CS 215 Assignment 2 Report

Diwan Anuj Jitendra - 170070005   Soumya Chatterjee - 170070010

August 22, 2018

1 Problem 1

1.1 PDF of $Z = X + Y$

In general:
Consider $Z = X + Y$. Since X and Y are random variables, so is Z. We are required to find the probability density function of Z. Consider z in the domain of Z. We must find pairs of x and y such that $z = x + y$. For every x, we get y=$z - x$. Thus, multiple pairs $(x, z - x)$ sum up to z, each with the joint probability $f_{XY}(x, z - x)$
For the discrete case:

$$f_Z(z) = \sum_{x \in \text{domain of X}} f_{XY}(x, z - x)$$

For the continuous case:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x)\,\mathrm{d}x$$

When X and Y are independent:
In this case, we can use $f_{XY}(x, y) = f_X(x)f_Y(y)$ by definition of independent random variables to write:
For the discrete case:

$$f_Z(z) = \sum_{x \in \text{domain of X}} f_X(x)f_Y(z - x)$$

For the continuous case:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)\,\mathrm{d}x$$

1.2 $P(X \leq Y)$

In general:
We can see that given $Y = y$, $P(X \leq Y)$ is the sum of the probabilities $P(X = x, Y = y) = f_{XY}(x, y)$ such that $x \leq y$. Then, we must also additionally sum over all the y. Thus,
For the discrete case:

$$P(X \leq Y) = \sum_{y \in \text{Domain of Y}} \sum_{x\ , x \leq y} f_{XY}(x, y)$$

For the continuous case:

$$P(X \leq Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{XY}(x, y)\ \mathrm{d}x\mathrm{d}y$$

When X and Y are independent:
In this case, we can use $f_{XY}(x, y) = f_X(x)f_Y(y)$ by definition of independent random variables to write:
For the discrete case:

$$P(X \leq Y) = \sum_{y \in \text{Domain of Y}} \sum_{x\ , x \leq y} f_X(x)f_Y(y)$$

For the continuous case:

$$P(X \leq Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_X(x) f_Y(y) \; \mathrm{d}x \mathrm{d}y$$

It can be simplified further using CDF, but since the problem asks to express only in terms of f_X, f_Y, f_{XY}, it has not been done here.

2 Problem 2

Given $X_1, X_2, \dots X_n$ are independent and identically distributed random variables with cdf $F_X(x)$ and pdf $f_X(x) = F'_X(x)$

$$Y_1 = \max(X_1, X_2, \dots X_n)$$

cdf of Y_1,

$$\begin{aligned}
F_{Y_1}(x) &= P(Y_1 \leq x) \\
&= P(\max(X_1, X_2, \dots X_n) \leq x) \\
&= P(X_1 \leq x, X_2 \leq x, \dots X_n \leq x) \\
&= P(X_1 \leq x) P(X_2 \leq x) \dots P(X_n \leq x) && (X_i \text{ are independent}) \\
&= F_X(x) F_X(x) \dots F_X(x) \\
&= [F_X(x)]^n
\end{aligned}$$

pdf of Y_1,

$$\begin{aligned}
f_{Y_1}(x) &= [F_{Y_1}(x)]' \\
&= [\{F_X(x)\}^n]' \\
&= n[F_X(x)]^{n-1}[F'_X(x)] \\
&= n[F_X(x)]^{n-1} f_X(x)
\end{aligned}$$

Now for

$$Y_2 = \min(X_1, X_2, \dots X_n)$$

cdf of Y_2,

$$\begin{aligned}
P(Y_2 > x) &= P(\min(X_1, X_2, \dots X_n) > x) \\
&= P(X_1 > x, X_2 > x, \dots X_n > x) \\
&= P(X_1 > x) P(X_2 > x) \dots P(X_n > x) && (X_i \text{ are independent}) \\
&= (1 - P(X_1 \leq x))(1 - P(X_2 \leq x)) \dots (1 - P(X_n \leq x)) \\
&= (1 - F_X(x))(1 - F_X(x)) \dots (1 - F_X(x)) \\
&= [1 - F_X(x)]^n \\
\therefore F_{Y_2}(x) &= P(Y_2 \leq x) \\
&= 1 - P(Y_2 > x) \\
&= 1 - [1 - F_X(x)]^n
\end{aligned}$$

pdf of Y_2,

$$\begin{aligned}
f_{Y_2}(x) &= [F_{Y_2}(x)]' \\
&= [1 - \{1 - F_X(x)\}^n]' \\
&= -n[1 - F_X(x)]^{n-1}[-F_X(x)]' \\
&= n[1 - F_X(x)]^{n-1}[F'_X(x)] \\
&= n[1 - F_X(x)]^{n-1} f_X(x)
\end{aligned}$$

