CS215 Assignment-2

Abhi Jain - 23b0903
Anushka Singhal - 23b0928
Sabil Ahmad - 23b1057

August 2024

Contents

1 Mathemagic

1.1 Task A

For a random variable X with a Probability mass function P , the PGF(probability-generating-function) of the distribution $P[X]$ of random variable X by

$$G(z) := P[z^X] = \sum_{n=0}^{\infty} P[X = n]z^n$$

for $X \sim Ber(p)$,

$$P[X = 0] = 1 - p \ , \ P[X = 1] = p \ , \ P[X = k] = 0 \ \forall \quad k > 1$$
$$PGF \ G_{Ber} = (1 - p).z^0 + p.z^1$$
$$PGF \ G_{Ber} = 1 - p + z.p$$

1.2 Task B

$For \ X \sim Bin(N, p) \ , \quad P[X = k] = \binom{N}{k}.p^k.(1-p)^k \quad P[X = k] = 0 \ \forall \quad k > N$

$$PGF \ G_{Bin} = \sum_{k=0}^{\infty} P[X = k]z^k$$
$$= \sum_{k=0}^{\infty} \binom{N}{k}.p^k.(1 - p)^{N-k}.z^k$$
$$= \sum_{k=0}^{\infty} \binom{N}{k}.(p.z)^k.(1 - p)^{N-k}$$

Since for $k > N$, the expected value of x= k is zero.

$$= \sum_{k=0}^{N} \binom{N}{k}.(p.z)^k.(1 - p)^{N-k}$$

We know the binomial theorem,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k$$

for $x \ = \ p.z \ and \ y \ = \ (1 - p)$

$$PGF \ G_{Bin} = \ \sum_{k=0}^{N} \binom{N}{k}.(p.z)^k.(1 - p)^{N-k} = (1 - p + p.z)^N$$
$$G_{Bin}(z) = (1 - p + p.z)^N$$

$$As \ we \ proved \ in \ part \ A \ , \quad G_{Ber}(z) = 1 - p + p.z$$
$$G_{Bin}(z) = G_{Ber}(z)^N$$

1.3 Task C

For $X_1, X_2 ....X_k$ independent random variables with nonnegative integer values with the same probability mass function P and common PGF G. We know,

$$G = \sum_{n=0}^{\infty} P[X = n] z^n$$

Now for the random variable $X = X_1 + X_2 ... + X_k$ defined on the cartesian product of the sample spaces underlying the random variable X_i.

$$G_{\sum} = \sum_{n=0}^{\infty} P[X = n] z^n$$

Now for the random variable X

$$P[X = n] = \sum_{i_1 + i_2 .... + i_k = n} P[X_1 = i_i].P[X_2 = i_2]......P[X_k = i_k]$$

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

$$= \sum_{n=0}^{\infty} \left(\sum_{i_1 + i_2 .... + i_k = n} P[X_1 = i_i].P[X_2 = i_2]......P[X_k = i_k] \right) z^n$$

Since , $X_1, X_2 ....X_k$ have the same probability mass function P and common PGF G ,replacing all with X_1 won't affect the sum.

$$= \sum_{n=0}^{\infty} \left(\sum_{i_1 + i_2 .... + i_k = n} P[X_1 = i_i].P[X_1 = i_2]......P[X_1 = i_k] \right) z^n$$

Let's Analyse $G(z)^k \ for \ X_1$

$$Since \ G(z) = \sum_{i=0}^{\infty} P[X_1 = i] z^i$$

$$G(z)^k = \left(\sum_{i=0}^{\infty} P[X_1 = i] z^i \right)^k$$

Clearly , the coefficient of z^n , is just picking k i's one from each of the k terms such that $i_1 + i_2 .... + i_k = n$

$$\sum_{i_1 + i_2 .... + i_k = n} P[X_1 = i_i].P[X_2 = i_2]......P[X_k = i_k]$$

Which is the same as coefficient of z^n in the expression of $G_{\sum}$ as we derived above as the coefficients are identical we can conclude that the

$$G_{\sum}(z) = G(z)^k$$

1.4 Task D

$$For\ X \sim Geo(p), \quad P[X=0]=0, \quad P[X=k] = p.(1-p)^{k-1}\ \forall \quad k>0$$

$$\begin{aligned} PGF\ G_{Geo} &= \sum_{k=0}^{\infty} P[X=k]z^k \\ &= \sum_{k=1}^{\infty} p.(1-p)^{k-1}.z^k \\ &= \frac{p}{1-p}.\sum_{k=1}^{\infty}(z-p.z)^k \\ &= \frac{p}{1-p}.\frac{z-p.z}{1-z+p.z} \\ &= \frac{p.z}{1-z+p.z} \end{aligned}$$

$$\implies G_{Geo}(z) = \frac{p.z}{1-z+p.z}$$

1.5 Task E

For $X \sim Bin(n,p)$ and $Y \sim NegBin(n,p)$, We derived the PGF for $Bin(n,p)$

$$G_X^{(n,p)}(z) = (1-p+p.z)^n$$

$$\begin{aligned} G_X^{(n,p^{-1})}(z^{-1}) &= (1-p^{-1}+p^{-1}.z^{-1})^n \\ (G_X^{(n,p^{-1})}(z^{-1}))^{-1} &= \left(\frac{p.z}{1-z+p.z}\right)^n \end{aligned}$$

Now we need to find $G_Y^{(n,p)}(z)$
We know that the $NegBin(n,p)$, The Negative Binomial is a sum of n Geometric random variables, we can write Y as

$$Y = Y_1 + Y_2......Y_n$$

Where $Y_1, Y_2....Y_n$, are $Y_i \sim Geo(p)\ \forall i's$, Thus they have the same PGF G and the Probability Mass function P.

