# CS215 ASSIGNMENT2

23B1032-SRIKAR NIHAL
23B0924-B.SOMANATH VASANTH
23B1055-ANIRUDH

September 2024

## Contents

**1 Mathemagic** **3**
1.1 Task A . . . . . . . . . . 3
1.2 Task B . . . . . . . . . . 3
1.3 Task C . . . . . . . . . . 3
1.4 Task D . . . . . . . . . . 4
1.5 Task E . . . . . . . . . . 4
1.6 Task F . . . . . . . . . . 5
1.7 Task G . . . . . . . . . . 6

**2 Normal Sampling** **6**
2.1 Task A . . . . . . . . . . 6
2.2 Task B . . . . . . . . . . 7
2.3 Task C . . . . . . . . . . 7
2.4 Task D . . . . . . . . . . 8
2.5 Task E . . . . . . . . . . 8

**3 Fitting Data** **10**
3.1 Task A . . . . . . . . . . 10
3.2 Task B . . . . . . . . . . 10
3.3 Task C . . . . . . . . . . 10
3.4 Task D . . . . . . . . . . 11
3.5 Task E . . . . . . . . . . 13
3.6 Task F . . . . . . . . . . 13

**4 Quality in Inequalities** **15**
4.1 Task A . . . . . . . . . . 15
4.2 Task B . . . . . . . . . . 16
4.3 Task C . . . . . . . . . . 16
4.4 Task D . . . . . . . . . . 17
4.5 Task E . . . . . . . . . . 19

**5 A Pretty "Normal" Mixture** **20**
5.1 Task A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
5.2 Task B . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
5.3 Task C . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
5.4 Task D . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25

# 1 Mathemagic

## 1.1 Task A

PGF when X is a bernoullis random variable: bernoullis random variable has only two possible values 1 and 0 by substituting this in given farmula we get
$pgf(Z) = p[x = 1]Z^1 + p[x = 0]Z^0 = pZ + 1 - p = 1 + p(Z - 1)$

## 1.2 Task B

PGF for Binomial distribution : for distribution $bin(n, p)$

$G(Z) = (1-p)^n p^0 Z^0 \binom{n}{0} + (1-p)^{n-1} p^1 Z^1 \binom{n}{1} + \cdots + (1-p)^0 p^n Z^n \binom{n}{n}$

$$G(Z) = \sum_{k=0}^{n} (1-p)^{n-k} p^k Z^k \binom{n}{k}$$

by binomial expansion

$G(Z) = (1 + p(Z-1))^n = (Gbern(Z))^n$
(Gbern from task A)

## 1.3 Task C

Take two independent random variables X,Y.
finding E(XY):

$$E(XY) = \sum_x \sum_y xy\, p(X = x, Y = y)$$

$$p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$$

$$E(XY) = \sum_x \sum_y xy\, p(X = x) \cdot p(Y = y)$$

$$E(XY) = \left( \sum_x x\, p(X = x) \right) \cdot \left( \sum_y y\, p(Y = y) \right)$$

$$E(XY) = E(X) \cdot E(Y)$$

Given $X = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent random variables. Consider the random variables $p_1 = Z^{X_1}$, $p_2 = Z^{X_2}$, ..., $p_n = Z^{X_n}$. Since $X_1, X_2, \ldots, X_n$ are independent, $p_1, p_2, \ldots, p_n$ are also independent random variables.

The expectation of the product $p_1 p_2 \cdots p_n$ can be computed as follows:

$$\mathbb{E}(p_1 p_2 \cdots p_n) = \mathbb{E}\left(Z^{X_1} Z^{X_2} \cdots Z^{X_n}\right) = \mathbb{E}\left(Z^X\right) = G_X(Z)$$

where $G_X(Z)$ is the probability generating function (PGF) of $X$.
By the independence of $p_i$, we have:

$$\mathbb{E}(p_1 p_2 \cdots p_n) = \mathbb{E}(p_1)\mathbb{E}(p_2)\cdots\mathbb{E}(p_n)$$

Furthermore, for each $i$, the expectation $\mathbb{E}(p_i)$ is given by:

$$\mathbb{E}(p_i) = \mathbb{E}\left(Z^{X_i}\right) = G_{X_i}(Z)$$

where $G_{X_i}(Z)$ is the PGF of $X_i$. Therefore:

$$G_X(Z) = \mathbb{E}(p_1 p_2 \cdots p_n) = \mathbb{E}(p_1)\mathbb{E}(p_2)\cdots\mathbb{E}(p_n) = (G_{X_i}(Z))^n$$

for each $i$, as $X_1, X_2, \ldots, X_n$ are identically distributed.

### 1.4 Task D

PGF for geometric distribution : for distribution $Geo(p)$
we Know $P(X = n) = p \cdot (1-p)^{n-1}$

$$G(Z) = p \cdot Z^1 + (1-p) \cdot p \cdot Z^2 + (1-p)^2 \cdot p \cdot Z^3 + \cdots$$

$$G(Z) = Z \cdot p \cdot \left(1 + (1-p) \cdot Z + (1-p)^2 \cdot Z^2 + \cdots\right)$$

it is an infinite gp series so

$$G(Z) = \frac{Z \cdot p}{1 - (1-p) \cdot Z}$$

### 1.5 Task E

now finding PGF for negative binomial random variable: for

$$Y \sim \text{NegBin}(n, p)$$

we know that negative binomial random variable is the sum of n independent random variables.

