Assignment 2: CS215

Satyam Sinoliya, 23B0958
Vaibhav Singh, 23B1068
Shaik Awez Mehtab, 23B1080

Solution 1

Task A

When $X \sim \text{Ber}(p)$, PGF of X is

$$\begin{aligned} G_{\text{Ber}}(z) &= \mathbb{E}(z^X) && (1.1) \\ &= \sum_{n=0}^{\infty} P[X = n] z^n && (1.2) \end{aligned}$$

Since $P[X = 0] = (1 - p)$, $P[X = 1] = p$, $P[X = n] = 0$ when $n > 1$,

$$\begin{aligned} G_{\text{Ber}}(z) &= P[X = 0]z^0 + P[X = 1]z^1 && (1.3) \\ &= (1 - p) + pz && (1.4) \end{aligned}$$

Task B

When $X \sim \text{Bin}(n, p)$, PMF of X is

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \text{ for } k \leq n. \quad (1.5)$$

and $P[X = k] = 0$ for $k > n$.

$$\begin{aligned} G_{\text{Bin}}(z) &= \sum_{k=0}^{\infty} P[X = k] z^k && (1.6) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} z^k && (1.7) \\ &= \sum_{k=0}^{n} \binom{n}{k} (pz)^k (1 - p)^{n-k} && (1.8) \\ &= (1 - p + pz)^n && (1.9) \end{aligned}$$

By equation 1.4,

$$G_{\text{Bin}}(z) = (1 - p + pz)^n = (G_{\text{Ber}}(z))^n \quad (1.10)$$

Task C

By the definition of PGF,

$$G(z) = \sum_{n=0}^{\infty} P[X_1 = n]z^n \tag{1.11}$$

Let $(G(z))^k = \sum_{n=0}^{\infty} a_n z^n$. Now, $G_\Sigma(z)$ is

$$G_\Sigma(z) = \sum_{n=0}^{\infty} P[X = n]z^n \tag{1.12}$$

$$= \sum_{n=0}^{\infty} P[X_1 + \cdots + X_k = n]z^n \tag{1.13}$$

$$= \sum_{n=0}^{\infty} \sum P[X_1 = i_1, \ldots, X_k = i_k]z^n \tag{1.14}$$

where $i_1 + \cdots + i_k = n$.

$$= \sum_{n=0}^{\infty} \sum P(i_1) \ldots P(i_n)z^n \tag{1.15}$$

Now, $G(z) = \sum_{n=0}^{\infty} P(n)z^n$. And since a_n is coefficient of z^n in $(G(z))^k = (\sum_{n=0}^{\infty} P(n)z^n)^k$.

$$a_n = \sum P(i_1)P(i_2) \ldots P(i_k) \text{ where } i_1 + \cdots + i_k = n \tag{1.16}$$

By equation 1.15

$$G_\Sigma(z) = \sum_{n=0}^{\infty} a_n z^n = (G(z))^k \tag{1.17}$$

Hence Proved.

Task D

When $X \sim \text{Geo}(p)$, PMF of X,

$$P[X = k] = (1-p)^{k-1}p \tag{1.18}$$

for $k > 0$. $P[X = 0] = 0$. Now, PGF of X,

$$
\begin{aligned}
G_{\text{Geo}}(z) &= \sum_{k=0}^{\infty} P[X = k] z^k && (1.19)\\
&= \sum_{k=1}^{\infty} P[X = k] z^k && (1.20)\\
&= \sum_{k=1}^{\infty} p(1-p)^{k-1} z^k && (1.21)\\
&= \sum_{k=1}^{\infty} pz(z - zp)^{k-1} && (1.22)\\
&= pz \sum_{k=0}^{\infty} (z - zp)^k && (1.23)\\
&= \frac{pz}{1 - z + pz} && (1.24)
\end{aligned}
$$

Task E

By equation 1.9,

$$G_{\text{Bin}}(z) = (1 - p + pz)^n = G_X^{(n,p)}(z). \tag{1.25}$$

For $Y \sim \text{NegBin}(n, p)$, Y represents the number of independent coin throws required to get n heads of a coin. Let X_i represents the number of throws of coin required after getting $(i-1)^{\text{th}}$ head to get the i^{th} head. Since all of the coin throws are independent, the outcome of a given throw doesn't depend on the previous coins' output. Thus, X_i is just the number of throws to get a head when a coin in thrown, where each $X_i \sim \text{Geo}(p)$ since each coin is same with probability of getting head as p.
Y can be written as $Y = X_1 + X_2 + \cdots + X_k$. Using equations 1.17 and 1.24,

$$
\begin{aligned}
G_Y^{(n,p)}(z) &= (G_{\text{Geo}}(z))^n && (1.26)\\
&= \left(\frac{pz}{1 - z + pz}\right)^n && (1.27)\\
& && (1.28)
\end{aligned}
$$

$$
\begin{aligned}
G_X^{(n,p^{-1})}(z^{-1}) &= \left(1 - \frac{1}{p} + \frac{1}{pz}\right)^n && (1.29)\\
&= \left(\frac{1 - z + pz}{pz}\right)^n && (1.30)\\
\left(G_X^{(n,p^{-1})}(z^{-1})\right)^{-1} &= \left(\frac{pz}{1 - z + pz}\right)^n && (1.31)\\
&= G_Y^{(n,p)}(z) && (1.32)
\end{aligned}
$$

Hence Proved.

Task F

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

$$P[Y = k] = \binom{k-1}{n-1} p^n (1-p)^{k-n} \text{ for } k \geq n \tag{1.33}$$

Otherwise, $P[Y = k] = 0$. PGF of Y is

$$G_Y^{(n,p)}(z) = \sum_{k=0}^{\infty} P[Y = k] z^k \tag{1.34}$$

$$= \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} z^k \tag{1.35}$$

$$= \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} p^n (1-p)^k z^{n+k} \tag{1.36}$$

$$= (pz)^n \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} (z - pz)^k \tag{1.37}$$

Using equation 1.28,

$$\left(\frac{pz}{1 - z + pz}\right)^n = (pz)^n \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} (z - pz)^k \tag{1.38}$$

$$(1 + pz - z)^{-n} = \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} (z - pz)^k \tag{1.39}$$

Since z, p are arbitrary, let $pz - z = x$.

$$(1 + x)^{-n} = \sum_{r=0}^{\infty} (-1)^r \binom{r+n-1}{n-1} x^r = \sum_{r=0}^{\infty} (-1)^r \binom{n+r-1}{r} x^r \tag{1.40}$$

Now,

$$(-1)^r \binom{n+r-1}{r} = (-1)^r \frac{(n+r-1)(n+r-2)\cdots n}{r!} \tag{1.41}$$

$$= \frac{(-n)(-n-1)\cdots(-n-r+1)}{r!} \tag{1.42}$$

$$= \binom{-n}{r} \tag{1.43}$$

Thus,

$$(1 + x)^{-n} = \sum_{r=0}^{\infty} (-1)^r \binom{n-r+1}{r} x^r = \sum_{r=0}^{\infty} \binom{-n}{r} x^r \tag{1.44}$$

Hence proved.

