Assignment 2: CS 215

Ayan Tanwar 22b0931

Nihar Hingrajia 22b1044

Sarthak Chauraisa 22b1014

September 4, 2023

1. Let $X_1, X_2, ..., X_n$ be $n > 0$ independent identically distributed random variables with cdf $F_X(x)$ and pdf $f_X(x) = F'_X(x)$. Derive an expression for the cdf and pdf of $Y_1 = \max(X_1, X_2, ..., X_n)$ and $Y_2 = \min(X_1, X_2, ..., X_n)$ in terms of $F_X(x)$. [10 points]

   Answer 1 :

   We are given $n > 0$ independent identically distributed random variables $X_1, X_2, ..., X_n$ with cdf $F_X(x)$ and pdf $f_X(x) = F'_X(x)$. Y_1 is given as the maximum of all the given random variables and Y_2 is their minimum.

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

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

   Let cumulative distribution function (cdf) of Y_1 be $F_{Y_1}(x)$. Using the definition of cdf, $F_{Y_1}(x) = \mathbb{P}(Y_1 \leq x)$.
   Let X_{i_0} be the maximum of all X_i. So, $F_{Y_1}(x) = \mathbb{P}(X_{i_0} \leq x)$.

   As $X_{i_0} \leq x$ and since X_{i_0} is the greatest, $X_i \leq x$ for all $i \in 1, 2, ..., n$.Hence,

$$F_{Y_1}(x) = \mathbb{P}(X_1 \leq x; X_2 \leq x; ...X_n \leq x)$$

   All the random variables are independent. So,

$$\begin{aligned}\mathbb{P}(X_1 \leq x; X_2 \leq x; ...X_n \leq x) &= \mathbb{P}(X_1 \leq x) \cdot \mathbb{P}(X_2 \leq x) \cdots \mathbb{P}(X_n \leq x) \\ &= (F_X(x)) \cdot (F_X(x)) \cdots (F_X(x)) \cdot \\ &= (F_X(x))^n\end{aligned}$$

   The cdf of Y_1, i.e., is $F_{Y_1}(x) = (F_X(x))^n$
   The probability density function (pdf) is the derivative of the cdf. The pdf of Y_1, i.e., $f_{Y_1}(x) = \frac{\partial F_{Y_1}(x)}{\partial x}$

$$\begin{aligned}f_{Y_1}(x) &= n \cdot (F_X(x))^{n-1} \cdot \frac{\partial F_X(x)}{\partial x} \\ &= n \cdot (F_X(x))^{n-1} \cdot f_X(x)\end{aligned}$$

   Cumulative distribution function (cdf) of Y_2, $F_{Y_2}(x) = \mathbb{P}(Y_2 \leq x)$.
   Let X_{j_0} be the minimum of all X_i . So, $F_{Y_2}(x) = \mathbb{P}(X_{j_0} \leq x)$.
   When $X_{j_0} \leq x$, all X_i greater than X_j can be greater than, less than or equal to x. Then the event when all X_i are such that $(X_{j_0} \leq x)$ is the complement of the event when all $X_i > x$. Hence, using the properties of independence of the random variables (as used above), we get

$$\begin{aligned}\mathbb{P}(X_{j_0} \leq x) &= 1 - \mathbb{P}(X_{j_0} > x) \\ &= 1 - \mathbb{P}(X_1 > x; X_2 > x; ...X_n > x) \\ &= 1 - \mathbb{P}(X_1 > x) \cdot \mathbb{P}(X_2 > x) \cdots \mathbb{P}(X_n > x) \\ &= 1 - (1 - \mathbb{P}(X_1 \leq x)) \cdot (1 - \mathbb{P}(X_2 \leq x)) \cdots (1 - \mathbb{P}(X_n \leq x)) \\ &= 1 - (1 - F_X(x)) \cdot (1 - F_X(x)) \cdots (1 - F_X(x)) \\ &= 1 - (1 - F_X(x))^n\end{aligned}$$

   The cdf of Y_2, i.e., is $F_{Y_2}(x) = 1 - (1 - F_X(x))^n$
   We get the pdf of Y_2 , i.e., $f_{Y_2}(x) = \frac{\partial(1-(1-F_X(x)))^n}{\partial x}$.

$$\begin{aligned}f_{Y_2}(x) &= -n \cdot (1 - F_X(x))^{n-1} \cdot (-\frac{\partial F_X(x)}{\partial x}) \\ &= n \cdot f_X(x) \cdot (1 - F_X(x))^{n-1}\end{aligned}$$

2. We say that a random variable X belongs to a Gaussian mixture model (GMM) if $X \sim \sum_{i=1}^{K} p_i \mathcal{N}(\mu_i, \sigma_i^2)$ where p_i is the 'mixing probability' for each of the K constituent Gaussians, with $\sum_{i=1}^{K} p_i = 1; \forall i, 0 \leq p_i \leq 1$. To draw a sample from a GMM, we do the following: (1) One of the K Gaussians is randomly chosen as per the PMF $\{p_1, p_2, ..., p_K\}$ (thus, a Gaussian with a higher mixing probability has a higher chance of being picked). (2) Let the index of the chosen Gaussian be (say) m. Then, you draw the value from $\mathcal{N}(\mu_m, \sigma_m^2)$.
   If X belongs to a GMM as defined here, obtain expressions for $E(X), \mathrm{Var}(X)$ and the MGF of X.