As , we derived in **Task C**

$$G_{\sum}(z) = G(z)^n$$

Here , $G_{\sum}(z) = G_Y^{(n,p)}(z)$ and $G(z) = G_{Geo}$
As we derived in Task D

$$G_{Geo}(z) = \frac{p.z}{1-z+p.z}$$

Therefore , $G_{\sum}(z) = G_Y^{(n,p)}(z) = (G_{Geo})^n$

$$\implies G_Y^{(n,p)}(z) = \left(\frac{p.z}{1 - z + p.z}\right)^n$$

$$\implies G_Y^{(n,p)}(z) = \left(G_X^{(n,p^{-1})}(z^{-1})\right)^{-1}$$

1.6 Task F

We know that the PGF is defined as,

$$G(z) := P[z^X] = \sum_{n=0}^{\infty} P[X = n]z^n$$

Now , for $X \sim NegBin(n, p)$
We know that

$$P[X = k] = \binom{k-1}{n-1} p^n (1-p)^{k-n} \quad \forall k \geq n \quad and \quad P[X = k] = 0 \; \forall k < n$$

Thus,

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

Substiuting $k = r + n$

$$G_Y^{(n,p)}(z) = \sum_{r=0}^{\infty} \binom{r+n-1}{n-1} p^n (1-p)^r z^{r+n}$$

$$G_Y^{(n,p)}(z) = (pz)^n \sum_{r=0}^{\infty} \binom{r+n-1}{n-1} (1-p)^r z^r$$

$$G_Y^{(n,p)}(z) = (pz)^n \sum_{r=0}^{\infty} \binom{r+n-1}{n-1} (z - pz)^r$$

As derived above , we also know that

$$G_Y^{(n,p)}(z) = \left(\frac{p.z}{1 - z + p.z}\right)^n$$

$$G_Y^{(n,p)}(z) = (pz)^n \left(1 - z + p.z\right)^{-n}$$

Equating it from what we wrote above

$$(pz)^n (1 - z + p.z)^{-n} = (pz)^n \sum_{r=0}^{\infty} \binom{r+n-1}{n-1} (z - pz)^r$$

Putting $p.z - z = x$

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

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

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

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

Now, as defined in the question we can write

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}$$

putting $\alpha = -n \; and \; k = r \; in$

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

$$\binom{-n}{r} := (-1)^r \binom{r+n-1}{r}$$

Substituting the value of $\binom{-n}{r}$ we get

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

1.7 Task G

We know,

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

Differentiating it wrt to z we get

$$G^{'}(z) = \sum_{n=1}^{\infty} P[X=n] z^{n-1} \cdot n$$

Putting z = 1

$$G^{'}(1) = \sum_{n=1}^{\infty} P[X=n].n$$

$$G^{'}(1) = \sum_{n=0}^{\infty} P[X = n].n$$

Since, the random variable X was distributed only for non-negative integer values it's expected value would be defined as

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} P[X = n].n$$

$$\mathbb{E}[X] = G^{'}(1)$$

$Y \sim Bernoulli(p)$

$$G_Y(z) = 1 - p + pz$$

$$G^{'}_Y(1) = p$$

$Y \sim Binomial(n, p)$

$$G_Y(z) = (1 - p + p.z)^n$$

$$G^{'}_Y(z) = n(1 - p + p.z)^{n-1}.p$$

$$G^{'}_Y(1) = np$$

$Y \sim Geometric(p)$

$$G_Y(z) = \frac{p.z}{1 - z + p.z}$$

$$G^{'}_Y(z) = \frac{p}{(1 - z + p.z)^2}$$

$$G^{'}_Y(1) = \frac{1}{p}$$

$Y \sim NegativeBinomial(n, p)$

$$G_Y(z) = \left(\frac{p.z}{1 - z + p.z}\right)^n$$

$$G_Y{'}(z) = n.\left(\frac{p.z}{1 - z + p.z}\right)^{n-1}.\frac{p}{(1 - z + p.z)^2}$$

$$G^{'}_Y(z) = n.\left(\frac{p.z}{1 - z + p.z}\right)^{n-1}.\frac{p}{(1 - z + p.z)^2}$$

$$G^{'}_Y(1) = \frac{n}{p}$$

2 Normal Sampling

2.1 Task A

Given:

- X is a continuous r.v. with CDF $F_X : \mathbb{R} \to [0, 1]$
- F_X is invertible and non-decreasing function.
- r.v. $Y = F_X(X) \in [0, 1]$

Let's define CDF of Y as $F_Y : [0, 1] \to [0, 1]$. Therefore,

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[F_X(X) \leq y] \\ &= P[X \leq F_X^{-1}(y)] \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

Now, $F_Y(y) = y$ implies PDF of y is constant and equal to 1. Therefore, Y is uniformly distributed over $[0, 1]$.

2.2 Task B

Let Y $U(0, 1)$ and F_X is the CDF of X.
Algorithm $\mathcal{A}$ - Apply the inverse of CDF of X to transform the uniform random variable Y.

$$\mathcal{A} = F_X^{-1}(y)$$

Proof:
CDF of $Z = F_X^{-1}(Y)$

$$F_Z^u = P(Z \leq u) = P(F_X^{-1}(Y) \leq u)$$

As F_X is continuous and invertible,

$$P(F_X^{-1}(Y) \leq u) = P(Y \leq F_X(u))$$

As Y is uniformly distributed over [0,1],

$$P(Y \leq F_X(u)) = F_X(u)$$

So, $F_Z(u) = F_X(u)$
Hence proved that Z, the output of the algorithm $\mathcal{A}$ has the same CDF as X, and upon taking derivative, has the same PDF.

2.3 Task C

Its code is in the `2c.ipynb` file, where we have used two in-built functions `np.random.uniform` and `norm.ppf` to sample from gaussian with mean equal to loc and standard deviation equal to scale. Here the first function is used to generate random numbers uniformly in [0,1] and as the no. of samples is 10^5, its size is also kept 10^5. Then the second function transform the random numbers generated by the first function into samples from gaussian distribution using the inverse CDF (Cumulative Density function) which is also called PPF (Percent Point Function).
The parameter choices (μ, σ^2) for which we have to generate 10^5 independent samples using the function `sample(loc,scale)` are:

- (0, 0.2)
- (0, 1)
- (0, 5)
- (-2, 0.5)

The plots of the samples for each parameter choice are shown below in the same plot.

