ASSIGNMENT 2

31 Aug - 13 Sept, 2024

SHAIK SAHIL CHANDA
23B0943

PALURI GNANA KOUSHIK REDDY
23B1000

DEVANGAM KISHAN TEJA
23B1061

CS215 Data Analysis and Interpretation

Contents

1 Mathemagic

1.1 Task A

Given, $X \sim \text{Ber}(p)$

$$P(X = x) = p^x(1-p)^x$$

Since Bernoulli random variable can only take values in $\{0, 1\}$, PGF is

$$G_{\text{Ber}}(z) = P(X = 0)z^0 + P(X = 1)z^1$$

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

1.2 Task B

Now let, $X \sim \text{Bin}(n, p)$

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

$$G_{\text{Bin}}(z) = \sum_{x=0}^{n} P(X = x)z^x = \sum_{x=0}^{n} \binom{n}{x} p^x(1-p)^{n-x} z^x$$

$$G_{\text{Bin}}(z) = \sum_{x=0}^{n} \binom{n}{x} (pz)^x(1-p)^{n-x}$$

From the binomial expansion,

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x$$

we have $a = (1-p)$ and $b = pz$,

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

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

1.3 Task C

Firstly let us state and prove a result which is required to solve the question.

Theorem: When X and Y are independent random variables, $E(XY) = E(X).E(Y)$.
Proof:
For discrete random variables X and Y,

$$E(XY) = \sum_i \sum_j x_i y_j f_{XY}(x_i, y_j)$$

$$E(XY) = \sum_i \sum_j x_i y_j f_X(x_i) f_Y(y_j)$$

And thus, we have:

$$E(XY) = \left(\sum_i x_i f_X(x_i)\right)\left(\sum_j y_j f_Y(y_j)\right) = E(X)E(Y)$$

For the given random variable $X = X_1 + X_2 + ... + X_k$, PGF is

$$G_\Sigma(z) = E(z^X) = E(z^{X_1+X_2+...+X_k})$$

$$G_\Sigma(z) = E(z^{X_1}z^{X_2}...z^{X_k})$$

If $X_1, X_2, .., X_k$ are independent random varibles, so are $z^{X_1}, z^{X_2}, ..., z^{X_k}$. So from the above result we can write,

$$G_\Sigma(z) = E(z^{X_1})E(z^{X_2})...E(z^{X_k})$$

Since each X_i is having the same distribution with G as their common PGF, we have

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

1.4 Task D

Now given that, $X \sim \text{Geo}(p)$

$$P(X = x) = (1-p)^{x-1}p$$

and PGF is

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

$$G_{\text{Geo}}(z) = \left(\frac{p}{1-p}\right)\sum_{x=1}^{\infty}(z(1-p))^x$$

by summation of infinite Geometric Series, we get

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

1.5 Task E

Since Negative Binomial is sum of n independent Geometric random variables, from the result of Task C we can write the PGF for $Y \sim \text{NegBin}(n,p)$ as

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

Because now we have PGF's for both Binomial and Negative Binomial the given result can be proved easily.

Consider the R.H.S.

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

1.6 Task F

Consider the PGF for Negative Binomial distribution,

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

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

Taking $k = n + r$ we get,

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

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

Further taking $x = -z(1-p)$, we get

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

From given definition we can write

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

Thus, we get:

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

which is the Negative Binomial theorem.

1.7 Task G

$$G(z) = \sum_{x=0}^{\infty} P(X = x) z^x$$

$$G'(z) = \sum_{x=0}^{\infty} P(X = x) x z^{x-1}$$

$$G'(1) = \left(\sum_{x=0}^{\infty} x P(X = x) \right) = E(X)$$

Bernoulli:$X \sim \text{Ber}(p)$

$$E(X) = G'_{\text{Ber}}(1) = p$$

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

$$E(X) = G'_{\text{Bin}}(1) = n(1 - p + p)^{n-1} p = np$$

Geometric: $X \sim \text{Geo}(p)$

$$E(X) = G'_{\text{Geo}}(1) = \frac{(1 - 1(1 - p))p - p(-(1 - p))}{(1 - (1 - p))^2} = \frac{1}{p}$$

