Assignment 2: CS 215

Devansh Jain	Harshit Varma
190100044	190100055

September 14, 2020

Contents

Question 1

Given that $X_1, \ldots, X_n$ are n Independent Identically distributed random variables with cdf $F_X(x)$ and pdf $f_X(x) = F'_X(x)$.

To determine cdf and pdf for $Y_1 = \max(X_1, \ldots, X_n)$

$$
\begin{aligned}
F_{Y_1}(x) &= P(Y_1 \le x) \\
&= \prod_{i=1}^{n} P(X_i \le x) \qquad (\text{As } Y_1 = \max(X_i\text{s}),\ Y_1 \ge X_i\ \forall i) \\
&= P(X \le x)^n \qquad (X_i\text{s are identically distributed}) \\
&\boxed{F_{Y_1}(x) = F_X(x)^n} \\
&\boxed{f_{Y_1}(x) = n \cdot F_X(x)^{n-1} \cdot f_X(x)} \qquad (\text{Differentiation w.r.t } x)
\end{aligned}
$$

To determine cdf and pdf for $Y_2 = \min(X_1, \ldots, X_n)$

$$
\begin{aligned}
F_{Y_2}(x) &= P(Y_2 \le x) = 1 - P(Y_2 \ge x) \\
&= 1 - \prod_{i=1}^{n} P(X_i \ge x) \qquad (\text{As } Y_2 = \min(X_i\text{s}),\ Y_2 \le X_i\ \forall i) \\
&= 1 - (P(X \ge x))^n \qquad (X_i\text{s are identically distributed}) \\
&= 1 - (1 - P(X \le x))^n \\
&\boxed{F_{Y_2}(x) = 1 - (1 - F_X(x))^n} \\
&\boxed{f_{Y_2}(x) = n \cdot (1 - F_X(x))^{n-1} \cdot f_X(x)} \qquad (\text{Differentiation w.r.t } x)
\end{aligned}
$$

The cdf and pdf of $Y_1 = \max(X_1, \ldots, X_n)$ are $F_X(x)^n$ and $n \cdot F_X(x)^{n-1} \cdot f_X(x)$ respectively.
The cdf and pdf of $Y_2 = \min(X_1, \ldots, X_n)$ are $1 - (1 - F_X(x))^n$ and $n \cdot (1 - F_X(x))^{n-1} \cdot f_X(x)$ respectively.

Question 2

Part 1

It's given that X belongs to a Gaussian Mixture Model.

$$f_X(x) = \sum_{i=1}^{k} p_i \mathcal{N}(\mu_i, \sigma_i^2)$$

Throughout the rest of this question, we shall use G_i for denoting $\mathcal{N}(\mu_i, \sigma_i^2)$.

Calculating E(X):

We know that for an arbitrary random variable Y which is distributed as a Gaussian, ($Y \sim G_i$),
$E(Y) = \int_{-\infty}^{\infty} xG_i dx = \mu_i$
Thus,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{-\infty}^{\infty} x \sum_{i=1}^{k} p_i G_i dx \\ &= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} x G_i dx \\ &= \boxed{\sum_{i=1}^{k} p_i \mu_i} \end{aligned}$$

Calculating Var(X):

We know that:

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

and,

$$Var(Y) = \sigma_i^2 = E(Y^2) - \mu_i^2 \text{ for an arbitrary random variable } Y \text{ distributed as a gaussian } (Y \sim G_i)$$

Thus, $E(Y^2) = \displaystyle\int_{-\infty}^{\infty} x^2 G_i dx = \sigma_i^2 + \mu_i^2$ (Here, $\mu_i = E(Y)$)

We shall now calculate $E(X^2)$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \sum_{i=1}^{k} p_i G_i dx \\ &= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} x^2 G_i dx \\ &= \sum_{i=1}^{k} p_i (\sigma_i^2 + \mu_i^2) \end{aligned}$$

Thus, now we have:

$$Var(X) = E(X^2) - \mu^2 = \boxed{\sum_{i=1}^{k} p_i(\sigma_i^2 + \mu_i^2) - (\sum_{i=1}^{k} p_i\mu_i)^2}$$

Calculating the MGF:

By the definition of the MGF, $\phi_X(t) = E(e^{tX})$
We also know that for an arbitrary random variable Y, which is distributed as a Gaussian G_i ($Y \sim G_i$), we have $\phi_Y(t) = \int_{-\infty}^{\infty} e^{tx} G_i dx = \exp(\mu_i t + \frac{(\sigma_i t)^2}{2})$
Thus for the given X,

$$\begin{aligned}\phi_X(t) &= E(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tX} \sum_{i=1}^{k} p_i G_i dx \\ &= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} e^{tX} G_i dx \\ &= \boxed{\sum_{i=1}^{k} p_i \exp(\mu_i t + \frac{(\sigma_i t)^2}{2})}\end{aligned}$$

Part II

It is given that $Z = \sum_{i=1}^{k} p_i X_i$ where $X_i \sim G_i$ are independent random variables.
By the properties of Gaussian distribution G_i we know that:

$$\begin{aligned}E(X_i) &= \mu_i \\ Var(X_i) &= \sigma_i^2\end{aligned}$$

Calculating E(Z):

$$\begin{aligned}E(Z) &= E(\sum_{i=1}^{k} p_i X_i) \\ &= \sum_{i=1}^{k} p_i E(X_i) && \text{(Linearity of Expectation)} \\ &= \boxed{\sum_{i=1}^{k} p_i \mu_i}\end{aligned}$$

