CS 215 - Data Analysis and Interpretation

Assignment 2

Saksham Rathi - 22B1003
Tanmay Gejapati - 22B0969
Aditya Singh - 22B1844

Contents

Question 1

Given: $X_1, X_2, ..., X_n$ $n > 0$, are independent identically distributed random variables, each having cdf $F_X(x)$ and pdf $f_X(x) = F'_X(x)$.
We need to find the cdf and pdf for $Y_1 = max(X_1, X_2, ..., X_n)$ and $Y_2 = min(X_1, X_2, ..., X_n)$.
We use the definition of cdf $F_X(x) = P(X \leq x)$ and pdf $f_X(x) = F'_X(x)$.

cdf and pdf of Y_1:

cdf $= P(max(X_1, X_2, ..., X_n) \leq x)$. For this to happen each $X_i \leq x$.
Thus,

$$cdf(Y_1) = F_{Y_1}(x) = P(\bigcap_{i=1}^{n}(X_i \leq x)) = \prod_{i=1}^{n} P(X_i \leq x) = (F_X(x))^n$$

because all X_i's are independent. Now we get,

$$pdf(Y_1) = f_{Y_1}(x) = F'_{Y_1}(x) = n(F_X(x))^{n-1}F'_X(x)$$

cdf and pdf of Y_2:

cdf $= P(min(X_1, X_2, ..., X_n) \leq x)$. This can be seen as the complement of the case when each $X_i > x$ $\forall$ i. Here we will use the fact that $P(X > x) = 1 - F_X(x)$.
Thus,

$$cdf(Y_2) = F_{Y_2}(x) = 1 - P(\bigcap_{i=1}^{n}(X_i > x)) = 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - (1 - F_X(x))^n$$

and,

$$pdf(Y_2) = f_{Y_2}(x) = F'_{Y_2}(x) = n(1 - F_X(x))^{n-1}F'_X(x)$$

Question 2

Case 1:

Given: X is a random variable which belongs to a Gaussian Mixture Model, i.e. X $\sim \sum_{i=1}^{K} p_i N(\mu_i, \sigma_i^2)$.

To find the various properties of X we need to find its pdf.

$$f_X(x) = \sum_{i=1}^{K} p_i \cdot \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right)$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx = \sum_{i=1}^{K} p_i \cdot \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} x \cdot \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) dx$$

Here the expression multiplied to p_i i.e. $\frac{1}{\sqrt{2\pi\sigma_i^2}}\int_{-\infty}^{\infty} x \cdot \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) dx$ evaluates to μ_i from lecture slides, thus,

$$E(X) = \sum_{i=1}^{K} p_i \cdot \mu_i$$

$Var(X) = E(X^2) - (E(X))^2$. We need the value of $E(X^2)$. By using LOTUS, we know that,

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \Rightarrow E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$

$$E(X^2) = \sum_{i=1}^{K} p_i \cdot \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} x^2 \cdot \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) dx$$

From lecture slides we know that on calculating the integral we obtain the following,

$$E(X^2) = \sum_{i=1}^{K} p_i \cdot (\sigma_i^2 + \mu_i^2)$$

$$Var(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^{K} p_i \cdot (\sigma_i^2 + \mu_i^2) - \left(\sum_{i=1}^{K} p_i \cdot \mu_i\right)^2$$

For MGF,

$$\Phi_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)\, dx = \sum_{i=1}^{K} p_i \cdot \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} e^{\left(tx - \frac{(x-\mu_i)^2}{2\sigma_i^2}\right)} dx$$

We observe that the expression multiplied to p_i is the MGF of X_i. Thus, we get,

$$\Phi_X(t) = \sum_{i=1}^{K} p_i \cdot \Phi_{X_i}(t)$$

Case 2:

Now $X_i = N(\mu_i, \sigma_i^2)$ are independent random variables. Z = $\sum_{i=1}^{K} p_i X_i = \sum_{i=1}^{K} p_i N(\mu_i, \sigma_i^2)$.

$$E(Z) = E\left(\sum_{i=1}^{K} p_i N(\mu_i, \sigma_i^2)\right) = \sum_{i=1}^{K} p_i E(N(\mu_i, \sigma_i^2)) = \sum_{i=1}^{K} p_i \mu_i$$

To find the variance, we use the fact that X_i are independent random variables, so Var(X + Y) = Var(X) + Var(Y). We also use the fact that $Var(aX) = a^2 V(X)$.

