CS215 : Assignment 2

Dion Reji (22B0029)
Rijul Bhat (22B0971)
Thomas Biju Cheeramvelil (22B1073)

Question 1

Question 1

$X_1, X_2, \dots, X_n \rightarrow$ independent identically distributed random variables with cumulative distribution function $F_X(x)$ and pdf $f_X(x) = F_X'(x)$

a) Consider the random variable $Y_1 = \max(X_1, X_2, \dots, X_n)$. The cdf of Y_1, $F_{Y_1}(y)$ is given by

$$
\begin{aligned}
F_{Y_1}(y) &= P(Y_1 \leq y) \quad \text{[Definition of cumulative distribution function]} \\
&= P(\max(X_1, X_2, \dots, X_n) \leq y) \\
&= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \\
&= P(X_1 \leq y)\, P(X_2 \leq y) \dots P(X_n \leq y) \quad \text{[Since, } X_1, X_2, \dots, X_n \text{ are independent]} \\
&= F_X(y)\, F_X(y) \dots F_X(y) \\
&= (F_X(y))^n
\end{aligned}
$$

$$
\therefore \boxed{
\begin{aligned}
\text{cdf:} \quad & F_{Y_1}(x) = (F_X(x))^n \\
\text{pdf:} \quad & f_{Y_1}(x) = F_{Y_1}'(x) \\
& = n(F_X(x))^{n-1} F_X'(x) \\
& = n(F_X(x))^{n-1} f_X(x)
\end{aligned}
}
$$

b) Consider the random variable $Y_2 = \min(X_1, X_2, \dots, X_n)$. The cdf of Y_2, denoted as $F_{Y_2}(y)$ is given by

$$
\begin{aligned}
F_{Y_2}(y) &= P(Y_2 \leq y) \\
&= 1 - P(Y_2 > y) \\
&= 1 - P(\min(X_1, X_2, \dots, X_n) > y) \\
&= 1 - P(X_1 > y, X_2 > y, \dots, X_n > y)
\end{aligned}
$$

By definition of minimum,

$$\min(x_1, x_2, \cdots, x_n) = x_k \quad , \quad k \in \{1, 2, \cdots, n\}$$

such that $x_k \leqslant x_i$ for all $i = 1, 2, \cdots, n$

<u>Claim</u>: $\min(x_1, x_2, \cdots, x_n) > y$

$$\Leftrightarrow \quad x_i > y \text{ for all } i = 1, 2, \cdots, n$$

<u>Proof</u>:

$$\min(x_1, x_2, \cdots, x_n) > y \Rightarrow x_i \geqslant x_k \text{ for all } i = 1, 2, \cdots, n$$

and $x_k > y$

$$\Rightarrow x_i \geqslant x_k > y$$

$$\Rightarrow x_i > y \text{ for all } i = 1, 2, \cdots, n$$

$x_i > y$ for all $i = 1, 2, \cdots, n$

$$\Rightarrow x_k > y \quad \text{where} \quad x_k = \min(x_1, x_2, \cdots, x_n)$$

$$F_{Y_2}(y) = 1 - P(X_1 > y)\,P(X_2 > y) \cdots P(X_n > y)$$

[Since $X_1, X_2, \cdots, X_n$ are independent]

$$= 1 - \prod_{i=1}^{n} P(X_i > y)$$

$$= 1 - \prod_{i=1}^{n} [1 - P(X_i \leq y)]$$

$$= 1 - \prod_{i=1}^{n} (1 - F_X(y)) = 1 - (1 - F_X(y))^n$$

$\therefore$ For random variable Y_2,

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

$$= n(1 - F_X(x))^{n-1} f_X(x)$$

Question 2

Question 2

Part (i) : Random variable X belongs to a Gaussian Mixture Model (GMM), i.e., $X \sim \sum_{i=1}^{k} p_i \mathcal{N}(\mu_i, \sigma_i^2)$ with $\sum_{i=1}^{k} p_i = 1$

Probability density function of X : $f_X(x) = \sum_{i=1}^{k} p_i f_{X_i}(x)$

- $$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty} x \sum_{i=1}^{k} p_i f_{X_i}(x)\,dx$$
$$= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} x f_{X_i}(x)\,dx = \sum_{i=1}^{k} p_i E(X_i)$$
$$\boxed{E(X) = \sum_{i=1}^{k} p_i \mu_i}$$

- $$Var(X) = E(X^2) - (E(X))^2$$
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_{-\infty}^{\infty} x^2 \sum_{i=1}^{k} p_i f_{X_i}(x)\,dx$$
$$= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} x^2 f_{X_i}(x)\,dx = \sum_{i=1}^{k} p_i E(X_i^2)$$
$$= \sum_{i=1}^{k} p_i \left(Var(X_i) + (E(X_i))^2 \right) = \sum_{i=1}^{k} p_i (\sigma_i^2 + \mu_i^2)$$

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

- $\phi_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$

[MGF of X]

$$= \int_{-\infty}^{\infty} e^{tx} \sum_{i=1}^{k} p_i f_{X_i}(x)\, dx$$

$$= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} e^{tx} f_{X_i}(x)\, dx = \sum_{i=1}^{k} p_i \phi_{X_i}(t)$$

$$\boxed{\phi_X(t) = \sum_{i=1}^{k} p_i\, e^{(\mu_i t + \sigma_i^2 t^2/2)}}$$

[Since, MGF of a Gaussian random variable with mean μ_i and variance σ_i^2 is $\phi_{X_i}(t) = e^{(\mu_i t + \sigma_i^2 t^2/2)}$]

<u>Part (ii)</u>: Random variable $Z = \sum_{i=1}^{k} p_i X_i$ (where $\{X_i\}_{i=1}^{k}$ are independent Gaussian random variables with mean μ_i and variance σ_i)

- $E(Z) = E\left(\sum_{i=1}^{k} p_i X_i\right) = \sum_{i=1}^{k} E(p_i X_i) = \sum_{i=1}^{k} p_i E(X_i)$ [Linearity of expectation operator]

$$\therefore \boxed{E(Z) = \sum_{i=1}^{k} p_i \mu_i}$$

- $Var(Z) = Var\left(\sum_{i=1}^{k} p_i X_i\right) = \sum_{i=1}^{k} Var(p_i X_i)$ [Since $X_1, X_2, \ldots, X_k$ are independent random variables]

$$= \sum_{i=1}^{k} p_i^2 Var(X_i)$$

$$\therefore \boxed{Var(Z) = \sum_{i=1}^{k} p_i^2 \sigma_i^2}$$

- MGF :

$$\phi_Z(t) = E(e^{tZ}) = E\left(e^{t\sum_{i=1}^{k} p_i X_i}\right)$$

$$= E\left(\prod_{i=1}^{k} e^{t p_i X_i}\right)$$

$$= \prod_{i=1}^{k} E(e^{t p_i X_i})$$ [Since $X_1, X_2, \ldots, X_k$ are independent, $\{e^{t p_i X_i}\}_{i=1}^{k}$ are independent.]

