HW2

Dorisala Srinivas Reddy(24b1000),Charan Bothamolla(24b0966)

August 2025

1. Consider you are testing samples (e.g., blood, urine, saliva, etc. During the COVID19 pandemic, the most common method for testing was RT-PCR on nasal or throat mucus.) of n different subjects for a rare disease. Testing each subject independently requires n tests which can be expensive and time-consuming (and unjustifiable since most tests would report negative, i.e. healthy or not diseased). Pooled testing is a technique where you take small portions of the sample of a subset of people and mix them together to create a pool. This pool is tested instead of testing the individual samples. If the pool tests negative (i.e. not diseased), then you can declare all subjects participating in that pool to be healthy– assuming test results are accurate. Overall, this greatly saves on the number of tests, if the disease is rare. If the pool tests positive (i.e. one or more participating subjects is/are diseased), then more work needs to be done, as we shall see below. Typically, the number of pools is taken to be less than n to save on testing time and resources. We will use our knowledge of probability and statistics to analyze some pooled testing methods.

   (a) In a popular technique of group testing, we divide the n subjects into n/s disjoint subsets each containing s subjects. For simplicity, assume that n is divisible by s. In round 1, the n/s pools are tested. Participating subjects in a negative pool are declared healthy. Subjects that participated in a positive pool are tested individually in round 2. Let the prevalence rate of the disease (the proportion of people who have the disease) be p where $p \in [0, 1]$. Now do as directed. Assume throughout that the test results are accurate.

      i. Prove that the expected total number of tests for this method (counting both round 1 and 2) is $T(s) = \frac{n}{s} + n(1 - (1 - p)^s)$.
      sol(i):
      E(Total number of tests)=Tests in Round 1+E(Tests in Round 2)
      The above statement is correct because the number of tests in Round 1 are always same i.e; it is a constant (E(a+X)=E(a)+E(X)=a+E(X))
      tests in round1
      number of tests in the first round are always equal to $\frac{n}{s}$ independent of the prevalence rate of the disease. which is the each group being tested is always has probability of **1** in round 1.
      Expected number of tests in round2
      The probability of randomly picked person to be +ve $=\frac{\text{no.of +ve persons}}{\text{total no.of persons}}=\frac{n*p}{n} = p$
      Let $P(r)$ = the probability for r tests to be happening.
      As the person can participate in only one pool in Round 1 and that if a pool tests $+ve$ every one of the pool should be tested in Round 2
      So we can say that the number of tests in Round 2 will be multiple of s(number of persons in each pool)
      from which we can say $P(r) = 0$ for r not equal to integer multiple of s then,

$$E(testsR2) = \sum_{i=0}^{n/s} (s * i)P(s * i)$$

      Now lets calculate the probability of i exactly i pools tests $+ve$ in Round 1
      $P(s * i)$ which says thar exactly i pools are tested $+ve$

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

      Here the term $\binom{\frac{n}{s}}{i}$ comes from the sum rule of probability that all combinations of i pools tests $+ve$.
      The term $(1 - (1 - p)^s)^i$ comes from that every pool from that i should contain atleast one $+ve$ subject which will be equal to 1 − probability that there all non +ve person in the pool
      The term $((1 - p)^s)^{\frac{n}{s}-i}$ comes from that the remaining $\frac{n}{s} - i$ pools have all non +ve persons

      By substituting the $P(s * i)$ in the E(testsR2) we get

$$E(testsR2) = \sum_{i=0}^{\frac{n}{s}} s * i * \binom{\frac{n}{s}}{i} * (1 - (1 - p)^s)^i((1 - p)^s)^{\frac{n}{s}-i}$$

$$E(testsR2)) = \sum_{i=0}^{\frac{n}{s}} s * i * \frac{(\frac{n}{s} * (\frac{n}{s} - 1)!)}{(i * (i-1)!) * (\frac{n}{s} - i)!} * (1 - (1-p)^s)^i ((1-p)^s)^{\frac{n}{s} - i}$$

$$E(testsR2)) = \sum_{i=0}^{\frac{n}{s}} s * \frac{n}{s} \frac{(\frac{n}{s} - 1)!}{(i-1)! * (\frac{n}{s} - i)!} * (1 - (1-p)^s)^i ((1-p)^s)^{\frac{n}{s} - i}$$

Excluding the $i = 0$ as it becomes zero.

$$E(testsR2)) = \sum_{i=1}^{\frac{n}{s}} n \binom{\frac{n}{s} - 1}{i - 1} * (1 - (1-p)^s)^i ((1-p)^s)^{\frac{n}{s} - i}$$

now take $(1 - (1-p)^s)$ outside of the summation

$$E(testsR2)) = (1 - (1-p)^s) \sum_{i=1}^{\frac{n}{s}} n \binom{\frac{n}{s} - 1}{i - 1} * (1 - (1-p)^s)^{i-1} ((1-p)^s)^{\frac{n}{s} - i}$$

$$E(testsR2)) = (1 - (1-p)^s) n * 1$$

combining both Round 1 and 2
now total no.of Expected tests are $\boldsymbol{\frac{n}{s} + n(1 - (1-P)^s)}$

ii. If p is very small, this equals $\frac{n}{s} + nps$. For what value of s, is this expected number the least? What is the expected number of tests in that case?
sol(ii):
Let $E = \frac{n}{s} + nps$ (where E is the expected no of tests)
we calculate by equating the differentiation of E with respect to s to 0

$$\frac{dE}{ds} = -\frac{n}{s^2} + np$$

$$\frac{dE}{ds} = 0$$

$$\frac{n}{s^2} = np$$

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

To prove that it is minimum we should also check the $\frac{d^2E}{ds^2}$ at $s = \frac{1}{\sqrt{p}}$

$$\frac{d^2E}{ds^2} = \frac{d}{ds}(\frac{dE}{ds})$$

$$\frac{d^2E}{ds^2} = \frac{d}{ds}(-\frac{n}{s^2} + np)$$

$$\frac{d^2E}{ds^2} = 2\frac{n}{s^3}$$

as n is $+ve$ $\frac{d^2E}{ds^2}$ will be +ve for all +ve values of s(here $s = \frac{1}{\sqrt{p}}$ is a +v3 value
when $\frac{d^2E}{ds^2}$ is +ve the graph of E is concave up which means the local point ($\frac{dE}{ds} = 0$) is a minimum.
now the $E = n * \sqrt{p} + \frac{np}{\sqrt{p}}$

$$E = 2n\sqrt{p}$$

iii. What is the maximum value of p for which $T(s) < n$? For this last part, you may need to do some simple numerical computation in MATLAB. [3+(2.5+2.5)+4=12 points]
sol(iii):
for this lets first simplyfy the given equation .