$$Var(Z) = Var\left(\sum_{i=1}^{K} p_i N(\mu_i, \sigma_i^2)\right) = \sum_{i=1}^{K} Var(p_i N(\mu_i, \sigma_i^2)) = \sum_{i=1}^{K} p_i^2 Var(N(\mu_i, \sigma_i^2)) = \sum_{i=1}^{K} p_i^2 \sigma_i^2$$

We can find the pdf using E(Z) and Var(Z). We know that the linear combination of Gaussian random variables is a Gaussian random variable. Thus Z is a Gaussian random variable with mean = E(Z), and variance = Var(Z) as calculated above, i.e. Z = $N(\sum_{i=1}^{K} p_i\mu_i, \sum_{i=1}^{K} p_i^2\sigma_i^2)$. Thus, the pdf of Z is of the form of a Gaussian distribution,

$$f_Z(x) = \frac{1}{\sqrt{\sum_{i=1}^{K} p_i^2\sigma_i^2}\sqrt{2\pi}} \cdot \exp\left(-\frac{(x - \sum_{i=1}^{K} p_i\mu_i)^2}{2\sum_{i=1}^{K} p_i^2\sigma_i^2}\right)$$

To find the MGF $\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$ which is true for independent random variables and for Gaussian distribution, $\Phi_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$. We also use the property of MGFs that $\Phi_{aX}(t) = \Phi_X(at)$ So,

$$\Phi_Z(t) = \Phi_{\left(\sum_{i=1}^{K} p_i N(\mu_i, \sigma_i^2)\right)}(t) = \prod_{i=1}^{K} \Phi_{p_i N(\mu_i,\sigma_i^2)}(t) = \prod_{i=1}^{K} \Phi_{N(\mu_i,\sigma_i^2)}(p_i t) = \prod_{i=1}^{K} \exp(\mu_i p_i t + \frac{\sigma_i^2 p_i^2 t_i^2}{2})$$

Question 3

Given: Random variable X with mean μ and variance σ^2. Let's define another random variable Y = X - μ.

$$E(Y) = E(X - \mu) = E(X) - \mu = \mu - \mu = 0$$
$$Var(Y) = E((Y - 0)^2) = E(Y^2) = E((X - \mu)^2) = Var(X) = \sigma^2$$

So mean of Y = 0, and variance of Y = σ^2. Let $n \geq 0$, then we have,

$$P(X - \mu \geq \tau) = P(Y \geq \tau) = P(Y + n \geq \tau + n) \leq P((Y + n)^2 \geq (\tau + n)^2),$$

when we square the expression then an extra inequality gets included which is: $P(Y+n \leq -(\tau+n))$ and hence we put the $\leq$ sign. As $(Y+n)^2$ is a non-negative random variable we can apply Markov's Inequality,

$$P(X - \mu \geq \tau) \leq \frac{E((Y + n)^2)}{(\tau + n)^2} = \frac{E(Y^2) + E(2Yn) + E(n^2)}{(\tau + n)^2} = \frac{\sigma^2 + n^2}{(\tau + n)^2}$$

where in the last line we used the fact that E(2Yn) = 2nE(Y) = 0, and that var(Y) = $E(Y^2) - (E(Y))^2 \Rightarrow E(Y^2) = \sigma^2$. Now for the value of n which minimizes this expression. We can find it using differentiation.

$$\frac{d}{dn}(\frac{\sigma^2 + n^2}{(\tau + n)^2}) = \frac{2n(\tau + n)^2 - 2(\sigma^2 + n^2)(\tau + n)}{(\tau + 4)^4} = 0,$$

this gives $n = \frac{\sigma^2}{\tau}$. Putting this in the expression we get the result which we had to prove for $\tau > 0$,

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

If $\tau < 0$, then we can slightly modify our previous argument. We take $\omega = -\tau$.

$$P(X - \mu \leq \tau) = P(Y \leq \tau) = P(Y \leq -\omega) = P(-Y \geq \omega)$$

In the above expressions, $Y \leq \tau$, i.e. Y is a negative random variable, thus $-Y$ is a non-negative random variable and w is some positive number. Thus, we have converted the case of $\tau < 0$ to the case where $\tau > 0$. So we can use the result obtained above,

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

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

Hence we have proved both the cases of One-Sided Chebyshev's Inequality using Markov's Inequality.