$$= \prod_{i=1}^{k} \phi_{X_i}(p_i t)$$

$$= \prod_{i=1}^{k} e^{\left(\mu_i p_i t + \frac{\sigma_i^2}{2} p_i^2 t^2\right)}$$ [Using MGF of Gaussian distribution]

$$= e^{\left(\left(\sum_{i=1}^{k} \mu_i p_i\right) t + \frac{\left(\sum_{i=1}^{k} \sigma_i^2 p_i^2\right) t^2}{2}\right)}$$

$$\boxed{\phi_Z(t) = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)} \quad \text{where } \mu = E(Z) = \sum_{i=1}^{k} p_i \mu_i, \quad \sigma^2 = Var(Z) = \sum_{i=1}^{k} p_i^2 \sigma_i^2}$$

- PDF :

Observe that the moment generating function of Z is same as that of a Gaussian random variable with mean μ and variance σ^2. Hence, the pdf of Z is

$$\therefore \boxed{f_Z(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{where } \mu = E(Z) = \sum_{i=1}^{k} p_i \mu_i, \quad \sigma^2 = Var(Z) = \sum_{i=1}^{k} p_i^2 \sigma_i^2}$$

because, the moment generating function of a probability distribution uniquely determines its pdf.

Question 3

Question 3

Consider a random variable X with mean μ and variance σ^2, and a real number τ.

Case (i) : $\tau \geq 0$

$$P(X-\mu \geq \tau) = P(X-\mu+b \geq \tau+b) \quad \forall\, b \geq 0$$

$$\leq P\left((X-\mu+b)^2 \geq (\tau+b)^2\right) \quad [\because a \geq b \geq 0 \Rightarrow a^2 \geq b^2]$$

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

$\left[\text{By Markov's inequality, } P(X \geq a) \leq \frac{E(X)}{a}. \text{ Take } X \leftarrow (X-\mu+b)^2,\ a \leftarrow (\tau+b)^2\right]$

$$\Rightarrow P(X-\mu \geq \tau) \leq \frac{\sigma^2+b^2}{(\tau+b)^2} \quad \forall\, b \geq 0$$

$$\begin{aligned} &E\left((X-\mu+b)^2\right) \\ &= E\left((X-\mu)^2 + 2b(X-\mu) + b^2\right) \\ &= E\left((X-\mu)^2\right) + 2b\,E(X-\mu) + b^2 \\ &= \sigma^2 + 0 + b^2 = \underline{\underline{\sigma^2+b^2}} \end{aligned}$$

since $\sigma^2 = E\left((X-\mu)^2\right)$, $E(X-\mu) = 0$

Consider the function $f : \mathbb{R}^+ \to \mathbb{R}^+$, $f(b) = \frac{\sigma^2+b^2}{(\tau+b)^2}$

If $f(b)$ is minimum at $b=b_0$, $f'(b_0) = 0$

$$\frac{2b_0}{(\tau+b_0)^2} - \frac{2(\sigma^2+b_0^2)}{(\tau+b_0)^3} = 0 \Rightarrow \tau b_0 + b_0^2 = \sigma^2 + b_0^2 \Rightarrow b_0 = \frac{\sigma^2}{\tau}$$

Observe that $f''(b_0) > 0$. So, f has a minimum at b_0.
$P(X-\mu \geq \tau)$ must be less than or equal to minimum value of $f(b)$.

$$\Rightarrow P(X-\mu \geq \tau) \leq \frac{\sigma^2+b_0^2}{(\tau+b_0)^2} = \frac{\sigma^2\left(1+\frac{\sigma^2}{\tau^2}\right)}{\left(\tau+\frac{\sigma^2}{\tau}\right)^2} = \underline{\underline{\frac{\sigma^2}{\sigma^2+\tau^2}}}$$

$$\therefore \boxed{\text{If } \tau \geq 0,\ P(X-\mu \geq \tau) \leq \frac{\sigma^2}{\sigma^2+\tau^2}} \quad — (1)$$

Case (ii): $\tau < 0$

$$P(X-\mu<\tau) = P(X-\mu+b<\tau+b) \quad \forall\, b\leq 0$$
$$\leq P\left((X-\mu+b)^2 > (\tau+b)^2\right) \quad [\because a\leq b\leq 0 \Rightarrow a^2\geq b^2]$$
$$\leq \frac{E\left((X-\mu+b)^2\right)}{(\tau+b)^2}$$

[By Markov's inequality,
$P(X\geq a) \leq \frac{E(X)}{a}$
Take $X \leftarrow (X-\mu+b)^2$
$a \leftarrow (\tau+b)^2$]

$$\Rightarrow P(X-\mu<\tau) \leq \frac{\sigma^2+b^2}{(\tau+b)^2} \quad [\because E((X-\mu+b)^2) = \sigma^2+b^2]$$
$$\forall\, b\leq 0$$

Consider the function $f:\mathbb{R}^- \to \mathbb{R}^+$, $f(b) = \frac{\sigma^2+b^2}{(\tau+b)^2}$

If $f(b)$ is minimum at $b=b_0$, $f'(b_0)=0$

$$\frac{2b_0}{(\tau+b_0)^2} - \frac{2(\sigma^2+b_0^2)}{(\tau+b_0)^3} = 0 \Rightarrow \tau b_0 + b_0^2 = \sigma^2 + b_0^2 \Rightarrow b_0 = \frac{\sigma^2}{\tau} \leq 0$$

Observe that $f''(b_0)>0$. So, f has a minimum at b_0.
$P(X-\mu<\tau)$ must be less than or equal to minimum value of $f(b)$

$$\Rightarrow P(X-\mu<\tau) \leq \frac{\sigma^2+b_0^2}{(\tau+b_0)^2} = \frac{\sigma^2\left(1+\frac{\sigma^2}{\tau^2}\right)}{\left(\tau+\frac{\sigma^2}{\tau}\right)^2} = \frac{\sigma^2}{\sigma^2+\tau^2}$$

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

$$\therefore \boxed{\text{If } \tau<0,\ P(X-\mu\geq\tau) \geq 1-\frac{\sigma^2}{\sigma^2+\tau^2}}$$

Question 4

Question 4

Consider the random variable $Y = e^{tX}$, where $t \in \mathbb{R}$.

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

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

for any random variable X and any real number t.