3 Problem 3

Markov's inequality states that for X takes take non-negative values and $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

If X is a random variable, $(X - \mu + b)^2$ is a non-negative random variable for some arbitrary $b \geq 0$. Since $(\tau + b)^2 > 0$, we have

$$P\left((X - \mu + b)^2 \geq (\tau + b)^2\right) \leq \frac{E\left((X - \mu + b)^2\right)}{(\tau + b)^2}$$

$$\begin{aligned}
E\left((X - \mu + b)^2\right) &= E\left((X - \mu)^2 + 2b(X - \mu) + b^2\right) \\
&= E((X - \mu)^2) + E(2b(X - \mu)) + b^2 \qquad \text{(By linearity of expectation)} \\
&= \sigma^2 + 2bE(X - \mu) + b^2 \\
&= \sigma^2 + b^2 \qquad (E(X - \mu) = E(X) - \mu = \mu - \mu = 0)
\end{aligned}$$

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

Consider,

$$\begin{aligned}
J(b) &= \frac{\sigma^2 + b^2}{(\tau + b)^2} \\
J'(b) &= \frac{(\tau + b)^2 \times 2b - (\sigma^2 + b^2) \times 2(\tau + b)}{(\tau + b)^4} \\
J'(b) &= 0 \\
\Rightarrow (\tau + b)b &= \sigma^2 + b^2 \\
\Rightarrow b &= \frac{\sigma^2}{\tau} \\
J\left(\frac{\sigma^2}{\tau}\right) &= \frac{\sigma^2}{\tau^2 + \sigma^2}
\end{aligned}$$

Thus, the RHS attains a minimum at $\frac{\sigma^2}{\tau^2+\sigma^2}$ (using the second derivative)
Now, we can show that for $z \geq 0$, $P(X \geq z) \leq P(X^2 \geq z^2)$, since:

$$\begin{aligned}
P(X^2 \geq z^2) &= P(X \geq z \text{ or } X \leq -z) \\
&\geq P(X \geq z)
\end{aligned}$$

For $\tau > 0$ and $b \geq 0$,

$$\begin{aligned}
P(X - \mu \geq \tau) &= P(X - \mu + b \geq \tau + b) \\
&\leq P\left((X - \mu + b)^2 \geq (\tau + b)^2\right) \\
&\leq \frac{E\left((X - \mu + b)^2\right)}{(\tau + b)^2}
\end{aligned}$$

Since b can be chosen arbitrarily, we can choose b such that the RHS is minimized. We have already found this minima at $b = \frac{\sigma^2}{\tau}$ Hence

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

For $\tau < 0$, consider $\alpha = -\tau > 0$

$$\begin{aligned}
P(X-\mu<\tau) &= P\left(-(X-\mu)>\alpha\right) \\
&\leq \frac{\sigma^2}{\alpha^2+\sigma^2} = \frac{\sigma^2}{\tau^2+\sigma^2} \quad \text{(Using the just proved result, on } -(X-\mu)) \\
\therefore P(X-\mu<\tau) &\leq \frac{\sigma^2}{\tau^2+\sigma^2} \\
P(X-\mu\geq\tau) &= 1-P(X-\mu<\tau) \\
\therefore P(X-\mu\geq\tau) &\geq 1-\frac{\sigma^2}{\tau^2+\sigma^2}
\end{aligned}$$

4 Problem 4

Markov's inequality states that for $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

We know that if X is a random variable then for some t, e^{tX} is also a random variable. Since $e^{tx} > 0$,

$$\begin{aligned}
P(e^{tX} \geq e^{tx}) &\leq \frac{E(e^{tX})}{e^{tx}} \\
P(e^{tX} \geq e^{tx}) &\leq e^{-tx}\phi_X(t) \\
P(tX \geq tx) &\leq e^{-tx}\phi_X(t) && \text{(Since } \forall_y, e^y > 0)
\end{aligned}$$

Now the following two cases arise, depending on the value of t

1. If $t > 0$, then $tX \geq tx \iff X \geq x$, hence
$$P(X \geq x) \leq e^{-tx}\phi_X(t) \tag{1}$$
2. If $t < 0$, then $tX \geq tx \iff X \leq x$, hence
$$P(X \leq x) \leq e^{-tx}\phi_X(t)$$

Given that $X_1, X_2 \dots X_n$ are independent Bernoulli random variables, their sum $X = X_1 + X_2 \dots X_n$ is also a random variable. Also $E(X_i) = p_i$ and $\mu = \sum_{i=1}^{n} p_i$.