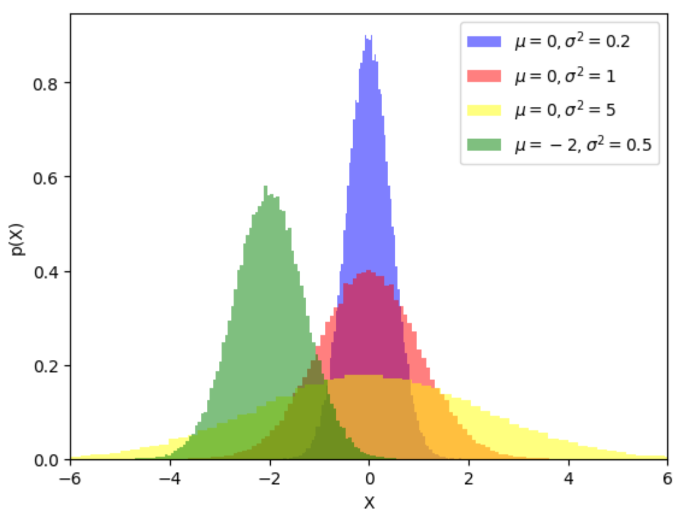

Figure 1: Samples for each parameter choice

2.4 Task D

Here first we wrote using loop, then as it was not fast enough, we omitted loops through numpy. So there is the commented code for loop too, which we wrote earlier. We used `np.random.choice` function for choosing between left and

right, then `np.sum` for finding the final positions by adding all the choices, then we shifted the positions by h, as it could be negative, then at last we counted the number of balls in each pocket by the function `np.bincount`.
The number of samples were $N = 10^5$. Different values of depth (h), resulted in different graphs:

- h = 10

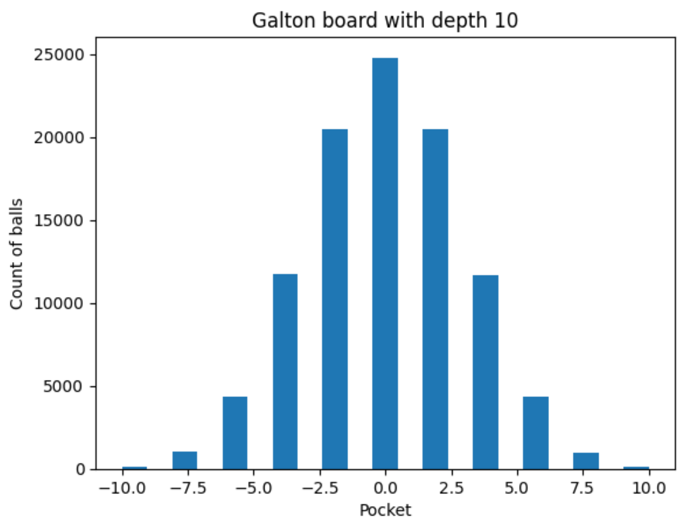

Figure 2: Galton Board with h=10

- h = 50

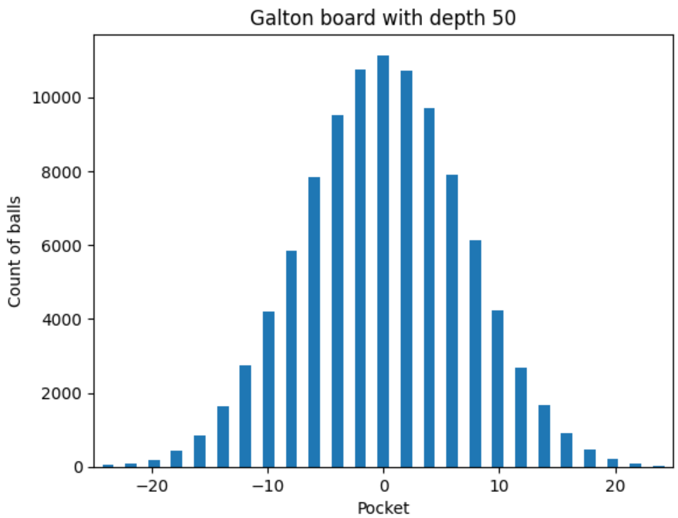

Figure 3: Galton Board with h=50

- h = 100

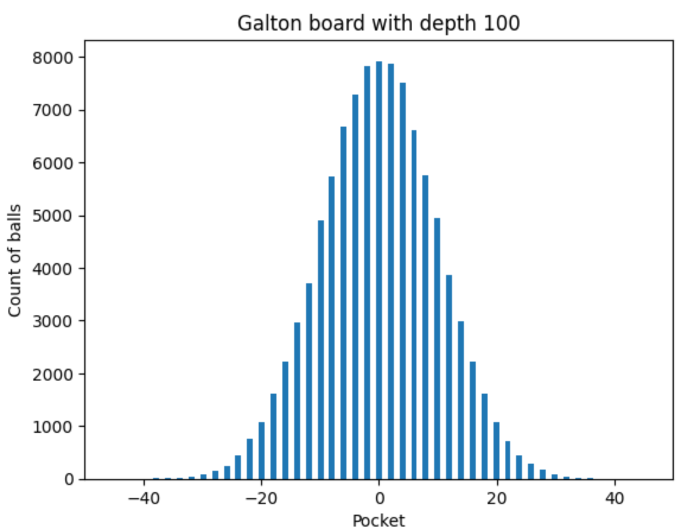

Figure 4: Galton Board with h=100

2.5 Task E

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

Now, when $i << \sqrt{h}$ [Using sterling approximation $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$]

$$\begin{aligned}
P_h[X = 2i] &= \frac{(2k)!}{(k+i)!(k-i)!2^{2k}} \\
&= \frac{\sqrt{4\pi k}\left(\frac{2k}{e}\right)^{2k}}{(\sqrt{2\pi(k+i)}\left(\frac{k+i}{e}\right)^{k+i}) \cdot (\sqrt{2\pi(k-i)}\left(\frac{k-i}{e}\right)^{k-i}) \cdot 2^{2k}} \\
&= \frac{\sqrt{k} \cdot k^{2k}}{\sqrt{\pi(k^2 - i^2)} \cdot (k+i)^{k+i} \cdot (k-i)^{k-i}} \\
&= \frac{k^{2k}}{\sqrt{\pi k} \cdot (k+i)^{k+i} \cdot (k-i)^{k-i}} \\
&= \frac{1}{\sqrt{\pi k} \cdot \left(1 + \frac{i}{k}\right)^{k+i} \cdot \left(1 - \frac{i}{k}\right)^{k-i}} \\
&= \frac{1}{\sqrt{\pi k} \cdot \left(1 - \left(\frac{i}{k}\right)^2\right)^k \cdot \left(1 + \frac{2i}{k-i}\right)^i}
\end{aligned}$$

Using approximation for small x, $(1+x)^n \approx 1 + nx$ and $e^x \approx 1 + x$

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

Therefore,

$$P[X = i] \approx \frac{1}{\sqrt{\pi k}} e^{-\frac{i^2}{4k}}$$

3 Fitting Data

3.1 Task A

Formula for the i^{th} moment is given by:

$$\mu_i = E[X^i] = \frac{1}{n}\sum_{j=1}^{n} x_j^i$$

Used `numpy.mean(data)` and `numpy.mean(data**2)` to calculate the first and second moments.