By Markov's inequality

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

Here, we are taking X as $Y = e^{tX}$ and a as e^{tx}

$$P(e^{tX} \geq e^{tx}) = P(\{\omega : e^{tX(\omega)} \geq e^{tx}\})$$
$$= P(\{\omega : tX(\omega) \geq tx\})$$

Since e^x is a strictly increasing function, i.e., $e^a \geq e^b \Leftrightarrow a \geq b$

If $t > 0$, $\{\omega : X(\omega) \geq x\} = \{\omega : tX(\omega) \geq tx\}$

If $t < 0$, $\{\omega : X(\omega) \leq x\} = \{\omega : tX(\omega) \geq tx\}$

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

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

Now consider $X = X_1 + X_2 + \dots + X_n$, where $X_1, X_2, \dots, X_n$ are n independent Bernoulli random variables such that $E(X_i) = p_i$

$$\sum_{i=1}^{n} p_i = \mu$$

The moment generating function of X is

$$\phi_X(t) = E(e^{tX}) = E(e^{tX_1+tX_2+\dots+tX_n}) = E(e^{tX_1}e^{tX_2}\dots e^{tX_n})$$

$$= E(e^{tX_1})E(e^{tX_2})\dots E(e^{tX_n}) \quad \left[\text{Since } X_1, X_2, \dots, X_n \text{ are independent}\right]$$

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

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

$\Big[X_i \rightarrow$ Bernoulli random variable with expectation p_i

$\phi_{X_i}(t) = 1-p_i+p_ie^t$

$= \underline{\underline{1+p_i(e^t-1)}}$ $\Big]$

$$\leqslant \left[\frac{\sum_{i=1}^{n}(1+p_i(e^t-1))}{n}\right]^n \quad \left[\text{by AM} \geqslant \text{GM inequality}\right]$$

$$\Rightarrow \phi_X(t) = \left(1+\frac{\mu}{n}(e^t-1)\right)^n \leqslant \left(e^{\frac{\mu}{n}(e^t-1)}\right)^n = \underline{\underline{e^{\mu(e^t-1)}}}$$

$$\left[\text{Using } 1+x \leqslant e^x\right]$$

$\therefore$ By using previously derived inequality $P(X \geqslant x) \leqslant e^{-tx}\phi_X(t)$, (if $t>0$)
we get

$$\underline{\underline{P(X \geqslant (1+\delta)\mu) \leqslant \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)\mu t}}}}$$

Hence proved.

To tighten this bound, we need to find t for which the upper bound is minimum.

Consider the function $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $f(t) = e^{\mu(e^t-1)-(1+\delta)\mu t}$

If $f(t)$ is minimum at $t=t_0$,

$$f'(t_0) = 0 \Rightarrow \left[\mu e^{t_0} - (1+\delta)\mu\right]f(t_0) = 0 \Rightarrow \underline{\underline{t_0 = \ln(1+\delta)}}$$

$$f(t_0) = e^{\mu(e^{t_0}-1-(1+\delta)t_0)} = e^{\mu(\delta-(1+\delta)\ln(1+\delta))} = \underline{\underline{\frac{e^{\mu\delta}}{(1+\delta)^{(1+\delta)}}}}$$

$$\therefore P(X \geqslant (1+\delta)\mu) \leqslant \min_t f(t) = f(t_0)$$

$$\Rightarrow \boxed{P(X \geqslant (1+\delta)\mu) \leqslant \frac{e^{\mu\delta}}{(1+\delta)^{(1+\delta)}}}$$

