Assignment-1

Data Analysis and Interpretation

NISCHAL 23B1024

PUSHPENDRA UIKEY 23B1023

NITHIN 23B0993

Professor: **Sunita Sarawagi**

September 9, 2024

1 Mathemagic

Task A

1. Consider a Bernoulli random variable $X \sim \text{Ber}(p)$, where X takes values 0 and 1. The probability mass function (PMF) is:

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

2. The probability generating function (PGF) is defined as:

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

Since X can only take the values 0 and 1, the sum reduces to:

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

3. Substitute the known values for $P(X=0)$ and $P(X=1)$:

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

4. Thus, the PGF for a Bernoulli random variable is:

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

Task B

1. Consider a binomial random variable $X \sim \text{Bin}(n,p)$. The PMF for a binomial distribution is given by:

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

2. The PGF is defined as:

$$G_{\text{Bin}}(z) = \mathbb{E}[z^X] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k.$$

3. Factor the terms:

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

4. Recognize this as a binomial expansion:

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

5. From Task A, we know that the PGF of a Bernoulli random variable is $G_{\text{Ber}}(z) = (1-p) + pz$. Therefore:

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

Task C

1. Suppose $X_1, X_2, \ldots, X_k$ are independent random variables with the same PGF $G(z)$. Let $X = X_1 + X_2 + \cdots + X_k$.

2. The PGF of the sum of independent random variables is the product of their individual PGFs. Therefore:

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

$$\Sigma(z) = [G(z)]^k$$

3. Simplifying, we get:

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

Thus, the PGF of the sum $X = X_1 + X_2 + \cdots + X_k$ is $G_\Sigma(z) = G(z)^k$.

Task D

1. Consider a geometric random variable X with parameter p, denoted $X \sim \text{Geo}(p)$. The probability mass function (PMF) is given by:

$$P(X = x) = p(1-p)^{x-1}, \quad \text{for } x \geq 1.$$

2. The Probability Generating Function (PGF) is defined as:

$$G(z) = \mathbb{E}[z^X] = \sum_{x=1}^{\infty} P(X = x) \cdot z^x.$$

3. Substituting the PMF into the definition:

$$G(z) = \sum_{x=1}^{\infty} p(1-p)^{x-1} \cdot z^x.$$

4. Factor out the constant p from the summation:

$$G(z) = p \sum_{x=1}^{\infty} \left[(1-p)^{x-1} \cdot z^x\right].$$

5. Notice that $z^x = z \cdot z^{x-1}$, and rewrite the series:

$$G(z) = pz \sum_{x=1}^{\infty} \left[(1-p)z\right]^{x-1}.$$

6. Recognize that the sum is a geometric series of the form $\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$, valid for $|a| < 1$. Here, the first term of the series corresponds to $n = 0$, so we adjust the index of summation:

$$G(z) = pz \sum_{n=0}^{\infty} \left[(1-p)z\right]^n.$$

7. Apply the geometric series formula:

$$G(z) = pz \cdot \frac{1}{1-(1-p)z}.$$

8. Simplify the expression to get the final form of the PGF:

$$G(z) = \frac{pz}{1-(1-p)z}.$$

Thus, the Probability Generating Function (PGF) for the geometric distribution is:

$$G(z) = \frac{pz}{1-(1-p)z}.$$

Task E

1. Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{NegBin}(n, p)$. We are asked to show that:

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

2. The PGF for a binomial random variable is:

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

3. The PGF for a negative binomial random variable is:

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

4. To prove the relationship, recognize that the negative binomial PGF is the inverse of the binomial PGF in a transformed form. By substituting $z^{-1}(z-1)$ into the binomial PGF, we can show that:

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

Task F

1. The generalized binomial coefficient is given by:

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

2. The negative binomial expansion for $(1+x)^{-\alpha}$ is:

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

3. This result can be derived using the binomial expansion theorem for real numbers with negative exponents. It generalizes the standard binomial theorem to allow for non-integer powers.