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

by the result of the task C we can now write

$$G_{\text{NegBin}}(Z) = (G_{\text{geo}}(Z))^n$$

$$G_{\text{NegBin}}(Z) = \left(\frac{pZ}{1 - Z(1-p)}\right)^n$$

from taskB

$$G_{\text{Bin}}\left(n, \frac{1}{p}\right)(Z^{-1}) = \left(1 + \frac{1}{p}\left(\frac{1}{Z} - 1\right)\right)^n = \left(\frac{1 + pZ - Z}{pZ}\right)^n$$

$$\left[G_{\text{Bin}}\left(n, \frac{1}{p}\right)(Z^{-1})\right]^{-1} = \left(\frac{pZ}{1 + pZ - Z}\right)^n = G_{\text{NegBin}}(Z)$$

proves the statement given in qn

## 1.6 Task F

we can write other form for the pdf of negative binomial rv by direct definition.

$$G_{\text{NegBin}}(Z) = \sum_{r=n}^{\infty} p^n (1-p)^{r-n} Z^r \binom{r-1}{n-1}$$

we know that $G_{\text{NegBin}}(Z) = \left(\frac{pZ}{1-Z(1-p)}\right)^n$ now put z=1 in the prev farmula,substitute p-1 as x and bring p outside..then
$G_{\text{NegBin}}(Z) = p^n \cdot \left(\frac{1}{1+x}\right)^n$
by equating above two verions of Gnegbin we get.

$$p^n \left(\frac{1}{1+x}\right)^n = \sum_{r=n}^{\infty} p^n (-x)^{r-n} \binom{r-1}{n-1}$$

by cancelling $p^n$ on both sides we get:

$$\left(\frac{1}{1+x}\right)^n = \sum_{r=n}^{\infty} (-x)^{r-n} \binom{r-1}{n-1}$$

now substitute q=r-n.

$$\left(\frac{1}{1+x}\right)^n = \sum_{q=0}^{\infty} (-x)^q \binom{q+n-1}{n-1}$$

now rename the variable q with r to get the form given in task.

$$\left(\frac{1}{1+x}\right)^n = \sum_{r=0}^{\infty} (-1)^r (x)^r \binom{r+n-1}{n-1}$$

given in qn

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$$

finding $\binom{-n}{r}$

$$\binom{-n}{r} = \frac{-n\cdot(-n-1)\cdots(-n-r+1)}{r!} = \frac{(-1)^r \cdot (n+r-1)!}{(n-1)!\cdot r!} = (-1)^r \binom{n+r-1}{r}$$

now substitue this in our original derived eqn .we get

$$\left(\frac{1}{1+x}\right)^n = \sum_{r=0}^{\infty} (x)^r \binom{-n}{r}$$

this proves the given farmula.

### 1.7 Task G

calculating $G'(Z)at(z=1) = (\sum_{n=0}^{\infty} P(X=n)z^n)'at(z=1) = (\sum_{n=0}^{\infty} P(X=n)(n))$
This is equal to E(X).this proves the given statement.
now calculating E(X) for bernoulis RV
$G(Z) = 1 + p(Z-1)$
by differentiating it and substituting 1 we get $E(X) = p$.
for binomial rv.
$G(Z) = (1 + p(Z-1))^n$

$$G'(Z) = n \cdot (1 + p \cdot (Z-1)) \cdot p$$

by substituting Z=1.

$$G'(Z)\Big|_{Z=1} = np$$

(expected value for binomial variable)
for geometric random variable

$$G(Z) = \frac{Z \cdot p}{1 - (1-p) \cdot Z}$$

$$g'(Z) = \frac{p}{(1 + pZ - Z)^2}$$

$$g'(Z)\Big|_{Z=1} = \frac{1}{p}$$

(expected value of geometric random variable) for negetive binomial random variable

$$G_{\text{negbin}}(Z) = (G_{\text{geo}}(Z))^n$$

$$G'\text{negbin}(Z) = n \cdot (G\text{geo}(Z))^{n-1} \cdot G'_{\text{geo}}(Z)$$

by substituting z=1

$$G'\text{negbin}(Z)\Big|Z = 1 = \frac{n}{p}$$

# 2 Normal Sampling

### 2.1 Task A

Finding CDF of RV Y:

$$Y_{\text{CDF}}(k) = \begin{cases} 0 & \text{if } k < 0 \\ 1 & \text{if } k > 1 \\ P(Y < k) = P(F_X(X) < k) = P(X < F_X^{-1}(k)) = (F_X(F_X^{-1}(k))) = k & \text{if } 0 \le k \le 1 \end{cases}$$

We used the fact that cdf is between 0 and one for the cases k¡0 and k¿1. We get the below step by the definition of cdf $F_X(X)$ which is for case where k is between zero and one.

$$P(X < F_X^{-1}(k)) = (F_X(F_X^{-1}(k)))$$

now differentiating cdf to get pdf.

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0 & \text{if } x > 1 \\ 1 & \text{if } 0 < x < 1 \end{cases}$$

This is the uniform distribution between 0 and 1. which proves the given statement.

## 2.2 Task B

Algorithm:
Given Y is a uniform random variable between 0 and 1. we know that $F_X(X)$ is a invertible function.
So we can get the inverse of this function.Let g be the inverse of the the $F_X(X)$. Define the new wanted random variable as Z=g(Y) where y is the given uniform random variable.

Proof for the newly chosen random variable is indeed the given random variable:

Now let us calculate the cdf of the new random variable Z:

$$\begin{aligned} \text{CDF}_Z(x) &= P(Z < x) \\ &= P(g(Y) < x) \\ &= P\left(Y < g^{-1}(x)\right) \\ &= P\left(Y < F_X(x)\right) \\ &= F_X(x) \end{aligned}$$

I used the fact that $P(Y < k)$ for a uniform random variable between zero and 1 is k if k is between 0 and 1. so now our newly defined random variable has same cdf so these both are same random variables.