Negative Binomial: $X \sim \text{NegBin}(n, p)$

$$E(X) = G'_{\text{NegBin}}(1) = n \left(\frac{p}{1 - 1 + p} \right)^{n-1} \left(\frac{1}{p} \right) = \frac{n}{p}$$

2 Normal Sampling

2.1 Task A

Given F_X is the CDF for continuous random variable X,

$$F_X(x) = Pr[X \leq x]$$

Let F_Y be the CDF of $Y = F_X(X)$. Then, for any $y \in [0, 1]$, we have:

$$F_Y(y) = \Pr[Y \leq y] = \Pr[F(X) \leq y]$$

assuming F_X is invertible,

$$= \Pr[X \leq F^{-1}(y)] = F(F^{-1}(y)) = y$$

$$F_Y(y) = y \quad \forall y \in [0, 1]$$

i.e, Y is uniformly distributed in $[0, 1]$.

2.2 Task B

Given Y is uniform random variable over $[0, 1]$ and X is some distribution we want to mimic through an algorithm $\mathcal{A}$ over Y, i.e, the outputs of $\mathcal{A}(Y) : [0, 1] \to R$ has the same distribution as the random variable X.

The algorithm is simple, it is just the inverse of the CDF of X. Let $F_X : R \to [0,1]$ be the CDF of X,

$$\mathcal{A} = F_X^{-1} : [0,1] \to R$$

We will show that the output of $\mathcal{A}$ has the same CDF as X, and hence the same distribution.

The CDF for $\mathcal{A}$ is,

$$\begin{aligned} F_{\mathcal{A}}(u) &= Pr[\mathcal{A}(Y) \leq u] \\ &= Pr[F_X^{-1}(Y) \leq u] \\ &= Pr[Y \leq F_X(u)] \end{aligned}$$

Since Y is a uniform distribution, $Pr[Y \leq y] = y$. And thus,

$$F_{\mathcal{A}}(u) = Pr[Y \leq F_X(u)] = F_X(u)$$

$$\Rightarrow F_{\mathcal{A}}(u) = F_X(u)$$

2.3 Task C

Refer to the code file **2c.py** and the image **2c.png** .

Libraries used are

- numpy
- matplotlib
- scipy

2.4 Task D

Refer to the code file **2d.py** and the images **2d1.png**, **2d2.png**, **2d3.png**. Shift between values(10/50/100) for the **parameter h** to obtain plots for different depths.

Libraries used are

- numpy
- matplotlib

2.5 Task E

part 1

For random variable X to take value $2i$, the ball should take $k+i$ steps to the right and the other $k-i$ to the left. Because at each step probability for it to take right is equal to that of left, we can write the Probability Mass Distribution for X as

$$P_h[X = 2i] = \binom{2k}{k+i}\frac{1}{2^{2k}} \quad i \in \{-k, -k+1, \cdots, k-1, k\}$$

part 2

Now consider,

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

Considering Stirling's approximation we can write

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

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

$$= \frac{1}{\sqrt{\pi k}}e^{k\log(1-\frac{i^2}{k^2})}$$

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

and hence,

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

3 Fitting Data

3.1 Task A

Check the code in **3.py**

3.2 Task B

Check the code in **3.py**

3.3 Task C

Check the code in **3.py**

Proof:

For the first two moments μ_1^{bin} and μ_2^{bin}:

Given,

$$\mu_1 = E[X]$$

$$\mu_2 = E[X^2]$$

As we know, for the binomial distribution:

$$E[X] = E\left(\sum_{i=1}^{n} Y_i\right)$$

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

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

$$\mu_1 = E[X] = np \tag{1}$$

Where $X_i = \sum_{i=1}^{n} Y_i$ and $Y_i \sim Bern(p)$.

$$E[X^2] = Var[X] + (E[X])^2$$
$$= np(1-p) + (np)^2$$
$$= np(1-p) + (np)^2$$
$$= np - np^2 + n^2p^2$$
$$\mu_2 = E[X^2] = np + n(n-1)p^2 \tag{2}$$

Where, $Var[X] = np(1-p)$ for the normal binomial distribution and $E[X] = np$.