$$T(S) < n$$

$$\frac{n}{s} + n(1-(1-P)^s) < n$$

$$\frac{1}{s} + 1 - (1-P)^s < 1$$

$$\frac{1}{s} < (1-P)^s$$

now apply power $\frac{1}{s}$ on both sides as s is positive the inequality sign doesn't changes.
we get

$$(\frac{1}{s})^{\frac{1}{s}} < 1 - P$$

$$P < 1 - (\frac{1}{s})^{\frac{1}{s}}$$

from the above equation we can say that P's possible maximum is near to $1-(\frac{1}{s})^{\frac{1}{s}}$. To calculate the max possible of $1-(\frac{1}{s})^{\frac{1}{s}}$ for some value of s, which is similar to minimum os $(\frac{1}{s})^{\frac{1}{s}}$.
This can be calculated by some rigorous differentiation But we can just calculate that value for some values of s to observe the minimum of $(\frac{1}{s})^{\frac{1}{s}}$ For that the code is given in **maxp.m**.
By that we will get s=3;and p's maximum possible as $1-(1/3)^{(1/3)}$.

(b) In another method, round 1 involves testing of T_1 pools. A subject participates in a pool with probability π independent of the other items and other pools (so the pools may be overlapping in this case). Similar to the earlier method, all subjects that participated in a positive pool are individually tested in round 2. Let the prevalence rate of the disease be p. Do as directed. Throughout this exercise, assume that all tests are accurate. Also use the identity $(1-x)^a = e^{-xa}$ for small $x > 0$
no of subjects who have positive from n subjects are =p*n

i. What is the probability that a genuinely healthy subject participates in a pool that tests negative?
sol:
let P= probability that a genuinely healthy subject participates in a pool that tests negative
no of subjects who are unhealthy is =n*p=np
for a genuinely healthy subject participates in a pool that tests negative to exist there shoudn't be any participation of unhealthy subject in that pool
probability for paricipation of healthy subject=π
probability for shoudn't be any participation of unhealthy subject in that pool=$(1-\pi)^{np}$ as no one from np subjects should participate
$\therefore$ P=$\pi(1-\pi)^{np}$

ii. For what value of π (in terms of n and p) is this probability maximized? (Note that we would like this probability to be as high as possible.)
sol:

$$
\begin{aligned}
P &= \pi(1-\pi)^{np} \\
\frac{\partial P}{\partial \pi} &= np(1-\pi)^{np-1} + (1-\pi)^{np} = 0 \\
&= -\pi np + (1-\pi) \\
1 &= \frac{1}{\pi(1+np)} \\
\pi &= \frac{1}{1+np} \\
\frac{\partial^2 P}{\partial \pi^2} &= -np - 1 \\
\text{for } \pi &= \frac{1}{(1+np)} \text{ probability is maximized}
\end{aligned}
$$

iii. Given this optimal π, what is the probability that all pools that a genuinely healthy subject participates in, test positive?
sol:

Let P(X)= probability when all participated tests of a particular healthy person should be positive
For this we consider all the possible cases when the participated pools are iterated from 0 to T1. Probability will be sum this all by Principle of sum.

$$P(X) = \sum_{i=0}^{T_1} \binom{T_1}{i} \pi^i * (1-\pi)^{(T_1-i)} P(y_i)$$

The term $\binom{T_1}{i}$ is for selecting the i pools in which the person participates ,This all possibilities have same probability so we these combinations all at once.
The terms π^i is for that the person will participate in selected i pools
The term $(1-\pi)^{(T_1-i)}$ represents that the oarticukar person will not participates in other (T_1-i) pools as we took it Exactly i pools
Where $P(y_i)$ means the Probability that all pools are $+ve$ when the particular healthy person participates in i pools exactly.
Now to calculate $P(y_i)$ we will do like

$$P(y_i) = (\text{Probability that atleast one +ve person participate in the pool})^i$$

The above equation says that every pool (fro the selected i pools) should contain atleast one +vely diseased person. We can say that,
Probability that atleast one +ve person participate in the pool=1-Probability that no +ve participates in the pool
The total number of +vely diseased are np

$$\text{Probability that at least one +ve person participate in the pool} = 1 - (1-\pi)^{np}$$

$$P(y_i) = (1-(1-\pi)^{np})^i$$

$$P(X) = \sum_{i=0}^{T_1} \binom{T_1}{i} \pi^i * (1-\pi)^{(T_1-i)} (1-(1-\pi)^{np})^i$$

$$P(X) = \sum_{i=0}^{T_1} \binom{T_1}{i} (1-\pi)^{(T_1-i)} (\pi - \pi(1-\pi)^{np})^i$$

$$P(X) = ((1-\pi) + (\pi - \pi(1-\pi)^{np}))^{T_1}$$
$$P(X) = (1 - \pi(1-\pi)^{np})^{T_i}$$

NOTE:Here we also considered number of pools as 0 Because when the person does not participate in any of the pools then we can vacuously say that all pools which he participated are tested +ve will be true.

(b)(iv) Expected total number of tests (round 1 + round 2)

Fix n, prevalence $p \in [0,1]$, number of round-1 pools T_1, and per-pool participation probability $\pi \in [0,1]$. Round 1 uses exactly T_1 tests.
Consider an arbitrary subject.

- If the subject is *positive* (probability p), they are tested in round 2 iff they participate in at least one pool. The probability of that is
$$1 - (1-\pi)^{T_1}.$$
Hence the expected number of round-2 tests of positive subjects is
$$np\big[1 - (1-\pi)^{T_1}\big].$$
- If the subject is *healthy* (probability $1-p$), then for a fixed pool the event "this subject is *not* individually tested because of that pool" is
$$(\text{not join the pool}) \quad \text{or} \quad (\text{join \& no other positive in pool}).$$
The probability of "join & no other positive in that pool" equals $\pi\,(1-\pi p)^{\,n-1}$ (each other of the $n-1$ subjects either does not join or is not positive; probability no other joining positive is $(1-\pi p)^{n-1}$). Thus for one pool the probability of *not* causing a test is
$$(1-\pi) + \pi(1-\pi p)^{\,n-1} = 1 - \pi\big[1 - (1-\pi p)^{\,n-1}\big].$$
Since the subject's participation/others' participations in different pools are independent, the probability the healthy subject is *never* in a positive pool (hence not tested individually) is
$$\Big[1 - \pi\big(1 - (1-\pi p)^{\,n-1}\big)\Big]^{T_1}.$$
Therefore the probability a healthy subject *is* individually tested in round 2 equals
$$1 - \Big[1 - \pi\big(1 - (1-\pi p)^{\,n-1}\big)\Big]^{T_1},$$
and the expected number of such tests is
$$n(1-p)\Big\{1 - \Big[1 - \pi\big(1 - (1-\pi p)^{\,n-1}\big)\Big]^{T_1}\Big\}.$$

