CS215 - Assignment 2

Ujwal L Shankar (23b1050)
Madhav R Babu (23b1060)
Marumamula Venkata Pranay (23b1073)

Contents

1 Mathemagic

1.1 Task A

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

$$P(X = 1) = p \quad \text{and} \quad P(X = 0) = 1 - p$$

The probability generating function (PGF) is:

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

Since the random variable takes values only 0 and 1:

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

1.2 Task B

Given $X \sim \text{Bin}(n, p)$, we have:

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

The PGF is:

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

This is the binomial expansion of $(pz + 1 - p)^n$. Thus:

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

1.3 Task C

Given independent non-negative integer-valued random variables $X_1, X_2, \ldots, X_k$ with common probability density function P and PGF $G(z)$, we have:

$$X = X_1 + X_2 + \cdots + X_k$$

The PGF of X is:

$$G_{\Sigma}(z) = E[z^X] = G(z)^k$$

This follows by applying the fact that the X_i are independent and using their individual PGFs.

1.4 Task D

Given $X \sim \text{Geo}(p)$, we have:

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

The PGF is:

$$G_{\text{Geo}}(z) = E[z^X] = \sum_{i=1}^{\infty} (1 - p)^{i-1} p z^i = pz \sum_{j=0}^{\infty} [(1 - p)z]^j$$

Using the infinite geometric series formula:

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

1.5 Task E

Given $X \sim \text{Bin}(n,p)$ and $Y \sim \text{NegBin}(n,p)$, the PGFs are $G_X^{(n,p)}(z)$ and $G_Y^{(n,p)}(z)$ respectively.

For Y, viewed as the sum of n geometric random variables, we have:

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

For X:

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

We also have:

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

1.6 Task F

The binomial coefficient for $\alpha \in \mathbb{R}$ and $k \in \mathbb{N}$ is defined as:

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

To calculate $G_Y^{(n,p)}(z)$ directly, we have:

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

Thus:

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

This simplifies to:

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

Equating this with the result from Task E gives:

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

Replace $-(1-p)z$ with x,

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

Expand $\binom{n+i-1}{i}$ and distribute $(-1)^i$ to each term in the product,

$$\binom{n+i-1}{i} \cdot (-1)^i = \frac{(n+i-1)(n+i-2)\dots(n)}{i!} \cdot (-1)^i = \frac{(-n)(-n-1)\dots(-n-i+1)}{i!} = \binom{-n}{i}$$

Hence,

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

1.7 Task G

$$G_X(z) = E[z^X] = \sum_{k=0}^{\infty} P(X=k) \cdot z^k = P(X=0) + \sum_{k=1}^{\infty} P(X=k) \cdot z^k$$
$$G_X^{'}(z) = \sum_{k=1}^{\infty} P(X=k) \cdot k \cdot z^{k-1}$$
$$G_X^{'}(1) = \sum_{k=1}^{\infty} P(X=k) \cdot k \cdot 1^{k-1} = \sum_{k=1}^{\infty} P(X=k) \cdot k = E[X]$$

$$E[X_{Ber}] = p$$
$$E[X_{Bin}] = n \cdot (pz + 1 - p)^{n-1} \cdot p\Big|_{z=1} = np$$
$$E[X_{Geo}] = \frac{(1-(1-p)z)p + pz(1-p)}{(1-(1-p)z)^2}\Bigg|_{z=1} = \frac{1}{p}$$
$$E[X_{NegBin}] = n \cdot \left(\frac{pz}{1-(1-p)z}\right)^{n-1}\Bigg|_{z=1} \cdot E[X_{Geo}] = \frac{n}{p}$$

2 Normal Sampling

2.1 Task A

Given X is a continuous real-valued random variable with invertible CDF $F_X : \mathbb{R} \to [0,1]$, Y is a random variable defined as $F_X(X) \in [0,1]$. Let $G(t)$ be the CDF of the random variable Y:

$$\begin{aligned}
G(t) &= P(Y \leq t) \\
G(t) &= P(F_X(X) \leq t) \quad \text{(by definition of } Y\text{)} \\
G(t) &= P(X \leq F_X^{-1}(t)) \quad \text{(because } F_X \text{ is increasing and invertible)} \\
G(t) &= F_X(F_X^{-1}(t)) \quad \text{(by definition of the CDF of } X\text{)} \\
G(t) &= t \\
P(Y = t) &= \frac{dG(t)}{dt} = 1
\end{aligned}$$

Hence, Y takes values in $[0,1]$ with equal probability and is uniform.