3.2 Task B

Used `plt.hist()` to draw the histogram, divided the data into 100 bins. The guess for mean for normal distribution from histogram is in between 6 and 8 more towards side of 6. So, my guess is 6.7.

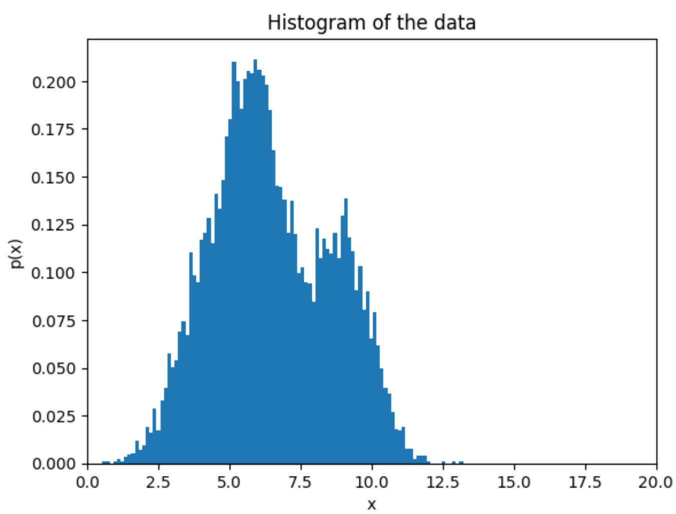

Figure 5: Histogram of the data

3.3 Task C

For binomial distribution X Bin(n, p). We have,

$$\mu_1^{\text{Bin}} = np$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = np(1-p)$$

This gives

$$\mu_2^{\text{Bin}} = \text{Var}(x) + (\mu_1^{\text{Bin}})^2 = n^2p^2 + np(1-p)$$

Now, used `fsolve` to solve the two equations for $\mu_1^{\mathrm{Bin}} = \mu_1$ and $\mu_2^{\mathrm{Bin}} = \mu_2$. Solving the equation gives the following:

$$n = 20 \text{ and } p = 0.3296865296375572$$

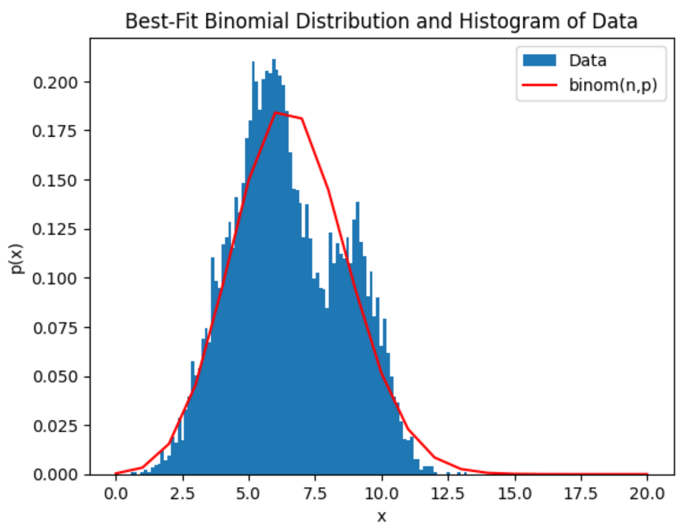

Figure 6: The best binomial distribution approximation to the true distribution

3.4 Task D

Let $X \sim \mathrm{Gam}(k, \theta)$ be the gamma random variable. We have gamma probability density distribution function defined as: [1]

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

,where $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$ is the gamma function.

[1]Gamma probability density is 0 for $x < 0$

Deriving $\mu_1^{\mathbf{Gam}}$ and $\mu_2^{\mathbf{Gam}}$ for Gamma Probability Distribution

For this we first prove the property of gamma function $\Gamma(k+1) = k\Gamma(k)$:

$$\begin{aligned}
\Gamma(k+1) &= \int_0^\infty t^k e^{-t} dt \\
&= t^k \int e^{-t} dt - \int kt^{k-1} \left(\int e^{-t} dt \right) dt \Big|_0^\infty \\
&= -t^k e^- t + k \int t^{k-1} e^{-t} dt \Big|_0^\infty \\
&= k \int_0^\infty t^{k-1} e^{-t} dt \\
&= k\Gamma(k)
\end{aligned}$$

Now, using this property of gamma function

$$\begin{aligned}
\mu_1^{\text{Gam}} = E[X] &= \int_0^\infty x \cdot \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{\frac{-x}{\theta}} dx \\
&= \int_0^\infty \frac{1}{\theta^k \Gamma(k)} x^k e^{\frac{-x}{\theta}} dx \\
&= \int_0^\infty \frac{k}{\theta^k \Gamma(k+1)} x^k e^{\frac{-x}{\theta}} dx \\
&= k\theta \int_0^\infty \frac{1}{\theta^{k+1} \Gamma(k+1)} x^k e^{\frac{-x}{\theta}} dx \\
&= k\theta \int_0^\infty f(x, k+1, \theta) dx \\
&= k\theta \cdot 1 \\
&= k\theta
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mu_2^{\text{Gam}} = E[X^2] &= \int_0^\infty x^2 \cdot \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{\frac{-x}{\theta}} dx \\
&= \int_0^\infty \frac{1}{\theta^k \Gamma(k)} x^{k+1} e^{\frac{-x}{\theta}} dx \\
&= \int_0^\infty \frac{k(k+1)}{\theta^k \Gamma(k+2)} x^{k+1} e^{\frac{-x}{\theta}} dx \\
&= k(k+1)\theta^2 \int_0^\infty \frac{1}{\theta^{k+2} \Gamma(k+2)} x^{k+1} e^{\frac{-x}{\theta}} dx \\
&= k(k+1)\theta^2 \int_0^\infty f(x, k+2, \theta) dx \\
&= k(k+1)\theta^2
\end{aligned}$$

