CS215 Assignments

Data Analysis and Interpretation

Paarth Jain
190050076

Sahasra Ranjan
190050102

Sibasis Nayak
190050115

September 15, 2020

Assignment 2

1. (a) $Y_1 = max(X_1, X_2, \cdots, X_n)$
   $Y_1 \leq y \implies X_1 \leq y, X_2 \leq y, \cdots, X_n \leq y$ (By definition of Y_1)

$$\begin{aligned} \therefore F_Y(y) &= P(Y_1 \leq y) = P(X_1 \leq y, X_2 \leq y, \cdots, X_n \leq y) \\ &= P(X_1 \leq y)P(X_2 \leq y)\cdots P(X_n \leq y) && (X_i\ are\ Independent\ Random\ Variables) \\ F_Y(y) &= F_X(y)^n && (X_i\ are\ Identically\ Distributed\ Random\ Variables) \\ F_Y(x) &= F_X(x)^n \end{aligned}$$

   Now, let $f_Y(y)$ be the pdf of Y_1, then,

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = nF_X(y)^{n-1}F'_X(y) = nF_X(y)^{n-1}f_X(y) \\ f_Y(x) &= nF_X(x)^{n-1}f_X(x) \end{aligned}$$

   (b) $Y_2 = min(X_1, X_2, \cdots, X_n)$
   $Y_2 \geq y \implies X_1 \geq y, X_2 \geq y, \cdots, X_n \geq y$ (By definition of Y_2)

$$\begin{aligned} \therefore F_Y(y) &= P(Y_2 \leq y) = (1 - P(Y_2 \geq y)) = 1 - P(X_1 \geq y, X_2 \geq y, \cdots, X_n \geq y) \\ &= 1 - P(X_1 \geq y)P(X_2 \geq y)\cdots P(X_n \geq y) && (X_i\ are\ Independent\ Random\ Variables) \\ F_Y(y) &= 1 - (1 - F_X(y))^n && (X_i\ are\ Identically\ Distributed\ Random\ Variables) \\ F_Y(x) &= 1 - (1 - F_X(x))^n \end{aligned}$$

   Now, let $f_Y(y)$ be the pdf of Y_1, then,

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = -n(1 - F_X(y))^{n-1}(-F'_X(y)) = n(1 - F_X(y))^{n-1}f_X(y) \\ f_Y(x) &= n(1 - F_X(x))^{n-1}f_X(x) \end{aligned}$$

2. The solution to this question is at the end in handwritten form.

3. Given: $E(X) = \mu, Var(X) = \sigma^2$
   We will define $Y = X - \mu$, then $E(Y) = 0, Var(Y) = Var(X) = \sigma^2$. We need to derive an upper bound for $P(Y \geq \tau), (\tau > 0)$ now consider t s.t. $\tau + t > 0$ i.e.

$$P(Y \geq \tau) = P(Y + t \geq \tau + t) = P(\frac{Y + t}{\tau + t} \geq 1)$$

   Now $P((\frac{Y+t}{\tau+t})^2 \geq 1) = P(\frac{Y+t}{\tau+t} \geq 1) + P(\frac{Y+t}{\tau+t} \leq -1)$
   as $P(A \cup B) = P(A) + P(B) - P(AB)$, here $P(AB) = 0$

$$\implies P(\frac{Y + t}{\tau + t} \geq 1) \leq P((\frac{Y + t}{\tau + t})^2 \geq 1)$$

$$\implies P(Y \geq \tau) \leq P((\frac{Y + t}{\tau + t})^2 \geq 1)$$

As the Random Variable $(\frac{Y+t}{\tau+t})^2$ is non negative, we can apply Markov's Inequality to the R.H.S. Thus we get-

$$\begin{aligned}
P((\frac{Y+t}{\tau+t})^2 \geq 1) &\leq \frac{E((\frac{Y+t}{\tau+t})^2)}{1} = \frac{E((Y+t)^2)}{(\tau+t)^2} \\
&\leq \frac{E(Y^2)+2tE(Y)+t^2}{(\tau+t)^2} \\
&\leq \frac{(Var(Y)-E(Y)^2)+2t\times 0+t^2}{(\tau+t)^2} = \frac{\sigma^2+t^2}{(\tau+t)^2}
\end{aligned}$$

