CS215 - Assignment 2

Ramachandran S - 23b1052
Keyaan KR - 23b0977
Harith S - 23b1085

September 3, 2024

Contents

1 Mathemagic

1.1 Task A

Let $X \sim \text{Ber}(p)$, meaning X is a Bernoulli random variable with parameter p. The probability mass function (PMF) for X is given by:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

The Probability Generating Function (PGF), $G_{\text{Ber}}(z)$, is defined as:

$$G_{\text{Ber}}(z) = \mathbb{E}[z^X] = \sum_{n=0}^{1} P(X = n) z^n$$

Substituting the values from the PMF, we get:

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

Result

Thus, the Probability Generating Function (PGF) for a Bernoulli random variable X with parameter p is:

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

1.2 Task B

Let $X \sim \text{Bin}(n, p)$, meaning X is a binomial random variable with parameters n and p. The probability mass function (PMF) for X is given by:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

The Probability Generating Function (PGF), $G_{\text{Bin}}(z)$, is defined as:

$$G_{\text{Bin}}(z) = \mathbb{E}[z^X] = \sum_{k=0}^{n} P(X = k) z^k$$

Substituting the values from the PMF, we get:

$$\begin{aligned} G_{\text{Bin}}(z) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} z^k \\ G_{\text{Bin}}(z) &= \sum_{k=0}^{n} \binom{n}{k} (pz)^k (1 - p)^{n-k} \\ G_{\text{Bin}}(z) &= (1 - p + pz)^n \end{aligned}$$

Result

From Task A, we know that the PGF for a Bernoulli random variable is given by $G_{\text{Ber}}(z) = 1 - p + pz$. Substituting this into the expression for the binomial PGF, we obtain:

$$G_{\text{Bin}}(z) = G_{\text{Ber}}(z)^n$$

1.3 Task C

Lemma

Let $X_1, X_2, \ldots, X_k$ be independent non-negative-integer-valued random variables, each distributed with the same probability mass function P and probability generating function $G(z)$. Then the PGF $G_\Sigma(z)$ of their sum $X = X_1 + X_2 + \cdots + X_k$ is given by:

$$G_\Sigma(z) = G(z)^k$$

The PGF $G(z)$ for a random variable X_i is defined as:

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

Proof: Consider the random variable $X = X_1 + X_2 + \cdots + X_k$. Let's find the PGF $G_\Sigma(z)$ of X:

$$G_\Sigma(z) = \mathbb{E}[z^X]$$
$$G_\Sigma(z) = \sum_{n=0}^{\infty} P(X = n) z^n$$
$$G_\Sigma(z) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) z^{n_1+n_2+\cdots+n_k}$$

Theorem (Joint Probability for Independent Variables)

If random variables $X_1, X_2, \ldots, X_k$ are independent, then their joint probability is the product of their individual probabilities:

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = P(X_1 = n_1) P(X_2 = n_2) \cdots P(X_k = n_k)$$

Applying this theorem to our sum:

$$G_\Sigma(z) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} P(X_1 = n_1) P(X_2 = n_2) \cdots P(X_k = n_k) z^{n_1+n_2+\cdots+n_k}$$
$$G_\Sigma(z) = \left(\sum_{n_1=0}^{\infty} P(X_1 = n_1) z^{n_1} \right) \cdot \left(\sum_{n_2=0}^{\infty} P(X_2 = n_2) z^{n_2} \right) \cdots \left(\sum_{n_k=0}^{\infty} P(X_k = n_k) z^{n_k} \right)$$

Each sum in parentheses is the PGF $G(z)$ of the random variable X_i, which is the same for each X_i. Therefore:

$$G_\Sigma(z) = G(z) \cdot G(z) \cdot \cdots \cdot G(z) \quad \text{(k times)}$$
$$G_\Sigma(z) = G(z)^k$$

Thus, we have proven that the PGF of the sum $X = X_1 + X_2 + \cdots + X_k$ of independent, identically distributed random variables is $G(z)^k$, where $G(z)$ is the common PGF of each X_i.

1.4 Task D

Let $X \sim \text{Geo}(p)$, where X is a Geometric random variable with parameter p. The probability mass function (PMF) of X is given by:

$$P(X = n) = (1 - p)^{n-1} p \quad \text{for } n = 1, 2, 3, \ldots$$

The Probability Generating Function (PGF), $G_{\text{Geo}}(z)$, is defined as:

$$G_{\text{Geo}}(z) = \mathbb{E}[z^X] = \sum_{n=0}^{\infty} P(X = n) z^n$$

Substituting the PMF of the Geometric distribution, we get:

$$G_{\text{Geo}}(z) = \sum_{n=1}^{\infty} (1-p)^{n-1} p z^n$$

$$G_{\text{Geo}}(z) = pz \sum_{n=1}^{\infty} ((1-p)z)^{n-1}$$

$$G_{\text{Geo}}(z) = pz \sum_{n=0}^{\infty} ((1-p)z)^{n}$$

The summation is a geometric series with ratio $(1-p)z$, which converges when $|(1-p)z| < 1$. Using the formula for the sum of an infinite geometric series $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ for $|r| < 1$, we get:

$$G_{\text{Geo}}(z) = pz \cdot \frac{1}{1-(1-p)z}$$

Result

Thus, the PGF of a Geometric random variable $X \sim \text{Geo}(p)$ is:

$$G_{\text{Geo}}(z) = \frac{pz}{1-(1-p)z}$$

1.5 Task E

From Task B, we know the PGF of a Binomial random variable $X \sim \text{Bin}(n, p)$ is given by:

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

Therefore, $G_X^{(n,p^{-1})}(z^{-1})$ is given by:

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

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

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

Let $Y \sim \text{NegBin}(n, p)$, which is the sum of n independent Geometric random variables:

$$Y = Y_1 + Y_2 + \cdots + Y_n,$$

where $Y_i \sim \text{Geo}(p)$ for all $i = 1, 2, \ldots, n$. From Task C, we know that the PGF of a sum of independent random variables is the product of their individual PGFs:

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

From Task D, we know the PGF of a Geometric random variable $X \sim \text{Geo}(p)$ is given by:

$$G_{\text{Geo}}(z) = \frac{pz}{1-(1-p)z}$$

Result

Therefore, the PGF of Y is:

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

1.6 Task F

For $\alpha \in \mathbb{R}$ and $k \in \mathbb{N}$, the generalized binomial coefficient is defined as:

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

Negative Binomial Theorem

The Negative Binomial Theorem states that for $\alpha \in \mathbb{N}$ and $|x| < 1$:

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

We use the general form of the binomial series expansion for real exponents:

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

Now, consider the case when we have $(1+x)^{-n}$. Rewriting the series expansion:

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

We now need to simplify the generalized binomial coefficient for negative exponents:

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

Thus, substituting this into the series expansion, we get:

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

1.7 Task G

Theorem (Expectation from PGF)

Let $G_X(z)$ be the probability generating function (PGF) of the random variable X. The expectation of X is given by the derivative of the PGF at $z = 1$:

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

Using this theorem, we can derive the expectations for various distributions:

Let $X \sim \text{Ber}(p)$, then:

$$\begin{aligned} G_{\text{Ber}}(z) &= 1 - p + pz \\ G'_{\text{Ber}}(z) &= p \\ G'_{\text{Ber}}(1) &= p \end{aligned}$$

Thus, the mean of a Bernoulli random variable is p.

