DAI
Assignment #2

Gurnoor Sohal (24B1042)
Sanvi Jain (24B1046)
Rishika Dhiman (24B1090)

September 2, 2025

Directory Structure & Running Instructions

The extracted folder contains six files:

```
A2-24B1042-24B1046-24B1090/
├── Assignment2_Report.pdf
├── Question4_1.m
├── Question4_2.m
├── Question4_3.m
├── Question1_a.m
└── Question1_c.m
```

Description of files:

- `Assignment2_Report.pdf`: The report file.
- `Question4_1.m`: Contains the code for Question 4 when I1 and I2 are the given images T1.jpg and T2.jpg respectively.
- `Question4_2.m`: Contains the code for Question 4 when I_1 is the given image and $I_2 = 255 - I_1$.
- `Question4_3.m`: Contains the code for Question 4 when I_1 is the given image and $I_2 = 255(I_1)^2 / \max((I_1)^2) + 1$.
- `Question1_a.m`: Contains code for Question1 part a(iii).
- `Question1_c.m`: Contains code for Question1 part c.

---

1 Question 1

1.1 (a)

1.1.1 (i)

We need to prove that

$$E[X] = \frac{n}{s} + n\left(1 - (1-p)^s\right),$$

where X is the random variable denoting the total number of tests.

The possible values of X are

$$X \in \left\{\tfrac{n}{s},\ \tfrac{n}{s}+s,\ \tfrac{n}{s}+2s,\ \ldots,\ \tfrac{n}{s}+\tfrac{n}{s}\cdot s\right\}.$$

X being equal to $\frac{n}{s}+is$ means that out of the $\frac{n}{s}$ pools tested, i pools were positive. Hence, all the s members of each positive group must be tested individually.
For $X = \frac{n}{s}+is$, where $0 \le i \le \frac{n}{s}$, the probability is given by

$$P(X) = \binom{\frac{n}{s}}{i}\Big(1-(1-p)^s\Big)^i\Big((1-p)^s\Big)^{\frac{n}{s}-i}.$$

Here, $(1-p)^s$ is the probability that all s members of a group are uninfected, so $1-(1-p)^s$ is the probability that at least one is infected.
Therefore,

$$E[X] = \sum_{i=0}^{\frac{n}{s}}\left(\tfrac{n}{s}+is\right)\binom{\frac{n}{s}}{i}\Big(1-(1-p)^s\Big)^i\Big((1-p)^s\Big)^{\frac{n}{s}-i}.$$

Splitting the summation:

$$E[X] = \frac{n}{s}\sum_{i=0}^{\frac{n}{s}}\binom{\frac{n}{s}}{i}\Big(1-(1-p)^s\Big)^i\Big((1-p)^s\Big)^{\frac{n}{s}-i} + s\sum_{i=0}^{\frac{n}{s}} i\binom{\frac{n}{s}}{i}\Big(1-(1-p)^s\Big)^i\Big((1-p)^s\Big)^{\frac{n}{s}-i}.$$

Using the binomial expansion:

$$E[X] = \frac{n}{s}\Big(1-(1-p)^s+(1-p)^s\Big)^{\frac{n}{s}} + s\frac{n}{s}\Big(1-(1-p)^s\Big)\Big(1-(1-p)^s+(1-p)^s\Big)^{(\frac{n}{s}-1)}$$

Therefore,

$$\boxed{E[X] = \frac{n}{s}+n\Big(1-(1-p)^s\Big).}$$

Alternate solution: *(Using linearity of expectation)*
Since, $T(s)$ is the expected total number of tests we can say that the total number of tests is the sum of tests from Round 1 and Round 2.

- Number of tests from Round 1 is a fixed constant = number of pools = $\frac{n}{s}$. $E(\text{Round 1 Tests}) = \frac{n}{s}$.
- Number of tests from Round 2 incorporates all s members of all the pools which test positive. We know that a pool of s subjects tests negative only if all s subjects are healthy. The probability of this is $(1-p)^s$. The probability that a pool tests positive (at least one subject is diseased) is $1-(1-p)^s$. Hence, the expected number of positive pools is $\frac{n}{s}\times(1-(1-p)^s)$. For each positive pool, all s subjects are tested individually. Thus, the expected number of tests in Round 2 is $s\times\frac{n}{s}(1-(1-p)^s) = n(1-(1-p)^s)$.