Task G

To prove: *Given PGF of a random variable X is $G(z)$, expectation of X i.e $\mathbb{E}(x) = G'(1)$*

Proof:

$$G(z) = \mathbb{E}(z^X) = \sum_{k=0}^{\infty} P[X = k] z^k \tag{1.45}$$

$$G'(z) = \sum_{k=0}^{\infty} k P[X = k] z^{k-1} \tag{1.46}$$

$$G'(1) = \sum_{k=0}^{\infty} k P[X = k] \tag{1.47}$$

$$= \mathbb{E}[X] \tag{1.48}$$

Hence Proved. Now, Let's derive means of Bernoulli, Binomial, Geometric and Negative Binomial distributions using this:

1. **Bernoulli Distribution:** Let $X \sim \text{Ber}(p)$,

$$G_{\text{Ber}}(z) = (1-p) + pz \tag{1.49}$$

$$G'_{\text{Ber}}(z) = p \tag{1.50}$$

$$G'_{\text{Ber}}(1) = p = \mathbb{E}[X] \tag{1.51}$$

   Thus, $\mathbb{E}[X] = p$.

2. **Binomial Distribution:** Let $X \sim \text{Bin}(n, p)$,

$$G_{\text{Bin}}(z) = (1 - p + pz)^n \tag{1.52}$$

$$G'_{\text{Bin}}(z) = np(1 - p + pz)^{n-1} \tag{1.53}$$

$$G'_{\text{Bin}}(1) = np = \mathbb{E}[X] \tag{1.54}$$

   Thus, $\mathbb{E}[X] = np$.

3. **Geometric Distribution:** Let $X \sim \text{Geo}(p)$,

$$G_{\text{Geo}}(z) = \frac{pz}{1 - z + pz} \tag{1.55}$$

$$G'_{\text{Geo}}(z) = \frac{p(1 - z + pz) - pz(p-1)}{(1 - z + pz)^2} \tag{1.56}$$

$$= \frac{p}{(1 - z + pz)^2} \tag{1.57}$$

$$G'_{\text{Geo}}(1) = \frac{p}{p^2} = \frac{1}{p} = \mathbb{E}[X] \tag{1.58}$$

   Thus, $\mathbb{E}[X] = \frac{1}{p}$.

4. **Negative Binomial Distribution:** Let $X \sim \text{NegBin}(n, p)$,

$$G_{\text{NegBin}}(z) = \left(\frac{pz}{1 - z + pz}\right)^n \tag{1.59}$$

$$G'_{\text{NegBin}}(z) = n\left(\frac{pz}{1 - z + pz}\right)^{n-1} \left(\frac{p}{(1 - z + pz)^2}\right) \tag{1.60}$$

$$G'_{\text{NegBin}}(1) = \frac{n}{p} = \mathbb{E}[X] \tag{1.61}$$

   Thus, $\mathbb{E}[X] = \frac{n}{p}$.

Solution 2

Task A

To prove:
Let X be a continuous real-valued random variable with CDF : $\mathbb{R} \to [0,1]$. *Assume that F_X is invertible. Then the random variable $Y := F_X(X) \in [0,1]$ is uniformly distributed in* $[0,1]$
Proof:
F_X by definition can also be written as

$$F_X(x) = P(X \leq x) \tag{2.1}$$

Define a new random variable Y,

$$Y = F_X(X) \tag{2.2}$$

Y is the result of applying CDF F_X to the random variable X. To prove the theorem, assume $y \in [0,1]$. So, the probablity that $Y \leq y$ is:

$$P(Y \leq y) = P(F_X(X) \leq y) \tag{2.3}$$

It is assumed that $F_X(x)$ is invertible, so,

$$P(Y \leq y) = P(X \leq F_X^{-1}(y)) \tag{2.4}$$

which is basically, probablity that X is less that or equal to $F_X^{-1}(y)$. This can be written in the CDF form, which is $F_X(F_X^{-1}(y))$. So,

$$P(Y \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y \tag{2.5}$$

So,

$$P(Y \leq y) = y \tag{2.6}$$

where $y \in [0,1]$, which is the CDF of uniform distribution in $[0,1]$. So, Y is a uniform distribution in $[0,1]$ regardless of X.

Task B

According to the theorem proved above, CDF of any random variable X mapped with itself gives a uniform random variable Y in $[0,1]$. So, let $Y \sim \text{Uniform}(0,1)$. Then for any random variable X,

$$F_X(X) = Y \tag{2.7}$$
$$X = F_X^{-1}(Y) \tag{2.8}$$

Algorithm $\mathcal{A}$:

1. Input: A sample y from the uniform distribution on $[0, 1]$.
2. Transformation:
   - Apply the inverse CDF to y to compute a sample u.
   - Define $\mathcal{A}(u) = u = F_X^{-1}(y)$
3. Output: The random variable $U = F_X^{-1}(Y)$

This gives us the correct required random variables as, CDF of U is $F_U(u)$,

$$P(U \leq u) = P(F_X^{-1}(Y) \leq u) \tag{2.9}$$
$$F_U(u) = P(Y \leq F_X(u)) \tag{2.10}$$
$$\tag{2.11}$$

Since, Y is a uniform random variable between 0 and 1,

$$F_U(u) = F_Y(F_X(u)) \tag{2.12}$$
$$F_U(u) = F_X(u) \tag{2.13}$$
$$\tag{2.14}$$

Since, for any uniform random variable P between 0 and 1,

$$F_P(x) = x \tag{2.15}$$

This proves that, U and X have the same CDF.