Let $X \sim \text{Bin}(n, p)$, then:

$$\begin{aligned} G_{\text{Bin}}(z) &= (1 - p + pz)^n \\ G'_{\text{Bin}}(z) &= np(1 - p + pz)^{n-1} \\ G'_{\text{Bin}}(1) &= np \end{aligned}$$

Thus, the mean of a binomial random variable is np.

Let $X \sim \text{Geo}(p)$, then:

$$\begin{aligned} G_{\text{Geo}}(z) &= \frac{pz}{1 - (1-p)z} \\ G'_{\text{Geo}}(z) &= \frac{p}{(1 - (1-p)z)^2} \\ G'_{\text{Geo}}(1) &= \frac{1}{p} \end{aligned}$$

Thus, the mean of a geometric random variable is $\frac{1}{p}$.

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

$$\begin{aligned} G_{\text{NegBin}}(z) &= \left(\frac{pz}{1 - (1-p)z}\right)^n \\ G'_{\text{NegBin}}(z) &= \frac{np^n z^{n-1}}{(1 - (1-p)z)^{n+1}} \\ G'_{\text{NegBin}}(1) &= \frac{n}{p} \end{aligned}$$

Thus, the mean of a negative binomial random variable is $\frac{n(1-p)}{p}$.

Results

1. For $X \sim \text{Ber}(p)$: $\mathbb{E}[X] = p$
2. For $X \sim \text{Bin}(n, p)$: $\mathbb{E}[X] = np$
3. For $X \sim \text{Geo}(p)$: $\mathbb{E}[X] = \frac{1}{p}$
4. For $X \sim \text{NegBin}(n, p)$: $\mathbb{E}[X] = \frac{n}{p}$

2 Normal Sampling

Instructions for running files: `python3 2c.py / 2d.py`

2.1 Task A

Theorem

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

Proof:

To find the CDF of the random variable $Y := F_X(X) \in [0, 1]$. We try to look at the $P(Y \leq y)$:

$$P(Y \leq y) = P(F_X(X) \leq y)$$

Since $F_X(X)$ is invertible (let $F_X^{-1}(x)$ represent the inverse), the required probability is the same as:

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

As $P(X \leq x) = F_X(x)$. Thus:

$$F_Y(y) = P(Y \leq y) = y$$

Taking the parital derivative with respect to y, we get:

$$\frac{\partial}{\partial y} F_Y(y) = f_Y(y) = 1$$

Since the probability density is 1, the distribution of the random variable Y is uniform(doesn't depend on the value that the random variable takes).

Result

The distribution of the random variable Y is uniform that is:

$$F_Y(y) = y$$

$$f_Y(y) = 1$$

2.2 Task B

The algorithm is as follows:

- Since we have a uniform number generator, we generate random numbers from that.
- Assuming we have access to the inverse function, we plug the randomly generated number, say r, into that. Let the output be u. This means $u = F_X^{-1}(r)$.
- The function gives a value u which satisfies $F_X(u)$ = r.
- We can then build a map, which maps u to r, which is nothing but the cumulative probability up to that point as seen above.
- Finally after sampling specified number of times, we return the map, which is nothing but $F_{\mathbb{A}}(X)$

This map which it returns is the cumulative probability up to any point. This is because at any point before the mapping, from u to $F_{\mathbb{A}}(X)(= r(\text{say}))$, we have $F_X(u) = r$. Thus, for all u which has been mapped to, we have $F_{\mathbb{A}}(u) = F_X(u)$. If we do enough sampling, we can extend this $\forall u \in \mathbb{R}$.

This is a neat way to sample from any distribution, even if the inverse is not known. Instead of plugging in the randomly generated number into the inverse CDF, we just return the min u such that $F_X(u) \geq r$. This works because probability density is always positive and thus $F_X(X)$ is an increasing function.

The sampling gives a faithful picture of the distribution behind it. This can be seen intuitively. For regions with greater probability density, $F_X(X)$ changes relatively more than X changes. Thus on average, the sampled value has a greater chance of returning the value $X = u$ in question(or a value close to that) as the region is less packed than other regions, but we still sample from a uniform distribution. (We have lesser things to choose from per unit space, thus representing a greater probability).

2.3 Task C

```python
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt

def sample(loc, scale, N=100000):
    uniform_samples = np.random.uniform(0, 1, N)
    return norm.ppf(uniform_samples, loc=loc, scale=scale)

def plot(params, N=100000):
    plt.figure(figsize=(10, 6))
    colors = ["blue", "red", "yellow", "green"]
    for (mean, var), color in zip(params, colors):
        std = np.sqrt(var)
        samples = sample(mean, std, N)
        plt.hist(
            samples,
            bins=500,
            density=True,
            alpha=0.5,
            color=color,
            label=f"μ={mean}, σ²={var}",
        )
    plt.title("Gaussian Distributions")
    plt.xlabel("x")
    plt.ylabel("p(x)")
    plt.legend()
    plt.tight_layout()
    plt.savefig("2c.png")
    plt.show()

params = [(0, 0.2), (0, 1.0), (0, 5.0), (-2, 0.5)]
plot(params)
```