2.2 Task B

- Construction of Algorithm $\mathcal{A}$:

  $\mathcal{A} = F_X^{-1}(y)$ where y is a uniform random variable taking values from $[0,1]$ and F_X^{-1} is the inverse CDF of the Gaussian random variable $\mathcal{N}(0,1)$.

- Correctness of $\mathcal{A}$

$$F_{\mathcal{A}}(u) = P(\mathcal{A} \leq u)$$
$$F_{\mathcal{A}}(u) = P(F_X^{-1}(y) \leq u) \quad \text{by definition of } \mathcal{A}$$
$$F_{\mathcal{A}}(u) = P(y \leq F_X(u)) \quad \text{since } F_X \text{ is an increasing function}$$
$$F_{\mathcal{A}}(u) = F_X(u) \quad \text{since y is uniform.}$$

  Since $\mathcal{A}$ and X have the same CDF, they will also have the same PDF (derivative of CDF).

2.3 Task C

Code for this task has been written in 2c.py. The plot obtained has been saved as 2c.png.

2.4 Task D

Code for this task has been written in 2d.py. The plots obtained have been saved as 2d1.png, 2d2.png and 2d3.png.

2.5 Task E

- Closed Form of $P_h[X = 2i]$ for each $i \in \{-k, -k+1, ...0...k-1, k\}$

  Firstly, the final pocket in which the ball falls is dependent only on the number of times the ball deflects to the right and left, and not on the order of those events.

  We can have $h+1$ different possible compositions and hence, $h+1$ pockets in which each ball can land.

  - 0 R's $h-1$ L's
  - 1 R's $h-2$ L's .

    .

    .

  - $h-1$ R's 1 L's
  - h R's 0 L's

  We know that

  - $R_0 + L_0 = h$
  - $R_0 - L_0 = 2i$

  Solving these equations, we obtain

$$R_0 = \frac{h}{2} + i$$
$$L_0 = \frac{h}{2} - i$$

Now, the number of ways to finally settle at position $2i$ is the number of arrangements of R_0 number of R's and L_0 number of L's.

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

- $P(X = i) = \binom{h}{\frac{h+i}{2}} \cdot \left(\frac{1}{2}\right)^h$
  Expanding the binomial coefficient in terms of factorials, and then using Stirling's approximation gives us

$$P(X = i) = \frac{\sqrt{2\pi h} \cdot \left(\frac{h}{e}\right)^h}{\sqrt{2\pi \frac{h+i}{2}} \cdot \sqrt{2\pi \frac{h-i}{2}} \cdot \left(\frac{h+i}{2e}\right)^{\frac{h+i}{2}} \cdot \left(\frac{h-i}{2e}\right)^{\frac{h-i}{2}}} \cdot \left(\frac{1}{2}\right)^h$$

Simplifying,

$$P(X = i) = \sqrt{\frac{h}{2\pi \left(\frac{h^2 - i^2}{4}\right)}} \cdot \frac{h^h}{(h+i)^{\frac{h+i}{2}} \cdot (h-i)^{\frac{h-i}{2}}}$$

$$P(X = i) = \sqrt{\frac{2}{\pi h}} \cdot \left(1 + \frac{i}{h}\right)^{\frac{-h-i}{2}} \cdot \left(1 - \frac{i}{h}\right)^{\frac{i-h}{2}}$$

Writing the 2^{nd} and 3^{rd} term as $e^{\text{power}\cdot\ln(\text{base})}$,

$$P(X = i) = \sqrt{\frac{2}{\pi h}} \cdot e^{-\frac{(h+i)}{2}\cdot\ln\left(1+\frac{i}{h}\right)} \cdot e^{-\frac{(h-i)}{2}\cdot\ln\left(1-\frac{i}{h}\right)}$$

Using the expansion of $\ln(1 + x)$ up to two terms,

$$P(X = i) = \sqrt{\frac{2}{\pi h}} \cdot e^{\frac{(h+i)}{2}\cdot\left(\frac{i^2}{2h^2}-\frac{i}{h}\right)} \cdot e^{\frac{(h-i)}{2}\cdot\left(\frac{i^2}{2h^2}+\frac{i}{h}\right)}$$