Calculating Var(Z)

$$\begin{aligned}
Var(Z) &= Var(\sum_{i=1}^{k} p_i X_i) \\
&= \sum_{i=1}^{k} p_i^2 Var(X_i) \qquad \text{(as } \{X_i\}_{i=1}^{k} \text{ are independent and } Var(aX) = a^2 Var(X)) \\
&= \boxed{\sum_{i=1}^{k} p_i^2 \sigma_i^2}
\end{aligned}$$

Calculating the MGF

We know that for a Gaussian $X_i \sim G_i$, we have:

$$\phi_{X_i}(t) = \int_{-\infty}^{\infty} e^{tx} G_i dx = \exp(\mu_i t + \frac{(\sigma_i t)^2}{2})$$

We also know the following properties of $\phi_X(t)$:

$$\phi_{(aX)}(t) = \phi_X(at)$$

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) \text{ for independent X, Y}$$

Thus, using the above 2 properties, and the fact that $\{X_i\}_{i=1}^{k}$ are independent,

$$\begin{aligned}
\phi_Z(t) &= \prod_{i=1}^{k} \phi_{(p_i X_i)}(t) \\
&= \prod_{i=1}^{k} \phi_{X_i}(p_i t) \\
&= \prod_{i=1}^{k} \exp(\mu_i p_i t + \frac{(\sigma_i p_i t)^2}{2}) \quad \text{(as } X_i \sim G_i) \\
&= \boxed{\exp(t \sum_{i=1}^{k} \mu_i p_i + \frac{t^2}{2} \sum_{i=1}^{k} p_i^2 \sigma_i^2} \quad \text{(as } \exp(a)\exp(b) = \exp(a+b))
\end{aligned}$$

Calculating the PDF

The obtained MGF of Z is same as that of a gaussian of mean $\mu = \sum_{i=1}^{k}(\mu_i p_i)$ and variance $\sigma^2 = \sum_{i=1}^{k} p_i^2 \sigma_i^2$
Thus, $Z \sim \mathcal{N}(\mu, \sigma^2)$. (Using the Moment Generating Function (MGF) Uniqueness Theorem[1])
Therefore, the PDF of Z will be:

$$\boxed{f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)}$$

[1] For a given random variable, the MGF and PMF **uniquely** determine each other.

Question 3

To prove:

$$P(X - \mu \geq \tau) \leq \frac{\sigma^2}{\sigma^2 + \tau^2} \qquad \text{for } \tau > 0$$

$$P(X - \mu \geq \tau) \geq 1 - \frac{\sigma^2}{\sigma^2 + \tau^2} \qquad \text{for } \tau < 0$$

We shall prove some lemmas before coming to the actual proof:
(Consider X to be a random variable and $a \geq 0$.)

1. $P(X \geq a) = P(X + b \geq a + b)$
   Note that $(X \geq a) \iff (X + b \geq a + b)$, thus this directly leads to the above statement.

2. $P(X^2 \geq a^2) \geq P(X \geq a)$
   This is because of the fact that:

$$X^2 \geq a^2 \Rightarrow |X| \geq a \Rightarrow (X > a) \vee (X < -a)$$

   Thus, $P(X^2 \geq a^2) = P(X > a) + P(X < -a)$.
   As probabilities are always ≥ 0, $P(X^2 \geq a^2) \geq P(X > a)$

Markov's Inequality states that, for $a > 0$ and a random variable X which always takes positive values, we have:

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

Consider a random variable $Y := X - \mu$ and $Z := Y + b$ and consider the case where $\tau > 0$.
Also consider $b \geq 0$, thus we have $(\tau + b) \geq 0$
Thus, as $Z^2 \geq 0$ and $(\tau + b)^2 > 0$, we can apply the Markov's inequality these:

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

Also, by Lemma 2, we have $P((Y + b)^2 \geq (\tau + b)^2) \geq P(Y + b \geq \tau + b)$.
By Lemma 1, $P((Y + b) \geq (\tau + b)) = P(Y \geq \tau)$
Now, using the linearity of the expectation operator,

$$\begin{aligned} E((Y + b)^2) &= E(Y^2) + E(2Yb) + E(b^2) \\ &= E(Y^2) + 2bE(Y) + b^2 \qquad \text{(expectation of a constant is the constant itself)} \\ &= \sigma^2 + 0 + b^2 = \sigma^2 + b^2 \qquad (E((X - \mu)^2) = \sigma^2 \text{ and } E(X - \mu) = 0) \end{aligned}$$

Thus we have,

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

To further tighten the inequality, we need to minimize the function $g(b) = \frac{\sigma^2 + b^2}{(\tau + b)^2}$, solving $g'(b) = 0$ will yield us an extrema b_0:

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

Equating it to zero, we get $b_0 = \frac{\sigma^2}{\tau}$, furthermore as $g''(b) > 0$, this indeed is the minima.
Thus, after putting $Y = X - \mu$ and $b = b_0$, we get:

$$\boxed{P((X-\mu) \geq \tau) \leq \frac{\sigma^2}{\sigma^2+\tau^2}} \qquad \text{(for } \tau > 0\text{)}$$

Now, as $(X-\mu)^2 = (\mu - X)^2$, the inequality $P((Y) \geq \tau) \leq \frac{\sigma^2}{\sigma^2+\tau^2}$ is valid even for $Y = \mu - X$.
Thus for $\tau < 0$ we can still use the above result for $-\tau$.

$$\begin{aligned} P((\mu - X) \geq (-\tau)) &\leq \frac{\sigma^2}{\sigma^2+\tau^2} \\ P((X-\mu) < \tau) &\leq \frac{\sigma^2}{\sigma^2+\tau^2} \\ 1 - P((X-\mu) \geq \tau) &\leq \frac{\sigma^2}{\sigma^2+\tau^2} \quad (P(X < a) = 1 - P(X \geq a)) \\ &\boxed{P((X-\mu) \geq \tau) \geq 1 - \frac{\sigma^2}{\sigma^2+\tau^2}} \quad \text{(for } \tau < 0\text{)} \end{aligned}$$