From equation 1, for $x = (1+\delta)\mu$ and some $t \geq 0$, we have,

$$P(X > (1+\delta)\mu) \leq \frac{\phi_X(t)}{e^{(1+\delta)t\mu}} \tag{2}$$

Since X_i are independent, we have

$$\begin{aligned}
\phi_X(t) &= \phi_{X_1}(t)\phi_{X_2}(t)\dots\phi_{X_n}(t) \\
\phi_{X_i}(t) &= 1-p_i+p_ie^t && \text{(MGF for a Bernoulli r.v.)} \\
&= 1+p_i(e^t-1) \\
&\leq e^{p_i(e^t-1)} && (1+x \leq e^x \text{ for } x = p_i(e^t-1)) \\
\therefore \phi_X(t) &\leq e^{p_1(e^t-1)}e^{p_2(e^t-1)}\dots e^{p_n(e^t-1)} \\
&\leq e^{\sum_{i=1}^{n} p_i(e^t-1)} \\
&\leq e^{\mu(e^t-1)}
\end{aligned}$$

So we obtain that

$$\phi_X(t) \le e^{\mu(e^t-1)} \tag{3}$$

From equation 2 and 3, we have

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

To find a tighter bound, consider

$$\begin{aligned}
J(t) &= \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}} \\
J'(t) &= \frac{e^{(1+\delta)t\mu} e^{\mu(e^t-1)} \mu e^t - e^{\mu(e^t-1)} e^{(1+\delta)t\mu}(1+\delta)\mu}{[e^{(1+\delta)t\mu}]^2} \\
&= \frac{\mu e^{\mu(e^t-1)}(e^t-(1+\delta))}{e^{(1+\delta)t\mu}} \\
J'(t) &= 0 \\
\Rightarrow e^t &= 1+\delta \\
t &= \ln(1+\delta) \\
J''(t) &= \frac{\mu e^{\mu\left(e^t-1\right)}\left(\mu e^t\left(e^t-2(1+\delta)\right)+e^t+(\delta+1)^2\mu\right)}{e^{(1+\delta)t\mu}} \\
J''\left(\ln(1+\delta)\right) &= \frac{\mu e^{\mu\delta}(1+\delta)}{(1+\delta)^{\mu(1+\delta)}} \\
&> 0
\end{aligned}$$

Therefore $t = \ln(1+\delta)$ is indeed a minimum.

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

5 Problem 5

In the first method, we have to do k tests irrespective of the value of p. Hence, for the random variable X_1 denoting the number of tests,

$$E(X_1) = k$$

Let X_2 be the random variable denoting the number of tests done in the second method and $P(x_i)$ be the probability that the i^{th} person does not suffer from the disease. Now the following two cases are possible:

1. If nobody suffers from the disease, then the required number of tests is 1 and the probability of this happening is given by $P(x_1x_2x_3\dots x_k)$. Since it is given that any person having the disease is independent of any other person,

$$\begin{aligned}
P(x_1x_2x_3\dots x_k) &= P(x_1)P(x_2)\dots P(x_k) \\
&= (1-p)(1-p)\dots(1-p) \\
&= (1-p)^k
\end{aligned}$$

2. If atleast one person suffers from the disease, $k+1$ tests will have to be performed. The probability of this happening is,

$$\begin{aligned}
P(x_1^C \cup x_2^C \dots \cup x_k^C) &= 1 - P(x_1x_2x_3\dots x_k) \\
&= 1-(1-p)^k
\end{aligned}$$

Hence, for the random variable X_2, we have,

$$X_2 = \begin{cases} 1 & \text{with probability } (1-p)^k \\ k+1 & \text{with probability } 1-(1-p)^k \end{cases}$$

Therefore,

$$\begin{aligned} E(X_2) &= 1 \times (1-p)^k + (k+1) \times (1-(1-p)^k) \\ &= (k+1) - k(1-p)^k \end{aligned}$$

We require that,

$$\begin{aligned} E(X_2) &< E(X_1) \\ (k+1) - k(1-p)^k &< k \\ 1 &< k(1-p)^k \\ 1/k &< (1-p)^k \\ (1/k)^{1/k} &< 1-p \\ p &< 1 - \left(\frac{1}{k}\right)^{\frac{1}{k}} \end{aligned} \tag{1}$$