Therefore, we have derived $\mu_1^{\text{Gam}} = k\theta$ and $\mu_2^{\text{Gam}} = k(k+1)\theta^2$. Now, again comparing these moments to the actual moments and solving the equaitons for k and θ. Solving the equations gives:

$$k = 20 \text{ and } \theta = 0.6703134703624472$$

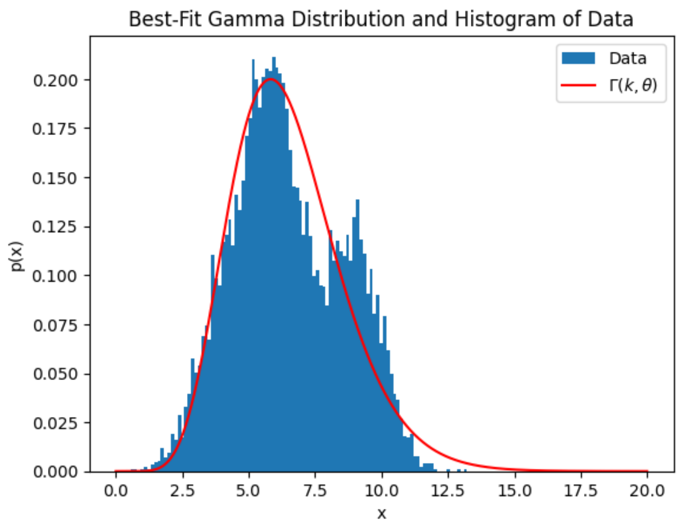

Figure 7: The best Gamma distribution approximation to the true distribution

3.5 Task E

Given a dataset S and a choice of parameter $\lambda = \lambda_0$ for a family of distributions $P[\lambda]$, parameterized by λ, define the likelihood of λ_0 by

$$\mathcal{L}(\theta|S) = P_{\lambda_0}[S] = \prod_{i=1}^{n} P_{\lambda_0}[X_i]$$

The average log-likelihood is typically calculated for parameter θ and dataset S is given as:

$$l(\lambda_0|S) = \frac{1}{n} log\mathcal{L}(\theta|S)$$

Therefore, log-likelihood of binomial and gamma distribution functions can be calculated using `binom.logpmf()` and `gamma.logpdf()`.

- Log likelihood of th e best-fit binomial distribution: -2.157068115434675
- Log likelihood of the best-fit gamma distribution: -2.1608217722066265

The best-fit distribution is the binomial distribution.

3.6 Task F

The GMM distribution is as follows:

$$P(x) = \frac{1}{\sqrt{2\pi}} \left[p_1 \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} \right) + p_2 \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2} \right) \right]$$

Here, are the first four moments:

$$\text{eq1: } p_1\mu_1 + p_2\mu_2 = \mu_1^{\text{GMM}}$$

$$\text{eq2: } p_1(\mu_1^2 + \sigma_1^2) + p_2(\mu_2^2 + \sigma_2^2) = \mu_2^{\text{GMM}}$$

$$\text{eq3: } p_1(\mu_1^3 + 3\mu_1\sigma_1^2) + p_2(\mu_2^3 + 3\mu_2\sigma_2^2) = \mu_3^{\text{GMM}}$$

$$\text{eq4: } p_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + p_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) = \mu_4^{\text{GMM}}$$

Solving the equations with $\sigma_1 = \sigma_2 = 1$, we get:

$$p_1 = 0.6118740341612672 \text{ and } \mu_1 = 5.129607694288399$$

$$p_2 = 0.38264565119245014 \text{ and } \mu_2 = 8.774363054420633$$

Log likelihood of the best-fit two-component gaussian distribution: -2.183038744911406. It was almost similar to the binomial and gamma distributions average log likelihoods.

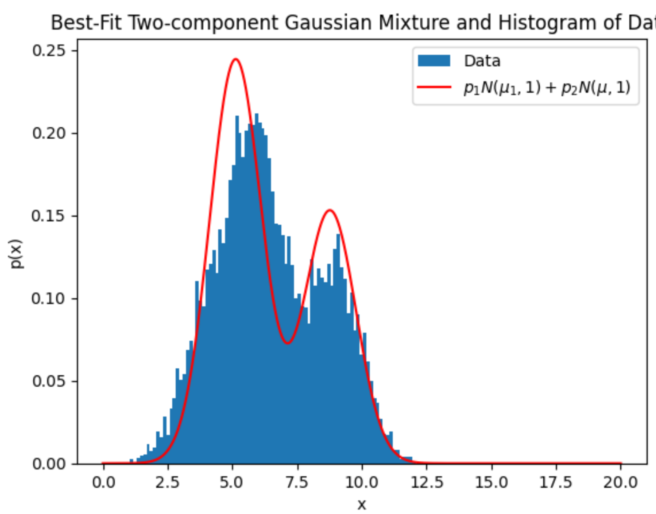

Figure 8: The best two-component unit-variance GMM approximation to the true distribution

Improving the number of parameters to 6 gives:

$p_1 = 0.7544987644483389$, $\mu_1 = 5.61551680544903$, $\sigma_1 = 1.5395974129232748$

$p_2 = 0.2461092217321142$, $\mu_2 = 9.179847512339403$, $\sigma_2 = 0.9751223502251171$

The Log likelihood of the best-fit two-component gaussian distribution with 6 parameters: -2.11464787612961. This was the best among all others.