Question 5

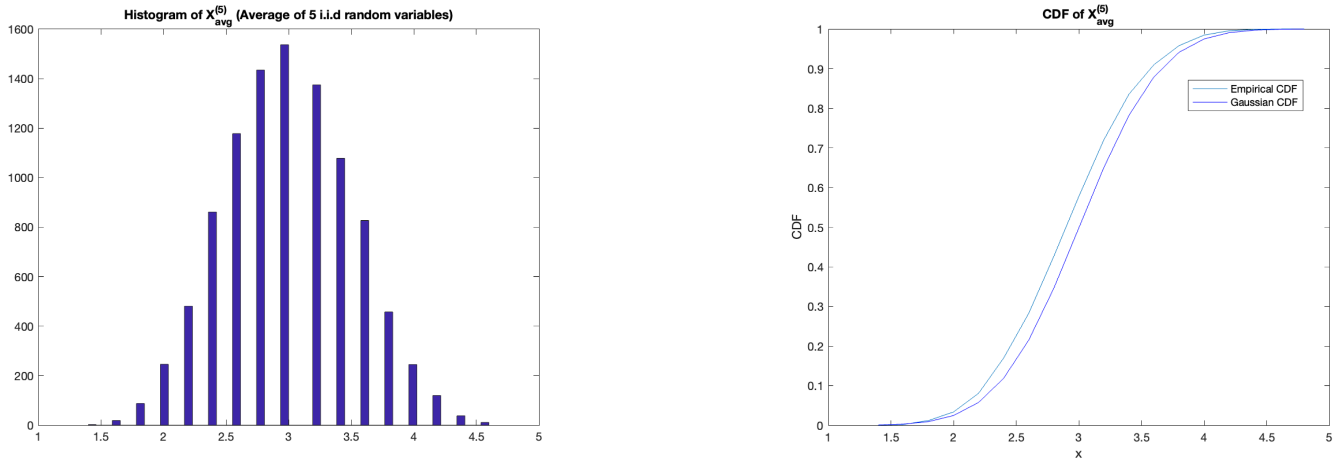


Figure 1: $N = 5, \text{MAD} = 0.0802$

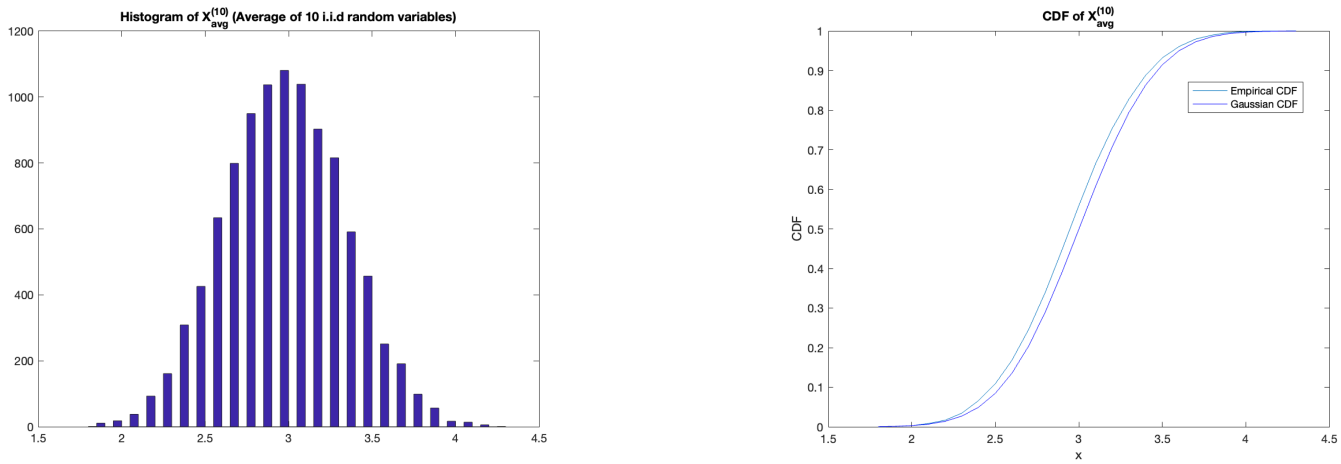


Figure 2: $N = 10, \text{MAD} = 0.0595$

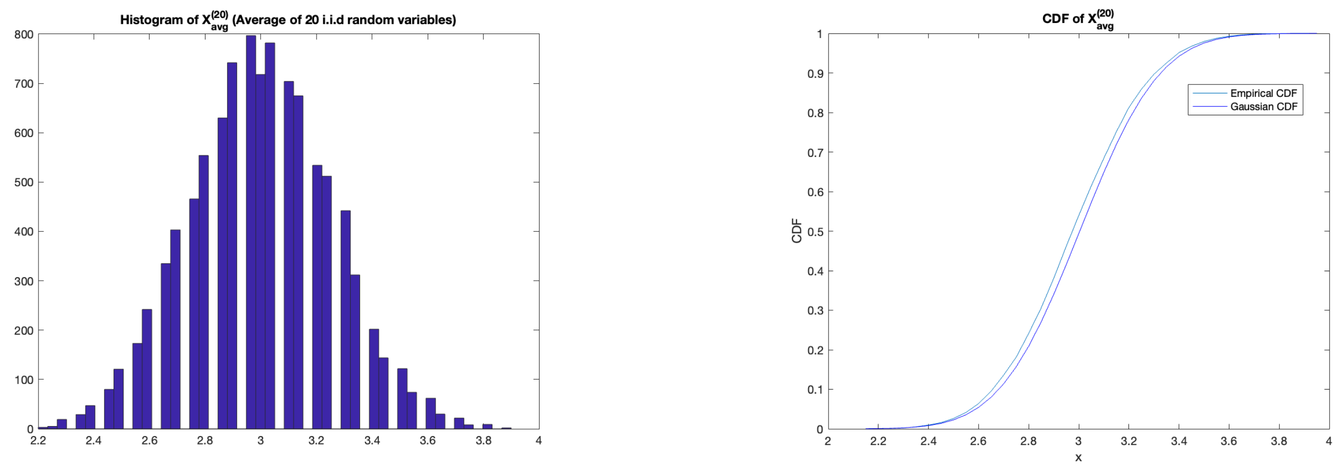


Figure 3: $N = 20, \text{MAD} = 0.454$

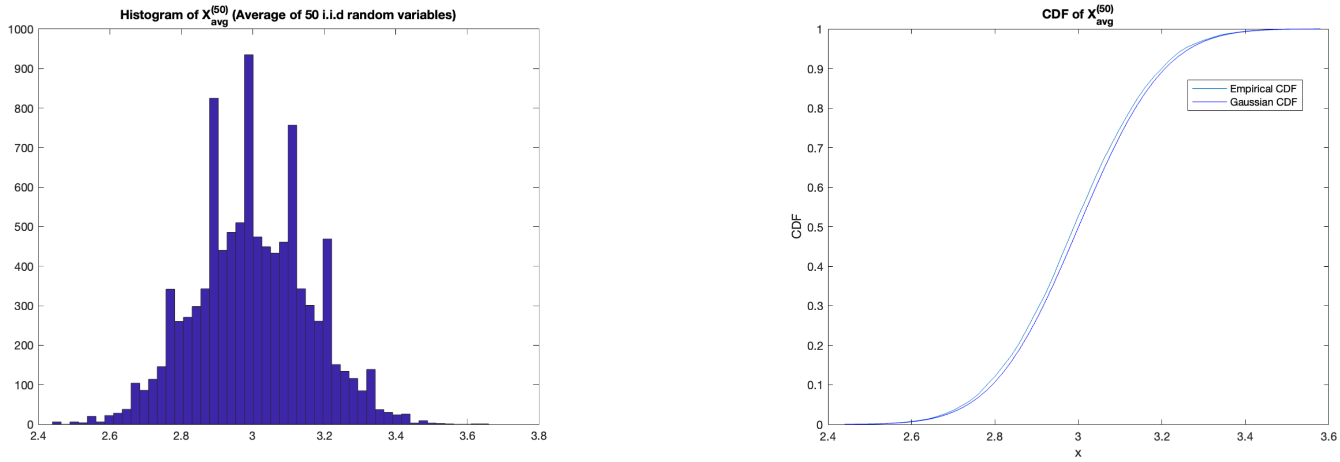


Figure 4: $N = 50, \mathrm{MAD} = 0.0306$

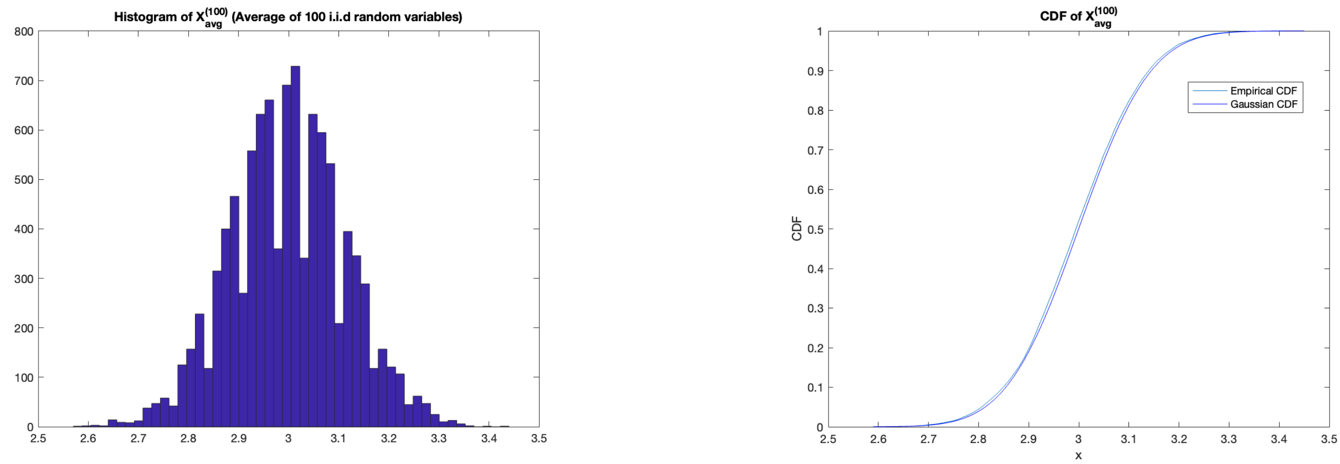


Figure 5: $N = 100, \mathrm{MAD} = 0.0221$

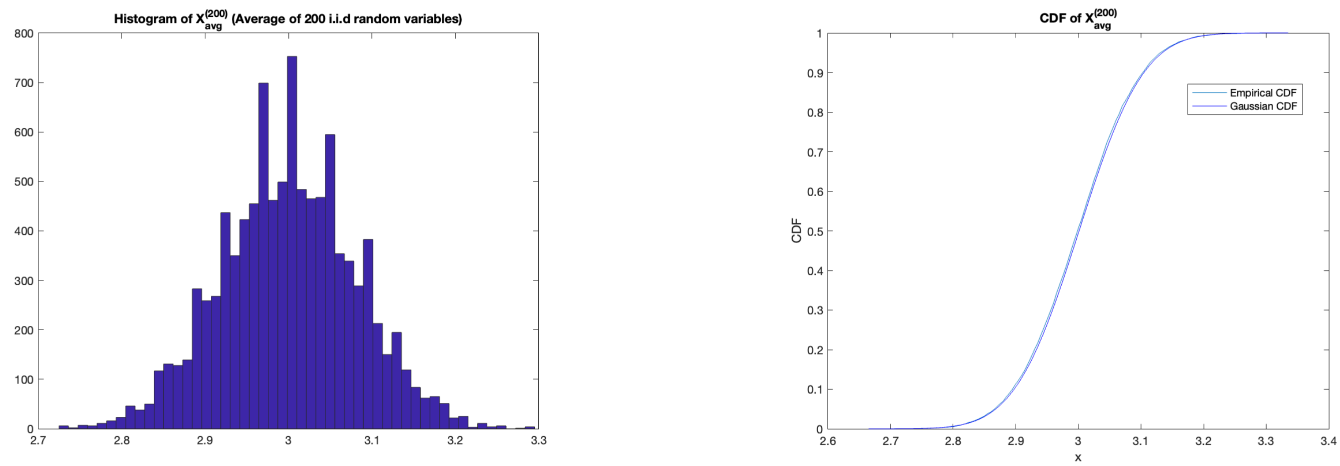


Figure 6: $N = 200, \mathrm{MAD} = 0.0161$

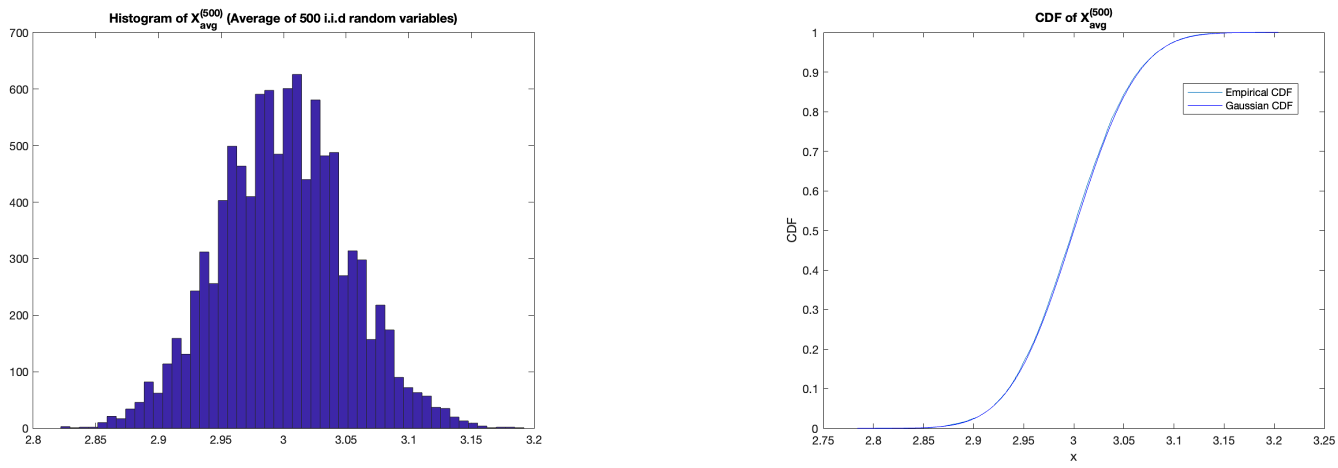


Figure 7: $N = 500, \mathrm{MAD} = 0.0119$

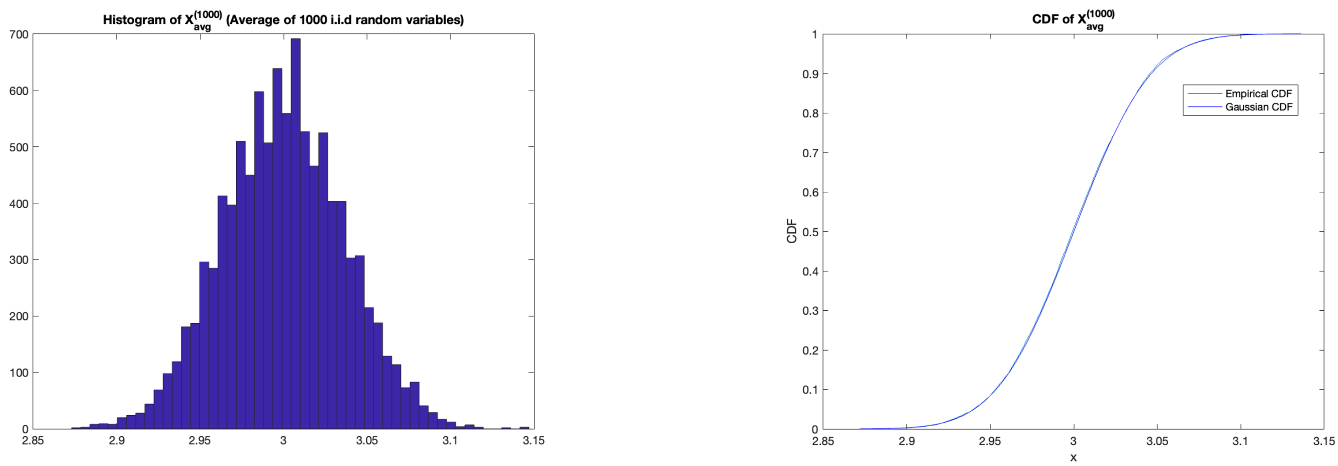


Figure 8: $N = 1000, \mathrm{MAD} = 0.0075$

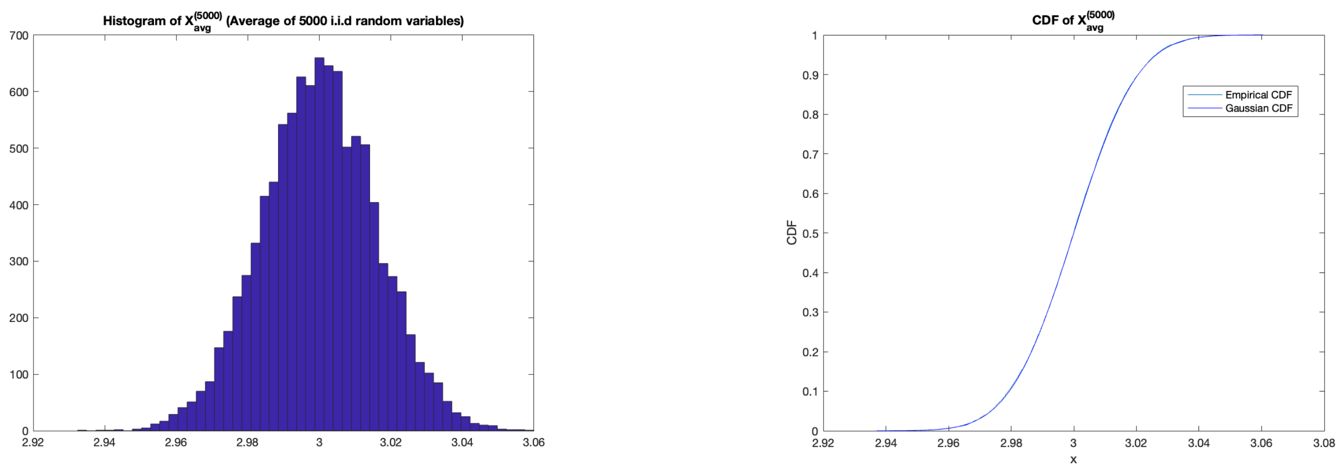


Figure 9: $N = 5000, \mathrm{MAD} = 0.0080$

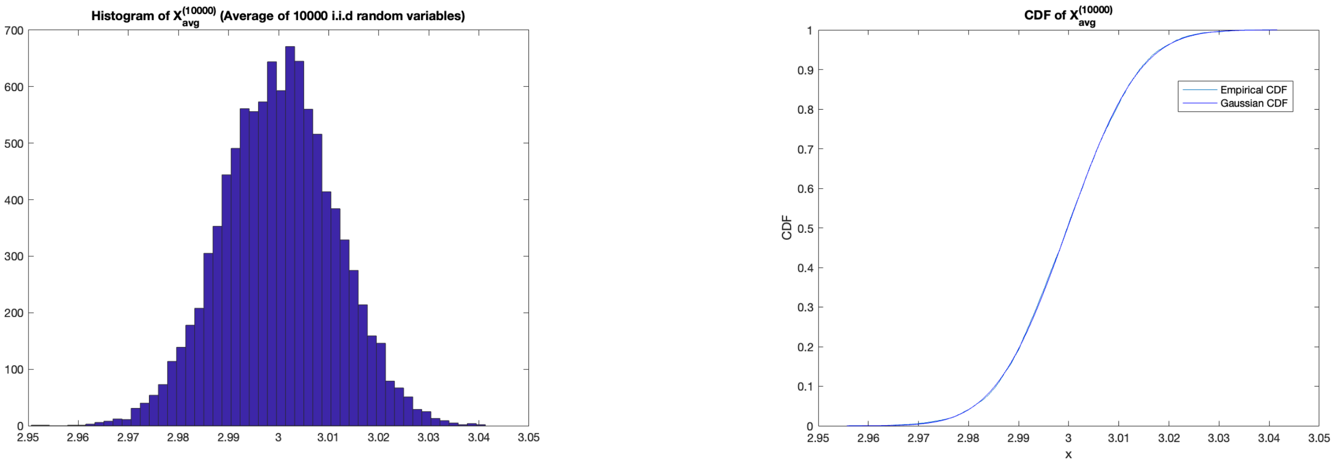


Figure 10: $N = 10000, \mathrm{MAD} = 0.0056$

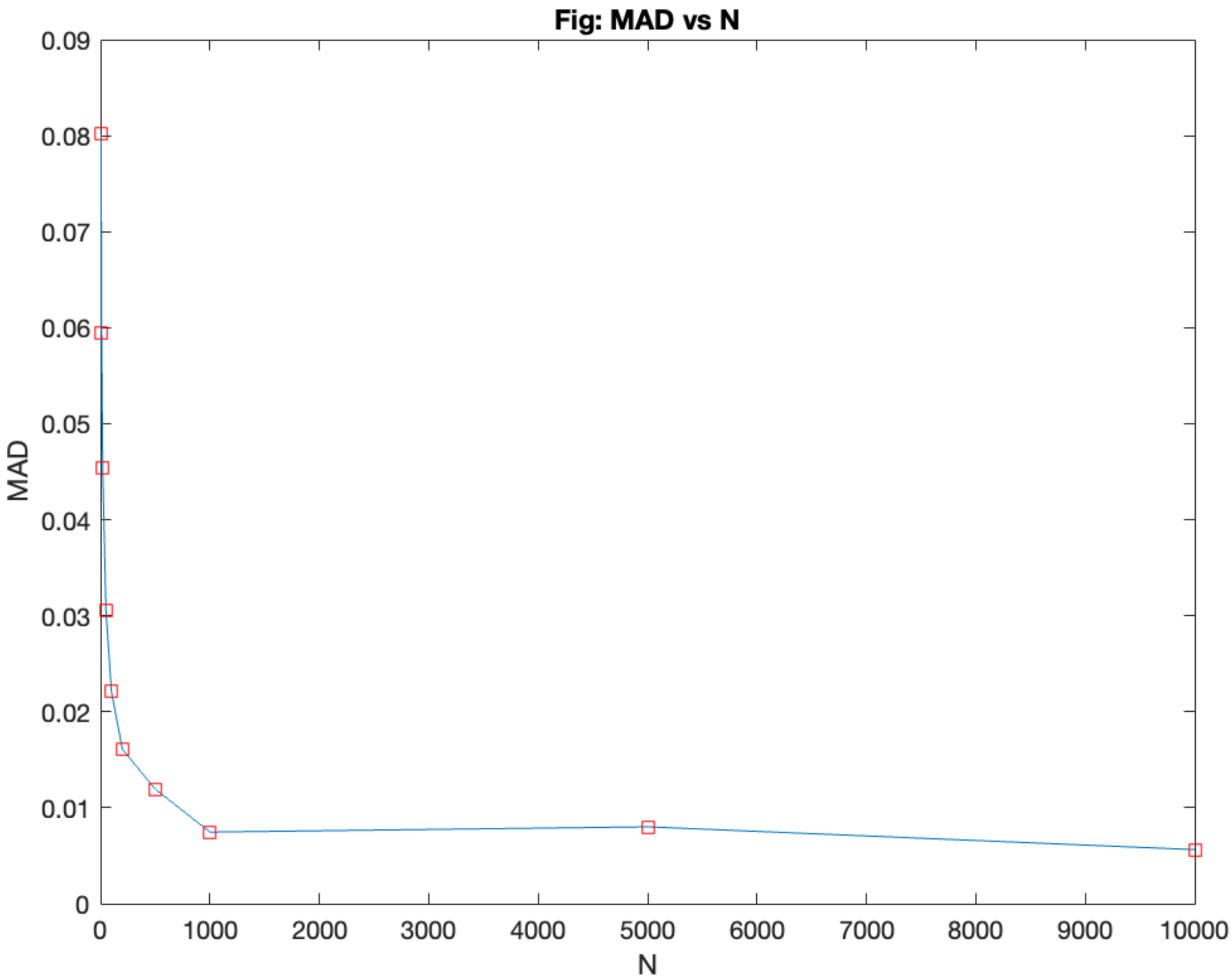