Question 4

The first part of the question asks to prove the bounds:

$$P(X \geq x) \leq e^{-tx}\phi_X(t), t > 0 \tag{1}$$

$$P(X \leq x) \leq e^{-tx}\phi_X(t), t < 0 \tag{2}$$

Case 1: $t > 0$

Markov's inequality states:

$$P(Y \geq a) \leq \frac{E[Y]}{a}, \forall a$$

Consider the random variable $Y = e^{tX}$ and $a = e^{tx}$, then $P(X \geq x)$ is equivalent to $P(Y \geq e^{tx})$ since $t > 0$. Using Markov's inequality, we can say that

$$P(Y \geq e^{tx}) \leq e^{-tx}E(e^{tX})$$

By definition, $\phi_X(t) = E[e^{tX}]$, plugging this into the above equation results in:

$$P(Y \geq e^{tx}) \leq e^{-tx}\phi_X(t)$$

But $Y = e^{tX}$ hence we get:

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

which is the desired result.

Case 2: $t < 0$

We prove this case in a similar way as the previous case. Here we have $X \leq x$ and we consider $Y = e^{tX}, a = e^{tx}$, which for $t < 0$ is equivalent to $Y \geq a$ and so are $P(X \leq x)$ and $P(Y \geq e^{tx})$ equivalent. Using this result in the equation:

$$P(Y \geq e^{tx}) \leq e^{-tx}\phi_X(t)$$

yields us the desired result:

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

Moving on to the next part of the question, requires us to prove:

$$P(X > (1+\delta)\mu) \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}, \forall t \geq 0, \delta > 0 \tag{3}$$

As $t \geq 0$ we can make use of the result we derived in the previous part, i.e. equation (1), by substituting $x = (1+\delta)\mu$ we get:

$$P(X > (1+\delta)\mu) \leq \frac{\phi_X(t)}{e^{(1+\delta)t\mu}} \tag{4}$$

Given that $X = X_1 + X_2 + \cdots + X_n$, using the properties of moment generating functions, as X_i are independent random variables, we can write

$$\phi_X(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdots \phi_{X_n}(t)$$

Let U be $Bernoulli(p)$, the MGF is

$$\phi_U(t) = 1 - p + p \cdot e^t$$
$$\phi_{X_i}(t) = 1 - p_i + p_i \cdot e^t$$
$$\phi_X(t) = \prod_{i=1}^{n}(1 - p_i + p_i \cdot e^t)$$

Using the inequality: $1 + x \leq e^x$, $1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$

$$\prod_{i=1}^{n}(1 - p_i + p_i \cdot e^t) \leq \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{(p_1+\cdots+p_n)(e^t-1)} = e^{\mu(e^t-1)}$$

Substituting this into the equation (4) gives us the desired result:

$$P(X > (1+\delta)\mu) \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}, \forall t \geq 0, \delta > 0$$

To have the tightest bound, we must minimize the right hand side of the inequality, we do that by choosing t such that the RHS's derivative with respect to t vanishes.

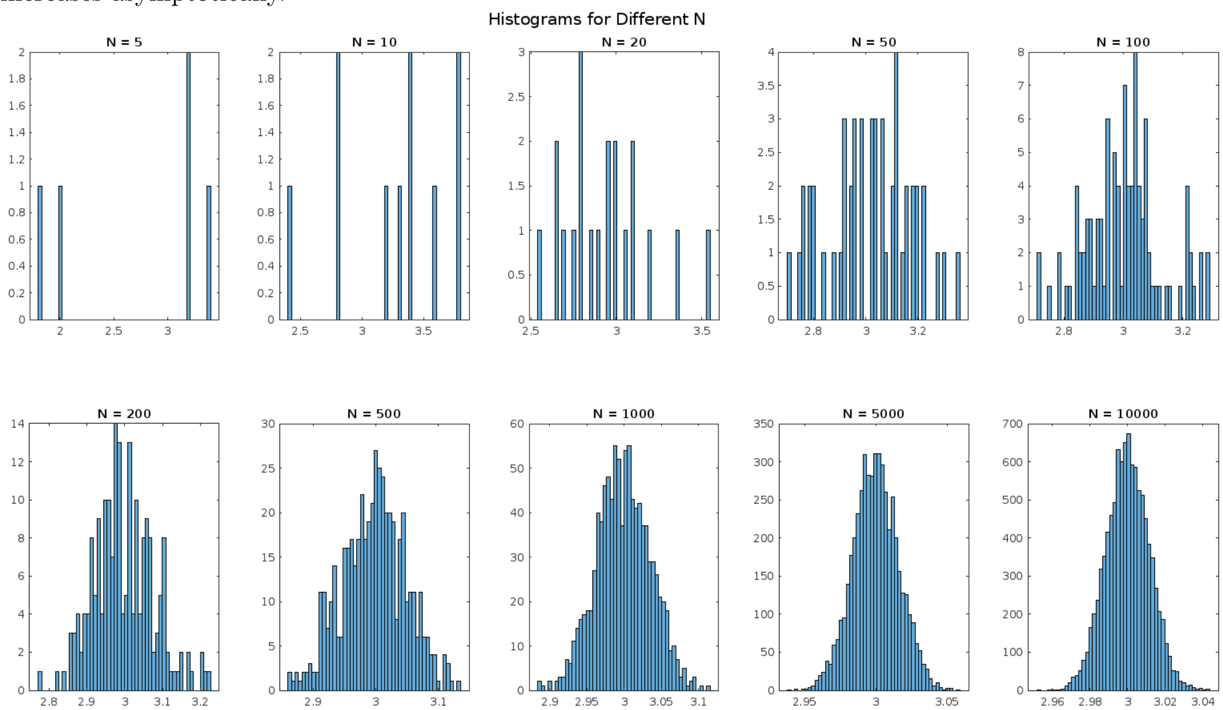
$$r(t) = \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$$
$$r'(t) = \mu(e^t - 1 - \delta) \cdot r(t)$$

