CS215 HomeWork Assignment - 2

Team members :-

- Kavya Gupta (22B1053)
- Aryan Katiyar (22B1205)
- Arvind Kumar Yadav (22B1025)

Honor Code : No scope of plagiarism and external help from other teams sought, all work distributed and worked on properly.

Solutions to all questions present below.
(Q1 to Q7 in order)

Note : In the case of Q5 and Q6, the very first section tells about the location of the code file and output pictures.

Q1 $X_1, X_2, X_3 \ldots X_n \rightarrow$ independent identically distributed random variables.

$$cdf = F_X(x)$$
$$pdf = f_X(x) = F_X'(x).$$

$$P(Y_1 \le y) = P(X_1 \le y, X_2 \le y \ldots X_n \le y)$$

As $Y_1 = \max\{X_1, X_2 \ldots X_n\}$.

$\therefore$ if Y_1 is smaller than/equal to some value, then all X_i are also smaller than/equal to that value as all X_i are smaller than/equal to Y_1.

As $X_1, X_2 \ldots X_n$ are independent.

$\therefore$ ~~$P(X_1 \le y) F$~~

$$\therefore P(X_1 \le y, X_2 \le y \ldots X_n \le y) = P(X_1 \le y).P(X_2 \le y) \ldots P(X_n \le y)$$

$$\therefore P(Y_1 \le y) = P(X_1 \le y).P(X_2 \le y) \ldots P(X_n \le y).$$

$$P(X_i \le y) = F_X(y) \quad \forall\, i \in \{1, 2 \ldots n\}.$$

$$\therefore P(Y_1 \le y) = [F_X(y)]^n$$

$$\boxed{F_{Y_1}(y) = [F_X(y)]^n} \rightarrow \text{cdf of } Y_1.$$

$$\text{Pdf of } Y_1 = F_{Y_1}'(y) = f_{Y_1}(y) = n.[F_X(y)]^{n-1}.f_X(y)$$

$$\therefore \boxed{f_{Y_1}(y) = n.f_X(y).[F_X(y)]^{n-1}} \rightarrow \text{pdf of } Y_1.$$

Similarly for Y_2.

$$P(Y_2 > y) = P(X_1 > y, X_2 > y \ldots X_n > y)$$

As $Y_2 \le X_i \quad \forall\, i \in \{1, 2 \ldots n\}$.

$\therefore$ if $y < Y_2 \Rightarrow y < X_i$

As $X_1, X_2 \ldots X_n$ are independent.

$$\therefore P(X_1 > y, X_2 > y \ldots X_n > y) = P(X_1 > y).P(X_2 > y) \ldots P(X_n > y)$$

$$P(X_i > y) = 1 - P(X_i \leq y)$$

$$P(X_i > y) = 1 - F_X(y) \quad \forall \quad i \in \{1, 2 \ldots n\}.$$

$$\therefore P(Y_2 > y) = [1 - F_X(y)]^n$$

$$1 - P(Y_2 \leq y) = [1 - F_X(y)]^n$$

$$\therefore \boxed{F_{Y_2}(y) = 1 - [1 - F_X(y)]^n} \rightarrow \text{cdf of } Y_2.$$

$$F'_{Y_2}(y) = f_{Y_2}(y) = (-n)[1 - F_X(y)]^{n-1}(-f_X(y))$$

$$\boxed{f_{Y_2}(y) = n \cdot f_X(y) \cdot [1 - F_X(y)]^{n-1}} \rightarrow \text{pdf of } Y_2.$$

Solution for Q2

Kavya Gupta

September 2023

Understanding the Gaussian Mixture Model (GMM) :

So interesting thing to note about GMM is that it is **not** a mere linear combination of random variables (r.v.). In latter, all the component r.v. X_is return a value whose linear combination is taken.
But in GMM, as clearly stated in the question, actually only one of the X_is are chosen with probability p_i and from that value is taken. Hence at a particular measurement, **only and exactly** one r.v. comes into action.

Now, for GMM, we will make use of **Law of Total Expectation/Law of Iterated Expectation/Adam's Rule** which states :-

$$X \sim \sum_i p_i X_i \Rightarrow E[g(X)] = \sum_i E[g(X)|(X = X_i)]P(X = X_i) = \sum_i E[g(X_i)]P(X = X_i) = \sum_i p_i E[g(X_i)]$$

Where g is a function on a r.v.

Solving the GMM :-

Before moving ahead, let's see what mean and variance of a Gaussian Variable $\mathcal{N}(\mu_i, \sigma_i^2)$

$$E[X_i] = E[\mathcal{N}(\mu_i, \sigma_i^2)] = \mu_i \tag{1}$$

$$E[(X_i - \mu_i)^2] = \sigma_i^2 \tag{2}$$

As $Var(Y) = E[Y^2] - E[Y]^2$ for any r.v. Y,

$$E[X_i^2] = Var(X_i) + E[X_i]^2 = \sigma_i^2 + \mu_i^2 \tag{3}$$

Mean E(X)

Using the above Law and equation (1)...

$$\mu = E[X] = \sum_i p_i E[X_i] = \sum_{i=1}^{K} p_i \mu_i \tag{4}$$

Variance Var(X)

$$Var(X) = E[(X-\mu)^2]$$

From the above Law,

$$E[(X-\mu)^2] = \sum_i p_i E[(X_i-\mu)^2]$$

Opening by the linearity of the expectation operator...

$$E[(X_i-\mu)^2] = E[X_i^2 - 2\mu X_i + \mu^2] = E[X_i^2] - 2\mu E[X_i] + E[\mu^2]$$

From equations (1), (3) and (4)...

$$E[X_i^2] - 2\mu E[X_i] + E[\mu^2] = \sigma_i^2 + \mu_i^2 - 2\mu\mu_i + \mu^2$$

$$\Rightarrow \sum_i p_i E[(X_i-\mu)^2] = \sum_i p_i(\sigma_i^2+\mu_i^2-2\mu\mu_i+\mu^2) = \sum_i [p_i(\sigma_i^2+\mu_i^2) - 2\mu(\sum_i p_i\mu_i) + \mu^2(\sum_i p_i)]$$

Given $\sum_i p_i = 1$, so...

$$\Rightarrow \sum_i p_i E[(X_i-\mu)^2] = \sum_i [p_i(\sigma_i^2+\mu_i^2)] - 2\mu(\mu) + \mu^2(1) = \sum_i [p_i(\sigma_i^2+\mu_i^2)] - \mu^2$$

Hence...

$$Var(X) = \sum_{i=1}^{K} [p_i(\sigma_i^2+\mu_i^2)] - \mu^2$$

MGF(X) $\Phi_X(t)$

For a Gaussian r.v. $X_i = \mathcal{N}(\mu_i, \sigma_i^2)$, its MGF is...

$$\Phi_{X_i}(t) = E[e^{tX_i}] = e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2}$$

So using the above Law...

$$\Phi_X(t) = E[e^{tX}] = \sum_i p_i E[e^{tX_i}] = \sum_i p_i \Phi_{X_i}(t) = \sum_{i=1}^{K} p_i e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2}$$

Solving the Linear Combination System :

Now we have a r.v. Z such that $Z = \sum_i p_i X_i$ and as said in the very beginning, all the component r.v. X_is return a value and their linear combination is taken. We shall make extensive use of the **linearity property** of the expectation operator...

Mean E(Z)

$$E[Z] = E[\sum_i p_i X_i] = \sum_i E[p_i X_i] = \sum_i p_i E[X_i] = \sum_{i=1}^{K} p_i \mu_i = \mu$$

Peculiarly it's same as the GMM case..

Variance Var(Z)

$$Var(Z) = E[(Z-\mu)^2] = E[(\sum_i p_i X_i - \mu)^2]$$

$$\Rightarrow E[(\sum_i p_i X_i)^2 - 2\mu Z + \mu^2] = E[(\sum_i p_i X_i)^2] - 2\mu E[Z] + \mu^2 = E[(\sum_i p_i X_i)^2] - 2\mu(\mu) + \mu^2 = E[(\sum_i p_i X_i)^2] - \mu^2$$

Now, using this equation : $(\sum_i p_i X_i)^2 = \sum_{1 \le i \le K} (p_i X_i)^2 + 2 \sum_{1 \le i < j \le K} p_i p_j X_i X_j$...

$$E[(\sum_i p_i X_i)^2] = E[\sum_{1 \le i \le K} (p_i X_i)^2 + 2 \sum_{1 \le i < j \le K} p_i p_j X_i X_j]$$

$$\Rightarrow \sum_{1 \le i \le K} p_i^2 E[X_i^2] + 2 \sum_{1 \le i < j \le K} p_i p_j E[X_i X_j] = \sum_{1 \le i \le K} p_i^2 E[X_i^2] + 2 \sum_{1 \le i < j \le K} p_i p_j E[X_i] E[X_j]$$

$E[X_i X_j] = E[X_i]E[X_j]$ is possible because for $i \neq j, X_i, X_j$ are **independent random variables**.
Using equation (1) and (3)...

$$\Rightarrow E[(\sum_i p_i X_i)^2] = \sum_{1 \le i \le K} p_i^2(\sigma_i^2 + \mu_i^2) + 2 \sum_{1 \le i < j \le K} p_i p_j \mu_i \mu_j = \sum_{1 \le i \le K} [p_i^2 \sigma_i^2] + (\sum_{1 \le i \le K} p_i^2 \mu_i^2 s + 2 \sum_{1 \le i < j \le K} p_i p_j \mu_i \mu_j)$$

$$\Rightarrow E[(\sum_i p_i X_i)^2] = \sum_{1 \le i \le K} [p_i^2 \sigma_i^2] + (\sum_i p_i \mu_i)^2 = \sum_{1 \le i \le K} [p_i^2 \sigma_i^2] + \mu^2$$

Finally...

$$Var(Z) = E[(\sum_i p_i X_i)^2] - \mu^2 = \sum_{1 \le i \le K} [p_i^2 \sigma_i^2] + \mu^2 - \mu^2 = \sum_{1 \le i \le K} p_i^2 \sigma_i^2$$

MGF(Z) $\Phi_Z(t)$:

Here we will make use of the fact that for r.v. $A = \sum_i A_i$ where A_is are independent r.v., then...

$$\Phi_A(t) = E[e^{tA}] = E[e^{t \sum_i A_i}] = E[\prod_i e^{tA_i}] = \prod_i E[e^{tA_i}] = \prod_i \Phi_{A_i}(t)$$

As A_is are independent r.v.s, then e^{tA_i}s are **also independent**... that's why they're expectation can be split into individual expectations.