3.4 Task D

Check the code in **3.py**

Proof:

The first two moments are given by $\mu_1^{Gamma} = E[X]$ and $\mu_2^{Gamma} = E[X^2]$.

Note that the random variable x is non-negative in gamma distributions.

$$E[X] = \int_0^\infty x f_X(x) dx$$
$$= \int_0^\infty x \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} dx$$
$$= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^k e^{-\frac{x}{\theta}} dx$$
$$= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty \theta^{k+1} t^k e^{-t} dt$$
$$= \frac{\theta^{k+1}}{\theta^k \Gamma(k)} \int_0^\infty t^k e^{-t} dt$$
$$= \frac{\theta^{k+1}\Gamma(k+1)}{\theta^k \Gamma(k)}$$
$$= \frac{\theta^{k+1} k \Gamma(k)}{\theta^k \Gamma(k)}$$
$$E[X] = k\theta$$

$$\mu_1^{Gamma} = E[X] = k\theta \tag{1}$$

We have used the Gamma function property $\Gamma(k+1) = k\Gamma(k)$.

$$
\begin{aligned}
E[X^2] &= \int_0^\infty x^2 f_X(x)dx \\
&= \int_0^\infty x^2 \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} dx \\
&= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty x^{k+1} e^{-\frac{x}{\theta}} dx \\
&= \frac{1}{\theta^k \Gamma(k)} \int_0^\infty \theta^{k+2} t^{k+1} e^{-t} dt \\
&= \frac{\theta^{k+2}}{\theta^k \Gamma(k)} \int_0^\infty t^{k+1} e^{-t} dt \\
&= \frac{\theta^{k+2}\Gamma(k+2)}{\theta^k \Gamma(k)} \\
&= \frac{\theta^{k+2}(k+1)k\Gamma(k)}{\theta^k \Gamma(k)} \\
E[X^2] &= \theta^2(k^2+k)
\end{aligned}
$$

$$
\mu_2^{Gamma} = E[X^2] = k(k+1)\theta^2 \tag{2}
$$

3.5 Task E

Check the code in **3.py**, Binomial distribution is better approximation than Gamma distribution.

3.6 Task F

Check the code in **3.py** , it is not a better approximation than the above two distributions.

4 Quality in Inequalities

4.1 Task A

Intuitive Proof

Let the random variable X take on one of three values: $0, 1, 2$. Suppose we know that $\mathbb{E}(X) = 1$.

We want to find the highest possible probability that could be assigned to a certain set of values without violating $\mathbb{E}(X) = 1$.

To calculate $\mathbb{E}(X)$, we sum $x \cdot P(x)$ for all x. This means that the maximum contribution of $2 \cdot P(2)$ cannot exceed $\mathbb{E}(X)$. Therefore,

$$
2 \cdot P(2) \leq 1, \quad \text{so} \quad P(2) \leq \frac{1}{2}.
$$

This leads to the conclusion that ”no more than $\frac{1}{a}$ of the population can have a value more than a times the average”. This gives Markov’s inequality, which states:

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

Proof

Proof when X is a continuous random variable.

Expectation

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

where $f_X(x)$ is probability density function of X.

Splitting the integral into two parts

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

For all $x \geq a$

$$\int_a^\infty x f_X(x)\,dx \geq \int_a^\infty a f_X(x)\,dx.$$

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

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

$$\mathbb{E}[X] \geq a\,P(X \geq a).$$

This gives gives Markov’s inequality

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

4.2 Task B

Let X be a real-valued random variable with variance σ^2 and expectation μ. We want to prove that for any $\tau > 0$,

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

By using Markov’s inequality.

Let $Y = X - \mathbb{E}[X]$ so that $\mathbb{E}[Y] = 0$ and $\mathrm{Var}(Y) = \sigma^2$.

