CS215 Assignment 2

Group Members :

B.Abhinav 23B1018

G.Abhiram 23B1084

U.Sai Likhith 23B1058

September 9, 2024

Contents

1 Mathemagic

1.1 Task A

Solution:
Lets derive the PGF for $X \sim \text{Ber}(p)$. PMF for Ber(p) is given by :

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

G(z) for Bernoulli random variable is given by:

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

$$G_{Ber} \;=\; (1-p)z^0 + pz^1 \;=\; 1 - p + pz \tag{1}$$

1.2 Task B

Solution:
PMF of Bin(n, p) is given by:

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

G(z) for Binomial random variable is given by:

$$G_{Bin} \;=\; \sum_{k=0}^{n} P[X \;=\; k]z^k$$

$$G_{Bin} \;=\; \sum_{k=0}^{n} \binom{n}{k} (zp)^k (1-p)^{n-k}$$

Using Binomial Theorem :

$$G_{Bin} \;=\; (1 - p + zp)^n \tag{1}$$

From G_{Ber} in Task A:

$$G_{Bin} \;=\; G_{Ber}(z)^n \tag{2}$$

1.3 Task C

Solution:
$X_1, X_2, X_3, \ldots, X_k$ are independent non-negative-integer-valued random variables, each distributed with the same probability mass function P and have a common PGF G.

$$X = X_1 + X_2 + X_3 + \cdots + X_k$$

PGF of X is given by :

$$G_\Sigma(z) = E(z^X) = E(z^{X_1+X_2+X_3+\cdots+X_k})$$

$$G_\Sigma(z) = E(z^{X_1} z^{X_2} z^{X_3} \ldots z^{X_k})$$

Using the fact that when a, b are 2 independent random variables we have:

$$E(f(a)g(b)) = E(f(a))E(g(b)) \tag{1}$$

The proof is as follows:

$$E(f(a)g(b)) = \sum_{a_i \in S(a)} \sum_{b_j \in S(b)} P_a[a = a_i] P_b[b = b_j] a_i b_j$$

As they are independent ie a,P(a) do not depend on b we get:

$$E(f(a)g(b)) = \sum_{a_i \in S(a)} P_a[a = a_i] a_i \sum_{b_j \in S(b)} P_b[b = b_j] b_j$$

This proves the result. As all X_i are independent we have :

$$G_\Sigma(z) = E(z^{X_1})E(z^{X_2})E(z^{X_3}) \ldots E(z^{X_k})$$

As all of them have common PGF G we get:

$$G_\Sigma(z) = G(z)^k \tag{2}$$

1.4 Task D

Solution:
PMF of Geo(p) is as follows:

$$P[X \ = \ n] \ = \ (1-p)^{n-1}p$$

where X = 1, 2, 3.... . PGF of Geo(p) is :

$$\sum_{n=1}^{\infty} P[X=n]z^n \ = \ \sum_{n=1}^{\infty}(1-p)^{n-1}pz^n$$

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

This summation only converges when:

$$|z(1-p)| < 1 \Rightarrow |z| < \frac{1}{|1-p|} \tag{1}$$

Using the formula for sum of infinite GP:

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

under the condition (1) Simplifies to:

$$G_{Geo} \ = \ \frac{zp}{1-z+zp} \tag{2}$$

1.5 Task E

Solution:
Consider $X \sim \text{Bin}(n,p)$ and $Y \sim \text{NegBin}(n,p)$.
Let us compute the PGF for Y using the result (2) from Task C. We know that NegBin(n, p) random variable can be written as sum of n independent Geo(p) random variables.

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

Using (2) from Task D with p and z replaced with p^{-1} and z^{-1} We ge :

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

This simplifies to:

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

Inverse on both sides and using (1) from Task B:

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

Replacing p with p^{-1} and z^{-1} we prove the result and inverting:

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

Note that z is bounded by condition (1) in Task D.

1.6 Task F

Solution:
PMF of $Y \sim \text{NegBin}(n,p)$ is given by(for k = n, n+1, ...):

$$P[Y \;=\; k] \;=\; \binom{k-1}{n-1}p^n(1-p)^{k-n} \tag{1}$$

The RHS of the equation (1) in Task E simplifies to:

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

The PGF of NegBin(n, p) by definition was:

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

Equate both of them and use k = n + r substitution in LHS:

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

Cancelling the common terms this simplifies to:

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

We have proved in Task D that PGF for Y is only defined under $|z(1-p) < 1$. Substitute $x = z(1-p)$ under $|x| < 1$ to get:

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

Using

$$\binom{n}{r} = \binom{n}{n-r}$$

and substituting x with $-x$ we get:

$$(1+x)^{-n} = \sum_{r=0}^{\infty} (-1)^r \binom{n+r-1}{r} x^r \qquad (2)$$

From the general definition of $\binom{n}{r}$ we have:

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

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

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

Using this in (2) we get
For $n \in N$ and $|x| < 1$:

$$(1+x)^{-n} = \sum_{r=0}^{\infty} (-1)^r \binom{n+r-1}{r} x^r = \sum_{r=0}^{\infty} \binom{-n}{r} x^r \qquad (3)$$

that is the Binomial Theorem for negative exponent.