Also, as A_is are independent, p_iA_is are also independent... So for $A = Z$ take $A_i = p_iX_i$, hence...

$$\Phi_Z(t) = \prod_i \Phi_{p_iX_i}(t)$$

Also, MGF(p_iX_i) is :

$$\Phi_{p_iX_i}(t) = E[e^{(p_it)X_i}] = e^{\mu_ip_it+\frac{1}{2}\sigma_i^2(p_it)^2}$$

$$\Rightarrow \Phi_Z(t) = \prod_i e^{\mu_ip_it+\frac{1}{2}\sigma_i^2(p_it)^2} = e^{\sum_i \mu_ip_it+\frac{1}{2}\sigma_i^2(p_it)^2} = e^{\sum_i(\mu_ip_i)t+\frac{1}{2}\sum_i(p_i^2\sigma_i^2)t^2}$$

And from results of mean and variance we calculated above...

$$\Rightarrow \Phi_Z(t) = e^{\mu t+\frac{1}{2}\sigma^2t^2}$$

This form of MGF is exactly same as of a Gaussian Variable with mean $= \mu$ and variance $= \sigma^2 = \sum_i p_i^2\sigma_i^2$, in other words, Z is also a Gaussian Variable : $Z = \mathcal{N}(\mu, \sigma^2)$.

PDF(Z):

Now that we know that Z is also a Gaussian Random Variable, we can calculate PDF for it easily.. it comes from the definition of the Gaussian itself...

$$PDF(Z) = f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

Where $\mu = \sum_i p_i\mu_i$ and $\sigma^2 = \sum_i p_i^2\sigma_i^2$ as we had calculated above.

22B1053- 22B1205-22B1025

Page No
Date

~~Sol~~

Q3 Given X is random Variable With mean μ, and variance σ^2.

(a) We first prove that for $\tau>0$,

$$P(X-\mu \geq \tau) \leq \frac{\sigma^2}{\sigma^2+\tau^2}$$

let Y is random variable given by

$$Y = X-\mu$$

then by

$$E(Y) = E(X-\mu) = E(X) - E(\mu) = E(X)-\mu = \mu-\mu=0.$$

$$Var(Y) = E(Y-E(Y))^2 = E(Y^2) = E((X-\mu)^2) = Var(X) = \sigma^2.$$

let $b>0$, be arbitary positive real number.

let Z be random variable given by,

$$Z = (Y+b)^2$$

then,

$$E(Z) = E((Y+b)^2) = E(Y^2+2Yb+b^2)$$

$$= E(Y^2)+E(2Yb)+E(b^2)$$

$$= E(Y^2)+E(b^2)+2bE(Y)$$

$$= \sigma^2+b^2+0 = \sigma^2+b^2$$

now we prove that $Y \geq \tau \Rightarrow Z \geq (\tau+b)^2$

Proof: Given $(Y-\tau) \geq 0$

now $Z-(\tau+b)^2 = (Y+b)^2-(\tau+b)^2$

$$= (Y-\tau)(Y+\tau+2b) \geq 0$$

$$\Rightarrow Z \geq (\tau+b)^2$$

$$\therefore P(Y\geq\tau) \leq P(Z\geq(\tau+b)^2)$$

using Markov's inequality, (as Z is non-negative)

$$P(Y\geq\tau) \leq P(Z\geq(\tau+b)^2) \leq \frac{E(Z)}{(\tau+b)^2} = \frac{\sigma^2+b^2}{(\tau+b)^2}$$

$$P(Y\geq\tau) \leq \frac{\sigma^2+b^2}{(\tau+b)^2}$$

$\because$ b is arbitiory positive real number.

we can minimize $\frac{\sigma^2+b^2}{(\tau+b)^2}$, let $l = \frac{\sigma^2+b^2}{(\tau+b)^2}$

One way is ~~not~~ to see l as function of b, and we differentiation

Page No.: Date:

But, we minimize it we using ~~AM-GM~~ ~~inequality~~ subsitution and then differentiation,

$$l = \frac{\sigma^2 + b^2}{(\tau + b)^2}$$

put $\alpha = \tau + b$, (note: $b > 0 \Rightarrow \alpha > \tau$)

$$l(\alpha) = l = \frac{\sigma^2 + (\alpha - \tau)^2}{\alpha^2} = \frac{\sigma^2 + \tau^2 - 2\alpha\tau + \alpha^2}{\alpha^2} = \frac{\sigma^2 + \tau^2}{\alpha^2} - \frac{2\tau}{\alpha} + 1$$

now as $\alpha > \tau \therefore$ l is differentiable w.r.t α.

$$l'(\alpha) = \frac{-2(\sigma^2 + \tau^2)}{\alpha^3} + \frac{2\tau}{\alpha^2} = \frac{2(\tau\alpha - (\sigma^2 + \tau^2))}{\alpha^3}$$

$$l'(\alpha) = 0 \Rightarrow \alpha = \frac{\sigma^2 + \tau^2}{\tau}$$

as for $\alpha > \frac{\sigma^2 + \tau^2}{\tau}$, $l'(\alpha) > 0$

$\therefore$ $\alpha = \frac{\sigma^2 + \tau^2}{\tau}$ is point of minimce

we put $\alpha = \frac{\sigma^2 + \tau^2}{\tau}$ in $\alpha = \tau + b$

$$\Rightarrow b = \frac{\sigma^2}{\tau}$$

now we put $b = \frac{\sigma^2}{\tau}$ in $l = \frac{\sigma^2 + b^2}{(\tau + b)^2}$

$$\therefore l \geq \frac{\sigma^2\left(1 + \frac{\sigma^2}{\tau^2}\right)}{\left(\tau + \frac{\sigma^2}{\tau}\right)^2} = \frac{\sigma^2\left(1 + \frac{\sigma^2}{\tau^2}\right)}{\tau^2\left(1 + \frac{\sigma^2}{\tau^2}\right)^2}$$

$$l \geq \frac{\sigma^2}{\tau^2\left(1 + \frac{\sigma^2}{\tau^2}\right)} = \frac{\sigma^2}{\tau^2 + \sigma^2}$$

$$\therefore P(Y \geq \tau) \leq \frac{\sigma^2}{\tau^2 + \sigma^2}$$ (as $b > 0$ was arbitrary)

$$\Rightarrow P(X - \mu \geq \tau) \leq \frac{\sigma^2}{\tau^2 + \sigma^2}$$

M T W T F S S

Page No.:

Date:

YOUVA

(b) Now we prove for $\tau < 0$, we have,

$$P(X-\mu \geq \tau) \geq 1 - \frac{\sigma^2}{\sigma^2+\tau^2}$$

we first oberse that $X-\mu \geq \tau \Rightarrow \mu - X \leq -\tau$

now $-\tau > 0$

$\therefore$ ~~from~~

$E(\mu-X) = E(\mu) - E(X) = \mu - E(X)$
$= 0$

$Var(\mu - X) = E[(\mu-X)^2]$
$= E[(X-\mu)^2] = \sigma^2$

$\because Var(\mu-X) = \sigma^2$ and $-\tau > 0$ when $E(\mu-X) = 0$,

$\therefore$ from previous argument, ~~we have~~, applying on $\mu - X$,

$$\therefore P(\mu - X \geq -\tau) \leq \frac{\sigma^2}{\sigma^2+(-\tau)^2} = \frac{\sigma^2}{\sigma^2+\tau^2} \quad \text{— (1)}$$

$$\text{now} \quad P(\mu - X \geq -\tau) = P(X-\mu \leq \tau) \quad \text{— (2)}$$

~~now $\because \{x=\mu\pm\tau$~~

now $\because \{x \mid x-\mu \leq \tau\}$ and $\{x \mid x-\mu > \tau\}$ are mutually exclusive and exhaustive sets.

$\therefore$

$$P(X-\mu \leq \tau) + P(X-\mu > \tau) = 1$$

$$1 - P(X-\mu > \tau) = P(X-\mu \leq \tau)$$

from (1) and (2)

$$1 - P(X-\mu > \tau) = P(X-\mu \leq \tau) \leq \frac{\sigma^2}{\sigma^2+\tau^2}$$

$$\Rightarrow P(X-\mu > \tau) \geq 1 - \frac{\sigma^2}{\sigma^2+\tau^2}$$

Now if we X to be continous in neighbourhood of $x = \mu + \tau$ then $P(X = \mu + \tau) = 0$

$$\Rightarrow P(X-\mu \geq \tau) \geq 1 - \frac{\sigma^2}{\sigma^2+\tau^2}$$

CS215 Assignment 2

Kavya Gupta

September 2023

Solution for Q4

Proving $P(X \geq x) \leq e^{-tx}\Phi_X(t), t > 0$

When X is a Discrete Random Variable

For a discrete r.v., its MGF

$$\Phi_X(t) = E(e^{tX}) = \sum_{i=1}^{n} e^{tx_i} P(X = x_i) \tag{1}$$

Where x_is are the outcomes of X. So any real x lies in $(-\infty, x_1]$ or (x_n, ∞) or $(x_i, x_{i+1}]$ for some i.
If $x \in (x_n, \infty)$, then there exists no x_i such that $X \geq x$, so trivially $P(X \geq x) = 0$ which is definitely lesser than RHS as RHS ≥ 0. Hence this case is trivially proved.
If $x \in (x_i, x_{i+1}]$, then $x_k \geq x$ for all $k \in \{i+1, i+2, ..., n\}$, and for $x \in (-\infty, x_1]$, $x_k \geq x$ for all $k \in \{1, 2, ..., n\}$, hence :-

$$P(X \geq x) = \sum_{r=i+1}^{n} P(X = x_r)$$

Also, $e^{t(x_r - x)} \geq 1$ as $t > 0$ and $x_r \geq x$, hence :-

$$P(X \geq x) = \sum_{r=i+1}^{n} P(X = x_r) \leq \sum_{r=i+1}^{n} e^{t(x_r - x)} P(X = x_r) = e^{-tx} \sum_{r=i+1}^{n} e^{tx_r} P(X = x_r)$$

It will be less than or equal to summation from 1 to n as all the terms are non-negative.

$$\Rightarrow P(X \geq x) \leq e^{-tx} \sum_{r=i+1}^{n} e^{tx_r} P(X = x_r) \leq e^{-tx} \sum_{r=1}^{n} e^{tx_r} P(X = x_r)$$

Comparing with MGF, we see that :-

$$P(X \geq x) \leq e^{-tx} \sum_{r=1}^{n} e^{tx_r} P(X = x_r) = e^{-tx}\Phi_X(t) \Rightarrow P(X \geq x) \leq e^{-tx}\Phi_X(t)$$

Hence proved for discrete r.v.