We use Markov’s inequality on Y^2 to bound $P(Y^2 \geq \tau^2)$. Markov’s inequality is

$$P(Z \geq a) \leq \frac{\mathbb{E}[Z]}{a}.$$

Applying it to Y^2 with $a = \tau^2$, we get

$$P(Y^2 \geq \tau^2) \leq \frac{\mathbb{E}[Y^2]}{\tau^2}.$$

Since $\mathbb{E}[Y^2] = \sigma^2$, we have

$$P(Y^2 \geq \tau^2) \leq \frac{\sigma^2}{\tau^2}.$$

This gives us a bound for $P(Y \geq \tau)$ because $Y \geq \tau$ implies $Y^2 \geq \tau^2$, so

$$P(Y \geq \tau) \leq P(Y^2 \geq \tau^2) \leq \frac{\sigma^2}{\tau^2}.$$

Let $u \geq 0$:

$$P(Y \geq \tau) = P(Y + u \geq \tau + u) \leq P((Y + u)^2 \geq (\tau + u)^2).$$

Applying Markov's inequality to $(Y + u)^2$, we get

$$P((Y + u)^2 \geq (\tau + u)^2) \leq \frac{\mathbb{E}[(Y + u)^2]}{(\tau + u)^2}.$$

Expectation is

$$\mathbb{E}[(Y + u)^2] = \mathbb{E}[Y^2] + 2u\mathbb{E}[Y] + u^2 = \sigma^2 + u^2,$$

since $\mathbb{E}[Y] = 0$. Thus, the bound becomes

$$P(Y \geq \tau) \leq \frac{\sigma^2 + u^2}{(\tau + u)^2}.$$

To minimize the right-hand side, differentiate $\frac{\sigma^2+u^2}{(\tau+u)^2}$ with respect to u and set the derivative equal to 0

$$\frac{d}{du}\left(\frac{\sigma^2 + u^2}{(\tau + u)^2}\right) = \frac{2u(\tau + u)^2 - 2(\sigma^2 + u^2)(\tau + u)}{(\tau + u)^4} = 0.$$

Solving this equation for u, we find

$$u_* = \frac{\sigma^2}{\tau}.$$

Substitute this back into the bound:

$$P(Y \geq \tau) \leq \frac{\sigma^2 + \left(\frac{\sigma^2}{\tau}\right)^2}{\left(\tau + \frac{\sigma^2}{\tau}\right)^2}.$$

Simplifying gives

$$P(Y \geq \tau) \leq \frac{\sigma^2}{\tau^2 + \sigma^2}.$$

Thus, we have shown that

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

This completes the proof of the Chebyshev-Cantelli inequality.

4.3 Task C

Let X be a random variable with moment generating function (MGF) $M_X(t) = \mathbb{E}[e^{tX}]$, which is valid for $-h < t < h$. We want to prove two inequalities using Markov's inequality.

Case 1: $P(X \geq x)$

Let $t > 0$, and let $u(X) = e^{tX}$. Since $u(X)$ is a positive and increasing function, we can apply Markov's inequality

$$P(X \geq x) = P(e^{tX} \geq e^{tx}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} = e^{-tx}\mathbb{E}[e^{tX}] = e^{-tx}M_X(t).$$

Thus, we obtain the bound

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

Case 2: $P(X \leq x)$

Now, let $t < 0$, and let $u(X) = e^{tX}$ again.$u(X)$ is a positive and decreasing function. Applying Markov's inequality gives

$$P(X \leq x) = P(e^{tX} \geq e^{tx}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} = e^{-tx}\mathbb{E}[e^{tX}] = e^{-tx}M_X(t).$$

Thus, we obtain the bound

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

This completes the proof.