1.7 Task G

Solution:
We have for a random variable X with PGF G(z)

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

$$G'(z) \; = \; \sum_{n=0}^{\infty} nz^{n-1}P[X=n]$$

Substituting z = 1 we get:

$$E(X) \; = \; \sum_{n=0}^{\infty} nP[X=n] \; = \; G'(1) \tag{1}$$

Mean for Bern(p):

$$E_{Ber} \; = \; G'_{Ber}(1), \; G_{Ber}(z) = 1 - p + zp \Rightarrow G'_{Ber}(z) = p$$

Hence :

$$E_{Ber} \; = \; p \tag{2}$$

Mean for Bin(n, p)

$$G_{Bin}(z) \; = \; (1-p+zp)^n \Rightarrow G'_{Bin}(z) = np(1-p+zp)^{n-1} \Rightarrow G'_{Bin}(1) = np$$

Hence

$$E_{Bin} \; = \; np \tag{3}$$

Mean for Geo(p)

$$G_{Geo}(z) \; = \; \frac{zp}{1-z+zp} \Rightarrow G'_{Geo}(z) = \frac{(1-z+zp)(p) - (zp)(p-1)}{(1-z+zp)^2}$$

$$G'_{Geo}(z) \; = \; \frac{p}{(1-z+zp)^2} \Rightarrow G'_{Geo}(1) = \frac{1}{p}$$

Hence

$$E_{Geo}(z) \; = \; \frac{1}{p} \tag{4}$$

Mean for NegBin(n, p)

$$G_{NegBin}(z) \; = \; G_{Geo}(z)^n \Rightarrow G'_{NegBin}(z) = nG_{Geo}(z)^{n-1}G'_{Geo}(z)$$

$$G'_{NegBin}(1) \; = \; \frac{n}{p} \quad as \;\; G_{Geo}(1) = 1$$

Hence

$$E_{NegBin} = \; \frac{n}{p} \tag{5}$$

2 Normal Sampling

2.1 Task A

Solution:
It suffices to prove that $f_Y(y) = 1$, Let us prove the following theorem first: Given random variables X and Y with PDFs $f_X(x)$ and $f_Y(y)$, under an invertible transformation $Y = g(X)$,

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

Proof:
Let $x_1, x_2, y_1, y_2 \in \mathbb{R}$ such that $y_1 = g(x_1)$ and $y_2 = g(x_2)$. We have assuming g is non decreasing::

$$P(x_1 \leq X \leq x_2) = P(y1 \leq Y \leq y2)$$

Since g is invertible, The probability of the event must be the same whether expressed in terms of X or Y. This can be written as:

$$\int_{x_1}^{x_2} f_X(x)\,dx = \int_{y_1}^{y_2} f_Y(y)\,dy$$

Using the transformation $y = g(x)$, we know that:

$$x = g^{-1}(y) \; and \; dx = (g^{-1})'(y)dy$$

Thus, the integral becomes:

$$\int_{y_1}^{y_2} f_Y(y)\,dy = \int_{y_1}^{y_2} f_X(g^{-1}(y))(g^{-1})'(y)dy$$

As the above is true $\forall y1, y2 \in R$, The integrands must be equal.This can also be shown by treating them as functions and differentiating both integrals.(Second Fundamental Theorem of Calculus)

$$f_Y(y) \;=\; f_X(g^{-1}(y)) \cdot (g^{-1})'(y)$$

From calculus we have $(g^{-1})'(y) = \frac{1}{g'(g^{-1}(y))}$. It follows that:

$$\boxed{f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}}$$

The modulus is accounting for the case when g is non increasing(Same proof with -ve sign) as PDF's are always non negative. Coming back to our question the invertible transformation g is F_X the CDF of X.

$$g(x) = \int_{-\infty}^{x} f_X(x)dx \;\Rightarrow\; g'(x) = f_X(x)$$

Therefore here we have $g'(g^{-1}(y)) = f_X(g^{-1}(y))$ and $g' \geq 0$ as CDF is non decreasing, which concludes through the above result :

$$f_Y(y) = 1$$

Therefore we proved that Y is uniformly distributed in $[0, 1]$.

2.2 Task B

From the theorem we have proved in Task A we can sample a standard normal distribution. Let X be a standard normal disrtribution. We have

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Since $F_X'(x) = f(x) > 0$, $F_X(x)$ is invertible. We know that $Y = F_X(X)$ is a uniform distribution in [0,1] from Task A.
Our algorithm $\mathcal{A}$ invloves computing the inverse $G_X(Y)$ of $F_X(X)$ and then mapping the uniform distribution Y using it.
Using this we can sample a standard normal distribution from a uniform distribution Y in [0,1].
This also implies that $\forall u \in R$ we have

$$F_{\mathcal{A}}(u) = F_X(u) \; and \; f_{\mathcal{A}}(u) = f_X(u)$$

since we are essentially recreating the distribution through a one one correlation to Y.