Task C

The code can be found in `2c.py`. In this script, we generate random samples from a Gaussian distribution using the inverse transform sampling method. The function `sample(loc, scale)` begins by creating a sample, `x`, which consists of uniformly distributed random variables between 0 and 1, with a sample size of `N` (where `N` = 10^5 or 100,000).
These uniform samples, `x`, are then transformed into samples from a Gaussian (normal) distribution by passing them through the inverse cumulative distribution function (CDF) of the normal distribution, implemented via `norm.ppf`. The arguments `loc` and `scale` represent the mean and standard deviation of the Gaussian distribution, respectively. This transformation effectively maps the uniform samples to samples drawn from the specified Gaussian distribution, as demonstrated theoretically in previous sections.
We then define the list `params`, which contains tuples of means and variances for the Gaussian distributions we wish to sample from and plot. For each set of parameters, the `sample()` function is called, and the resulting samples are stored in the `samples` list.
Finally, we loop over the pairs of parameters (`params`) and their corresponding samples (`samples`) to create a histogram for each distribution. The histograms are plotted with 500 bins, and the density is normalized. A legend is included to indicate the mean (μ) and standard deviation (σ) for each distribution.
Below is the resulting histogram:

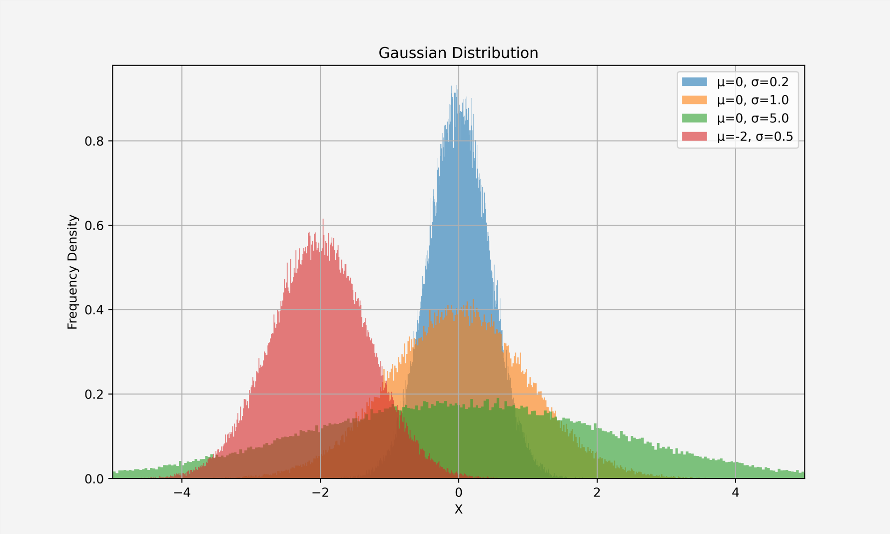

Figure 2.1: Histograms of Gaussian distributions for different means and variances

Task D

The code for simulating the Galton board is found in `2d.py`. We implemented a function called `galton_board`, which takes two parameters: `h`, representing the depth of the Galton board (number of pegs the ball hits), and `N`, the number of balls dropped.
Inside the function, the variable steps is a matrix with `N` rows and `h` columns. Each row corresponds to a ball, and each column represents a decision the ball makes at a peg. At each step, the ball either moves left, represented by -1, or right, represented by 1. The final position of each ball is determined by summing all its steps, which is stored in the 1-D array positions. This array records the final pocket into which each ball falls.
For the simulation, 1000 balls are dropped for different values of `h` (the depth of the board). The resulting distributions for each value of h are visualized as histograms, with the x-axis representing the final positions (pockets) and the y-axis showing the probability of balls falling into each pocket. The results for different values of h are shown below.

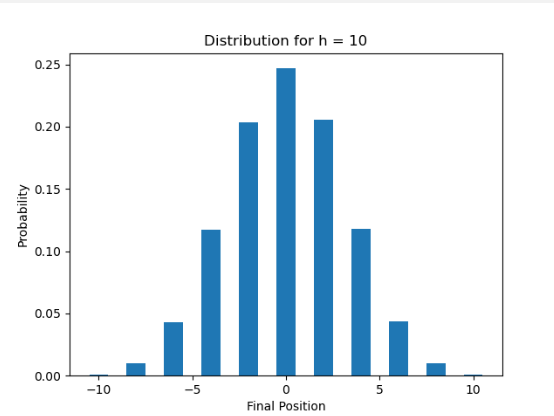

Figure 2.2: Height = 10

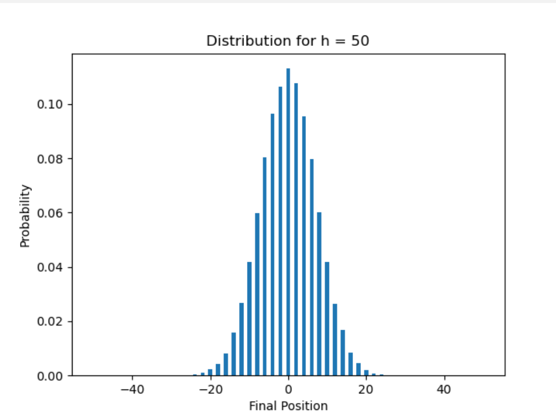

Figure 2.3: Height = 50

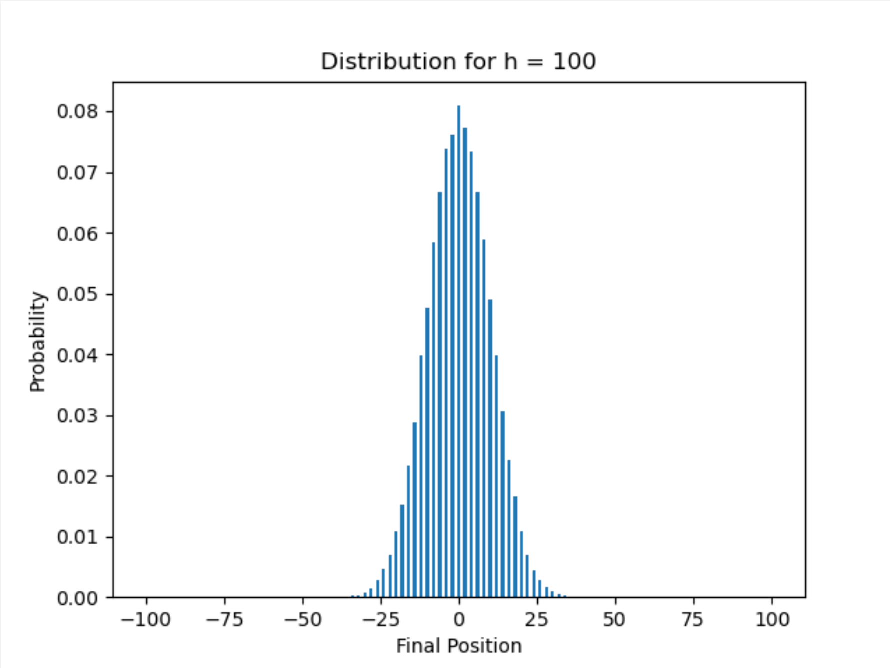

Figure 2.4: Height = 100

The shape of the tops of the histogram closely resembles a Gaussian distribution. As the height (h) of the Galton board increases, the distribution becomes smoother and more bell-shaped, indicating that most balls tend to fall around the center pockets.