Now consider a random variable of the form $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for each $i \in \{1, 2, ..., K\}$. Define another random variable $Z = \sum_{i=1}^{K} p_i X_i$ **where** $\{X_i\}_{i=1}^{K}$ **are independent random variables**. Derive an expression for $E(Z), \text{Var}(Z)$ and the PDF, MGF of Z. [2+2+2+2+2+2+3=15 points]

Answer 2 :

At the end.

3. Using Markov's inequality, prove the following one-sided version of Chebyshev's inequality for random variable X with mean μ and variance σ^2: $P(X - \mu \geq \tau) \leq \dfrac{\sigma^2}{\sigma^2 + \tau^2}$ if $\tau > 0$, and $P(X - \mu \geq \tau) \geq 1 - \dfrac{\sigma^2}{\sigma^2 + \tau^2}$ if $\tau < 0$. [15 points]

Answer 3 :

The one-sided version of Chebyshev's inequality for random variable X with mean μ and variance σ^2 states that for

$$S_k = \{x_i : x_i - \mu \geq \tau\}$$

the probability that a given x_i belongs to this set S_k is given by

$$\frac{|S_k|}{N} \leq \frac{\sigma^2}{\sigma^2 + \tau^2}$$

given $\tau > 0$ and if $\tau < 0$ then,

$$\frac{|S_k|}{N} \geq 1 - \frac{\sigma^2}{\sigma^2 + \tau^2}$$

Markov's inequality states that for $a \geq 0$,

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

Lets first try to derive the Chebyshav's Inequality for $\tau > 0$.

Let us consider a variable $b \geq 0$ and $Y = X - \mu$, then

$$\mathbb{P}(X - \mu \geq \tau) = \mathbb{P}((X - \mu) + b \geq \tau + b) = \mathbb{P}(Y + b \geq \tau + b)$$

We know

$$\mathbb{P}((Y + b)^2 \geq (\tau + b)^2) \geq \mathbb{P}(Y + b \geq \tau + b)$$

because the L.H.S. includes all those x_i such that $Y + b \geq \tau + b$ or $Y + b \leq -(\tau + b)$ Therefore,

$$\mathbb{P}(X - \mu \geq \tau) \leq \mathbb{P}((Y + b)^2 \geq (\tau + b)^2)$$

Applying Markov's Inequality to the R.H.S. of the inequality,

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

We get

$$\begin{aligned} E((Y + b)^2 &= E(Y_2 + b^2 + 2Yb) \\ &= E(Y^2) + E(b^2) + 2 \cdot E(Yb) \end{aligned}$$

We know

$$E(Y^2) = \sigma^2$$
$$E(b^2) = b^2$$

Since $E(Y) = E(X - \mu) = E(X) - E(\mu) = \mu - \mu = 0$
Therefore,

$$E((Y + b)^2) = \sigma^2 + b^2$$

Hence,

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

To get the tightest bound possible, we will try to minimize the R.H.S. of the inequality with respect to the variable b. On differentiating the R.H.S. with respect to b and equating it to zero,

$$\frac{\partial(\frac{\sigma^2+b^2}{(\tau+b)^2})}{\partial b} = 0$$

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

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

Therefore, $b = \frac{\sigma^2}{\tau}$ is when the R.H.S. is minimum. Replacing the value of b on the R.H.S., we finally get

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

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

Now, lets derive the Chebyshav's Inequality for $\tau < 0$

$$\mathbb{P}(X-\mu \geq \tau) = \mathbb{P}((-X)-(-\mu) \leq (-\tau))$$

Let us consider a new random variable $A = -X$. The mean of the new random variable A is obviously $\mu_A = -\mu$ and the variance of A, $(\sigma_A)^2 = (-1)^2 \cdot (\sigma)^2 = \sigma^2$. Let $k = -\tau$ where $k \geq 0$ as $\tau < 0$. Therefore,

$$\begin{aligned}\mathbb{P}(X-\mu \geq \tau) &= \mathbb{P}(A-\mu_A < k)\\ &= 1-\mathbb{P}(A-\mu_A \geq k)\end{aligned}$$

In a similar manner as done above, we get

$$\mathbb{P}(A-\mu_A \geq k) \leq \frac{(\sigma_A)^2}{(\sigma_A)^2+k^2}$$

Therefore,

$$\mathbb{P}(X-\mu \geq \tau) \geq 1 - \frac{(\sigma_A)^2}{(\sigma_A)^2+k^2}$$

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

Hence Proved!