Hence:

$$T(s) = E(\text{Round 1 Tests}) + E[\text{Round 2 Tests}] = \frac{n}{s}+n\Big(1-(1-p)^s\Big).$$

1.1.2 (ii)

We are given that if p is very small then the above expression of $T(s)$ can be reduced to:

$$T(s) = \frac{n}{s}+nps.$$

To minimize $T(s)$, we differentiate with respect to s:

$$T'(s) = -\frac{n}{s^2}+np.$$

Equating to 0,

$$-\frac{1}{s^2} + p = 0 \implies s^2 = \frac{1}{p}.$$

Hence,

$$s = \sqrt{\frac{1}{p}}.$$

Thus, $T(s)$ will be the least when $s = \sqrt{\frac{1}{p}}$.
Using this optimal value of s and substituting this value back into the approximate formula:

$$T(s) = \frac{n}{1/\sqrt{p}} + np\left(\frac{1}{\sqrt{p}}\right) = n\sqrt{p} + n\sqrt{p} = 2n\sqrt{p}.$$

1.1.3 (iii)

We need to maximise the value of p so that $T(s) < n$.
We'll first derive the conditions on s, p such that $T(s) < n$.

$$\begin{aligned} T(s) < n &\implies \frac{n}{s} + n\big(1 - (1-p)^s\big) < n \\ &\implies \frac{1}{s} + 1 - (1-p)^s < 1 \\ &\implies \frac{1}{s} < (1-p)^s \\ &\implies \left(\frac{1}{s}\right)^{\frac{1}{s}} < 1 - p \\ &\implies p < 1 - \frac{1}{s^{1/s}}. \end{aligned}$$

$$\boxed{p_{\max}(s) = 1 - \frac{1}{s^{1/s}}}$$

The maximum value of p for a given s such that $T(s) < n$ is given by $p_{max}(s)$ as we derived above.
However, to determine the value of s that maximises p_{max}, we use MATLAB:

```
% Range of s: adjusted such that maxima can be viewed in graph
s = 1:1000;
% Note s has to be integer though p is graphed on other real
    values too
% p_max(s):
p_max = 1 - (1 ./ (s.^(1./s)));

figure;
plot(s, p_max, 'LineWidth', 2);
xlabel('s');
ylabel('p_{max}(s)');
title('Plot of p_{max}(s) v/s s');
grid on;

% Finding the value of s when p becomes maximum
[p, i] = max(p_max);
```

```
s_max = s(i);

fprintf('Maximum p is %f and it occurs when s is %d', p, s_max
    );
```

We find that

$$\boxed{p_{\max}(s) \approx 0.306639 \quad \text{at } s = 3}$$

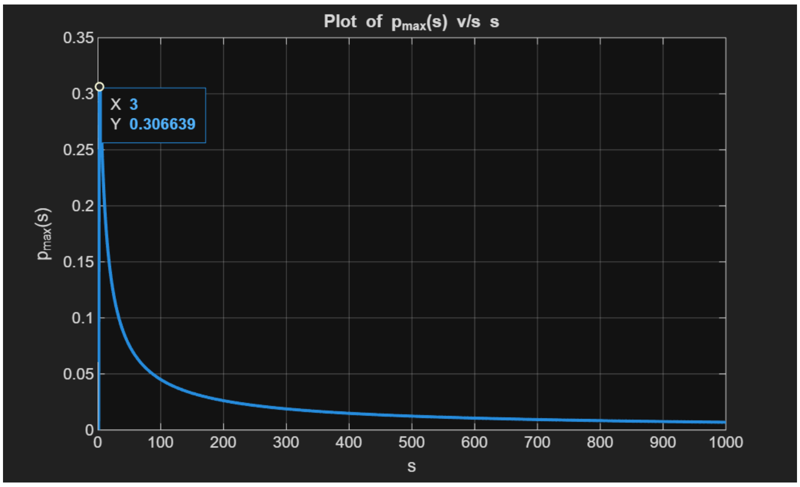


Figure 1.1: Plotting $p_{max}(s)$ versus s

We could have alternately differentiated $p_{max}(s)$ from s and, by equating it to zero, found the value of s that maximises it (noting that $p_{max}(s)$ is differentiable).
Let $f(s) = \frac{1}{s^{1/s}} = s^{-1/s}$. then taking log both sides we get $\ln f(s) = -\frac{\ln s}{s}$. Now differentiating both sides and using chain rule we get $\frac{f'(s)}{f(s)} = -\frac{1}{s^2}(1-\ln s)$. Now using that $p_{max}(s) = 1 - f(s)$, we get:

$$p'_{\max}(s) = -s^{-1/s} \cdot \frac{\ln s - 1}{s^2}.$$

At the maxima, $p'_{\max}(s) = 0$ but because $s^{-1/s} > 0$ and $s^2 > 0$ $(\ln s - 1) = 0$. Hence,

$$p'_{\max}(s) = 0 \quad \implies \quad \ln s = 1 \quad \implies \quad s = e$$

But s is an integer, therefore, the closest value is $s = 3$ which gives $p_{\max}(3) \approx 0.306639$. As a sanity check at $s = 2$ we get $p_{\max}(2) \approx 0.29289$ which is less than $p_{max}(3)$. Therefore,

$$\boxed{\max_s p_{\max}(s) \approx 0.306639 \quad \text{at } s = 3}$$

1.2 (b)