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

This holds for all $t > -a$. Now we would like to derive a bound as tight as possible, thus we shall find the minimum value of the R.H.S., using calculus (let $f(t) = \dfrac{\sigma^2+t^2}{(\tau+t)^2}$)

$$f'(t) = \frac{(\tau+t)^2(2t) - 2(\tau+t)(\sigma^2+t^2)}{(\tau+t)^4} = 0$$

$$2t\tau - 2\sigma^2 = 0, \quad t = \sigma^2/\tau$$

$$\begin{aligned}
f'(t) &= \frac{2(\tau t - \sigma^2)}{(\tau+t)^3} \\
f"(t) &= \frac{2r(\tau+t) - 6(\tau t - \sigma^2)}{(r+t)^4} \\
\textit{for } t = \sigma^2/\tau, & \\
f"(t) &= \frac{2(\sigma^2+\tau^2)}{(r+t)^2} > 0
\end{aligned}$$

Therefore, t $= \sigma^2/\tau$ represents the point of minima of the function f

This satisfies $\tau + t > 0$(as $\tau > 0$ for this value of t), $f(t) = \dfrac{\sigma^2}{\sigma^2+\tau^2}$
Thus we get -

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

Which is the one sided Chebyshev's Inequality

For $\tau < 0$ we shall evaluate $P(Y < \tau)$ and then use the formula $P(Y \geq \tau) = 1 - P(Y < \tau)$ to get our answer. Consider t s.t. $\tau + t < 0$

$$P(Y < \tau) = P(Y + t < \tau + t) = P(\frac{Y+t}{\tau+t} > 1)$$

Now $P(\frac{Y+t}{\tau+t} > 1) \leq P((\frac{Y+t}{\tau+t})^2 > 1)$ (Similar to the first case) We can now apply Markov's Inequality to the R.H.S. (even though it has a strictly greater than condition), as this would result in a weaker bound as compared to the original Inequality (as we are excluding the equal to case).

Thus following all the subsequent steps (as in first case ($\tau > 0$), we get the value of t as σ^2/τ (satisfies $f''(t) > 0$) which satisfies our intial condition $t + \tau < 0$ as $\tau < 0$ for this part. So we finally have-

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

hence,

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

4. Let us chose a random variable $Y = e^{t(X-x)}$. Clearly Y is a non negative random variable. So applying markov's inequality with $a = 1$ we get,

$$\begin{aligned}
P(Y \geq 1) &\leq E(Y)/1 \\
&\leq E(e^{t(X-x)}) \\
&\leq E(e^{tX} \times e^{-tx}) \\
&\leq E(e^{tX})e^{-tx} && \because e^{-tx} \textit{ is a constant} \\
&\leq \phi_X(t)e^{-tx}
\end{aligned}$$

Now we see,
$P(Y \geq 1) = P(e^{t(X-x)}) \geq 1) = P(t(X-x) \geq 0)$

If $t > 0$
$P(t(X-x) \geq 0) = P(X - x \geq 0) = P(X \geq x)$
Therefore $P(X \geq x) = P(Y \geq 1) \leq \phi_X(t)e^{-tx}$
or $P(X \geq x) \leq \phi_X(t)e^{-tx}$

If $t < 0$
$P(t(X-x) \geq 0) = P(X - x \leq 0) = P(X \leq x)$
Therefore $P(X \leq x) = P(Y \geq 1) \leq \phi_X(t)e^{-tx}$
or $P(X \leq x) \leq \phi_X(t)e^{-tx}$

Given $X = X_1 + X_2 + ... + X_n$, where X_is are n independent Bernoulli variables with $E(X_i) = p_i$ and $\mu = \sum_{i=1}^n p_i$.
For $t > 0$ and $\delta > 0$ from the above proven first inequality with $x = (1+\delta)\mu$ we have,

$$\begin{aligned}
P(X > (1+\delta)\mu) &\leq P(X \geq (1+\delta)\mu) \\
&\leq e^{-t(1+\delta)\mu}\phi_X(t) \\
&\leq \frac{\phi_X(t)}{e^{t(1+\delta)\mu}} \\
&\leq \frac{\phi_{\sum_{i=1}^n X_i}(t)}{e^{t(1+\delta)\mu}} \\
&\leq \frac{\prod_{i=1}^n \phi_{X_i}(t)}{e^{t(1+\delta)\mu}} && \because X_1, X_2, ...X_n \textit{ are independent} \\
&\leq \frac{\prod_{i=1}^n (1 - p_i + p_i e^t)}{e^{t(1+\delta)\mu}} \\
&\leq \frac{\prod_{i=1}^n (1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\
&\leq \frac{\prod_{i=1}^n e^{p_i(e^t-1)}}{e^{t(1+\delta)\mu}} && \because 1 + k \leq e^k \\
&\leq \frac{e^{\sum_{i=1}^n p_i(e^t-1)}}{e^{t(1+\delta)\mu}} \\
&\leq \frac{e^{(e^t-1)\sum_{i=1}^n p_i}}{e^{t(1+\delta)\mu}} \\
&\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} && \because \sum_{i=1}^n p_i = \mu
\end{aligned}$$

