Statistical Analysis

V.S.S. Vardhan : 23b0909
Shasank Reddy P : 23b1015
Sree Vamshi Madhav Nenavath: 23b1039

September 29, 2024

Contents

Chernoff bound

Lemma 1. *Let X be any non-negative random variable and $\dashv > 0$*

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

Proof. Let's assume the random variable X be a discrete random variable, and extend to continous random variable. The Expected value $E(X)$ can be expressed with $x \in X$ (value in the range) and its probability p_x as

$$E(X) = \sum p_x \cdot x$$

Now split this sum into $x < a$ and $x \geq a$

$$E(X) = \sum_{x<a} p_x \cdot x + \sum_{x \geq a} p_x \cdot x$$

$$E(X) \geq \sum_{x \geq a} p_x \cdot x$$

Now since $x \geq a$, put this inequality in the above equation and we get

$$E(X) \geq \sum_{x \geq a} p_x \cdot a$$

$$E(X) \geq a \cdot \sum p_x$$

Since all the x are discrete, they are disjoint and thus

$$P[X \geq a] = \sum p_x$$

Thus we get

$$E(X) \geq a \cdot P[X \geq a]$$

For the continous case, the expectation can be expressed with the PDF $f(x)$ as

$$\begin{aligned}
E(X) &= \int_{-\infty}^{a} x f(x) dx + \int_{a}^{\infty} x f(x) dx \\
&\geq \int_{a}^{\infty} x f(x) dx \\
&\geq \int_{a}^{\infty} a \cdot f(x) dx \text{ ; since } x \geq a \\
&\geq a \int_{a}^{\infty} f(x) dx \\
&\geq a P[X \geq a]
\end{aligned}$$

Therefore, for discrete or continous random variables,

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

□

Lemma 2. *Now take n independent Bernoulli random variables $X_1, X_2, \cdots X_n$ where $E[X_i] = p_i$ Let us define a new random variable $\mathcal{Y}$ to be the sum of these random variables, that is, $\mathcal{Y} = \sum_{i=1}^{n} X_i$.*

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

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

Proof. The MGF function for a random variable X is given by

$$M_X(t) = E(e^{Xt})$$

Now consider the set $\{X | X \geq x\}$ and the function $f(x) = e^t x$. Let's construct a new random variable $Y = e^{tX}$ and its bound $y = e^{tx}$

$$P[Y \geq y] \leq \frac{E(Y)}{y}$$

$$P[Y \geq y] \leq \frac{E(e^{tX})}{e^{tx}}$$

Here, by susbtituting the definition of the MGF,

$$P[Y \geq y] \leq e^{-tx} \cdot M_X(t)$$

Now consider two cases for the sign of t.

case 1 : $t > 0$

This set is equivalent to $\{X | e^{tX} \geq e^{tx}\}$ since the function $f(x) = e^{tx}$ is a one-one, monotonic **increasing** function.

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

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

case 2 : $t < 0$

This set is equivalent to $\{X | e^{tX} \leq e^{tx}\}$ since the function $f(x) = e^{tx}$ is a one-one, monotonic **decreasing** function.

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

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

Thus, we can conclude that

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

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

□

Let's use the formula derived above

$$P[Y \geq (1+\delta)\mu] \leq \frac{E(e^{tY})}{e^{t((1+\delta)\mu)}} \forall t > 0$$

By expanding e^{tY}, we get

$$\begin{aligned}
E(e^{tY}) &= E\left(e^{t\sum_{i=1}^{n} X_i}\right) \\
&= E\left(\prod_{i=1}^{n} e^{tX_i}\right) \\
&= \prod_{i=1}^{n} E(e^{tX_i})
\end{aligned}$$

For the expected value $E(e^{tX_i})$, we can compute its values and probabilities as

$$\begin{aligned}
E(e^{tX_i}) &= (1-p_i) \cdot e^{0 \cdot t} + p_i \cdot e^{1 \cdot t} \\
E(e^{tX_i}) &= 1 + (e^t - 1)p_i
\end{aligned}$$

