

DAI Assignment-2

Aditya Neeraje, Balaji Karedla, Moulik Jindal

September 9, 2024

Contents

1	Mathemagic	1
1.1	Task A	1
1.2	Task B	1
1.3	Task C	1
1.4	Task D	2
1.5	Task E	3
1.6	Task F	3
1.7	Task G	4
2	Normal Sampling	5
2.1	Task A	5
2.2	Task B	5
2.3	Task C	6
2.4	Task D	6
2.5	Task E	8
3	Fitting Data	9
3.1	Task A	9
3.2	Task B	9
3.3	Task C	9
3.4	Task D	11
3.5	Task E	12
3.6	Task F	12
4	Quality in Inequalities	13
4.1	Task A	13
4.2	Task B	14
4.3	Task C	14
4.4	Task D	15
4.5	Task E	17

5	A Pretty “Normal” Mixture	19
5.1	Task A	19
5.2	Task B	19
5.2.1	Part 1	19
5.2.2	Part 2	19
5.2.3	Part 3	20
5.3	Task C	20
5.3.1	Part 1	20
5.3.2	Part 2	20
5.3.3	Part 3	21
5.3.4	Part 4	21
5.3.5	Part 5	21
5.3.6	Part 6	21
5.4	Task D	21

1 Mathemagic

1.1 Task A

For a Bernoulli random variable $X \sim \text{Ber}(p)$, $P[X = 0] = 1 - p$, $P[X = 1] = p$ and $P[X = n] = 0 \forall n \geq 2$.

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

1.2 Task B

When $X \sim \text{Bin}(n, p)$, $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \forall k \in \mathbb{Z}, 0 \leq k \leq n$. $P[X = k] = 0 \forall k > n$.

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

1.3 Task C

Let X_1 and X_2 be two random variables which take up non-negative integers and let $X = X_1 + X_2$.

$$\begin{aligned}
G_X(z) &= \sum_{n=0}^{\infty} P[X = n]z^n \\
&= \sum_{n=0}^{\infty} \sum_{i=0}^n (P[X_1 = i]P[X_2 = n - i])z^n \\
&= \sum_{i=0}^{\infty} \sum_{n=i}^{\infty} (P[X_1 = i]P[X_2 = n - i])z^n \\
&= \sum_{i=0}^{\infty} \sum_{n=i}^{\infty} (P[X_1 = i]z^i)(P[X_2 = n - i]z^{n-i}) \\
&= \sum_{i=0}^{\infty} P[X_1 = i]z^i \sum_{n=i}^{\infty} P[X_2 = n - i]z^{n-i} \\
&= \sum_{i=0}^{\infty} P[X_1 = i]z^i \sum_{n=0}^{\infty} P[X_2 = n]z^n \\
G_X(z) &= G_{X_1}(z)G_{X_2}(z)
\end{aligned}$$

For $k = 1$, $G_{\Sigma}(z) = G(z)^k$.

If for $k = n - 1$, $G_{\Sigma}(z) = G(z)^k$, then for $k = n$,

$$\begin{aligned}
G_{\Sigma}(z) &= G_{X_1+X_2+\dots+X_{n-1}}(z)G_{X_n}(z) \\
&= \prod_{i=1}^n G_{X_i}(z) \\
G_{\Sigma}(z) &= G(z)^n
\end{aligned}$$

From the *Principle of Mathematical Induction*, for all $k \in \mathbb{N}$, $G_{\Sigma}(z) = G(z)^k$

1.4 Task D

When $X \sim \text{Geo}(p)$, $P[X = n] = (1 - p)^{n-1}p$ and $P[X = 0] = 0$.

$$\begin{aligned}
G_{\text{Geo}}(z) &= \sum_{n=0}^{\infty} P[X = n]z^n \\
&= \sum_{n=1}^{\infty} (1 - p)^{n-1}pz^n \\
&= pz \sum_{n=1}^{\infty} ((1 - p)z)^{n-1} \\
&= \frac{pz}{1 - (1 - p)z}
\end{aligned}$$

1.5 Task E

$X \sim \text{Bin}(n, p)$ and $Y \sim \text{NegBin}(n, p)$. $Y = \sum_{k=1}^n Y_k$ where $Y_k \sim \text{Geo}(p)$.

$$\begin{aligned}
 G(Y) &= G(Y_1 + Y_2 + \cdots + Y_n) \\
 &= G(Y_1)G(Y_2) \cdots G(Y_n) \\
 &= \left(\frac{pz}{1 - (1-p)z} \right)^n \\
 G_Y^{(n,p)}(z) &= \left(\frac{pz}{1 - (1-p)z} \right)^n \\
 &= \left(\frac{1}{\frac{1}{pz} + (1 - \frac{1}{p})} \right)^n \\
 G_X^{(n,p^{-1})}(z^{-1}) &= \left(\frac{1}{pz} + (1 - \frac{1}{p}) \right)^n \\
 \implies G_Y^{(n,p)}(z) &= \left(G_X^{(n,p^{-1})}(z^{-1}) \right)^{-1}
 \end{aligned}$$

1.6 Task F

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

$$\begin{aligned}
 G_Y^{(n,p)}(z) &= \sum_{k=0}^{\infty} P[Y = k] z^k \\
 &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} z^k \\
 \left(\frac{1}{pz} + \left(1 - \frac{1}{p} \right) \right)^{-n} &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} z^k
 \end{aligned}$$