As the equation is valid for any $t \geq 0$, to tighten the bound we must choose t such that the RHS of the inequality gets minimised.
We have RHS =

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

Since e^x is an increasing function in $\mathbb{R}$, to minimise the above we must minimise the exponent term i.e. $(e^t-1)\mu - t(1+\delta)\mu$.

$$\begin{aligned} Let\ P(t) &= (e^t-1)\mu - t(1+\delta)\mu \\ P'(t) &= \mu(e^t-1) - (\delta)\mu \\ P''(t) &= \mu e^t > 0 \end{aligned}$$

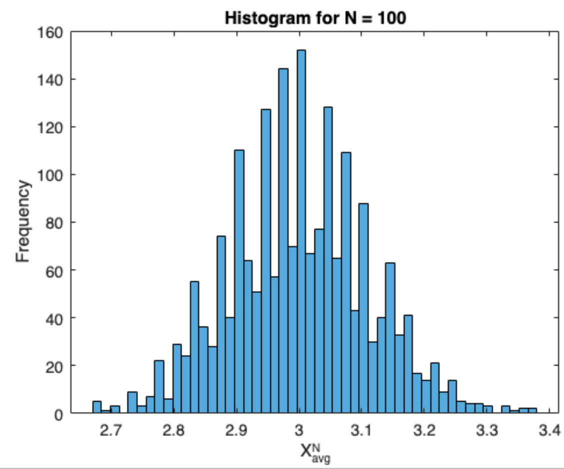
Setting $P'(t) = 0$ we have $t = ln(1+\delta)$
Since $P''(t) > 0$ we have indeed obtained a minima in P(t)
Putting the optimal $t = ln(1+\delta)$ in the equation we get the tighter bound,

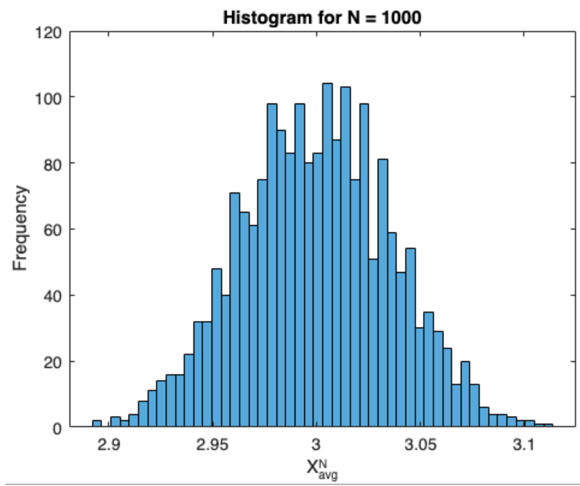
$$P(ln(1+\delta)) = \delta\mu - \mu(1+\delta)ln(1+\delta)$$

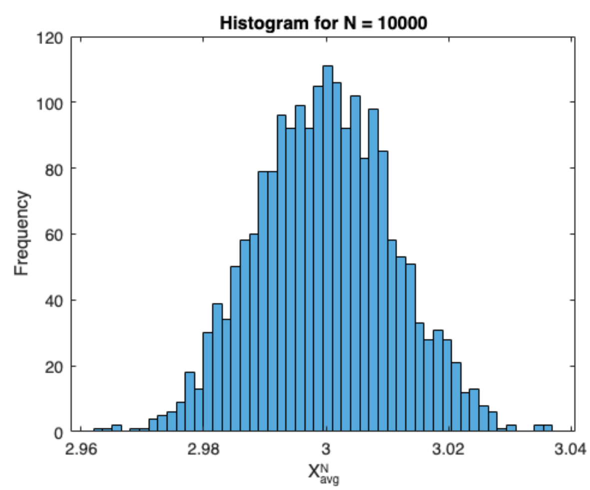
Putting it in original equation we have

$$\begin{aligned} P(X > (1+\delta)\mu) &\leq e^{\delta\mu - \mu(1+\delta)ln(1+\delta)} \\ &\leq \frac{e^{\delta\mu}}{e^{\mu(1+\delta)ln(1+\delta)}} \\ &\leq \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \end{aligned}$$