Question 4

By definition, MGF of a random variable X for parameter t is $\phi_X(t) = \int_{-\infty}^{\infty} e^{tu} f_X(u)du$.

$$
\begin{aligned}
e^{-tx}\phi_X(t) &= \int_{-\infty}^{\infty} e^{t(u-x)} f_X(u)du \\
&= \int_{-\infty}^{x} e^{t(u-x)} f_X(u)du + \int_{x}^{\infty} e^{t(u-x)} f_X(u)du
\end{aligned}
$$

As the integrand is always non-negative,

$$
e^{-tx}\phi_X(t) \geq \int_{-\infty}^{x} e^{t(u-x)} f_X(u)du \quad \textbf{and,} \quad e^{-tx}\phi_X(t) \geq \int_{x}^{\infty} e^{t(u-x)} f_X(u)du \tag{1}
$$

For $t \geq 0$,

$$
\begin{aligned}
e^{-tx}\phi_X(t) &\geq \int_{x}^{\infty} e^{t(u-x)} f_X(u)du && \text{(From 1)} \\
&\geq \int_{x}^{\infty} (1 + t(u-x)) f_X(u)du && (e^x \geq (1+x)) \\
&\geq \left(\int_{x}^{\infty} f_X(u)du \right) + \left(t \cdot \int_{x}^{\infty} (u-x) f_X(u)du \right) \\
&\geq P(X \geq x) + \left(t \cdot \int_{x}^{\infty} (u-x) f_X(u)du \right)
\end{aligned} \tag{2}
$$

As $t \geq 0$; $u \geq x$; $f_X(u) \geq 0 \ \forall u \in (-\infty, \infty)$, we get

$$
\boxed{P(X \geq x) \leq e^{-tx}\phi_X(t) \quad \forall t \geq 0} \tag{3}
$$

For $t \leq 0$,

$$
\begin{aligned}
e^{-tx}\phi_X(t) &\geq \int_{-\infty}^{x} e^{t(u-x)} f_X(u)du && \text{(From 1)} \\
&\geq \int_{-\infty}^{x} (1 + t(u-x)) f_X(u)du && (e^x \geq (1+x)) \\
&\geq \left(\int_{-\infty}^{x} f_X(u)du \right) + \left(t \cdot \int_{-\infty}^{x} (u-x) f_X(u)du \right) \\
&\geq P(X \leq x) + \left(t \cdot \int_{-\infty}^{x} (u-x) f_X(u)du \right)
\end{aligned} \tag{4}
$$

As $t \leq 0$; $u \leq x$; $f_X(u) \geq 0 \ \forall u \in (-\infty, \infty)$, we get

$$
\boxed{P(X \leq x) \leq e^{-tx}\phi_X(t) \quad \forall t \leq 0} \tag{5}
$$

Now, $X = \sum_{i=1}^{n} X_i$, where X_i are independent Bernoulli random variables with mean p_i.
$E(X) = \mu = \sum_{i=1}^{n} p_i$.

$$
\begin{aligned}
\phi_X(t) &= \prod_{i=1}^{n} \phi_{X_i}(t) && (X_i\text{s are independent random variables}) \\
&= \prod_{i=1}^{n} (1 - p_i + p_i e^t) && (X_i\text{s are Bernoulli random variables}) \\
&\leq \prod_{i=1}^{n} (e^{p_i(e^t - 1)}) && (e^x \geq (1+x)) \\
&\leq exp(\sum_{i=1}^{n} p_i(e^t - 1)) \\
&\leq exp(\mu(e^t - 1))
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
P(X > (1+\delta)\mu) &\leq e^{-t\cdot((1+\delta)\mu)} \phi_X(t) \quad \forall t \geq 0 && \text{(From 3)} \\
&\leq exp((-(1+\delta)t\mu)) + \mu(e^t - 1)) && \text{(From 6)}
\end{aligned} \tag{7}
$$

Therefore,

$$
\boxed{(X > (1+\delta)\mu) \leq \frac{e^{\mu(e^t - 1)}}{e^{(1+\delta)t\mu}} \quad \forall t \geq 0,\ \delta \geq 0} \tag{8}
$$

To get a tighter bound, we need the minimum attainable value of the RHS of Eqn 8.
As e^x is a monotonically increasing function, we need to minimize $(-(1+\delta)t\mu)) + \mu(e^t - 1)$.
$\frac{d}{dt}((-(1+\delta)t\mu)) + \mu(e^t - 1)) = (\mu e^t - (1+\delta)\mu) = 0 \implies \boxed{t = ln(1+\delta)}$

$$
\boxed{P(X > (1+\delta)\mu) \leq \frac{e^{\mu\delta}}{e^{\mu(1+\delta)(ln(1+\delta))}} \quad \forall \delta \geq 0} \tag{9}
$$

Question 5

Instructions for running the code:

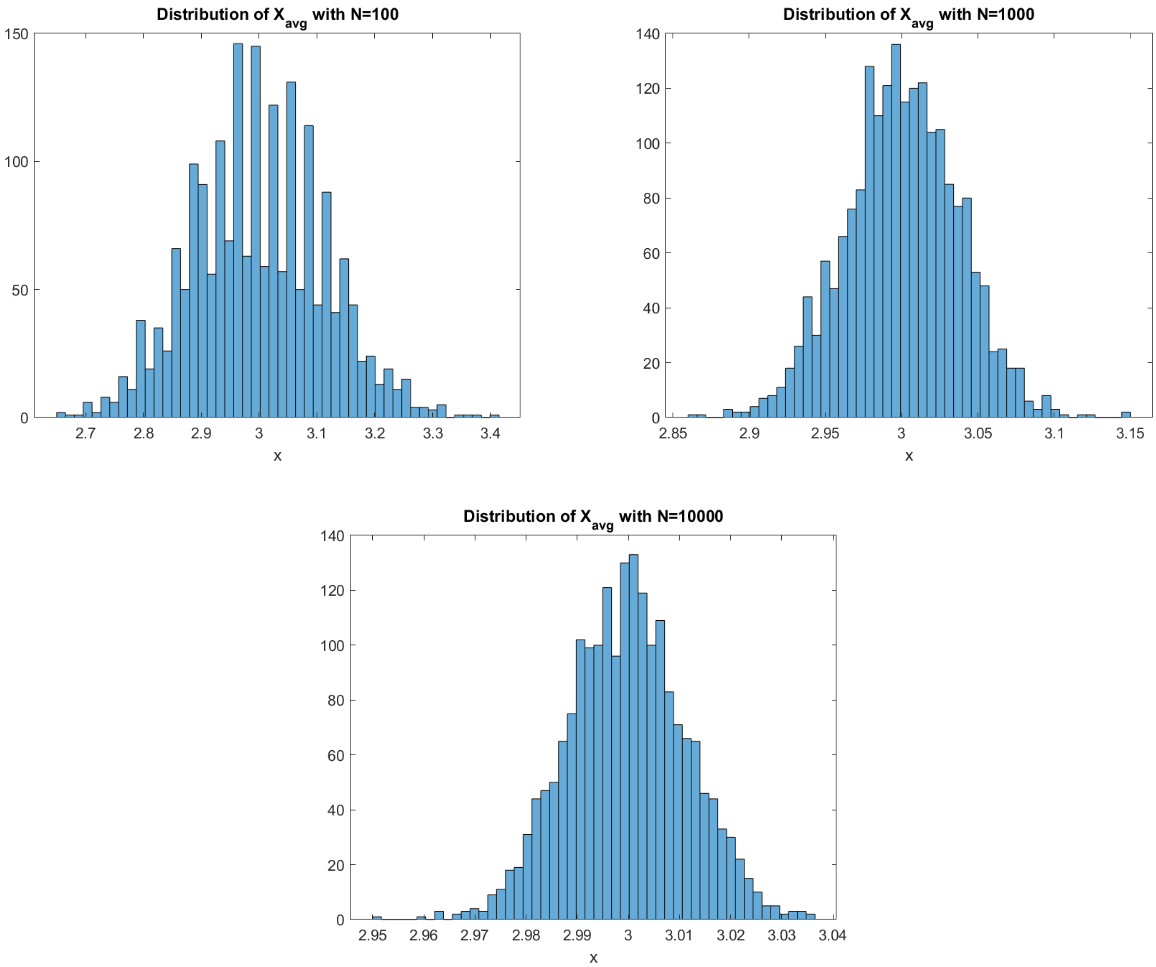
- After extracting submitted file, look for a directory named `code`.
- Within this, the code for this question is contained in a directory named `q5`.
- Run the file `q5.m` (this will create a few plots in `./plots/`).
- The plots can be found in `./plots/`
- a_*.png, b_*.png and c.png are plots corresponding part (a), (b) and (c) respectively.

We also had to calculate the true mean μ and std. dev. σ for getting the Gaussian curve.

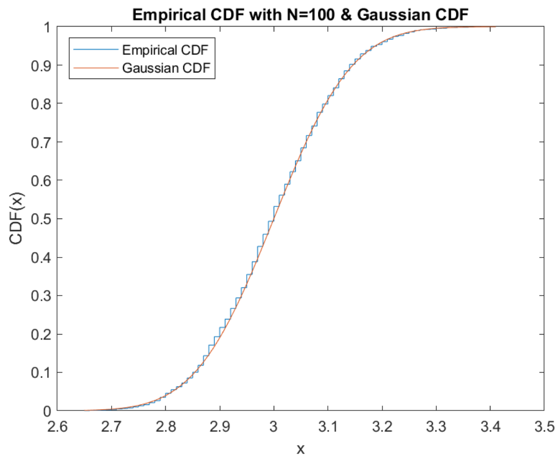
$$\mu_i = E(X_i) = \sum_{k=1}^{5} k \cdot f_{X_i}(X_i = k) = 1 * 0.05 + 2 * 0.4 + 3 * 0.15 + 4 * 0.3 + 5 * 0.1 = 3$$

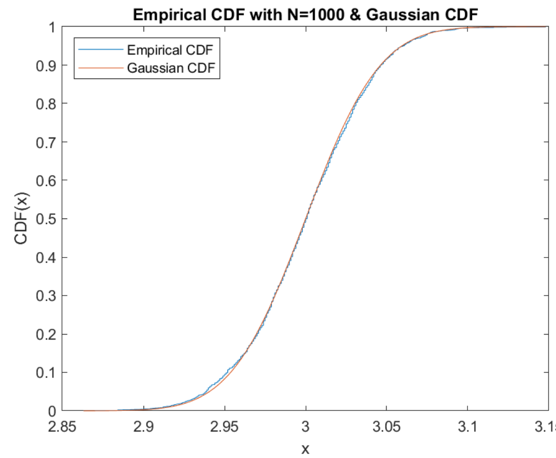
$$\begin{aligned}
\sigma_i^2 = Var(X_i) = E((X_i - \mu_i)^2) = E((X_i - 3)^2) &= \sum_{k=1}^{5} (k-3)^2 \cdot f_{X_i}(X_i = k) \\
&= 4 * 0.05 + 1 * 0.4 + 0 * 0.15 + 1 * 0.3 + 4 * 0.1 = 1.3 \\
\mu &= N * \frac{\mu_i}{N} = 3 \qquad (X_{avg} = \sum_{i=1}^{N} \frac{X_i}{N}) \\
\sigma^2 &= N * \frac{\sigma_i^2}{N^2} = 1.3/N \qquad (X_{avg} = \sum_{i=1}^{N} \frac{X_i}{N}) \\
\sigma &= \sqrt{\frac{1.3}{N}}
\end{aligned}$$