For such values of p, $E(X_2) < E(X_1)$ Now, $\left(\frac{1}{k}\right)^{\frac{1}{k}}$ increases with k for $k \geq 3$, hence the minimum value of $1 - \left(\frac{1}{k}\right)^{\frac{1}{k}}$ is achieved at either 2 or 25. For $k = 2$, we have

$$p < 1 - \left(\frac{1}{2}\right)^{\frac{1}{2}} = 0.292893218813$$

And for $k = 25$,

$$p < 1 - \left(\frac{1}{25}\right)^{\frac{1}{25}} = 0.120810688459$$

Hence for any $p < 0.1208$, the second method will have

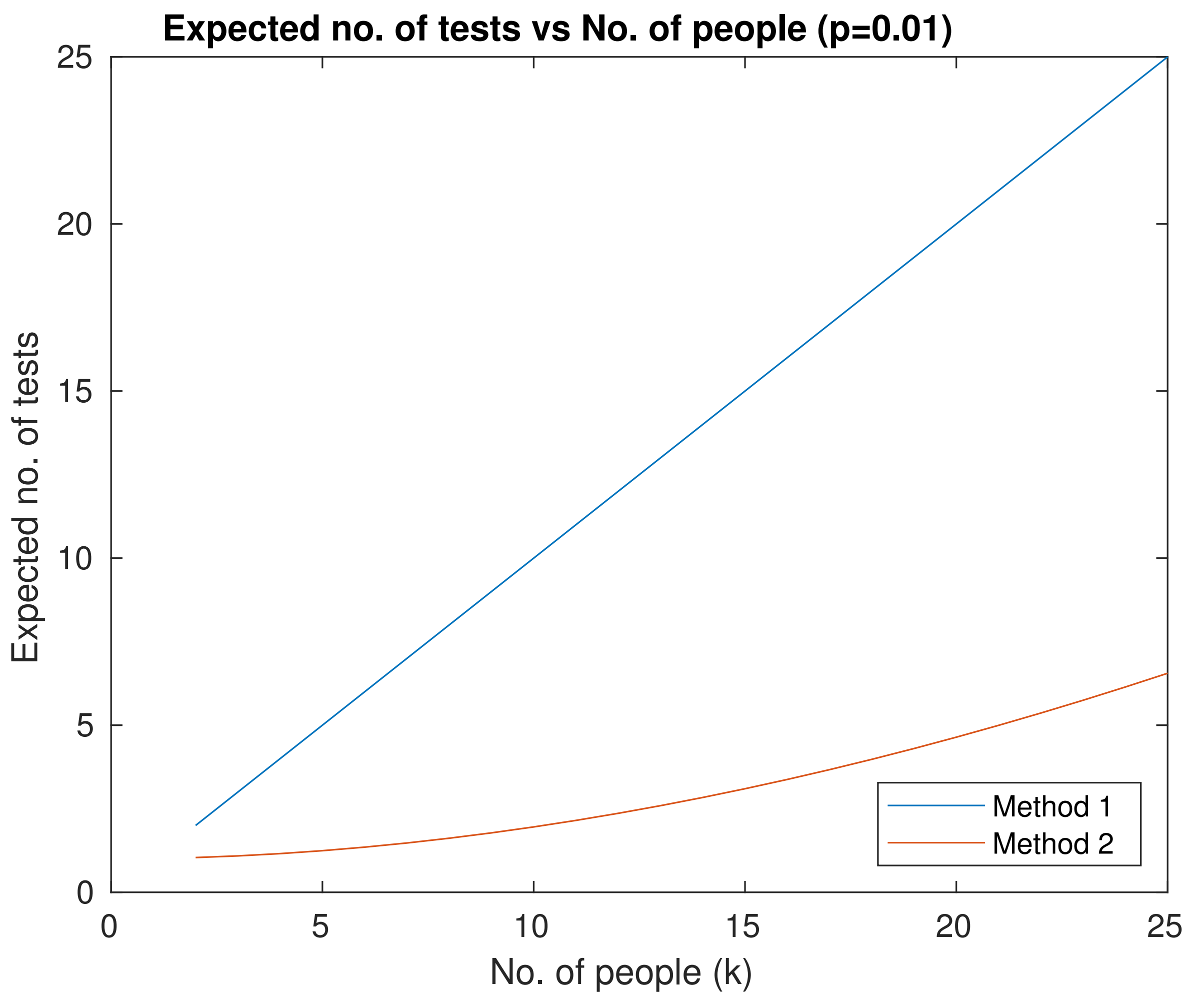
$$E(X_2) < E(X_1) \quad \forall k \in [2, 25]$$

From the following plots, we can observe that as the probability of a person having the disease p decreases, the advantage of the second method (pooled method) increases.