Simplifying,

$$P(X = i) = \sqrt{\frac{2}{\pi h}} \cdot e^{\frac{-i^2}{2h}}$$

3 Fitting Data

3.1 Task A

Using the numpy mean() function we computed the first and second moment.
First Moment = 6.496145618324817
Second Moment= 46.554361807879815

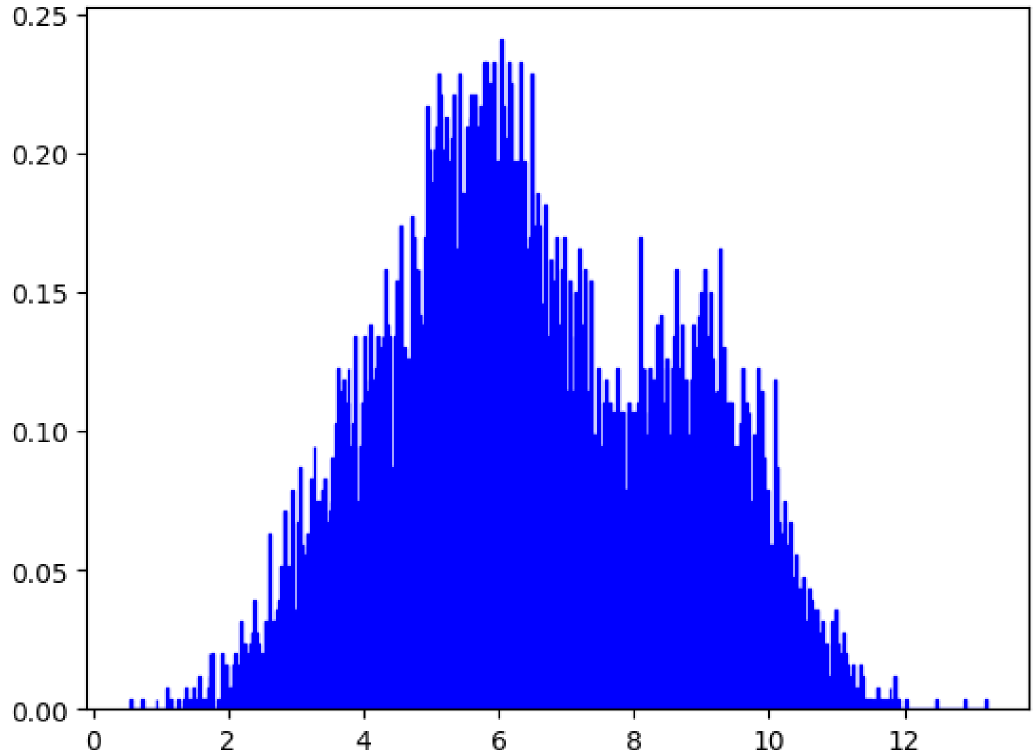

Figure 1: Histogram

3.2 Task B

From the figure, we can estimate the mode to be 6.

3.3 Task C

1. - Computing μ_1^{bin}
     $\mu_1 = \text{E[X]} = \Sigma_{k=0}^{n} k \cdot^n C_k \cdot p^k \cdot (1-p)^{n-k}$
     Binomial distribution can be treated as the sum of Bernoulli distribution hence
     $E[X]_{bin} = n \cdot E[X]_{bernouli} = n \cdot p$
     $\mu_1^{bin} = np$
   - Computing μ_2^{bin}
     $\mu_1 = E[X^2] = \Sigma_{k=0}^{n} k^2 \cdot^n C_k \cdot p^k \cdot (1-p)^{n-k}$
     {
     $k^2 \cdot^n C_k \cdot p^k \cdot (1-p)^{n-k} = (k(k-1)+k) \cdot \frac{n(n-1)(n-2)!}{k(k-1)(k-2)!(n-k)!} \cdot p^k \cdot (1-p)^{n-k}$
     $=n(n-1) \cdot^{n-2} C_{k-2} \cdot p^k (1-p)^{n-k} + k \cdot^n C_k \cdot p^k \cdot (1-p)^{n-k}$
     }

$$E[X^2] = p^2 \cdot n(n-1) \sum_{k=2}^{n} \binom{n-2}{k-2} \cdot p^k (1-p)^{n-k} + \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$= p^2 \cdot n(n-1)(p+1-p)^{n-2} + E[X]_{bin}$$