4. 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]

Answer 4 :

For the first part of the question
Markov's inequality can be stated as follows:

$$P(X \geq x) \leq \frac{E(X)}{x} \quad \text{for all} \quad x > 0$$

lets first solve for $t > 0$, here e^{tx} would be an increasing function, and we can write

$$P(X \geq x) = P(tX \geq tx) = P(e^{tX} \geq e^{tx})$$

Now using Markov's inequality

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

we know

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

Finally leading to

$$P(X \geq x) \leq e^{-tx}\phi_X(t) \quad for \quad t > 0$$

now lets solve the second one for $t < 0$, here e^{tx} would be an decreasing function, and we can write

$$P(X \geq x) = P(tX \geq tx) = P(e^{tX} \geq e^{tx})$$

Now using Markov's inequality

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

we know

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

Finally leading to

$$P(X \geq x) \leq e^{-tx}\phi_X(t) \quad for \quad t < 0$$

Now moving to the second part of the question
Here we intend to show

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

As in first part we can again write

$$P(X > (1+\delta)\mu) = P(tX > t(1+\delta)\mu) = P(e^{tX} > e^{t(1+\delta)\mu})$$

Again using Markov's inequality

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

now we need to solve for $E(e^{tX})$ for given n independent Bernoulli random variables, as they are independent their joint probability n variables can be written as product of probability of those n variables

$$E(e^{\sum_{i=1}^{n} tX_i}) = \sum_{x_1} e^{tx_1}P(X_1 = x_1) \cdot \sum_{x_2} e^{tx_2}P(X_2 = x_2) \cdots \sum_{x_n} e^{tx_n}P(X_n = x_n)$$

We can write each term as

$$\sum_{x_i} e^{tx_i}P(X_i = x_i) = (1 - p_i + p_i e^t)$$

Now the product of sum can be written as

$$\prod_i = 1^n(1 - p_i + p_i e^t)$$

Simplifying for each term

$$1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$$

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

Now summing up all we have proved the required inequality

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

5. Consider N independent random variables $X_1, X_2, ..., X_N$, such that each variable X_i takes on the values $1, 2, 3, 4, 5$ with probability $0.05, 0.4, 0.15, 0.3, 0.1$ respectively. For different values of $N \in \{5, 10, 20, 50, 100, 200, 500, 1000, 5000, 10000\}$, do as follows:

1. Plot the (empirically determined) distribution of the average of these random variables ($X_{avg}^{(N)} = \sum_{i=1}^{N} X_i/N$) in the form of a histogram with 50 bins.

2. Empirically determine the CDF of $X_{avg}^{(N)}$ using the `ecdf` command of MATLAB (this is called the empirical CDF). On a separate figure, plot the empirical CDF. On this, overlay the CDF of a Gaussian having the same mean and variance as $X_{avg}^{(N)}$. To get the CDF of the Gaussian, use the `normcdf` function of MATLAB.

3. Let $E^{(N)}$ denote the empirical CDF and $\Phi^{(N)}$ denote the Gaussian CDF. Compute the maximum absolute difference (MAD) between $E^{(N)}(x)$ and $\Phi^{(N)}(x)$ numerically, at all values x returned by `ecdf`. For this, read the documentation of `ecdf` carefully. Plot a graph of MAD as a function of N. [3+3+4 = 10 points]

Answer 5 :

1. Run the program **CS215_HW2_5.m** to generate the images given below, which represent the plot of Frequency value of the averages vs Average value of the randomly generated data.