Task E(B)

- $P_h[X = 2i]$ calculation

We have a random variable X which can take values from $\{-h, -h+2, \ldots, h-2, h\}$. Each ball makes h random binary decisions(left or right) as it descends. If we let Y be the number of times the ball moves right, the final position of the ball will be given by,

$$X = -h + 2Y \tag{2.16}$$

where Y is a **binomial variable** because in simple terms it is the summation of h bernoulli decisions each with probability $\frac{1}{2}$.

$$Y \sim \text{Bin}(h, \frac{1}{2}) \tag{2.17}$$

For a particular pocket $X = 2i$, the corresponding value of Y is:

$$Y = \frac{h+2i}{2} \tag{2.18}$$

Thus, the probability that the ball lands in the pocket $X = 2i$ is the probability that $Y = \frac{h+2i}{2}$. Using Binmial distribution, this is:

$$P_h[X = 2i] = P_h\left[Y = \frac{h+2i}{2}\right] = \binom{h}{\frac{h+2i}{2}}\left(\frac{1}{2}\right)^h \tag{2.19}$$

-$P_h[X = i]$ approximates to normal distribution

Now, we need to show $P_h[X = i]$ approximates to normal distribution for very large h.
Using **Stirling's approximation** for large n, which states that:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{2}\right)^n \tag{2.20}$$

P_h in terms of k can be written as

$$P_h[X = i] = \binom{2k}{k + \frac{i}{2}}\left(\frac{1}{2}\right)^{2k} \tag{2.21}$$

Where $h = 2k$. Let $j = \frac{i}{2}$. So, $j \in \{-k, -k+1, \dots, k-1, k\}$. So now,

$$P_h[X = 2j] = \binom{2k}{k + j}\left(\frac{1}{2}\right)^{2k} \tag{2.22}$$

Using Stirling's approximations,

$$\begin{aligned}
P_h[X = i] &= \frac{(2k)!}{(k+j)!(k-j)!}\left(\frac{1}{2}\right)^{2k} && (2.23)\\
&\approx \frac{\sqrt{2\pi(2k)}\left(\frac{2k}{e}\right)^{2k}}{\sqrt{2\pi(k+j)}\left(\frac{k+j}{e}\right)^{k+j}\sqrt{2\pi(k-j)}\left(\frac{k-j}{e}\right)^{k-j}}\left(\frac{1}{2}\right)^{2k} && (2.24)\\
&= \frac{\sqrt{4\pi k}}{\sqrt{2\pi(k+j)}\sqrt{2\pi(k-j)}}\left(\frac{2k}{e}\right)^{2k}\left(\frac{e}{k+j}\right)^{k+j}\left(\frac{e}{k-j}\right)^{k-j}\left(\frac{1}{2}\right)^{2k} && (2.25)\\
&= \frac{\sqrt{k}}{\sqrt{\pi(k+j)(k-j)}}\left(\frac{2^{2k}\cdot e^{k+j+k-j}}{e^{2k}\cdot 2^{2k}}\right)\left(\frac{k^{2k}}{(k+j)^{k+j}\cdot(k-j)^{k-j}}\right) && (2.26)\\
&= \frac{\sqrt{k}}{\sqrt{\pi(k^2-j^2)}}\left(\frac{(k-j)^j}{(k+j)^j}\right)\left(\frac{k^2}{(k+j)(k-j)}\right)^k && (2.27)\\
&= \frac{\sqrt{k}}{\sqrt{\pi(k^2-j^2)}}\left(\frac{k^j\left(1-\frac{j}{k}\right)^j}{k^j\left(1+\frac{j}{k}\right)^j}\right)\left(\frac{k^2}{k^2-j^2}\right)^k && (2.28)
\end{aligned}$$

As $i \ll \sqrt{h}$, so $j \ll \sqrt{h}$ and as $h = 2k$, so $j \ll \sqrt{k}$ and $j \ll k$.
I will use the following well known approximation in my proof next:

$$[(1+x)^n \approx 1 + nx \tag{2.29}$$

where x is close to 0.

$$\begin{aligned}
P_h[X = i] &= \frac{\sqrt{k}}{\sqrt{\pi(k^2-j^2)}}\left(1-\frac{j^2}{k}\right)\left(1-\frac{j^2}{k}\right)\left(\frac{k^2-j^2+j^2}{k^2-j^2}\right)^k && (2.30)\\
&= \frac{\sqrt{k}}{\sqrt{\pi(k^2-j^2)}}\left(1-\frac{2j^2}{k}\right)\left(1+\frac{j^2}{k^2-j^2}\right)^k && (2.31)\\
& && (2.32)
\end{aligned}$$

We are going to use few more well known approximation. We know that:

$$\lim_{x\to\infty y\to 0}(1+y)^x = e^{xy} \tag{2.33}$$

$$\lim_{y \to 0} e^y \approx 1 + y \tag{2.34}$$

The second approximation comes from taylor series. Continuing the proof:

$$P_h[X = i] = \frac{\sqrt{k}}{\sqrt{\pi(k^2 - j^2)}} e^{\frac{-2j^2}{k}} e^{k \cdot \frac{j^2}{k^2 - j^2}} \tag{2.35}$$

$$\tag{2.36}$$

Since $j \ll k$, $j^2 \ll k^2$, we can write $k^2 - j^2 = k^2$

$$P_h[X = i] = \frac{\sqrt{k}}{\sqrt{\pi k^2}} e^{\frac{-2j^2}{k}} e^{k \cdot \frac{j^2}{k^2}} \tag{2.37}$$

$$= \frac{1}{\sqrt{\pi k}} e^{\frac{-2j^2}{k} + \frac{j^2}{k}} \tag{2.38}$$

$$= \frac{1}{\sqrt{\pi k}} e^{\frac{-j^2}{k}} \tag{2.39}$$

Substituting back the value of $j = \frac{i}{2}$

$$P_h[X = i] = \frac{1}{\sqrt{\pi k}} e^{\frac{-i^2}{4k}} \tag{2.40}$$

Solution 3

Task A

For compputing the first two moments, we have used the **np.mean()** function. From that, $\hat{\mu}_1$ is 6.496145618324817 and $\hat{\mu}_2$ is 46.554361807879815.