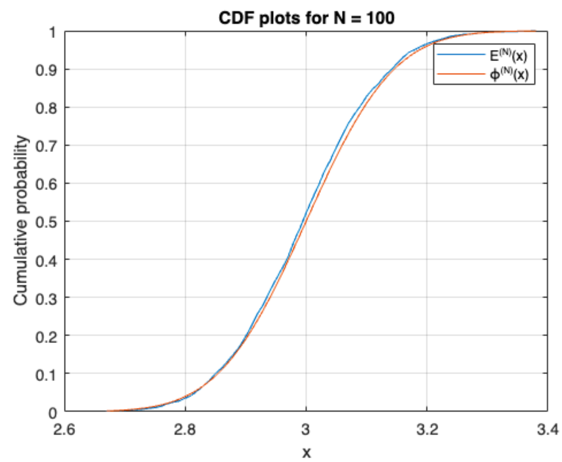
5. (a) The histogram plots for N corresponding to 100,1000 and 10000 are-

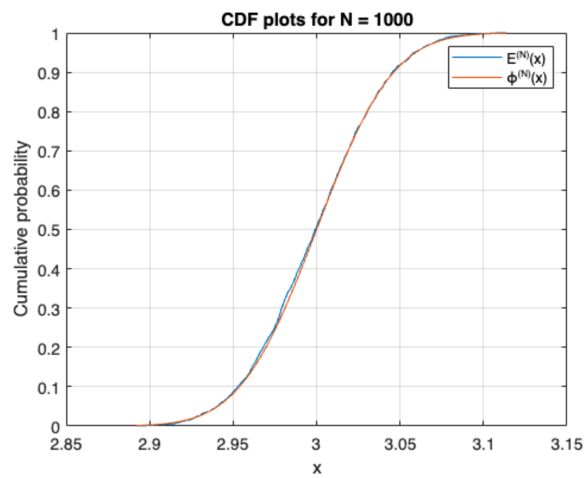


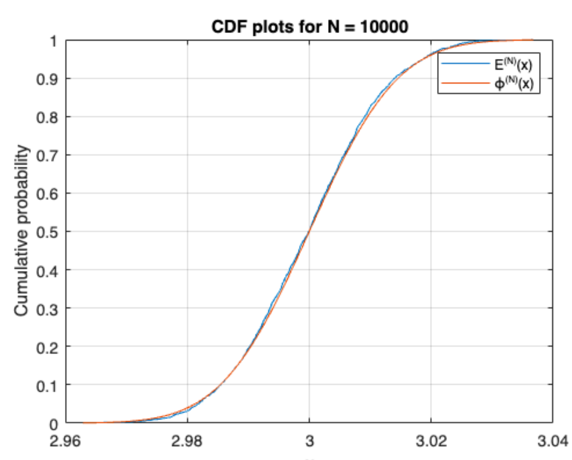




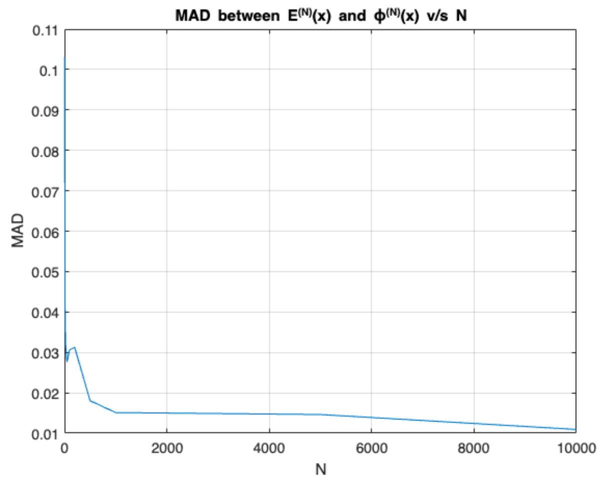
(b) The empirical CDF and gaussian CDF plots for N corresponding to 100,1000 and 10000 are-







(c) MAD as a function of N calculated over $N \in 5, 10, 20, 50, 100, 200, 500, 1000, 5000, 10000$