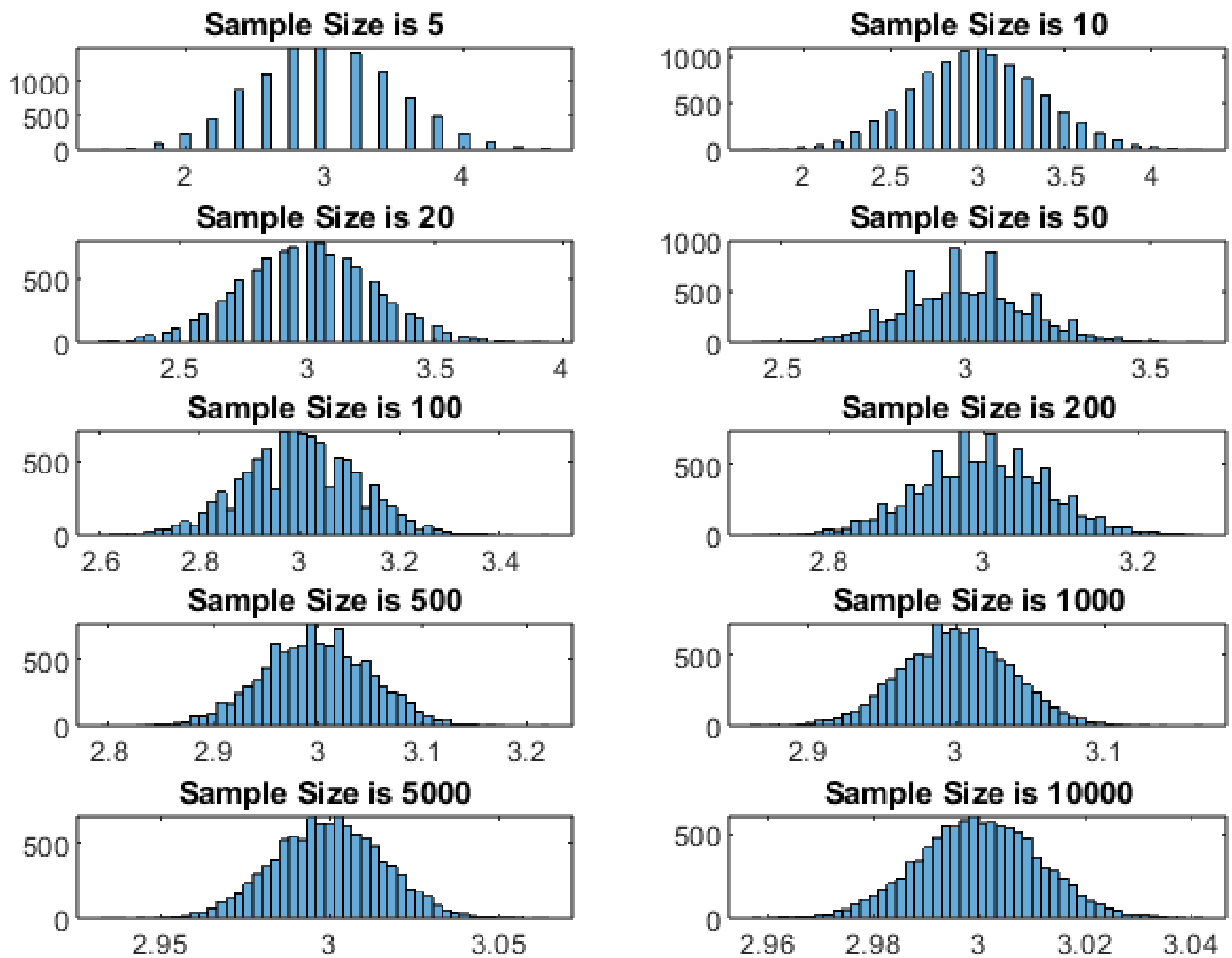

Figure 1: Histograms for Various Sample Size

2. Below are the Cummulative Frequency Distribution (CDF) plots which contains both. the empirically calculated CDF and the Gaussian CDF overlayed on same plot for various sample sizes. Which also shows that increasing sample size made the empirical calculation better as it approaches to gaussian CDF as Sample Size is increasing.