## 2.3 Task C

code is in 2c.py you have to do ”python3 2c.py” to get the plot and plot will be saved in the same directory with name 2c.png. In the code first i calculated a uniform random variable between 0 and 1 by using np.random.uniform.Then by using the algorithm of prev task i calculated the guassian random variable.(by applying inverse of cdf of gaussian variable to the calculated uniform random variable it is present as ppf in scipy.stats).then i plotted graph for the asked samples.i used density=true in graph which gives the p(x) for the given gaussian distribution.The graphs came as expected.if the variance is more the graph is spread more.if the mean is higher the graphs height is more.

### 2.4 Task D

Code is in 2d.py. "python3 2d.py" to run command and the images are in 2d1.png,2d2.png,2d3.png. In code I exactly applied what told in question.I took a random variable which becomes o or 1 with equal probability and then i calculated the position by moving forward if 1 and backward if 0.so then i plotted graph. By looking at the shape of the tops of the histogram we can see it is symmetric and maximum at middle,closely resemble normal distribution. As the number of balls and heights are increasing the graph will become closer to the normal distribution. sum of a large number of independent random variables give a normal distribution.

### 2.5 Task E

Given: h = 2k

We have to calculate the probability:

$$P_h(X = 2i) = ?$$

If the position is 2i and let the number of forward steps be n, then:
Backward steps = h - n.
Totally,

$$n + (-1) \times (h - n) = 2i$$

$$2n - h = 2i$$

$$n = \frac{h}{2} + i$$

For h times steps with probability $\frac{1}{2}$ to go forward and backward, it is like a binomial distribution. Therefore,

$$P(X = 2i) = \binom{h}{\frac{h}{2} + i} \left(\frac{1}{2}\right)^h$$

Finally,

$$P(X = i) = \binom{h}{\frac{h+i}{2}} \left(\frac{1}{2}\right)^h$$

$$P_h(x = 2i) = \binom{h}{k + i} \left(\frac{1}{2}\right)^h \quad \text{substitute} \quad k = \frac{h}{2}$$

$$= \frac{h!}{(k + i)!(k - i)!} \left(\frac{1}{2}\right)^h \text{since} \quad k = \frac{h}{2}$$

Using Stirling's approximation:

$$P_n(x=2i) \approx \frac{\sqrt{2\pi h}\left(\frac{h}{e}\right)^h}{\sqrt{2\pi(k+i)2\pi(k-i)}\left(\frac{k+i}{e}\right)^{k+i}\left(\frac{k-i}{e}\right)^{k-i}}\left(\frac{1}{2}\right)^h$$

For $h \gg i$ and $k = \frac{h}{2} \gg i$, we have:

$$k - i = k + i \approx k$$

By simplifying:

$$p(x=2i) = \frac{1}{\sqrt{\pi k}} \frac{\left(\frac{h}{1}\right)^h}{\left(\frac{k+i}{1}\right)^{k+i}\left(\frac{k-i}{1}\right)^{k-i}}\left(\frac{1}{2}\right)^h$$

$$= \frac{1}{\sqrt{\pi k}} \frac{\left(\frac{h}{1}\right)^h}{\left(\frac{h+2i}{1}\right)^{\frac{h+2i}{2}}\left(\frac{h-2i}{1}\right)^{\frac{h-2i}{2}}}$$

Take ln of this function:

$$\ln\left(p(x=2i)\right) = -\ln\left(\sqrt{\pi k}\right) + h\ln h - \frac{h+2i}{2}\ln(h+2i) - \frac{h-2i}{2}\ln(h-2i)$$

$$\ln(h+2i) = \ln h + \ln\left(1 + \frac{2i}{h}\right) \quad (h >> i)$$

$$= \ln h + \frac{2i}{h} - \frac{2i^2}{h^2}$$

$$\ln(h-2i) = \ln h - \frac{2i}{h} - \frac{2i^2}{h^2}$$

$$\begin{aligned}\ln\left(P(X=2i)\right) = -\ln(\sqrt{\pi k}) + \ln h - \frac{h+2i}{2}\ln h - \frac{(h+2i)i}{h} + \frac{(h+2i)i^2}{h^2} \\ - \frac{(h-2i)}{2}\ln h + \frac{(h-2i)i}{h} + \frac{(h-2i)i^2}{h^2}\end{aligned}$$

$$\ln\left(P(X=2i)\right) = -\ln\left(\sqrt{\pi k}\right) - \frac{2i^2}{h}$$

$$\begin{aligned}P(X=2i) &\approx \frac{1}{\sqrt{\pi k}}e^{-\frac{2i^2}{h}} = \frac{1}{\sqrt{\pi k}}e^{-\frac{i^2}{k}} \\ P(X=i) &\approx \frac{1}{\sqrt{\pi k}}e^{-\frac{2i^2}{4h}} = \frac{1}{\sqrt{\pi k}}e^{-\frac{i^2}{4k}}\end{aligned}$$

this proves the given statement.

# 3 Fitting Data

## 3.1 Task A

This code loads numerical data from a file using **np.loadtxt** and calculates two key statistical moments. The first moment, computed using **np.mean(data)**, is the average (mean) of the data points. The second moment, calculated by **np.mean(np.square(data))**, represents the mean of the squared data values, which gives insight into the spread of the data. These moments are printed as part of the output. The first moment indicates central tendency, while the second moment relates to the data's dispersion.

## 3.2 Task B

This code loads data from a file and plots its histogram using matplotlib. The **plt.hist()** function creates a histogram with 100 bins, showing the probability density (density=True) of the data. The x-axis represents the data values, while the y-axis shows the probability density. The histogram is labeled with appropriate axis labels and a title. Finally, the plot is saved as **'3b.png'**.

The data is centered around a value of approximately 6, with most of the probability density concentrated between 3 and 9.