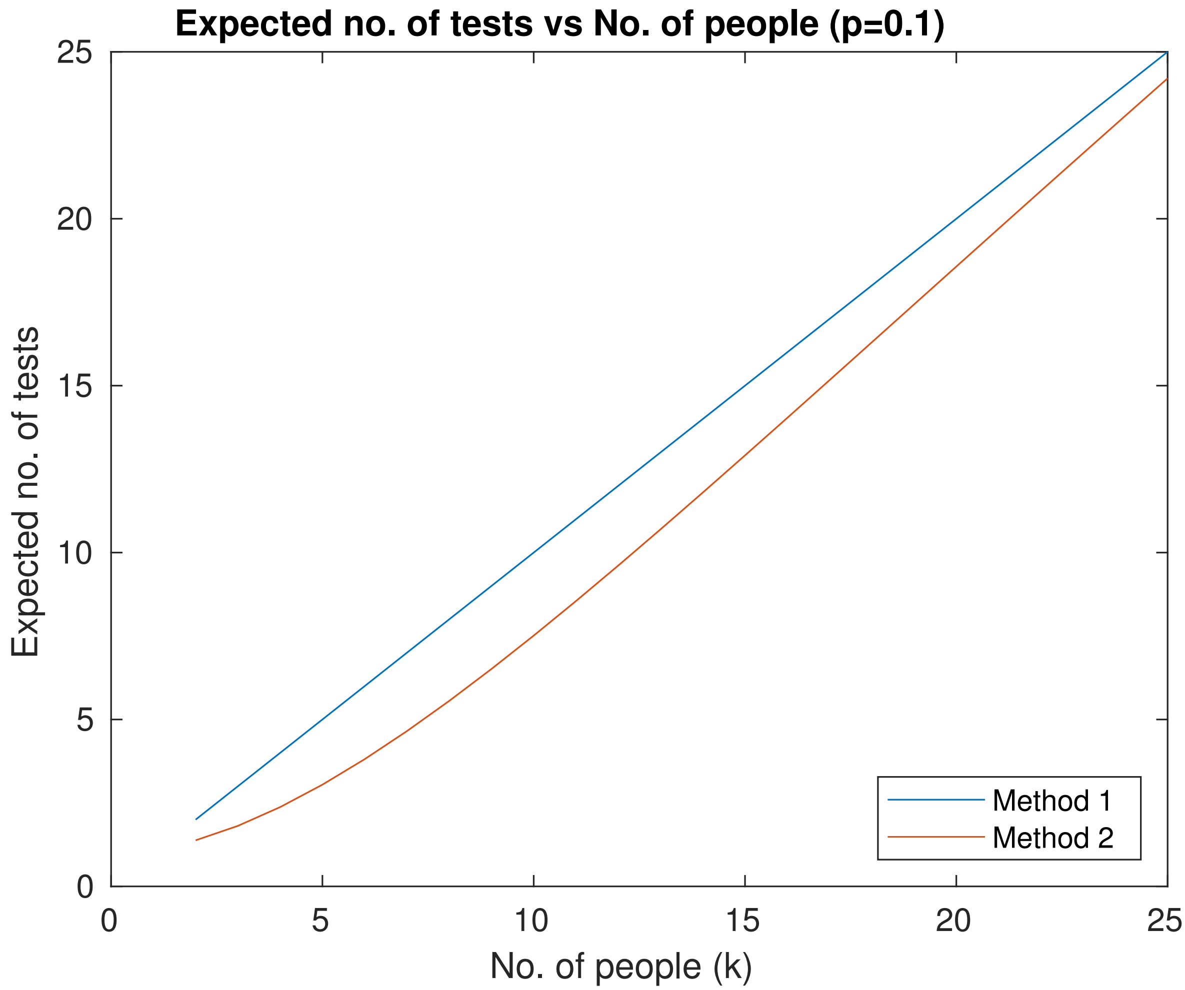
For example, with $p = 0.01$ and $k = 25$, we can infer from the plot that the expected number of tests for the first method is $E(X_1) = 25$ while for second method $E(X_2) \approx 7$ which is significantly lesser. However, for larger values of $p(\approx 0.1)$, the difference is negligible. This happens because the lower expected value of the second test arises from the fact that if none of the people had the disease, it takes just 1 test. As the probability of none of the people having the disease decreases, the second test's advantage decreases.
Steps to run the code for generating plots:

1. In the code, set the two values of p to the desired values. By default, they have been set to 0.01 and 0.1.

2. Run the code `p5.m`



Expected no. of tests vs No. of people (p=0.01)
Expected no. of tests
No. of people (k)
Method 1
Method 2
0
5
10
15
20
25


Expected no. of tests vs No. of people (p=0.1)
Expected no. of tests
No. of people (k)
0
5
10
15
20
25
Method 1
Method 2