6. For the same reason, Quadratic Mean Information too is mimimum for t_x = -1.

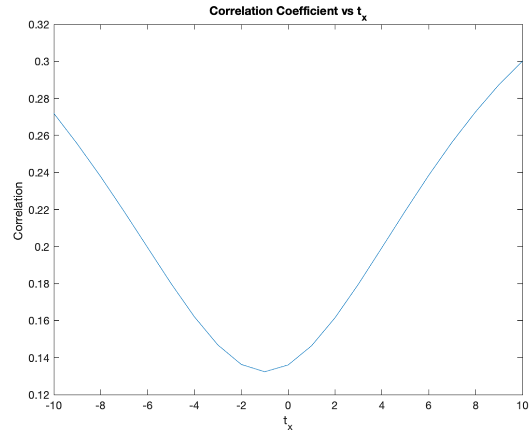


Figure 1: MATLAB plot for Correlation Coefficient as the shift in the image

Both images, even though representing the same physical information, have very low correlation because the correlation coefficient is a pixel by pixel comparison which is highly misleading when such detailed comparison does not yield a realistic picture. The photos given are negatively correlated in their middle portion(not middle portion of graph, middle portion of the images) but positively in the outer portion thus the overall effect is that the images are less correlated (according to correlation coefficient). Thus when shifted, the high correlation of the outer part is barely affected but the low correlation in inner parts is broken (again due to pixel by pixel comparison) and thus the overall correlation goes up slightly.

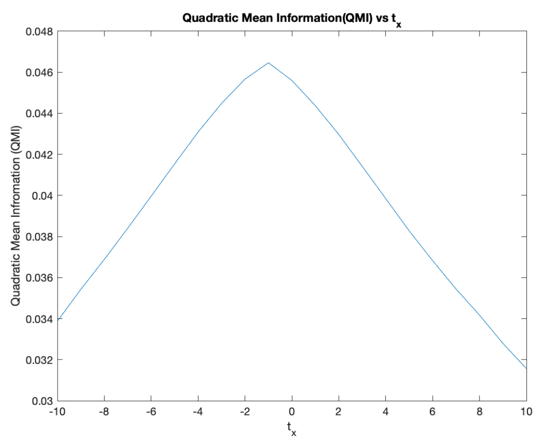


Figure 2: MATLAB plot for Quadratic Mean Information(QMI) as the shift in the image

QMI is a measure of the mutual dependence of two variables, i.e. it will be zero in the case of two mutually independent variables. We can see that QMI decreases as the shift increases(though the change is not much), i.e. mutual dependence decreases, which in some sense is correlation. Thus QMI is a better measure for measuring dependence in such cases.

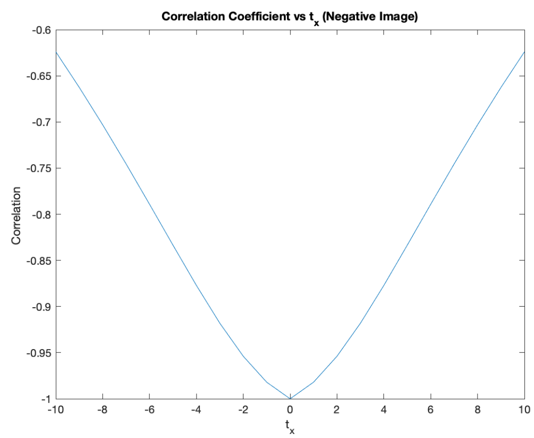


Figure 3: MATLAB plot for Correlation Coefficient as the shift in the image

Here the second image is an exact negative of the first image hence the correlation coefficient a pixel by pixel comparison measure, shows a correlation of -1 which is the expected result as 0 shift. As the shift increases the negative correlation decreases, which is also fairly obvious, and expected.

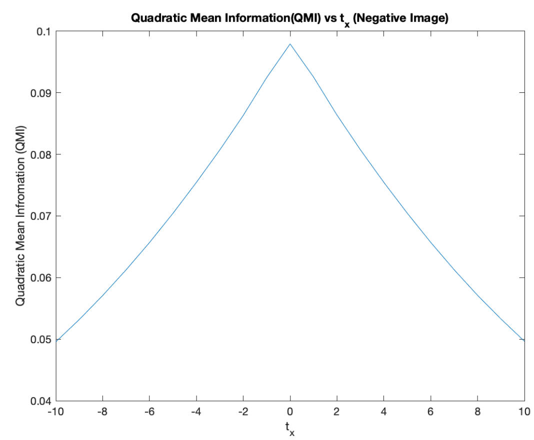


Figure 4: MATLAB plot for Quadratic Mean Information(QMI) as the shift in the image

Here we see that the QMI is still giving the expected results i.e. high dependence at 0 shift and decreasing dependence on increasing shift. So, even in "exact" (sort of) scenarios, where correlation coefficient serves as a good measure(In fact, works better than in the previous case, as is evident from the change), QMI is also a good measure.
Thus, we can say that QMI is a better dependence measure than the correlation coefficient in all aspects.