## 3.3 Task C

For a binomial distribution $X \sim \mathrm{Bin}(n, p)$, the first two moments $\mu_{\mathrm{Bin},1}$ and $\mu_{\mathrm{Bin},2}$ as a function of $n$ and $p$ are given by:

The first moment, or mean, of a binomial distribution is:

$$\mu_{\mathrm{Bin},1} = \mathbb{E}[X] = n \cdot p$$

‘

The second moment of a binomial distribution can be calculated as:

$$\mu_{\mathrm{Bin},2} = \mathbb{E}[X^2] = \mathrm{Var}(X) + (\mathbb{E}[X])^2$$

where the variance of the binomial distribution is given by:

$$\mathrm{Var}(X) = n \cdot p \cdot (1 - p)$$

Substituting the expressions for $\mathrm{Var}(X)$ and $\mathbb{E}[X]$, we get:

$$\mu_{\mathrm{Bin},2} = n \cdot p \cdot (1 - p) + (n \cdot p)^2$$
$$\mu_{\mathrm{Bin},2} = n \cdot p \cdot (1 - p) + n^2 \cdot p^2$$

The histogram appears to closely follow the binomial distribution's shape, especially around the central values, indicating that the data may be well-modeled

by a binomial distribution with the peak around 6.
The value of n* obtained from the given data 20.
The value of p* obtained from the given data is 0.32968652963756934.

### 3.4 Task D

For a Gamma distribution $X \sim \text{Gamma}(k, \theta)$, where $k$ is the shape parameter and $\theta$ is the scale parameter, the first two moments $\mu_{\text{Gamma},1}$ and $\mu_{\text{Gamma},2}$ as a function of $k$ and $\theta$ are given by: To derive the first moment (mean) of a gamma-distributed random variable $X$ with shape parameter $k$ and scale parameter $\theta$, we use the following PDF:

$$f_X(x) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}$$

The mean is given by:

$$\mathbb{E}[X] = \int_0^\infty x \cdot f_X(x)\, dx$$

Substitute the PDF into the integral:

$$\mathbb{E}[X] = \int_0^\infty x \cdot \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}\, dx$$

$$\mathbb{E}[X] = \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^k e^{-x/\theta}\, dx$$

Use the substitution $u = \frac{x}{\theta}$, hence $du = \frac{dx}{\theta}$:

$$\int_0^\infty x^k e^{-x/\theta}\, dx = \theta^{k+1} \int_0^\infty u^k e^{-u}\, du$$

The integral $\int_0^\infty u^k e^{-u}\, du$ is the gamma function $\Gamma(k+1)$:

$$\int_0^\infty x^k e^{-x/\theta}\, dx = \theta^{k+1}\Gamma(k+1)$$

Thus:

$$\mathbb{E}[X] = \frac{\theta^{k+1}\Gamma(k+1)}{\theta^k \Gamma(k)} = k\theta$$

$$\mu_{\text{Gamma},1} = \mathbb{E}[X] = k \cdot \theta$$

To derive the second moment of a gamma-distributed random variable $X$ with shape parameter $k$ and scale parameter $\theta$, we use the following PDF:

$$f_X(x) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}$$

The second moment is given by:

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot f_X(x)\, dx$$

Substitute the PDF into the integral:

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}\, dx$$

$$\mathbb{E}[X^2] = \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^{k+1}e^{-x/\theta}\, dx$$

Use the substitution $u = \frac{x}{\theta}$, hence $du = \frac{dx}{\theta}$:

$$\int_0^\infty x^{k+1}e^{-x/\theta}\, dx = \theta^{k+2} \int_0^\infty u^{k+1}e^{-u}\, du$$

The integral $\int_0^\infty u^{k+1}e^{-u}\, du$ is the gamma function $\Gamma(k+2)$:

$$\int_0^\infty x^{k+1}e^{-x/\theta}\, dx = \theta^{k+2}\Gamma(k+2)$$

Thus:

$$\mathbb{E}[X^2] = \frac{\theta^{k+2}\Gamma(k+2)}{\theta^k \Gamma(k)} = \theta^2 \frac{\Gamma(k+2)}{\Gamma(k)}$$

Using the property of the gamma function $\Gamma(k+2) = (k+1)\Gamma(k+1)$ and $\Gamma(k+1) = k\Gamma(k)$:

$$\Gamma(k+2) = (k+1)k\Gamma(k)$$

$$\mathbb{E}[X^2] = \theta^2 \frac{(k+1)k\Gamma(k)}{\Gamma(k)} = \theta^2(k^2 + k)$$

$$\mu_{\text{Gamma},2} = \mathbb{E}[X^2] = \text{Var}(X) + (\mathbb{E}[X])^2$$

where the variance of the Gamma distribution is given by:

$$\text{Var}(X) = k \cdot \theta^2$$

Substituting the expressions for $\text{Var}(X)$ and $\mathbb{E}[X]$, we get:

$$\mu_{\text{Gamma},2} = k \cdot \theta^2 + (k \cdot \theta)^2$$

$$\mu_{\text{Gamma},2} = k(k+1) \cdot \theta^2$$

The value of k* is 9.691205541246163.
The value of $\theta$* is 0.6703134703624323.

### 3.5 Task E

First, the sample's first and second moments are calculated. The parameters n and p for the Binomial distribution are found by solving two equations based on these moments. Similarly, the parameters k and $\theta$ for the Gamma distribution are obtained. After rounding the data, the log-likelihoods for both distributions are computed and averaged.
Average Log-Likelihood (Binomial): -2.157068115434681.
Average Log-Likelihood (Gamma): -2.1608217722067953
A larger likelihood is typically attributed to a better fit.Hence,the Binomial distribution provides a better fit.