Substitute $x = z, p = 2$

$$\begin{aligned}
\left(\frac{1}{2x} + \frac{1}{2}\right)^{-n} &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} 2^n (-1)^{k-n} x^k \\
x^n (1+x)^{-n} &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} (-1)^{k-n} x^k \\
(1+x)^{-n} &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} (-1)^{k-n} x^{k-n} \\
(1+x)^{-n} &= \sum_{r=0}^{\infty} \binom{n+r-1}{n-1} (-1)^r x^r \\
(1+x)^{-n} &= \sum_{r=0}^{\infty} \binom{n+r-1}{r} (-1)^r x^r
\end{aligned}$$

1.7 Task G

$$\begin{aligned}
G(z) &= \sum_{n=0}^{\infty} P[X = n] z^n \\
G'(z) &= \sum_{n=0}^{\infty} n P[X = n] z^{n-1} \\
G'(1) &= \sum_{n=0}^{\infty} n P[X = n] \\
G'(1) &= E[X]
\end{aligned}$$

For Bernoulli random variable $X \sim \text{Ber}(p)$,

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

For Binomial random variable $X \sim \text{Bin}(n, p)$,

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

For Geometric random variable $X \sim Geo(p)$,

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

For Negative Binomial random variable $X \sim NegBin(n, p)$,

$$\begin{aligned} G_{NegBin}(z) &= \left(\frac{pz}{1 - (1-p)z} \right)^n \\ G'_{NegBin}(z) &= n \left(\frac{pz}{1 - (1-p)z} \right)^{n-1} \frac{p}{(1 - (1-p)z)^2} \\ G'_{NegBin}(1) &= n \left(\frac{p}{p} \right)^{n-1} \frac{p}{p^2} \\ E[X] &= \frac{n}{p} \end{aligned}$$

2 Normal Sampling

2.1 Task A

To prove that Y is uniformly distributed in $[0,1]$, we need to prove that for an arbitrary interval $[a,b]$ in $[0, 1]$, Y lies in $[a, b]$ with probability $b - a$.

The probability that $a \leq Y \leq b$ is the probability that the randomly chosen X is such that $a \leq F_X(x) \leq b$. The cumulative probability of x being in the interval $[a_0, b_0]$ such that $F_X(a_0) = a$ and $F_X(b_0) = b$ is $b - a$, since the cumulative probability that $x \leq b_0$ is b , and the cumulative probability of $x \leq a_0$ is a , hence probability that $a_0 \leq x \leq b_0$ is $b - a$. Our condition that $F_X()$ is invertible guarantees that the \leq can be replaced with $<$, since it cannot be that $\lim_{x \rightarrow a_0^-} \neq \lim_{x \rightarrow a_0^+}$, otherwise between a_0^- and a_0^+ , all elements would be mapped to by a_0 .

2.2 Task B

Given a sample y , let us return $F_X^{-1}(y)$. To prove that this has the same CDF as X , we need to prove that if $F_X(a_0) = a$ and $F_X(b_0) = b$, the probability of getting $a_0 \leq F_X^{-1}(y) \leq b_0$ is $b - a$.

But this is fairly straightforward, since $a_0 \leq F_X^{-1}(y) \leq b_0 \iff a \leq y \leq b$, and the probability of getting $a \leq y \leq b$ is $b - a$.

Thus, this is the required function.

2.3 Task C

Code provided in 2c.ipynb.

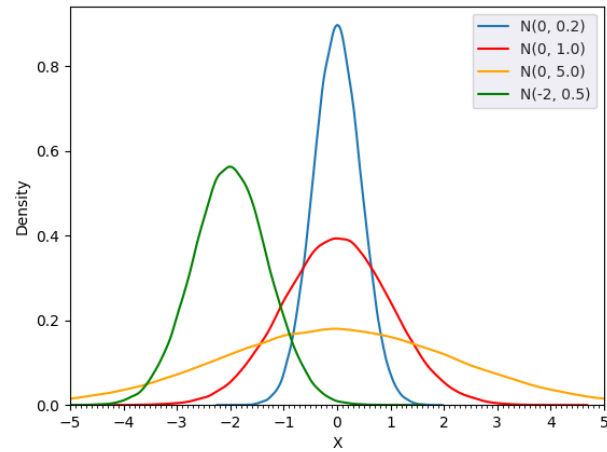


Figure 1: Histogram of the samples generated

2.4 Task D

Code provided in 2d.ipynb.

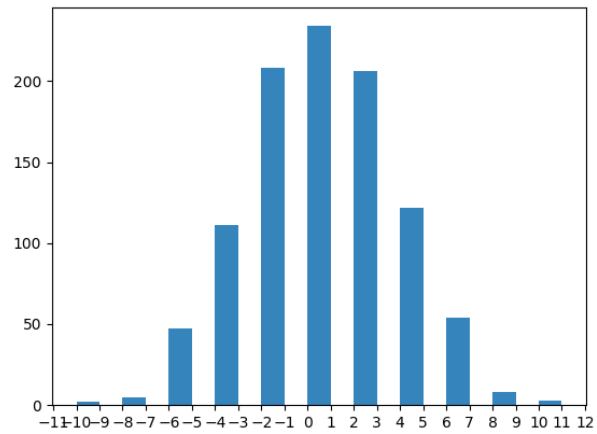
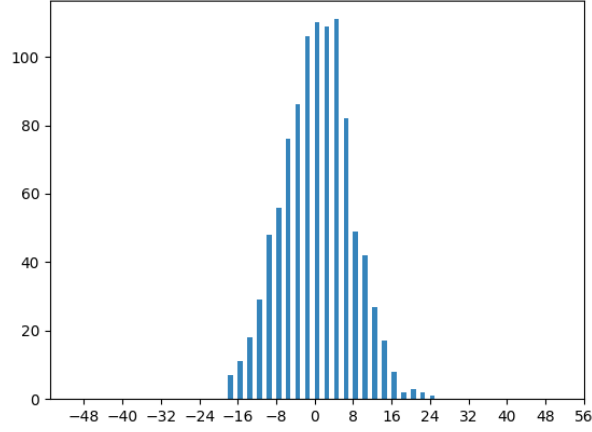
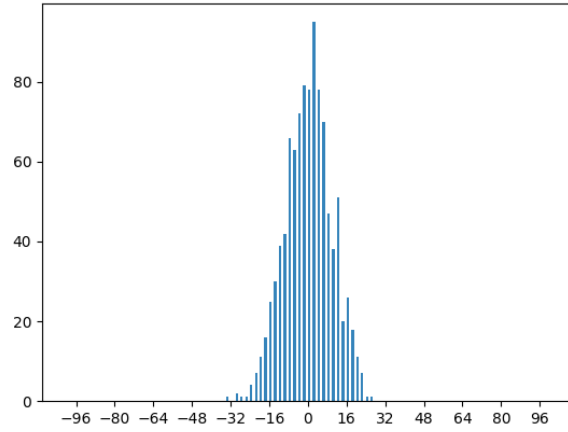