Figure 11: MAD vs N plot

Comments:

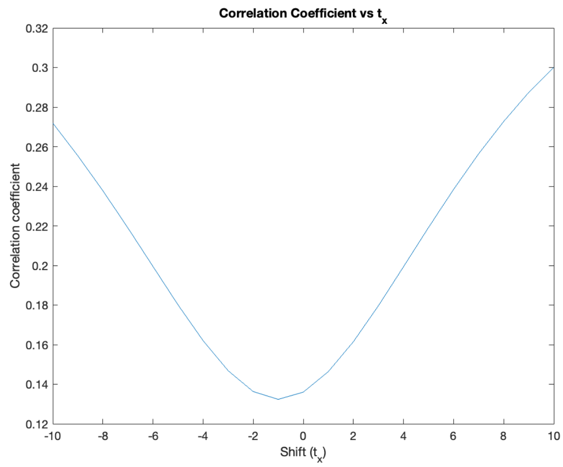
1. *Histograms of Averages:* For small values of N, the histograms of the average may appear quite irregular and skewed, reflecting the characteristics of the individual random variables $X_1, X_2, \ldots, X_N$. As N increases, the shape of the histograms tends to become more symmetric and bell-shaped. This aligns with the CLT, which states that as N approaches infinity, the distribution of the sample mean $X_{avg}^{(N)}$ approaches a normal (*Gaussian*) distribution, regardless of the original distribution of X.

2. *Empirical CDFs vs. Gaussian CDFs:* When comparing the empirical CDFs of the averages to Gaussian CDFs, for small N, there may be noticeable discrepancies, especially in the tails of the distributions. As N increases, the empirical CDFs become increasingly similar to the Gaussian CDFs. This is consistent with the CLT, which asserts that as N grows, the distribution of the sample mean becomes increasingly Gaussian.