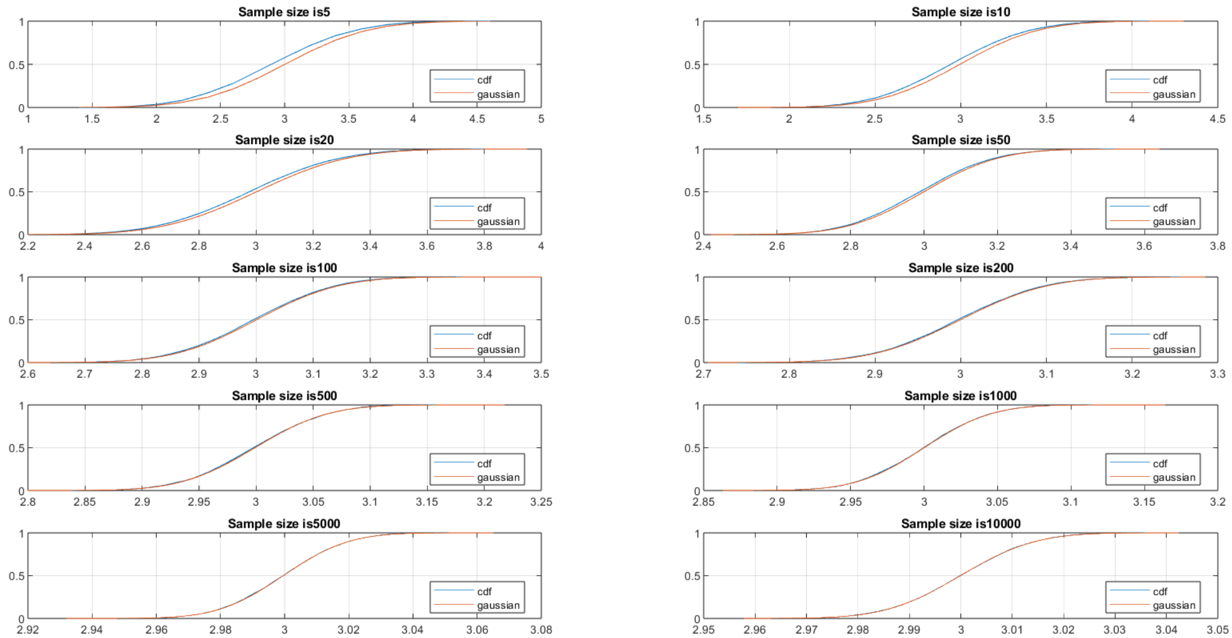

Figure 2: CDF for Various Sample Size

3. Below is the plot for Maximum Absolute Difference (MAD) vs Sample size (N).

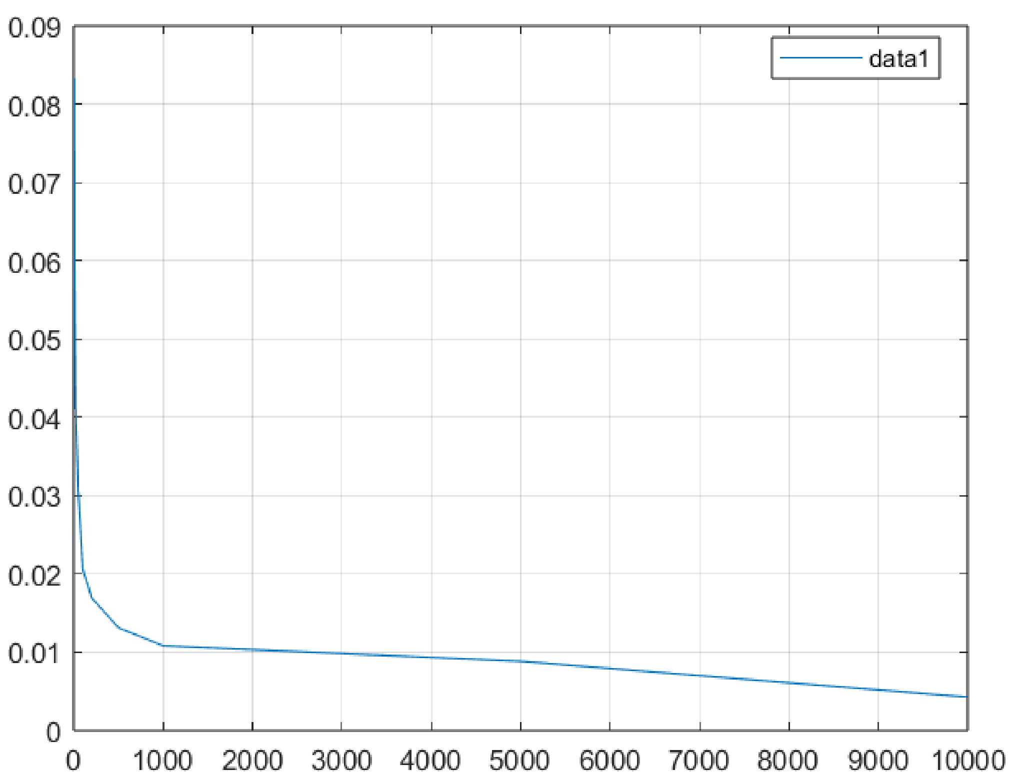

Figure 3: MAD vs Sample Size plot

6. 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

```
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 I_1, I_2 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:

- the correlation coefficient ρ,
- a measure of dependence called quadratic mutual information (QMI) defined as $\sum_{i_1}\sum_{i_2}(p_{I_1 I_2}(i_1, i_2) - p_{I_1}(i_1)p_{I_2}(i_2))^2$, where $p_{I_1 I_2}(i_1, i_2)$ represents the normalized joint histogram (*i.e.*, joint pmf) of I_1 and I_2 ('normalized' means that the entries sum up to one).

For computing the joint histogram, use a bin-width of 10 in both I_1 and I_2. For computing the marginal histogram, you need to integrate the joint histogram along one of the two directions respectively. You should write your own joint histogram routine in MATLAB - do not use any inbuilt functions for it. Plot a graph of the values of ρ versus t_x, and another graph of the values of QMI versus t_x.

Repeat exactly the same steps when the second image is a negative of the first image, i.e. $I_2 = 255 - I_1$.

Comment on all the plots. In particular, what do you observe regarding the relationship between the dependence measures and the alignment between the two images? Your report should contain all four plots labelled properly, and the comments on them as mentioned before. [25 points]

Answer 6 :

1. To quantitatively assess this correlation between these two correlated images T1 and T2, we employ two metrics: the correlation coefficient and the QMI.

   From the correlation coefficient graph, we notice that the images exhibit their lowest correlation when tx = 1. Consequently, this plot might suggest that the images are poorly correlated, especially when the second image is shifted by -1 pixel. However, the QMI graph reaches its maximum at the same tx value.

   Given our knowledge that the two images are highly similar, we can argue that the QMI metric is a superior choice for quantifying image similarity compared to the correlation coefficient. This is because the QMI metric offers a more sophisticated analysis by considering complex relationships within the images. QMI accurately assesses the two images and indicates a high correlation when the second image is shifted by 1 pixel to the left. This implies that image 1 aligns perfectly with image 2 when the latter is moved one pixel to the left.