2.3 Task C

Check the code in 2c.py

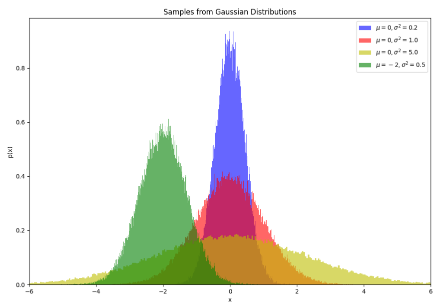

Figure 1: The Bell curve

2.4 Task D

Check the code in 2d.py

The shape of the graph, which appears to be approximately bell-shaped for large depth(h), suggests that the distribution of the final positions of the balls is similar to normal distribution.

This is because :

1. The majority of the balls tend to fall in the middle pockets.
2. The number of balls in each pocket decreases as you move away from the middle.
3. The distribution is symmetric around the middle.

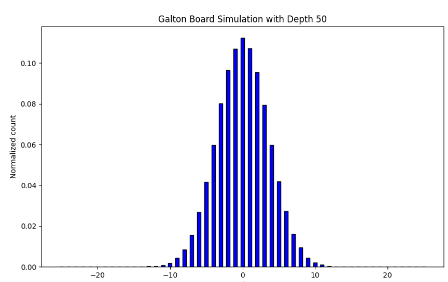

Figure 2: Plot for h = 50 and N = 100000

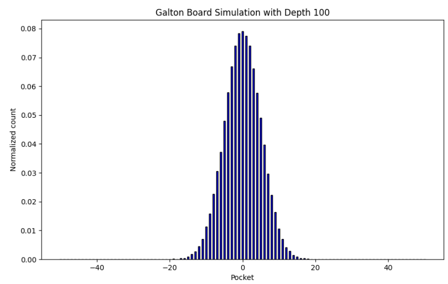

Figure 3: Plot for h = 100 and N = 100000

2.5 Task E(B)

Part A

Lets compute $P_h[X = 2i]$. For each collision we have equally probable outcomes 1 or -1. For landing in pocket 2i let the number of 1's and -1's be m, n. We have $m+n = 2k \; and \; m-n = 2i$. Solving them gives $m = k+i, \; n = k-i$. We need to chose $k + i$ collisions and give them 1 ie the probability will be given by:

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

Part B

We will use stirling's approximation on $(2k)!, (k+i)!, (k-i)!$. The approximation being:

$$n! \;\; = \;\; \sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}$$

$$\begin{aligned} P_h[X = 2i] &= \binom{2k}{k+i}\left(\frac{1}{2}\right)^{2k} \\ &= \frac{(2k)!}{(k+i)!(k-i)!(2^{2k})} \\ &= \frac{\sqrt{2\pi 2k}\left(\frac{2k}{e}\right)^{2k}}{\sqrt{2\pi(k+i)}\left(\frac{k+i}{e}\right)^{k+i}\sqrt{2\pi(k-i)}\left(\frac{k-i}{e}\right)^{k-i}2^{2k}} \\ &= \frac{1}{\sqrt{\pi}}\sqrt{\frac{k}{k^2-i^2}}\left(\frac{k}{k+i}\right)^{k+i}\left(\frac{k}{k-i}\right)^{k-i} \Rightarrow approx \; k^2 - i^2 \approx k^2 \; as \; k >> i^2 \end{aligned} \tag{2}$$

We are going to use the The infinite series expansion of $log(1+x)$ that is:

$$log(1+x) \;=\; x - \frac{x^2}{2} + \frac{x^3}{3} \ldots$$

$$\begin{aligned} P_h[X=2i] &= \frac{1}{\sqrt{\pi k}} \frac{1}{(1+\frac{i}{k})^{k+i}(1-\frac{i}{k})^{k-i}} \\ &= \frac{1}{\sqrt{\pi k}} \left(\frac{1}{e^{(k+i)log(1+\frac{i}{k})+(k-i)log(1-\frac{i}{k})}} \right) \end{aligned} \tag{3}$$

The exponent simplifies to using the series expansion and ignoring higher order terms(that is $(\frac{i^2}{k})^2$ and above as we have been given $\frac{i^2}{k} << 1$).

$$\begin{aligned} (k+i)log(1+\frac{i}{k}) + (k-i)log(1-\frac{i}{k}) &= (k+i)(\frac{i}{k} - \frac{i^2}{2k^2} + \frac{i^3}{3k^3}...) - (k-i)(\frac{i}{k} + \frac{i^2}{2k^2} + \frac{i^3}{3k^3}...) \\ &= \frac{i(k+i-k+i)}{k} + \frac{i^2(-k+i-k-i)}{2k^2} + \frac{i^3(2i)}{3k^3}... \\ &\approx \frac{2i^2}{k} - \frac{i^2}{k} \;=\; \frac{i^2}{k} \end{aligned} \tag{4}$$

Substituting in 3 we get:

$$\begin{aligned} P_h[X=2i] &= \frac{1}{\sqrt{\pi k}} e^{-\frac{i^2}{k}} \\ P_h[X=i] &= \frac{1}{\sqrt{\pi k}} e^{-\frac{i^2}{4k}} \end{aligned} \tag{5}$$

Which is clearly the normal distribution $N(\mu = 0, \sigma^2 = 2k)$.