Adding round 1 and both contributions from round 2 yields the exact expectation

$$\boxed{E[\text{tests}] = T_1 + np\big[1 - (1-\pi)^{T_1}\big] + n(1-p)\Big\{1 - \Big[1 - \pi\big(1 - (1-\pi p)^{\,n-1}\big)\Big]^{T_1}\Big\}.}$$

(b)(v) Minimization over T_1

We now analyze how $E[\text{tests}]$ varies with T_1 (treating π fixed for the moment).
Differentiate the exact expression with respect to the continuous parameter T_1 (this shows monotonicity; for integer T_1 the same conclusion follows by the discrete difference):

$$\frac{d}{dT_1}E[\text{tests}] = 1 + np \cdot \big[-(1-\pi)^{T_1}\ln(1-\pi)\big] + n(1-p)\cdot\big[-A^{T_1}\ln A\big],$$

where we set

$$A := 1 - \pi\big(1-(1-\pi p)^{\,n-1}\big).$$

Note that $0 < 1-\pi < 1$ and $0 < A < 1$ for $0 < \pi < 1$, hence $\ln(1-\pi) < 0$ and $\ln A < 0$. Consequently each bracketed term $-(\cdot)^{T_1}\ln(\cdot)$ is strictly positive, so every summand on the right is positive. Therefore

$$\frac{d}{dT_1}E[\text{tests}] > 0 \qquad \text{for all } T_1 > 0.$$

Thus $E[\text{tests}]$ is strictly increasing in T_1.
Hence the expected number of tests is minimized at the smallest admissible T_1. If the protocol requires at least one pool in round 1, the minimizer is

$$\boxed{T_1^\star = 1.}$$

(If one allows $T_1 = 0$ — i.e. skip round 1 and individually test everyone — then $T_1^\star = 0$.)

Expected tests at the minimizing T_1

For the practically relevant case $T_1^\star = 1$, the exact expression simplifies (put $T_1 = 1$):

$$\begin{aligned}
E[\text{tests} \mid T_1 = 1] &= 1 + np\big[1-(1-\pi)\big] + n(1-p)\Big\{1 - \Big[1-\pi\big(1-(1-\pi p)^{\,n-1}\big)\Big]\Big\} \\
&= 1 + np\pi + n(1-p)\,\pi\big(1-(1-\pi p)^{\,n-1}\big) \\
&= 1 + n\pi\Big\{p + (1-p)\big(1-(1-\pi p)^{\,n-1}\big)\Big\}.
\end{aligned}$$

If you want the further step of choosing π to minimize this (or using the earlier approximation $\pi^\star \approx 1/(pn)$), I can expand/substitute and produce series approximations; tell me whether you prefer the exact optimization in π or the usual small-p, large-n approximations (e.g. $(1-x)^a \approx e^{-ax}$).

(c) For n = 1000 and $p \in \{10^{-4}, 5*10^{-4}, 0.001, 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2\}$, plot a graph of the expected number of tests for both methods versus n. For the first method, the expected number should be for optimal s and for the second method, it should be for optimal π and optimal T_1. Include this plot in your report and comment on it. [2 points]
For Part(a):

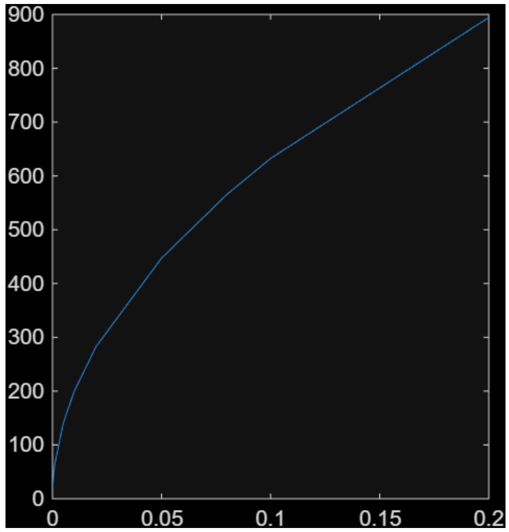

Figure 1: Graph btw Expected and Prevalence rate of the disease

The Above graph shows that the expected number of tests increases with the prevekence rate of the disease.This plot is Drawn for the optimal value of s from the a(ii) part of the question which assumes p is very less. for part(b):

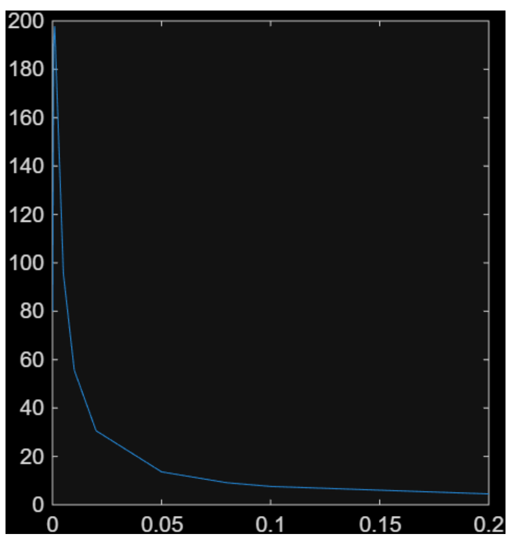

Figure 2: Graph btw Expected and Prevalence rate of the disease

The Below graph shows that the expected number of tests are decreasing for the optimal $T_i = 1$ may due to the independence of the subject to participate in the pools. The above plots are drawn by running the **plots.m** in the code given.

2. Let X and Y be two independent random variables with PDFs fX(.) and fY (.) respectively. Derive the PDF of the random variable Z = XY in terms of fX, fY . (Hint: Look at the discussion forum on moodle where the PDF of X + Y was derived. Follow similar steps.)
sol:
X and Y are random variables with probability density functions $f_X(a)$ and $f_Y(y)$, let $F_X(a)$ and $F_Y(y)$ are its cumulative distribution functions . We aim to find the probability density function $f_Z(z)$ of the product $Z = XY$.