### 3.6 Task F

To derive the third moment $E[x^3]$ of a random variable $x$ that follows a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, we start by computing the expectation $E[x^3]$. The third moment of a random variable $x$ is defined as:

$$E[X^3] = \int_{-\infty}^{\infty} x^3 f(x)\, dx$$

where $f(x)$ is the probability density function (PDF) of the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. The PDF for a Gaussian distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Thus, the third moment is:

$$E[X^3] = \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

To simplify the integration, we change variables to the standard normal form by defining $z = \frac{x-\mu}{\sigma}$, or equivalently, $x = \mu + z\sigma$. The differential $dx$ becomes $dz\sigma$, and the limits of integration stay from $-\infty$ to $+\infty$ because the transformation is linear.

Thus, the integral becomes:

$$E[X^3] = \int_{-\infty}^{\infty} (\mu + z\sigma)^3 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

Now, we expand $(\mu + z\sigma)^3$:

$$(\mu + z\sigma)^3 = \mu^3 + 3\mu^2 z\sigma + 3\mu z^2\sigma^2 + z^3\sigma^3$$

Substitute this expansion into the integral:

$$E[X^3] = \int_{-\infty}^{\infty} \left(\mu^3 + 3\mu^2 z\sigma + 3\mu z^2\sigma^2 + z^3\sigma^3\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

Each term in this expression needs to be integrated separately:
1. For $\mu^{3**}$:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 1$$

So, the contribution of this term is $\mu^3$.
2.For $3\mu^2 z\sigma^{**}$:

$$\int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 0$$

This integral is zero because the integrand is an odd function and the limits are symmetric about zero. Hence, this term vanishes.
3. For $3\mu z^2 \sigma^{2**}$:

$$\int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 1$$

Thus, the contribution of this term is $3\mu\sigma^2$.
4. For $z^3\sigma^{3**}$:

$$\int_{-\infty}^{\infty} z^3 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 0$$

This integral is also zero for the same reason as the second term (it is an odd function integrated over a symmetric interval). Now, summing up the non-zero contributions:

$$E[X^3] = \mu^3 + 3\mu\sigma^2$$

Thus, the third moment of a Gaussian distribution is:

$$E[X^3] = \mu^3 + 3\mu\sigma^2$$

$$E[X^3] = \sum_{i=1}^{k} p_i E(X_i^3) \tag{1}$$

$$\mathbb{E}[X^3] = p_1(\mu_1^3 + 3\mu_1\sigma_1^2) + p_2(\mu_2^3 + 3\mu_2\sigma_2^2)$$

Similarly we can write $E[X^4]$ as

$$\mathbb{E}[X^4] = p_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + p_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4)$$

Average Negative Log-Likelihood of GMM: 2.183038744911308.
This is not a better approximation since average Log-Likelihood is smaller as compared to binomial and gaussian.

# 4 Quality in Inequalities

## 4.1 Task A

**Markov's Inequality:**Let X be any non-negative random variable and a>0.

$$P[X \geq a] \leq \frac{E[X]}{a}$$

**Proof for Discrete Random Variables**

Consider $E(X)$:

$$E(X) = \sum_{x \in D} xP(X = x)$$

Since $X \geq 0$:

$$E(X) \geq \sum_{x \in D, x \geq a} xP(X = x)$$

Note that $X \geq a$ for all terms in the sum, so:

$$E(X) \geq aP(X \geq a)$$

Thus, we conclude that:

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

**Proof for Continuous Random Variables**

Consider $E(X)$:

$$E(X) = \int_D xdx$$

Where $f(x)$ is the probability density function (PDF).

$$E(X) \geq \int_{D, x \geq a} af(x)dx$$

Thus:

$$E(X) \geq aP(X \geq a)$$

This implies:

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

## 4.2 Task B

Given that Chebyshev-Cantelli inequality

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

$$\text{As we know that} \quad P(X \geq a) \leq P(X^2 \geq a^2),$$

$$P(X-\mu \geq \tau) = P(X-\mu+b \geq \tau+b)$$

As $\quad (X-\mu+b)^2 \quad$ is a positive random variable, from Markov's Inequality

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

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

$$P(X-\mu \geq \tau) \leq \frac{E[(X-\mu)^2]+b^2+2bE(X-\mu)}{(\tau+b)^2}$$

As $Var(X)$ is $\sigma^2$ and the value of $E(X-\mu)$ is 0

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

As b can be any value we choose $b = \frac{\sigma^2}{\tau}$
By substituting the value of b the above the above equation will be

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

Hence proved.

## 4.3 Task C

**Part 1:**For t>0
Given

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

**Proof:** Consider $M_X(t)$

$$M_X(t) = E(e^{tX}) \quad \text{(Moment generating function (MGF))}$$

As $t > 0$:

$$M_X(t) = \sum_{x \in D} e^{tx} P(X=x) \geq \sum_{x \in D, x \geq a} e^{tx} P(X=x)$$

$$M_X(t) \geq \sum_{x \in D, x \geq a} e^{ta} P(X = x) = e^{at} P(X \geq a)$$

$$\Rightarrow P(X \geq a) \leq e^{-at} M_X(t)$$

Now, put $a = x$:

$$P(X \geq x) \leq e^{-xt} M_X(t) \quad \forall t > 0$$

**Part 2:**For t<0
Given

$$P(X \leq x) \leq e^{-tx} M_X(t) \quad \forall t < 0$$

**Proof:** Consider $M_X(t)$

$$M_X(t) = E(e^{tX}) \quad \text{(Moment generating function (MGF))}$$

As $t < 0$:

$$M_X(t) = \sum_{x \in D} e^{tx} P(X = x) \geq \sum_{x \in D, x \leq a} e^{tx} P(X = x)$$

$$M_X(t) \geq \sum_{x \in D, x \leq a} e^{ta} P(X = x) = e^{at} P(X \leq a)$$

$$\Rightarrow P(X \leq a) \leq e^{-at} M_X(t)$$

Now, put $a = x$:

$$P(X \leq x) \leq e^{-xt} M_X(t) \quad \forall t < 0$$

## 4.4 Task D

Given $Y = \sum_{i=1}^{n} X_i$ where each $X_i$ has the same distribution and is independent of all other $X_j$'s