3 Fitting Data

3.1 Task A

Solution:

Check the code in 3.ipynb

3.2 Task B

<u>Solution:</u>

Check the code in 3.ipynb

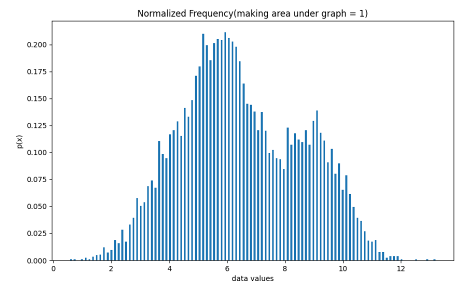

Figure 4: Histogram

3.3 Task C

<u>Solution:</u>

Check the code in 3.ipynb

<u>Proof:</u>

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

Given,

$$\begin{aligned} \mu_1 &= E[X] \\ \mu_2 &= E[X^2] \end{aligned} \tag{1}$$

As we know that, for binomial distribution

$$\begin{aligned} E[X] &= E[\sum_{i=1}^{n} Y_i] \\ &= \sum_{i=1}^{n} E[Y_i] \\ &= \sum_{i=1}^{n} p \end{aligned} \tag{2}$$

$$\mu_1 = E[X] = np$$

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

$$\begin{aligned} E[X^2] &= Var[X] + (E[X])^2 \\ &= np(1-p) + (np)^2 \\ &= np - np^2 + n^2p^2 \\ \mu_2 = E[X^2] &= np + n(n-1)p^2 \end{aligned} \tag{3}$$

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

3.4 Task D

Solution:

Check the code in 3.ipynb
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.

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x) dx \\ &= \int_0^\infty x(\frac{1}{\theta^k \tau(k)} x^{k-1} e^{-\frac{x}{\theta}}) dx \\ &= \frac{1}{\theta^k \tau(k)} \int_0^\infty x^k e^{-\frac{x}{\theta}} dx \\ &= \frac{1}{\theta^k \tau(k)} \int_0^\infty \theta^{k+1} t^k e^{-t} dt \\ &= \frac{1}{\theta^k \tau(k)} (\theta^{k+1} \tau(k+1)) \\ &= \frac{\theta^{k+1}(k\tau(k))}{\theta^k \tau(k)} \\ \mu_1^{Gamma} = E[X] &= k\theta \end{aligned} \tag{1}$$

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

$$\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_{0}^{\infty} x^2 (\frac{1}{\theta^k \tau(k)} x^{k-1} e^{-\frac{x}{\theta}})dx \\
&= \frac{1}{\theta^k \tau(k)} \int_{0}^{\infty} x^{k+1} e^{-\frac{x}{\theta}} dx \\
&= \frac{1}{\theta^k \tau(k)} \int_{0}^{\infty} \theta^{k+2} t^{k+1} e^{-t} dt \\
&= \frac{1}{\theta^k \tau(k)} (\theta^{k+2} \tau(k+2)) \\
&= \frac{\theta^{k+2}(k(k+1)\tau(k))}{\theta^k \tau(k)} \\
\mu_2 = E[X^2] &= k(k+1)\theta^2
\end{aligned} \tag{2}$$

3.5 Task E

Solution:

Check the code in 3.ipynb

3.6 Task F

Solution:

Check the code in 3.ipynb

4 Quality in Inequalities

Given, Markov's Inequality

$$P[X \geq a] \leq \frac{E[X]}{a} \tag{1}$$

Where, $X \geq 0, a > 0$

4.1 Task A

<u>Solution:</u>

Intuitive proof :
Imagine a jar containing a large number of balls(N_t) each having some weight. The random variable here is the weight of a ball.Now for some weight a, this is what we will do:
1) We will remove all balls with weight less than a,
2) We will replace all the balls with weight greater than a with a ball of weight a.
Obviously the weight of jar can never increase. ie $N_t E(x) \geq N_l a \Rightarrow E(x) \geq a * \frac{N_l}{N_t}$.($N_l$ is number of balls left.)

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

Rigorous proof :
Let X be discrete Random variable and x_i be in non decreasing order.
Let x_m be the smallest value such that $x_m \geq a$ To get $P[X \geq a]$:

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} x_i P[x_i] \\ &\geq \sum_{i=m}^{n} x_i P[x_i] \\ &\geq \sum_{i=m}^{n} a P[x_i] \\ &\geq a \sum_{i=m}^{n} P[x_i] \\ E[X] &\geq a P[X \geq a] \end{aligned} \tag{1}$$

Therefore, we get

$$P[X \geq a] \leq \frac{E[X]}{a} \tag{2}$$

We can similarly for continuous Random variables by limiting n to infinity.