6 Problem 6

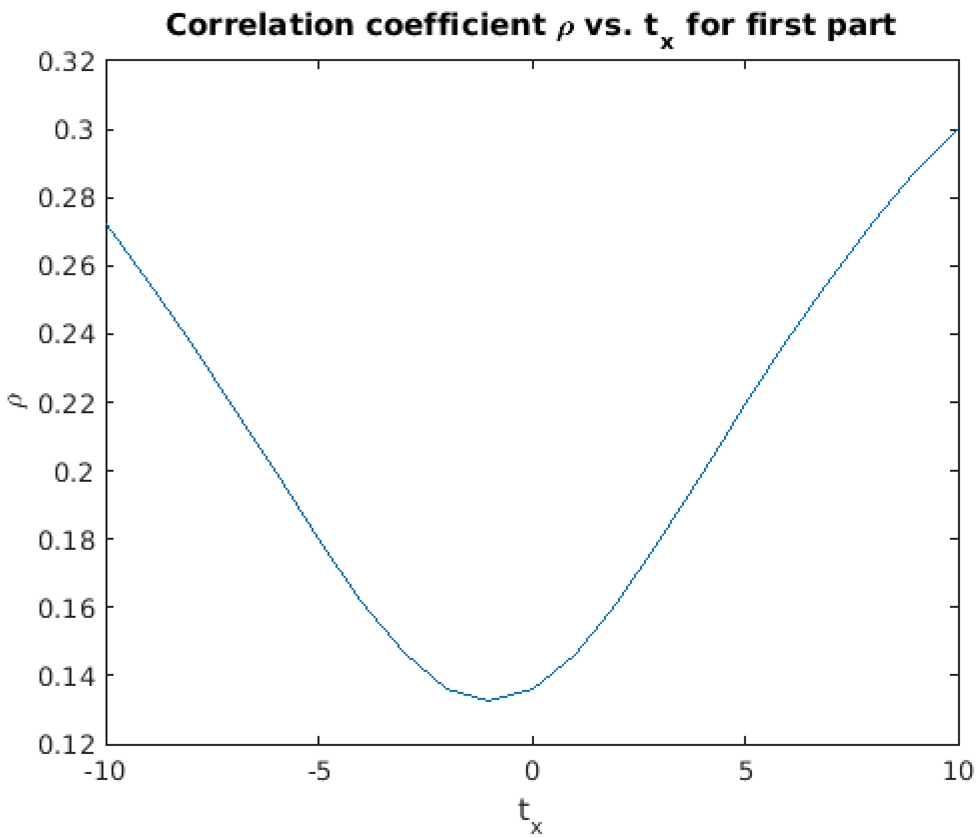
6.1 How to run

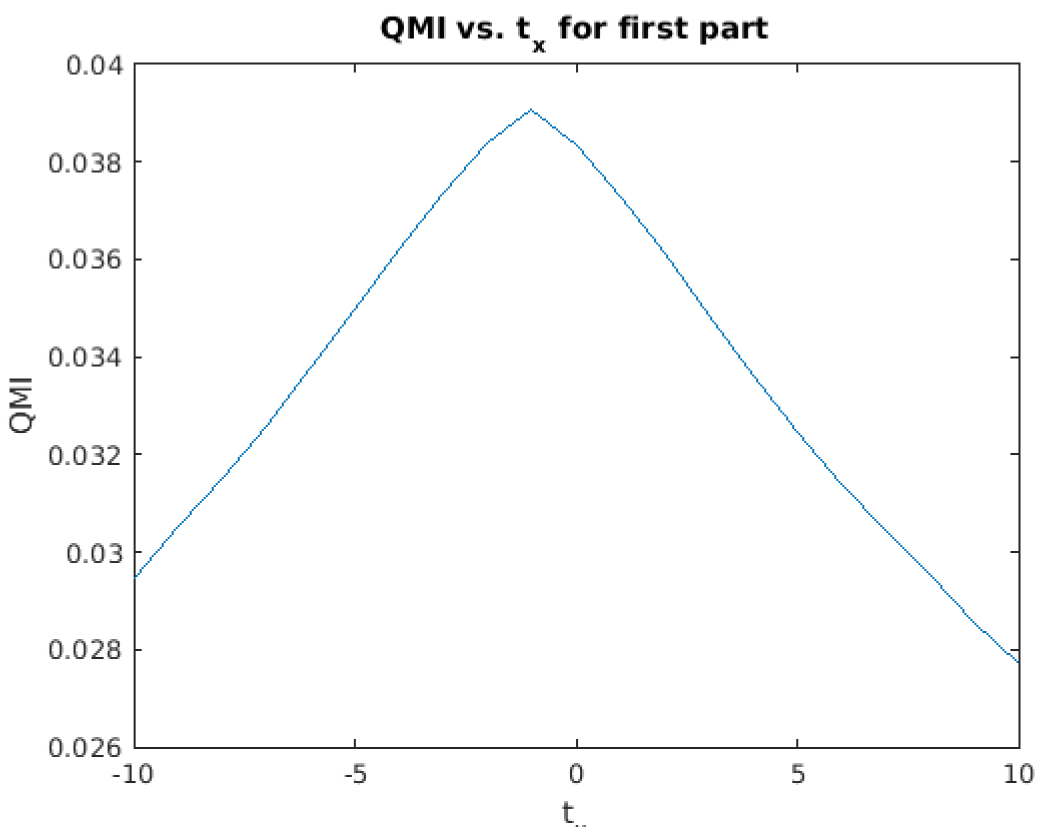
Run `p6.m` in MATLAB. The four plots will appear. Ensure that `T1.png` and `T2.png` are in the same directory as `p6.m`.