3. *Maximum Absolute Difference (MAD):* The MAD between the empirical CDF and Gaussian CDF quantifies how closely the distribution of the averages resembles a Gaussian distribution. For small N, MAD values can be relatively large, indicating significant deviations fro Gaussian behavior. As N increases, MAD values tend to decrease, reflecting the convergence of the sample mean distribution to a Gaussian distribution as predicted by the CLT.

Question 6

T1.jpg and T2.jpg



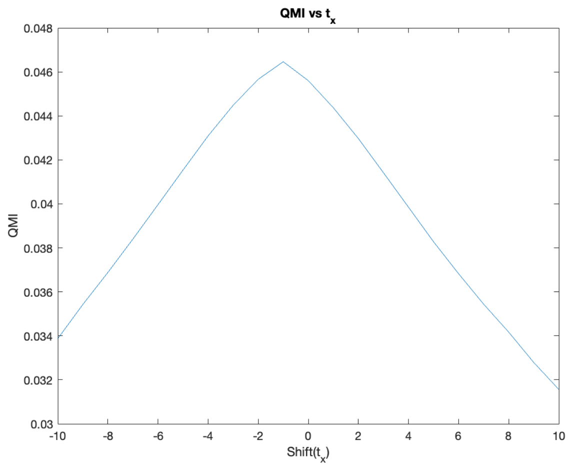


Figure 12: Plots correlating T1 and shifted T2

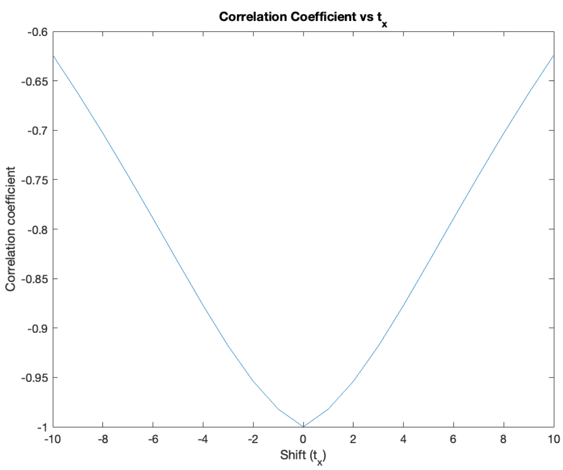
Comments:

1. The part of the images with the *Brain* are almost inverted and hence correspond to a correlation coefficient close to -1, while the *Black backgrounds* in both images have correlation coefficient of 1. This dominates over the -1 of the *Brain* region, thus making the net value of correlation coefficient positive, and of low magnitude.

2. The correlation coefficient, ρ attains its minimum close to zero. This is because when T2 was shifted, some part of original image was replaced by columns of black pixels. Hence, the correlation coefficient will increase in shifted T2 as compared to one with zero shift.

3. QMI quantifies the shared information between the images, and it is maximized when the images match perfectly. Here, QMI attains maximum close to zero, because both images represent the same anatomical structures and are perfectly aligned at shift 0. Note that if the pixel intensities from the two images I1 and I2 were independent, $p_{I_1 I_2}(i_1, i_2) = p_{I_1}(i_1, i_2) \cdot p_{I_2}(i_1, i_2)$, and thus QMI would have been zero. Thus, maximum QMI is attained when images are most dependent.