1.2.1 (i)

Fix a genuinely healthy subject and consider a pool. This pool is negative if no other subject is infected and is a subject of this particular pool.
For each of the other subjects $n-1$, the probability that they are infected and present in a given pool (complement of our desired event) is $p \times \pi$ (since the event that a subject is infected has probability p and he participates in a pool has probability π and these events are independent $\implies P(\text{infected} \cup \text{subject participates}) = P(\text{infected}) \times P(\text{subject participates}) = p \times \pi$.
The probability that a subject is not infected and participates in a pool is, therefore, $1 - p\pi$.
Hence:

$$P(\text{pool negative} \mid \text{subject participates}) = (1 - p\pi)^{n-1}$$

If $p\pi$ is small we can use the identity $(1-x)^a = e^{-xa}$ to find the above found probability to be:

$$P(\text{pool negative} \mid \text{subject participates}) \approx e^{-p\pi(n-1)} \quad \text{for small } p\pi$$

1.2.2 (ii)

Note: *The following derivation is for a general case without any assumptions on the size of $p\pi$.*
To find the value of π that maximises the above found probability we differentiate it w.r.t π. Let $f(\pi) = \pi(1-p\pi)^{n-1}$ where $\pi \in [0,1]$. Differentiating both sides w.r.t π we get:

$$f'(\pi) = (1-p\pi)^{n-1} - p\pi(n-1)(1-p\pi)^{n-2}$$

When $f(\pi)$ is at its maximum, its derivative has a value of zero. Hence, equating the above equation to 0, we get:

$$(1-p\pi)^{n-1} - p\pi(n-1)(1-p\pi)^{n-2}$$
$$\implies (1-p\pi) = p\pi(n-1)$$
$$\implies \pi = \frac{1}{np}$$

Note that $np \geq 1$ (since np approximately depicts the number of people who test positive which must be a positive integer) hence $\pi \in [0,1]$.
Note: *Assuming that $p\pi$ is very small.* In order to maximise the probability from part (i), $e^{-n\pi p}$, leads to the trivial solution $\pi = 0$.

1.2.3 (iii)

We need to find the probability that all pools in which a healthy subject participates in are positive. We know using the results derived from (i) that the probability that an entered pool is positive is $1-(1-p\pi)^{n-1}$ (the event is the complement of the event we found in (i)).
Note that the number of pools that a given healthy subject is part of follows a binomial distribution (~Binomial(T_1, π)).
Hence:

$$P(\text{all pools that the subject enters are positive}) = \Sigma_{i=0}^{i=T_1} \binom{T_1}{i} \pi^i (1-\pi)^{T_1-i} (1-(1-p\pi)^{n-1})^i$$

$$= \Sigma_{i=0}^{i=T_1} \binom{T_1}{i} (\pi \times (1-(1-p\pi)^{n-1}))^i (1-\pi)^{T_1-i}$$
$$= (1-\pi(1-p\pi)^{n-1})^{T_1} \quad \text{(using binomial expansion)}$$

Using the optimal $\pi = \frac{1}{np}$ from the previous question we get:

$$P(\text{all pools that the subject enters are positive}) = (1 - \frac{(n-1)^{n-1}}{pn^n})^{T_1} \quad \text{with optimal } \pi$$

Note that if $p\pi$ is small, then using the identity $(1-x)^a = e^{-xa}$ to find the above found probability to be:

$$P(\text{all pools that the subject enters are positive}) = (1-\pi e^{-p\pi(n-1)})^{T_1} \quad \text{for small } p\pi$$

However, if $p\pi$ is small, then, using $\pi = \frac{1}{np}$ we get:

$$P(\text{all pools that the subject enters are positive}) = (1-\frac{e^{-(1-\frac{1}{n})}}{np})^{T_1} \quad \text{for small } p\pi \text{ and optimal value of } \pi$$

1.2.4 (iv)

We need to find the expected total number of tests (round 1 and round 2). Let X be the total number of tests. Then:

$$X = T_1 + (\text{number of subjects tested individually in round 2}).$$

A subject is tested individually in round 2 if:

- They are truly infected (probability p), or
- They are healthy (probability $1-p$) but all pools they participate in are positive.