4.4 Task D

We are given n independent Bernoulli random variables $X_1, X_2, \ldots, X_n$ where $\mathbb{E}[X_i] = p_i$. Since each X_i is independent and identically distributed (i.i.d.), Let a random variable Y is:

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

1. Expectation of Y

The expectation of Y is the sum of the expectations of the individual Bernoulli random variables. Since each X_i is Bernoulli, $\mathbb{E}[X_i] = p_i$. Therefore, the expectation of Y is:

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

Let $p_i = p$ then

$$\mathbb{E}[Y] = np.$$

Let $\mu = \mathbb{E}[Y]$, then

$$\mu = np.$$

2. Chernoff Bound for Y

We need to prove

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

Proof:
Let $t > 0$ be a variable. The Chernoff bound for Y is

$$P(Y \geq (1+\delta)\mu) = P(e^{tY} \geq e^{t(1+\delta)\mu}).$$

Using Markov's inequality

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

Calculating $\mathbb{E}[e^{tY}]$:
Since Y is the sum of independent Bernoulli random variables, we have

$$\mathbb{E}[e^{tY}] = \mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}].$$

For a Bernoulli random variable X_i with probability p, we know:

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

Thus:

$$\mathbb{E}[e^{tY}] = \prod_{i=1}^{n} \left(pe^t + (1-p)\right) = \left(pe^t + (1-p)\right)^n.$$

$$\mathbb{E}[e^{tY}] = \left(1 + p(e^t - 1)\right)^n.$$

Using $\mu = np$

$$\mathbb{E}[e^{tY}] \leq e^{\mu(e^t-1)}.$$

We have

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

This completes the proof.

3. Improving the Bound

To improve this bound, Let a variable t such that minimizes the right-hand side of the inequality. For calculating t

$$\frac{d}{dt}\left(\frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}}\right) = 0.$$

Taking the derivative and solving for t, we can calculate the value of t is:

$$t = \log(1+\delta).$$

Substituting this value of t into the bound gives the better bound.

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

This is the better bound obtained by choosing the optimal t.

4.5 Task E

We want to prove that for i.i.d. random variables $X_1, X_2, \ldots, X_n$ with mean μ, the probability that the sample mean $A_n = \frac{1}{n}\sum_{i=1}^n X_i$ deviates from the true mean μ by more than ϵ approaches zero as $n \to \infty$.

$$\lim_{n\to\infty} P\left(|A_n - \mu| > \epsilon\right) = 0 \quad \text{for all } \epsilon > 0.$$

Let $X_1, X_2, \ldots, X_n$ be i.i.d. Bernoulli random variables with mean μ. The sample mean A_n is

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

$$P\left(|A_n - \mu| > \epsilon\right) = P\left(A_n > \mu + \epsilon\right) + P\left(A_n < \mu - \epsilon\right).$$

By symmetry, it suffices to bound $P\left(A_n > \mu + \epsilon\right)$ and then multiply by 2. We'll use the Chernoff bound to do this.

Define Y as the sum of the random variables:

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

The sample mean is:

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

probability:

$$P\left(A_n > \mu + \epsilon\right) = P\left(Y > n(\mu + \epsilon)\right).$$

From the Chernoff bound, for any $\delta > 0$

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

substituting δ in terms of ϵ gives

$$\delta = \frac{\epsilon}{\mu}.$$

probability becomes:

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

Since $P\left(|A_n - \mu| > \epsilon\right)$ is the sum of the probabilities of the two tails (both $A_n > \mu + \epsilon$ and $A_n < \mu - \epsilon$), we multiply by 2 to get:

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

As $n \to \infty$, the exponential term goes to 0. Therefore, we can conclude:

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

This completes the proof of the Weak Law of Large Numbers using the Chernoff bound.

5 A Pretty "Normal" Mixture

5.1 Task A

To show that the sampling algorithm produces a random variable with the same distribution as the Gaussian Mixture Model (GMM), we will check the probability density function (PDF) of the random variable $\mathcal{A}$ sampled by the algorithm matches the PDF of the GMM variable X.

Let the randomly selected one of the K Gaussian distributions with probability p_i. The i is chosen with probability p_i.