$$= p^2 \cdot n \cdot (n-1) + n \cdot p$$

$$\mu_2^{bin} = p^2 \cdot n \cdot (n-1) + n \cdot p$$

2. n= 20
   p=0.32968652963756023

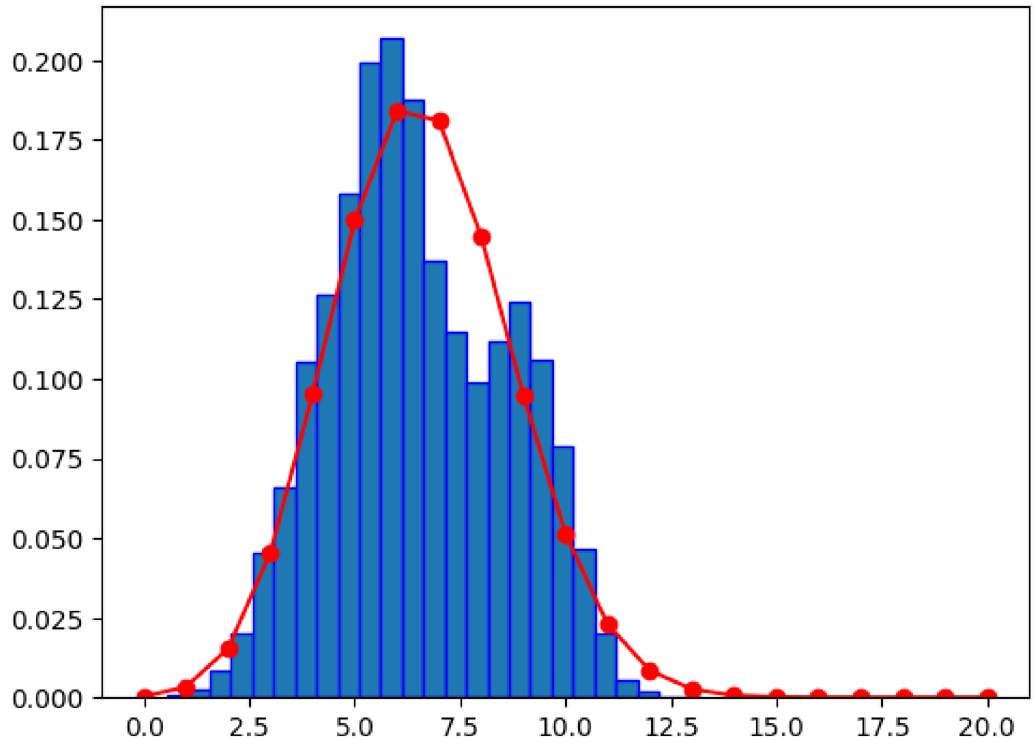

Figure 2: Bin(n,p) on top of Histogram

3.

3.4 Task D

1. • Computing μ_1^{Gamma}
   $\mu_1^{Gamma} = E[X]_{Gamma} = \int \frac{x \cdot x^{k-1} \cdot e^{\frac{-x}{\theta}}}{\theta^k \cdot \Gamma(k)} dx$
   putting $\frac{x}{\theta} = t$ gives:

$$\mu_1^{Gamma} = \int \frac{\theta \cdot t^k \cdot e^{-t}}{\Gamma(k)} dt$$

$$= \theta \cdot \frac{\Gamma(k+1)}{\Gamma(k)} = \theta \cdot \frac{k!}{(k-1)!} = \theta \cdot k$$

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

- Computing μ_2^{Gamma}
  $\mu_2^{Gamma} = E[X^2]_{Gamma} = \int \frac{x^2 \cdot x^{k-1} \cdot e^{\frac{-x}{\theta}}}{\theta^k \cdot \Gamma(k)} dx$
  putting $\frac{x}{\theta} = t$ gives:

$$\mu_2^{Gamma} = \int \frac{\theta^2 \cdot t^{k+1} \cdot e^{-t}}{\Gamma(k)} dt$$

$$= \theta^2 \cdot \frac{\Gamma(k+2)}{\Gamma(k)} = \theta^2 \cdot \frac{(k+1)!}{(k-1)!} = \theta^2 \cdot (k+1)(k)$$