From (iii), for a healthy subject, the probability that all pools they participate in are positive is:

$$P_{(\text{all pools that the subject enters are positive})} = (1-\pi(1-p\pi)^{n-1})^{T_1}$$

Thus, the probability that a subject is tested in round 2 is:

$$P(\text{tested in round 2}) = p + (1-p)\cdot P_{(\text{all pools that the subject enters are positive})}.$$

Since there are n subjects, the expected number of individual tests in round 2 is:

$$n\left[p + (1-p)\cdot P_{(\text{all pools that the subject enters are positive})}\right].$$

Therefore, the expected total number of tests is:

$$E[X] = T_1 + n\left[p + (1-p)\cdot (1-\pi(1-p\pi)^{n-1})^{T_1}\quad\right].$$

1.2.5 (v)

We minimize $\mathbb{E}[X]$ with respect to T_1. From (iv):

$$\mathbb{E}[X] = T_1 + np + n(1-p)\left[1-\pi(1-p\pi)^{n-1}\right]^{T_1}$$

Let:

$$\alpha = \pi(1-p\pi)^{n-1}$$

Then:

$$f(T_1) = T_1 + np + n(1-p)(1-\alpha)^{T_1}$$

Differentiate with respect to T_1:

$$f'(T_1) = 1 + 0 + n(1-p)(1-\alpha)^{T_1}\ln(1-\alpha)$$

Set $f'(T_1) = 0$:

$$1 + n(1-p)(1-\alpha)^{T_1}\ln(1-\alpha) = 0$$

$$n(1-p)(1-\alpha)^{T_1}(-\ln(1-\alpha)) = 1$$

For small α, $-\ln(1-\alpha)\approx\alpha$, so:

$$n(1-p)(1-\alpha)^{T_1}\alpha \approx 1$$

$$(1-\alpha)^{T_1} \approx \frac{1}{n(1-p)\alpha}$$

Take natural logarithms:

$$T_1\ln(1-\alpha) \approx -\ln(n(1-p)\alpha)$$

Since $\ln(1-\alpha)\approx -\alpha$:

$$T_1\alpha \approx \ln(n(1-p)\alpha)$$

$$T_1 \approx \frac{1}{\alpha}\ln(n(1-p)\alpha)$$

Substitute back $\alpha = \pi(1-p\pi)^{n-1}$:

$$\boxed{T_1 \approx \frac{1}{\pi(1-p\pi)^{n-1}}\ln\left(n(1-p)\pi(1-p\pi)^{n-1}\right)}$$

Now, the minimal expected number of tests:

$$\mathbb{E}[X]^* \approx T_1 + np + n(1-p)(1-\alpha)^{T_1}$$

From earlier:

$$(1-\alpha)^{T_1} \approx \frac{1}{n(1-p)\alpha}$$

So:

$$\mathbb{E}[X]^* \approx T_1 + np + n(1-p)\cdot\frac{1}{n(1-p)\alpha} = T_1^* + np + \frac{1}{\alpha}$$

Substitute $\alpha = \pi(1-p\pi)^{n-1}$:

$$\boxed{\mathbb{E}[X]^* \approx T_1 + np + \frac{1}{\pi(1-p\pi)^{n-1}}}$$

1.3 (c)

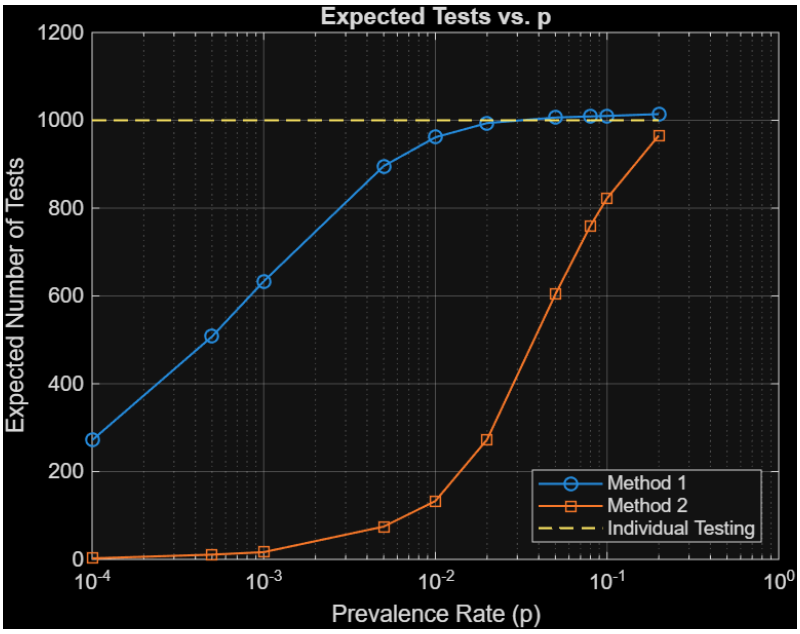


Figure 1.2: Plotting Expected number of tests vs prevalence rate p

From the figure it is easily observable that the expected number of tests in method 2 is lesser than method 1. The difference is considerably high for lower values of p and becomes almost equal for higher values of p.
For example,
At p=0.2, the expected number of tests for method 1 is 1014.1 and for method 2 is 965.4.
While at p = 0.0001, expected number of tests for method 1 is 271.4 and for method 2 is 2.4

This indicates that for very low values of p, method 2 is a better testing method.