Figure 2: $h=10$ Figure 3: $h=50$ Figure 4: $h=100$

For $h=100$ alone, for better visibility, I have also written code with bins set to increment every 2 pockets. To observe the property of only even indexed bins being filled, simply replace `range(-100, 102, 2)` with `range(-100, 102, 1)`.

2.5 Task E

$$P_h[X = 2i] = \frac{\binom{2k}{i+k}}{2^{2k}}$$

Since the likelihood of a ball arriving at pocket i is equal to the likelihood of having $i+k$ collisions which move it to the right within the $2k$ possible collisions overall.

Applying Stirling's approximations,

$$\begin{aligned} P_h[X = 2j] &\approx \frac{k^{2k} \sqrt{4\pi k}}{(k+j)^{k+j} \sqrt{2\pi(k+j)} (k-j)^{k-j} \sqrt{2\pi(k-j)}} \\ &\approx \frac{k^{2k} \sqrt{2k} (k-j)^j}{(k^2 - j^2)^k \sqrt{2\pi(k^2 - j^2)} (k+j)^j} \\ &\approx \frac{1}{\sqrt{\pi k}} \frac{k^{2k} (k-j)^j}{(k^2)^k (k+j)^j} && \text{(By approximating } k^2 - j^2 \approx j^2) \\ &= \frac{1}{\sqrt{\pi k}} \left(1 - \frac{2j}{k+j}\right)^j \end{aligned}$$

Now, we independently approximate $\left(1 - \frac{2j}{k+j}\right)^j$ by taking its logarithm and using the property that for small x , $\ln(1-x) \approx -x$.

$$\begin{aligned} \ln \left(\left(1 - \frac{2j}{k+j}\right)^j \right) &= j \ln \left(1 - \frac{2j}{k+j}\right) \\ &\approx -\frac{2j^2}{k+j} \\ \left(1 - \frac{2j}{k+j}\right)^j &\approx e^{-\frac{2j^2}{k}} \end{aligned}$$

Setting $i = 2j$ as required, we get

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

As required.

3 Fitting Data

3.1 Task A

$$\mu_1 = \frac{\sum x_i}{N} = 6.50$$
$$\mu_2 = \frac{\sum x_i^2}{N} = 46.55$$

3.2 Task B

The histogram of the data is shown in Figure 5.

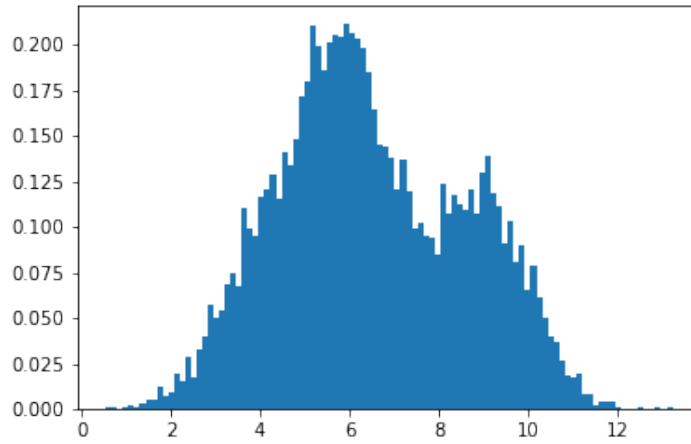


Figure 5: Histogram of the data

3.3 Task C

The binomial distribution is given by

$$\begin{aligned}
\mu_1^{Bin} &= \sum_{r=0}^n \binom{n}{r} p^r (1-p)^{n-r} r \\
&= np \\
\mu_2^{Bin} &= \sum_{r=0}^n \binom{n}{r} p^r (1-p)^{n-r} r^2 \\
&= np \sum_{r=1}^n \binom{n-1}{r-1} p^{r-1} (1-p)^{n-r} r \\
&= np \sum_{r=0}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r-1} (r+1) \\
&= n(n-1)p^2 + np \sum_{r=0}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r-1} \\
&= n(n-1)p^2 + np
\end{aligned}$$

The values of μ_1^{Bin} and μ_2^{Bin} are 6.50 and 46.55 respectively. These values are the same as the values of μ_1 and μ_2 calculated in Task A.

When the values of n and p are fit the the data and the integers next to n are taken, the values of n and p are 20 and 0.33 respectively.