7.

$$\mathbf{X} = [X_1, X_2, \cdots, X_n]$$
$$\mathbf{t} = [t_1, t_2, \cdots, t_n]$$

$$\Phi_X(t) = E(e^{t_1X_1+t_2X_2+\cdots+t_nX_n})$$

For non-diagonal elements consider the partial derivative of $\Phi_X(t)$ wrt t_i and then t_j, we get:

$$\frac{\partial\Phi}{\partial t_i} = E(X_i e^{t^TX})$$

Now taking partial derivative wrt to t_j:

$$\frac{\partial}{\partial t_j}\frac{\partial\Phi}{\partial t_i} = E(X_iX_j e^{t^TX})$$

For $\mathbf{t} = [0]$,

$$\frac{\partial}{\partial t_j}\frac{\partial\Phi}{\partial t_i} = E(X_iX_j)$$

We already know, MGF of Multinomial distribution is

$$\Phi_X(t) = (\sum_i p_i e^{t_i})^n$$

where p_i is the probability of ith category.

$$\frac{\partial\Phi}{\partial t_i} = n(\sum p_i e^{t_i})^{n-1} p_i e^{t_i}$$
$$\frac{\partial}{\partial t_j}\frac{\partial\Phi}{\partial t_i} = n(n-1)(\sum p_i e^{t_i})^{n-2} p_i e^{t_i} p_j e^{t_j}$$
$$\frac{\partial}{\partial t_j}\frac{\partial\Phi}{\partial t_i}|_{t=0} = n(n-1)p_ip_j + 0$$
$$E[X_i, X_j] = n(n-1)p_ip_j$$

$$E[X_i] = \frac{\partial\Phi}{\partial t_i}|_{t=0} = np_i$$
$$E[X_j] = \frac{\partial\Phi}{\partial t_j}|_{t=0} = np_j$$
$$E[X_i, X_j] = n(n-1)p_ip_j$$

For the diagonal elements we need to take the partial derivatives with respect to i twice,

$$\frac{\partial\Phi}{\partial t_i} = E(X_i e^{t^TX})$$
$$\frac{\partial^2\Phi}{\partial t_i^2} = E(X_i^2 e^{t^TX})$$

For $\mathbf{t} = [0]$,

$$\frac{\partial^2 \Phi}{\partial t_i^2} = E(X_i^2)$$

From the multinomial distribution MGF,

$$\begin{aligned}
&\Phi_X(t) = (\sum_i p_i e^{t_i})^n \\
&\frac{\partial \Phi}{\partial t_i} = n(\sum p_i e^{t_i})^{n-1} p_i e^{t_i} \\
&\frac{\partial^2 \Phi}{\partial t_i^2} = n(n-1)(\sum p_i e^{t_i})^{n-2} p_i^2 e^{2t_i} + n(\sum p_i e^{t_i})^{n-1} p_i e^{t_i} \\
&\frac{\partial^2 \Phi}{\partial t_i^2}|_{t=0} = n(n-1)p_i^2 + np_i \\
&E[X_i^2] = n(n-1)p_i^2 + np_i
\end{aligned}$$

Let $\mathbf{C}$ be the co-variance matrix,

$$\begin{aligned}
C(i,j) &= E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= E[X_i X_j] - E[X_i]E[X_j] \\
&= n(n-1)p_i p_j - n^2 p_i p_j \\
&= -np_i p_j \\
C(i,i) &= E[(X_i - \mu_i)(X_i - \mu_i)] \\
&= E[(X_i - \mu_i)^2] \\
&= E[X_i^2] - E[X_i]E[X_i] \\
&= E[X_i^2] - E[X_i]^2 \\
&= n(n-1)p_i^2 + np_i - n^2 p_i^2 \\
&= np_i - np_i^2 \\
&= np_i(1 - p_i)
\end{aligned}$$