2 QUESTION 2

We have been given two independent random variables X and Y with PDFs $f_X(.)$ and $f_Y(.)$. We have to find $f_Z(z)$ where $Z = X \cdot Y$.
For some z,

$$P(Z \in [z, z+dz]) = f_Z(z)\,dz$$

Also as X and Y are independent,

$$P(Z \in [z, z+dz]) = \int_S P(X = x) \cdot P(Y = y)$$

where S is the set of all (x, y) such that $z \le xy \le z + dz$

Hence,

$$P(Z \in [z, z+dz]) = \int_{-\infty}^{\infty} f_X(x)\,P(Y = y)\,dx$$

for all x such that $(x, y) \in S$.

Now, if $x > 0$,

$$z \le xy \le z + dz \quad \Rightarrow \quad \frac{z}{x} \le y \le \frac{z}{x} + \frac{dz}{x}$$

As we know for $y \le Y \le y + dy$

$$P(Y = y) \; = f_Y(y)dy$$

Thus,

$$f_Z(z)\,dz = \int_0^{\infty} f_X(x)\; f_Y\left(\frac{z}{x}\right)\frac{dz}{x}dx, \qquad x > 0.$$

So,

$$f_Z(z) = \int_0^{\infty} f_X(x)\,f_Y\left(\frac{z}{x}\right)\frac{1}{x}\,dx, \qquad x > 0.$$

Similarly, for $x < 0$,

$$f_Z(z) = -\int_{-\infty}^{0} f_X(x)\,f_Y\left(\frac{z}{x}\right)\frac{1}{x}\,dx$$

Hence for $x \neq 0$,

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\,f_Y\left(\frac{z}{x}\right)\frac{1}{|x|}\,dx}$$

3 QUESTION 3

A random variable X has PDF $f_X(.)$ where x_i are independent samples from the PDF. The two ways the student has for estimating $E(x)$ are:

1. $\hat{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. $\hat{x}_2 = \frac{1}{n}\sum_{i=1}^{n} f_X(x_i)x_i$

Of these the correct estimate is $\hat{x}_1$.

3.1 Why $\hat{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ is correct

$E(x)$ is calculated as

$$E[X] = \int_{-\infty}^{\infty} x \cdot (f_X(x))dx$$

where x is taken from the random variable X and hence it needs to be multiplied with its frequency which is $f_X(x)$.
Since x_i are taken from the PDF, they naturally occur with frequency proportional to $f_X(x_i)$ and no further weighting with $f_X(x_i)$ is needed.
Since $\hat{x}_1$ is basically the mean of samples x_i, as we increase the n, by Strong Law of Large Numbers, the mean converges to $E(x)$.

$$\hat{x}_{1[n\to\infty]} = E[X]$$

3.2 Why $\hat{x}_2 = \frac{1}{n}\sum_{i=1}^{n} f_X(x_i)x_i$ is incorrect

Since x_i are samples from PDF, they don't need to be weighted with $f_X(x_i)$, as it already gets embedded when x_i are taken from PDF. This measure will allow double-counting the probability.
When $\hat{x}_2$ is being calculated, it is calculating the expectation of another variable Y which can be represented in terms of X as:

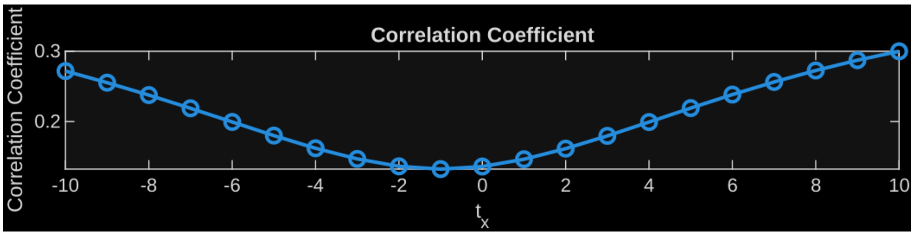
$$Y = f_X(X) \cdot X$$

$$E[Y] = \int_{-\infty}^{\infty} x \cdot (f_X(x))^2 dx$$

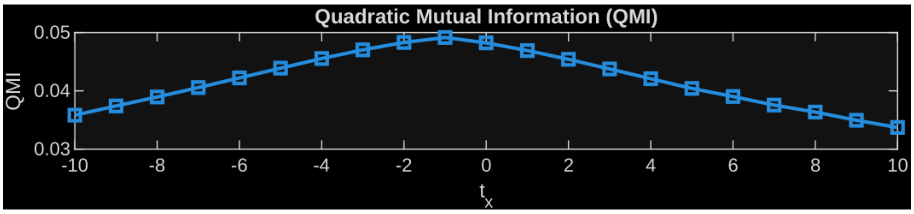
This shows that $\hat{x}_2$ is not calculating the expectation of X and instead for another random variable Y or it is the expectation of X under the density proportional to $(f_X(x))^2$. This can be equal to $E(X)$ in some cases like if $(f_X(x))$ is uniform, but, not in general.

4 Question 4

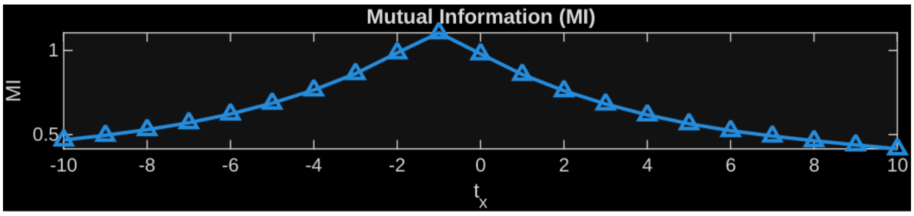
4.1 $Image_1$ v/s shifted $Image_2$



(a) Correlation Coefficient vs. t_x



(b) QMI vs. t_x



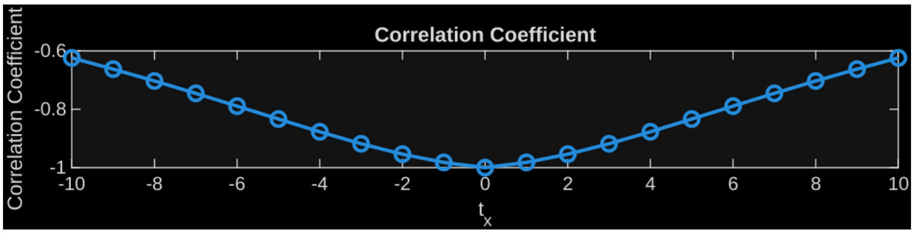
(c) MI vs. t_x

Figure 4.1: Measures of dependance for I_1 vs. shifted I_2.

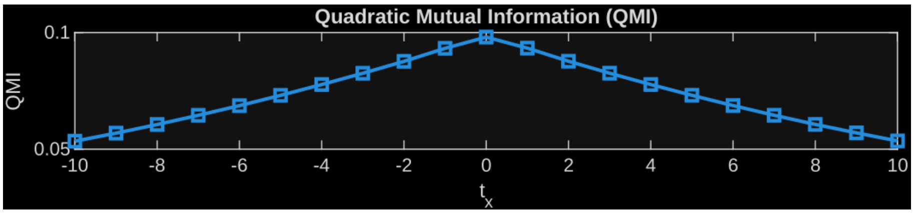
- **MI and QMI:** Both MI and QMI have a sharp peak at $t_x \approx 0$, which corresponds to the point of perfect alignment. As the second image is shifted away from this position, the statistical dependence between the images decreases, causing the MI and QMI values to drop symmetrically.
- **Correlation Coefficient:** The correlation coefficient has the minimum value at $t_x \approx 0$ and increases symmetrically on both sides as the image is misaligned. This suggests that the linear correlation between the pixel intensities of the two images is weakest when they are perfectly aligned.

4.2 $Image_1$ v/s $Image_2$ as the Photographic Negative

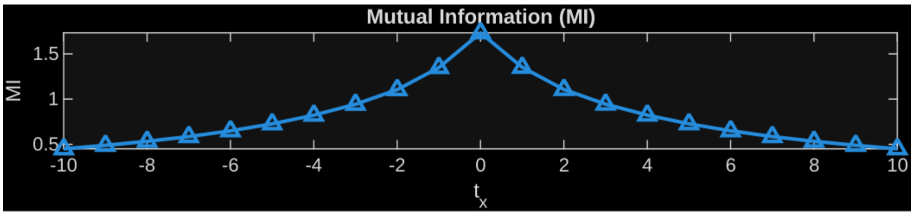
$Image_2$ is the photographic negative of $Image_1$, that is, we have a linear relationship between I_1 and I_2.