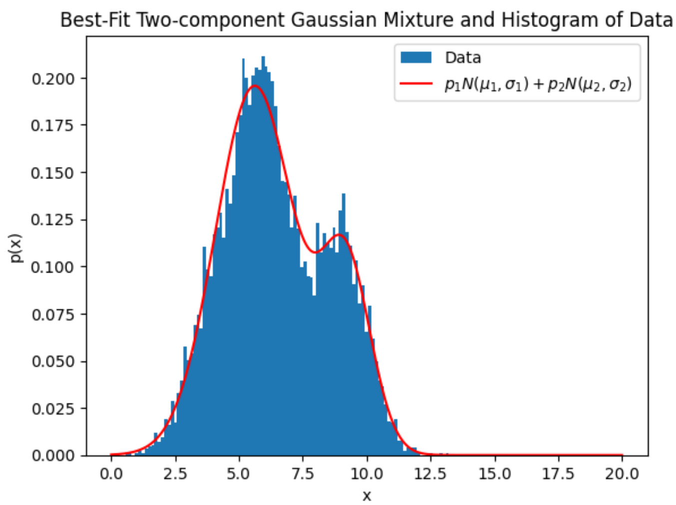

Figure 9: Two more parameters bring the total to 6 parameters estimated using the first six moments

4 Quality in Inequalities

4.1 Task A

Intutive Proof
One way we can think intuitively is that say , that the $P[X \geq a]$ increases and say it does cross the $\frac{\mathbb{E}[X]}{a}$ but as we increase the $P[X \geq a]$ the value of $\mathbb{E}[X]$ will also increase as

$$\mathbb{E}[X] = \int_0^a p(x).x\ dx + \int_a^\infty p(x).x\ dx$$

since overall we're increasing the $p(x)$ for $x > a$ and the value of $p(x)$ decreases as $x < a$ the overall value of $\mathbb{E}[X]$ increases since we're increasing the $p(x)$ coefficients of the larger x terms overall and intuitively we can convince ourselves that the inequality holds itself.
Rigorous Proof
Rigorous Proof for assuming non negative continous Random Variable X ,with the probability Density Function p

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

$$We\ know,\ P[X \geq a]\ = \int_a^\infty p(x)\ dx$$

$$Also,\ \mathbb{E}[X] = \int_0^\infty p(x).x\ dx$$

,

$$\mathbb{E}[X] = \int_0^a p(x).x\ dx + \int_a^\infty p(x).x\ dx$$

$$Since\ a > 0,\ \int_a^\infty p(x).x\ dx \geq a.\int_a^\infty p(x)\ dx$$

$$\implies \int_0^a p(x).x\ dx + \int_a^\infty p(x).x\ dx \geq a.\int_a^\infty p(x)\ dx$$

$$\implies \mathbb{E}[X] \geq a.\int_a^\infty p(x)$$

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

Similar Proof for Discrete Random Variables can be achieved by removing the integral by the summation over all the values and the probability density function by the probability mass function.

4.2 Task B

To Prove , for a random variable X with mean μ and variance σ^2 for every $\tau > 0$

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

For any $a \geq 0$

$$P[X - \mu \geq \tau] = P[X - \mu + a \geq \tau + a] \leq P[(X - \mu + a)^2 \geq (\tau + a)^2]$$

Using Markov's Inequality from above

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

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

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

To obtain a bound we see the minima of the term on the rhs for $\forall a \geq 0$

$$f(a) = \frac{\sigma^2 + a^2}{(\tau + a)^2}$$

$$f^{`}(a) = \frac{2(\tau + a)(a\tau - \sigma^2)}{(\tau + a)^4}$$

Clearly , the function achieves it's minimum at $a = \frac{\sigma^2}{\tau}$

$$f(\frac{\sigma^2}{\tau}) = \frac{\sigma^2}{\sigma^2 + \tau^2}$$

Thus, plugging the value of f(a) in the above inequality we get the desired result as,

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

4.3 Task C

The MGF for a random variable X is defined as $\mathbb{E}[e^{tX}]$,We know

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tr} f_X(r)dr$$

$$M_X(t) = \int_{-\infty}^{x} e^{tr} f_X(r)dr + \int_{x}^{\infty} e^{tr} f_X(r)dr$$

Also we , Know

$$P[X \geq x] = \int_{x}^{\infty} f_X(r)dr \quad and \quad P[X \leq x] = \int_{-\infty}^{x} f_X(r)dr$$

Clearly $\forall t > 0,\ e^{tx}\ increases\ as\ the\ value\ of\ x\ increases$

$$\int_{-\infty}^{x} e^{tr} f_X(r)dr \leq e^{tx} \int_{-\infty}^{x} f_X(r)dr$$

$$\implies M_X(t) \leq e^{tx} \int_{-\infty}^{x} f_X(r)dr$$

$$\implies e^{-tx} M_X(t) \leq \int_{-\infty}^{x} f_X(r)dr$$

$$\implies e^{-tx} M_X(t) \leq P[X \leq x]$$

Clearly $\forall t < 0,\ e^{tx}\ decreases\ as\ the\ value\ of\ x\ increases$

$$\int_{x}^{\infty} e^{tr} f_X(r)dr \leq e^{tx} \int_{x}^{\infty} f_X(r)dr$$

$$\implies M_X(t) \leq e^{tx} \int_{-\infty}^{x} f_X(r)dr$$

$$\implies e^{-tx} M_X(t) \leq \int_{-\infty}^{x} f_X(r)dr$$

$$\implies e^{-tx} M_X(t) \leq P[X \leq x]$$

4.4 Task D

4.4.1 1

Since , $Y = \sum_{i=1}^{i=n} X_i$

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

By Linearity of expectations

$$\mathbb{E}[Y] = \sum_{i=1}^{i=n} \mathbb{E}[X_i]$$

Given, that $\mathbb{E}[X_i] = p_i$

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

4.4.2 2

Claim : *Moment generating function of the sum of independent random variables is just the product of the individual moment generating functions*
To, prove this suppose X and Y to independent random variables and $Z = X+Y$

$$\mathbb{E}[e^{tZ}] = \mathbb{E}[e^{t(X+Y)}]$$

$$\mathbb{E}[e^{tZ}] = \mathbb{E}[e^{tX}.e^{tY}]$$

Now, since for two independent random variables $\mathbb{E}[X.Y] = \mathbb{E}[X].\,\mathbb{E}[Y]$

$$\mathbb{E}[e^{tZ}] = \mathbb{E}[e^{tX}].\,\mathbb{E}[e^{tY}]$$

$$M_{X+Y}(t) = M_X(t).M_Y(t)$$

Now,

$$P[Y \geq (1+\delta)\mu] = P[e^{tY} \geq e^{t(1+\delta)\mu}] \;\forall t > 0$$

Now , we have from Markov's Inequality

$$P[Y \geq (1+\delta)\mu] = [e^{tY} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tY}]}{e^{t(1+\delta)\mu}}$$

$$Since,\; Y = \sum_{i=1}^{i=n} X_i$$

$$P[Y \geq (1+\delta)\mu] \leq \frac{\mathbb{E}[e^{t(\sum_{i=1}^{i=n} X_i)}]}{e^{t(1+\delta)\mu}}$$

$$P[Y \geq (1+\delta)\mu] \leq \frac{\mathbb{E}[\prod_{i=1}^{i=n} e^{tX_i}]}{e^{t(1+\delta)\mu}}$$