The above quantity becomes zero when $(e^t - 1 - \delta)$ becomes zeros, as $r(t) > 0, \forall t \geq 0$. Hence, $t = \ln(1+\delta)$ gives the minima as t^- makes $r'(t)$ negative and t^+ makes $r'(t)$ positive, therefore it gives the tightest bound.

Question 5

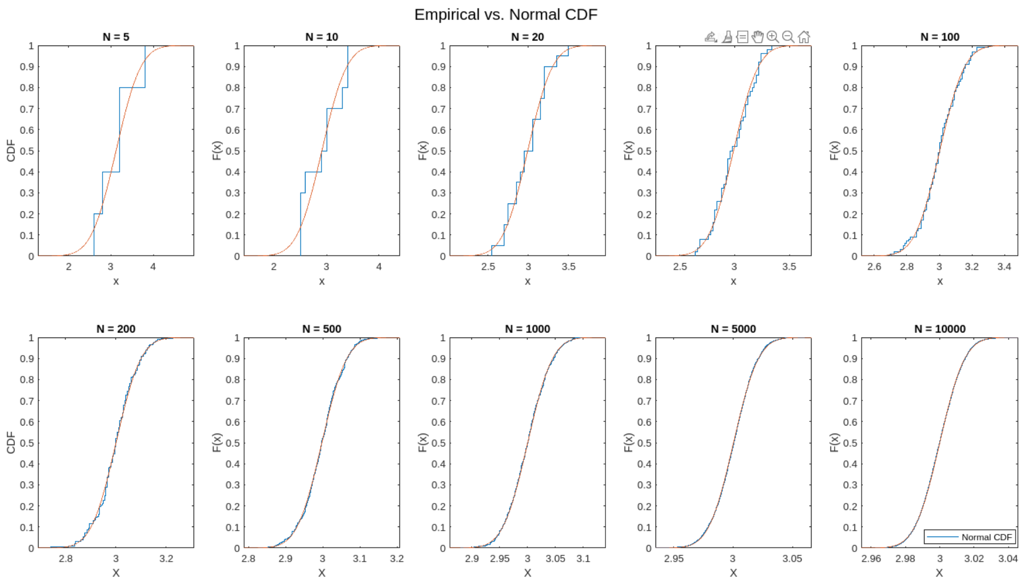
Part (a)

The plots below empirically indicate that the histogram tends to a normal distribution as n increases asymptotically.