4. The sum represents an infinite series, with the binomial coefficient extending to non-integer values of α.

Task G

1. The mean of a random variable X is obtained by differentiating the PGF and evaluating at $z = 1$:

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

2. **Bernoulli:** From Task A, the PGF is:

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

Differentiating:

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

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

3. **Binomial:** From Task B, the PGF is:

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

Differentiating:

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

Evaluating at $z = 1$:

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

4. **Geometric:** From Task D, the PGF is:

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

Differentiating:

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

Evaluating at $z = 1$:

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

5. **Negative Binomial:** The PGF is:

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

Differentiating:

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

Evaluating at $z = 1$:

$$\mathbb{E}[X] = \frac{n(1-p)}{p}.$$

2 Normal Sampling

Task A

To prove that if X is a continuous real-valued random variable with CDF F_X that is invertible, then the random variable $Y = F_X(X)$ is uniformly distributed in $[0,1]$, follow these steps:

1. **Definition of Uniform Distribution:** A random variable Y is uniformly distributed in $[0,1]$ if its CDF $F_Y(y)$ satisfies $F_Y(y) = y$ for $y \in [0,1]$.

2. **CDF of Y:** To find the CDF of Y, denote $F_Y(y)$ as the probability $P(Y \leq y)$:

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

3. **Relationship Between X and $F_X(X)$:** Since F_X is a strictly increasing function (because it is invertible), the inequality $F_X(X) \leq y$ is equivalent to $X \leq F_X^{-1}(y)$. Therefore:

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

4. **Using the Definition of F_X:** By the definition of the CDF F_X, we have:

$$P(X \leq x) = F_X(x)$$

Substituting $x = F_X^{-1}(y)$ into the above equation:

$$F_Y(y) = F_X(F_X^{-1}(y))$$

5. **Simplification:** Since F_X and F_X^{-1} are inverse functions:

$$F_X(F_X^{-1}(y)) = y$$

6. **Conclusion:** Thus, the CDF of Y is $F_Y(y) = y$ for $y \in [0, 1]$, which means that Y is uniformly distributed over $[0, 1]$.

Task B

To construct an algorithm A that converts a uniformly distributed random variable $Y \in [0, 1]$ into a random variable $t \in \mathbb{R}$ with the same distribution as X, follow these steps:

1. **Define the Input and Goal:**
   - **Input:** A random variable Y uniformly distributed on $[0, 1]$.
   - **Goal:** Transform Y into a random variable t that follows the same distribution as X.

2. **Algorithm Definition:**
   - Define the algorithm A as:
   $$t = A(Y) = F_X^{-1}(Y)$$
   Here, F_X^{-1} is the inverse of the CDF of X.

3. **Justification:**
   - The inverse CDF function F_X^{-1} converts a uniform random variable Y into a value t such that $F_X(t)$ equals Y.
   - This transformation ensures that t has the same distribution as X because $F_X(t)$ represents the probability that $X \leq t$.

4. **Verify the CDF:**

   - To show that t follows the same distribution as X, check the CDF of t:

     $$F_A(t) = P(A(Y) \leq t)$$

     Substitute $A(Y) = F_X^{-1}(Y)$:

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

   - Since F_X^{-1} is an increasing function, $F_X^{-1}(Y) \leq t$ is equivalent to $Y \leq F_X(t)$:

     $$F_A(t) = P(Y \leq F_X(t))$$

   - Given that Y is uniformly distributed on $[0, 1]$:

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

5. **Conclusion:**

   - Thus:

     $$F_A(t) = F_X(t)$$

   - This means that t has the same CDF as X, so t follows the same distribution as X.

Task E (Sub-task 1)

1. **Setup:**

   - Let $h = 2k$ be the depth of the Galton board, where h is even.
   - After h steps, the ball lands in one of the positions $\{-h, -h + 2, \ldots, h - 2, h\}$.

2. **Binomial Distribution:**

   - Each step of the ball is equally likely to be left or right. Let X be the number of right steps.
   - The number of right steps X follows a binomial distribution:

     $$X \sim \text{Binomial}(h, 0.5)$$

3. **Final Position Calculation:**

   - The final position X_f of the ball can be expressed as:

     $$X_f = 2X - h$$

   - Therefore, for X_f to be $2i$:

     $$2i = 2X - h \quad \text{or} \quad X = k + i$$

4. **Probability Calculation:**

- The probability that $X_f = 2i$ is the probability that $X = k + i$:

$$P_h[X_f = 2i] = P(X = k + i)$$

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

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

Task E (Sub-task 2)

1. **Normal Approximation:**

   - For large h, the binomial distribution can be approximated by a normal distribution due to the Central Limit Theorem (CLT).