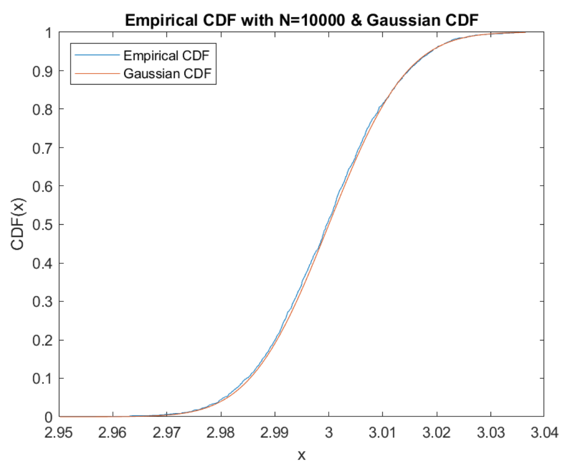
(a)


Distribution of X_{avg} with N=100
150
100
50
0
2.7
2.8
2.9
3
3.1
3.2
3.3
3.4
x
Distribution of X_{avg} with N=1000
140
120
100
80
60
40
20
0
2.85
2.9
2.95
3
3.05
3.1
3.15
x
Distribution of X_{avg} with N=10000
140
120
100
80
60
40
20
0
2.95
2.96
2.97
2.98
2.99
3
3.01
3.02
3.03
3.04
x

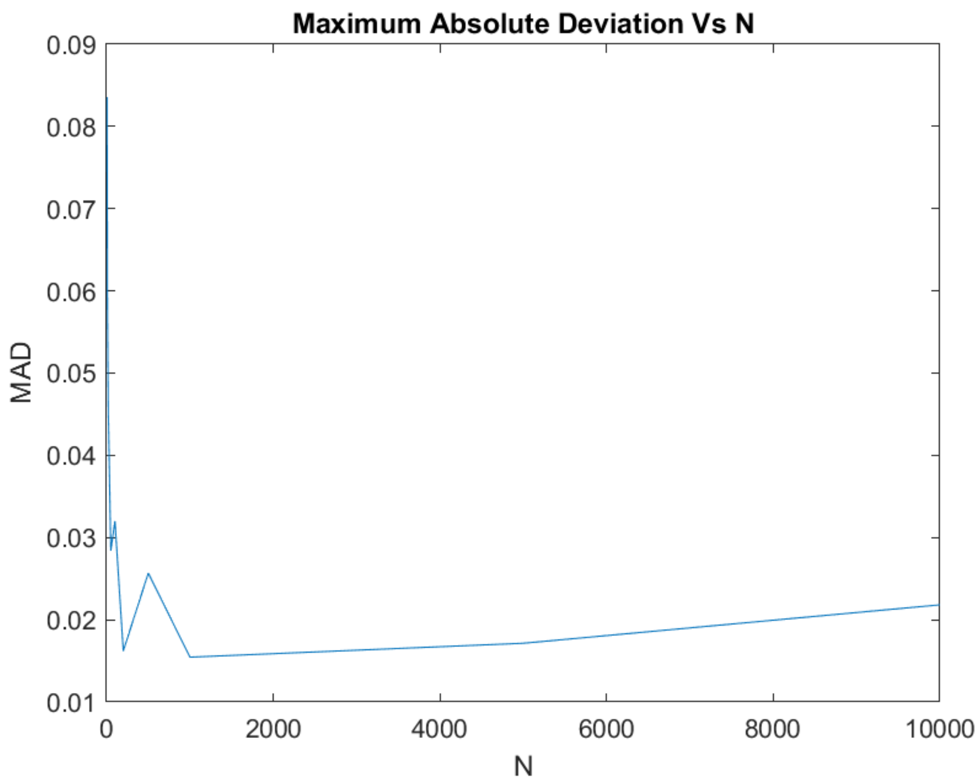
(b)


Empirical CDF with N=100 & Gaussian CDF
Empirical CDF
Gaussian CDF
CDF(x)
x


Empirical CDF with N=1000 & Gaussian CDF
Empirical CDF
Gaussian CDF
CDF(x)
x


Empirical CDF with N=10000 & Gaussian CDF
Empirical CDF
Gaussian CDF
CDF(x)
x

(c)