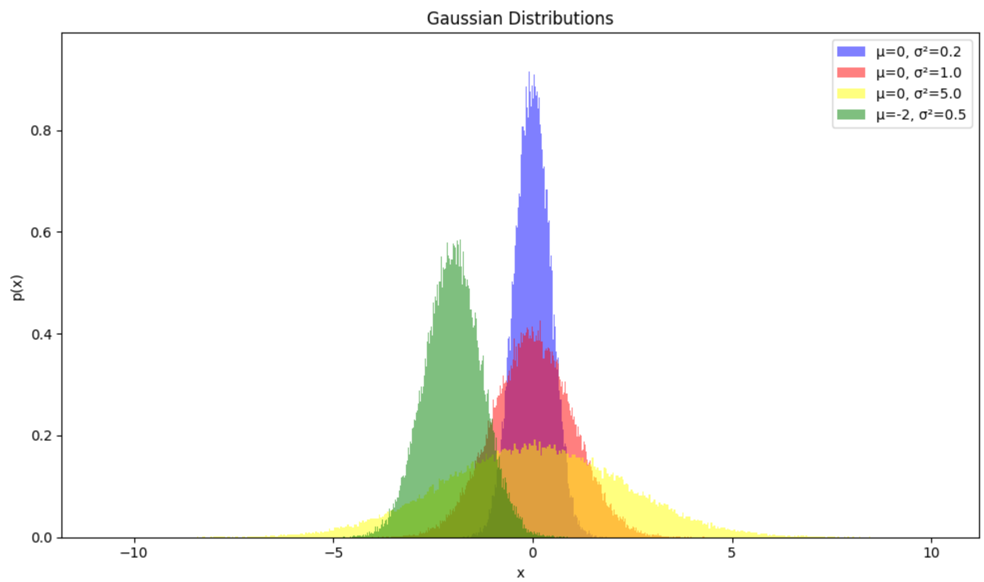

Figure 1: Gaussian Distributions with Different Parameters

2.4 Task D

```python
import numpy as np
import matplotlib.pyplot as plt


def simulate(N, h):
    # generates random left/right moves for each ball at each level
    moves = np.random.randint(0, 2, size=(N, h))
    # converts left to decrement and right to increment
    steps = 2 * moves - 1
    # computes the final position of each ball by adding up the increments/decrements
    final_pos = np.sum(steps, axis=1)
    # counts the number of balls at each final position
    pos, counts = np.unique(final_pos, return_counts=True)
    # creates a array with all possible positions
    all_pos = np.arange(-h, h + 1)
    # creates a array with zeros and puts the counts at the correct positions
    distribution = np.zeros(2 * h + 1, dtype=float)
    # calculates the indices for each position
    indices = pos + h
    # updates the normalized counts at the correct positions
    distribution[indices] = counts / N

    return all_pos, distribution


def plot(positions, distribution, h, N, filename):
    plt.figure(figsize=(12, 6))
    plt.bar(positions, distribution, align="center", alpha=0.5)
    plt.title(f"Galton Board Simulation (h={h}, N={N})")
    plt.xlabel("Pocket")
    plt.ylabel("Normalized count")
    plt.tight_layout()
    plt.savefig(filename)
    plt.close()


N = 100000
```

```python

for i, h in enumerate([10, 50, 100], start=1):
    positions, distribution = simulate(N, h)
    plot(positions, distribution, h, N, f"2d{i}.png")
```

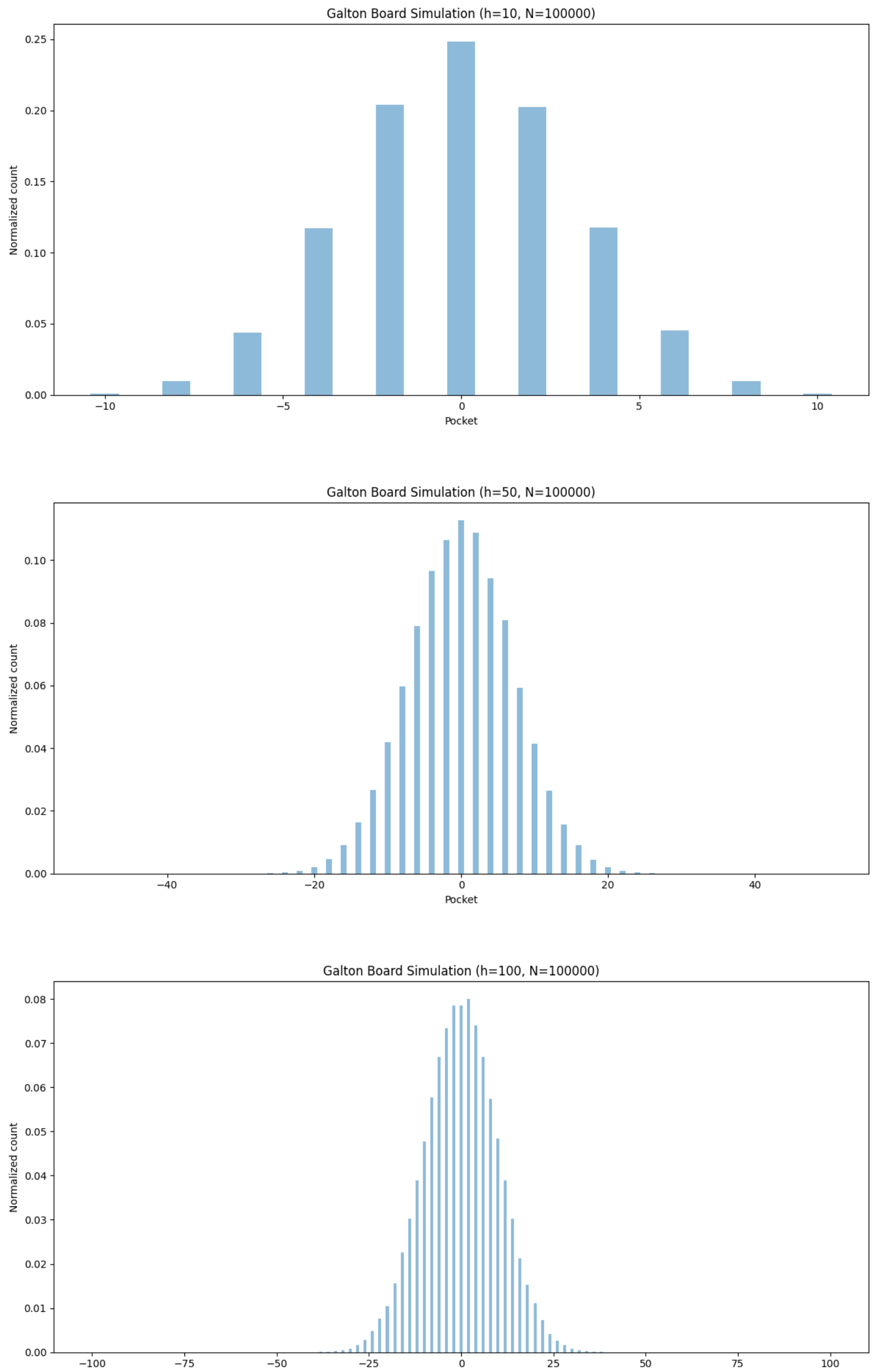

Figure 2: Galton Board Simulations.

2.5 Task E

Consider a Galton board of depth $h = 2k$ where k is a positive integer. Let random variable $X \in \{-h, -h+2, \ldots, 0, 2, \ldots, h-2, h\}$ describe the pocket in which a ball finally lands.

To compute $P_h[X = 2i]$ for $i \in \{-k, -k+1, \ldots, k-1, k\}$:

The ball must make $k+i$ right moves and $k-i$ left moves out of $2k$ total moves as $(k+i)+(k-i) = 2i$. And this follows a binomial distribution with $n = 2k$ attempts and with probability $p = \frac{1}{2}$ for each direction.

The probability mass function for a binomial distribution is:

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

Substituting our values and simplifying:

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

Result

$P_h[X = 2i]$ for $i \in \{-k, -k+1, \ldots, k-1, k\}$:

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

Theorem

For large k and $i \ll \sqrt{h}$:

$$P_k[X = i] \approx \frac{1}{\sqrt{\pi k}} e^{-\frac{i^2}{4k}} = \mathcal{N}\left(\mu = 0, \sigma^2 = k\right)(i)$$

where $k = \frac{h}{2}$ and i is even.