Given the chosen index i, sample from the Gaussian distribution $\mathcal{N}(\mu_i, \sigma_i^2)$. This provides the final value sampled from the GMM.

Let:

- The random variable $\mathcal{A}$ be the result of this sampling process.
- X_i be the random variable following the Gaussian distribution $\mathcal{N}(\mu_i, \sigma_i^2)$.
- The PDF of $\mathcal{A}$ be $f_{\mathcal{A}}(u)$.
- The PDF of the GMM variable X be $f_X(u)$.

To find $f_{\mathcal{A}}(u)$, we consider that $\mathcal{A}$ can be any of the Gaussian distributions $\mathcal{N}(\mu_i, \sigma_i^2)$ depending on the chosen index i. The probability of selecting index i is p_i, and if i is chosen, $\mathcal{A}$ follows $\mathcal{N}(\mu_i, \sigma_i^2)$.

PDF of $\mathcal{A}$ is:

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

where $f_{X_i}(u)$ is the PDF of $\mathcal{N}(\mu_i, \sigma_i^2)$:

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

$$f_{\mathcal{A}}(u) = \sum_{i=1}^{K} p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(u-\mu_i)^2}{2\sigma_i^2}\right)$$

The PDF of the GMM variable X is given by:

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

$$f_{\mathcal{A}}(u) = \sum_{i=1}^{K} p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(u-\mu_i)^2}{2\sigma_i^2}\right)$$

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

Thus:

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

This confirms that the sampling algorithm for $\mathcal{A}$ produces a random variable with the same distribution as the GMM variable X.

5.2 Task B

Let X be a random variable following a Gaussian Mixture Model (GMM) where each Gaussian component X_i follows $\mathcal{N}(\mu_i, \sigma_i^2)$ and is chosen with probability p_i. We need to compute:

1. Expected Value $E[X]$

The expected value of X is:

$$E[X] = \sum_{i=1}^{K} p_i E[X_i]$$

Since $E[X_i] = \mu_i$ for each Gaussian component X_i, then

$$E[X] = \sum_{i=1}^{K} p_i \mu_i$$

2. Variance Var$[X]$

The variance is calculated by

$$\mathrm{Var}[X] = E[\mathrm{Var}[X|i]] + \mathrm{Var}[E[X|i]]$$

$$E[\mathrm{Var}[X|i]] = \sum_{i=1}^{K} p_i \sigma_i^2$$

$$\mathrm{Var}[E[X|i]] = \mathrm{Var}[\mu_i] = \sum_{i=1}^{K} p_i (\mu_i - E[X])^2$$

Then we have

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

3. Moment Generating Function (MGF) $M_X(t)$

The Moment Generating Function (MGF) of X is:

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

Since X follows a GMM, the MGF is:

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

where $M_{X_i}(t)$ is the MGF of X_i, which is:

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

Therefore:

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

5.3 Task C

Let Z be a random variable defined as a weighted sum of K independent Gaussian random variables:

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

where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ are independent Gaussian random variables. We need to find:

1. Expected Value $E[Z]$

The expected value of Z is:

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

$$E[Z] = \sum_{i=1}^{K} p_i E[X_i]$$

For each X_i, $E[X_i] = \mu_i$:

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

2. Variance $\mathbf{Var}[Z]$

The variance of Z is

$$\mathrm{Var}[Z] = \mathrm{Var}\left[\sum_{i=1}^{K} p_i X_i\right]$$

Since X_i are independent:

$$\mathrm{Var}[Z] = \sum_{i=1}^{K} p_i^2 \mathrm{Var}[X_i]$$

For each X_i, $\mathrm{Var}[X_i] = \sigma_i^2$:

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

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

To find the PDF of Z, we note that Z is a weighted sum of independent Gaussian random variables. The weighted sum of independent Gaussian random variables is itself Gaussian. Specifically, Z follows a Gaussian distribution with:
- Mean $E[Z] = \sum_{i=1}^{K} p_i \mu_i$ - Variance $\mathrm{Var}[Z] = \sum_{i=1}^{K} p_i^2 \sigma_i^2$
So, the PDF of Z is:

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

4. The MGF $M_Z(t)$ of Z

The Moment Generating Function (MGF) of Z is:

$$M_Z(t) = E[e^{tZ}]$$

Since Z is a Gaussian random variable

$$M_Z(t) = \exp\left(tE[Z] + \frac{t^2 \mathrm{Var}[Z]}{2}\right)$$

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

5. Conclusion

Based on the properties calculated

- **Gaussian Distribution**: Z is Gaussian since it is a linear combination of independent Gaussian random variables.
- **Mean and Variance**: The mean and variance of Z are:
  - Mean: $\sum_{i=1}^{K} p_i\mu_i$
  - Variance: $\sum_{i=1}^{K} p_i^2\sigma_i^2$
- **PDF**: The PDF of Z is Gaussian, similar to normal distribution.
- **MGF**: The Moment Generating Function (MGF) of Z confirms that Z follows a Gaussian distribution, as the MGF matches the form expected for a Gaussian random variable.

Thus, Z exhibits properties consistent with a Gaussian distribution.

6. Distribution of Z

Since Z is a weighted sum of independent Gaussian random variables, Z itself follows a Gaussian distribution.

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

5.4 Task D

Theorem: For a random variable X:

- If X is finite and discrete, or
- If X is continuous and its Moment Generating Function (MGF), $\phi_X(t)$, is known over some interval,

then the MGF and the probability distribution function (PDF) of X uniquely determine each other.

We will prove this theorem for the case where X is a finite discrete random variable.

Proof:

Let X be a finite discrete random variable with possible values $x_1, x_2, \ldots, x_n$ and corresponding probabilities $p_1, p_2, \ldots, p_n$, where $p_i = \Pr[X = x_i]$ for each $i = 1, 2, \ldots, n$. The MGF of X, denoted as $\phi_X(t)$, is defined by:

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

To show that $\phi_X(t)$ uniquely determines the probabilities p_i, we will show that the p_i values can be recovered from $\phi_X(t)$.

Consider the expression for $\phi_X(t)$:

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

Since $\phi_X(t)$ is a generating function for the probabilities p_i, we can extract these probabilities by using certain properties of generating functions:

1. **Uniqueness of Coefficients:** The function $\phi_X(t)$ is a finite sum of exponential terms, each multiplied by a coefficient p_i. By differentiating $\phi_X(t)$ multiple times and evaluating the result at $t = 0$, we can gather each coefficient p_i.

   The k-th derivative of $\phi_X(t)$ is:

   $$\frac{d^k}{dt^k}\phi_X(t) = \sum_{i=1}^{n} p_i x_i^k e^{tx_i}$$

   Evaluating this at $t = 0$ gives:

   $$\frac{d^k}{dt^k}\phi_X(0) = \sum_{i=1}^{n} p_i x_i^k$$

   This represents the k-th moment of the distribution. From the moments, we can recover the probabilities p_i using techniques like the method of moments or polynomial interpolation.

2. **Invertibility:** The MGF $\phi_X(t)$ is a sum of distinct exponential terms. Each probability p_i is associated with a unique exponential term e^{tx_i}. Because of this distinct structure, we can uniquely determine each p_i by matching the coefficients in the MGF.

Thus, knowing the MGF $\phi_X(t)$ allows us to uniquely determine the probabilities p_i.

Conclusion:

Now, let's apply this theorem to the variables X and Z from Task C:

1. **Uniqueness:** Recall that X follows a Gaussian Mixture Model (GMM), so its MGF is:

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

   On the other hand, Z is a weighted sum of independent Gaussian random variables, so its MGF is:

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

2. **Comparison:** The MGFs of X and Z are different. The MGF of X includes a sum over multiple exponential terms, each weighted by p_i, while the MGF of Z has a simpler form that reflects the weighted sum of Gaussian variables.

   Since their MGFs are not the same, X and Z cannot have the same distribution.

Therefore, X and Z do not have the same properties. This is evident from their different MGFs, confirming they have different distributions.