The histogram of the data and the binomial distribution with $n = 20$ and $p = 0.33$ is shown in Figure 6.

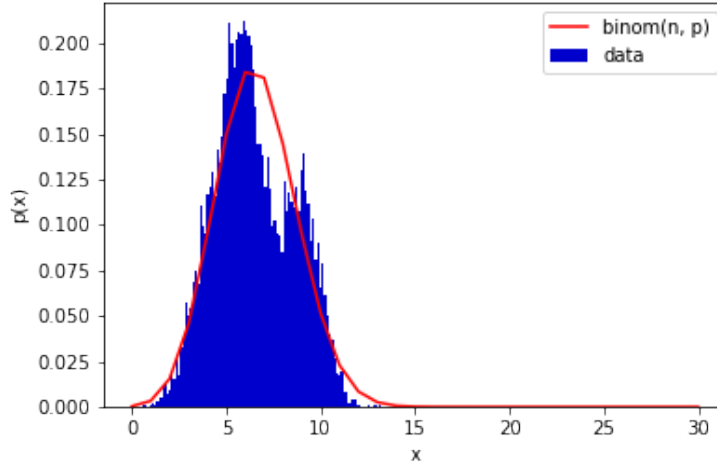


Figure 6: Histogram of the data and the binomial distribution with $n = 20$ and $p = 0.33$

3.4 Task D

The gamma distribution is given by

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

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

The values of μ_1^{Gamma} and μ_2^{Gamma} are 6.50 and 46.55 respectively. These values are the same as the values of μ_1 and μ_2 calculated in Task A.

When the values of k and θ are fit the the data, the values of k and θ are 9.70 and 0.67 respectively.

The histogram of the data and the gamma distribution with $k = 9.70$ and $\theta = 0.67$ is shown in Figure 7.

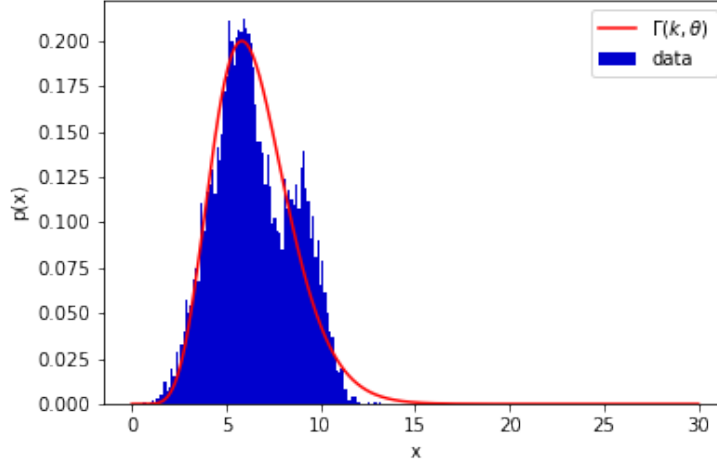


Figure 7: Histogram of the data and the gamma distribution with $k = 9.70$ and $\theta = 0.67$

3.5 Task E

Definition: *Likelihood*

Given a dataset S and a choice of parameter $\lambda = \lambda_0$ for a family of distributions $P[\lambda]$ parameterized by λ , define the likelihood of λ_0 by

$$\mathcal{L}(\lambda_0 | S) = P_{\lambda_0}[S] = \prod_{i=1}^N P_{\lambda_0}[x_i]$$

For large datasets, since each probability is a small number, the likelihood can underflow. Thus, the average log-likelihood is typically calculated:

$$\ell(\theta | S) = \frac{\log \mathcal{L}(\theta | S)}{N}$$

The average log-likelihood of the data under the binomial and gamma distributions are -2.157 and -2.161 respectively. Binomial distribution has a higher average log-likelihood than the gamma distribution.

3.6 Task F

$$\begin{aligned} \mu_1^{gmm} &= p_1\mu_1 + p_2\mu_2 \\ \mu_2^{gmm} &= p_1(\sigma_1^2 + \mu_1^2) + p_2(\sigma_2^2 + \mu_2^2) \\ \mu_3^{gmm} &= p_1(\mu_1^3 + 3\mu_1\sigma_1^2) + p_2(\mu_2^3 + 3\mu_2\sigma_2^2) \\ \mu_4^{gmm} &= p_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + p_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) \end{aligned}$$

When the values of μ_1 , p_1 , μ_2 , and p_2 are fit to the data, with $\sigma_1 = \sigma_2 = 1$, the values of are 5.13, 0.61, 8.77, and 0.38 respectively.

The histogram of the data and the GMM with $\mu_1 = 5.13$, $p_1 = 0.61$, $\mu_2 = 8.77$, and $p_2 = 0.38$ is shown in Figure 8.

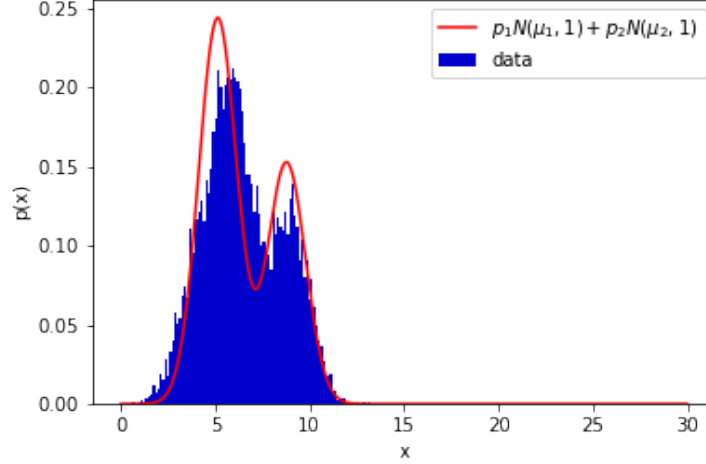


Figure 8: Histogram of the data and the GMM with $\mu_1 = 5.13$, $p_1 = 0.61$, $\mu_2 = 8.77$, and $p_2 = 0.38$