Proof:

We start with the binomial coefficient:

$$\binom{h}{h/2 + i/2} = \frac{h!}{(h/2 + i/2)!(h/2 - i/2)!}$$

Applying Stirling's approximation:

$$\binom{h}{h/2 + i/2} \approx \frac{\sqrt{2\pi h}(h/e)^h}{\sqrt{2\pi(h/2 + i/2)}((h/2 + i/2)/e)^{h/2+i/2} \cdot \sqrt{2\pi(h/2 - i/2)}((h/2 - i/2)/e)^{h/2-i/2}}$$

$$\binom{h}{h/2 + i/2} \approx \frac{\sqrt{h}}{\sqrt{2\pi(h^2/4 - i^2/4)}} \cdot \frac{h^h}{(h/2 + i/2)^{h/2+i/2}(h/2 - i/2)^{h/2-i/2}}$$

Taking the logarithm:

$$\ln\binom{h}{h/2 + i/2} \approx \frac{1}{2}\ln\left(\frac{h}{2\pi(h^2/4 - i^2/4)}\right) + h\ln h - (h/2 + i/2)\ln(h/2 + i/2) - (h/2 - i/2)\ln(h/2 - i/2)$$

Next, applying the Taylor expansion for $\ln(1+x)$ for small x:

$$\ln(h/2 \pm i/2) = \ln(h/2) \pm \frac{i}{h} - \frac{1}{2}\frac{i^2}{h^2} + O\left(\frac{i^3}{h^3}\right)$$

Substitute the Taylor expansions into the expression:

$$\begin{aligned}\ln\binom{h}{h/2+i/2} \approx \frac{1}{2}\ln\left(\frac{h}{2\pi(h^2/4-i^2/4)}\right) + h\ln h \\ -(h/2+i/2)\left(\ln(h/2)+\frac{i}{h}+\frac{1}{2}\frac{i^2}{h^2}\right) \\ -(h/2-i/2)\left(\ln(h/2)-\frac{i}{h}+\frac{1}{2}\frac{i^2}{h^2}\right)\end{aligned}$$

After expanding and combining like terms:

$$\begin{aligned}\ln\binom{h}{h/2+i/2} &\approx \frac{1}{2}\ln\left(\frac{h}{2\pi(h^2/4-i^2/4)}\right) + h\ln h - h\ln(h/2) - \frac{i^2}{2h} \\ &= \frac{1}{2}\ln\left(\frac{2}{(\pi h)(1-i^2/h^2)}\right) + h\ln h - h\ln(h/2) - \frac{i^2}{2h}\end{aligned}$$

Approximating the logarithmic term and simplifying:

$$\begin{aligned}\ln\binom{h}{h/2+i/2} &\approx \frac{1}{2}\ln\left(\frac{2}{(\pi h)(1-i^2/h^2)}\right) + h\ln h - h\ln(h/2) - \frac{i^2}{2h} \\ &= \frac{1}{2}\ln\frac{2}{\pi h} - \frac{1}{2}\ln(1-i^2/h^2) + h\ln h - h\ln(h/2) - \frac{i^2}{2h} \\ &= \frac{1}{2}\ln\frac{2}{\pi h} + h\ln h - h\ln(h/2) + \frac{1}{2}\frac{i^2}{h^2} - \frac{i^2}{2h} \\ &\approx \frac{1}{2}\ln\frac{2}{\pi h} + h\ln h - h(\ln h - \ln 2) - \frac{i^2}{2h} \\ &= \frac{1}{2}\ln\frac{2}{\pi h} + h\ln 2 - \frac{i^2}{2h}\end{aligned}$$

Exponentiating both sides:

$$\binom{h}{h/2+i/2} \approx \frac{\sqrt{2}}{\sqrt{\pi h}}2^h e^{-i^2/2h}$$

Finally, substituting this into the probability mass function:

$$\begin{aligned}P_h[X=i] &\approx \frac{\sqrt{2}}{\sqrt{\pi h}}e^{-i^2/2h} \\ P_k[X=i] &= \frac{1}{\sqrt{\pi k}}e^{-i^2/4k}\end{aligned}$$

3 Fitting Data

Instructions for running file: Run it in Jupyter(it is submitted as a .ipynb notebook)

3.1 Task A

We have the i^{th} moment $\hat{\mu}_i = E[X^i]$. This is nothing but

$$\hat{\mu}_i = \frac{x_1^i + x_2^i + \cdots + x_n^i}{n}$$

The first moment is the mean and the second moment is the mean of the squares. For this, we use the `numpy.mean()` and `numpy.square()` function available in numpy.

3.2 Task B

We use the histogram function available in `matplotlib.pyplot` to plot the histogram. The normal distribution parameter μ seems to be around 6. Since there is data unequally distributed, the parameter would lie somewhere to the right of 6.

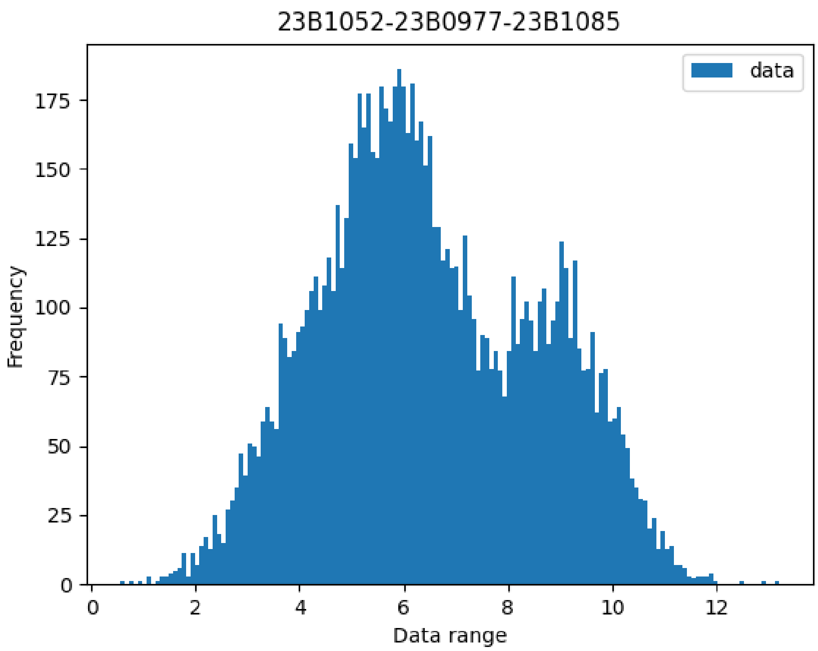

Figure 3: Histogram of the data

3.3 Task C

3.3.1 Subpart 1

The probability mass function of the binomial distribution is given by:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

where:

- n is the number of trials,
- k is the number of successes,
- p is the probability of success on a single trial.

The binomial coefficient is:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Using the below relation which can be verified by expanding the binomial co-efficient:

$$k \cdot \binom{n}{k} = n \cdot \binom{n-1}{k-1}$$

The first moment μ_1^{Bin} is given by:

$$\begin{aligned}
\mu_1^{\text{Bin}} &= \sum_{k=0}^{n} P(X=k)\cdot k \\
&= \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot k \\
&= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&= np \cdot (p+1-p)^n \\
&= np
\end{aligned} \tag{1}$$

For the second moment μ_2^{Bin}, we have:

$$\begin{aligned}
\mu_2^{\text{Bin}} &= \sum_{k=0}^{n} P(X=k)\cdot k^2 \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot k^2 \\
&= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} p^k (1-p)^{n-k} \cdot (1+k-1) \\
&= np + n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} \cdot (k-1) \\
&= np + n \sum_{k=2}^{n} (n-1) \cdot \binom{n-2}{k-2} p^k (1-p)^{n-k} \\
&= np + n(n-1)p^2 \sum_{k=0}^{n-2} \cdot \binom{n-2}{k} p^k (1-p)^{n-2-k} \\
&= np + n(n-1)\, p^2
\end{aligned} \tag{2}$$

We use the result we got from μ_1^{Bin} when splitting k into (1 + k - 1).