When X is a Continuous Random Variable

MGF of a continuous r.v. is

$$\Phi_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \tag{2}$$

Where $f_X(x)$ is the pdf of X, also :-

$$P(X \geq x_0) = \int_{x_0}^{\infty} f_X(x)\,dx$$

Again $e^{t(x-x_0)} \geq 1$ as $t > 0$ and $x \geq x_0$ in the above integral, so $f_X(x) \leq e^{t(x-x_0)} f_X(x)$ for all $x \geq x_0$, hence :-

$$P(X \geq x_0) = \int_{x_0}^{\infty} f_X(x)\,dx \leq \int_{x_0}^{\infty} e^{t(x-x_0)} f_X(x)\,dx = e^{-tx_0} \int_{x_0}^{\infty} e^{tx} f_X(x)\,dx$$

Also, since $e^{tx} f_X(x) \geq 0$ for all x, the above will be $\leq$ integral from $-\infty$ to ∞.

$$P(X \geq x_0) \leq e^{-tx_0} \int_{x_0}^{\infty} e^{tx} f_X(x)\,dx \leq e^{-tx_0} \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = e^{-tx_0} \Phi_X(t)$$

Hence this is proved for continuous random variable as well, for $t > 0$:-

$$P(X \geq x) \leq e^{-tx} \Phi_X(t)$$

Proving $P(X \leq x) \leq e^{-tx} \Phi_X(t), t < 0$

When X is a Discrete Random Variable

So any real x lies in $(-\infty, x_1)$ or $[x_n, \infty)$ or $[x_i, x_{i+1})$ for some i.
If $x \in (-\infty, x_1)$, then there exists no x_i such that $X \leq x$, so trivially $P(X \leq x) = 0$ which is definitely lesser than RHS as RHS ≥ 0. Hence this case is trivially proved.
If $x \in [x_i, x_{i+1})$, then $x_k \leq x$ for all $k \in \{1, 2, ..., i\}$, and for $x \in [x_n, \infty)$, $x_k \leq x$ for all $k \in \{1, 2, ..., n\}$, hence :-

$$P(X \leq x) = \sum_{r=1}^{i} P(X = x_r)$$

Also, $e^{t(x_r - x)} \geq 1$ as $t < 0$ and $x_r \leq x$, hence :-

$$P(X \leq x) = \sum_{r=1}^{i} P(X = x_r) \leq \sum_{r=1}^{i} e^{t(x_r - x)} P(X = x_r) = e^{-tx} \sum_{r=1}^{i} e^{tx_r} P(X = x_r)$$

It will be less than or equal to summation from 1 to n as all the terms are non-negative.

$$\Rightarrow P(X \leq x) \leq e^{-tx} \sum_{r=1}^{i} e^{tx_r} P(X = x_r) \leq e^{-tx} \sum_{r=1}^{n} e^{tx_r} P(X = x_r)$$

Comparing with MGF (from (1)), we see that :-

$$P(X \leq x) \leq e^{-tx} \sum_{r=1}^{n} e^{tx_r} P(X = x_r) = e^{-tx} \Phi_X(t) \Rightarrow P(X \leq x) \leq e^{-tx} \Phi_X(t)$$

Hence proved for discrete r.v.

When X is a Continuous Random Variable

Considering $f_X(x)$ is the pdf of X, then :-

$$P(X \leq x_0) = \int_{-\infty}^{x_0} f_X(x)\, dx$$

Again $e^{t(x-x_0)} \geq 1$ as $t < 0$ and $x \leq x_0$ in the above integral, so $f_X(x) \leq e^{t(x-x_0)} f_X(x)$ for all $x \leq x_0$, hence :-

$$P(X \leq x_0) = \int_{-\infty}^{x_0} f_X(x)\, dx \leq \int_{-\infty}^{x_0} e^{t(x-x_0)} f_X(x)\, dx = e^{-tx_0} \int_{-\infty}^{x_0} e^{tx} f_X(x)\, dx$$

Also, since $e^{tx} f_X(x) \geq 0$ for all x, the above will be $\leq$ integral from $-\infty$ to ∞, comparing MGF from (2).

$$P(X \leq x_0) \leq e^{-tx_0} \int_{-\infty}^{x_0} e^{tx} f_X(x)\, dx \leq e^{-tx_0} \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx = e^{-tx_0} \Phi_X(t)$$

Hence this is proved for continuous random variable as well, for $t < 0$:-

$$P(X \leq x) \leq e^{-tx} \Phi_X(t)$$

Now other part of the question...

Given that X is a sum of n independent Bernoulli variables $X_1, X_2, ..., X_n$. Each Bernoulli variable is a discrete r.v, hence its sum = X is also a discrete r.v.
Let's use $P(X \geq x) \leq e^{-tx} \Phi_X(t), t \geq 0$ proved above for discrete r.v.
Put $(1 + \delta)\mu$ as x in this formula, we get:-

$$P(X \geq (1 + \delta)\mu) \leq e^{-t(1+\delta)\mu} \Phi_X(t) \tag{3}$$

We know that for a Bernoulli variable X_i, its MGF is (where $E[X_i] = p_i$):-

$$\Phi_{X_i}(t) = 1 - p_i + p_i e^t$$

Also, as $X = \sum_{i=1}^{n} X_i$ and all X_is are independent, then:-

$$\Phi_X(t) = \prod_{i=1}^{n} \Phi_{X_i}(t) = \prod_{i=1}^{n} (1 - p_i + p_i e^t)$$

Here we used that if $X = X_1 + X_2$ and X_1, X_2 are independent, then $\Phi_X(t) = \Phi_{X_1}(t).\Phi_{X_2}(t)$
Using $1 + x \leq e^x$, put $x = p_i(e^t - 1)$:-

$$1 - p_i + p_i e^t = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$$

$$\Rightarrow \Phi_X(t) = \prod_{i=1}^{n}(1 - p_i + p_i e^t) \leq \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{\sum_{i=1}^{n} p_i(e^x-1)}$$

$$\Rightarrow \Phi_X(t) = e^{(e^t-1)\sum_{i=1}^{n} p_i} = e^{\mu(e^t-1)}$$

Put this in equation (3)...

$$P(X \geq (1+\delta)\mu) \leq e^{-t(1+\delta)\mu}\Phi_X(t) \leq e^{-t(1+\delta)\mu}e^{\mu(e^t-1)} = \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$$

Final Touch...

$$P(X > (1+\delta)\mu) \leq P(X \geq (1+\delta)\mu) \Rightarrow P(X > (1+\delta)\mu) \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$$

Hence formula proved.

Tightening the bound...

Since the above formula is true for all $t \geq 0$, it should be true for that t_0 too for which RHS is minimum. Hence we minimize the RHS:-

$$RHS = g(t) = \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}} = e^{\mu(e^t-1)-(1+\delta)t\mu} = e^{\mu(e^t-(1+\delta)t-1)}$$

Since e^x is an increasing function, minimum value of expression exists at minimum value of power. Differentiating $f(t) = e^t - (1+\delta)t - 1$:-

$$f'(t) = \mu(e^t - (1+\delta)), f''(t) = e^t > 0$$

$$f'(t_0) = 0 \Rightarrow e^{t_0} - (1+\delta) \Rightarrow t_0 = \ln(1+\delta)$$

Hence $f(t_0)$ is minimum or $g(t_0)$ is also minimum as stated above...

$$g(t_0) = g(\ln(1+\delta)) = e^{\mu(1+\delta-(1+\delta)\ln(1+\delta)-1)} = \frac{e^{\mu\delta}}{(1+\delta)^{\mu(1+\delta)}}$$

Finally...

$$P(X > (1+\delta)\mu) \leq \frac{e^{\mu\delta}}{(1+\delta)^{\mu(1+\delta)}}$$

Solution for Q5

Kavya Gupta

September 3, 2023

Instructions to Run the Code

The MATLAB code for this analysis is provided in the file `A2Q5.m` present in main zip. This code generates histograms, empirical cumulative distribution functions (ECDFs), and compares them to Gaussian cumulative distribution functions (CDFs) for different sample sizes `N`. The code also calculates and prints the Mean Absolute Deviation (MAD) values for each `N`.
Important Note: MATLAB Online displays max of 20 images at once, so there maybe some warning at 21st image, please ignore that.

The zip also contains a folder `Q5` which contains all the pictures of histograms and cdfs related with this question along with MAD vs N.

Results

For $N = 5$:

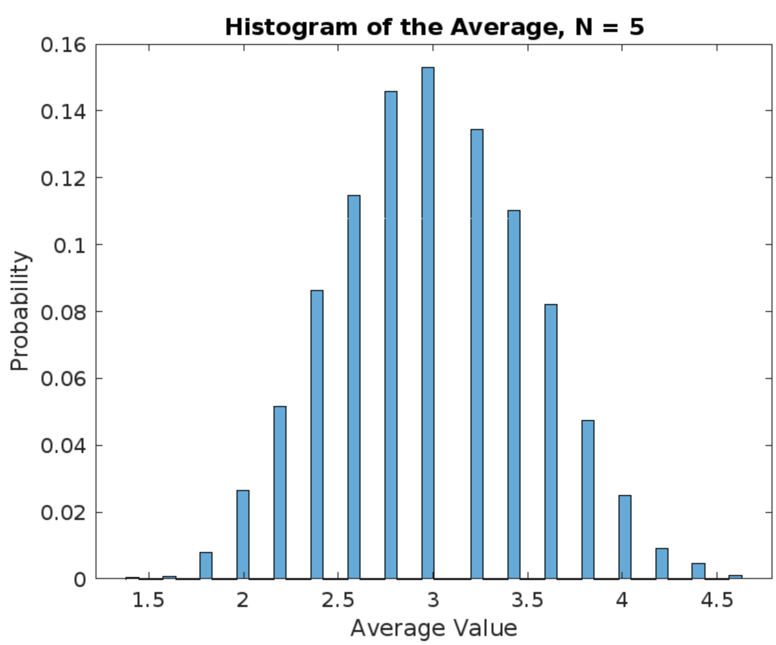

Figure 1: Histogram of the Average

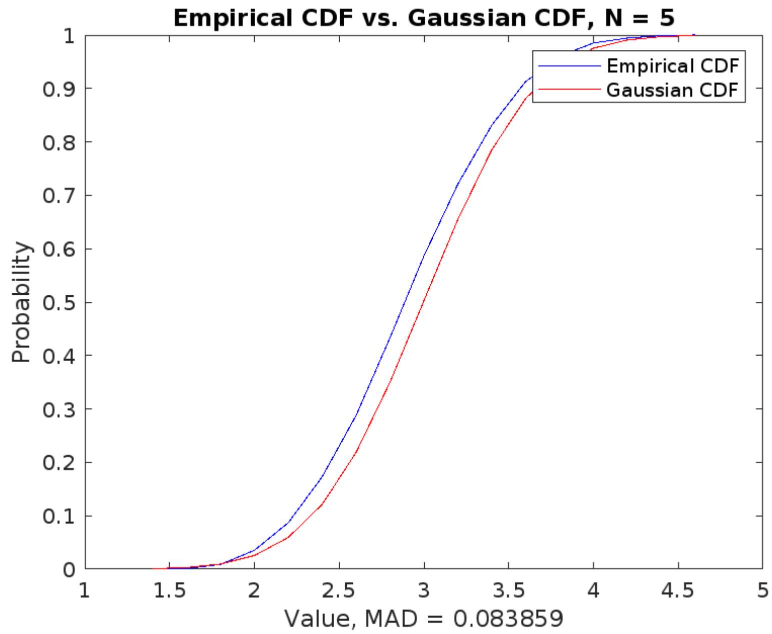

Figure 2: Empirical CDF vs. Gaussian CDF

The MAD for $N = 5$ is 0.083859.

