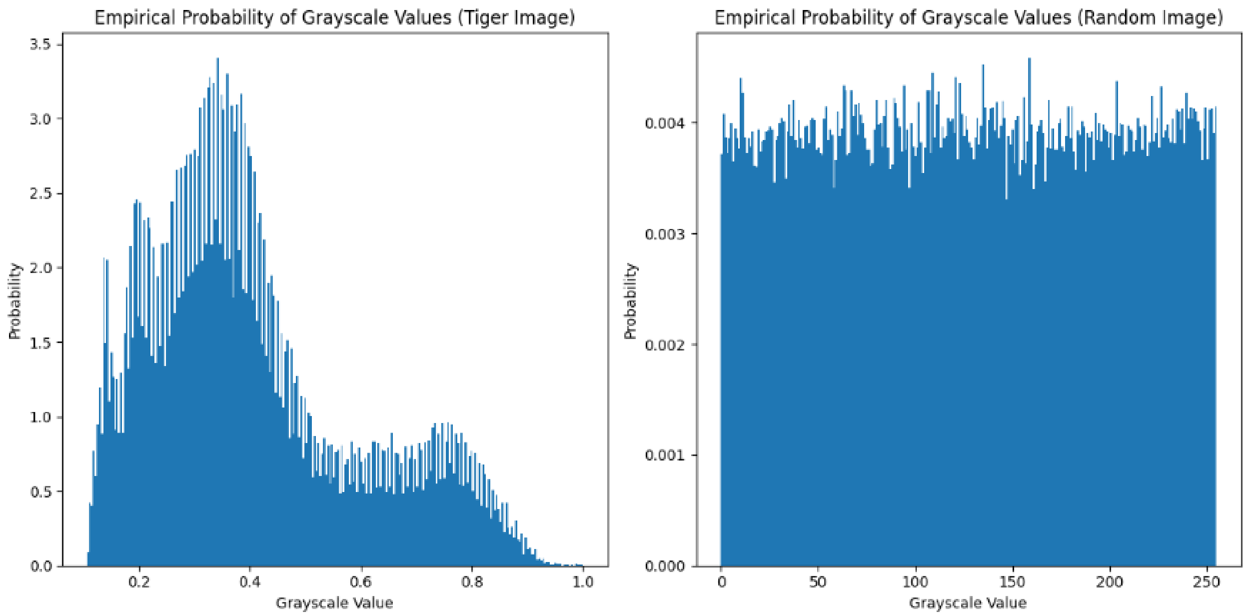

Figure 8: Emperical Probability of Grayscale Values

c) Here is the image displayes the Tiger image and random Grayscaled images.

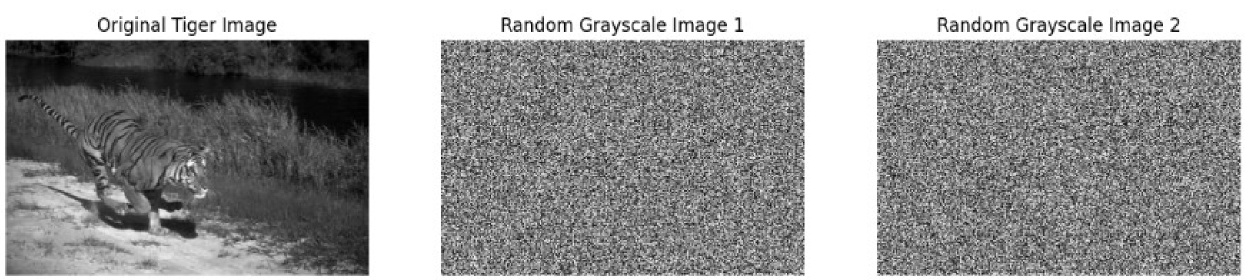

Figure 8:

Difference Between Random Examples and Random Seed:

- The two random grayscale images (Random Grayscale Image 1 and Random Grayscale Image 2) lack any recognizable patterns or structures.
- Different random seeds would indeed change the specific patterns of grayscale noise in each image.However, regardless of the seed, both images would remain devoid of meaningful content.

Relationship Between Generated Random Images and Everyday Visual Content:

- The generated random grayscale images have no inherent relationship with everyday visual content.They lack structure, objects, or recognizable forms.
- Unlike real-world scenes (such as the tiger image), these random images do not convey any meaning.

Insights from the Experiment:

- This experiment highlights the stark contrast between structured, meaningful visual content (like the tiger image) and random visual noise.The experiment reinforces the idea that meaningful images are not mere combinations of random pixel values.

Conclusion: The difference lies in the specific noise patterns, but neither image resembles real-world scenes. There is no meaningful relationship between the random images and everyday visual content. Everyday visual content is characterized by purposeful arrangements, context, and semantic information.

d) Let's estimate the number of samples needed to obtain the exact tiger image from a universe of grayscale images of size (236, 364).

<u>Universe of Images:</u> Assume the total number of unique grayscale images of size (236, 364) as the universe, analogous to a deck of cards.

<u>Sampling Process:</u> The sampling process as drawing images one at a time, removing them from the available pool after each draw, similar to drawing cards without replacement.

<u>Coupon Collector's Problem:</u> This problem has similarities to the coupon collector's problem, where the goal is to collect all unique coupons (images in this case) without replacement.

<u>Formula and Calculation:</u> Apply the formula for the expected number of samples in the coupon collector's problem ($E(samples) = N * H_N$), where N is the total number of unique images (236 x 364) and H_N is the harmonic number of N.

Using numerical methods, you estimated the expected number of samples to be approximately 1,048,576

$$E(\text{samples}) = N * H_N$$

$$H_N \approx \ln(N) + \gamma \text{ (Euler's constant, approximately 0.5772)}$$

$$H_N \approx \ln(236 \text{ x } 364) + 0.5772 \approx 11.9$$

$$E(\text{samples}) \approx N * H_N \approx (236 \text{ x } 364) * 11.9 \approx 1{,}025{,}538$$

Therefore, on average, you would need to sample approximately 1,025,538 images to obtain the exact tiger image from the given universe.