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

2. k= 9.691205540717183
   theta =0.6703134703938897

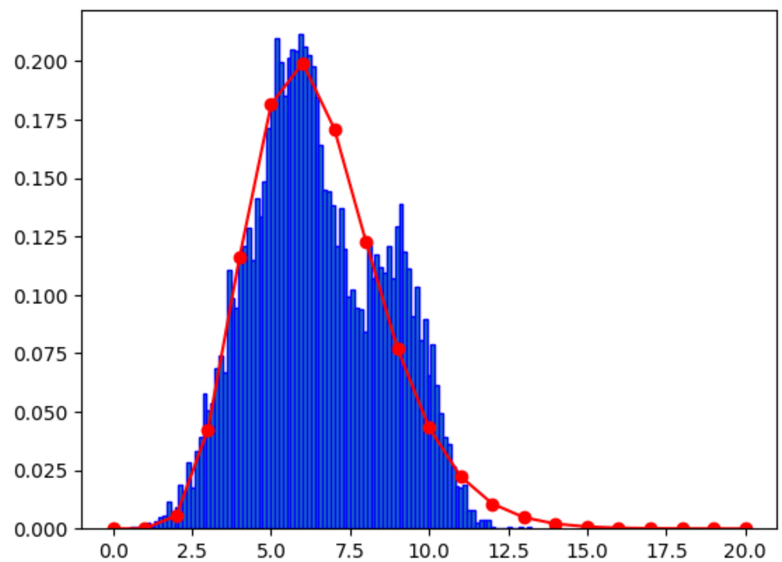

Figure 3: Gamma(k,theta) on top of Histogram

3.

3.5 Task E

Log Likelihood for Binomial= -2.1570681154346736
Log Likelihood for Gamma= -2.160821772204266
Binomial is a better fit for the given curve

3.6 Task F

1. Third Moment = 360.56586952543273
   Fourth Moment= 2968.068491427333

2. $\mu_1^* = 5.129607694312159$
   $p_1^* = 0.6118740341704901$
   $\mu_2^* = 8.774363054465352$
   $p_2^* = 0.38264565118317123$

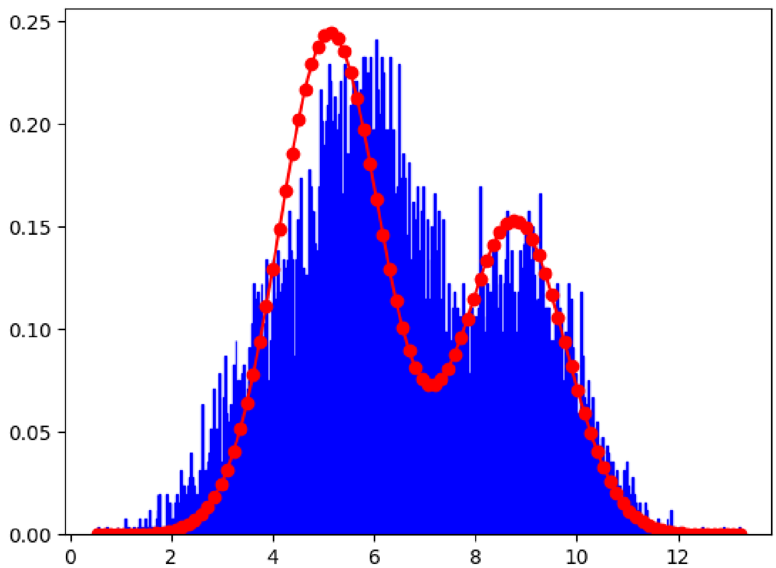

Figure 4: GMM on top of Histogram

3.

4. log-likelihood = -2.18303387449112685
   Hence this is worse than both binomial and gamma approximations

4 Quality in Inequalities

4.1 Task A

Markov's Inequality :

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

where X is any non-negative random variable and $a > 0$.
Intuitive Proof :
The expected value is like the weighted average of P(X=x) with the weight being x. Now consider all x greater than or equal to a, for this part of the sum, x will be at least a, taking this out of the summation ($E[X] \geq a \cdot P(X \geq a)$)

Proof for Continuous Random Variables:
Let $f(x)$ be the PDF of the given R.V X. Thus, $\forall a > 0$,

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

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

$$\implies a \cdot P[X \geq a] \leq \int_{-\infty}^\infty x \cdot f(x)\, dx$$

$$\implies a \cdot P[X \geq a] \leq E[X]$$

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

Hence Proved.