(a) Correlation Coefficient vs. t_x



(b) QMI vs. t_x

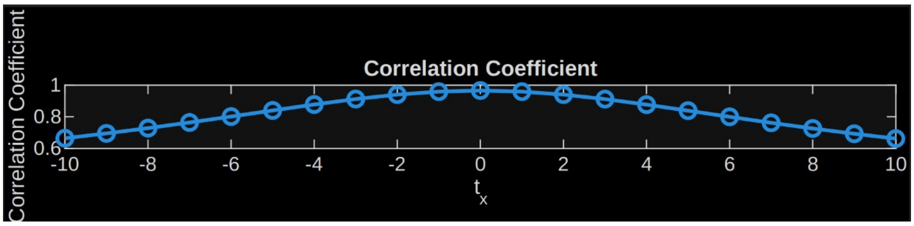


(c) MI vs. t_x

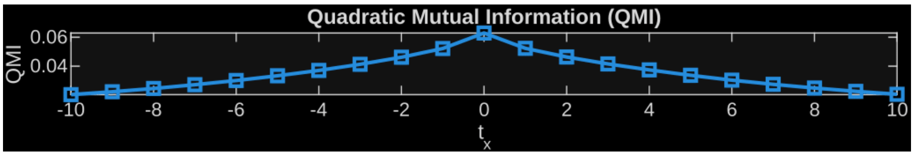
Figure 4.2: Measures of dependence for I_1 v/s its shifted negative I_2

- **Correlation Coefficient:** The correlation coefficient reaches its minimum of $\rho = -1$ at $t_x = 0$, correctly indicating a perfect negative linear relationship at alignment.
  We know that if $y_i = a + bx_i$ where $b < 0$ then $r(x, y) = -1$. Here, $y_i = I_2, a = 255, b = -1, x_i = I_1$ and, therefore, $\rho = r(I_1, I_2 = 255 - I_1) = 1$ when $t_x = 0$.
  As the image is shifted, this pixel-wise relationship is broken, and the correlation magnitude decreases (i.e., moves towards 0).
- **QMI and MI:** Both QMI and MI show sharp peaks at $t_x = 0$ and slope symmetrically on both sides. The maximum value is at $t_x = 0$ showing that this measure of dependence is the highest when they are perfectly aligned.

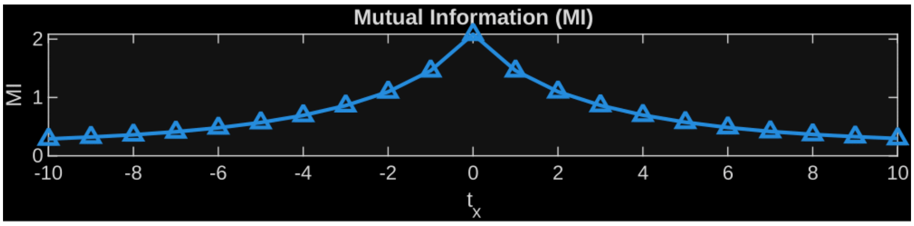
4.3 $Image_1$ v/s Non-linear Transformation in $Image_2$



(a) Correlation Coefficient vs. t_x



(b) QMI vs. t_x



(c) MI vs. t_x

Figure 4.3: Measures of dependence for I_1 v/s its shifted non-linear transformation.

- **MI and QMI:** Both QMI and MI show sharp peaks at $t_x = 0$ and slope symmetrically on both sides. The maximum value is at $t_x = 0$ showing that this measure of dependence is the highest when they are perfectly aligned.
- **Correlation Coefficient:** The correlation coefficient reaches its maximum of $\rho \approx 1$ at $t_x = 0$, which indicates a perfect positive linear relationship at alignment. As the image is shifted, this pixel-wise relationship is broken, and the correlation magnitude decreases (i.e., moves towards 0).
  As $I_2 = 255 \times (I_1)^2/max(I_1)^2 + 1$, we can say that I_2 increases when I_1 increases. Hence, the magnitude of ρ is positive.

4.4 Conclusion

- **Correlation Coefficient:** Correlation depicts dependence the best when there is a linear relationship, as can be seen in (2), relative to non-linear relationships (viz, quadratic - in (3) *for e.g.*)
- **QMI:** It can capture non-linear dependence relationships well successfully. It shows a sharp peak when $t_x = 0$, hence, depicting it's sensitivity to alignment.
- **MI:** Like QMI, it can capture non-linear dependence relationships well successfully. Also, mutual dependence can be observed to be the most reliable measure of dependence. While other measures of dependence had relatively flatter peaks, MI exhibited the sharpest peak of them all, indicating it's relatively higher sensitivity to alignment.