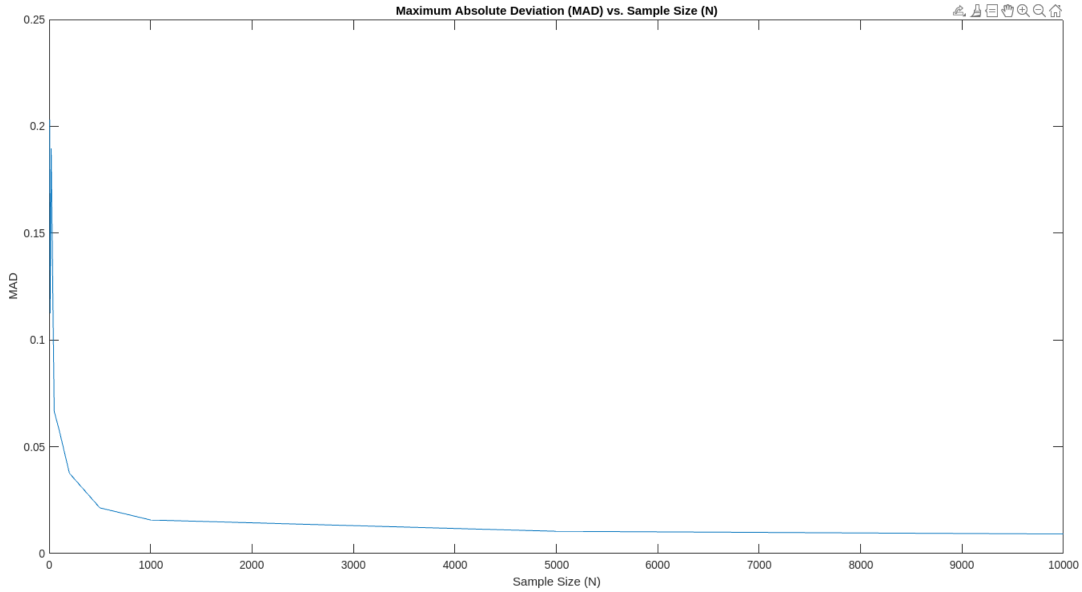
Part (b)

The plots below show that empirically determined CDF approaches normal CDF as n increases asymptotically.



Part (c)

The maximum absolute difference between the empirically determined CDF and normal CDF which converges to 0 as n tends to infinity.



Question 6

The MATLAB code is present in the file **a2_6.m**.

The graph for the relation between correlation coefficient and the shift in the second image is as follows 6:

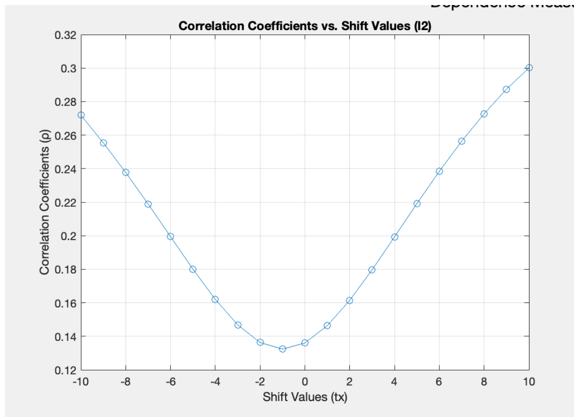


Figure 1: Correlation vs Shift of I2 image

From the graph, it is clear that the correlation coefficients have their minima at shift value of 0. This behaviour seems contradictory to our expectations. For perfectly aligned images, the value of the correlation coefficients should be close to one, but this is clearly not the case, as our plot depicts. the correlation coefficient is neither maximum at shift value of 0, nor the values are closer to 1. The images have been taken from two different settings of the Magnetic Resonance Imaging Machine. Therefore the images aren't exactly the same and hence, the values of correlation coefficients being not close to 1 seems justifiable. But the nature of the graph (minima at 0) is not acceptable.

Let us see the plot for the relation between the correlation coefficient and the shift in the negative of the first image 6:

In this plot, we have first calculated the negative of image 1. $(255 - array(I_1))$ The negative values of the correlation coefficients clearly demonstrates this fact. As we can see from the graph the minima is attained at 0 and is exactly equal to -1. This also demonstrates the correctness of the code. This behaviour is quite expected because we are just comparing a image and its negative through correlation coefficient, and it came out to be -1. As we shift the second image, the correlation will decrease between them and hence the coefficient will become closer to 0. The symmetry of the graph is also self-explanatory.

One thing which we can deduce from the first graph is that correlation coefficients is not a good index to compare two images. Therefore, this index should not be preferred for the various applications in which image processing is applied.