Result

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

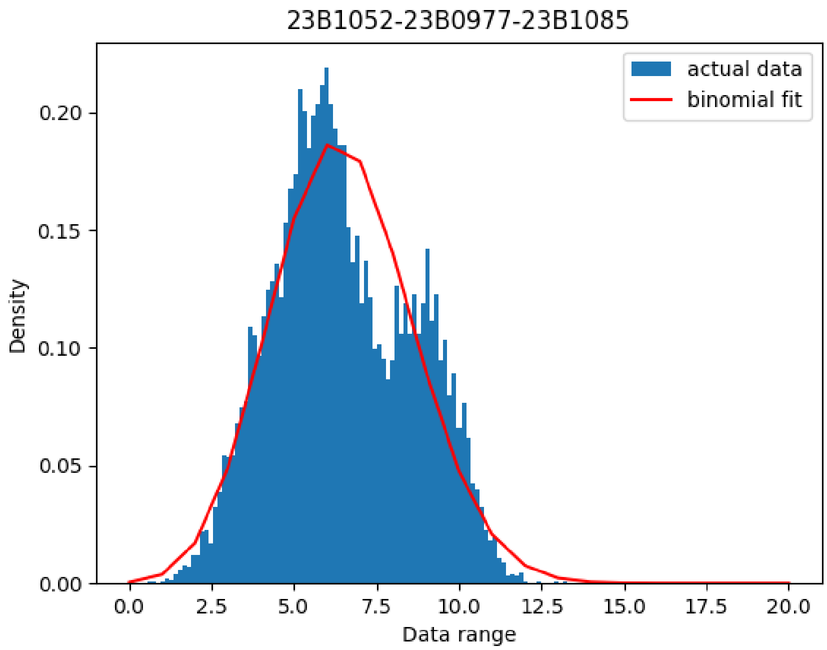

Figure 4: Fitting using Binomial

3.4 Task D

3.4.1 Subpart 1

$$f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta} \tag{3}$$

where:

- $x > 0$ is the variable,
- $k > 0$ is the shape parameter,
- $\theta > 0$ is the scale parameter,
- $\Gamma(\alpha)$ is the Gamma function, which is defined as:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt \tag{4}$$

To calculate the first moment μ_1^{Gamma} of the gamma distribution, we use:

$$\begin{aligned}
\mu_1^{\text{Gamma}} &= \int_0^\infty x \, f(x; k, \theta) \, dx \\
&= \int_0^\infty \frac{1}{\theta^k \Gamma(k)} x^k e^{-x/\theta} \, dx \\
&= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^k e^{-x/\theta} \, dx
\end{aligned}$$

Substitute $x/\theta = t$ (so $x = \theta t$ and $dx = \theta \, dt$):

$$
\begin{aligned}
\mu_1^{\text{Gamma}} &= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty (\theta t)^k e^{-t} \cdot \theta \, dt \\
&= \frac{1}{\theta^k \Gamma(k)} \cdot \theta^k \cdot \theta \int_0^\infty t^k e^{-t} \, dt \\
&= \frac{\theta}{\Gamma(k)} \int_0^\infty t^k e^{-t} \, dt
\end{aligned}
$$

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

$$
\begin{aligned}
\mu_1^{\text{Gamma}} &= \frac{\theta}{\Gamma(k)} \cdot \Gamma(k+1) \\
&= \theta \cdot \frac{\Gamma(k+1)}{\Gamma(k)}
\end{aligned}
$$

Using the property of the Gamma function:

$$
\frac{\Gamma(k+1)}{\Gamma(k)} = k
$$

Thus:

$$
\mu_1^{\text{Gamma}} = \theta \cdot k
$$

Similarly, to calculate the second moment μ_2^{Gamma} of the gamma distribution, we use:

$$
\begin{aligned}
\mu_2^{\text{Gamma}} &= \int_0^\infty x^2 \, f(x; k, \theta) \, dx \\
&= \int_0^\infty \frac{1}{\theta^k \Gamma(k)} x^{k+1} e^{-x/\theta} \, dx \\
&= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^{k+1} e^{-x/\theta} \, dx
\end{aligned}
$$

We do a similar substitution and end up with:

$$
\begin{aligned}
\mu_2^{\text{Gamma}} &= \frac{1}{\theta^k \Gamma(k)} \cdot \theta^{k+1} \cdot \theta \int_0^\infty t^{k+1} e^{-t} \, dt \\
&= \frac{\theta^2}{\Gamma(k)} \int_0^\infty t^{k+1} e^{-t} \, dt
\end{aligned} \tag{5}
$$

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

$$
\begin{aligned}
\mu_1^{\text{Gamma}} &= \frac{\theta}{\Gamma(k)} \cdot \Gamma(k+2) \\
&= \theta \cdot \frac{\Gamma(k+2)}{\Gamma(k)}
\end{aligned}
$$

Using the property mentioned above, we have:

$$
\begin{aligned}
\frac{\Gamma(k+2)}{\Gamma(k)} &= \frac{\Gamma(k+2)}{\Gamma(k+1)} \cdot \frac{\Gamma(k+1)}{\Gamma(k)} \\
&= (k+1) \cdot k
\end{aligned}
$$

Thus:

$$\mu_2^{\text{Gamma}} = \theta^2 \cdot k(k+1)$$

Result

$$\mu_1^{\text{Gamma}} = \theta \cdot k$$
$$\mu_2^{\text{Gamma}} = \theta^2 \cdot k(k+1)$$

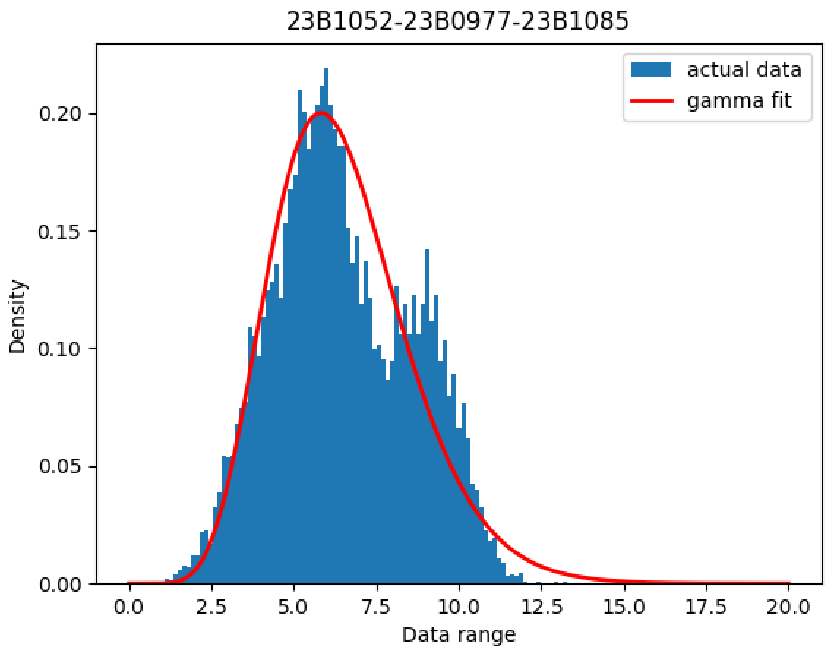

Figure 5: Fitting using Gamma distribution

3.5 Task E

For calculating the likelihood, we send in the numpy array into `numpy.mean(numpy.log())`. For Binomial, we would have to use `numpy.round()` as well. On comparing, we find the **Binomial Distribution** to be a better fit.