Task D

Given, PDF of Gamma-distribution $\text{Gamma}(k, \theta)$ is $f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$. First moment of it is

$$
\begin{aligned}
\mu_1^{\text{Gamma}} &= \mathbb{E}[X] && (3.1)\\
&= \int_{-\infty}^{\infty} x \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} dx && (3.2)\\
&= \frac{1}{\theta^k \Gamma(k)} \int_{-\infty}^{\infty} x^k e^{-\frac{x}{\theta}} dx && (3.3)
\end{aligned}
$$

Let $u = \frac{x}{\theta}$, then $\theta du = dx$

$$
\begin{aligned}
&= \frac{\theta}{\Gamma(k)} \int_{-\infty}^{\infty} u^k e^{-u} du && (3.4)\\
&= \frac{\theta \Gamma(k+1)}{\Gamma(k)} && (3.5)
\end{aligned}
$$

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

$$
= k\theta \tag{3.6}
$$

Second moment of it is

$$
\begin{aligned}
\mu_2^{\text{Gamma}} &= \mathbb{E}[X^2] && (3.7)\\
&= \int_{-\infty}^{\infty} x^2 \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} dx && (3.8)\\
&= \frac{1}{\theta^k \Gamma(k)} \int_{-\infty}^{\infty} x^{k+1} e^{-\frac{x}{\theta}} dx && (3.9)
\end{aligned}
$$

Let $u = \frac{x}{\theta}$, then $\theta du = dx$

$$
\begin{aligned}
&= \frac{\theta^2}{\Gamma(k)} \int_{-\infty}^{\infty} u^{k+1} e^{-u} du && (3.10)\\
&= \frac{\theta^2 \Gamma(k+2)}{\Gamma(k)} && (3.11)\\
&= (k+1)k\theta^2 && (3.12)
\end{aligned}
$$

Thus, $\mu_1^{\text{Gamma}} = k\theta$, $\mu_2^{\text{Gamma}} = (k+1)k\theta^2$.
Estimate the best gamma distribution approximation to the given data by equating first and second moments. For $X \sim \text{Gamma}(n, p)$, density function is

$$f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \tag{3.13}$$

$\mu_1 = k\theta$ and $\mu_2 = k(k+1)\theta^2$.

Task E

- Log likelihood for binomimal distribution ≈ -2.157
- Log likelihood for gamma distribution $= -$`inf`

Thus binomial distribution is a better fit for the data.

Task F

- Log likelihood for two-component Gaussian mixture is -2.183

Thus, binomial distribution is slightly better fit than the given two-component gaussian mixture (whose variance is assumed to be 1).

Solution 4

Task A

We can write
$$\mathbb{E}[X] = P\{X < a\} \cdot \mathbb{E}[X|X < a] + P\{X \geq a\} \cdot \mathbb{E}[X|X \geq a]. \tag{4.1}$$
where $\mathbb{E}[X|X < a]$ is greater than zero and as random variable X is non negative and $\mathbb{E}[X|X \geq a]$ is greater than or equal to a since conditional expression only accounts for values greater than equal to a. Therefore, we can say

$$\mathbb{E}[X] \geq P\{X \geq a\} \cdot \mathbb{E}[X|X \geq a] \tag{4.2}$$

which leads to

$$\mathbb{E}[X] \geq a \cdot P\{X \geq a\} \tag{4.3}$$
$$P\{X \geq a\} \leq \frac{\mathbb{E}[X]}{a} \tag{4.4}$$

Formal Proof
For non negative continuous random variable with density f, we can write the expected value as

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f(x) dx && (4.5)\\
&= \int_0^a x f(x) dx + \int_a^\infty x f(x) dx && (4.6)\\
&\geq \int_a^\infty x f(x) dx && (4.7)\\
&\geq \int_a^\infty a f(x) dx && (4.8)\\
&= a \int_a^\infty f(x) dx && (4.9)\\
&= aP\{X \geq a\} && (4.10)\\
\implies \mathbb{E}[X] &\geq aP\{X \geq a\} && (4.11)\\
\implies P\{X \geq a\} &\leq \frac{\mathbb{E}[X]}{a} && (4.12)
\end{aligned}$$

Hence we arrive at our required relation.

Task B

To prove:

$$P\left[X-\mu \geq \tau\right] \leq \frac{\sigma^2}{\sigma^2+\tau^2} \tag{4.13}$$

for every $\tau > 0$ where X is a random variable with mean μ and variance σ^2.

Let $Y = X - \mu$. So,we have

$$P[Y \geq \tau] = P[Y+\mu \geq \tau+\mu] \leq P[(Y+\mu)^2 \geq (\tau+\mu)^2] \tag{4.14}$$

$$\implies P[Y \geq \tau] \leq \frac{\mathbb{E}[(Y+\mu)^2]}{(\tau+\mu)^2} \tag{4.15}$$

$$\implies P[Y \geq \tau] \leq \frac{\sigma^2+\mu^2}{(\tau+\mu)^2} \text{ since } E[(Y+\mu)^2] \text{ is } E[X^2] \tag{4.16}$$

Since this probability holds for all μ and τ, we can put $\mu = \frac{\sigma^2}{\tau}$ to minimise the function by differentiating and we have

$$P[Y \geq \tau] \leq \frac{\sigma^2(\tau^2+\sigma^2)}{(\sigma^2+\tau^2)^2} \tag{4.17}$$

$$\implies P[Y \geq \tau] \leq \frac{\sigma^2}{\sigma^2+\tau^2} \tag{4.18}$$

Hence proved.

Task C

To show:

$$P[X \geq x] \leq e^{-tx} M_x(t) \ \forall t > 0 \tag{4.19}$$

$$P[X \leq x] \leq e^{-tx} M_x(t) \ \forall t < 0 \tag{4.20}$$

From markov inequality, we know that for a non negative random variable X and value $x > 0$, we can write

$$P[X \geq x] \leq \frac{E[X]}{x} \tag{4.21}$$

Using the fact that e^{tx} is always positive, we can write $P(e^{tX} \geq e^{tx})$ same as $P(X >= x)$, we get:

$$P(X \geq x) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} \tag{4.22}$$

By definition, we know $\mathbb{E}[e^{tX}] = M_X(t)$, therefore

$$P(X \geq x) \leq \frac{M_X(t)}{e^{tx}} = e^{-tx} M_X(t) \tag{4.23}$$

For second part, since we need to prove for $t < 0$ and we want to bound the probability $P(X \leq x)$, we can rewrite it as

$$P(X \leq x) = P(-X \geq -x) \tag{4.24}$$

Using markov's inequality on $e^{t(-X)} = e^{-tX}$ (Since $t < 0$)

$$P(e^{-tX} \geq e^{-tx}) \leq \frac{\mathbb{E}[e^{-tX}]}{e^{-tx}} \tag{4.25}$$

Now substituting $\mathbb{E}[e^{-tX}] = M_X(-t)$ and $-t$ with t since $t < 0$ we get

$$P(X \leq x) \leq e^{-tx} M_X(t) \tag{4.26}$$

Hence both proofs are finished under given conditions.