   The apparent contradiction between QMI and the correlation coefficient arises from the fact that the correlation coefficient relies solely on mean and variance, while QMI examines the joint probability density function. As a result, the correlation coefficient may yield a lower correlation value despite the actual high correlation between the images.

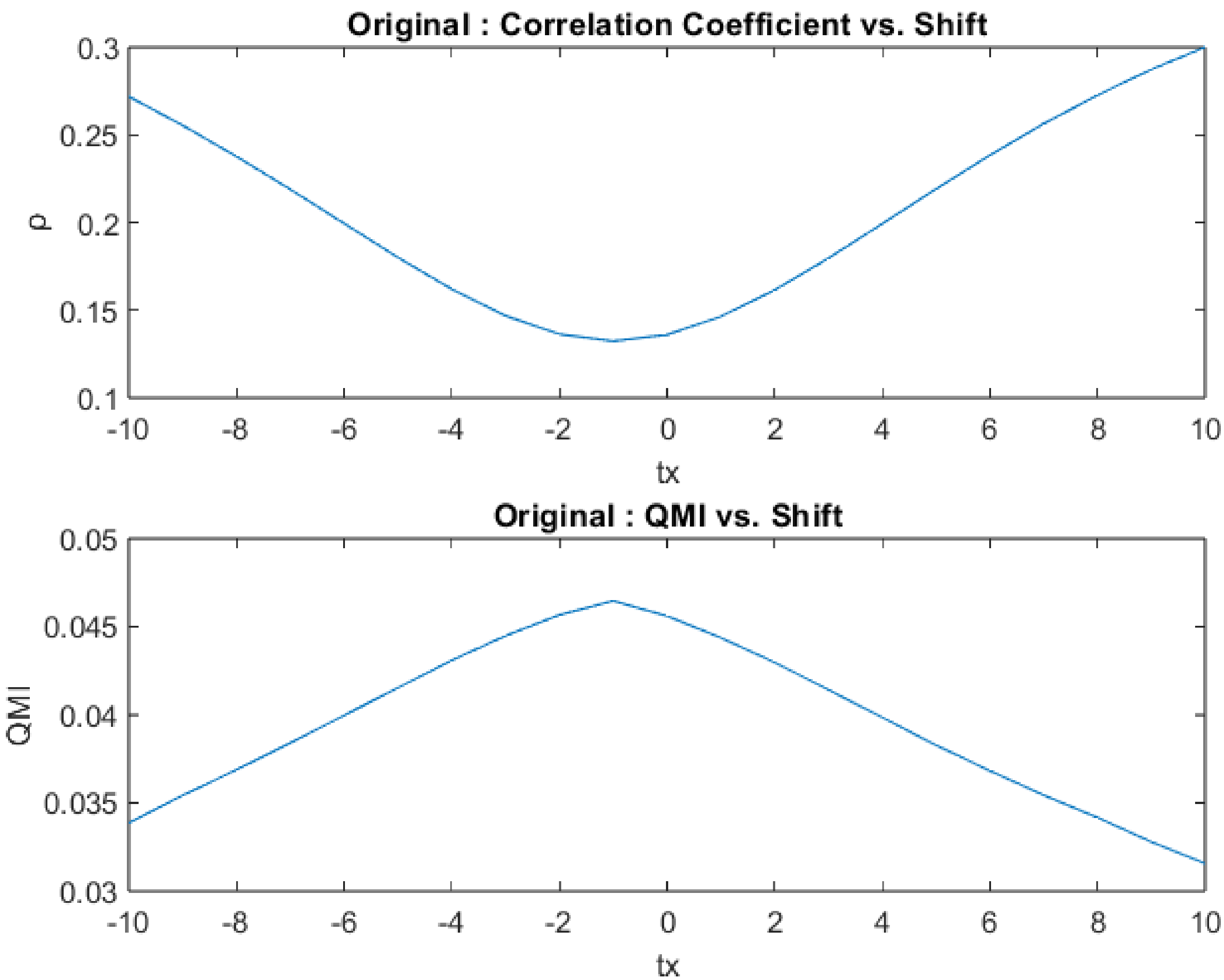

Figure 4: Original Image plots

2. In this context, the correlation coefficient plot exhibits symmetry and demonstrates a complete negative correlation when tx = 0. This outcome is expected because the other image is essentially the negative of the other, resulting in a high correlation with T1 but in the opposite direction. For other values of tx, the images deviate from being exact opposites, leading to an increase in the correlation coefficient.

   The QMI values remain positive as both images still exhibit similarity. There is also a peak in QMI at tx = 0, indicating that the images are most similar in this instance. As the images become less similar due to offset, the QMI decreases. It's important to note that QMI is always non-negative and only represents the magnitude of correlation, without indicating whether it's a positive or negative correlation. In this case, since alignment was achieved at tx = 0, both plots have their extrema at this point.

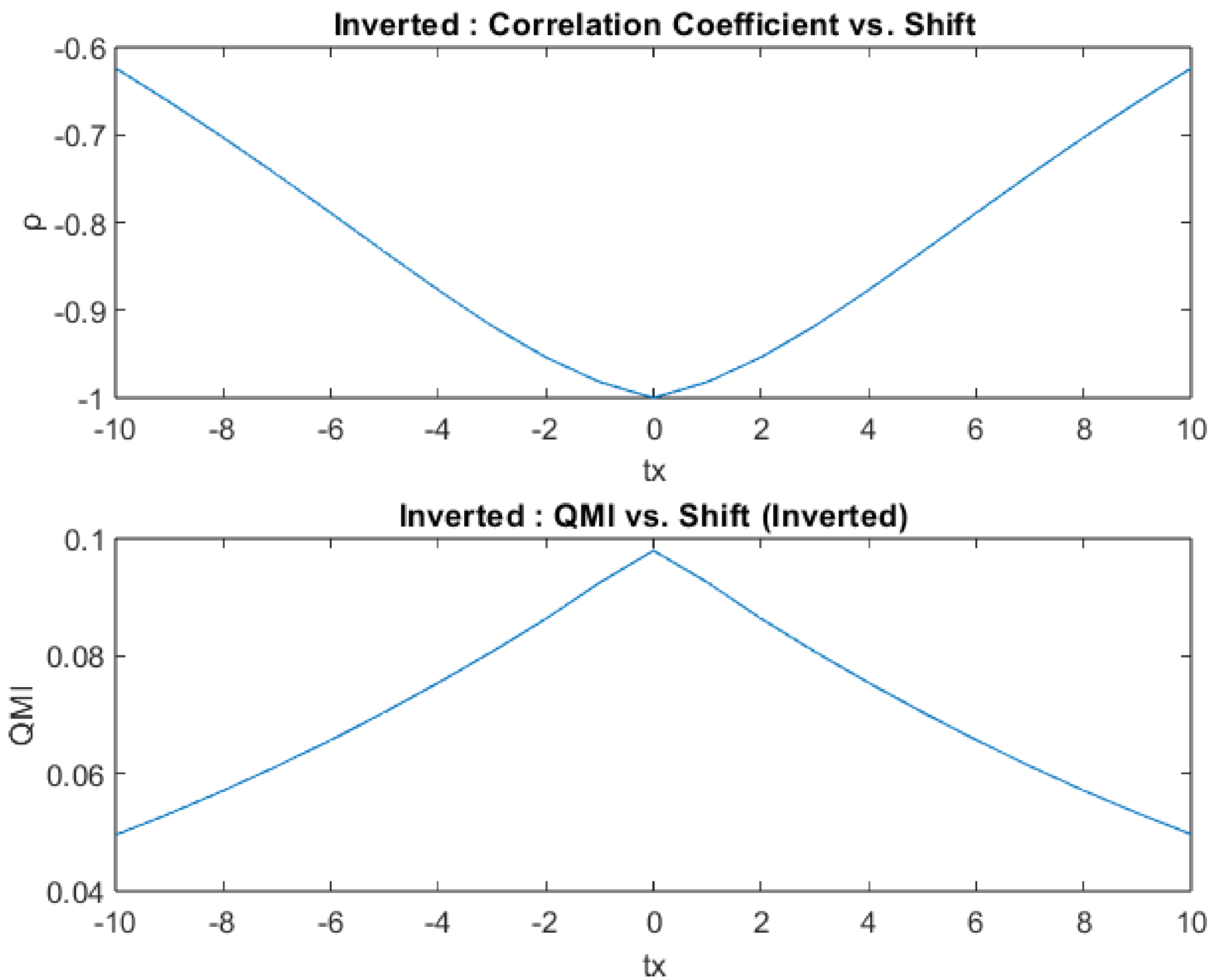

Figure 5: Inverted Image plots

7. Derive the covariance matrix of a multinomial distribution using moment generating functions. You are not allowed to use any other method. Since a covariance matrix $\boldsymbol{C}$ is square and symmetric, it is enough to derive expression for the diagonal elements C_{ii} and the off-diagonal elements $C_{ij}, i \neq j$. [10 points]

Answer 7 :

To derive the covariance matrix C of a multinomial distribution using moment generating function (MGF), we will first assume some random variables X_1,$X_2 \cdots$X$_k$
Lets also assume a random variable X as:

$$X = t_1X_1 + t_2X_2 \cdots + t_kX_k$$

Now the moment generating function of a binomial distribution can be represented as:

$$\phi_{X_i}(t_i) = E[e^{t_iX_i}] = (1 - p_i + p_ie^{t_i})^n \tag{7.1}$$

Now from this the moment generating function of multinomial distributions can be given as:

$$\phi_X(t) = E[e^{t_1X_1}]E[e^{t_1X_1}] \cdot \cdot [e^{t_1X_1}] = \prod_{i=1}^{k}(1 - p_i + p_ie^{t_i})^n$$

Now deriving $\phi_{X_i+X_j}(t_i, t_j)$

$$\begin{aligned}