The cumulative distribution function of Z is defined as:

$$F_Z(z) = P(Z < z) = P(XY < z).$$

For a fixed $a \le X \le a + da$, the condition $XY < z$ implies:

- If $a > 0$, then $Y \le \frac{z}{a}$, so $P\left(Y \le \frac{z}{a}\right) = F_Y\left(\frac{z}{a}\right)$.
- If $a < 0$, then $Y > \frac{z}{a}$, so $P\left(Y > \frac{z}{a}\right) = 1 - F_Y\left(\frac{z}{a}\right)$.

Thus, we express $F_Z(z)$ by integrating over all possible values of a:

$$F_Z(z) = \int_{-\infty}^{\infty} f_X(a) P\left(Y < \frac{z}{a}\right) da.$$

Splitting the integral based on the sign of a:

$$F_Z(z) = \int_{-\infty}^{0} f_X(a)\left(1 - F_Y\left(\frac{z}{a}\right)\right) da + \int_{0}^{\infty} f_X(a) F_Y\left(\frac{z}{a}\right) da.$$

To find the probability density function $f_Z(z)$, we differentiate $F_Z(z)$ with respect to z:

$$f_Z(z) = \frac{\partial}{\partial z}\left(\int_{-\infty}^{0} f_X(a)\left(1 - F_Y\left(\frac{z}{a}\right)\right) da + \int_{0}^{\infty} f_X(a) F_Y\left(\frac{z}{a}\right) da\right).$$

Applying the derivative inside the integrals:

$$f_Z(z) = \int_{-\infty}^{0} f_X(a)\frac{\partial}{\partial z}\left(1 - F_Y\left(\frac{z}{a}\right)\right) da + \int_{0}^{\infty} f_X(a)\frac{\partial}{\partial z} F_Y\left(\frac{z}{a}\right) da.$$

Compute the derivatives:

$$\frac{\partial}{\partial z} F_Y\left(\frac{z}{a}\right) = f_Y\left(\frac{z}{a}\right) \cdot \frac{1}{a},$$
$$\frac{\partial}{\partial z}\left(1 - F_Y\left(\frac{z}{a}\right)\right) = -f_Y\left(\frac{z}{a}\right) \cdot \frac{1}{a}.$$

Thus:

$$f_Z(z) = \int_{-\infty}^{0} f_X(a)\left(-\frac{1}{a} f_Y\left(\frac{z}{a}\right)\right) da + \int_{0}^{\infty} f_X(a)\left(\frac{1}{a} f_Y\left(\frac{z}{a}\right)\right) da.$$

Combine the integrals using the absolute value of a:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(a) f_Y\left(\frac{z}{a}\right)\frac{1}{|a|} da.$$

3. Consider a random variable X with PDF $f_X(.)$. Consider independent samples $\{x_i\}_{i=1}^n$ from this PDF. A student wants to estimate E(X) given these samples. Let the estimated value be $\hat{x}$. Should (s)he consider $\hat{x} := \sum_{i=1}^n x_i/n$ or $\hat{x} := \sum_{i=1}^n f_X(x_i)x_i/n$ as the estimate for E(X)? Which one of these is correct and why? To answer the why part, you explain how the other option is incorrect (eg: maybe it is estimating the expectation of some other random variable). [10 points]
sol:
first lets assume that **n is large** now then by weak law of Large numbers says that (for any +ve value ϵ)

$$P(|\frac{\sum_{i=1}^n x_i}{n} - E(X)| > \epsilon) \to 0$$

which means that the E(X) will be approximate to $\frac{\sum_{i=0}^n x_i}{n}$ for large values of n, so, the student should consider $\frac{\sum_{i=0}^n x_i}{n}$ as the expected value for the E(X).
similarly if we apply the weak law of large numbers for the random variable $Y = f_X(X)X$ we get $E(Y) = \sum_{i=1}^n f_X(x_i)x_i/n$.

4. Read in the images T1.jpg and T2.jpg from the homework folder using the MATLAB function imread and cast them as a double array using the code
report:
`im = double(imread('T1.jpg'));`. These are magnetic resonance images of a portion of the human brain, acquired with different settings of the MRI machine. They both represent the same anatomical structures and are perfectly aligned (i.e. any pixel at location (x, y) in both images represents the exact same physical entity). Consider random variables I1, I2 which denote the pixel intensities from the two images respectively. Write a piece of MATLAB code to shift the second image along the X direction by t_x pixels where t_x is an integer ranging from -10 to +10. While doing so, assign a value of 0 to unoccupied pixels. For each shift, compute the following measures of dependence between the first image and the shifted version of the second image: report: for runing code type name of codefile in terminal which is code1,code2,code3

   (a) **Case 1**

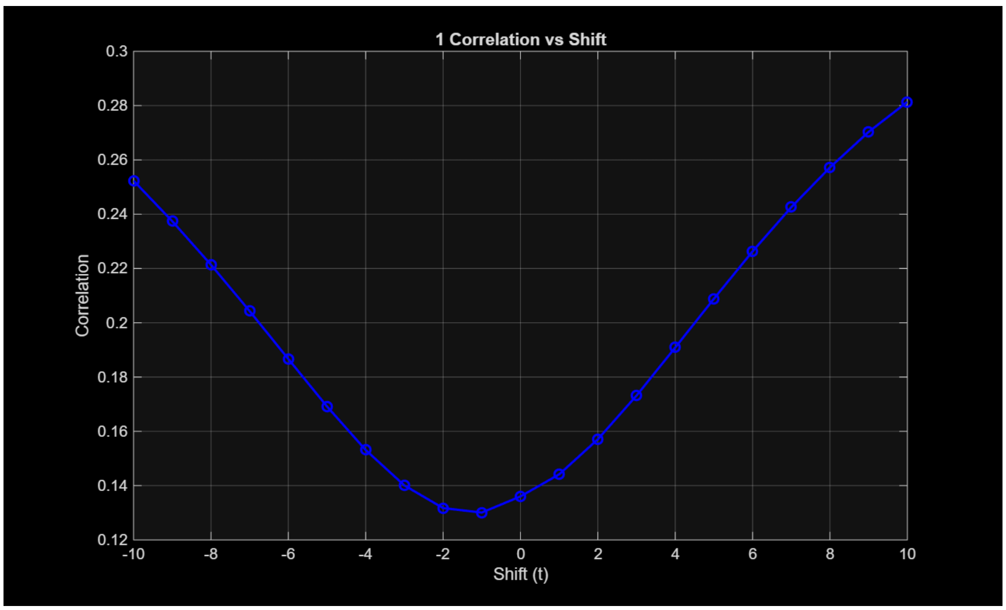

Figure 3: Correlation vs Shift (Case 1)