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right)$$

1.As E follows the linearity property

$$E(Y) = \sum_{i=1}^{n} E(X_i)$$

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

As all random variables have the same distribution, let all $p_i = p$, so:

$$E(Y) = \sum_{i=1}^{n} p = np$$

2. Given to show that

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

$$M_{X_i}(t) = \sum_{x \in \{0,1\}} e^{tx} P(X_i = x)$$

$$M_{X_i}(t) = e^0 P(X_i = 0) + e^t P(X_i = 1)$$

$$M_{X_i}(t) = e^{t(1-p)} + e^t p$$

$$M_{X_i}(t) = p(e^t - 1) + 1$$

$$M_Y(t) = E\left(e^{t \sum_{i=1}^{n} X_i}\right)$$

$$M_Y(t) = E\left(e^{tX_1 + tX_2 + \cdots + tX_n}\right)$$

$$M_Y(t) = \mathbb{E}(e^{tX_1} e^{tX_2} e^{tX_3} \cdots e^{tX_n})$$

As $X_i$ and $X_j$ are independent where $i \neq j$

$$M_Y(t) = \mathbb{E}(e^{tX_1}) \cdot \mathbb{E}(e^{tX_2}) \cdot \mathbb{E}(e^{tX_3}) \cdots \mathbb{E}(e^{tX_n})$$

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

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

$$= (1 + p(e^t - 1))^n \quad \text{(as all have the same } p_i\text{)}$$

From Task (C), we have shown that

$$P(Y \geq x) \leq e^{-tx} M_X(t), \quad \forall t > 0$$

So,

$$P\left(Y \geq (1+\delta)\mu\right) \leq e^{-(1+\delta)\mu} M_Y(t)$$

$$P\left(Y \geq (1+\delta)\mu\right) \leq e^{-(1+\delta)\mu} \left(1 + p(e^t - 1)\right)^n$$

$$P\left(Y \geq (1+\delta)\mu\right) \leq e^{-(1+\delta)\mu} \left[1 + \binom{n}{1} p(e^t - 1) + \binom{n}{2} p^2 (e^t - 1)^2 + \cdots + \binom{n}{n} p^n (e^t - 1)^n\right]$$

As $\binom{n}{r} \leq \frac{n^r}{r!}$

Replacing each $\binom{n}{r}$ with $\frac{n^r}{r!}$

$$P\left(Y \geq (1+\delta)\mu\right) \leq e^{-(1+\delta)\mu}\left(1 + \frac{np(e^t-1)}{1!} + \frac{n^2p^2(e^t-1)^2}{2!} + \cdots\right)$$

As

$$e^n = 1 + n + \frac{n^2}{2!} + \frac{n^3}{3!} + \cdots$$

$$P\left(Y \geq (1+\delta)\mu\right) \leq e^{-(1+\delta)\mu} \cdot e^{np(e^t-1)}$$

As $\mu = np$

$$P\left(Y \geq (1+\delta)\mu\right) \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)\mu}}$$

3.To improve the bound of $P(Y \geq (1+\delta)\mu)$, we have to substitute $t$ such that

$$\frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}} \text{ becomes minimum.}$$

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

For this to be minimum, minimizing $et - (1+\delta)t$ will be sufficient. Now, for finding the minimum of $e^t - (1+\delta)t$, we differentiate it and equate to zero and double differention at the obtained value of t will be positive: Hence it is minimum

$$e^t - (1+\delta) = 0$$

Thus, we have

$$t = \ln(1+\delta)$$

Substituting it in our original equation, we get

$$\frac{e^{\mu\delta}}{e^{(1+\delta)\ln(1+\delta)\mu}}$$

This will be the improved upper bound for Chernoff bound.

## 4.5 Task E

From the previous question we have

$$P(Y \geq (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$$

Here we can substitute $\ln(1+x)$ with $\frac{2x}{2+x}$ as $\ln(1+x)$ is greater than $\frac{2x}{2+x}$ for all $x > 0$ which can be proved by taking $f(x) = \ln(1+x) - \frac{2x}{2+x}$ and differentiating

it.
This further simplifies to

$$P(Y \geq (1+\delta)\mu) \leq e^{\frac{-\delta^2\mu}{2+\delta}}$$

Now,

$$P(Y \geq (1+\delta)E(X)) \leq e^{\frac{-\delta^2 E(X)}{2+\delta}}$$

Now, put $\delta = \frac{\epsilon}{\mu}$ and $E(X) = n\mu$, as we know $E(X) = n\mu$,

$$P(Y \geq (\mu+\epsilon)n) \leq e^{\frac{-\epsilon^2 n}{\mu^2+\epsilon}}$$

Now,

$$P(Y \geq (\mu+\epsilon)n) = P\left(\frac{Y}{n} \geq (\mu+\epsilon)\right)$$

$$P(A_n \geq \mu+\epsilon) \leq e^{\frac{-\epsilon'^2 n}{\mu(\mu+\epsilon)}}$$

As $n \to \infty$, RHS $\to 0$. For that, we can conclude

$$\lim_{n\to\infty} P(A_n \geq \mu+\epsilon) = 0$$

$$\Rightarrow \lim_{n\to\infty} P(A_n - \mu \geq \epsilon) = 0$$

$$\lim_{n\to\infty} P(A_n - \mu \geq \epsilon) = 0$$

Similarly, we can prove

$$\lim_{n\to\infty} P(\mu - A_n \geq \epsilon) = 0$$

using Chernoff tail bound. From the above equation, we can write

$$\lim_{n\to\infty} P(|A_n - \mu| \geq \epsilon) = 0$$

# 5 A Pretty "Normal" Mixture

## 5.1 Task A

We want to show that the PDF of both the GMM and the output random variable $\mathcal{A}$, generated by the algorithm, are equal.
The probability density function (PDF) of the Gaussian Mixture Model (GMM) is given by:

$$f_X(u) = \sum_{i=1}^{k} p_i f_{X_i}(u)$$

where: $p_i$ is the probability of selecting the $i$-th Gaussian component, $f_{X_i}(u)$ is the PDF of the $i$-th Gaussian component, which is $\mathcal{N}(\mu_i, \sigma_i^2)$.