3.6 Task F

Again on comparing, **Binomial Distribution** seems to do better at approximating the distribution than the Gausian Mixture Model.

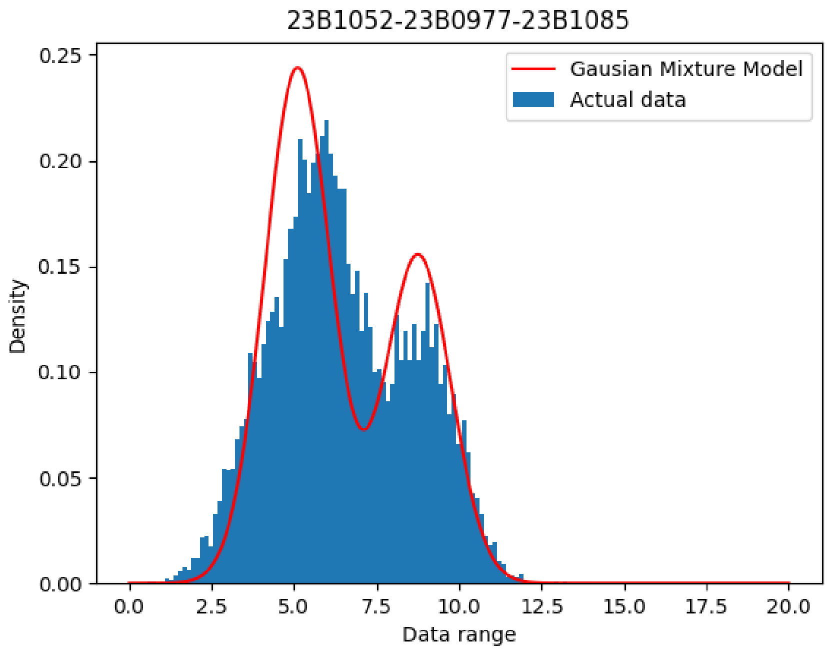

Figure 6: Fitting using GMM using 4 parameters

3.7 Ungraded Bonus :)

Since the distribution has two modes, it could've been bimodal distribution. Since we restricted ourselves to a variance of 1 in part F, if we send in both sigma's as parameters, perhaps we would get a better fit. In order to solve for 6 parameters, we further calculate the 5th and 6th moments. On solving with `fsolve` however did not return any better results. So, it was fitted with log likelihood convergence. The results gotten were as follows: $\mu_1 = 5.61$, $\sigma_1 = 1.52$, $p_1 = 0.75$, $\mu_2 = 9.21$, $\sigma_2 = 0.94$, and $p_2 = 0.25$.

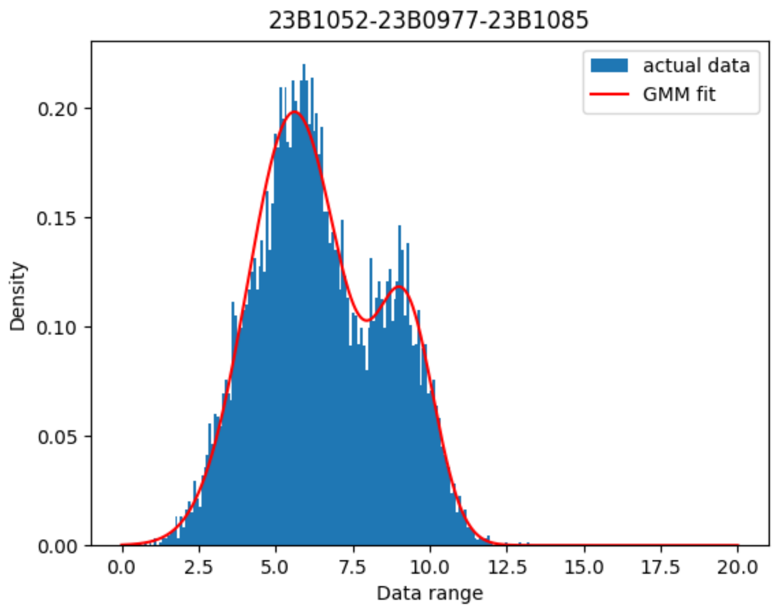

Figure 7: GMM fit with 6 parameters

4 Quality in Inequalities

4.1 Task A

Let a subset sum represent the sum of terms which are a subset of the set of all weighted terms. Intuitively, since expectation is just the weighted sum of the values which the random variable takes, where the weights are the probabilities themselves, we can't have any subset sum of weighted terms to be greater than the expectation since each probability associated with a value is positive and X itself takes only positive values. So, all terms are positive. So the expectation must at least be as much as any subset sum.

Now, we'll turn to a more rigorous proof.

Theorem

(Markov's Inequality). *Let X be any non-negative random variable and $a > 0$,*

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

Proof:

Let X be a non-negative random variable with PDF $f(X)$.

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x f(x)\, dx \\ &= \int_0^a x f(x)\, dx + \int_a^\infty x f(x)\, dx \end{aligned} \tag{6}$$

In the first integral, since $X = x > 0$, and $f(x) > 0$, we have the integral also be greater than 0. Thus we can say:

$$\mathbb{E}[X] \geq \int_a^\infty x f(x)\, dx \tag{7}$$

In the interval $[a, \infty)$:

$$xf(x) \geq af(x)$$

Thus:

$$\int_a^\infty xf(x)\,dx \geq \int_a^\infty af(x)\,dx$$

Combining this result and equation (2), we get:

$$\mathbb{E}[X] \geq a \int_a^\infty f(x)\,dx$$

Using

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

We have:

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

Thus:

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

4.2 Task B

Let $Y = X - \mu$. Here, Y is a Random Variable with mean 0 and variance σ^2.
We know that adding a constant on both sides of the inequality $Y > \tau$ doesn't make a difference.

$$P[Y \geq \tau] = P[Y + b \geq \tau + b]$$

We also know that,

$$P(X \geq k) \leq P(X^2 \geq k^2)$$

Using the above inequality, we get

$$P[Y + b \geq \tau + b] \leq P[(Y + b)^2 \geq (\tau + b)^2]$$

Using Markov Inequality, we get:

$$\begin{aligned} P[(Y + b)^2 \geq (\tau + b)^2] &\leq \frac{E((Y + b)^2)}{(\tau + b)^2} \\ &\leq \frac{E(Y^2) + 2E(Y)b + b^2}{(\tau + b)^2} \\ &\leq \frac{\sigma^2 + b^2}{(\tau + b)^2} \quad \text{Since E(X) = Mean = 0} \end{aligned}$$