4.2 Task B

To be proved :

$$P[X - \mu \geq \tau] \leq \frac{\sigma^2}{\sigma^2 + \tau^2} \forall \tau > 0 \tag{1}$$

Where, μ = mean of R.V, σ = standard deviance of R.V

Solution:
We can write the following

$$\begin{aligned} P[X - \mu \geq \tau] &= P[X - \mu + \alpha \geq \tau + \alpha] \\ &\leq P[(X - \mu + \alpha)^2 \geq (\tau + \alpha)^2] \end{aligned} \tag{2}$$

Where, $\alpha \geq -\tau$

we can apply the Markov's inequality to (2)

$$\begin{aligned} P[(X - \mu + \alpha)^2 \geq (\tau + \alpha)^2] &\leq \frac{E((X - \mu + \alpha)^2)}{(\tau + \alpha)^2} \\ &\leq \frac{E((X - \mu)^2 + 2(X - \mu)\alpha + \alpha^2}{(\tau + \alpha)^2} \\ &\leq \frac{E((X - \mu)^2) + 2E((X - \mu))\alpha + \alpha^2}{(\tau + \alpha)^2} \end{aligned} \tag{3}$$

$$P[(X - \mu + \alpha)^2 \geq (\tau + \alpha)^2] \leq \frac{\sigma^2 + \alpha^2}{(\tau + \alpha)^2} \tag{4}$$

$$P[X - \mu \geq \tau] \leq \frac{\sigma^2 + \alpha^2}{(\tau + \alpha)^2} \tag{5}$$

We can further decrease the range in (5) by differentiating with α. So that we get minimum value in right term.

$$\begin{aligned} \frac{d(\frac{\sigma^2 + \alpha^2}{(\tau + \alpha)^2})}{d\alpha} &= 0 \\ \frac{2\alpha(\tau + \alpha) - 2(\sigma^2 + \alpha^2)}{(\tau + \alpha)^3} &= 0 \\ 2\alpha(\tau + \alpha) &= 2(\sigma^2 + \alpha^2) \\ \alpha &= \frac{\sigma^2}{\tau} \end{aligned} \tag{6}$$

By substituting (6), we will get

$$(\frac{\sigma^2+\alpha^2}{(\tau+\alpha)^2})_{min} = \frac{\sigma^2}{\sigma^2+\tau^2} \tag{7}$$

So, from (7) and (5), we can conclude that

$$P[X-\mu \geq \tau] \leq \frac{\sigma^2}{\sigma^2+\tau^2} \tag{8}$$

4.3 Task C

To be proved :

$$P[X \geq x] \leq e^{-xt} M_X(t) \forall t > 0 \tag{1}$$

$$P[X \leq x] \leq e^{-xt} M_X(t) \forall t < 0 \tag{2}$$

Solution:
We can write MGF for X as

$$M_X(t) = \int_0^\infty e^{ty} P[X=y]dy \tag{3}$$

- **Case 1 :** t>0

  Similar to the Method used in Task A, we can extend it to continuous and do the following

$$\begin{aligned}
\int_0^\infty e^{ty} P[X=y]dy &\geq \int_x^\infty e^{ty} P[X=y]dy \\
\int_0^\infty e^{ty} P[X=y]dy &\geq e^{tx} \int_x^\infty P[X=y]dy \\
M_X(t) &\geq e^{tx} \int_x^\infty P[X=y]dy \\
e^{-tx} M_X(t) &\geq \int_x^\infty P[X=y]dy \\
e^{-tx} M_X(t) &\geq P[X \geq x]
\end{aligned} \tag{4}$$

  Hence, proved.

- **<u>Case 2 :</u>** t<0

  Similarly,

$$
\begin{aligned}
\int_0^\infty e^{ty} P[X=y]dy &\geq \int_0^x e^{ty} P[X=y]dy \\
\int_0^\infty e^{ty} P[X=y]dy &\geq e^{tx} \int_0^x P[X=y]dy \\
M_X(t) &\geq e^{tx} \int_0^x P[X=y]dy \\
e^{-tx} M_X(t) &\geq \int_0^x P[X=y]dy \\
e^{-tx} M_X(t) &\geq P[X \geq x]
\end{aligned} \tag{5}
$$

  Hence, proved.