For $N = 10$**:**

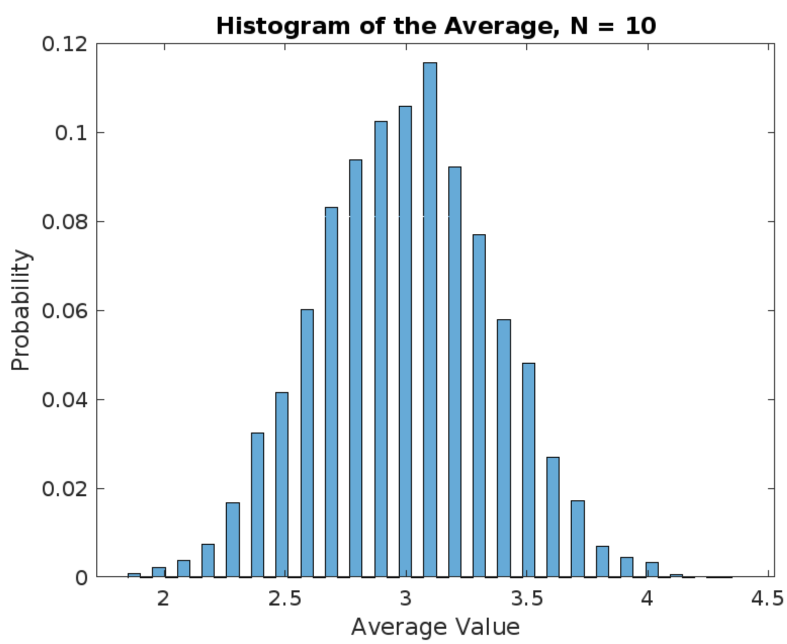

Figure 3: Histogram of the Average

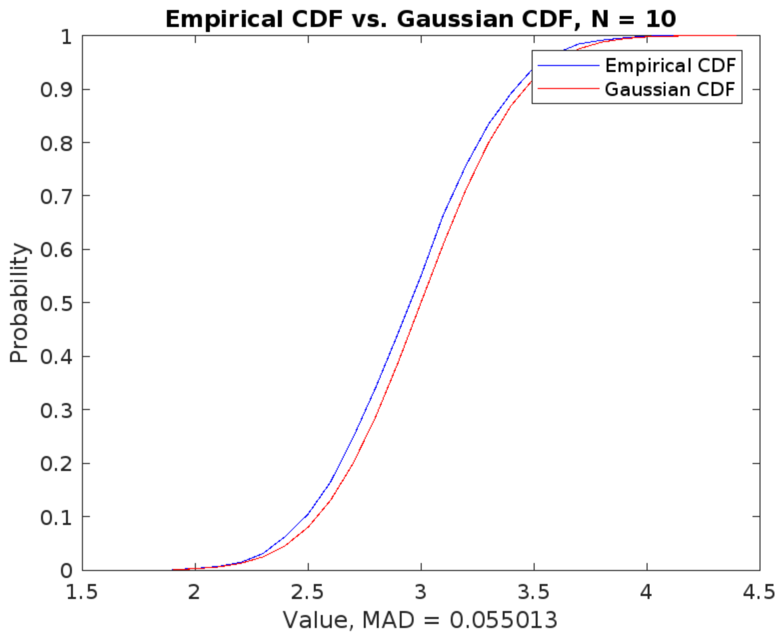

Figure 4: Empirical CDF vs. Gaussian CDF

The MAD for $N = 10$ is 0.055013.

For $N = 20$:

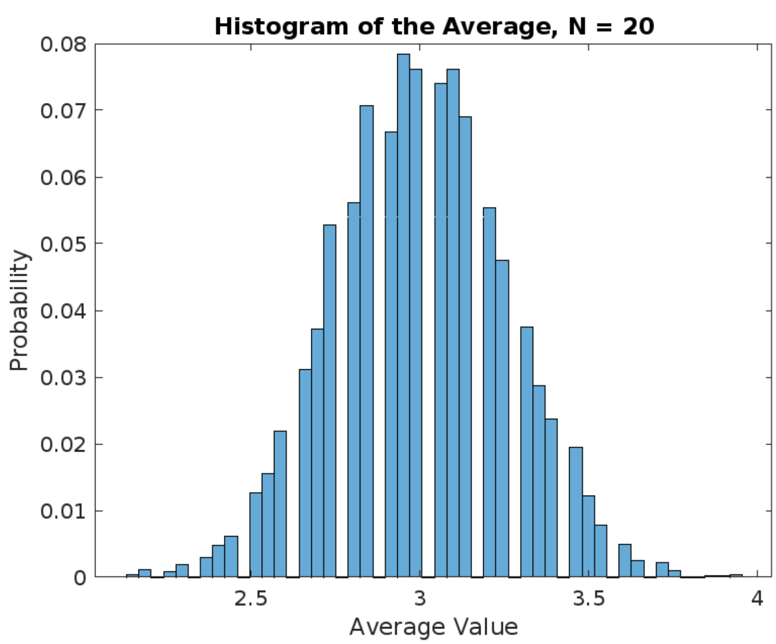

Figure 5: Histogram of the Average

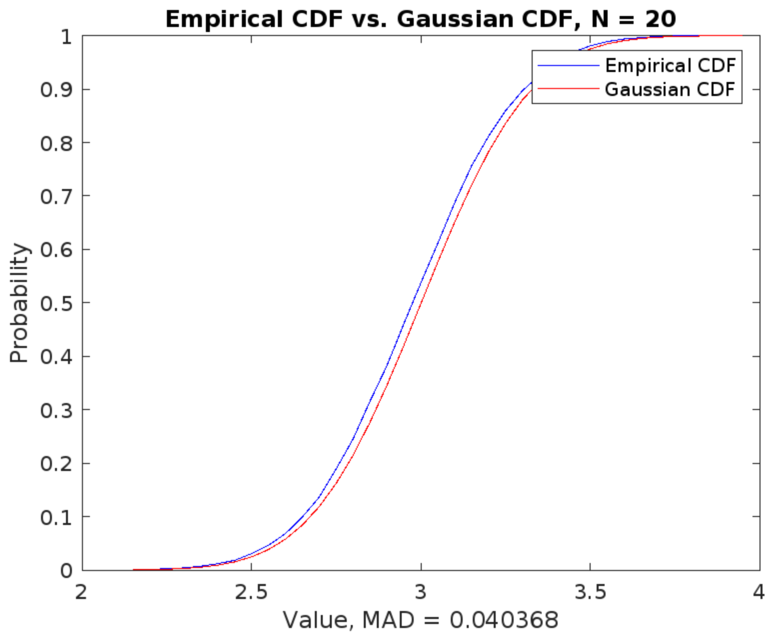

Figure 6: Empirical CDF vs. Gaussian CDF

The MAD for $N = 20$ is 0.040368.

For $N = 50$:

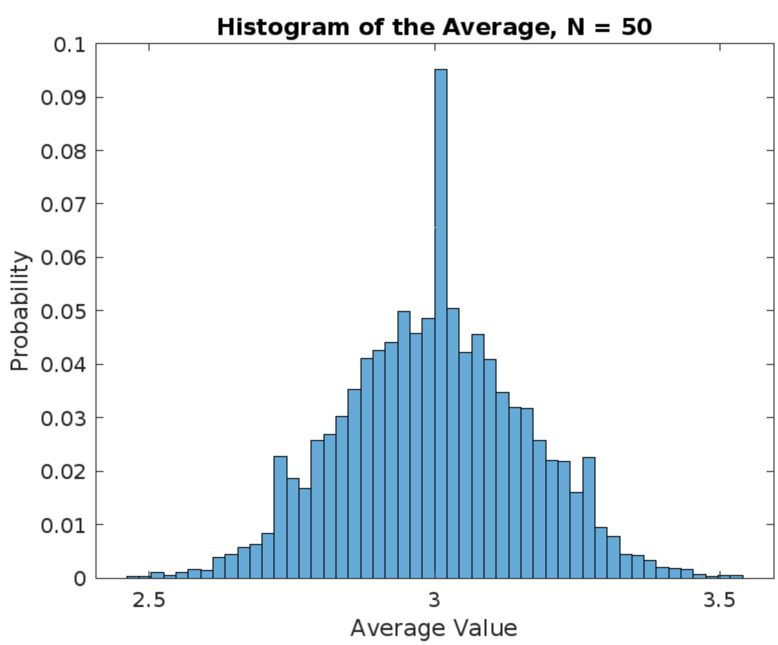

Figure 7: Histogram of the Average

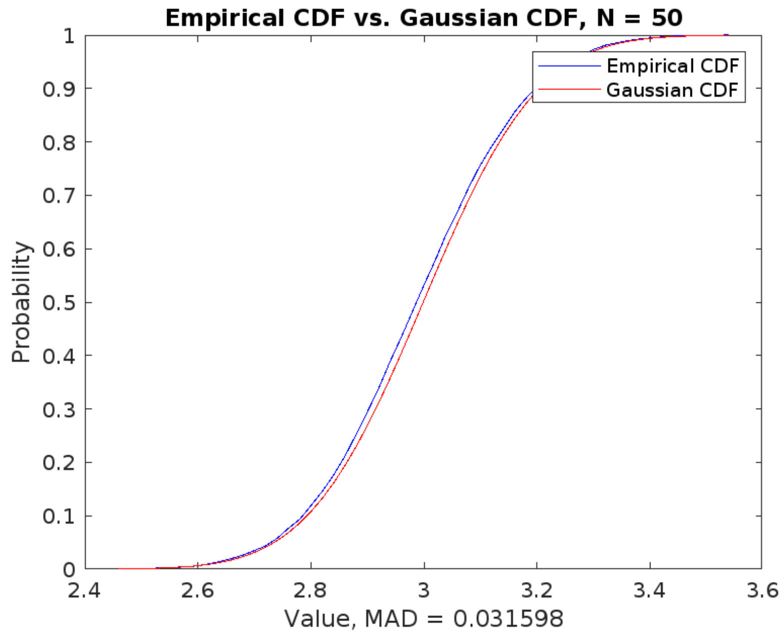

Figure 8: Empirical CDF vs. Gaussian CDF

The MAD for $N = 50$ is 0.031598.