Task D

We are given $Y = \sum_{i=1}^{n} X_i$, where each X_i is an independent Bernoulli random variable with $\mathbb{E}[X_i] = p_i$

1. Expectation of Y

The new random variable Y is given by:

$$Y = \sum_{i=1}^{n} X_i \tag{4.27}$$

Using the liniearity of expectation, the expectation of Y is:

$$\mathbb{E}[X_i] = \mathbb{E}\left[\sum_{i=1}^{n}\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p_i \tag{4.28}$$

Let $\mu = \sum_{i=1}^{n} p_i$, So $\mathbb{E}[Y] = \mu$

2. Proof

From previous subparts, we know

$$P[X \geq x] \leq e^{-tx} M_X(t) \text{ for all } t > 0 \tag{4.29}$$

where $M_X(t)$ is the MGF.

$$M_Y(t) = \mathbb{E}[e^{tY}] \tag{4.30}$$

$$= \mathbb{E}[e^{t\sum_{i=1}^{n} X_i}] \tag{4.31}$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \tag{4.32}$$

Using the independence of each X_i, we can write the expecatation of product as product of expectation,

$$M_Y(t) = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] \tag{4.33}$$

$$= \prod_{i=1}^{n} M_{X_i} \tag{4.34}$$

MGF of a Bernoulli random variable is

$$M_{X_i} = p_i e^t + (1 - p_i) \tag{4.35}$$

So,

$$M_Y(t) = \prod_{i=1}^{n}(p_i e^t + (1-p_i)) \tag{4.36}$$

Putting it back into the inequality,

$$P[X \geq (1+\delta)\mu] \leq \frac{\prod_{i=1}^{n}(p_i e^t + (1-p_i))}{e^{t(1+\delta)\mu}} \tag{4.37}$$

We know that, $1 + x \leq e^x$. Substituting x with $p_i(e^t - 1)$, we get:

$$1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)} \tag{4.38}$$

Rearranging terms,

$$p_i e^t + (1-p_i) \leq e^{p_i(e^t-1)} \tag{4.39}$$

So,

$$P[X \geq (1+\delta)\mu] \leq \frac{\prod_{i=1}^{n} e^{p_i(e^t-1)}}{e^{t(1+\delta)\mu}} \tag{4.40}$$

$$P[X \geq (1+\delta)\mu] \leq \frac{e^{\sum_{i=1}^{n} p_i(e^t-1)}}{e^{t(1+\delta)\mu}} \tag{4.41}$$

$$P[X \geq (1+\delta)\mu] \leq \frac{e^{(e^t-1)\sum_{i=1}^{n} p_i}}{e^{(1+\delta)t\mu}} \tag{4.42}$$

$$P[X \geq (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{(1+\delta)t\mu}} \tag{4.43}$$

This completes the proof

3. Improving bound

To improve bound, we have to minimise the right hand term in 4.43. As it is a continuous functions, we can differentiate it to get a minima. Let the RHS be $f(t)$

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\mathrm{d}\frac{e^{(e^t-1)\mu}}{e^{(1+\delta)t\mu}}}{\mathrm{d}t} \tag{4.44}$$

For minima, $\dfrac{\mathrm{d}f}{\mathrm{d}t}$ should be 0.

$$\frac{\mathrm{d}\frac{e^{(e^t-1)\mu}}{e^{(1+\delta)t\mu}}}{\mathrm{d}t} = 0 \tag{4.45}$$

$$\frac{(e^{(1+\delta)t\mu})\frac{\mathrm{d}e^{(e^t-1)\mu}}{\mathrm{d}t} - e^{(e^t-1)\mu}\frac{\mathrm{d}e^{(1+\delta)t\mu}}{\mathrm{d}t}}{(e^{(1+\delta)t\mu})^2} = 0 \tag{4.46}$$

$$(e^{(1+\delta)t\mu})(e^{(e^t-1)\mu})(\mu e^t) - (e^{(e^t-1)\mu})((e^{(1+\delta)t\mu}))(1+\delta)\mu = 0 \tag{4.47}$$

$$e^t - (1+\delta) = 0 \tag{4.48}$$

$$t = \ln(1+\delta) \tag{4.49}$$

The improved bound will be,

$$f(\ln(1+\delta)) = \frac{e^{(e^{\ln(1+\delta)}-1)\mu}}{e^{(1+\delta)\ln(1+\delta)\mu}} \tag{4.50}$$

On further calculation, we get,

$$f(\ln(1+\delta)) = \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \tag{4.51}$$

Finally,

$$P[X \geq (1+\delta)\mu] \leq \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \tag{4.52}$$

Task E(B)

We need to prove Weak Law of Large Numbers using Chernoff bound, that is, for any $\epsilon > 0$, the probability that sample average deviates from the true mean μ by more than ϵ tends to zero as $n \to \infty$, that is

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| > \epsilon) = 0 \tag{4.53}$$

Applying the Chernoff bound, we need to bound A_n.

$$P(|A_n - \mu| > \epsilon) = P(A_n > \mu + \epsilon) + P(A_n < \mu - \epsilon) \tag{4.54}$$

Considering the probability $P(A_n > \mu + \epsilon)$. The moment generating function for the i.i.d. random variable X_i as $M_X(t) = \mathbb{E}[e^{tX_i}]$. Using Chernoff's bound on this probability,keeping $\epsilon = \delta\mu$ we get

$$\lim_{n\to\infty} P(A_n > \mu + \epsilon) < \inf_{t>0}(e^{n[(e^t-1)\mu-(1+\delta)t\mu]}) \tag{4.55}$$

Putting the value of t from the earlier task we get,

$$\lim_{n\to\infty} P(A_n > \mu + \epsilon) < \lim_{n\to\infty} e^{nu[\delta-(1+\delta)ln(1+\delta)]} \tag{4.56}$$

For $t > 0$, δ must be greater than 1 and hence the exponential power becomes negative and thus as n grow larger the term tends to zero.
For the other part, i.e., $P(A_n < \mu - \epsilon)$ using the chernoff bound on the left tail and putting $\epsilon = \delta\mu$,

$$\lim_{n\to\infty} P(A_n < \mu - \epsilon) < \inf_{t<0}(e^{n[(e^t-1)\mu-(1-\delta)t\mu]}) \tag{4.57}$$

Since minimising while $t < 0$ we will be using $-1 < \delta < 0$ and putting value of t, we get

$$\lim_{n\to\infty} P(A_n < \mu - \epsilon) < (e^{nu(\delta-(1-\delta)ln(1+\delta))}) \tag{4.58}$$

Again, we can see that the exponential power becomes negative and therefore as n rises, the probability tends to zero.
Therefore we can conclude as n tends to infinity, the probability tends to zero from both sides.