PS - Instructions to run MATLAB codes for 5th and 6th questions

The 5th question MATLAB code can be run as follows-
Press F5 to print all the plots (histograms and CDFs) for all the given N. The histograms and the CDFs make up the first 20 plots, each N having it's histogram and CDF plot. The 21th plot is the MAD plot.The name of the file is `question5.m`

For question 6 there are two MATLAB scripts (viz. `q6.m`, `q6negative.m`). q6.m is for comparing T1.jpg and T2.jpg and `q6_negative.m` for comparing T1.jpg with it's negative. Press F5 to print both the plots and save them in the working directory. First plot is of Co-variance vs t_x and the second is of QMI vs t_x.

2) X is a mixture of K Normal R.V.

$$\Rightarrow \text{pdf}(x) = f_X(x) = \sum_{i=1}^{K} p_i f_{X_i}(x) \qquad \left[f_{X_i}(x) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu_i)^2}{\sigma_i^2}\right)} \right]$$

$$\Rightarrow E(X) = \int_{-\infty}^{\infty} x \sum_{i=1}^{K} p_i f_{X_i}(x)\,dx$$

$$E(X) = \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} \frac{x}{\sigma_i\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}\,dx$$

→ This expression is the same as $E(X_i)$

$$\boxed{\Rightarrow E(X) = \sum_{i=1}^{K} (p_i \mu_i)}$$

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \left(\int_{-\infty}^{\infty} x^2 \sum_{i=1}^{K} p_i f_{X_i}(x)\,dx\right) - \left(\sum p_i x_i\right)^2$$

$$= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} x^2 f_{X_i}(x)\,dx - \left(\sum p_i x_i\right)^2$$

↳ same as $E(X_i^2)$

Now, $E(X_i^2) = \text{Var}(X_i) + (E(X_i))^2$
$= \sigma_i^2 + \mu_i^2$

$$\boxed{\Rightarrow \text{Var}(X) = \left[\sum_{i=1}^{K} p_i (\sigma_i^2 + \mu_i^2)\right] - \left(\sum_{i=1}^{K} p_i x_i\right)^2}$$

- $MGF(X) = \phi_X(t) = \sum_{i=1}^{k} p_i \, \phi_{X_i}(t)$ (as X is X_i with prob. p_i)

$$= \sum_{i=1}^{k} p_i \, e^{\mu_i t + \sigma_i^2 t^2/2}$$ (using formula of MGF of a Normal R.V.)

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

$$Y = \sum_{i=1}^{K} p_i X_i \qquad (X_i \text{ are i.i.d. random variable})$$

$$E(Y) = E\left(\sum_{i=1}^{K} p_i X_i\right) = \sum_{i=1}^{K} p_i \, E(X_i) = \boxed{\sum_{i=1}^{K} p_i \mu_i}$$

$$Var(Y) = Var\left(\sum_{i=1}^{K} p_i X_i\right)$$

$$= \sum_{i=1}^{K} p_i^2 Var(X_i) \qquad (\text{as } X_i \text{ are i.i.d})$$

$$= \sum_{i=1}^{K} p_i^2 \sigma_i^2$$

$$MGF(Y) = \phi_Y(t) = E\left(e^{t \sum p_i X_i}\right)$$

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

Now, as X_i are independent, so are $e^{t p_i X_i}$, thus –

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

$$\phi_Y(t) = \prod_{i=1}^{K} e^{\mu_i (p_i t) + p_i^2 \sigma_i^2 t^2/2} \qquad \left(\text{using } \phi_{X_i}(p_i t)\right)$$

$$\boxed{\phi_Y(t) = e^{t\left(\sum_{i=1}^{K} p_i \mu_i\right) + \left(\sum_{i=1}^{K} p_i^2 \sigma_i^2\right) t^2/2}}$$

Thus comparing this with the MGF of a Normal Random Variable, we can see that Y is a normal Random Variable with mean $= \sum_{i=1}^{K} p_i \mu_i$ and variance $= \sum_{i=1}^{K} p_i^2 \sigma_i^2$

$$\Rightarrow \boxed{pdf(Y) = f_Y(y) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{-(y-\mu_y)^2 / 2\sigma_y^2}}$$

where, $\mu_y = \sum_{i=1}^{K} p_i \mu_i$

$\sigma_y^2 = \sum_{i=1}^{K} p_i^2 \sigma_i^2$