4.2 Task B

For a random variable X with mean μ, variance σ^2 and any $\tau > 0$, the Chebyshev-Cantelli inequality is:

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

Proof using Markov's inequality: Let us define a new R.V $Y = (X - \mu)^2$. This means that $E[Y] = \sigma^2$.

$$P[Y \geq \tau^2] \leq \frac{E[Y]}{\tau^2} = \frac{\sigma^2}{\tau^2} \qquad \text{By Markov's Inequality}$$

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

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

Now,

$$P[|X - \mu| \geq \tau] = P[X - \mu \geq \tau] + P[\mu - X \geq \tau]$$

Hence,

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

For Chebyshev-Cantell inequality, we need a tighter bound on the probability. For this, we can find the inequality in terms of a function of an arbitrary variable v and minimize the function to find the tighter bound.

$\forall v > 0$,

$$P[X - \mu \geq \tau] = P[X - \mu + y \geq \tau + y]$$

$$\implies P[X - \mu \geq \tau] \leq P[|X - \mu + y| \geq \tau + y]$$

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

$$P[X - \mu \geq \tau] \leq \frac{E[(X - \mu + y)^2]}{(\tau + y)^2} \qquad \text{By Markov's Inequality}$$

Let

$$g(y) = \frac{E[(X-\mu+y)^2]}{(\tau+y)^2} = \frac{E[(X-\mu)^2] + E[y^2] + E[2\cdot(X-\mu)\cdot y]}{(\tau+y)^2} \qquad \text{By Linearity of Expectation}$$

$$\implies g(y) = \frac{\sigma^2 + y^2 + 2y(E[X]-\mu)}{(\tau+y)^2} = \frac{\sigma^2+y^2}{(\tau+y)^2}$$

If the inequality holds $\forall y > 0$, then it should naturally hold for the minimum value of $g(y)$. On differentiating $g(y)$ w.r.t y and equating to zero,

$$g'(y) = \frac{2ty - 2\sigma^2}{(\tau+y)^3} = 0$$

$$\implies y = \frac{\sigma^2}{t}$$

So,

$$g(y)_{min} = \frac{\sigma^2}{\sigma^2+\tau^2}$$

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

Hence Proved.

4.3 Task C

For a Random Variable X with moment-generating function $M_X(t) = E[e^{tX}]$,

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

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

Proof:

We define a new random variable $Y = e^{tX}$. Choose an $x \in \mathbb{R}$. We know that $\forall x$, $e^{tx} > 0$. Thus, for any $t \in \mathbb{R}$,

$$P[Y \geq e^{tx}] \leq \frac{E[Y]}{e^{tx}} = \frac{M_X(t)}{e^{tx}} \qquad \text{By Markov's Inequality}$$

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

Inequality remains unchanged above as e^x is a strictly increasing function.
Now we have two cases:

1. $t > 0$:
$$P[X \geq x] \leq e^{-tx} \cdot M_X(t)$$
as inequality remains unchanged by dividing both sides by a positive number.

2. $t < 0$:
$$P[X \leq x] \leq e^{-tx} \cdot M_X(t)$$
as inequality is flipped when both sides are divided by a negative number.

Hence Proved.

4.4 Task D

Y is a random variable s.t. $Y = \sum_{x=1}^{n} X_i$, where $\{X_i | 1 \leq i \leq n\}$ is a set of independent and identically distributed random variables (Bernoulli in this case) with $E[X_i] = p_i$.

4.4.1 Part 1

Expectation of Y:

$$E[Y] = E[\sum_{x=1}^{n} X_i] = \sum_{x=1}^{n} E[X_i] \qquad \text{By linearity of expectation}$$

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

4.4.2 Part 2

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

where $\mu = E[Y]$.

Proof: $\forall t > 0$, by Markov's inequality

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

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

Since X_i are independent,

$$E[\prod_{i=1}^{n} e^{tX_i}] = \prod_{i=1}^{n} E[e^{tX_i}]$$

This can be derived by equating the correlation-coefficient formula to zero for independent random variables.

For a Bernoulli R.V,

$$M_X(t) = E[e^{tX}] = pe^t + (1-p)e^0 = 1 + p(e^t - 1)$$

We know that $1 + x \leq e^x$. This can be proved by graph comparison or differentiation. So,

$$E[e^{tX}] = 1 + p(e^t - 1) \leq e^{p(e^t-1)}$$

Now,

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

Coming back to the inequality,

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

Hence Proved.

4.4.3 Part 3

We can improve this bound further by minimizing the RHS w.r.t. t but note that $t > 0$. Let $g(t) = e^{\mu(e^t-1)-t(1+\delta)\mu}$. On differentiating,

$$g'(t) = e^{\mu(e^t-1)-t(1+\delta)\mu} \cdot (\mu e^t - \mu(1+\delta)) = 0$$

$$\implies e^t = 1+\delta$$

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

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

by plugging $t = \ln(1+\delta)$ into the above inequality.
This stronger bound is known as the "Chernoff" Bound.