2. **Mean and Variance:**

   - For $X \sim \text{Binomial}(h, 0.5)$, the mean μ and variance σ^2 are:

$$\mu = \frac{h}{2}$$

$$\sigma^2 = \frac{h}{4}$$

   - Therefore, X can be approximated by the normal distribution:

$$X \approx N\left(\frac{h}{2}, \frac{h}{4}\right)$$

3. **Probability Approximation:**

   - The final position X_f is:

$$X_f \approx 2X - h$$

   - Hence, X_f follows approximately:

$$X_f \approx N\left(2 \cdot \frac{h}{2} - h, 4 \cdot \frac{h}{4}\right) \Rightarrow X_f \approx N(0, h)$$

4. **Probability Density Function:**

   - The probability that $X_f \approx 2i$ can be approximated by the normal density function:

$$P_h[X_f = 2i] \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(2i)^2}{2\sigma^2}\right)$$

   - Substituting $\sigma^2 = \frac{h}{4}$:

$$P_h[X_f = 2i] \approx \frac{1}{\sqrt{2\pi\frac{h}{4}}} \exp\left(-\frac{(2i)^2}{2 \cdot \frac{h}{4}}\right)$$

$$P_h[X_f = 2i] \approx \frac{1}{\sqrt{\pi h}} \exp\left(-\frac{4i^2}{h}\right)$$

5. **Conclusion:**

   - For large h and when i is much smaller than h, the distribution of the final positions of the balls approximates a normal distribution.

3 Quality in Inequalities

Task A

Markov's Inequality states that for any non-negative random variable X and any positive number a, the probability that X is at least a is at most the expected value of X divided by a:

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

First, we need to see why this makes sense.

1. **Expectation as an Average**: Visualize the expectation $\mathbb{E}[X]$ as the typical range of X. At this point, if X tends to get small values, then the average value will be small as well.

2. **Large Values of X**: When the value of X is large, i.e. at least a, it greatly affects the average value $\mathbb{E}[X]$. However, if X is almost never this high, then the likelihood the event will happen is the small $P(X \geq a)$. In other words, when X is mostly low, the rest of the numbers it can be are lesser than or equal to a.

3. **Bounding the Probability**: Suppose the very worst thing: every time X is at least a, it's exactly a. However, in this case, the proportional value of the long tail numbers contributes to the average which is $a \times P(X \geq a)$. Since this must be a maximum of the actual average $\mathbb{E}[X]$, we get:

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

Dividing both sides by a, we find:

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

This tells us that the probability $P(X \geq a)$ cannot be larger than $\frac{\mathbb{E}[X]}{a}$, which is exactly what Markov's Inequality states.

This intuitive reasoning helps us understand why the inequality holds without diving into formal mathematical proofs. We see that the expectation $\mathbb{E}[X]$ acts as a kind of upper bound on the probability that X takes on large values relative to a.

Task A: Second part

Let X be a non-negative continuous random variable, and let $a > 0$. Markov's Inequality tells us that the probability of X being at least a is at most the expected value of X divided by a:

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

Here's how we can rigorously prove this.

- **Step 1: Understanding Expectation**

  The expectation $\mathbb{E}[X]$ is essentially the "average" value of the random variable X. For a continuous random variable X with probability density function (PDF) $f_X(x)$, the expectation is defined as:

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

  This integral sums up all possible values of X, weighted by how likely they are (given by the PDF $f_X(x)$).

- **Step 2: Splitting the Expectation**

  To analyze the probability that X is large (specifically, at least a), let's break the expectation into two parts:

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

  The first part accounts for values of X that are less than a, while the second part accounts for values of X that are at least a. Since we're interested in the probability $P(X \geq a)$, we focus on the second integral.

- **Step 3: Bounding the Integral**

  For $x \geq a$, it's clear that x is at least a. Therefore, we can say:

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

  This inequality holds because $x \geq a$ for all x in the interval $[a, \infty)$.

- **Step 4: Simplifying the Bound**

  Notice that the right-hand side is simply the constant a multiplied by the probability that X is at least a:

  $$\int_a^\infty a f_X(x)\, dx = a \int_a^\infty f_X(x)\, dx = a \cdot P(X \geq a)$$

  So we have:

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

- **Step 5: The Final Inequality**

  Finally, by dividing both sides of the inequality by a, we obtain Markov's Inequality:

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

  This completes the proof. The key idea was to break down the expectation into parts that allowed us to bound the probability $P(X \geq a)$ using the expectation $\mathbb{E}[X]$.