T1.jpg and inverted T1.jpg



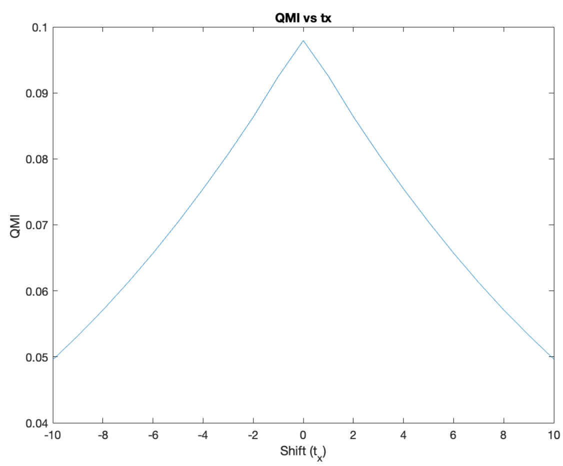


Figure 13: Plots correlating T1 and shifted inverted T1

Comments:

1. In this case, the correlation coefficient (ρ) becomes negative, approaching -1 at zero shift. This negative correlation indicates ideal uncorrelation or anti-correlation. When one image has high pixel intensity, the other has low intensity, resulting in a strong negative relationship.

2. The value of correlation coefficient increases on either sides of zero, since as the pixel intensities are shifted in inverted T1, its correlation with T1 increases.

3. The measure QMI quantifies the shared information between the images. Here, QMI attains maximum exactly at zero, because both images represent the same anatomical structures and are perfectly aligned at shift $t_x = 0$. QMI is most dependent at shift 0 since if we are given the pixel intensity $I_1(x, y)$ of any pixel in the first image, then we will be able to find the intensity of the corresponding pixel $I_2(x, y)$ of the inverted image accurately. As the shift increases, the QMI decreases.

Question 7

Question 7

Consider a multinomial random variable $\vec{X} = (X_1, X_2, \cdots, X_k)$. The ith element (X_i) represents the number of trials that produced the ith outcome, and p_i represents the probability of a single trial producing the ith outcome among the k outcomes.

$$\sum_{i=1}^{k} p_i = 1$$

The moment generating function for the multinomial random variable $\vec{X}$, taking n (total number of trials), $p_1, p_2, \cdots, p_k$ as parameters is given by:

$$\phi_{\vec{X}}(\vec{t}) = E(e^{\vec{t}\cdot\vec{x}})$$
$$= (p_1e^{t_1} + p_2e^{t_2} + \cdots + p_k t_k)^n$$

- Note that $E\left(\prod_{i=1}^{k} X_i^{\alpha_i}\right) = \left(\begin{array}{l}\text{Coefficient of } \prod_{i=1}^{k} t_i^{\alpha_i} \\ \text{in MGF expansion}\end{array}\right) \times \prod_{i=1}^{k} (\alpha_i!)$

$$= \left.\frac{\partial^{(\alpha_1+\alpha_2+\cdots)} \phi_{\vec{X}}(\vec{t})}{\partial t_1^{\alpha_1}\partial t_2^{\alpha_2}\cdots\partial t_k^{\alpha_k}}\right|_{\vec{t}=\vec{0}}$$

- $E(X_i) = \left.\frac{\partial}{\partial t_i}\phi_{\vec{X}}(\vec{t})\right|_{\vec{t}=0} = \left.\frac{\partial}{\partial t_i}(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^n\right|_{\vec{t}=\vec{0}}$

$$= \left. np_ie^{t_i}(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^{n-1}\right|_{\vec{t}=\vec{0}}$$

$$E(X_i) = \underline{\underline{np_i}} \qquad \forall\, i \in \{1, 2, \ldots, k\}$$

- $E(X_i^2) = \frac{\partial^2 \phi_{\vec{X}}(\vec{t})}{\partial t_i^2} = \left.\frac{\partial^2}{\partial t_i^2}(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^n\right|_{\vec{t}=\vec{0}}$

$$= \left.\frac{\partial}{\partial t_i}\left[np_ie^{t_i}(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^{n-1}\right]\right|_{\vec{t}=\vec{0}}$$

$$= np_i\left[\begin{array}{l} e^{t_i}(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^{n-1} \\ + (n-1)p_i\, e^{2t_i}(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^{n-2}\end{array}\right]_{\vec{t}=\vec{0}}$$

$$= \underline{\underline{np_i[1 + (n-1)p_i]}}$$

$$\Rightarrow C_{ii} = \operatorname{Cov}(X_i, X_i) = E((X_i-\mu_i)(X_i-\mu_i)) = E((X_i-\mu_i)^2)$$
$$= \operatorname{Var}(X_i)$$
$$= E(X_i^2) - (E(X_i))^2 = [np_i + n(n-1)p_i^2] - [np_i]^2$$
$$= np_i - np_i^2$$
$$\boxed{C_{ii} = np_i(1-p_i)}$$

- If $i \neq j$,

$$E(X_iX_j) = \frac{\partial^2 \phi_{\vec{X}}(\vec{t})}{\partial t_i \partial t_j}\Bigg|_{\vec{t}=\vec{0}} = \frac{\partial^2}{\partial t_i \partial t_j}\left(p_1e^{t_1} + \cdots + p_ke^{t_k}\right)^n\Bigg|_{\vec{t}=\vec{0}}$$
$$= \frac{\partial}{\partial t_i}\left[np_j\left(p_1e^{t_1}+\cdots+p_ke^{t_k}\right)^{n-1}\right]\Bigg|_{\vec{t}=\vec{0}}$$
$$= \left[n(n-1)p_ip_j\left(p_1e^{t_1}+\cdots+p_ke^{t_k}\right)\right]_{\vec{t}=\vec{0}}$$
$$= n(n-1)p_ip_j$$

$$\Rightarrow C_{ij} = \operatorname{cov}(X_i,X_j) = E[(X_i-\mu_i)(X_j-\mu_j)] = E[X_iX_j - \mu_iX_j - \mu_jX_i + \mu_i\mu_j]$$
$$= E(X_iX_j) - \mu_iE(X_j) - \mu_jE(X_i) + \mu_i\mu_j$$
$$= E(X_iX_j) - \mu_i\mu_j$$
$$= E(X_iX_j) - E(X_i)E(X_j)$$
$$= n(n-1)p_ip_j - (np_i)(np_j) = -np_ip_j$$

$$\therefore \boxed{C_{ij} = -np_ip_j, \quad i \neq j}$$

Hence, the covariance matrix C of a multinomial distribution is given by $C = [C_{ij}]_{k\times k}$ where

$$C_{ij} = \begin{cases} np_i(1-p_i) & \text{if } i=j \\ -np_ip_j & \text{if } i\neq j \end{cases}$$

Attached MATLAB code files

1. Q5_A.m
2. Q5_B.m
3. Q5_C.m
4. Q6.m

Instructions to run the files

1. Simply run each script to observe the outputs shown in the report.
2. For Q5_A.m,
   Simply run the script and the histogram plots will be displayed and saved in the default path set in MATLAB.