Since this is true for any b, we will find the b that can minimize the RHS to get the best bound. That can be done differentiating the RHS with respect to b and equating it to zero. By doing so, we get $b = \sigma^2/\tau$. Substituting it in the equation, we get

$$\begin{aligned} P[(Y + b)^2 \geq (\tau + b)^2] &\leq \frac{\sigma^2 + (\sigma^2/\tau)^2}{(\tau + (\sigma^2/\tau))^2} \\ &\leq \frac{\sigma^2}{\sigma^2 + \tau^2} \end{aligned}$$

Now combining the inequalities, we get:

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

Thus:

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

4.3 Task C

The Moment Generating Function(MGF) of a Random Variable X is defined as:

$$M_X(t) = E[e^{tX}]$$

where $t \in R$

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

By Markov's inequality, for a non-negative Random Variable X and a ≥ 0:

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

Now, let $Y = e^{tX}$, which is always non-negative since $e^{tx} \geq 0$. Choose $a = e^{tx}$. The inequality gives us:

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

Since $t > 0$,

$$P[e^{tX} \geq e^{tx}] = P[tX \geq tx] = P[X \geq x]$$

Using the above equality, we get:

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

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

We follow the same proof technique as the previous part. But,

$$P(tX \geq tx) = P(X \leq x)$$

because t is a negative number.
From previous part,

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

Using the earlier inequality,

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

4.4 Task D

4.4.1 SubPart-1

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

Since Expectation operator is linear, we get:

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

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

4.4.2 SubPart-2

Let us use the exponential Markov Inequality we have proved in task C. From the result of exponential Markov inequality, we get:

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

Now, let's compute$M_Y(t)$

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E(e^{t(\sum_{i=1}^{n} X_i)}) \\ &= E(\prod_{i=1}^{n} e^{tX_i}) \end{aligned}$$

Since the X_i's are independent. the expectation of products is the product of expectations.

$$M_Y(t) = \prod_{i=1}^{n} E(e^{tX_i})$$

Now, let's compute $E(e^{tX_i})$ for a Bernoulli Random Variable.

$$\begin{aligned} E(e^{tX_i}) &= p_i \cdot e^{t \cdot i} + (1-p_i)e^{t \cdot 0} \\ &= 1 + p_i(e^t - 1) \end{aligned}$$

We know, for $x \geq 0$:

$$1 + x \leq e^x$$

Using the above inequality,

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

Now, let's use the above inequality in $M_Y(t)$ expression we have obtained.

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} E(e^{tX_i}) \\ &\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} \\ &\leq e^{(\sum_{i=1}^{n} p_i)(e^t-1)} \\ &\leq e^{(e^t-1)\mu} \end{aligned}$$

Now let's use the above inequality in exponential Markov Inequality.

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

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

4.4.3 SubPart-3

To find the best bound, we find the value of t that minimizes the RHS. To evaluate that, we take the derivative of the exponent of e and equate i to 0. By doing so, we get:

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

will minimize the RHS.
Substituting $t = ln(1+\delta)$ back into the inequality and using another inequality $(1+\delta)ln(1+\delta) \geq \frac{\delta^2}{2+\delta} \quad \forall \delta > 0$, we get:

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

4.5 Task E

Let's take the random variable $Y = \sum_{i=1}^{n} X_i$. The sample mean A_n in terms of Y as:

$$A_n = \frac{Y}{n}.$$

The Chernoff bound gives us:

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

where $\delta > 0$ and $\mu n = E[Y]$.
Now, we set $Y = nA_n$, and the event $A_n \geq \mu + \epsilon$ corresponds to:

$$Y = nA_n \geq n(\mu + \epsilon) = (1+\delta)\mu n,$$

where $\delta = \frac{\epsilon}{\mu}$. Thus:

$$P(A_n \geq \mu + \epsilon) = P\left(Y \geq n(\mu + \epsilon)\right) \leq e^{-\frac{\mu n (\epsilon/\mu)^2}{2+\epsilon/\mu}} = e^{-\frac{n\epsilon^2}{\mu(2+\epsilon/\mu)}}.$$

Similarly, for the event $A_n \leq \mu - \epsilon$:

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

Thus:

$$P(|A_n - \mu| > \epsilon) = P(A_n \geq \mu + \epsilon) + P(A_n \leq \mu - \epsilon) \leq 2e^{-\frac{n\epsilon^2}{\mu(2+\epsilon/\mu)}}.$$

Taking the Limit $n \to \infty$, we can notice that the expression $2e^{-\frac{n\epsilon^2}{\mu(2+\epsilon/\mu)}}$ tends to 0 as $n \to \infty$ because the exponential term $e^{-\frac{n\epsilon^2}{\mu(2+\epsilon/\mu)}}$ decreases to 0 extremely rapidly. Therefore, we have:

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

5 A Pretty "Normal" Mixture

5.1 Task A

Understanding the Gaussian Mixture Model (GMM)

A Gaussian Mixture Model (GMM) with k components is a probabilistic model assuming data is generated from a mixture of k Gaussian distributions, each with mean μ_i and variance σ_i^2, where $i = 1, 2, \ldots, k$. The PDF of the GMM is given by:

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

where:

- p_i is the weight (mixing coefficient) of the i-th Gaussian distribution, with $\sum_{i=1}^{k} p_i = 1$ and $p_i \geq 0$.
- $f_{\{}X_i\}(u, i)$ is the PDF of the i-th Gaussian distribution, defined as:

$$f_{X_i}(u) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \, e^{-\frac{(u-\mu_i)^2}{2\sigma_i^2}}.$$

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

1. **Calculate the PDF of $\mathcal{A}$ (Output of the Algorithm):**
   The random variable $\mathcal{A}$ is generated by:
   - First selecting an index i with probability p_i.
   - Then sampling from the Gaussian distribution $\mathcal{N}(\mu_i, \sigma_i^2)$.

   The PDF of $\mathcal{A}$ is therefore the weighted sum of the PDFs of each Gaussian component:
   $$f_{\mathcal{A}}(u) = \sum_{i=1}^{k} p_i \, f_{X_i}(u).$$
   Here, p_i is the probability of selecting the i-th Gaussian, and $f_{X_i}(u)$ is the PDF of the selected Gaussian.

2. **Compare with the PDF of the GMM:**
   The PDF of the GMM variable X is defined as:
   $$f_X(u) = \sum_{i=1}^{k} p_i \, f_{X_i}(u).$$
   Notice that $f_{\mathcal{A}}(u)$ has the same form as $f_X(u)$. Since both $f_{\mathcal{A}}(u)$ and $f_X(u)$ are represented by the same weighted sum of Gaussian components with the same weights p_i, we conclude:
   $$f_{\mathcal{A}}(u) = f_X(u), \quad \text{for all } u \in \mathbb{R}.$$

5.2 Task B

5.2.1 E(X)

Using the above algorithm for sampling, we get

$$f_{\mathcal{A}}(u) = f_X(u) \quad \forall u \in \mathbb{R}$$

And,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_{\mathcal{A}}(x)\, dx \\ &= \int_{-\infty}^{\infty} x \left(\sum_{i=1}^{K} p_i f_{X_i}(x) \right) dx \\ &= \sum_{i=1}^{K} p_i \left(\int_{-\infty}^{\infty} x f_{X_i}(x)\, dx \right) \end{aligned} \tag{8}$$

It is given that each $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ is a Gaussian Random Variable. For a Gaussian random variable, the Expected Value is it's mean.