Task B

To prove the Chebyshev-Cantelli inequality:

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

we proceed with the following steps:

1. Start with the given inequality:

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

where $\mathbb{E}[(X - \mu + a)^2] = \sigma^2 + a^2$.

2. Substitute the expectation into the inequality:

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

3. To minimize the right-hand side, differentiate with respect to a:

$$\frac{d}{da}\left(\frac{\sigma^2 + a^2}{(\tau + a)^2}\right) = \frac{2a(\tau + a)^2 - 2(\tau + a)(\sigma^2 + a^2)}{(\tau + a)^4}$$

4. Set the derivative to zero:

$$a(\tau + a) = \sigma^2 + a^2$$

5. Simplify the equation to find a:

$$a\tau = \sigma^2 \quad \Rightarrow \quad a = \frac{\sigma^2}{\tau}$$

6. Substitute $a = \frac{\sigma^2}{\tau}$ back into the original inequality:

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

7. Simplify the right-hand side:

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

8. Thus, we derive the Chebyshev-Cantelli inequality:

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

Task C

Given the moment-generating function (MGF) $M_X(t) = \mathbb{E}[e^{tX}]$, we want to show:

- For $t > 0$:

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

- For $t < 0$:

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

Proof

- **For** $t > 0$:

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

Applying Markov's Inequality:

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

Thus:

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

- **For** $t < 0$:

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

Applying Markov's Inequality:

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

Since $M_X(-t) = e^{-tx} M_X(t)$:

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

Task D

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables where $\mathbb{E}[X_i] = p_i$. Define Y as the sum of these random variables:

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

1. Expectation of Y

The expectation of Y is:

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

2. Bound on $\mathbb{P}(Y \geq (1+\delta)\mu)$

To bound $\mathbb{P}(Y \geq (1+\delta)\mu)$, we use the moment-generating function (MGF).

1. **MGF of Y:**

   The MGF of Y is:

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

   Since $Y = \sum_{i=1}^{n} X_i$ and the X_i are independent:

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

For each Bernoulli random variable X_i:

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

Therefore:

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

2. **Applying Markov's Inequality:**

   Using Markov's Inequality with e^{tY}:

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

   Applying Markov's Inequality:

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

   Thus:

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

3. **Substitute MGF and Simplify:**

   Using $M_Y(t)$ and $\mu = \sum_{i=1}^n p_i$:

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

   To simplify, note that:

$$\prod_{i=1}^n \left[(1 - p_i) + p_i e^t\right] \leq e^{\mu(e^t - 1)}$$

   Therefore:

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

3. Choosing the Optimal t

To minimize the bound, choose t that minimizes the exponent:

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

Taking the derivative with respect to t and setting it to zero:

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

Solving for t:

$$e^t = 1+\delta \quad \Rightarrow \quad t = \ln(1+\delta)$$

Substitute $t = \ln(1+\delta)$ into the bound:

$$\mathbb{P}(Y \geq (1+\delta)\mu) \leq e^{\mu(\ln(1+\delta)-1-\ln(1+\delta))} = e^{-\mu\ln(1+\delta)} = (1+\delta)^{-\mu}$$

Task E

Theorem Statement (WLLN)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ. Define the sample average as:

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

The WLLN asserts that for any $\epsilon > 0$:

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

This implies that as the sample size n grows, the probability that the sample average deviates from the true mean by more than ϵ approaches zero.

Proof Using the Chernoff Bound

To prove the WLLN, we bound the probability $P\left(|A_n - \mu| > \epsilon\right)$ using the Chernoff bound. We focus on bounding $P\left(A_n > \mu + \epsilon\right)$; the bound for $P\left(A_n < \mu - \epsilon\right)$ follows similarly by symmetry.