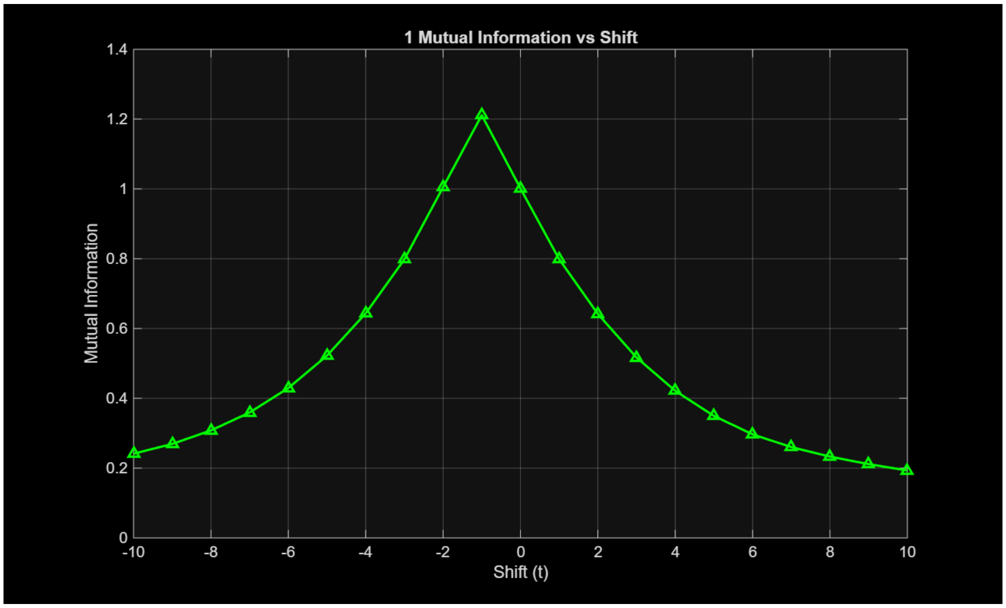

Figure 4: Mutual Information vs Shift (Case 1)

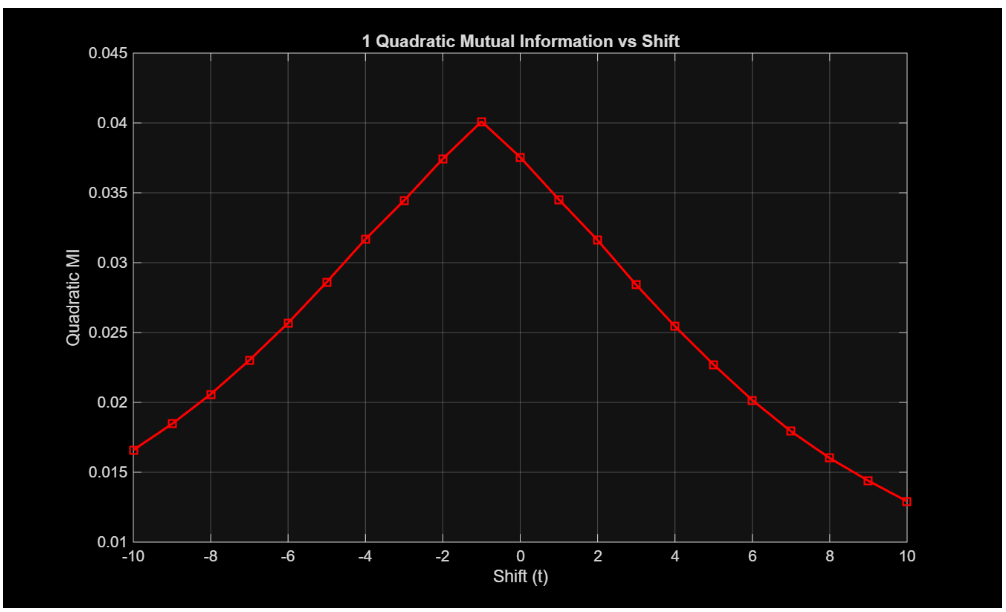

Figure 5: Quadratic Mutual Information vs Shift (Case 1)

Observation: All three measures peak near zero shift, showing best alignment at $t = 0$.
For plot 1 (1 Quadratic Mutual Information vs Shift), the Quadratic Mutual Information peaks at a shift of 0, indicating maximum dependence when the images are perfectly aligned, and decreases symmetrically as the shift moves away from 0, suggesting reduced alignment.
For plot 1 (1 Mutual Information vs Shift), the Mutual Information peaks at a shift of 0, showing maximum dependence with perfect alignment, and decreases symmetrically as the shift increases or decreases, indicating reduced alignment.
For plot 1 (1 Correlation vs Shift), the Correlation shows a U-shaped curve with a minimum at shift 0, suggesting that perfect alignment results in the lowest correlation, increasing as the shift moves

away from 0.

(b) **Case 2**

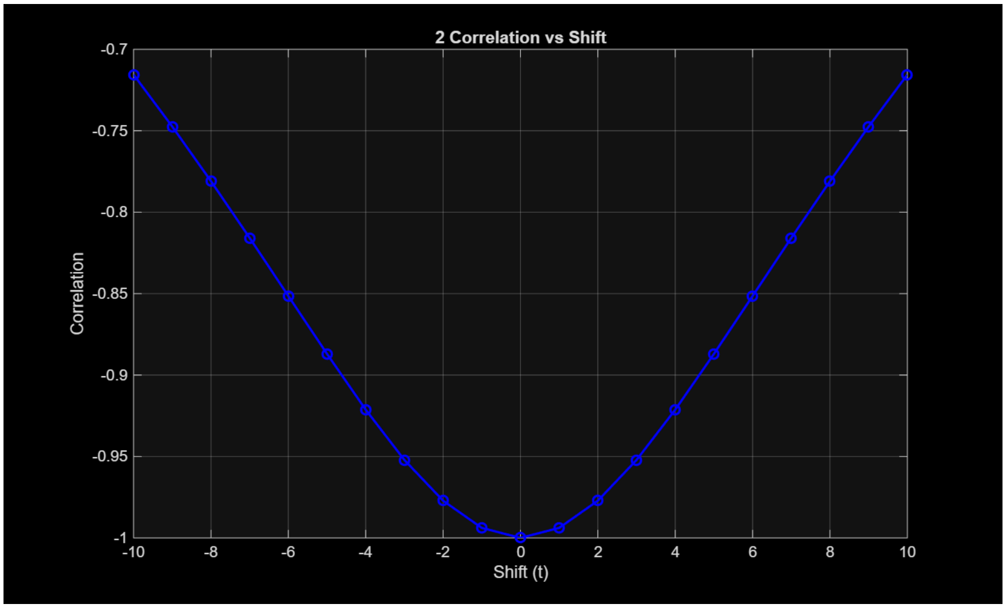

Figure 6: Correlation vs Shift (Case 2)