4 Quality in Inequalities

4.1 Task A

$$\begin{aligned}
 \mathbb{E}[X] &= \int_0^{\infty} X P(X) dX \\
 &= \int_0^a X P(X) dX + \int_a^{\infty} X P(X) dX \\
 &\geq \int_0^a X P(X) dX + \int_a^{\infty} a P(X) dX \\
 &\geq \int_0^a X P(X) dX + a P[X \geq a] \\
 &\geq a P[X \geq a]
 \end{aligned}$$

Here, we have used the fact that $\int_0^a X P(X) dX$ is ≥ 0 since $X \geq 0$ and $P(X) \geq 0$ everywhere in that interval.

From the above result, we get

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

Intuitively, this corresponds to saying that if there is a high probability that $X \geq a$, then we can also expect that a random sample will give us a value greater than a .

4.2 Task B

First, we will prove a result which is useful later

$$\mathbb{E}\left[\left(X - \mathbb{E}[X] + \frac{\sigma^2}{\tau}\right)^2\right] = \text{Var}(X - \mathbb{E}[X] + \frac{\sigma^2}{\tau}) + \left(\mathbb{E}[X - \mathbb{E}[X] + \frac{\sigma^2}{\tau}]\right)^2$$

Now, since the variance of a distribution does not change on adding a constant to the distribution, and since the expectation value of the sum of terms is the sum of expectations of the terms, we get that this is equal to

$$\mathbb{E}\left[\left(X - \mathbb{E}[X] + \frac{\sigma^2}{\tau}\right)^2\right] = \sigma^2 + \left(\frac{\sigma^2}{\tau}\right)^2$$

Now, we notice that whenever $X - \mu \geq \tau$, it also holds that $\left(X - \mu + \frac{\sigma^2}{\tau}\right)^2 \geq \left(\tau + \frac{\sigma^2}{\tau}\right)^2$. Since we are assuming that $x - \mu + \frac{\sigma^2}{\tau} \geq \tau + \frac{\sigma^2}{\tau} \geq 0$, this result is proven by the fact that x^2 is an increasing function over the non-negative reals.

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

Using Markov's inequality on the above result, we get

$$\begin{aligned} P[X - \mu \geq \tau] &\leq P\left[\left(X - \mu + \frac{\sigma^2}{\tau}\right)^2 \geq \left(\tau + \frac{\sigma^2}{\tau}\right)^2\right] \leq \frac{\sigma^2 + \left(\frac{\sigma^2}{\tau}\right)^2}{\left(\tau + \frac{\sigma^2}{\tau}\right)^2} \\ &= \frac{\sigma^2}{\sigma^2 + \tau^2} \end{aligned}$$

Thus proving our desired result.

4.3 Task C

$$P[X \geq x] = \int_x^\infty P(X) dX$$

$$\begin{aligned} e^{-tx} E[e^{tx}] &= e^{-tx} \int_{-\infty}^\infty e^{tX} P(X) dX \\ &= e^{-tx} \int_{-\infty}^x e^{tX} P(X) dX + e^{-tx} \int_x^\infty e^{tX} P(X) dX \\ &\geq e^{-tx} \int_x^\infty e^{tX} P(X) dX && \text{(since both } P(X) \text{ and } e^{tX} \text{ are positive for all } X) \\ &\geq e^{-tx} e^{tx} \int_x^\infty P(X) dX && \text{(since } e^{tX} \text{ is increasing if } t \geq 0) \\ &= e^{-tx} e^{tx} P[X \geq x] \\ &= P[X \geq x] \end{aligned}$$

This proves that

$$P[X \geq x] \leq e^{-tx} E[e^{tx}] \quad (1)$$

Similarly,

$$P[X \leq x] = \int_{-\infty}^x P(X) dX$$

$$\begin{aligned} e^{-tx} E[e^{tx}] &= e^{-tx} \int_{-\infty}^{\infty} e^{tX} P(X) dX \\ &= e^{-tx} \int_{-\infty}^x e^{tX} P(X) dX + e^{-tx} \int_x^{\infty} e^{tX} P(X) dX \\ &\geq e^{-tx} \int_{-\infty}^x e^{tX} P(X) dX && \text{(since both } P(X) \text{ and } e^{tX} \text{ are positive for all } X) \\ &\geq e^{-tx} \cdot e^{tx} \cdot \int_{-\infty}^x P(X) dX && \text{(since } e^{tx} \leq e^{tx'} \text{ where } x \geq x' \text{ when } t \text{ is negative)} \\ &= \int_{-\infty}^x P(X) dX \\ &= P[X \leq x] \end{aligned}$$

as required.

4.4 Task D

1. We know, by the linearity of expectation, that

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

2. For this question, we make use of the first result proved in Task C above.

$$\begin{aligned} P[Y \geq (1 + \delta)\mu] &\leq \frac{\mathbb{E}[e^{tY}]}{e^{(1+\delta)\mu t}} \\ &= \frac{\mathbb{E}[e^{(t \sum_{i=1}^n X_i)}]}{e^{(1+\delta)\mu t}} \end{aligned}$$

options We will now prove that for independent random variables, $\mathbb{E}[e^{(t \sum_{i=1}^n X_i)}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}]$.