5 Question 5

By Markov's inequality,

$$P(X \geq a) \leq \frac{E[X]}{a}, \quad \text{for } a > 0.$$

Now, for $t > 0$,

$$P(X \geq x) = P\left(e^{tX} \geq e^{tx}\right) \leq \frac{E\left[e^{tX}\right]}{e^{tx}} = e^{-tx}\phi_X(t),$$

where $\phi_X(t) = E[e^{tX}]$ is the moment generating function of X.
Similarly, for $t < 0$,

$$P(X \leq x) = P\left(e^{tX} \geq e^{tx}\right) \leq \frac{E\left[e^{tX}\right]}{e^{tx}} = e^{-tx}\phi_X(t).$$

Now consider $X = \sum_{i=1}^{n} X_i$, where X_i are independent Bernoulli random variables with $E[X_i] = p_i$, and let $\mu = \sum_{i=1}^{n} p_i$.
We want to show:

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

From the previous result,

$$P(X > (1+\delta)\mu) \leq e^{-(1+\delta)\mu t}\phi_X(t).$$

Now,

$$\phi_X(t) = E\left[e^{tX}\right] = E\left[e^{t\sum_{i=1}^{n} X_i}\right] = E\left[\prod_{i=1}^{n} e^{tX_i}\right].$$

Since the X_i are independent,

$$\phi_X(t) = \prod_{i=1}^{n} E\left[e^{tX_i}\right].$$

For each Bernoulli X_i,

$$E[e^{tX_i}] = 1 - p_i + p_i \cdot e^t = 1 + p_i(e^t - 1).$$

Thus,

$$\phi_X(t) = \prod_{i=1}^{n} \left(1 + p_i(e^t - 1)\right).$$

As $p_i(e^t - 1) \geq 0$ (since $t \geq 0$),

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

As $e^t - 1$ is constant and $\sum_{i=1}^{n} p_i = \mu$. Therefore,

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

Finally, applying $1 + x \leq e^x$,

$$1 + (e^t - 1)\mu \leq e^{\mu(e^t-1)}.$$

Hence,

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

Tightening the bound.
Choosing the optimal value of t to minimize the value of $f(t)$. Where $f(t)$ is given as,

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

Taking derivative and setting it to zero:

$$\frac{d}{dt}\left(\mu(e^t-1)-(1+\delta)\mu t\right) = \mu e^t - (1+\delta)\mu = 0.$$

This gives

$$e^t = 1+\delta, \quad \text{so } t = \ln(1+\delta).$$

Substituting back,

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

6 Question 6

Given that there are n independent coin tosses that take place.
For a given coin toss at some trial number T we define

$$\Omega_T = \{H, T\}.$$

It is given that

$$P(H) = p$$

and

$$P(\Omega_T) = P(H) + P(T).$$

By the axioms of probability,

$$1 = p + P(T) \quad \Rightarrow \quad P(T) = 1-p.$$

If T is the random variable that denotes the trial number at which heads first appears, and i denotes a trial number, then

$$P(T=i) = (1-p)^{i-1}\cdot p.$$

(This is because the probability of no head in the first $i-1$ trials is $(1-p)^{i-1}$, and the probability of head appearing in the i-th trial is p.)
The random variable T can range from 1 (head appears in the first trial itself) to n (head appears at the last trial). Note that $T=0$ does not contribute to the expectation at all, hence, can be safely omitted for the calculation of $E[T]$.
Therefore, the expected value is

$$E[T] = \sum_{i=1}^{n} i\,P(T=i) = \sum_{i=1}^{n} i(1-p)^{i-1}p.$$

$$= p\sum_{i=1}^{n} i(1-p)^{i-1}.$$

Let

$$S = 1\cdot(1-p)^0 + 2\cdot(1-p)^1 + \cdots + n\cdot(1-p)^{n-1}.$$

Multiply by $(1-p)$:

$$S(1-p) = 1\cdot(1-p)^1 + 2\cdot(1-p)^2 + \cdots + (n-1)(1-p)^{n-1} + n(1-p)^n.$$

Subtracting, we get:

$$pS = 1 + (1-p) + (1-p)^2 + \cdots + (1-p)^{n-1} - n(1-p)^n.$$

The geometric series gives:

$$1 + (1-p) + \cdots + (1-p)^{n-1} = \frac{1-(1-p)^n}{p}.$$

So,

$$pS = \frac{1-(1-p)^n}{p} - n(1-p)^n.$$

Therefore,

$$E[T] = pS = \frac{1-(1-p)^n}{p} - n(1-p)^n.$$

$$\boxed{E[T] = \frac{1-(1-p)^n}{p} - n(1-p)^n}$$

$$\boxed{E[T] = \frac{1-(1-p)^n}{p} - n(1-p)^n}.$$

As $n \to \infty$ the above expected value becomes:

$$\boxed{E[T] \approx \frac{1}{p}} \quad \text{as } n \to \infty$$