For $N = 100$:

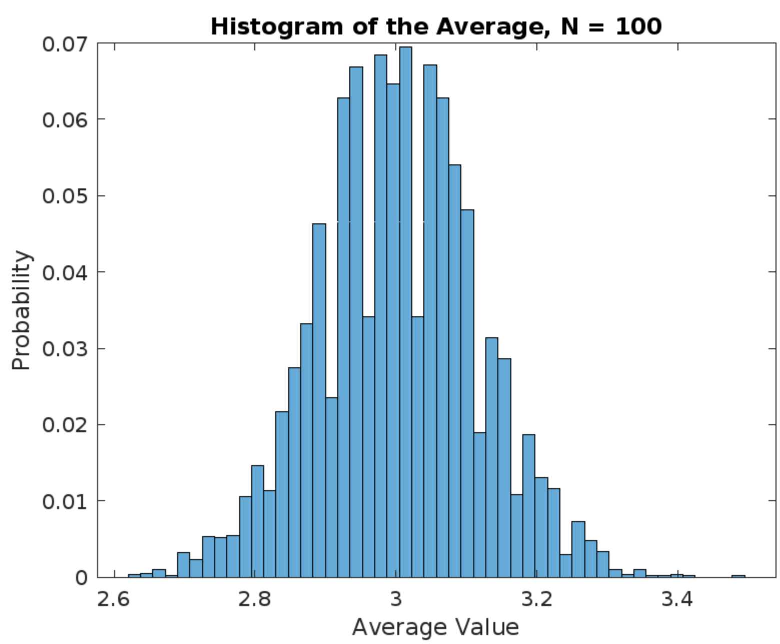

Figure 9: Histogram of the Average

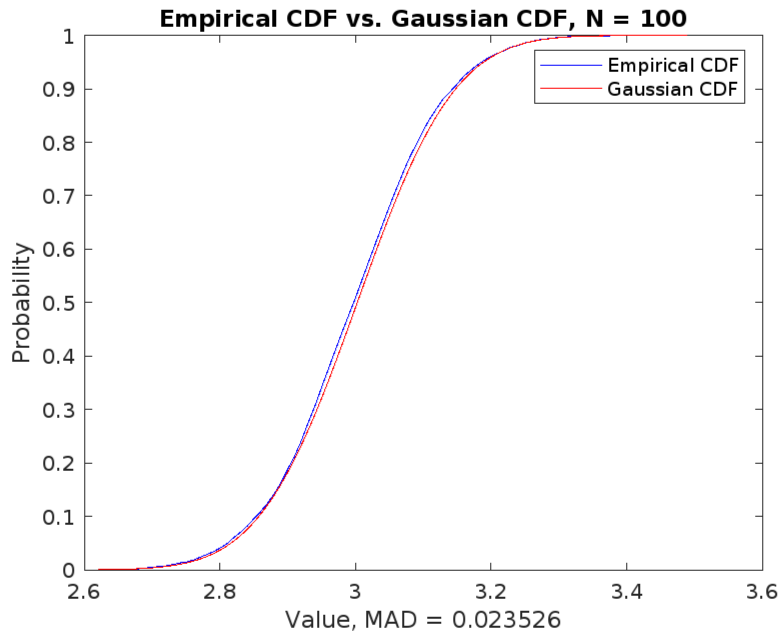

Figure 10: Empirical CDF vs. Gaussian CDF

The MAD for $N = 100$ is 0.023526.

For $N = 200$:

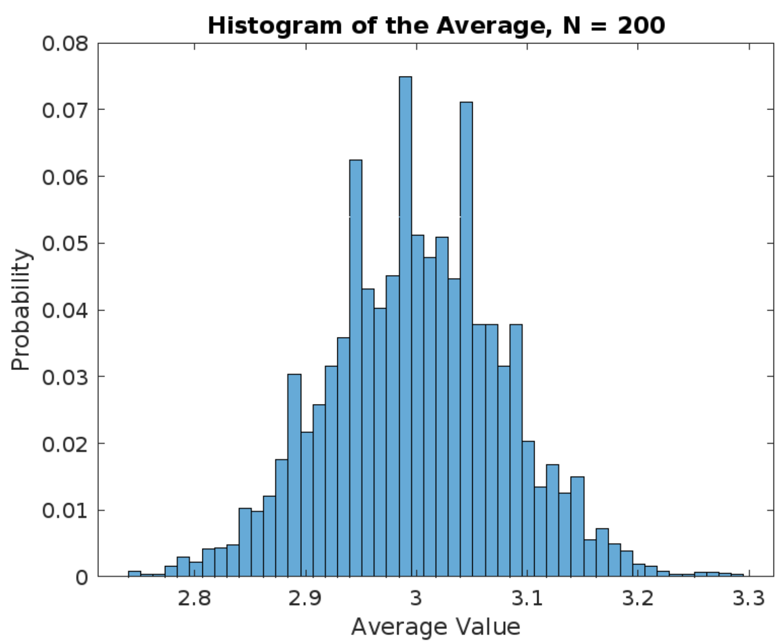

Figure 11: Histogram of the Average

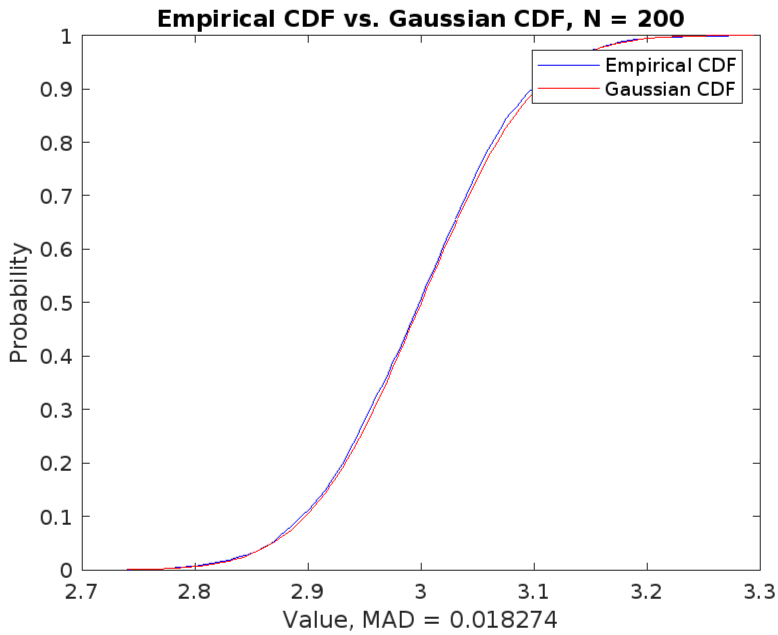

Figure 12: Empirical CDF vs. Gaussian CDF

The MAD for $N = 200$ is 0.018274.

For $N = 500$:

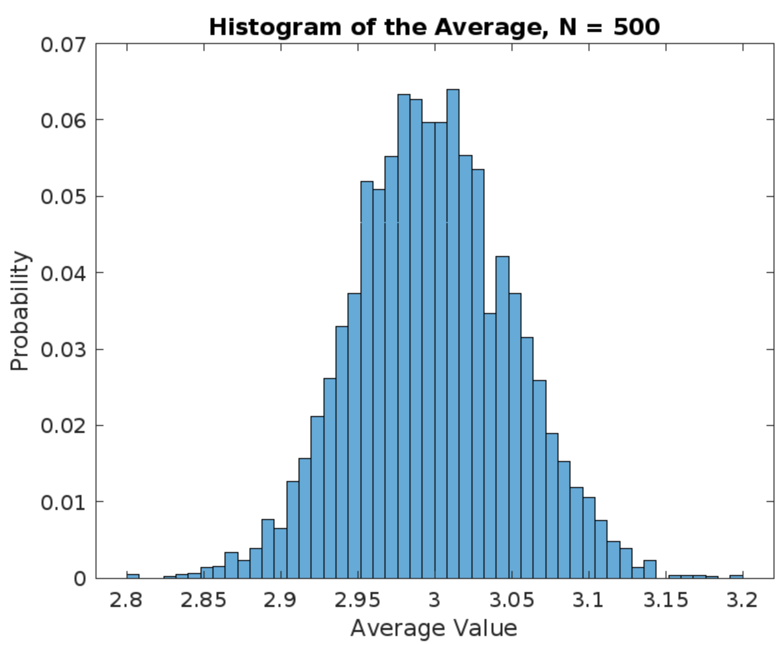

Figure 13: Histogram of the Average

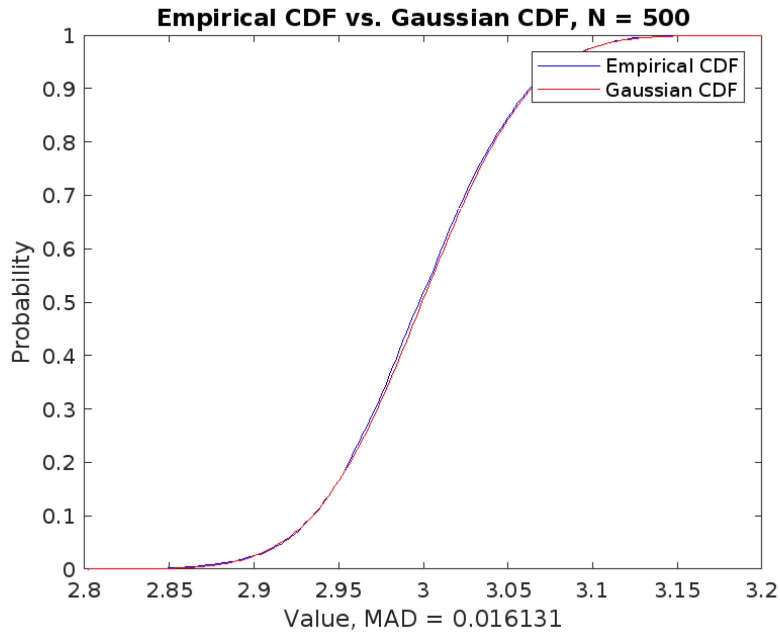

Figure 14: Empirical CDF vs. Gaussian CDF

The MAD for $N = 500$ is 0.016131.

For $N = 1000$**:**

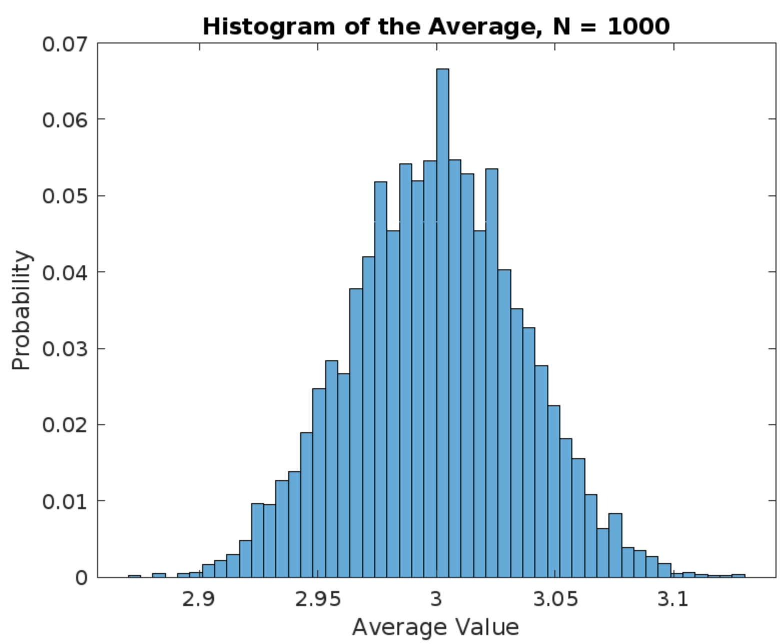

Figure 15: Histogram of the Average

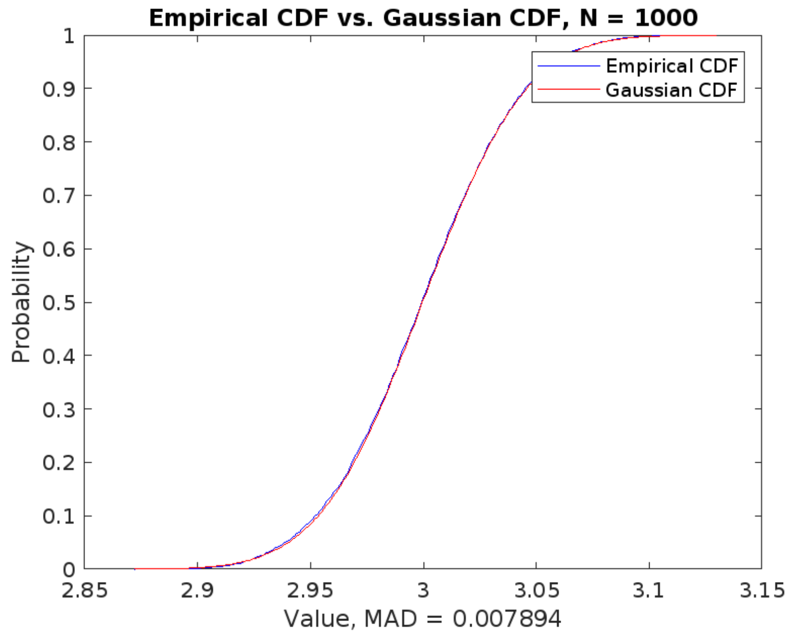

Figure 16: Empirical CDF vs. Gaussian CDF

The MAD for $N = 1000$ is 0.007894.