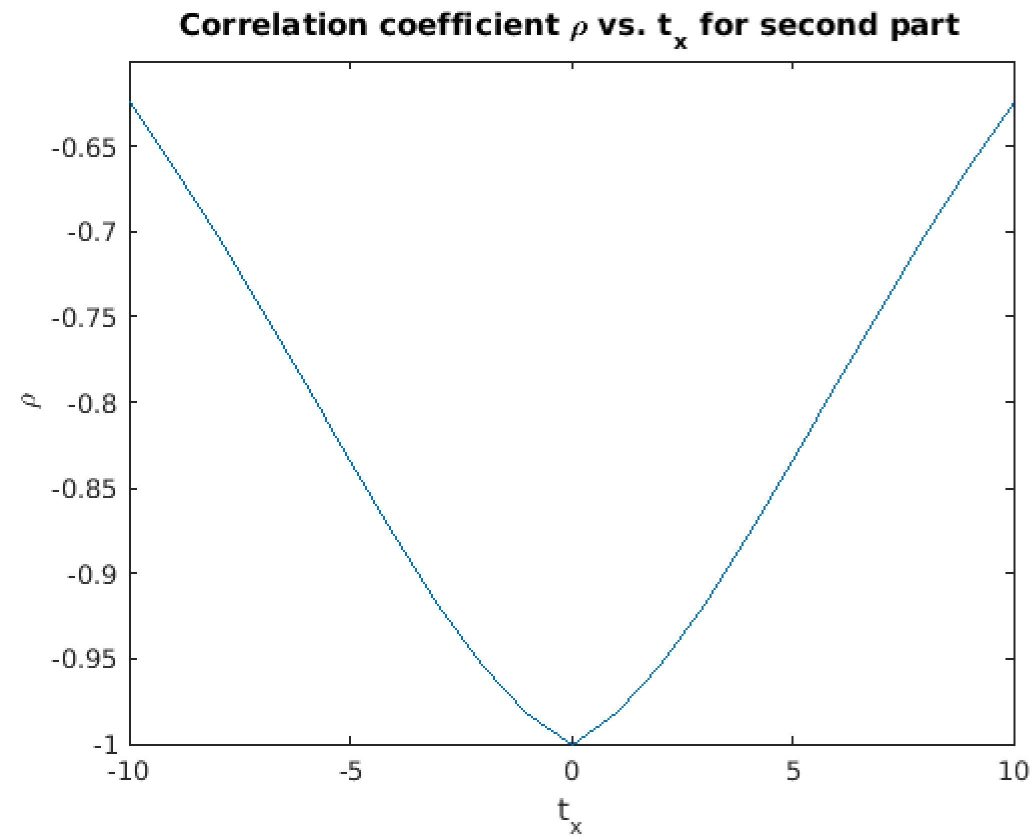
6.2 Plots and comments

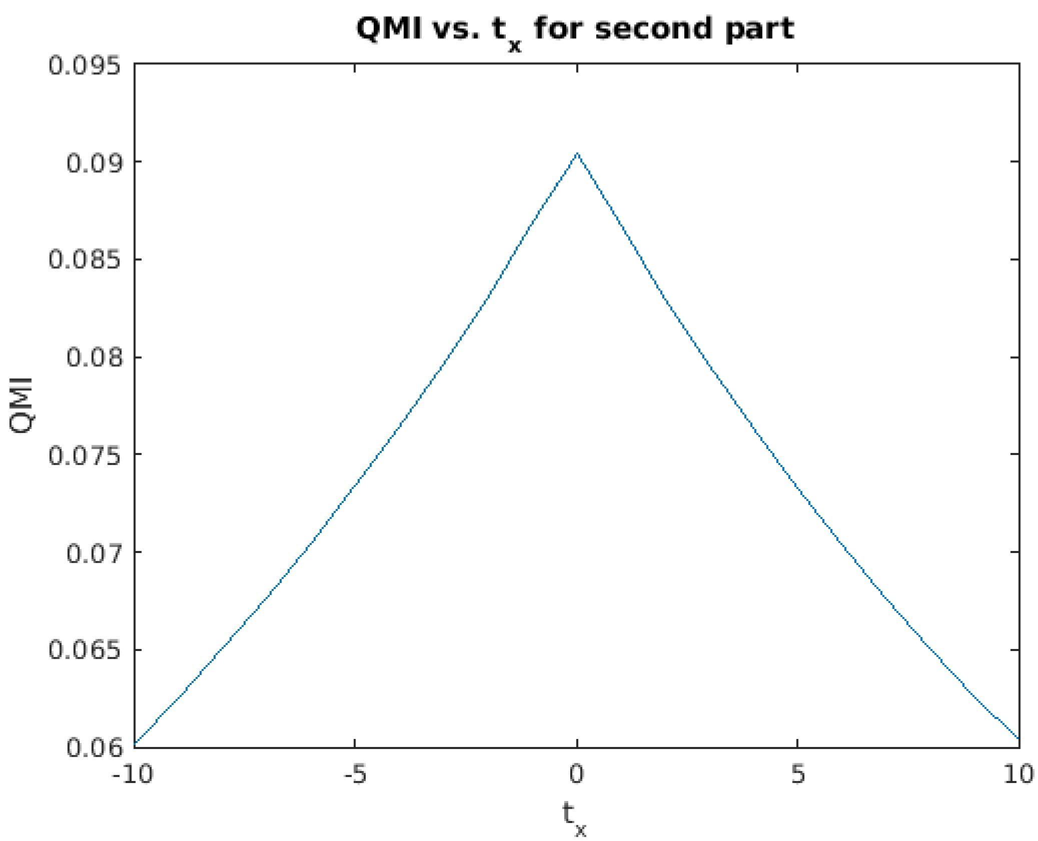
(Plots on the next 2 pages)
First, we consider the first part i.e using the images `T1.png` and `T2.png`. The two images intuitively have a very strong correlation between them(they are images of the same human brain, taken with different settings). The two images are misaligned by a few pixels. To capture this correlation quantitatively, and to find whether the images are not completely aligned, we use two measures: the correlation coefficient and the QMI and observe the plots.
The correlation coefficient plot shows positive values of ρ for all values of t_x, is concave upwards, and has a minimum of around 0.13 at a value around -1. The correlation coefficient plot thus seems to show low correlation between the two images(near 0) and implies the images are least correlated when the second image is offset by -1 pixel.
The QMI plot, on the other hand, is concave downward and shows a maximum near -1 ! The QMI measures how much information the two images share more accurately than the correlation coefficient, and thus correctly analyzes the two images and gives a high correlation when the second image is offset by -1 pixel. This implies that the image 1 is perfectly aligned with image 2 moved to the left by 1 pixel.
Here the QMI and correlation coefficient seem to give contradictory results. In fact, the correlation coefficient only looks at the mean and the variance while the QMI uses the joint pdf and hence looks at the entire distribution 'intelligently'. The QMI is