4.4 Task D

<u>Solution:</u>
For, n i.i.ds, $X_1, X_2, \ldots, X_n$, which have Bernouli distribution with E[X_i]= p_i, Y can be written as $Y = \sum_{i=1}^{n} X_i$

1. **Expectation of Y**

   As all X's are independent and identically distributed, we can do as follows

$$
\begin{aligned}
E[Y] &= E[X_1 + X_2 + \cdots + X_n] \\
&= E[X_1] + E[X_2] + \cdots + E[X_n] \\
&= \sum_{i=1}^{n} p_i \\
&= np_i
\end{aligned} \tag{1}
$$

2. **Prove:**

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

Where, μ = expectation of Y, $\delta > -1, t > 0$

Using (1) of Task C, we can write that

$$\begin{aligned} P[Y \geq (1+\delta)\mu] &\leq e^{-t(1+\delta)\mu} M_Y(t), t > 0 \\ &\leq e^{-t(1+\delta)\mu} E[e^{Yt}] \\ &\leq e^{-t(1+\delta)\mu} E[e^{t\sum_{i=1}^{n} X_i}] \end{aligned} \tag{3}$$

As X's are i.i.d, we can write

$$\begin{aligned} P[Y \geq (1+\delta)\mu] &\leq e^{-t(1+\delta)\mu} \prod_{i=1}^{n} E[e^{tX_i}] \\ P[Y \geq (1+\delta)\mu] &\leq e^{-t(1+\delta)\mu} \prod_{i=1}^{n} (1 + p_i(e^t - 1)) \\ P[Y \geq (1+\delta)\mu] &\leq e^{-t(1+\delta)\mu} (1 + p_i(e^t - 1))^n \end{aligned} \tag{4}$$

Now using the inequality $1 + a < e^a, \forall a > 0$, we can write

$$\begin{aligned} e^{-t(1+\delta)\mu}(1 + p_i(e^t - 1))^n &\leq e^{-t(1+\delta)\mu} e^{p_i(e^t-1)n} \\ e^{-t(1+\delta)\mu}(1 + p_i(e^t - 1))^n &\leq e^{-t(1+\delta)\mu} e^{\mu(e^t-1)} \\ P[Y \geq (1+\delta)\mu] &\leq e^{-t(1+\delta)\mu} e^{\mu(e^t-1)} \end{aligned} \tag{5}$$

Therefore, we can conclude that

$$P[Y \geq (1+\delta)\mu] \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}} \tag{6}$$

3. **Improve the bound with appropriate t**

   Using (2), to improve bound, we need to minimise right term. We can do so by differentiating right term with t and equating it with 0.

$$\begin{aligned} \frac{d(\frac{e^{\mu(e^t-1)}}{e^{(1+\delta)t\mu}})}{dt} &= 0 \\ \mu(e^t) - (1+\delta)\mu &= 0 \\ e^t &= 1 + \delta \end{aligned} \tag{7}$$

From (3), we get t = ln$(1+\delta)$. By substituting it in (2), we get

$$P[Y \geq (1+\delta)\mu] \leq \frac{e^{\mu\delta}}{(1+\delta)^{(1+\delta)\mu}}$$
$$\frac{e^{\mu\delta}}{(1+\delta)^{(1+\delta)\mu}} = e^{\mu\delta-(1+\delta)ln(1+\delta)\mu} \tag{8}$$

- **Case 1:** $\delta > 0$
  We can use $ln(1+\delta) \geq \frac{2\delta}{2+\delta}$ and get:

  $$e^{\mu\delta-(1+\delta)ln(1+\delta)\mu} \leq e^{\mu\delta-(1+\delta)\frac{2\delta}{2+\delta}\mu} \tag{9}$$

  Therefore,
  $$P[Y \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2+\delta}} \tag{10}$$

- **Case 2:** $-1 < \delta < 0$
  Similarly, we can use $ln(1+\delta) \geq \frac{\delta(\delta+2)}{2(\delta+1)}$ and get:

  $$e^{\mu\delta-(1+\delta)ln(1+\delta)\mu} \leq e^{\mu\delta-(1+\delta)\frac{\delta(\delta+2)}{2(\delta+1)}} \tag{11}$$

  Therefore,
  $$P[Y \leq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}} \tag{12}$$

4.5 Task E

<u>Solution:</u>
To be Proved:
Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with each having mean μ.We define $A_n = \frac{\sum_{i=1}^n X_i}{n}$. Then for all $\epsilon > 0$,we have

$$\lim_{n\to\infty} P[|A_n - \mu| > \epsilon] = 0 \tag{1}$$

First lets remove modulus and bound it to the minimum range(just like chernoff's bound)

$$P[|A_n - \mu| > \epsilon] = P[A_n - \mu > \epsilon] + P[A_n - \mu < -\epsilon] \tag{2}$$

Lets first assume the sum $S_n = \sum_{i=1}^n X_i$, so $A_n = \frac{S_n}{n}$

- First lets focus on $P[A_n > \mu + \epsilon]$

$$P[A_n > \mu + \epsilon] = P[S_n > n(\mu + \epsilon)] \leq e^{-tn(\mu+\epsilon)} E[tS_n] \tag{3}$$

Since $X_1, X_2, \dots, X_n$ are independent , the M.G.F of S_n is a product of M.G.Fs of $X_i's$

$$\begin{aligned} E[e^{tS_n}] &= (E[e^{tX_i}])^n \\ P(S_n > n(\mu + \epsilon)) &\leq (e^{-t(\mu+\epsilon)} E[e^{tX_i}])^n \end{aligned} \tag{4}$$

To get the minimum range in the bound,we need to minimize the R.H.S W.R.T t.