Using, the claim we proved above this is the same as

$$P[Y \geq (1+\delta)\mu] \leq \frac{\prod_{i=1}^{i=n} \mathbb{E}[e^{tX_i}]}{e^{t(1+\delta)\mu}}$$

$$\mathbb{E}[e^{tX_i}] = 1 - p_i + e^t p_i = 1 + p_i(e^t - 1)$$

Since , $1 + x \leq e^x \;\forall x$

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

$$\prod_{i=1}^{i=n} \mathbb{E}[e^{tX_i}] \leq \prod_{i=1}^{i=n} e^{p_i(e^t-1)}$$

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

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

4.4.3 3

To get the optimal value we differentiate it with respect to t and obtain a minima by analysing the sign of the second derivative to obtain a better bound for the inequality.
On Differentiating we obtain a minima at $t = \ln(1+\delta)$ (note that $\delta > 0$ as we assumed it to hold for $t > 0$) which yields the minimum value as,

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

This,yields the tighter bound namely the Chernoff bound for $\delta > 0$

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

4.5 Task E

For $X_1, X_2, X_3.....X_n$ bernoulli Random Variables each with mean μ.

$$A_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

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

We can rewrite this as,

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

So, basically we need to calculate,

$$\lim_{n\to\infty} P[\ nA_n \geq n\mu + n\epsilon] + \lim_{n\to\infty} P[\ nA_n \leq n\mu - n\epsilon]$$

Or,

$$\lim_{n\to\infty} P[\ nA_n \geq n\mu(1+\frac{\epsilon}{\mu})] + \lim_{n\to\infty} P[\ nA_n \leq n\mu(1-\frac{\epsilon}{\mu})]$$

We'll come back to this later first let's prove that for $Y = X_1 + X_2 + X_3....X_n$ where each X_i is a Bernoulli Random Variable,

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

We , go about this the same way as we did for the upper part , as follows

$$P[Y \leq (1+\delta)\mu] = P[e^{tY} \geq e^{t(1+\delta)\mu}]\ \forall t < 0$$

Now , we have from Markov's Inequality

$$P[Y \leq (1+\delta)\mu] = [e^{tY} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tY}]}{e^{t(1+\delta)\mu}}$$

Now everything is the same as we did in the above part and we end up with

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

To get the optimal value we differentiate it with respect to t and obtain a minima by analysing the sign of the second derivative to obtain a better bound for the inequality.
On Differentiating we obtain a minima at $t = \ln(1+\delta)$ (note that $\delta < 0$ as we assumed it to hold for $t < 0$) which yields the minimum value as,

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

Thus , we can say for $\delta < 0$

$$P[Y \leq (1+\delta)\mu] \leq \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}$$

Now , having proved this let's come back to the original problem

$$\lim_{n\to\infty}(P[\ nA_n \geq n\mu(1+\frac{\epsilon}{\mu})] + P[\ nA_n \leq n\mu(1-\frac{\epsilon}{\mu})])$$

Now as we derived above , since nA_n is a sum of n Bernoulli Random Variables with mean of each μ, $\implies mean(nA_n) = n\mu$, we can write

$$P\left[\ nA_n \geq n\mu\left(1+\frac{\epsilon}{\mu}\right)\right] \leq \frac{e^{n\epsilon}}{(1+\frac{\epsilon}{\mu})^{(n(\mu+\epsilon))}}$$

Similarly we can write and the final expression comes out to be

$$\lim_{n\to\infty} P[\ |A_n-\mu| \geq \epsilon] \leq \lim_{n\to\infty}\left(\left(\frac{e^{\frac{\epsilon}{\mu}}}{(1+\frac{\epsilon}{\mu})^{(1+\frac{\epsilon}{\mu})}}\right)^{n\mu} + \left(\frac{e^{-\frac{\epsilon}{\mu}}}{(1-\frac{\epsilon}{\mu})^{(1-\frac{\epsilon}{\mu})}}\right)^{n\mu}\right)$$

Now, we consider the function, with domain $(-1,\infty)$

$$f(\delta) = \frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}$$

$$f^{'}(\delta) = -e^{\delta}(1+\delta)^{-(1+\delta)}\ln(1+\delta)$$

Since , $f(0) = 1$ and $f^{'}(\delta) > 0\ for\ \delta < 0\ and\ f^{'}(\delta) < 0\ for\ \delta > 0$, we can say that the function f takes it maximum value at $\delta = 0$ in it's domain and $\forall\ \delta \neq 0\ 0 < f(\delta) < 1$. Now,

$$\lim_{n\to\infty} P[\ |A_n-\mu| \geq \epsilon] \leq \lim_{n\to\infty}\left(\left(\frac{e^{\frac{\epsilon}{\mu}}}{(1+\frac{\epsilon}{\mu})^{(1+\frac{\epsilon}{\mu})}}\right)^{n\mu} + \left(\frac{e^{-\frac{\epsilon}{\mu}}}{(1-\frac{\epsilon}{\mu})^{(1-\frac{\epsilon}{\mu})}}\right)^{n\mu}\right)$$

Clearly the above expression tends to 0 $\forall \epsilon > 0$ as both the sums in the limits are $f(\frac{\epsilon}{\mu})$ and $f(-\frac{\epsilon}{\mu})$ where $\epsilon > 0$ which are both clealrly < 1 as we proved above and we know $\lim_{n\to\infty}(c)^n = 0,\ when\ |c| < 1$ Thus ,

$$\lim_{n\to\infty} P[\ |A_n - \mu| \geq \epsilon] \leq 0$$

Also Since , $P[X] \geq 0$

$$0 \leq \lim_{n\to\infty} P[\ |A_n - \mu| \geq \epsilon] \leq 0$$

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

5 A Pretty "Normal" Mixture

5.1 Task A

By definiton of GMM, we have:

$$P[X = x] = \sum_{i=1}^{k} P[X_i = x]$$

,where each $X_i \sim N(\mu_i, \sigma_i^2)$ is gaussian random variable $\forall i \in \{1, 2, \ldots, k\}$. Moreover, each $p_i \geq 0$ and $\sum_{i=1}^{k} p_i = 1$. This can be rewritten as:

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

Algorithm's Output PDF

By Law of Total Probability,

$$f_{\mathcal{A}}(u) = \sum_{i=1}^{k} Pr(\text{choose i}) \cdot f_{\mathcal{A} \text{ corresponding to chosen } i}(u)$$

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

Clearly, $f_{\mathcal{A}} = f_X$.