For $N = 5000$:

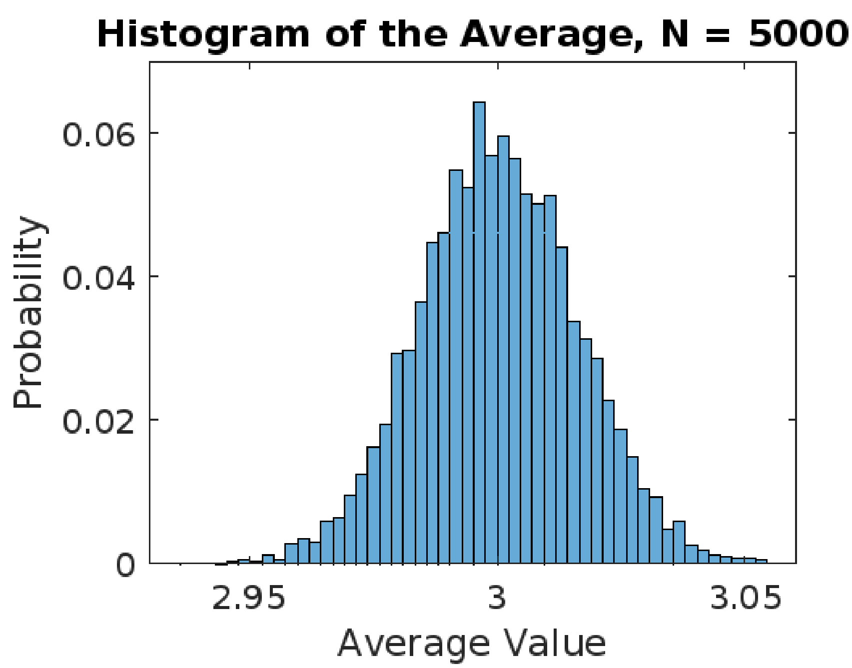

Figure 17: Histogram of the Average

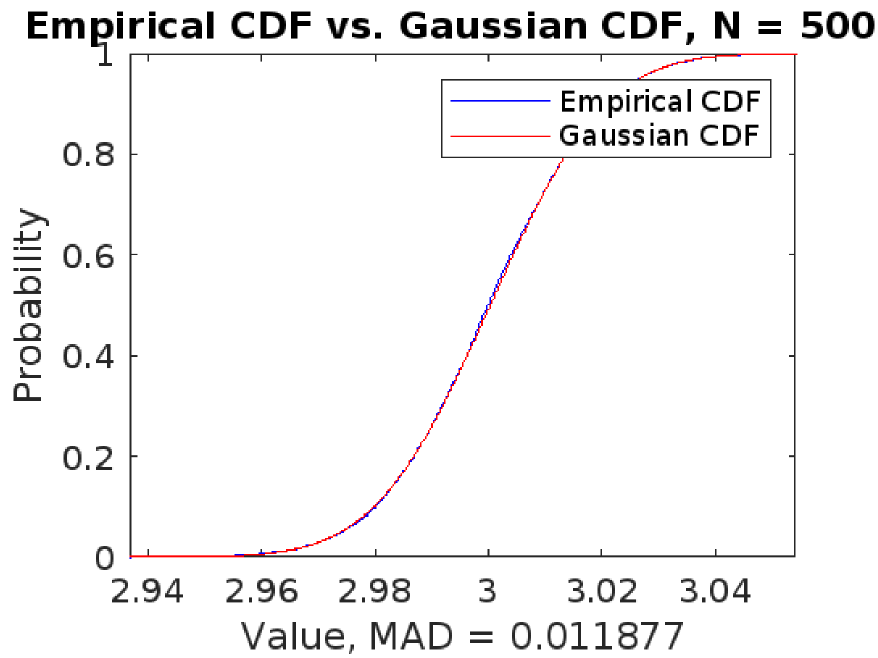

Figure 18: Empirical CDF vs. Gaussian CDF

The MAD for $N = 5000$ is 0.011877.

For $N = 10000$:

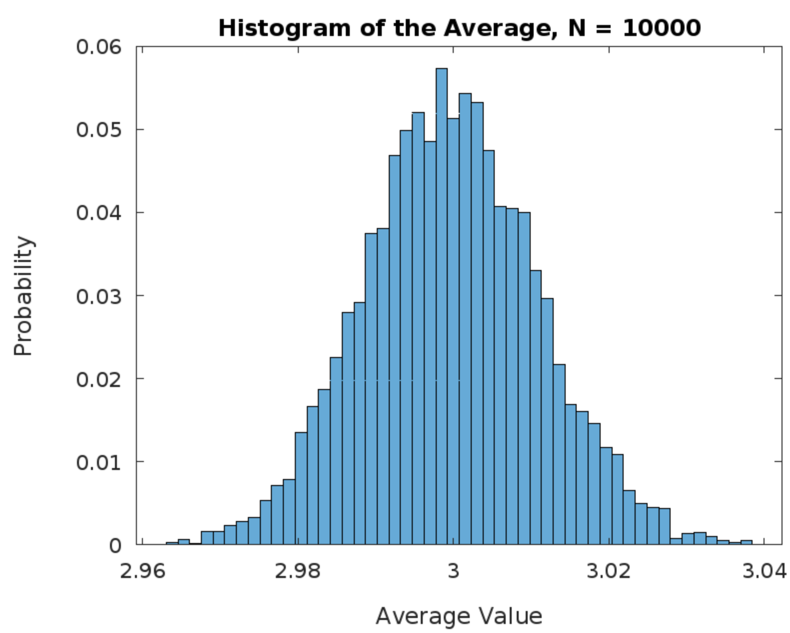

Figure 19: Histogram of the Average

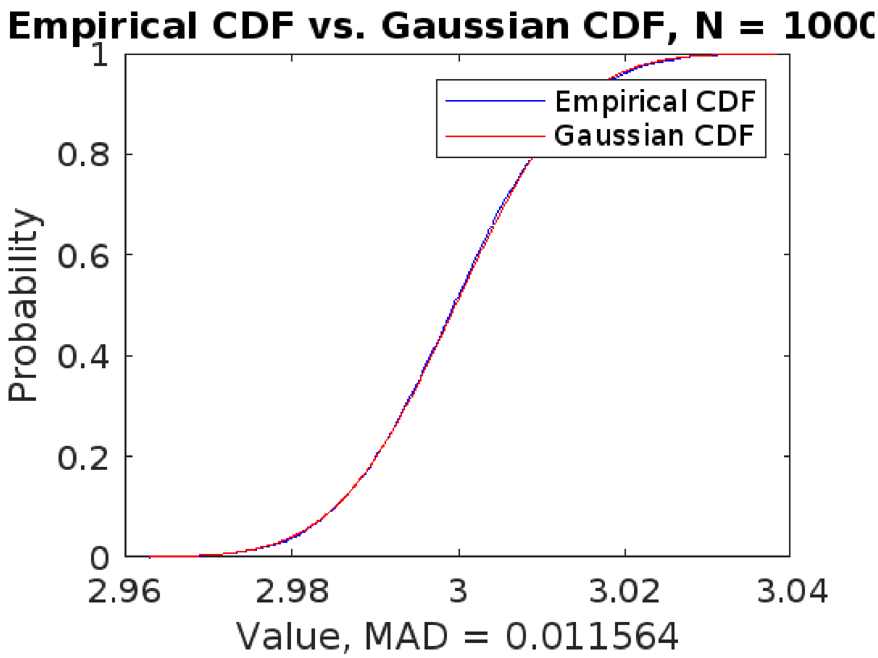

Figure 20: Empirical CDF vs. Gaussian CDF

The MAD for $N = 10000$ is 0.011564.

MAD vs. Sample Size (N)

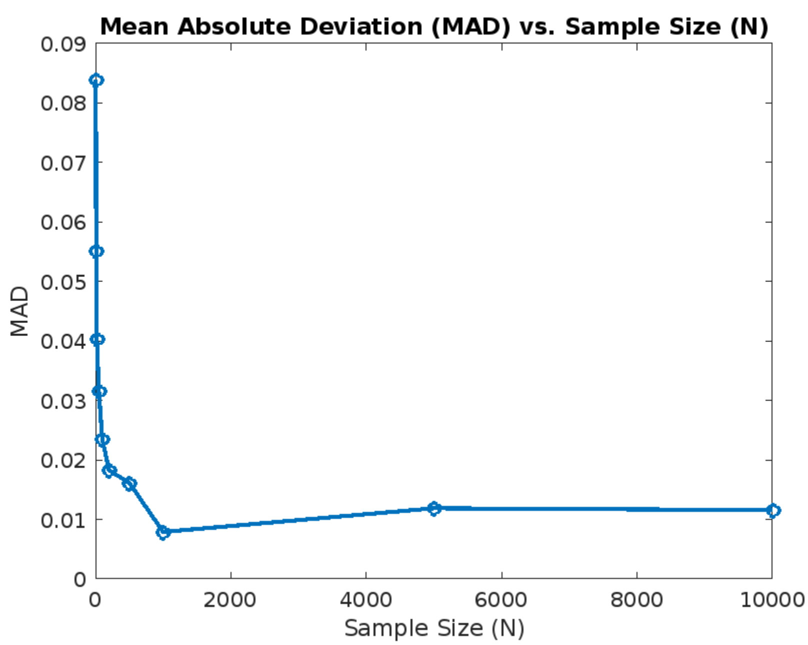

Figure 21: Mean Absolute Deviation (MAD) vs. Sample Size (N)

The plot above shows how the Mean Absolute Deviation (MAD) changes with increasing sample size (N). As N increases, the empirical cumulative distribution function (ECDF) gets closer to the Gaussian cumulative distribution function (CDF), resulting in lower MAD values.
This was an important observation owing to **Central Limit Threorem**.

Solution for Q6 … (By Kavya Gupta)

Instructions to run the code :-

So the code for all the 4 plots is saved in **A2Q6.m** MATLAB file saved in the main .zip file. Running it will create those 4 plots. First two plots will be for Image 2 = T2.jpg, then other two will be for Image 2 = Negative of T1.jpg

Also in the folder **Q6** in the main .zip file you will find the pics of 4 plots {**corr1.png, qmi1.png, corr2.png, qmi2.png**} which are also attached in this pdf.