To prove that, we prove the simpler result that for independent variables, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy) p_{X,Y}(x, y) dy dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy) p_X(x) p_Y(y) dy dx \quad (\text{For independent RVs, } p(xy)=p(x)p(y)) \\
&= \int_{-\infty}^{\infty} x p_X(x) \int_{-\infty}^{\infty} y p_Y(y) dy dx \\
&= \int_{-\infty}^{\infty} x p_X(x) dx \int_{-\infty}^{\infty} y p_Y(y) dy \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Since above, all X_i 's are independent, so are e^{tX_i} for all i . Thus,

$$\begin{aligned}
E[e^{(t \sum_{i=1}^n X_i)}] &= \mathbb{E}[e^{tX_1}] E[e^{(t \sum_{i=2}^n X_i)}] \\
&= \mathbb{E}[e^{tX_1}] \mathbb{E}[e^{tX_2}] \mathbb{E}[e^{(t \sum_{i=3}^n X_i)}] \\
&\dots \\
&= \prod_{i=1}^n \mathbb{E}[e^{tX_i}]
\end{aligned}$$

For each Bernoulli random variable X_i , it is easily observed that $E[e^{tX_i}] = 1 + (e^t - 1)p_i$. Using this result,

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

We will now prove $\prod_{i=1}^n (1 + x_i) \leq e^{\sum_{i=1}^n x_i}$ for positive x_i 's.

$$\begin{aligned}
\ln \left(\prod_{i=1}^n (1 + x_i) \right) &= \sum_{i=1}^n \ln(1 + x_i) \\
&\leq \sum_{i=1}^n x_i \quad (\text{Well known result that } \ln(1 + x) \leq x \text{ in these conditions}) \\
&= \ln \left(e^{\sum_{i=1}^n x_i} \right)
\end{aligned}$$

Since our initial assumption in using 1 required $t > 0$ and hence $e^t > 1$, all $(e^t - 1)p_i$ values are positive and we can use the above result. Hence, $\mathbb{E}[e^{tY}] \leq e^{\mu(e^t - 1)}$. Plugging this into 1, we get

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

as required.

3. Since this holds for arbitrary positive t , we can find the t for which the value of $e^{\mu(e^t-1-(1+\delta)t)}$ is minimum. By differentiation, we observe the minima of the function is when $t = \ln(1 + \delta)$. At the minima, we have

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

Using the expansion $\ln(1 + \delta) = \sum_{i=1}^{\infty} \frac{\delta^i (-1)^{i-1}}{i}$,

$$\begin{aligned} \delta - (1 + \delta) \ln(1 + \delta) &= \delta + (1 + \delta) \left(\sum_{i=1}^{\infty} \frac{\delta^i (-1)^i}{i} \right) \\ &= \delta + \sum_{i=1}^{\infty} \frac{\delta^i (-1)^i}{i} + \sum_{i=2}^{\infty} \frac{\delta^i (-1)^{i-1}}{i-1} \\ &= \sum_{i=2}^{\infty} \frac{\delta^i (-1)^{i-1}}{i(i-1)} \end{aligned}$$

This is approximately $-\frac{\delta^2}{2} + \frac{\delta^3}{6}$ for small delta. Thus we get a good upper bound of $e^{\frac{-\mu\delta^2(3-\delta)}{6}}$ for $P[Y \geq (1 + \delta)\mu]$.

If $\delta > 0$, this can be simplified further, in a way that is more useful for us in question 5. Observe that for $x > 0$

$$\ln(1 + x) \geq \frac{2x}{2 + x}$$

This can be seen by observing that at $x=0$, $\ln(1+x) - \frac{2x}{2+x}$ evaluates to 0 and that its derivative, $\frac{4}{(1+x)(2+x)^2}$ is always non-negative. Thus,

$$\begin{aligned} \mu(\delta - (1 + \delta) \ln(1 + \delta)) &\leq \mu\left(\delta - \frac{2\delta(1 + \delta)}{2 + \delta}\right) \\ &= \frac{-\mu\delta^2}{2 + \delta} \end{aligned}$$

Thus, we get,

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

For the last few steps of simplification, <https://math.mit.edu/~goemans/18310S15/chernoff-notes.pdf> was used as a guide.

4.5 Task E

First, let us consider the corner cases where $\mu = 0$ or $\mu = 1$. In these cases, $A_n = 0$ or $A_n = 1$ always, and the result is trivial.

Now we can consider $0 < \mu < 1$.

Consider the random variable $B_n = nA_n$.

We need to prove $\lim_{n \rightarrow \infty} P[|B_n - n\mu| > n\epsilon] = 0$.

Note that we are done if we prove $\lim_{n \rightarrow \infty} P[B_n - n\mu \geq n\epsilon] = 0$ and $\lim_{n \rightarrow \infty} P[n\mu - B_n \geq n\epsilon] = 0$.

We will prove the first part first, assuming μ is not 0.

$$P[B_n \geq n\epsilon + n\mu] \leq e^{-\frac{n\epsilon^2}{2\mu + \epsilon}}$$

For a positive μ (since the distribution is the sum of Bernoulli RVs) and ϵ (given), $\frac{\epsilon^2}{2\mu + \epsilon}$ is a positive coefficient. e^{-n} tends to 0 as n tends to ∞ , which we will use henceforth.