\phi_{X_i+X_j}(t_i, t_j) &= (1 - p_i + p_i e^{t_i})^n (1 - p_j + p_j e^{t_j})^n \\
&= (1 + p_i(e^{t_i} - 1) + p_j(e^{t_j} - 1) + p_i p_j(e^{t_i+t_j} - e^{t_i} - e^{t_j} + 1))^n
\end{aligned}$$

We can further calculate:

$$\begin{aligned}
E[X_i] = \left.\frac{\partial \phi_{X_i}(t_i)}{\partial t_i}\right|_{t_i=0} &= np_i e^{t_i}(1 - p_i + p_i e^{t_i})^{n-1} \\
&= np_i \cdot 1 \cdot (1 - p_i + p_i \cdot 1)^{n-1} \\
&= np_i
\end{aligned}$$

$$\begin{aligned}
E[X_i^2] = \left.\frac{\partial^2 \phi_{X_i}(t_i)}{\partial t_i^2}\right|_{t_i=0} &= np_i e^{t_i}(1 - p_i + p_i e^{t_i})^{n-1} + np_i e^{t_i}((n-1)p_i e^{t_i})(1 - p_i + p_i e^{t_i})^{n-2} \\
&= np_i + n(n-1)p_i^2
\end{aligned}$$

$$\begin{aligned}
E[X_i \cdot X_j] = \left.\frac{\partial^2 \phi_{X_i+X_j}(t_i, t_j)}{\partial t_i \partial t_j}\right|_{t_i=0} &= \frac{\partial}{\partial t_j} \frac{\partial(1 + p_i(e^{t_i} - 1) + p_j(e^{t_j} - 1) + p_i p_j(e^{t_i+t_j} - e^{t_i} - e^{t_j} + 1))^n}{\partial t_i} \\
&= \frac{\partial(n(p_i e^{t_i} + p_i p_j(e^{t_i+t_j} - e^{t_i}))^{n-1})}{\partial t_j} \\
&= n(n-1)(p_i p_j(e^{t_i+t_j}))^{n-2} \quad i \neq j
\end{aligned}$$

Now we can write the general terms of the covariance matrix C as:

$$\begin{aligned}
C_{ij} &= E[X_i \cdot X_j] - E[X_i] \cdot E[X_j] \\
&= n(n-1)(p_i p_j(e^{t_i+t_j}))^{n-2} - np_i \cdot np_j \\
&= -n \cdot p_i \cdot p_j \\
C_{ij} &= -n \cdot p_i \cdot p_j \qquad\qquad i \neq j
\end{aligned}$$

$$\begin{aligned}
C_{ii} &= E[X_i^2] - E[X_i]^2 \\
&= np_i + n(n-1)p_i^2 - (np_i)^2 \\
C_{ii} &= n \cdot p_i \cdot (1 - p_i)
\end{aligned}$$

So now the covariance matrix can be written as:

$$\begin{aligned}
C_{ij} &= n \cdot p_i \cdot (1 - p_i) \qquad for \quad i = j \\
&= -n \cdot p_i \cdot p_j \qquad\qquad for \quad i \neq j
\end{aligned}$$

ANSWER: 2

Part: 1

It is given that a random variable X belongs to a Gaussian mixture model, if

$$X \sim \sum_{i=1}^{K} p_i \mathcal{N}(\mu_i, \sigma_i^2)$$

where p_i is the mixing probability for each of the K constituent Gaussians with

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

$$\forall i, \quad 0 \le p_i \le 1$$

Expectation of X

The pdf of X can be obtained using the concept of conditional probability on X_i.

$$P(X=x) = \sum_{i=1}^{K} P(X=X_i) P(X_i=x \mid X=X_i)$$

$$= \sum_{i=1}^{K} (p_i) \cdot \left(\frac{1}{\sqrt{2\pi}\,\sigma_i} e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}} \right)$$

$$E(X) = \int_{-\infty}^{\infty} x \, P(X=x) \, dx$$

$$= \int_{-\infty}^{\infty} x \sum_{i=1}^{k} \left(P(X=X_i) \, P(X_i=x \mid X=X_i) \right) dx$$