- Let $Y = \sum_{i=1}^{n}(X_i - \mu)$. Then $A_n > \mu + \epsilon$ is equivalent to $Y > n\epsilon$.
- Applying the Chernoff bound:
$$P(Y > n\epsilon) \leq \frac{\mathbb{E}\left[e^{tY}\right]}{e^{tn\epsilon}}.$$
- Since Y is a sum of i.i.d. random variables with zero mean, the moment generating function (MGF) $\mathbb{E}\left[e^{tY}\right]$ is:
$$\mathbb{E}\left[e^{tY}\right] = \left(\mathbb{E}\left[e^{t(X_1-\mu)}\right]\right)^n.$$

- Expanding the MGF for small t:

$$\mathbb{E}\left[e^{t(X_i-\mu)}\right] \approx 1 + \frac{t^2\sigma^2}{2},$$

where σ^2 is the variance of X_i. Thus,

$$\mathbb{E}\left[e^{tY}\right] \approx \exp\left(\frac{nt^2\sigma^2}{2}\right).$$

- The Chernoff bound becomes:

$$P(Y > n\epsilon) \leq \exp\left(\frac{nt^2\sigma^2}{2} - tn\epsilon\right).$$

- Optimizing the bound by choosing $t = \frac{\epsilon}{\sigma^2}$, we obtain:

$$P(Y > n\epsilon) \leq \exp\left(-\frac{n\epsilon^2}{2\sigma^2}\right).$$

- Thus, the probability that the sample average deviates from the mean by more than ϵ is:

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

- As n increases, this probability approaches zero, thereby proving the Weak Law of Large Numbers.

4 A Pretty "Normal" Mixture

Task A

We want to show that the algorithm samples from the same distribution as the Gaussian Mixture Model (GMM).

- **Gaussian Mixture Model (GMM):**

  The GMM is defined as a mixture of K Gaussian distributions. The probability density function (PDF) of the GMM variable X is given by:

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

  where:

  - p_i is the probability of choosing the ith Gaussian component,
  - $f_{X_i}(u)$ is the PDF of the ith Gaussian X_i, which is $\mathcal{N}(\mu_i, \sigma_i^2)$.

- **The Sampling Algorithm:**

  The algorithm samples from the GMM as follows:

  1. **Choose a Component**: Select index i with probability p_i.
  2. **Sample from the Chosen Component**: Sample a value u from the Gaussian distribution $\mathcal{N}(\mu_i, \sigma_i^2)$.

- **PDF of the Sampled Variable:**

  Let $\mathcal{A}$ be the random variable representing the output of the algorithm. The PDF of $\mathcal{A}$ at u can be calculated by considering the two steps of the algorithm:

  1. The probability of choosing the ith component is p_i.
  2. The probability of sampling the value u from the chosen Gaussian is $f_{X_i}(u)$.

  Thus, the PDF of $\mathcal{A}$ is:

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

- **Conclusion:**

  We observe that the PDF $f_{\mathcal{A}}(u)$ is the same as the PDF $f_X(u)$ of the GMM:

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

  Therefore, the algorithm correctly samples from the GMM distribution.

Task B

- **Expectation** $\mathbb{E}[X]$:

  The expectation $\mathbb{E}[X]$ of the GMM can be computed as the weighted sum of the expectations of the individual Gaussian components:

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

- **Variance Var**(X):

  The variance $\mathrm{Var}(X)$ of the GMM can be computed using the law of total variance:

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid I)] + \mathrm{Var}(\mathbb{E}[X \mid I])$$

  Here:

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

  Thus, the total variance is:

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

- **Moment Generating Function** $M_X(t)$:

  The moment generating function (MGF) $M_X(t)$ of the GMM is given by:

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

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 each $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Let's find the properties of Z.

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

   The expected value of Z is the weighted sum of the means of the X_i's:

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

2. **Variance Var(Z):**

   The variance of Z is the sum of the variances of the X_i's, each weighted by p_i^2:

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

3. **Probability Density Function (PDF) $f_Z(u)$:**

   Since Z is a weighted sum of independent Gaussians, it is itself Gaussian. Its PDF is:

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

4. **Moment Generating Function (MGF) $M_Z(t)$:**

   The MGF of Z, which is also Gaussian, is:

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

5. **Comparison:**

   Comparing Z with X from Task B:

   - Both X and Z have the same expected value:

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

   - The variance of X is generally not the same as Var(Z). The variance of X is:

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

   Thus, X (a GMM) and Z (a weighted sum of Gaussians) do not have the same variance and hence do not have the same distribution properties.

6. **Distribution of Z:**

   Z is a Gaussian random variable. Specifically:

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