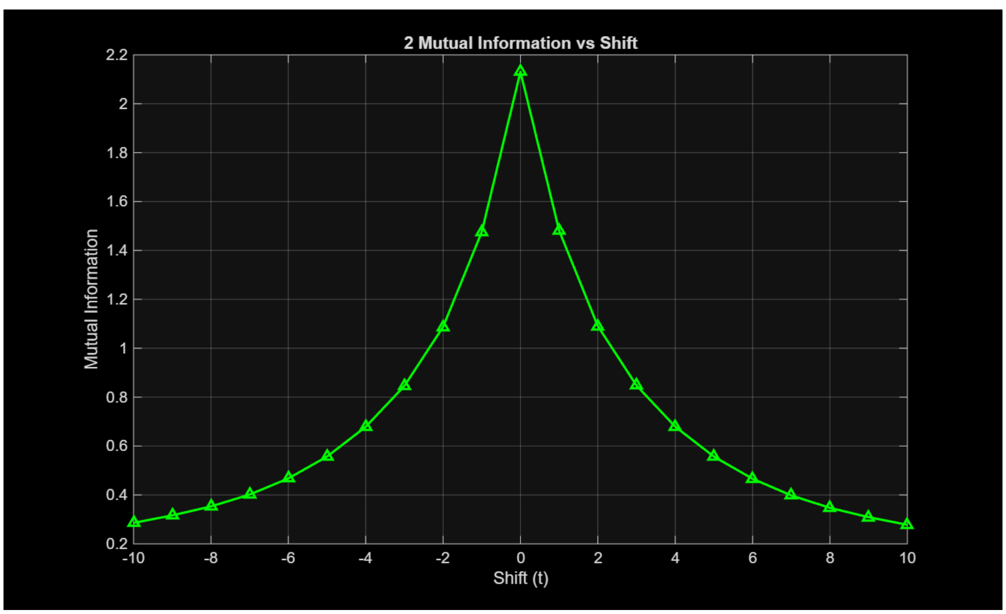

Figure 7: Mutual Information vs Shift (Case 2)

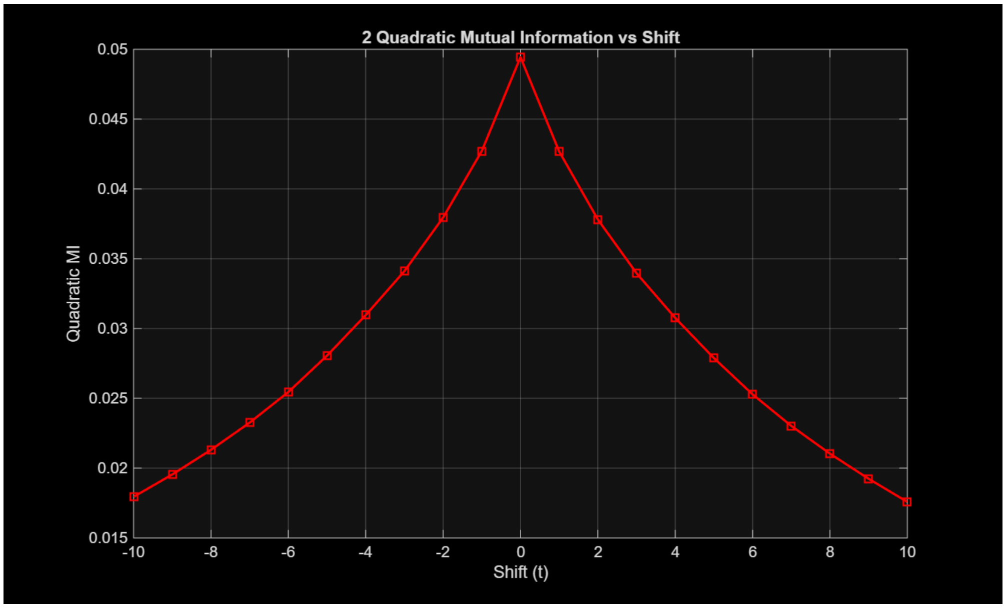

Figure 8: Quadratic Mutual Information vs Shift (Case 2)

Observation: Similar to Case 1, maximum dependence is around zero shift.
For plot 2 (2 Quadratic Mutual Information vs Shift), the Quadratic Mutual Information again peaks at a shift of 0, reflecting maximum dependence with perfect alignment, and decreases symmetrically with increasing shift, indicating reduced alignment.
For plot 2 (2 Mutual Information vs Shift), the Mutual Information peaks at a shift of 0, showing maximum dependence when images are aligned, and decreases symmetrically as the shift deviates, suggesting reduced alignment.
For plot 2 (2 Correlation vs Shift), the Correlation forms a parabolic curve with a minimum at shift 0, indicating that perfect alignment results in the lowest correlation, which increases as the shift moves away from 0.