Now, let us have a look at another index of dependence - **Quadratic Mutual Information (QMI)**. The graph of QMI values vs the shift in image 2 is as follows 6:

The expression for Quadratic Mutual Information is as follows 5:

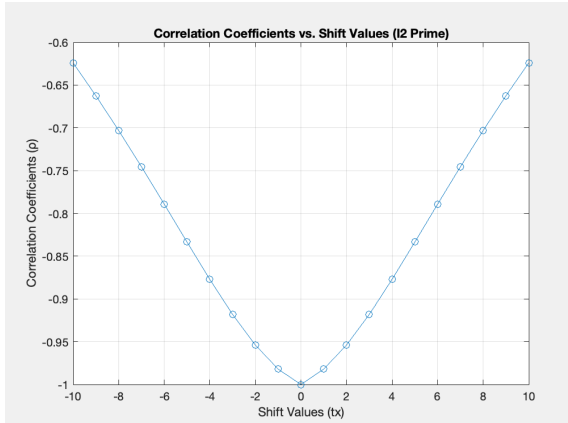


Figure 2: Correlation vs Shift of negative of I1 image

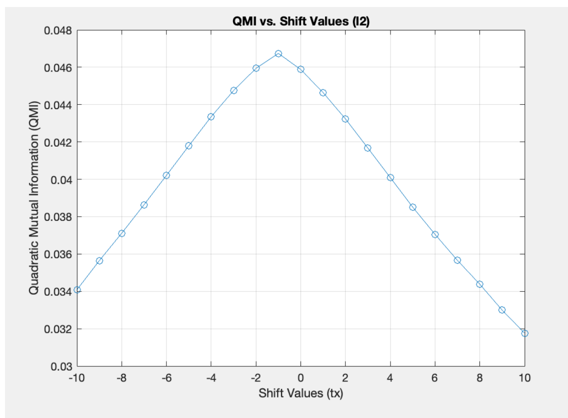


Figure 3: QMI vs Shift of I2 image

$$\mathbf{QMI} = \sum_{i1} \sum_{i2} (p_{I_1 I_2}(i_1, i_2) - p_{I_1}(i_1) p_{I_2}(i_2))^2 \tag{5}$$

where $p_{I_1 I_2}(i_1, i_2)$ represents the normalized joint histogram of I_1 and I_2. For computing the joint histogram, we used a bin-width of 10 in both I_1 and I_2.

The QMI graph attains a maxima around 0, thereby symbolizing the maximum dependence between the two images for a shift of 0. This is a good signal for our research as compared to the graphs of the correlation coefficients at 0. The maxima is not exactly at 0, because of the minor differences between the two images. And moreover, we are doing a padding of 0 for the values which are remaining after performing a shift. So, this small inconsistency seems justifiable.

Let us also see the plot for the QMI values of the negative of Image $I1$. 6

The graph attains a steep maxima at 0. This graph demonstrates the correctness of the code. In this case, the second image is negative of the first one. Thus, they are expected to be strongly depen-

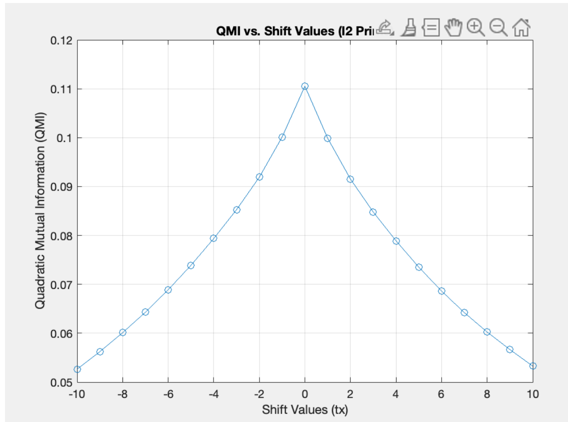


Figure 4: QMI vs Shift of negative of I2 image

dent. The QMI maxima is steeper than the correlation minima, thereby signalling the advantages of QMI index.

One may get a bigger picture by seeing plots for larger shifts of 20 and 50. Here are the plots 6 and 6:

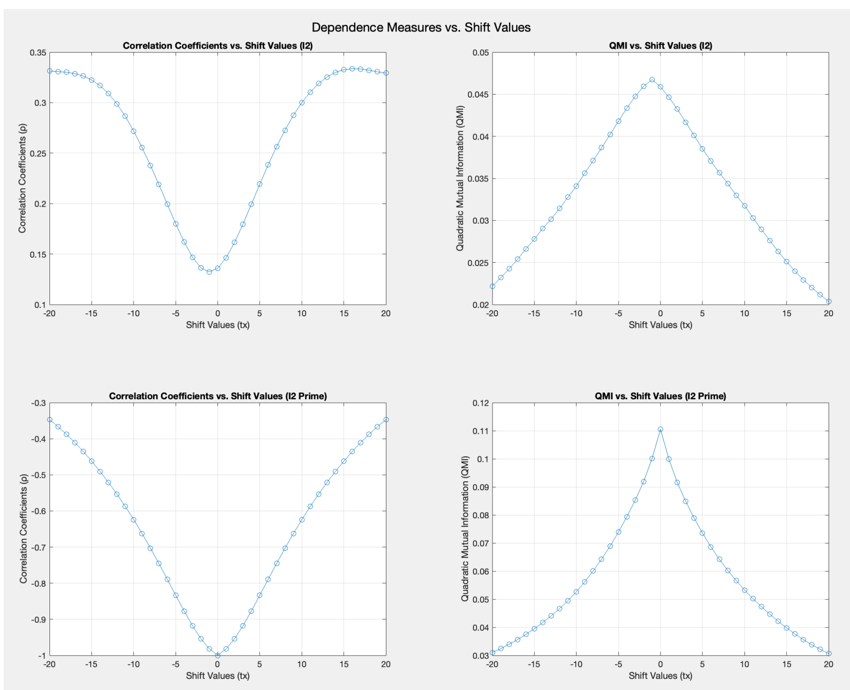


Figure 5: Maximum Shift of 20

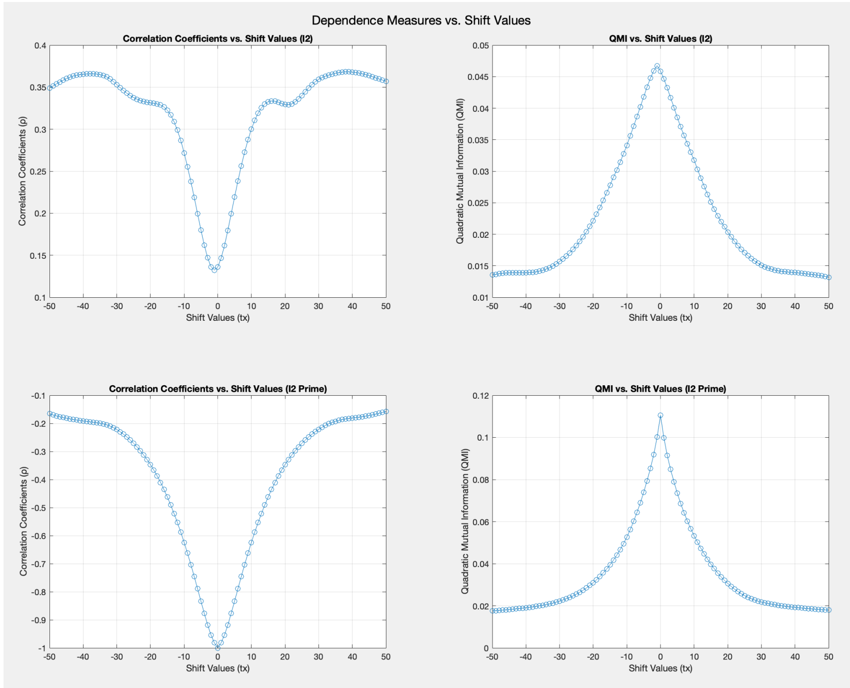


Figure 6: Maximum Shift of 50

From the graphs, we can clearly see that the correlation coefficients has a huge fluctuation over the shifts, but the QMI values are comparatively smoother. Therefore, we can deduce that QMI is a better index as compared to correlation coefficients. QMI is generally more robust to transformations like scaling, translation, and rotation because it measures information content rather than direct pixel relationships. QMI can capture nonlinear dependencies between images because it considers all possible relationships, not just linear ones. This makes it more suitable for cases where the relationship between images is nonlinear or complex.

Question 7

Objective - To derive the covariance matrix of a multinomial distribution using Moment Generating Functions.

For vector random variables, the variance is replaced by the **covariance matrix**. The number of trials n and the success probabilities for each category, i.e. $p_1, p_2, \ldots, p_k$ are all parameters of the multinomial pmf (probability mass function). The covariance matrix in this case will have size $k \times k$, where we have:

$$C(i,j) = E[(X_i - \mu_i)(X_j - \mu_j)] = Cov(X_i, X_j)$$

The covariance matrix is square and symmetric, therefore, it is enough to derive the expression for the diagonal elements C_{ii} and the off-diagonal elements $C_{ij}, i \neq j$.