Since we know that

$$1 + c \cdot x \leq e^{cx}$$

By the above two inequalities, we can say that

$$\begin{aligned}
E(e^{tY}) = \prod_{i=1}^{n} E(e^{tX_i}) &= \prod_{i=1}^{n} (1 + (e^t - 1)p_i) \\
&\leq \prod_{i=1}^{n} e^{(e^t-1)p_i} \\
&\leq e^{(e^t-1)\sum_{i=1}^{n} p_i}
\end{aligned}$$

But we have proved that $E(Y) = \mu = \sum_{i=1}^{n} p_i$, by all this, we can conclude that

$$\boxed{P[Y \geq (1+\delta)\mu] \leq \frac{e^{\mu(e^t-1)}}{e^{t\mu(1+\delta)}} \forall t > 0}$$

To improve this bound, we have to find a t such that $\frac{e^{\mu(e^t-1)}}{e^{t\mu(1+\delta)}}$ is minimised. we can use the fact that if

$$\frac{\partial}{\partial t}\frac{f(t)}{g(t)} = 0 \text{ then, } f'(t)g(t) = f(t)g'(t)$$

When applied to the above term,

$$\begin{aligned}
\frac{\partial}{\partial t} e^{\mu(e^t-1)} \cdot e^{t\mu(1+\delta)} &= e^{\mu(e^t-1)} \cdot \frac{\partial}{\partial t} e^{t\mu(1+\delta)} \\
e^{\mu(e^t-1)} \mu(e^t) \cdot e^{t\mu(1+\delta)} &= e^{\mu(e^t-1)} \cdot \mu(1+\delta) e^{t\mu(1+\delta)} \\
e^t &= 1 + \delta
\end{aligned}$$

when substituted in $\frac{e^{\mu(e^t-1)}}{e^{t\mu(1+\delta)}}$,

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

$$= (\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^{\mu}$$

Thus the final expression becomes

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

Weak Law of Large Numbers

From the above theorm, lets define for $Y = \sum_{i=1}^{n} X_i$. We have

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

Now, lets define a new random variable $Y' = \frac{Y}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$ with a new mean $\mu' = \mu/n$.

$$P[Y - \mu \geq \delta\mu] = P[Y' - \mu' \geq \delta\mu']$$

now, substitute $\delta \cdot \mu \to \epsilon$,$\mu \to n \cdot \mu'$

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

Apply the limit for $\delta \to 0$ which corresponds to $\epsilon \to 0$

$$\begin{aligned}(\frac{e^\delta}{(1+\delta)^{1+\delta}})^{\mu' \cdot n} &= (\frac{e^\delta}{e^{(1+\delta)\ln(1+\delta)}})^{\mu' \cdot n} \\ &= (e^{\delta - (1+\delta)\ln(1+\delta)})^{\mu' \cdot n}\end{aligned}$$

By expanding the ln with taylor series,

$$\begin{aligned}&= (e^{\delta - (1+\delta)(\delta - \frac{\delta^2}{2} + \cdots)})^{\mu' \cdot n} \\ &\leq (e^{-\frac{\delta^2}{2}})^{\mu' \cdot n}\end{aligned}$$

Thus we this expression.

$$P[Y' - \mu' \geq \epsilon] \leq (e^{-\frac{\delta^2}{2}})^{\mu' \cdot n}$$

Now $n \to \infty$ and making sure that $\epsilon > 0$ so that $e^{-\frac{\delta^2}{2}} < 1$,

$$\lim_{n\to\infty} P[Y' - \mu' > \epsilon] \leq \lim_{n\to\infty} (e^{-\frac{\delta^2}{2}})^{\mu' \cdot n} = 0 \tag{2.1}$$

For the lower bounding, consider another random variable Y'' constructed by $X_1' = 1 - X_1, X_2' = 1 - X_2, \ldots$. We can relate to Y' By

$$Y'' = 1 - Y'$$

$$\mu'' = 1 - \mu'$$

since $X_1', X_2', \ldots$ are aslo Bernoulli variables, we can construct a similar equation as above as