4.5 Task E

Weak Law of Large Numbers:
Define $A_n = \frac{\sum_{i=1}^n X_i}{n}$ where X_i are i.i.d random Bernoulli variables, each with mean μ. Then, for all $\epsilon > 0$,

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

Proof:

$$P[|A_n - \mu| > \epsilon] = P[A_n > \epsilon + \mu] + P[A_n < \mu - \epsilon]$$

We prove both these probabilities tend to zero as n tends to infinity and thus their sum tends to zero as well.

We define a new R.V. $Y = \sum_{i=1}^n X_i = nA_n$. So, $\mu_y = n\mu$

1. **Case 1:** $\lim_{n\to\infty} P[A_n > \epsilon + \mu] = 0$
   We apply the Chernoff bound for Y, where $\mu_y = n\mu$ and $\delta = \epsilon/\mu$.

$$P[Y > n(\epsilon+\mu)] = P[Y > n\mu(\epsilon/\mu + 1)] \leq \frac{e^{n\epsilon}}{e^{n(\mu+\epsilon)\ln(1+\epsilon/\mu)}}$$

$$\implies P[Y > n(\epsilon+\mu)] \leq e^{n\epsilon - n(\mu+\epsilon)\ln(1+\epsilon/\mu)} = e^{n(\epsilon(1-\ln(1+\epsilon/\mu)) - \mu\ln(1+\epsilon/\mu))}$$

2. **Case 2:**$\lim_{n\to\infty} P[A_n < \mu - \epsilon] = 0$ From Task C, we know that,

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

$$\implies P[Y < n(\mu - \epsilon)] \leq e^{-tn(\mu-\epsilon)} \cdot M_Y(t)$$

   Now,

$$M_Y(t) = E[e^{tY}] = E[e^{t(\sum_{i=1}^n X_i)}]$$

$$M_Y(t) = \prod_{i=1}^n (E[e^{tX_i}]) = \prod_{i=1}^n (\mu(e^t - 1) + 1) = (\mu(e^t-1)+1)^n$$

Now,

$$P[Y < n(\mu - \epsilon)] \leq e^{-tn(\mu-\epsilon)} \cdot (\mu(e^t - 1) + 1)^n$$

Here,

$$\lim_{n\to\infty} e^{-tn(\mu-\epsilon)} \cdot (\mu(e^t - 1) + 1)^n = 0$$

because

$$\lim_{n\to\infty} (\mu(e^t - 1) + 1)^n = 0$$

as $(\mu(e^t - 1) + 1) < 1$

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

Thus, by combining both cases,

$$P[|A_n - \mu| > \epsilon] = 0$$

5 A Pretty "Normal" Mixture

5.1 Task A

Proving that $\forall u \in \mathbb{R}, f_{\mathcal{A}}(u) = f_X(u)$:
Probability of $\mathcal{N}(\mu_i, \sigma_i^2)$ getting picked is p_i. If the i^{th} Gaussian Model is chosen the probability of u getting chosen from it will be $\mathcal{N}(u, \mu_i, \sigma_i^2) = P[X_i = u]$. Hence the overall probability of $f_{\mathcal{A}}(u) = \Sigma_{i=1}^K p_i \cdot P[X_i = u] = f_X(u)$

5.2 Task B

X is a Gaussian Mixture Model.

5.2.1 Part 1

$$E[X] = \sum_x x \cdot P[X = x] = \sum_x x \cdot (\sum_{i=1}^k p_i \cdot P[X_i = x])$$

$$\implies E[X] = \sum_{i=1}^k p_i \cdot (\sum_x x \cdot P[X_i = x])$$

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

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

5.2.2 Part 2

$$Var[X] = E[X^2] - (E[X])^2$$

Now,

$$E[X^2] = \sum_x x^2 \cdot P[X = x] = \sum_x x^2 \cdot (\sum_{i=1}^{k} p_i \cdot P[X_i = x])$$

$$\implies E[X^2] = \sum_{i=1}^{k} p_i \cdot (\sum_x x^2 \cdot P[X_i = x])$$

$$\implies E[X^2] = \sum_{i=1}^{k} p_i E[X_i^2]$$

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

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

5.2.3 Part 3

For any Gaussian R.V. A, $M_A(t) = e^{t\mu_A + t^2\sigma_A^2/2}$.
Thus,

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

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} \cdot P[X = x] = \sum_x e^{tx} \cdot (\sum_{i=1}^{k} p_i \cdot P[X_i = x])$$

$$\implies E[e^{tX}] = \sum_{i=1}^{k} p_i \cdot (\sum_x e^{tx} \cdot P[X_i = x])$$

$$\implies E[e^{tX}] = \sum_{i=1}^{k} p_i E[e^{tX_i}]$$

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

5.3 Task C

Z is a weighted sum of k independent Gaussian variables.