Solution 5

Task A

Given, PDF of GMM variable X is $f_X = \sum_{i=1}^{K} p_i P[X_i = x]$. Let it's CDF be F_X. Then $F_X(x)$ is given by

$$F_X(x) = P[X \le x] \tag{5.1}$$
$$= \int_{-\infty}^{x} f_X(t)dt \tag{5.2}$$
$$= \int_{-\infty}^{x} \sum_{i=1}^{K} p_i P[X_i = t]dt \tag{5.3}$$
$$= \sum_{i=1}^{K} p_i \int_{-\infty}^{x} P[X_i = t]dt \tag{5.4}$$
$$= \sum_{i=1}^{K} p_i P[X_i \le x] \tag{5.5}$$
$$= \sum_{i=1}^{K} p_i F_{X_i}(x) \tag{5.6}$$

Where $F_{X_i}(x) = P[X_i \le x]$ is CDF of X_i.
Now, let CDF of output of the given algorithm be $F_{\mathcal{A}}(x) = P[\mathcal{A} \le x]$. Since the events that we choose $\mathcal{A}$ to be from the distribution i (say E_i) are disjoint for $i = 1, \ldots, k$.

$$F_{\mathcal{A}}(x) = P[\mathcal{A} \le x] \tag{5.7}$$
$$= \sum_{i=1}^{K} P[E_i] \cdot P[\mathcal{A} \le x | E_i] \tag{5.8}$$
$$= \sum_{i=1}^{K} p_i F_{X_i}(x) \tag{5.9}$$
$$= F_X(x) \tag{5.10}$$

We know that PDF of a random variable X with CDF $F_X(x)$ is $\frac{\partial F_X}{\partial x}$. Thus,

$$f_{\mathcal{A}}(x) = \frac{\partial F_{\mathcal{A}}}{\partial x} \tag{5.11}$$
$$= \frac{\partial F_X}{\partial x} \tag{5.12}$$
$$= f_X \tag{5.13}$$

Since x was arbitrary, for every $u \in \mathbb{R}$, $f_{\mathcal{A}}(u) = f_X(u)$. i.e the algorithm indeed samples from the GMM variable's distribution.

Task B

Since

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot P[X = t]dt \tag{5.14}$$
$$= \int_{-\infty}^{\infty} t \cdot \sum_{i=1}^{K} p_i P[X_i = t]dt \tag{5.15}$$
$$= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} P[X_i = t]dt \tag{5.16}$$
$$= \sum_{i=1}^{K} p_i \mathbb{E}[X_i] \tag{5.17}$$
$$= \sum_{i=1}^{K} p_i \mu_i \tag{5.18}$$

Let $\mu = \mathbb{E}[X]$.

$$\text{Var}[X] = \int_{-\infty}^{\infty} (t-\mu)^2 P[X = t]dt \tag{5.19}$$
$$= \int_{-\infty}^{\infty} (t-\mu)^2 \sum_{i=1}^{K} p_i P[X_i = t]dt \tag{5.20}$$
$$= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} (t-\mu)^2 P[X_i = t]dt \tag{5.21}$$
$$= \sum_{i=1}^{K} p_i \text{Var}[X_i] \tag{5.22}$$
$$= \sum_{i=1}^{K} p_i \sigma_i^2 \tag{5.23}$$

Let $\sigma^2 = \text{Var}[X]$.

$$\text{MGF}_X(t) = \int_{-\infty}^{\infty} e^{tX} P[X = x]dx \tag{5.24}$$
$$= \int_{-\infty}^{\infty} e^{tX} \sum_{i=1}^{K} p_i P[X_i = x]dx \tag{5.25}$$
$$= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} e^{tX} P[X_i = x]dx \tag{5.26}$$
$$= \sum_{i=1}^{K} p_i \text{MGF}_{X_i}(t) \tag{5.27}$$
$$= \sum_{i=1}^{K} p_i e^{t\mu_i + \frac{1}{2}t^2\sigma_i^2} \tag{5.28}$$

Task C

1. Given $Z = \sum_{i=1}^{K} p_i X_i$, where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=1}^{K} p_i X_i\right] \tag{5.29}$$

$$= \sum_{i=1}^{K} p_i \mathbb{E}[X_i] \tag{5.30}$$

$$= \sum_{i=1}^{K} p_i \mu_i \tag{5.31}$$

2. For $\mathrm{Var}[Z]$,

$$\mathrm{Var}[Z] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 \tag{5.32}$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{k} p_i X_i\right)^2\right] - \left(\sum_{i=1}^{k} p_i \mu_i\right)^2 \tag{5.33}$$

$$= \mathbb{E}\left[\sum_{i=1}^{N} p_i^2 X_i^2\right] + \mathbb{E}\left[2\sum_{i\neq j} p_i p_j X_i X_j\right] - \sum_{i=1}^{K} p_i^2 \mu_i^2 - 2\sum_{i\neq j} p_i p_j \mu_i \mu_j \tag{5.34}$$

$$= \sum_{i=1}^{K} p_i^2 \mathbb{E}[X_i^2] + 2\sum_{i\neq j} p_i p_j \mathbb{E}[X_i X_j] - \sum_{i=1}^{K} p_i^2 \mu_i^2 - 2\sum_{i\neq j} p_i p_j \mu_i \mu_j \tag{5.35}$$

Since each X_i, X_j ($i \neq j$) are independent

$$= \sum_{i=1}^{K} p_i^2 (\sigma_i^2 + \mu_i^2) + 2\sum_{i\neq j} p_i p_j \mu_i \mu_j - \sum_{i=1}^{K} p_i^2 \mu_i^2 - 2\sum_{i\neq j} p_i p_j \mu_i \mu_j \tag{5.36}$$

$$= \sum_{i=1}^{K} p_i^2 \sigma_i^2 \tag{5.37}$$

Thus, $\mathrm{Var}[Z] = \sum_{i=1}^{K} p_i^2 \sigma_i^2$.

3. Let $\Phi(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right)$.
For PDF $f_Z(u)$ of Z, Let $\hat{\mu} = \sum_{i=1}^{K} p_i \mu_i$ and $\hat{\sigma}^2 = \sum_{i=1}^{K} p_i^2 \sigma^2$. Then,

$$f_Z(u) = \Phi(\hat{\mu}, \hat{\sigma}^2) \tag{5.38}$$

This seems intuitive from the above results. Let's try to prove it using induction on i. For $i = 1$, $Z = p_1 X_1$, $\hat{\mu} = p_1 \mu_1$, $\hat{\sigma}^2 = p_1^2 \sigma_1^2$

$$f_Z(u) = \Phi(p_1 \mu_1, p_1 \mu_1^2) \tag{5.39}$$

$$= \Phi(\hat{\mu}, \hat{\sigma}^2) \tag{5.40}$$

Now, let $\forall j$ s.t $j < K$ if $Z = \sum_{i=1}^{j} p_i X_i$ then $f_Z(u) = \Phi(\hat{\mu}, \hat{\sigma}^2) = \Phi(\sum p_i \mu_i, \sum p_i^2 \sigma_i^2)$.