$f(t) = e^{-t(\mu+\epsilon)} E[e^{tX_i}]$

Here we don't need the exact value of t because the limits we are applying are on 'n' whereas the function f(t) doesnt depend on n, we just need to know that for some choice of t:

$$P(S_n > n(\mu + \epsilon)) \leq e^{-cn} \tag{5}$$

Where, c is a constant W.R.T n.

- Similarly for $P[A_n - \mu < -\epsilon]$:

$$\begin{aligned} P[A_n < \mu - \epsilon] &= P[S_n < n(\mu - \epsilon)] \\ P[S_n < n(\mu - \epsilon)] &\leq e^{-cn} \end{aligned} \tag{6}$$

Now by combining the bounds:

$$P[|A_n - \mu| > \epsilon] \leq (e^{-cn} + e^{-cn} = 2e^{-cn}) \tag{7}$$

Now applying $\lim_{n\to\infty}$ on B.S:

$$\lim_{n\to\infty} P[|A_n - \mu| > \epsilon] \leq \lim_{n\to\infty} 2e^{-cn} \tag{8}$$

As $n \to \infty$, $\lim_{n\to\infty} 2e^{-cn} \to 0$ and we know that $P[|A_n - \mu| > \epsilon] \geq 0$
Therefore:

$$\lim_{n\to\infty} P[|A_n - \mu| > \epsilon] = 0 \tag{9}$$

5 A Pretty "Normal" Mixture

5.1 Task A

Solution: Lets find the pdf for random variable $\mathcal{A}$, that is $f_{\mathcal{A}}(x)$.
Lets find the probability that our algorithm samples a value between $x \; and \; dx$.
The probability for this is $f_{\mathcal{A}}(x)dx$.
It can also be computed using conditional probability .Let $P(X_i)$ be probability that X_i is chosen in first step and $f_i(x)$ be the PDF of X_i.As all such events(that a specific X_i is selected) are mutually exclusive probability can be added.

$$\begin{aligned} f_{\mathcal{A}}(x)dx &= \sum_{i=0}^{k} P(X_i)P\left(\frac{X_i = x \; to \; dx}{X_i}\right) \\ f_{\mathcal{A}}(x)dx &= \sum_{i=0}^{k} p_i f_i(x)dx \\ f_{\mathcal{A}}(x) &= \sum_{i=0}^{k} p_i f_i(x) \end{aligned}$$

From question $f_X(x) = \sum_{i=0}^{k} p_i f_i(x)$. Therefore $\forall u \in \mathcal{R}$

$$f_X(u) = f_{\mathcal{A}}(u) \tag{1}$$

5.2 Task B

Solution:

1. **E[X]**

   As X is GMM sampled, we can write it's expectation as

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

   Also, we know that

   $$E[X_i] = \mu_i \tag{2}$$

From (1) and (2),

$$\mu = E[X] = \sum_{i=1}^{k} p_i \mu_i \tag{3}$$

2. **Var[X]**

   For a GMM, acc to Law of Total Variance, Var[X] can be written as

$$\begin{aligned} Var[X] &= E[Var[X_i]] + Var[E[X_i]] \\ &= \sum_{i=1}^{k} p_i Var[X_i] + Var[\sum_{i=1}^{k} p_i \mu_1 | \mu] \end{aligned} \tag{4}$$

   For the second term,

$$Var[\sum_{i=1}^{k} p_i \mu_i | \mu] = \sum_{i=1}^{k} p_i (\mu_i - \mu)^2 \tag{5}$$

   From (4) and (5),

$$Var[X] = \sum_{i=1}^{k} p_i \sigma_i^2 + \sum_{i=1}^{k} p_i (\mu_i - \mu)^2 \tag{6}$$

3. **The MGF, M(t) of X**

   MGF of a GMM is

$$M_X(t) = \sum_{i=1}^{k} p_i M_{X_i}(t) \tag{7}$$

   For a Normal Guassian variable,

$$\begin{aligned} M_{X_i}(t) &= E[e^{tX_i}] \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}} dx \\ &= e^{t\mu_i + \frac{1}{2}t^2\sigma_i^2} \frac{1}{\sqrt{2\pi}\sigma_i} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu_i+\sigma_i^2 t))^2}{2\sigma_i^2}} dx \\ M_{X_i}(t) &= e^{t\mu_i + \frac{1}{2}t^2\sigma_i^2} \end{aligned} \tag{8}$$

From (7) and (8),

$$M_X(t) = \sum_{i=1}^{k} p_i e^{t\mu_i + \frac{1}{2}t^2\sigma_i^2} \tag{9}$$

5.3 Task C

Solution: Given random variable Z as

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

1. **E[Z]**

   We can write expectation of Z similar to the GMM above.