5.2 Task B

5.2.1 $E[X]$

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\
&= \int_{-\infty}^{\infty} x \cdot \left(\sum_{i=1}^{k} p_i \cdot f_{X_i}(x) \right) dx \\
&= \sum_{i=1}^{k} p_i \left(\int_{-\infty}^{\infty} x \cdot f_{X_i}(x) dx \right) \\
&= \sum_{i=1}^{k} p_i \cdot E[X_i] \\
&= \sum_{i=1}^{k} p_i \cdot \mu_i
\end{aligned}$$

5.2.2 Var$[X]$

$$
\begin{aligned}
&\mathrm{Var}[X] = E[(X-\mu)^2] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f_X(x)dx \\
&= \sum_{i=1}^{k} p_i \left(\int_{-\infty}^{\infty} (x-\mu_i+\mu_i-\mu)^2 \cdot f_{X_i}(x)dx \right) \\
&= \sum_{i=1}^{k} p_i \left(\int_{-\infty}^{\infty} (x-\mu_i)^2 f_{X_i}(x)dx + 2(\mu_i-\mu) \int_{-\infty}^{\infty} (x-\mu_i) f_{X_i}(x)dx + (\mu_i-\mu)^2 \int_{-\infty}^{\infty} f_{X_i}(x)dx \right) \\
&= \sum_{i=1}^{k} p_i \left(\sigma^2 + 2(\mu_i-\mu)\cdot(\mu_i-\mu_i) + (\mu_i-\mu)^2 \right) \\
&= \sum_{i=1}^{k} p_i \left(\sigma^2 + (E[X_i]-\mu)^2 \right) \\
&= \sum_{i=1}^{k} p_i \left(\sigma^2 + \mu_i^2 \right) - 2\mu \sum_{i=1}^{k} p_i \cdot \mu_i + \mu^2 \sum_{i=1}^{k} p_i \\
&= \sum_{i=1}^{k} p_i \left(\sigma^2 + \mu_i^2 \right) - 2\mu \cdot \mu + \mu^2 \\
&= \sum_{i=1}^{k} p_i \left(\sigma^2 + \mu_i^2 \right) - \mu^2
\end{aligned}
$$

5.2.3 $M_X(t)$

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)dx \\
&= \sum_{i=1}^{k} p_i \left(\int_{-\infty}^{\infty} e^{tx} \cdot f_{X_i}(x)dx \right) \\
&= \sum_{i=1}^{k} p_i \cdot M_{X_i}(t) \\
&= \sum_{i=1}^{k} p_i \cdot \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2 \right)
\end{aligned}
$$

5.3 Task C

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

Now, Moment generating function of Z is:

$$M_Z(t) = E[e^{tZ}] = E[e^{t\sum_{i=1}^{k} p_i X_i}]$$

$$\implies M_Z(t) = E[\prod_{i=1}^{k} e^{tp_i X_i}]$$

Since, we are given that all X_i are independent random variables. Therefore,

$$M_Z(t) = \prod_{i=1}^{k} E[e^{tp_i X_i}] = \prod_{i=1}^{k} M_{p_i X_i}(t)$$

For, a gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. We have,

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

and if $Y = a \cdot X + b$, where a and b are scalars then, $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ Therefore,

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

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

So, this is now MGF of a Normal distribution with $\mu = \sum_{i=1}^{k} p_i \mu_i$ and $\sigma^2 = \sum_{i=1}^{k} p_i^2 \sigma_i^2$. Therefore, $Z \sim \mathcal{N}(\mu, \sigma^2)$ by using the theorem from D part of the problem. So, we can write the following results as per the results of normal distribution as Z is a normal distribution.

1. $E[X] = \mu = \sum_{i=1}^{k} p_i \mu_i$
2. $\text{Var}[X] = \sigma^2 = \sum_{i=1}^{k} p_i^2 \sigma_i^2$
3. $f_Z(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
4. $M_Z(t) = exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
5. No, X and Z are don't have the same properties. Z is a normal gaussian random vairable while X is GMM. The probability distribution of X is weighted sum of probability distribution of normal probability distributions while Z is weighted sum of random variables.
6. Z follows the normal gaussian distribution.

5.4 Task D (B)

To Prove: For a random variable X, if it is finite and discrete, then its MGF and PDF uniquely determine each other.

First proving that if PDF are same then MGF are same. Let's say we have two random variables X and Y, s.t. $\forall u, f_X(u) = f_Y(u)$. Since both PDF are same so their domains are also same and hence the r.v. X and Y can take the same set of values U only. Therefore, we can write:

$$\forall t, \sum_u e^{tu} f_X(u) = \sum_u e^{tu} f_Y(u) \implies M_X(t) = M_Y(t)$$

Hence, both X and Y have same MGF.

Now, proving that if MGF are same then we have the same MGF: Let's say we have two random variable X and Y taking values from the finite set $S = \{s_1, s_2, \ldots, s_n\}$. Therefore,

$$\forall t, \sum_{i=1}^{n} e^{ts_i} F_X(s_i) = \sum_{i=1}^{n} e^{ts_i} F_Y(s_i)$$

$$\implies \sum_{i=1}^{n} e^{ts_i} [F_X(s_i) - F_Y(s_i)] = 0$$

Let $\forall i, k_i = F_X(s_i) - F_Y(s_i)$. Therefore,

$$\sum_{i=1}^{n} e^{ts_i} k_i = 0$$

Now, this is a polynomial in e^t with k_i as coefficients and s_i as powers. So, this finite polynomial has infinite zeros which isn't possible. Therefore,

$$\forall i, k_i = 0 \implies F_X(s_i) = F_Y(s_i)$$

Hence, we have PDF for both X and Y same.

Final conclusion of X and Z

Finally, we can conclude about X and Z is that X is that Z is weighted sum of normal gaussian randoma variables, which finally comes out to be a normal gaussian random variable. While for X we have the probability distribution of X as the weighted mean of probability distributions of normal gaussian variables.