The moment generating function for a multinomial for an arbitrary k is:

$$\Phi_x(t) = (p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k e^{t_k})^n \tag{6}$$

The above equation has been adopted from the lecture slides. Moment generating functions satisfy:

$$\Phi_x^{(k)}(0) = E(X^k) \tag{7}$$

This result says that the expectation value of X^k is the value of the k^{th} derivative of the moment generating function at $t = 0$. Let us first calcuate the diagonal elements of the covariance matrix:

$$\begin{aligned}
Cov(X_i, X_i) &= E[(X_i - \mu_i)(X_i - \mu_i)] \\
&= E(X_i^2) - 2\mu_i E(X_i) + \mu_i^2 \\
E(X_i) &= \frac{\partial \Phi_X(t)}{\partial t_i}(0) && \text{(From the properties of MGFs)} \\
&= np_i && \text{(From lecture slides)} \\
E(X_i^2) &= \frac{\partial^2 \Phi_X(t)}{\partial t_i^2}(0) && \text{(From the properties of MGFs)} \\
\frac{\partial^2 \Phi_X(t)}{\partial t_i^2} &= \frac{\partial^2((p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k e^{t_k})^n)}{\partial t_i^2} \\
&= \frac{\partial}{\partial t_i}(n((p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k e^{t_k})^{n-1}) p_i e^{t_i}) \\
&= n(n-1)((p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k e^{t_k})^{n-2}) p_i^2 e^{2t_i} \\
&\quad + n((p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k e^{t_k})^{n-1} p_i e^{t_i}) \\
\frac{\partial^2 \Phi_X(t)}{\partial t_i^2}(0) &= n(n-1)p_i^2 + np_i && (\sum_{i=1}^{k} p_i = 1) \\
Cov(X_i, X_i) &= E(X_i^2) - 2 * \mu_i E(X_i) + \mu_i^2 \\
&= n(n-1)p_i^2 + np_i - 2(np_i)^2 + (np_i)^2 && (\mu_i = E(X_i) = np_i) \\
&= np_i(1 - p_i)
\end{aligned}$$

And for independent variables X, Y, we have:

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t) \tag{8}$$

$X_i$andX_j are two independent variables, so let us know calculate the value of the off-diagonal elements of the covariance matrix:

$$\begin{aligned}
Cov(X_i, X_j) &= E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= E(X_i)E(X_j) - \mu_i E(X_j) - \mu_j E(X_i) + \mu_i\mu_j && \text{(Independence of the two variables)} \\
E(X_i) &= \frac{\partial \Phi_X(t)}{\partial t_i}(0) && \text{(From the properties of MGFs)} \\
&= np_i && \text{(From lecture slides)} \\
E(X_j) &= \frac{\partial \Phi_X(t)}{\partial t_j}(0) && \text{(From the properties of MGFs)} \\
&= np_j && \text{(From lecture slides)} \\
E(X_i X_j) &= \frac{\partial^2 \Phi_X(t)}{\partial t_i t_j}(0) && \text{(From the properties of MGFs)} \\
\frac{\partial^2 \Phi_X(t)}{\partial t_i t_j} &= \frac{\partial^2((p_1e^{t_1} + p_2e^{t_2} + \cdots + p_{k-1}e^{t_{k-1}} + p_ke^{t_k})^n)}{\partial t_i t_j} \\
&= \frac{\partial}{\partial t_j}(n((p_1e^{t_1} + p_2e^{t_2} + \cdots + p_{k-1}e^{t_{k-1}} + p_ke^{t_k})^{n-1})p_ie^{t_i}) \\
&= n(n-1)((p_1e^{t_1} + p_2e^{t_2} + \cdots + p_{k-1}e^{t_{k-1}} + p_ke^{t_k})^{n-2})p_ip_je^{t_i}e^{t_j} \\
\frac{\partial^2 \Phi_X(t)}{\partial t_i t_j}(0) &= n(n-1)p_ip_j && (\sum_{i=1}^{k} p_i = 1) \\
Cov(X_i, X_j) &= E(X_i)E(X_j) - \mu_i E(X_j) - \mu_j E(X_i) + \mu_i\mu_j \\
&= n(n-1)p_ip_j - np_i \times np_j - np_i \times np_j + np_i \times np_j && (\mu_i = E(X_i) = np_i) \\
&= -np_ip_j
\end{aligned}$$

Thus, we have calculated the covariance of both the diagonal elements and the off-diagonal elements.

The covariance matrix will look like: $\begin{bmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \dots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \dots & np_k(1-p_k) \end{bmatrix}$