For $Z = \sum_{i=1}^{K} p_i X_i$, let $Z' = \sum_{i=1}^{K-1} p_i X_i$, $\hat{\mu}' = \sum_{i=1}^{K-1} p_i \mu_i$ and $\hat{\sigma}'^2 = \sum_{i=1}^{K-1} p_i^2 \sigma_i^2$

$$\begin{aligned} f_Z(u) &= \frac{\partial}{\partial u} P[Z \leq u] && (5.41)\\ &= \frac{\partial}{\partial u} \int_{-\infty}^{\infty} P[x \leq Z' \leq x + dx \text{ and } p_K X_k \leq u - x] dx && (5.42)\\ &= \frac{\partial}{\partial u} \int_{-\infty}^{\infty} f_{Z'}(x) P[p_K X_K \leq u - x] dx && (5.43)\\ &= \int_{-\infty}^{\infty} f_{Z'}(x) f_{X_K}(u - x) dx && (5.44)\\ &= \frac{1}{2\pi\sigma' \sigma_K} \int_{-\infty}^{\infty} \exp\left(-\frac{(x - \hat{\mu}')^2}{2\sigma'^2} - \frac{(x - \mu_K)^2}{2\sigma_K^2} \right) dx && (5.45) \end{aligned}$$

This can be simplified as it's an exponential of a quadratic equation, gaussial integral has to be used. This simplifies to the gaussian $\Phi(\hat{\mu}, \hat{\sigma}^2)$. Thus $\forall i \in \mathbb{N}$, $f_Z(u) = \Phi(\sum_{i=1}^{K} p_i \mu_i, \sum_{i=1}^{K} p_i^2 \sigma_i^2)$

We've proved that

$$\begin{aligned} Z &\sim \mathcal{N}\left(\sum_{i=1}^{K} p_i \mu_i, \sum_{i=1}^{K} p_i^2 \sigma_i^2 \right) && (5.46)\\ &\sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2) && (5.47) \end{aligned}$$

4. Now, MGF of Z, $M_Z(t)$ is

$$\begin{aligned} M_Z(t) &= \exp\left(\hat{\mu} t + \frac{\hat{\sigma}^2 t^2}{2} \right) && (5.48)\\ &= \exp\left(\sum_{i=1}^{K} p_i \mu_i t + \frac{\sum_{i=1}^{K} p_i^2 \sigma_i^2 t^2}{2} \right) && (5.49) \end{aligned}$$

5. X and Z are different. X is a mixture of gaussian distributions whereas Z itself is a gaussian distribution. Their PDFs would be different, X would have i peaks whereas Z would have only one peak. X is like a weighted sum of gaussian distributions. Z is a gaussian distribution whose mean and variance are weighted sum of means and variances of other gaussian distributions.
6. Z follows a gaussian distribution with mean $\sum_{i=1}^{K} p_i \mu_i$ and variance $\sum_{i=1}^{K} p_i^2 \sigma_i^2$.

Task D(B)

Suppose we have two arbitrary random variables X_1 and X_2. Let's try to prove the theorem in one direction first. Suppose that we know PDF/PMF of X_1 and X_2 i.e $f_{X_1}(x)$ and $f_{X_2}(x)$ are equal. Let MGF of X_1 and X_2 be $M_{X_1}(t)$ and $M_{X_2}(t)$ respectively. Let their common sample space be S (PDFs are equal). We have two cases:

1. X_1 **and** X_2 **are finite or discrete**:

$$\begin{aligned} f_{X_1}(x) &= f_{X_2}(x) \text{ for all } x \in S && (5.50)\\ \sum e^{tx} f_{X_1}(x) &= \sum e^{tx} f_{X_2}(x) \text{ for all } t \in S && (5.51)\\ M_{X_1}(t) &= M_{X_2}(t) \text{ for all } t \in S && (5.52) \end{aligned}$$

2. X_1 **and** X_2 **are continuous**:

$$f_{X_1}(x) = f_{X_2}(x) \text{ for all } x \tag{5.53}$$
$$\int_S e^{tx} f_{X_1}(x) = \int_S e^{tx} f_{X_2}(x) \text{ for all } t \tag{5.54}$$
$$M_{X_1}(t) = M_{X_2}(t) \text{ for all } t \tag{5.55}$$

Let's try to prove the other direction, i.e suppose MGFs of X_1 and X_2 are equal, i.e $M_{X_1}(t) = M_{X_2}(t)$. We again have cases

1. X_1 **and** X_2 **are finite or discrete**:

$$M_{X_1}(t) = M_{X_2}(t) \text{ for all } t \in S \tag{5.56}$$
$$\sum_{x \in S} e^{tx} f_{X_1}(x) = \sum_{x \in S} e^{tx} f_{X_2}(x) \text{ for all } t \tag{5.57}$$
$$\sum_{x \in S} e^{tx} (f_{X_1}(x) - f_{X_2}(x)) = 0 \tag{5.58}$$
$$\sum_{x \in S} a_x (e^t)^x = 0 \tag{5.59}$$

where $a_x = f_{X_1}(x) - f_{X_2}(x)$. This has infinitely many solutions, which is only possible if all the coefficients a_x are 0, Thus

$$a_k = 0 \text{ for all } k \tag{5.60}$$
$$f_{X_1}(x) = f_{X_2}(x) \text{ for all } x \tag{5.61}$$

Hence, PMFs of X_1 and X_2 are equal.

2. X_1 **and** X_2 **are continuous**:

$$M_{X_1}(t) = M_{X_2}(t) \text{ for all } t \in \mathbb{R} \in S \tag{5.62}$$
$$\int_{x \in S} e^{tx} f_{X_1}(x)dx = \int_{x \in S} e^{tx} f_{X_2}(x)dx \text{ for all } t \in \mathbb{R} \tag{5.63}$$
$$\int_{x \in S} e^{tx} (f_{X_1}(x) - f_{X_2}(x))dx = 0 \text{ for all } t \in \mathbb{R} \tag{5.64}$$
$$\int_{x \in S} a(x)(e^t)^x dx = 0 \text{ for all } t \in \mathbb{R} \tag{5.65}$$

where $a(x) = f_{X_1}(x) - f_{X_2}(x)$. This integral is 0 for all $t \in \mathbb{R}$ if and only if $a(x) = 0$ for all $x \in S$. Thus, $f_{X_1}(x) = f_{X_2}(x)$ for all $x \in S$. Hence, PDFs of X_1 and X_2 are equal.

Hence, we've shown that PDFs/PMFs of distributions (continuous/discrete respectively) are equal iff their MGFs are equal.