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

Using the above equation in (8),

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

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

5.2.2 Var(X)

$$\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
Var(X) &= E(X^2) - 2\mu E(X) + \mu^2 \qquad (9)
\end{aligned}$$

For the GMM sampled using the above algorithm,

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

Now, let's compute $E(X^2)$.

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f_{\mathcal{A}}(x)\,dx \\
&= \int_{-\infty}^{\infty} x^2 \left(\sum_{i=1}^{K} p_i f_{X_i}(x)\right) dx \\
&= \sum_{i=1}^{K} p_i \left(\int_{-\infty}^{\infty} x^2 f_{X_i}(x)\,dx\right) \qquad (10)
\end{aligned}$$

It is given that each $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ is a Gaussian Random Variable. For a Gaussian random variable,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \sigma^2 + \mu^2 \quad \text{(derived in class using MGF)}$$

Using the above equation in (10), we get

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

For the above sampling method,

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

Now, let's use the results we have got in (9).

$$\begin{aligned}
Var(X) &= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2
\end{aligned}$$

Substituting $E(X^2)$ and μ, we get

$$Var(x) = \sum_{i=1}^{K} p_i(\mu_i^2 + \sigma_i^2) - \left(\sum_{i=1}^{K} p_i \mu\right)^2$$

5.2.3 MGF $M_X(t)$ of X

We know,

$$\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_{-\infty}^{\infty} e^{tx} f_{\mathcal{A}}(x)\, dx \\
&= \int_{-\infty}^{\infty} e^{tx} \left(\sum_{i=1}^{K} p_i f_{X_i}(x) \right) dx \\
&= \sum_{i=1}^{K} p_i \left(\int_{-\infty}^{\infty} e^{tx} f_{X_i}(x)\, dx \right) \qquad (11)
\end{aligned}$$

It is given that each $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ is a Gaussian Random Variable. For a Gaussian random variable,

$$M_{X_i}(t) = \int_{-\infty}^{\infty} e^{tx} f_{X_i}(x)\, dx$$

where M_{X_i} is it's MGF. Now, Substituting the above equation in (11), we get

$$M_X(t) = \sum_{i=1}^{K} p_i M_{X_i}(t) \qquad (12)$$

For a Gaussian Random Variable,

$$M_{X_i}(t) = e^{t\mu_i + 1/2 t^2 \sigma_i^2}$$

Substituting the above equation in (12), we get

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

5.3 Task C

5.3.1 E(Z)

The expected value of a weighted sum of random variables is the weighted sum of their expected values. Since each $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$:

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

Substitute the expectation $E[X_i] = \mu_i$:

$$E[Z] = \sum_{i=1}^{k} p_i \mu_i.$$

5.3.2 Var(Z)

The variance of a weighted sum of independent random variables is the sum of the variances of the individual random variables, each multiplied by the square of its corresponding weight. Since X_i are independent:

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

Substitute the variance $\mathrm{Var}(X_i) = \sigma_i^2$:

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

5.3.3 PDF $f_Z(u)$ of Z

Since Z is a linear combination of independent Gaussian random variables, Z itself is also Gaussian. The parameters of the normal distribution for Z are given by the expressions for the mean and variance we derived above:

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

Thus, the PDF $f_Z(u)$ of Z is:

$$f_Z(u) = \frac{1}{\sqrt{2\pi \mathrm{Var}(Z)}} \exp\left(-\frac{(u - E[Z])^2}{2\mathrm{Var}(Z)}\right),$$

where:

$$E[Z] = \sum_{i=1}^{k} p_i \mu_i,$$

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

5.3.4 MGF $M_Z(t)$ of Z

The MGF of a random variable Z is defined as:

$$M_Z(t) = E\left[e^{tZ}\right].$$

Since $Z = \sum_{i=1}^{k} p_i X_i$ and X_i are independent:

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

Because the X_i are independent, we can factor the expectation:

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

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

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

Substitute t with tp_i for each X_i:

$$E\left[e^{tp_i X_i}\right] = \exp\left(p_i t \mu_i + \frac{(p_i t)^2 \sigma_i^2}{2}\right).$$

Thus, the MGF of Z is:

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

Simplifying, we get:

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

Combine the terms:

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

5.3.5 Comparison of X and Z

No. X and Z do not have the same properties. And our thought of X being the same is weighted sum of Gaussians is false. It can be seen from the expressions of Var, PDF and MGF of X and Z.

5.3.6 Distribution of Z

From the results obtained, we conclude that Z follows a **Normal (Gaussian) distribution** with:

- **Mean:** $\sum_{i=1}^{k} p_i\mu_i$
- **Variance:** $\sum_{i=1}^{k} p_i^2\sigma_i^2$

Therefore:

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

5.4 Task D (B)

Before proving the theorem, let us see what do MGF and PMF mean for a Random Variable X that is finite and discrete.

- **PMF:** $p_X(x) = P(X = x)$
- **MGF:** $\phi_X(t) = E[e^{tX}] = \sum_x e^{tx} p_X(x)$

where summation is over all possible values x that X can take.

PMF uniquely determines MGF

Given the PMF, we can construct the MGF directly from the definition of MGF mentioned above. And by the definition, substituting x and $p_X(x)$ would result in a $\phi_X(t)$ and that is the unique MGF of the given PMF.

MGF uniquely determines PMF

Let's suppose X takes values $x_1, x_2, \ldots, x_n$ with corresponding probabilities $p_1, p_2, \ldots, p_n$. The MGF can be expressed as:

$$\phi_X(t) = \sum_{i=1}^{n} e^{tx_i} p_i$$

Using the power series representation of e^{tx_i}, we get:

$$\phi_X(t) = \sum_{i=1}^{n} p_i \left(\sum_{k=0}^{\infty} \frac{(tx_i)^k}{k!} \right)$$
$$= \sum_{i=1}^{n} p_i \left(\sum_{k=0}^{\infty} \frac{t^k x_i^k}{k!} \right)$$

Now differentiating the above equation 'm' times and evaluating at t=0, we get:

$$\frac{d^m \phi_X(t)}{dt^m}\bigg| t = 0 = \sum i = 1^n x_i^m p_i$$

For different values of m, we get different linear equations on p_i. Solving the linear equations, we can obtain the values of p_i's that uniquely determine a PMF.

Comments on X and Z

We can see that the MGF of X and Z are clearly different as derived from the above tasks. And using the above theorem, we can say that the PDF's of X and Z are not the same. This result also follows with what we have derived in the above tasks. One simple logical reason is that the Sampling algorithm is dependent on Choosing a Gaussian distribution and then sampling a value from it. But in case of Z, it is given that it is a weighted sum of Gaussian Random Variables and in that case, we know that Z is also a Gaussian Random Variable and it's PDF is determined using it's E(Z) and Var(Z), which are different from E(X) and Var(X).