(c) **Case 3**

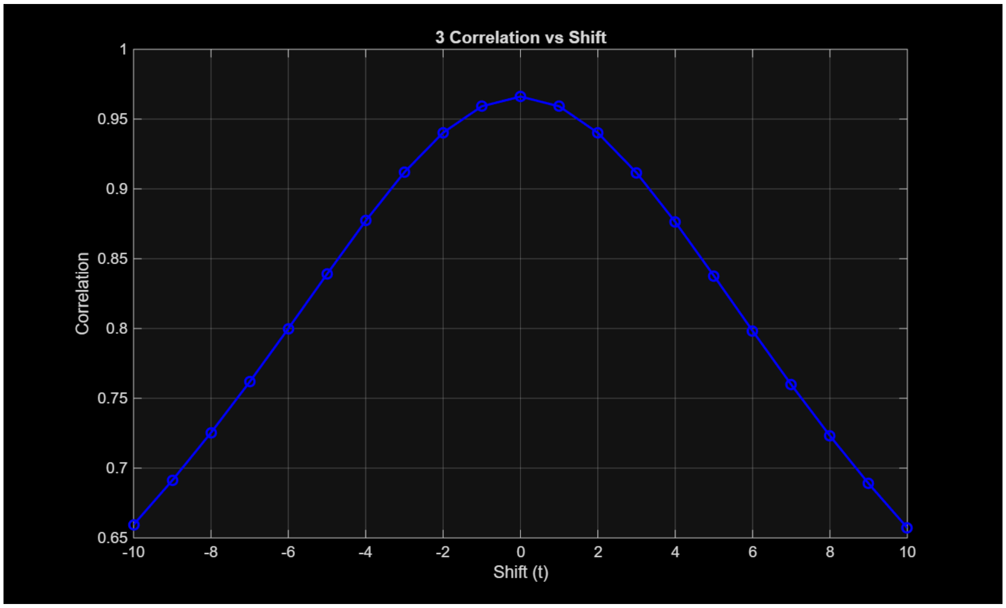

Figure 9: Correlation vs Shift (Case 3)

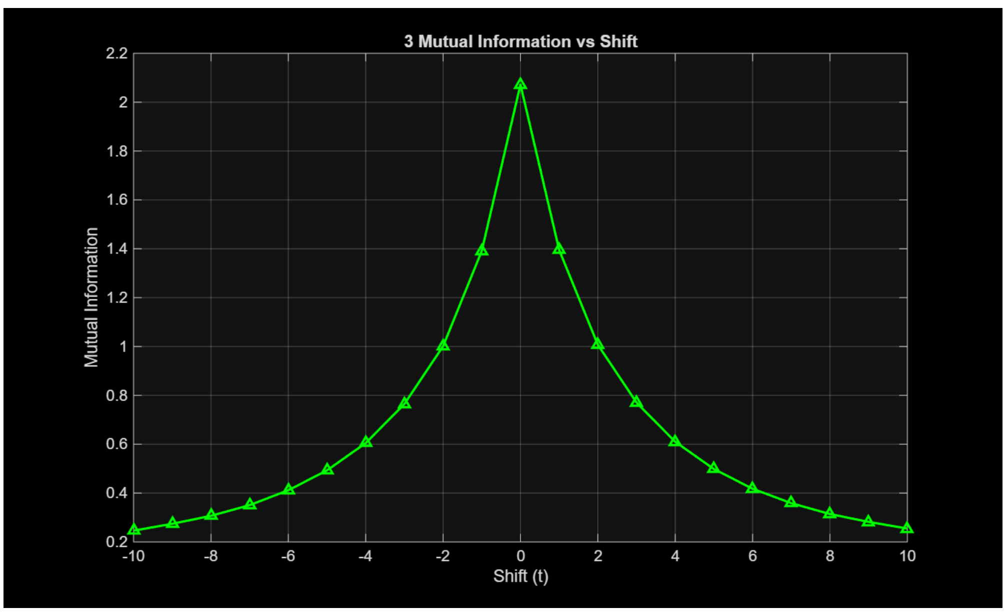

Figure 10: Mutual Information vs Shift (Case 3)

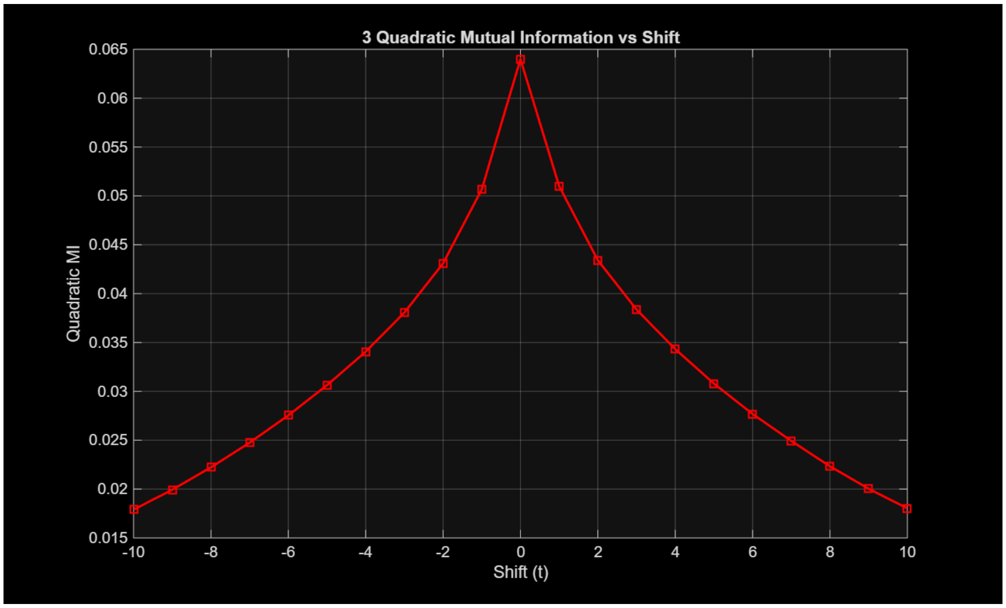

Figure 11: Quadratic Mutual Information vs Shift (Case 3)

Observation: Again, measures confirm correct alignment at zero shift.
For plot 3 (3 Quadratic Mutual Information vs Shift), the Quadratic Mutual Information peaks at a shift of 0, indicating maximum dependence with perfect alignment, and decreases symmetrically with increasing shift, reflecting reduced alignment.
For plot 3 (3 Mutual Information vs Shift), the Mutual Information peaks at a shift of 0, showing maximum dependence with perfect alignment, and decreases symmetrically as the shift increases or decreases, indicating reduced alignment.
For plot 3 (3 Correlation vs Shift), the Correlation peaks at a shift of 0, suggesting that perfect alignment results in maximum correlation, which decreases as the shift moves away from 0.
relationship between the dependence measures
The plots reveal the relationships between dependence measures (Quadratic Mutual Information, Mutual Information, and Correlation) across different shifts. For plot 1, Quadratic Mutual Information and Mutual Information both peak at a shift of 0, indicating maximum dependence with perfect alignment, while Correlation shows a minimum at shift 0, suggesting an inverse relationship with the information-based measures. In plot 2, Quadratic Mutual Information and Mutual Information again peak at shift 0, while Correlation exhibits a minimum, reinforcing the contrasting behavior where information measures align with perfect alignment, but Correlation does not. For plot 3, Quadratic Mutual Information and Mutual Information peak at shift 0, showing consistent dependence with alignment, whereas Correlation also peaks at shift 0, indicating a shift-specific alignment where all three measures align positively with perfect alignment. Overall, Quadratic Mutual Information and Mutual Information consistently indicate maximum dependence at zero shift, while Correlation's relationship varies, either inversely or positively, depending on the image pair