The PDF of a Gaussian distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ is:

$$f_{X_i}(u) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(u-\mu_i)^2}{2\sigma_i^2}\right)$$

Thus, the GMM PDF is a weighted sum of these Gaussian PDFs.
Let $\mathcal{A}$ be the random variable generated by this process. To compute the PDF of $\mathcal{A}$, note that: - With probability $p_i$, the variable $\mathcal{A}$ is sampled from the Gaussian $X_i$, which has PDF $f_{X_i}(u)$.

Therefore, the PDF of $\mathcal{A}$ is also a weighted sum of the individual Gaussian PDFs:

$$f_{\mathcal{A}}(u) = \sum_{i=1}^{k} p_i f_{X_i}(u)$$

Since both the GMM PDF and the PDF of $\mathcal{A}$ are given by the same expression, we conclude that:

$$f_{\mathcal{A}}(u) = f_X(u)$$

Thus, the algorithm correctly samples from the GMM distribution.

## 5.2 Task B

1.Given $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$

$$P_X(x) = \sum_{i=1}^{k} p_i P_{X_i}(x)$$

where $P_X$ is the probability density of $x$, and $p_i$ is the probability.

$$E(X) = \int x P_X(x) dx$$

$$E(X) = \int x \sum_{i=1}^{k} p_i P_{X_i}(x) dx$$

$$E(X) = \sum_{i=1}^{k} p_i \int x P_{X_i}(x) dx$$

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

$$E(X) = \sum_{i=1}^{k} P p_i \mu_i$$

Thus,

$$E(X) = \sum_{i=1}^{k} P_i \mu_i$$

2. Given $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$

$$P_X(x) = \sum_{i=1}^{k} p_i P_{X_i}(x)$$

where $P_X(x)$ is the probability density function of $X$, and $p_i$ is the probability.

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

$$E[X^2] = \int x^2 P_X(x)\, dx$$

$$E[X^2] = \int x^2 \sum_{i=1}^{k} p_i P_{X_i}(x)\, dx$$

$$E[X^2] = \sum_{i=1}^{k} p_i \int x^2 P_{X_i}(x)\, dx$$

$$E[X^2] = \sum_{i=1}^{k} p_i E(X_i^2) \tag{1}$$

$$E(X_i^2) = \mathrm{Var}[X_i] + (E[X_i])^2$$

$$E(X_i^2) = \sigma_i^2 + \mu_i^2$$

Substituting this equation into (1),

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

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

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

3.Moment generating function of $X$ is

$$M_X(t) = \int e^{tx} \left(\sum_{i=1}^{k} p_i P(X_i = x)\right) dx$$

$$M_X(t) = \sum_{i=1}^{k} \int e^{tx} p_i P(X_i = x) dx$$

$$M_X(t) = \int e^{tx} \sum_{i=1}^{k} p_i P(X_i = x) dx$$

$$M_X(t) = \sum_{i=1}^{k} p_i \int e^{tx} P(X_i = x) dx$$

$$M_X(t) = \sum_{i=1}^{k} p_i M_{X_i}(t)$$

$$M_{X_i}(t) = e^{\mu_i t + (\sigma_i^2 t^2)/2} \quad \text{(As } X_i \text{ is Gaussian random variable)}$$

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

## 5.3 Task C

Let the random variable $Z$ be defined as a weighted sum of $k$ independent Gaussian random variables $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$:

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

where $p_i$ are constants.

**1.** $\mathbb{E}[Z]$

The expectation of a linear combination of independent random variables is the linear combination of their expectations:

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=1}^{k} p_i X_i\right] = \sum_{i=1}^{k} p_i \mathbb{E}[X_i] = \sum_{i=1}^{k} p_i \mu_i$$

**2. Var**$[Z]$

The variance of a linear combination of independent random variables is the sum of the variances of the individual random variables, each scaled by the square of its coefficient:

$$\text{Var}(Z) = \text{Var}\left(\sum_{i=1}^{k} p_i X_i\right) = \sum_{i=1}^{k} p_i^2 \text{Var}(X_i) = \sum_{i=1}^{k} p_i^2 \sigma_i^2$$

**3. The PDF $f_Z(u)$ of $Z$**

Since $Z$ is a linear combination of independent Gaussian random variables, $Z$ is also normally distributed. The distribution of $Z$ is:

$$Z \sim \mathcal{N}\left(\sum_{i=1}^{k} p_i\mu_i, \sum_{i=1}^{k} p_i^2\sigma_i^2\right)$$

The PDF of a normally distributed random variable with mean $\mu$ and variance $\sigma^2$ is given by:

$$f_Z(u) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left(-\frac{(u-\mu_Z)^2}{2\sigma_Z^2}\right)$$

where

$$\mu_Z = \sum_{i=1}^{k} p_i\mu_i, \quad \sigma_Z^2 = \sum_{i=1}^{k} p_i^2\sigma_i^2$$

Thus, the PDF of $Z$ is:

$$f_Z(u) = \frac{1}{\sqrt{2\pi\sum_{i=1}^{k} p_i^2\sigma_i^2}} \exp\left(-\frac{\left(u-\sum_{i=1}^{k} p_i\mu_i\right)^2}{2\sum_{i=1}^{k} p_i^2\sigma_i^2}\right)$$

**4. The Moment Generating Function (MGF) $M_Z(t)$ of $Z$**

The moment generating function (MGF) of a random variable $X$ is defined as $M_X(t) = \mathbb{E}[e^{tX}]$.