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

$$= \sum_{i=1}^{K} p_i E(X_i)$$

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

Variance of X

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

$$E(X^2) = \int_{-\infty}^{\infty} x^2 P(X=x)\, dx$$

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

Similar to the above approach,

$$E(X^2) = \sum_{i=1}^{K} p_i E(X_i^2)$$

$$= \sum_{i=1}^{k} p_i (\sigma_i^2 + \mu_i^2)$$

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

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

MGF of X

$$\text{MGF of } X = \phi_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\, P(X=x)\, dx$$

$$= \int_{-\infty}^{\infty} e^{tx} \left(\sum_{i=1}^{k} p_i \cdot \frac{1}{\sqrt{2\pi}\,\sigma_i} \cdot e^{\frac{(x-\mu_i)^2}{2\sigma_i^2}} \right) dx$$

$$= \sum_{i=1}^{k} p_i \left(\int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_i} \cdot e^{\frac{(x-\mu_i)^2}{2\sigma_i^2}}\, dx \right)$$

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

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

Part : 2

It is given, $Z = \sum_{i=1}^{k} p_i X_i$ where $\{X_i\}_{i=1}^{k}$ are independent random variables.

$$X_i \sim N(\mu_i, \sigma_i^2) \text{ for } i \in \{1, 2, \ldots, k\}$$

Expectation of Z

Using the linear properties of expectation operator,

$$E(Z) = E\left(\sum_{i=1}^{k} p_i X_i\right)$$

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

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

Variance of Z

$$Var(Z) = E\left(\sum_{i=1}^{k} p_i X_i\right)$$

We know that where the random variables are independent, variance of their sum is same as the sum of their variance (given in slides)

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

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

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

MGF of Z

$$\text{MGF of } Z = E(e^{tz}) = E\left(e^{\sum_{i=1}^{k} t p_i x_i}\right)$$

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

(given in slides, $E(XY) = E(X)E(Y)$ if X, Y are independent)

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

$$= \prod_{i=1}^{K} \phi_{x_i}(p_i t)$$

$$= \prod_{i=1}^{K} e^{p_i \mu_i t + \frac{\sigma_i^2 p_i^2 t^2}{2}}$$

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

$$\boxed{\phi_Z(t) = e^{\left(t E(Z) + \frac{t^2}{2} Var(Z)\right)}}$$

PDF of Z

Notice that the MGF of Z is that of a gaussian random variable.

$$\boxed{\therefore \text{ PDF of } Z = P(Z=z) = \frac{1}{\sqrt{2\pi \, Var(Z)}} \cdot e^{-\frac{(x - E(Z))^2}{2 Var(Z)}}}$$