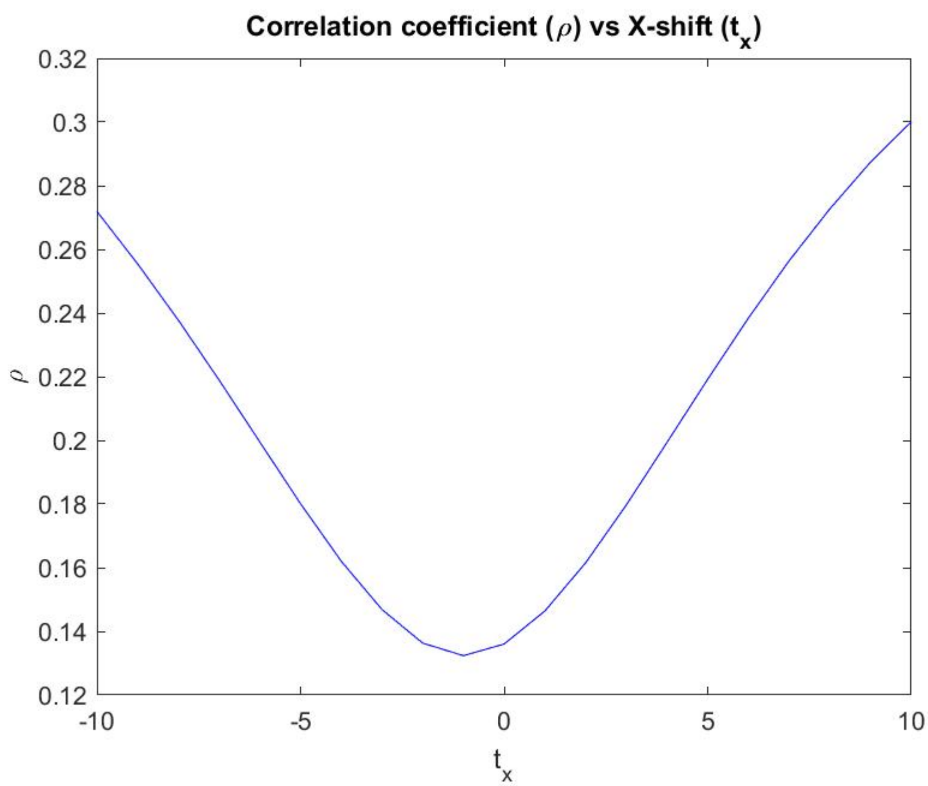

Maximum Absolute Deviation Vs N
MAD
N
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
2000
4000
6000
8000
10000

Question 6

Instructions for running the code:

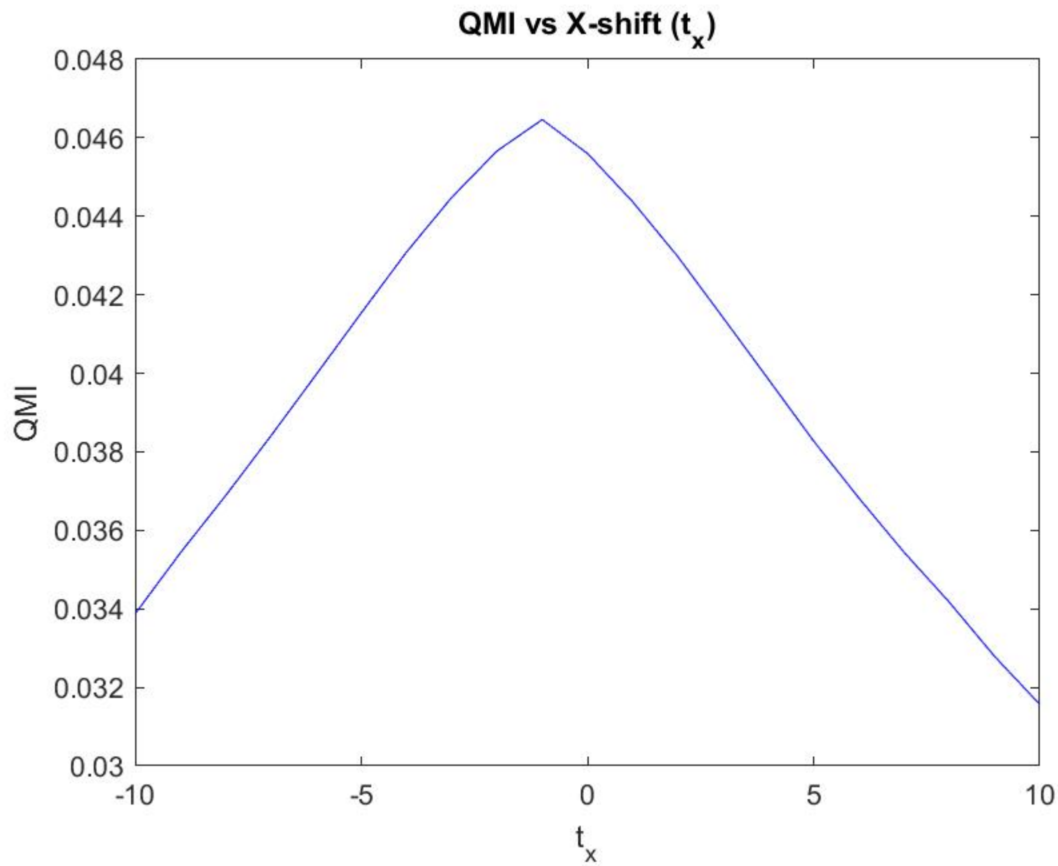
- After extracting submitted file, look for a directory named `code`.
- Within this, the code for this question is contained in a directory named `q6` while the 2 images are contained in the `img` directory.
- Run the file `q6i.m` for plotting the given dependence measures between `T1.jpg` and `T2.jpg` and run `q6ii.m` for plotting the given dependence measures between `T1.jpg` and it's negative.
- Both of these will create two plots and save them at `./plots/`

For T1.jpg and T2.jpg



Comments:

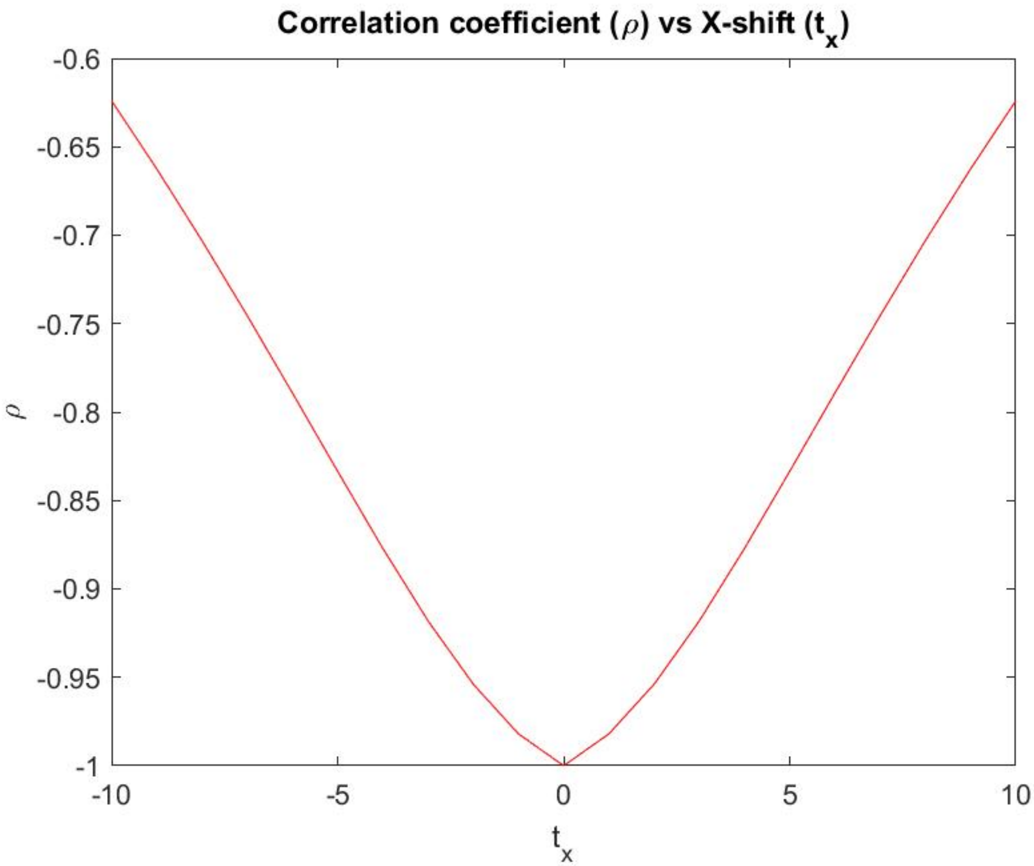
- The correlation is the least when the 2 images are almost exactly aligned ($t_x \approx 0$), a minima is achieved at $t_x = -1$.
- As the shift starts to increase in either direction, the images start to become more positively correlated, with the correlation increasing approximately linearly.
- The correlation is always positive in the given range of the shifts.



Comments:

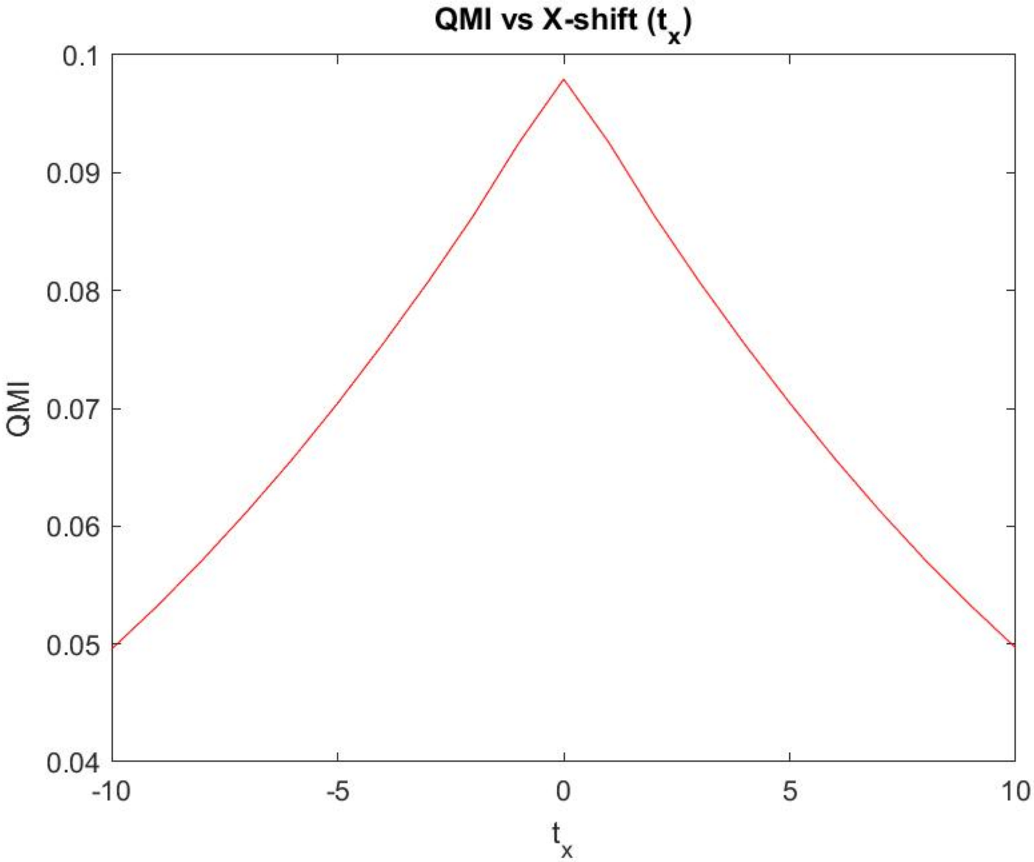
- QMI essentially measures the "dependence" of two random variables on each other. In our case, it is a measure of the mutual information contained between the 2 images.
- The QMI value reaches a maximum at $t_x = -1$ (This is also the point where the correlation coefficient is the least)
- As we further increase the shift in either direction, the QMI value start decreasing approximately linearly.
- On close observation you can see that T1 and T2 have positively correlated background (both black, even on shifting) while the brain image (central part) are negatively correlated (almost inverted color, this correlation diminishes in magnitude as we shift).
  Due to this, $t_x \approx 0 \ \ (= -1)$ has minimum correlation as both of these part's correlation try to cancel out, but it still has high dependence.