Analysis of Image 1 vs Image 2 :-

1) **Correlation Coefficient Plot (corr1.png)**

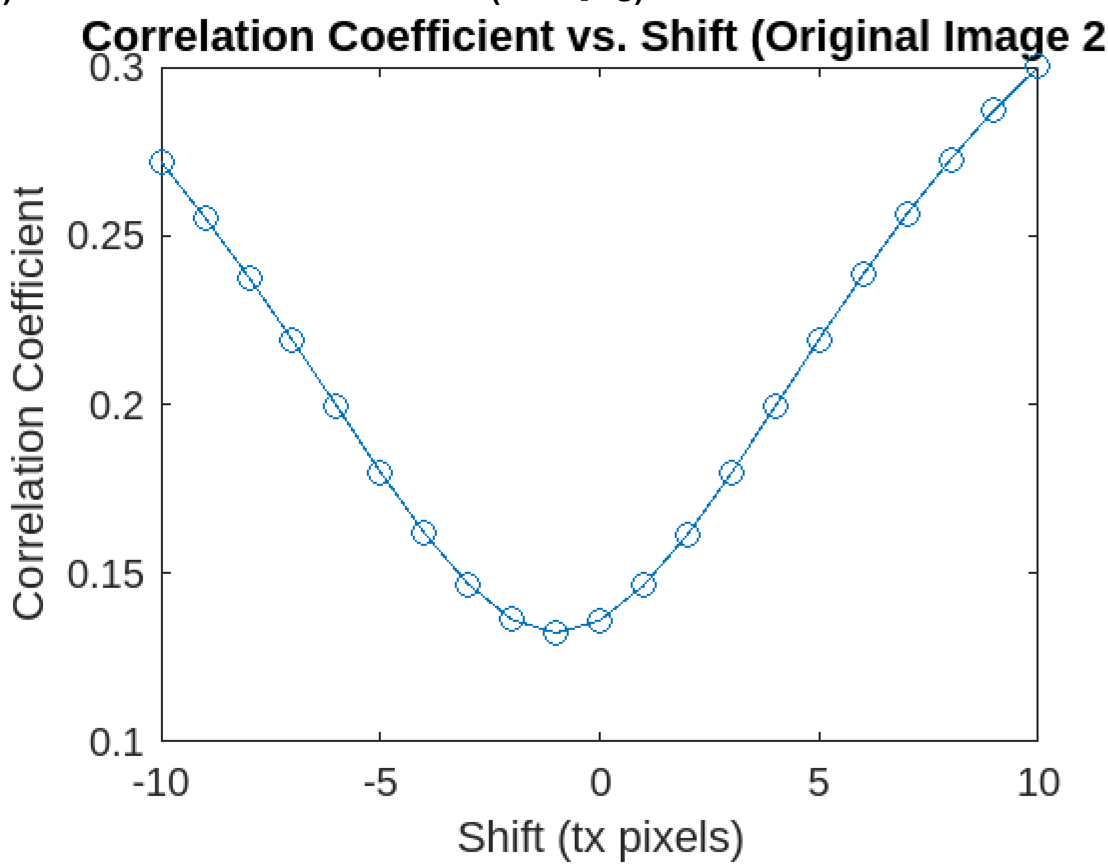

Figure 1 : Correlation Coeff. v/s tx for Original Image 2

Commentary :

It's a V-shaped graph. Surprisingly, at tx close to 0, correlation coefficient (CC) is minimum (around 0.13). It is surprising as one would expect the CC to be maximum at tx=0 because as said in the question, each point (x, y) represents the same entity and is perfectly aligned. If that was so then the correlation coefficient must have been close to 1. But we get to see the almost opposite. As tx deviates from 0, i.e. misalignment increases, then correlation coefficient is increasing 🤯, which sort of indicates that relation between images increases, which is obviously not the case.

2) **Quadratic Mutual Information (QMI) Plot (qmi1.png)**

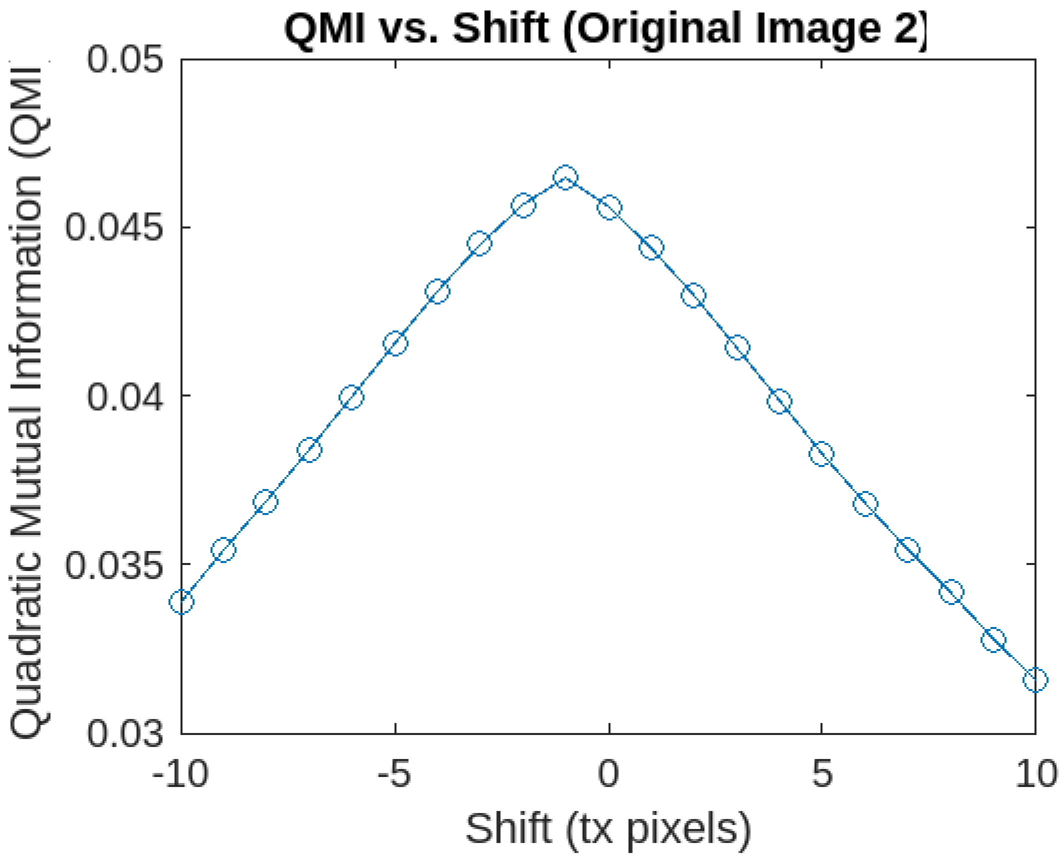

Figure 2 : QMI v/s tx for Original Image 2

Commentary :

To our relief, here the QMI works as we expected. It is max at tx close to 0 and decreases as |tx| increases. As misalignment increases, QMI decreases. Sharp change at tx close to 0.

Analysis of Image 1 vs Negative of Image 1 :-

1) **Correlation Coefficient Plot (corr2.png)**

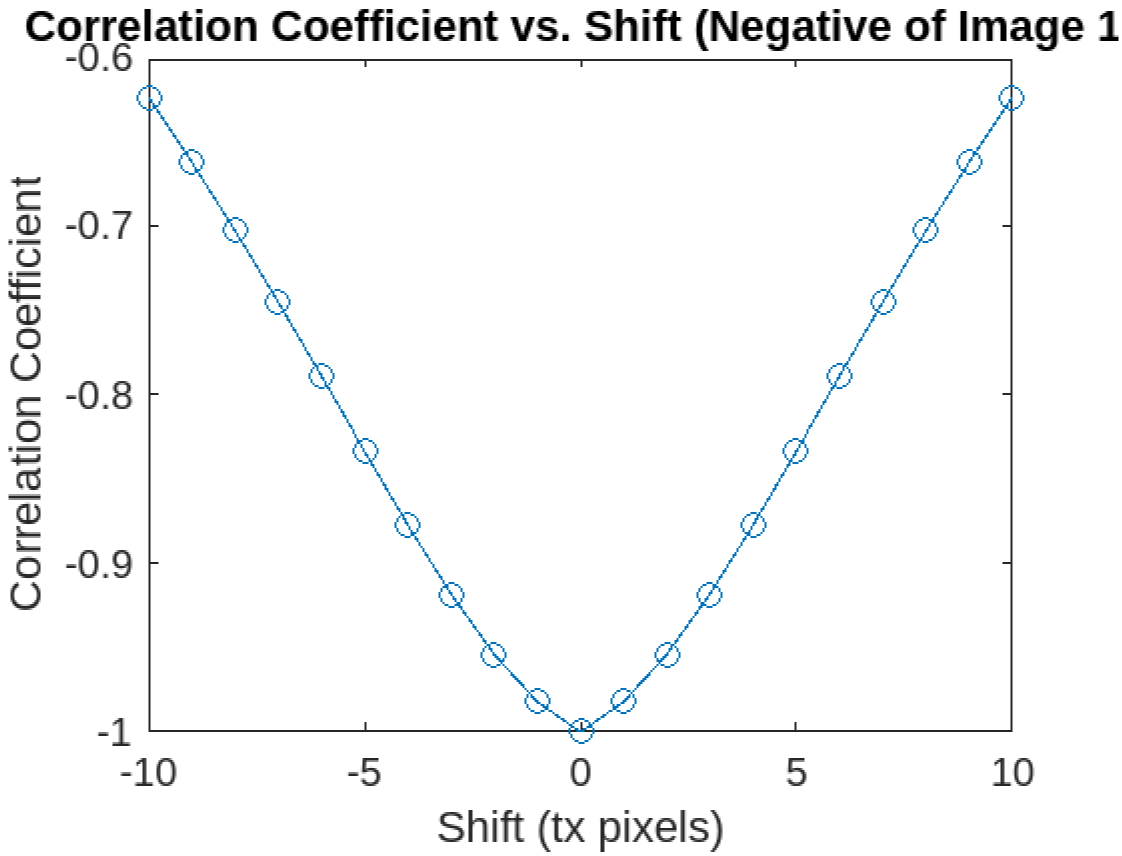

Figure 3 : Correlation Coeff. v/s tx for Negative of Image 1

Commentary :

At tx=0, the CC is -1 exact. Here, it may look like its same as first plot, but see here that all the values are completely negative, hence when I take modulus of this graph, I will get an inverted V-graph, there CC = 1 at tx=0, meaning best relation at no misalignment which is what we expected hence CC works out here well. As misalignment increases, modulus of CC decreases, which indicates lesser relation between images, hence matching our intuition.