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

where each $X_i \sim \mathbb{N}(\mu_i, \sigma_i^2)$.

5.3.1 Part 1

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

5.3.2 Part 2

To calculate this, we use the results from Task C, part 4 and Task D. Please read those sections first.

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

5.3.3 Part 3

To calculate this, we use the results from Task C, part 4 and Task D. Please read those sections first.

$$f_Z(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(u-\mu)^2}{2\sigma^2}}$$

5.3.4 Part 4

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

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

as $X_i (0 \leq i \leq k)$ are independent variables.
Now, we know that, for any Gaussian R.V. A, $M_A(t) = e^{t\mu_A + t^2\sigma_A^2/2}$.
Hence,

$$M_Z(t) = \prod_{i=1}^{k} e^{t\mu_i + t^2\sigma_i^2/2}$$

$$\implies M_Z(t) = e^{t(\sum_{i=1}^{k} p_i\mu_i) + t^2(\sum_{i=1}^{k} p_i^2\sigma_i^2)/2}$$

From the theorem proved in Task D, we can say that for a random variable X, its MGF and PDF uniquely determine each other. This means that different MGFs of a particular distribution share resemblance.
The MGF of Z is of the same form as the MGF of a Gaussian random variable. Hence, we can conclude that $\underline{Z \text{ is a Gaussian distribution}}$.

From this, we can derive the values of mean and variance of Z directly by comparing terms with the equation:

$$M_Z(t) = e^{t\mu + t^2\sigma^2/2}$$

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

$$\therefore \sigma^2 = \sum_{i=1}^{k} p_i^2 \sigma_i^2$$

$$\therefore f_Z(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(u-\mu)^2}{2\sigma^2}}$$

5.3.5 Part 5

Although X and Z have the same formula for mean, they differ in terms of Variance and MGF.
As the MGF uniquely determines R.V.'s PDF, we can conclude that X and Z have different properties.

5.3.6 Part 6

Z follows the Gaussian(Normal) Distribution as shown in Task C Part 4.

5.4 Task D

Theorem: For a random variable X, either finite and discrete or continuous, and its MGF is known for some (non-infinitesimal) interval, then its MGF and PDF uniquely determine each other.

Proof for discrete variables:

1. MGF uniquely determines PDF:

   Let us say two finite, discrete R.Vs X and Y have different MGFs but the same PMF. Let the union of their sample spaces be $S = \{s_1, s_2, s_3 \dots s_n\}$.
   So,
   $$P[X = s] = P[Y = s] \quad \text{for all } s \in S$$
   The MGFs are:
   $$M_X(t) = \sum_{s \in S} P[X = s] e^{ts}$$
   $$M_Y(t) = \sum_{s \in S} P[Y = s] e^{ts}$$
   The MGFs are equal because the coefficients of all the e^{ts} terms are the same for all $s \in S$. $\therefore$ **Contradiction**.
   Thus, MGF uniquely determines PDF.

2. PDF uniquely determines MGF:

   Let us say two finite, discrete R.Vs X and Y have different PMFs but the same MGF. Let the union of their sample spaces be $S = \{s_1, s_2, s_3 \dots s_n\}$.
   So,
   $$M_X(t) = M_Y(t) \quad \forall t \in \mathbb{R}$$
   $$\implies \sum_{s \in S} P[X = s] e^{ts} = \sum_{s \in S} P[Y = s] e^{ts}$$
   $$\implies \sum_{s \in S} e^{ts} (P[X = s] - P[Y = s]) = 0$$
   This is possible only if $P[X = s] = P[Y = s]$ for all $s \in S$. Thus, the PMF of X and Y must be same. $\therefore$ **Contradiction**.
   Therefore, PMF(or PDF) uniquely determines MGF

To conclude, the theorem stands valid.