Note that the result that $P[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu\delta^2}{2 + \delta}}$ was not proven without referring to the internet, so I will prove the claim above using an alternative method which assumes only what has been proven by me.

$$P[X \geq (1 + \delta)\mu] \leq e^{\mu(\delta - (1 + \delta)\ln(1 + \delta))}$$

We want to prove that $f(\delta) = \delta - (1 + \delta)\ln(1 + \delta)$ is strictly negative if δ is positive. For that, notice that $f(0) = 0$ and $f'(\delta) = -\ln(1 + \delta)$, which is non-positive for all non-negative δ . Thus, $f(\delta)$ is strictly negative for all positive δ .

This again means that $P[B_n \geq n\epsilon + n\mu] \leq e^{-nk}$ for some positive coefficient k . Thus, $P[B_n \geq n\epsilon + n\mu]$ tends to 0 as n tends to ∞ .

Similarly, $P[B_n \leq n\mu - n\epsilon]$ also tends to 0 as n tends to ∞ .

We can prove this by considering the RV $1 - A_n$ instead of A_n and using the same logic as above.

$P[n - B_n - n - n\mu \geq n\delta] = P[n\mu - nB_n \geq n\delta] \leq e^{-nk}$ for some positive coefficient k .

Now, combining the two results,

$$\begin{aligned} P[|B_n - n\mu| > n\epsilon] &\leq P[B_n - n\mu \geq n\epsilon] + P[n\mu - B_n \geq n\epsilon] \\ &\leq e^{-nk} + e^{-nk'} \quad (\text{for some positive coefficients } k \text{ and } k') \end{aligned}$$

If we get any positive value σ for the value of $P[|B_n - n\mu| > n\epsilon]$, simply set $n > \max(\frac{\ln(\frac{2}{\sigma})}{k}, \frac{\ln(\frac{2}{\sigma})}{k'})$ and we have

$$\begin{aligned} P[|B_n - n\mu| > n\epsilon] &\leq e^{-nk} + e^{-nk'} \\ &< \frac{\sigma}{2} + \frac{\sigma}{2} \\ &= \sigma \end{aligned}$$

which is a contradiction.

But, for all finite n , it holds that whenever $|B_n - n\mu| > n\delta$, we also know by the definition of B_n that $|A_n - \mu| > \delta$. Thus, $P[|B_n - n\mu| > n\delta] \geq P[|A_n - \mu| > \delta] \geq 0$. As this is a sandwiched limit, by the Sandwich theorem for limits, we have proven that $\lim_{n \rightarrow \infty} P[|B_n - n\mu| > n\epsilon] = \lim_{n \rightarrow \infty} P[|A_n - \mu| > \epsilon] = 0$.

5 A Pretty “Normal” Mixture

5.1 Task A

Let A_i be the event in which you choose i in the first step of the algorithm. Then the set of all the A_i s is a valid partition, since they are completely disjoint and their union represents all the possibilities.

Hence we can use the *law of Total Probability* to get,

$$\begin{aligned}
 P[\mathcal{A} = x] &= \sum_{i=1}^K P[A_i] \cdot P[X_i = x] \\
 &= \sum_{i=1}^K p_i \cdot P[X_i = x] \\
 &= P[X = x] \\
 \implies f_{\mathcal{A}}(x) &= f_X(x)
 \end{aligned}$$

5.2 Task B

5.2.1 Part 1

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{x \in \mathbb{Z}} \left(x \cdot \sum_{i=1}^K p_i \cdot P[X_i = x] \right) \\
 &= \sum_{i=1}^K \left(p_i \cdot \sum_{x \in \mathbb{Z}} x \cdot P[X_i = x] \right) \\
 &= \sum_{i=1}^K (p_i \cdot \mathbb{E}[X_i]) = \sum_{i=1}^K (p_i \cdot \mu_i)
 \end{aligned}$$

5.2.2 Part 2

- Calculating $\mathbb{E}[X^2]$

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{x \in \mathbb{Z}} \left(x^2 \cdot \sum_{i=1}^K p_i \cdot P[X_i = x] \right) \\
 &= \sum_{i=1}^K \left(p_i \cdot \sum_{x \in \mathbb{Z}} x^2 \cdot P[X_i = x] \right) \\
 &= \sum_{i=1}^K (p_i \cdot \mathbb{E}[X_i^2]) = \sum_{i=1}^K \left(p_i \cdot \left(\text{Var}(X_i) + (\mathbb{E}[X_i])^2 \right) \right) \\
 &= \sum_{i=1}^K (p_i \cdot (\sigma_i^2 + \mu_i^2))
 \end{aligned}$$

- Calculating $(\mathbb{E}[X])^2$

$$\begin{aligned}
 (\mathbb{E}[X])^2 &= \left(\sum_{i=1}^K p_i \cdot \mathbb{E}[X_i] \right)^2 \\
 &= \sum_{i=1}^K (p_i \cdot \mathbb{E}[X_i])^2 + 2 \sum_{1 \leq i < j \leq K} (p_i \cdot \mathbb{E}[X_i] \cdot p_j \cdot \mathbb{E}[X_j]) \\
 &= \sum_{i=1}^K (p_i \mu_i)^2 + 2 \sum_{1 \leq i < j \leq K} (p_i \mu_i p_j \mu_j)
 \end{aligned}$$

- Calculating $\text{Var}[X]$

$$\begin{aligned}
 \text{Var}[X] &= (\mathbb{E}[X])^2 - \mathbb{E}[X^2] \\
 &= \left(\sum_{i=1}^K (p_i \mu_i)^2 + 2 \sum_{1 \leq i < j \leq K} (p_i \mu_i p_j \mu_j) \right) - \sum_{i=1}^K (p_i \cdot (\sigma_i^2 + \mu_i^2))
 \end{aligned}$$