5. Given stuff you've learned in class, prove the following bounds:
$P(X \geq x) \leq e^{-tx}\phi_X(t)$ for $t > 0$, and $P(X \leq x) \leq e^{-tx}\phi_X(t)$ for $t < 0$. Here $\phi_X(t)$ represents the MGF of random variable X for parameter t. Now consider that X denotes the sum of n independent Bernoulli random variables $X_1, X_2, ..., X_n$ where $E(X_i) = p_i$. Let $\mu = \sum_{i=1}^{n} p_i$. Then show that $P(X > (1+\delta)\mu) \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$ for any $t \geq 0, \delta > 0$. You may use the inequality $1 + x \leq e^x$. Further show how to tighten this bound by choosing an optimal value of t. [15 points]
sol:

(a) proving $P(X \geq x) \leq e^{-tx}\phi_X(t)$ for $t > 0$
As e^{xt} is increasing function when $t > 0$ we can say

$$P(X \geq x) = P(e^{Xt} \geq e^{xt})$$

we know that $P(X \geq a) \leq \frac{E(X)}{a}$ for $+ve$ Random variable distribution(proved in classroom) As we see e^{Xt} is a random variable which only takes $+ve$ values so, we can apply above formula for that distribution,we get

$$P(e^{Xt} \geq e^{xt}) \leq \frac{E(e^{Xt})}{e^{xt}}$$

$$P(e^{Xt} \geq e^{xt}) \leq e^{-tx}E(e^{Xt})$$

from one of the above conclusions we can say that

$$P(X \geq x) \leq e^{-tx}E(e^{Xt})$$

as we know $\phi_X(t) = E(e^{Xt})$ by substituting this value in the above equation we get

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

(b) proving $P(X \leq x) \leq e^{-tx}\phi_X(t)$ for $t < 0$
As e^{xt} is decreasing function when $t < 0$ we can say

$$P(X \leq x) = P(e^{Xt} \geq e^{xt})$$

we know that $P(X \geq a) \leq \frac{E(X)}{a}$ for $+ve$ Random variable distribution(proved in classroom) As we see e^{Xt} is a random variable which only takes $+ve$ values so, we can apply above formula for that distribution,we get

$$P(e^{Xt} \geq e^{xt}) \leq \frac{E(e^{Xt})}{e^{xt}}$$

$$P(e^{Xt} \geq e^{xt}) \leq e^{-tx}E(e^{Xt})$$

from one of the above conclusions we can say that

$$P(X \leq x) \leq e^{-tx}E(e^{Xt})$$

as we know $\phi_X(t) = E(e^{Xt})$ by substituting this value in the above equation we get

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

(c) proving $P(X > (1+\delta)\mu) \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}$ for any $t \geq 0, \delta > 0$u

$$P(X > (1+\delta)\mu) \leq P(X \geq (1+\delta)\mu)$$

from the (a) part we can write ,

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

we know that $X = \sum_{i=1}^{n} X_i$ and X_i are independent of each other from which we can say that e^{X_i}) are also independent of each other

$$\phi_X(t) = E(e^{Xt})$$

$$\phi_X(t) = E(e^{t(\sum_{i=1}^{n} X_i)})$$

$$\phi_X(t) = E(\prod_{i=1}^{n} e^{X_i t})$$

As they are independent

$$\phi_X(t) = \prod_{i=1}^{n} E(e^{X_i t})$$

Given that X_i is Bernoulli random variable which has $\phi_X(t) = E(e^{Xt}) = (1 - p + pe^t)$ where p is the expected value of the random variable(from class).
By this formula we get,

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

we are given a inequality $1 + x \leq e^x$ by applying this over $(1 + p_i(e^t - 1))$ we get,

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

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

$$\phi_X(t) \leq e^{(e^t-1)\mu} \tag{2}$$

from both equation 1 and 2 we get ,

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

6. Consider that there are n independent coin tosses with heads probability p. Let T be the trial number at which you get the first heads. Derive a formula for E(T) in terms of p and/or n. [10 points]
sol:
Let probability that we get first heads at i^{th} toss $= P(i)$
For first head to come at i^{th} toss gives only two constraints that are

   (a) all the coin tosses before the i^{th} toss should be tails and
   (b) and the i^{th} toss should be head.

There are no constraints on the tosses after i^{th} one as they can be head or toss it won't change that first head is at i^{th} toss.
From the above constraints we can say that

$$P(i) = (1-p)^{i-1} * p$$

As the $(1-p)^{i-1}$ is for the constraint (a) , and the p is for the constraint (b).
Then

$$E(T) = \sum_{i=1}^{n} i * P(i)$$

by substituting the value of $P(i)$ in the $E(T)$ we get

$$E(T) = \sum_{i=1}^{n} i * (1-p)^{i-1} * p$$

Just expanding the above summation we get

$$E(T) = p(1 * (1-p)^{(1-1)} + 2(1-p)^{(2-1)} + ... + n(1-p)^{(n-1)})$$

by simplifying the above equation more we get the below equation
The below equation is of the format A.G.P(Arithmetic and Geometric Progression) we solve by subtracting the $E(T)(1-p)$ to the $E(T)$ by shifting one term of $E(T)(1-p)$ to right

$$\frac{E(T)}{p} = 1 * (1-p)^0 + 2(1-p)^1 + ... + n(1-p)^{(n-1)} + 0$$

$$-$$

$$\frac{E(T) * (1-p)}{p} = \quad 0 + 1 * (1-p)^1 + 2 * (1-p)^2 + ... + n * (1-p)^n$$

By doing this we get

$$\begin{aligned}\frac{E(T) - E(T) * (1-p)}{p} &= \{1-0\} + \{2(1-p)^1 - 1 * (1-p)^1\} + ...+ \\ &\{n(1-p)^{n-1} - (n-1) * (1-p)^{n-1}\}+ \\ &\{0 - n * (1-p)^n\}\end{aligned}$$

by further simplification we get

$$\frac{E(T)(1-1+p)}{p} = \{1 + (1-p)^1 + (1-p)^2 + .. + (1-p)^{n-1}\} - \{n * (1-p)^n\}$$

using the G.P formula which is $\sum_{i=0}^{r-1} a^i = \frac{1-a^r}{1-a}$

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

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