$$\begin{aligned} E[Z] &= E[\sum_{i=1}^{k} p_i X_i] \\ &= \sum_{i=1}^{k} p_i E[X_i] \end{aligned} \tag{2}$$

   Therefore,

$$E[Z] = \sum_{i=1}^{k} p_i \mu_i \tag{3}$$

2. **Var[Z]**

   acc to the formula for variance calculation of sum of independent random variables

$$\begin{aligned} Var[Z] &= p_1^2 Var[X_1] + p_2^2 Var[X_2] + \dot{+} p_k^2 Var[X_k] \\ &= \sum_{i=1}^{k} p_i^2 \sigma_i^2 \end{aligned} \tag{4}$$

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

   As Z is a Normal Gaussian Distribution which we can say by comparing MGF, PDF of Z is

$$f_Z(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-\mu)^2}{2\sigma^2}} \tag{5}$$

Where, $\mu = \sum_{i=1}^{k} p_i\mu_i, \sigma^2 = \sum_{i=1}^{k} p_i^2\sigma_i^2$

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

   For MGF,

$$\begin{aligned} M_Z(t) &= E(e^{Zt}) \\ &= E(e^{(p_1X_1+p_1X_2+\cdots+p_kX_k)t}) \\ &= E(e^{p_1X_1t})E(e^{p_2X_2t})\ldots E(e^{p_2X_kt}) \\ &= \prod_{i=1}^{k} e^{tp_i\mu_i+\frac{1}{2}p_i^2\sigma_i^2t^2} \\ &= e^{t\sum_{i=1}^{k} p_i\mu_i+\frac{1}{2}t^2\sum_{i=1}^{k} p_i^2\sigma_i^2} \end{aligned} \tag{6}$$

$$M_Z(t) = e^{t\mu+t^2\frac{1}{2}\sigma^2} \tag{7}$$

   Where, $\mu = \sum_{i=1}^{k} p_i\mu_i, \sigma^2 = \sum_{i=1}^{k} p_i^2\sigma_i^2$

   Here, MGF is similar to Normal Gaussian randam variables.

5. **Conclusion**

   Weighted sum of Normal Gaussian variables, Z, is different from GMM, X. They have same expectation but remaining properties are different.

   Therefore, we can conclude that Z and X have different properties.

6. **Type of Distribution**

   Z follows Normal Gaussian Distribution.

5.4 Task D(B)

Solution:

For a finite discrete random variable X,we need to show that its MGF and PDF uniquely determine each other.
To do so, we need to complete proving 2 parts,i.e., given a PDF, there exist a unique MGF and vice-versa.

Let's take a finite discrete random variable X with PDF, P[X=x_i] = p_i, and MGF, M_X(t) = $\sum_{i=1}^{n} e^{x_it}p_i$.

Proof of uniqueness :

1. **PDF to MGF :**

   It is simpler to prove this direction.

   - For a given PDF P, we can always write MGF,M_X(t), as $\sum_{i=1}^{n} e^{x_i t} p_i$.
   - The MGF can be uniquely defined for a particular X and its corresponding PDF,P.

2. **MGF to PDF :**

   Now let's take two random variables X, Y with $M_X(t) = M_Y(t)$.

   - We can write $M_X(t)$ as $\sum_{i=1}^{n} e^{x_i t} p(x_i)$ and $M_Y(t)$ as $\sum_{i=1}^{n} e^{y_i t} p(y_i)$.
   - Using Taylor series we can split them into powers of t as follows:

   $$M_X(t) = \sum_{i=1}^{n} (\sum_{k=0}^{\infty} \frac{t^k x_i^k}{k!}) p(x_i) \tag{1}$$

   $$M_Y(t) = \sum_{i=1}^{n} (\sum_{k=0}^{\infty} \frac{t^k y_i^k}{k!}) p(y_i) \tag{2}$$

   - As we already assumed $M_X(t) = M_Y(t)$, we get

   $$\sum_{i=1}^{n} (\sum_{k=0}^{\infty} \frac{t^k x_i^k}{k!}) p(x_i) = \sum_{i=1}^{n} (\sum_{k=0}^{\infty} \frac{t^k y_i^k}{k!}) p(y_i) \tag{3}$$

   - By comparing t^k coefficients, we get

   $$\sum_{i=1}^{n} \frac{x_i^k}{k!} p(x_i) = \sum_{i=1}^{n} \frac{y_i^k}{k!} p(y_i), \forall k \in W \tag{4}$$

   - As the given equation satisfy for all k, from the properties of linear equations with p(x_i) and p(y_i) as unknowns, equation has unique solutions.

- Hence, we can say that there is unique PDF for a given MGF.

From 1 and 2, we can conclude that for a finite discrete random variable X,its MGF and PDF uniquely determine each other.
Conclusion for X and Z:

- Z is a simple Gaussian because its MGF corresponds to that of a normal distribution,confirming it has a unique Gaussian PDF.
- X is not Gaussian even though its MGF uniquely determines its distribution.X's distribution is a more complex mixture of Gaussians which depends on the probability selection of different Gaussian components.