2) **Quadratic Mutual Information (QMI) Plot (qmi2.png)**

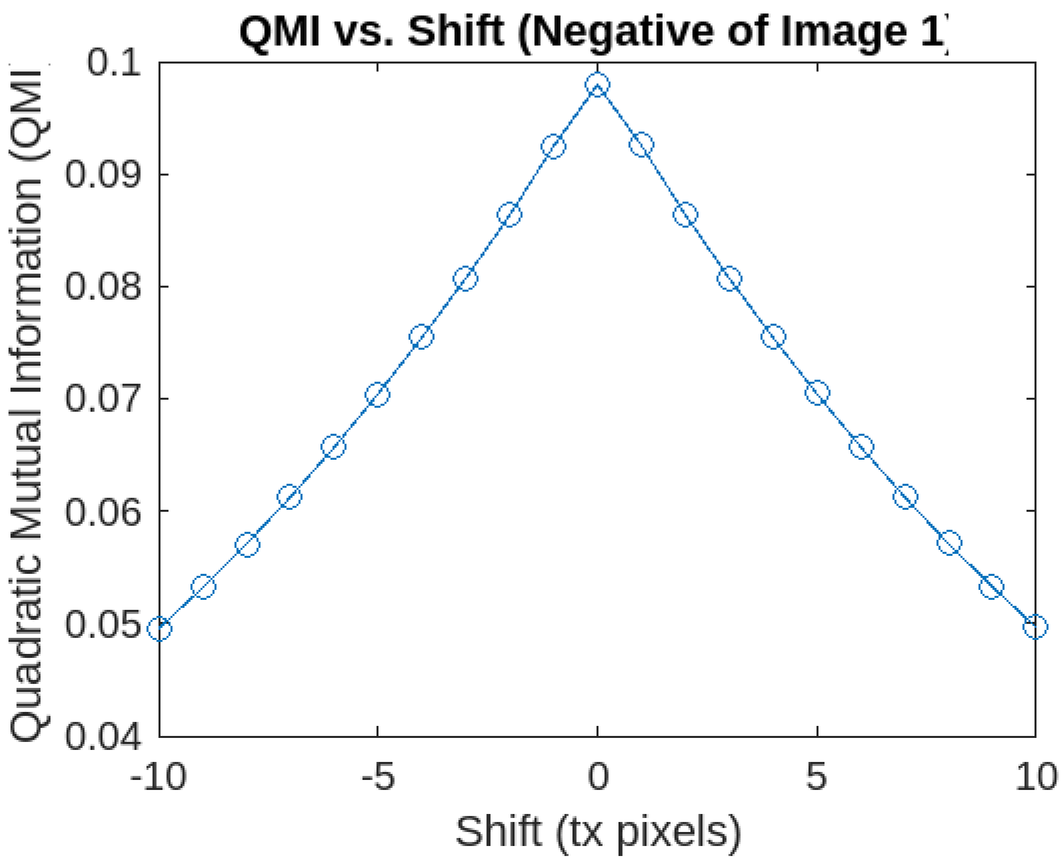

Figure 4 : QMI v/s tx for Negative of Image 1

Commentary :

Same like the first QMI graph, this is an inverted V graph. QMI max at no alignment (tx close to 0) and it decreases as misalignment increases (|tx| > 0) indicating that the images are less related, matching our intuition. In this case specifically, the graph looks nearly symmetrical.

Final Observations :-

So the Correlation Coefficient finds the relation between the images depending on the ***mean*** and ***variance*** of *all the points at the same time* **(DOESN'T MAKE USE OF FULL INFORMATION OF PDF)**.. So in certain cases it could happen that the misalignment causes changes at all the neighboring points, but it may happen the net result of all the misalignments cancels out, and we may see less change in "**mean**" even when images become more and more misaligned OR even that the **mean** starts approaching the mean of the Image 1 like in case of **corr1.png..** That's why we see unexpected graphs.

In case of **corr2.png**, since the second image is exact negative of Image 1, we see that "**mean**" becomes exact negative and all the happenings due to

misalignment happen in one direction, hence correlation coefficient gives a very good result. As |tx| increased the "**mean**" started moving toward the original mean of Image 1, hence CC started becoming less negative.
If we take the modulus of CC, then we see that the two images were maximally correlated for tx close to 0, as we expected !! Most optimal at least misalignment in corr2.png, but not in corr1.png

This happened because here the intensities were "linearly dependent" on each other (I2 = 255 - I1), which directly affects the mean hence the CC was able to give accurate results here.

Hence Correlation Coefficient is highly dependent on the variation of mean and linear relation between variables than actual distortion at each point.

This is where QMI comes to save the day… it is less dependent on mean and rather depends on distortion at small neighborhoods at points.

We expected that there should be a maxima near 0 and QMI meets our expectations in both the graphs, meaning it rightly said that the two images were maximally correlated for tx close to 0.

This is because QMI uses the **Joint Histogram Bin method,** nearby neighbors of a point will most probability will fall into the same point as the point itself. So if any misalignment happens, the new neighbors of a point in the new misaligned image start going in different bins and hence the QMI decreases.
*It does not depend on "**mean**" at all unlike CC and gives the same nature of graph everytime* **qmi1.png, qmi2.png***, unlike CC.*
Also QMI takes square of the difference between joint pmf and marginal pmfs, hence there is no issue of **"Net Result = 0** due to canceling each other" as in CC.

It catches the linear as well as the non-linear relation between the intensities and hence gives better results….

Hence QMI is a much superior method to detect misalignments than Correlation Coefficient !!

22B1053 - 22B1205 - 22B1025

Q.7 let X be multinomial distribution random varible, with paramaters $(P_1, P_2, \dots, P_m, n)$.

let $X = (X_1, X_2, X_3, \dots, X_m)$

where X_i represent the no. of trials that produced the i^{th} outcome, which are also random varidbles.

let C be covariance matrix of X, where C_{ij} is $Cov(X_i, X_j)$.

now by def,

$$Cov(X_i, X_j) = \cancel{Cov(X_i} E[(X_i - \mu_i)(X_j - \mu_j)]$$ (where μ_i, μ_j are $E(X_i)$ & $E(X_j)$ respectively)

$$= E[X_i X_j - \mu_i X_j - X_i \mu_j + \mu_i \mu_j]$$

$$= E[X_i X_j] - E[\mu_i X_j] - E[X_i \mu_j] + E[\mu_i \mu_j]$$

$$= E[X_i X_j] - \mu_i E[X_j] - E[X_i]\mu_j + \mu_i \mu_j$$

$$= E[X_i X_j] - \mu_i \mu_j - \mu_i \mu_j + \mu_i \mu_j$$

$$= E[X_i X_j] - E[X_i]E[X_j] \quad \text{— } 4$$

Now to get $E[X_i]$'s and $E[X_i X_j]$'s we use MGF.

First we drive the MGF for X.

$$\phi_X(t) = \phi_X(t_1, t_2, \dots, t_m)$$

$$= \sum_{k_1 + k_2 + k_3 \dots + k_m} \frac{n!}{k_1! k_2! k_3! \dots k_m!} P_1^{k_1} P_2^{k_2} \dots P_m^{k_m} e^{t_1 k_1} e^{t_2 k_2} \dots e^{t_m k_m}$$

$$= \sum_{k_1 + k_2 \dots + k_m} \frac{n!}{k_1! k_2! \dots k_m!} (P_1 e^{t_1})^{k_1} (P_2 e^{t_2})^{k_2} \dots (P_m e^{t_m})^{k_m}$$

now by multinomial theorem,

$$\phi_X(t) = (P_1 e^{t_1} + P_2 e^{t_2} \dots + P_m e^{t_m})^n$$

Now as $\phi_X(t) = E(e^{tX}) = E(e^{\sum t_i X_i})$

$$\therefore \frac{\partial \phi_X(t)}{\partial t_i} = E[X_i e^{\sum t_i x_i}]$$

$$\therefore \left.\frac{\partial \phi_X(t)}{\partial t_i}\right|_{t=0} = E[X_i] \quad \text{— } 1$$

we also have,

for $i \neq j$,

$$\frac{\partial^2 \phi_X(t)}{\partial t_i \partial t_j} = \frac{\partial \left(E[x_i e^{\sum t_i x_i}]\right)}{\partial t_j} = E[(x_i)(x_j) e^{\sum t_i x_i}]$$

$$\therefore \left.\frac{\partial^2 \phi_X(t)}{\partial t_i \partial t_j}\right|_{t=0} = E[X_i X_j] \quad \text{— } 2$$

and

$$\frac{\partial^2 \phi_x(t)}{\partial^2 t_i} = \frac{\partial(E[x_i e^{\Sigma t_i x_i}])}{\partial t_i} = E[x_i^2 e^{\Sigma t_i x_i}]$$

$$\left.\frac{\partial^2 \phi_x(t)}{\partial^2 t_i}\right|_{t=0} = E[x_i^2] \quad — \quad (3)$$

using (1), (2), (3) we find C_{ij} using (4).

for ~~C_{ij}~~ we first find $\frac{\partial^2 \phi_x(t)}{\partial t_i \partial t_j}$, $\frac{\partial \phi_x(t)}{\partial t_i}$

$$\frac{\partial \phi_x(t)}{\partial t_i} = \frac{\partial(\Sigma P_j e^{t_j})^n}{\partial t_i} = n(\Sigma P_j e^{t_j})^{n-1} P_i e^{t_i}$$

for $i \neq j$,

$$\frac{\partial^2 \phi_x(t)}{\partial t_i t_j} = \frac{\partial(nP_i e^{t_i}(\Sigma P_k e^{t_k})^{n-1})}{\partial t_j} = nP_i e^{t_i}(n-1)(\Sigma P_k e^{t_k})^{n-2} P_j$$

$$\frac{\partial^2 \phi_x(t)}{\partial^2 t_i} = \frac{\partial(nP_i e^{t_i}(\Sigma P_k e^{t_k})^{n-1})}{\partial t_i} = nP_i e^{t_i}(\Sigma P_k e^{t_k})^{n-1} + n(n-1)P_i^2 e^{2t_i}(\Sigma P_k e^{t_k})^{n-2}$$

$$\therefore E(X_i) = \left.\frac{\partial \phi_x(t)}{\partial t_i}\right|_{t=0} = nP_i$$

for $i \neq j$, $E(X_i X_j) = \left.\frac{\partial^2 \phi_x(t)}{\partial t_i \partial t_j}\right|_{t=0} = n(n-1)P_i P_j$

$$E(X_i^2) = \left.\frac{\partial^2 \phi_x(t)}{\partial^2 t_i}\right|_{t=0} = nP_i + n(n-1)P_i^2$$

$\therefore$ for $i \neq j$,

$$C_{ij} = E(X_i X_j) - E(X_i)E(X_j)$$
$$= n(n-1)P_i P_j - n^2 P_i P_j$$
$$C_{ij} = P_i P_j(n^2 - n - n^2) = -nP_i P_j$$

$\because$ C_{ij} is symmetry $(P_i P_j = P_j P_i)$

$\therefore C_{ij} = C_{ji}$.

$$C_{ii} = E(X_i^2) - E^2(X_i)$$
$$= nP_i + (n^2 - n)P_i^2 - n^2 P_i^2$$
$$C_{ii} = nP_i(1 - P_i)$$

$\therefore$ The final covariance matrix $C = \begin{cases} C_{ij} = -nP_i P_j & \text{for } i \neq j \\ C_{ii} = nP_i(1-P_i) \end{cases}$