For a linear combination of independent Gaussian random variables $Z = \sum_{i=1}^{k} p_iX_i$, the MGF is the product of the MGFs of the individual Gaussian random variables:

$$M_Z(t) = \mathbb{E}\left[e^{tZ}\right] = \mathbb{E}\left[e^{t\sum_{i=1}^{k} p_iX_i}\right] = \prod_{i=1}^{k} \mathbb{E}\left[e^{tp_iX_i}\right]$$

The MGF of a Gaussian random variable $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ is:

$$M_{X_i}(t) = \exp\left(t\mu_i + \frac{t^2\sigma_i^2}{2}\right)$$

Therefore, the MGF of $Z$ is:

$$M_Z(t) = \prod_{i=1}^{k} \exp\left(tp_i\mu_i + \frac{t^2p_i^2\sigma_i^2}{2}\right)$$

Simplifying:

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

**5. What can you conclude? Do $X$ and $Z$ have the same properties?**

Yes, $Z$ inherits the properties of a Gaussian distribution from the individual random variables $X_i$. Specifically:

- $Z$ is normally distributed.
- The mean of $Z$ is a weighted sum of the means of the $X_i$'s.
- The variance of $Z$ is a weighted sum of the variances of the $X_i$'s, with the weights squared.

Thus, $Z$ has the same properties as a normal random variable, but with parameters that depend on the weights $p_i$.

**6. What distribution does $Z$ seem to follow?**

Since $Z$ is a linear combination of independent Gaussian random variables, $Z$ follows a normal (Gaussian) distribution:

$$Z \sim \mathcal{N}\left(\sum_{i=1}^{k} p_i\mu_i, \sum_{i=1}^{k} p_i^2\sigma_i^2\right)$$

## 5.4 Task D

**Theorem 8 (Finite Discrete Case)**

For a finite discrete random variable $X$, its Moment Generating Function (MGF) uniquely determines its Probability Distribution Function (PDF), and vice versa.

**Proof:**

Let $X$ be a finite discrete random variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with probabilities $P(X = x_i) = p_i$, where $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ for all $i$.

The Moment Generating Function (MGF) of $X$ is defined as:

$$\phi_X(t) = \mathbb{E}\left[e^{tX}\right] = \sum_{i=1}^{n} p_i e^{tx_i}$$

Since the values $\{x_1, x_2, \ldots, x_n\}$ are distinct, the function $\phi_X(t)$ is a linear combination of the exponentials $e^{tx_i}$, with the probabilities $p_i$ as coefficients.

To recover the probabilities $p_i$, observe that the system of equations formed by evaluating $\phi_X(t)$ at different values of $t$ is a linear system in the unknowns $p_1, p_2, \ldots, p_n$. More formally, consider the system:

$$\begin{aligned}
\phi_X(t_1) &= p_1 e^{t_1 x_1} + p_2 e^{t_1 x_2} + \cdots + p_n e^{t_1 x_n} \\
\phi_X(t_2) &= p_1 e^{t_2 x_1} + p_2 e^{t_2 x_2} + \cdots + p_n e^{t_2 x_n} \\
&\vdots \\
\phi_X(t_n) &= p_1 e^{t_n x_1} + p_2 e^{t_n x_2} + \cdots + p_n e^{t_n x_n}
\end{aligned}$$

This system is uniquely solvable for $p_1, p_2, \ldots, p_n$, provided that the values $\{x_1, x_2, \ldots, x_n\}$ are distinct, because the matrix formed by $e^{t_i x_j}$ is a Vandermonde matrix, which is invertible when the $x_i$ values are distinct.

Thus, knowing the MGF $\phi_X(t)$, we can uniquely determine the probabilities $p_1, p_2, \ldots, p_n$, which uniquely determine the PDF of $X$.

Conversely, if the PDF of $X$ is known, then the MGF is simply the expected value of $e^{tX}$, which is uniquely determined by the probabilities $p_i$.

**Conclusion about $X$ and $Z$:**

- If $X$ is defined as a mixture of random variables $X_1, X_2, \ldots, X_K$ with weights $p_i$, then the MGF of $X$ is:

  $$\phi_X(t) = \sum_{i=1}^{K} p_i \phi_{X_i}(t)$$

  where $\phi_{X_i}(t)$ is the MGF of $X_i$.

- For $Z$, the MGF is:

  $$\phi_Z(t) = \mathbb{E}[e^{tZ}] = \mathbb{E}\left[e^{t\left(\sum_{i=1}^{K} p_i X_i\right)}\right] = \prod_{i=1}^{K} \mathbb{E}\left[e^{tp_i X_i}\right] = \prod_{i=1}^{K} \phi_{X_i}(tp_i)$$

- If the MGFs of $X$ and $Z$ are the same, then $X$ and $Z$ have the same distribution. This is because the MGF uniquely determines the distribution of a random variable (assuming the MGF exists and is finite over some interval).

- Since $X$ is a mixture of $X_i$'s and $Z$ is a weighted sum of the same $X_i$'s, if their MGFs are equal, it implies that $X$ and $Z$ have the same distribution.

**Logical Explanation:**

The MGF of a random variable encodes all the moments of that variable. Since $X$ and $Z$ are defined in terms of the same underlying random variables $X_i$, if their MGFs are the same, it implies that $X$ and $Z$ have the same distribution.