5.2.3 Part 3

$$\begin{aligned}
 M_X(t) &= G(e^t) = \int_{-\infty}^{\infty} e^{xt} P[X = x] dx = \int_{-\infty}^{\infty} e^{xt} \sum_{i=1}^K p_i P[X_i = x] dx \\
 &= \int_{-\infty}^{\infty} \sum_{i=1}^K e^{xt} p_i P[X_i = x] dx = \sum_{i=1}^K p_i \int_{-\infty}^{\infty} e^{xt} P[X_i = x] dx \\
 &= \sum_{i=1}^K p_i M_{X_i}(t) = \sum_{i=1}^K p_i e^{\mu_i t + \frac{1}{2} \sigma_i^2 t^2}
 \end{aligned}$$

5.3 Task C

5.3.1 Part 1

$$\mathbb{E}[Z] = \mathbb{E} \left[\sum_{i=1}^K p_i X_i \right] = \sum_{i=1}^K p_i \cdot \mathbb{E}[X_i] = \sum_{i=1}^K (p_i \mu_i)$$

5.3.2 Part 2

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

5.3.3 Part 3

Since we already calculated μ_Z and $\sigma_Z = \sqrt{\text{Var}(Z)}$ We can just use the expression for Gaussian Random Variable (See Part 6 for the proof that Z is a Gaussian random variable)

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

5.3.4 Part 4

Using the expression for Moment generating function of a Gaussian Random Variable (See Part 6 for the proof that Z is a gaussian random variable)

$$M_Z(t) = e^{\mu_Z t + \frac{1}{2} \sigma_Z^2 t^2}$$

5.3.5 Part 5

No, It can be seen that even though the expected value is the same in both the case, The values of both variance and MGF are different for both X and Z hence these are different random variables

5.3.6 Part 6

We can show that linear combination of independent random variables is also a random variable.

For any random Gaussian variable X is given by $M_X(t) = \mathbb{E}[e^{tX}] = e^{\mu t + \frac{1}{2} \sigma^2 t^2}$

Let $Y = \sum_{i=1}^K a_i X_i$ where each X_i is an independent Gaussian random variable. Then

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] = \mathbb{E}\left[e^{t \sum_{i=1}^K a_i X_i}\right] = \mathbb{E}\left[\prod_{i=1}^K e^{ta_i X_i}\right] \\ &= \prod_{i=1}^K \mathbb{E}[e^{ta_i X_i}] \quad , \text{Since all the } X_i \text{ are mutually independent} \\ &= \prod_{i=1}^K M_{a_i X_i}(t) = \prod_{i=1}^K \left(e^{a_i \mu_i t + \frac{1}{2} (a_i \sigma_i)^2 t^2}\right) = \exp\left(\sum_{i=1}^K (a_i \mu_i) t + \frac{1}{2} \sum_{i=1}^K ((a_i \sigma_i)^2) t^2\right) \end{aligned}$$

If we define $\mu_y = \sum_{i=1}^K (a_i \mu_i)$ and $\sigma_y^2 = \sum_{i=1}^K ((a_i \sigma_i)^2)$ Then we get

$$M_y(t) = e^{\mu_y t + \frac{1}{2} \sigma_y^2 t^2} \quad (3)$$

This is just the Moment generating function of a Gaussian random variable, Hence Y is a Gaussian random variable. Hence Z is also a Gaussian random variable.

5.4 Task D

To prove that a PDF uniquely determines the MGF, we simply use the construction of the MGF. The MGF is defined as

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} p_X(x) dx = \sum_{i=1}^n e^{tx_i} p(x_i)$$

where the support of the PDF is $\{x_1, \dots, x_n\}$.

To prove the other direction, let us assume that we can find two distinct solutions $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ from an arbitrary MGF which we know is of a finite discrete PDF. Then, by the above result, we know that

$$\sum_{i=1}^n e^{tx_i} p(x_i) = \sum_{j=1}^m e^{ty_j} p(y_j)$$

If any of the x_i 's are not equal to any of the y_j 's, and given that all $p(x_i)$'s and $p(y_j)$'s are non-zero, then we can rearrange the terms to get that for some a_i 's

$$e^{a_0 t} b_0 = \sum_{i=1}^k e^{a_i t} b_i$$

for some non-zero coefficients b_0 and for some k , and all a_i 's being distinct. Now, note that the LHS has an annihilating polynomial $D - a_0$, where D is the derivative operator w.r.t t . The RHS has an annihilating polynomial $(D - a_1) \dots (D - a_k)$. Since the LHS equals the RHS, if an operator makes the LHS 0, it also makes the RHS zero. Thus, the LHS $(e^{a_0 t} b_0)$ is also annihilated by the annihilator of the RHS.

$$\begin{aligned} & (D - a_1) \dots (D - a_k) (e^{a_0 t} b_0) = 0 \\ \implies & (D - a_1) \dots (D - a_{k-1}) ((a_0 - a_k) e^{a_0 t} b_0) = 0 \\ \implies & (D - a_1) \dots (D - a_{k-2}) (a_0 - a_{k-1}) (a_0 - a_k) e^{a_0 t} b_0 = 0 \\ & \dots \\ \implies & (a_0 - a_1) \dots (a_0 - a_k) e^{a_0 t} b_0 = 0 \end{aligned}$$

Since $e^{a_0 t}$ is never 0, and we know $b_0 \neq 0$, $(a_0 - a_1) \dots (a_0 - a_k) = 0$. But this is only possible if $a_0 = a_i$ for some $i \in \{1, \dots, k\}$. But that contradicts our assumption that the a_i 's are distinct.

Thus, it is not possible for two PDFs of finite discrete RVs to correspond to the same MGF.