hence a more robust measure than correlation coefficient. The correlation coefficient here fails and associates lower values of correlation with actual higher values of correlation. The QMI thus also predicts that the two images can be aligned by moving image 2 to the left by approx 1 pixel.

Now, in the second part($I_2 = 255 - I_1$), the correlation coefficient plot is symmetric, and shows complete negative correlation ($\rho = 1$) when $t_x = 0$ and shows lesser negative correlation for other values of t_x. This is as expected, since the image is a negative of the other, it should have very high correlation(in the opposite i.e negative direction), and for offsets this correlation decreases since both images are perfectly aligned and correlated when there's no offset by a simple relation $I_2 = 255 - I_1$.
The QMI plot is also as expected, showing a maxima at $t_x = 0$ i.e. indicating maximum correlation, similar to the correlation coefficient. Note that QMI is always non-negative and shows only magnitude of correlation, not the sign(positively or negatively correlated). Here since alignment was achieved at 0, both plots had extrema at 0.


Correlation coefficient ρ vs. t_x for first part
ρ
0.32
0.3
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
-10
-5
0
5
10
t_x


QMI vs. t_x for first part
QMI
0.04
0.038
0.036
0.034
0.032
0.03
0.028
0.026
-10
-5
0
5
10
t_x


Correlation coefficient ρ vs. t_x for second part
ρ
-0.65
-0.7
-0.75
-0.8
-0.85
-0.9
-0.95
-1
-10
-5
0
5
10
t_x


QMI vs. t_x for second part
QMI
0.095
0.09
0.085
0.08
0.075
0.07
0.065
0.06
-10
-5
0
5
10
t_x