For T1.jpg and it's negative



Comments:

- As the second image is the negative of the first and the images are exactly aligned at $t_x = 0$, the correlation coefficient is exactly equal to (-1). Thus, at this point ($t_x = 0$), the two images are completely anti-correlated.
- As the shift starts to increase in either direction, the correlation coefficient also starts to become less negative. The increase is approximately linear.
- The correlation is always negative in the given range of the shifts for the given image and the negative.



Comments:

- QMI essentially measures the "dependence" of two random variables on each other. In our case, it is a measure of the mutual information contained between the 2 images.
- In our case, the second image is the negative of the first, thus both of them contain the exact same information about each other at $t_x = 0$. ($(255 - I_1 = I_2)$ and $(255 - I_2) = I_1$)
- As we increase the shift in either direction, the QMI values start to decrease, as when we are shifting an image, we are also "deleting" information by blacking out the unoccupied pixels.
- The rate of decrease of QMI is higher in this case as compared to the previous case.
- Although the QMI values appear quite low at $t_x = \pm 10$, these are still higher than the previous case, where the maxima itself was about 0.046. This is because even after deleting some information from the negative, there is still a lot of common information between the 2 images.

Question 7

MGF of Multinomial distribution of $\mathbf{X} = (X_1, \ldots, X_k)$, where $\sum_{i=1}^{k} X_i = n$ is $\phi_{\mathbf{X}}(\mathbf{t}) = (p_1e^{t_1}+\cdots+p_ke^{t_k})^n$, where $\mathbf{t} = (t_1, \ldots, t_k)$ and p_i represent probability of X_i.

$$\begin{aligned}
\frac{\partial}{\partial t_i}\phi_{\mathbf{X}}(\mathbf{t}) &= n \cdot (p_1e^{t_1} + \cdots + p_ke^{t_k})^{n-1} \cdot (p_ie^{t_i}) \\
\mu_i = E(X_i) &= \frac{\partial}{\partial t_i}\phi_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=\mathbf{0}=(0,\ldots,0)} \\
&= n \cdot (p_1 + \cdots + p_k)^{n-1} \cdot p_i \\
&= n \cdot p_i \qquad (\sum_{n=1}^{k} p_i = 1)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial t_i^2}\phi_{\mathbf{X}}(\mathbf{t}) &= n(n-1) \cdot (p_1e^{t_1} + \cdots + p_ke^{t_k})^{n-2} \cdot (p_ie^{t_i})^2 + n \cdot (p_1e^{t_1} + \cdots + p_ke^{t_k})^{n-1} \cdot (p_ie^{t_i}) \\
E(X_i^2) &= \frac{\partial^2}{\partial t_i^2}\phi_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=\mathbf{0}=(0,\ldots,0)} \\
&= n(n-1) \cdot (p_1 + \cdots + p_k)^{n-2} \cdot p_i^2 + n \cdot (p_1 + \cdots + p_k)^{n-1} \cdot p_i \\
&= n(n-1) \cdot p_i^2 + n \cdot p_i \qquad (\sum_{n=1}^{k} p_i = 1) \\
Cov(X_i, X_i) &= Var(X_i) \\
&= E(X_i^2) - E(X_i)^2 \\
&= n(n-1) \cdot p_i^2 + n \cdot p_i - (n \cdot p_i)^2 \\
&= n \cdot p_i(1 - p_i)
\end{aligned}$$

For $i \neq j$,

$$\begin{aligned}
\frac{\partial^2}{\partial t_j \partial t_i}\phi_{\mathbf{X}}(\mathbf{t}) &= n(n-1) \cdot (p_1e^{t_1} + \cdots + p_ke^{t_k})^{n-2} \cdot (p_ie^{t_i}) \cdot (p_je^{t_j}) \\
E(X_i \cdot X_j) &= \frac{\partial^2}{\partial t_j \partial t_i}\phi_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=\mathbf{0}=(0,\ldots,0)} \\
&= n(n-1) \cdot (p_1 + \cdots + p_k)^{n-2} \cdot p_i \cdot p_j \\
&= n(n-1) \cdot p_i \cdot p_j \qquad (\sum_{n=1}^{k} p_i = 1) \\
Cov(X_i, X_j) &= E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= E[(X_i - E(X_i))(X_j - E(X_j))] \\
&= E(X_i \cdot X_j) - E(X_i)E(X_j) \\
&= n(n-1) \cdot p_i \cdot p_j - (n \cdot p_i)(n \cdot p_j) \\
&= (-n) \cdot p_i \cdot p_j
\end{aligned}$$

The co-variance matrix **C** of **X** is given by:

$$\boxed{C_{ii} = n \cdot p_i(1 - p_i)}$$

$$\boxed{C_{ij} = (-n) \cdot p_i \cdot p_j \quad \forall\, i \neq j}$$