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

$$\lim_{n\to\infty} P[(1-Y')-(1-\mu')>\epsilon]=0$$

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

So by combining, we get

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

Gaussian Mixture Model

3.1 General Idea

The output of the random variable $\mathcal{A}$ is obtained by

1. pick a gaussian distribution and

2. get the value of the distribution

Let the event A_i is picking the i^{th} Gaussian distribution with $P(A_i) = p_i$. The probability of an ouput x for the random variable $\mathcal{A}$ is given by

$$P[X = x] = \sum_{i=1}^{k} P(A_i)P(X = x|A_i)$$

Here, $P(X = x|A_i)$ is the probability for the Gaussian distribution $X_i = x$

$$P[X = x] = \sum_{i=1}^{k} p_i P[X_i = x]$$

3.2 Parameter Estimation

We employ an algorithm called **expectation maximization** to find the parameters of the GMM. A general expectation maximisation algorithm involves 2 major steps :the Expectation Step (E) and the Maximisation step (M) . These steps repeat until convergence.

Step-by-Step Procedure

Step 1: Initialize Parameters

First, the parameters for each of the K components of the mixture model must be initialized. These parameters include:

- μ_k (the **mean** for each component),

- σ_k (the **variance** for each component),

- π_k (the mixing coefficient(**wieghts**) for each component, where $\sum_{k=1}^{K} \pi_k = 1$).

These can be initialized using random values, k-means clustering, or any other heuristic.

Step 2: Expectation Step (E-step)

In the E-step, we calculate the responsibility γ_{ik}, which represents the probability that a given data point x_i was generated by component k. This step involves computing the posterior probabilities using the current parameters:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \sigma_j^2)}$$

Where:

- γ_{ik} is the responsibility of component k for data point x_i,
- π_k is the mixing coefficient for component k,
- $\mathcal{N}(x_i \mid \mu_k, \Sigma_k)$ is the Gaussian probability density function for component k with mean μ_k and covariance matrix Σ_k.

This step computes the expected value of the latent variables (i.e., which Gaussian generated each data point).

Step 3: Maximization Step (M-step)

In the M-step, we update the parameters μ_k, σ_k, and π_k based on the responsibilities γ_{ik} computed in the E-step. The update equations are as follows:

- Update the means μ_k:

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma_{ik} x_i}{\sum_{i=1}^{N} \gamma_{ik}}$$

  where μ_k is the weighted average of the data points assigned to component k.

- Update the variances σ_k:

$$\sigma_k^2 = \frac{\Sigma_{i=1}^{N} \gamma_{ik} (x_i - \mu_k)^2}{\Sigma_{i=1}^{N} \gamma_{ik}}$$

  where Σ_k is the weighted covariance of data points assigned to component k.

- Update the weights π_k:

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} \gamma_{ik}$$

  where π_k is the fraction of data points assigned to component k, ensuring that $\sum_{k=1}^{K} \pi_k = 1$.

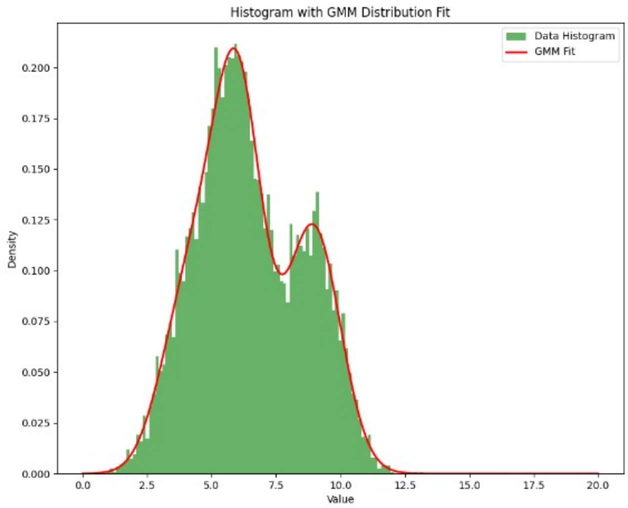
These updates maximize the expected complete log-likelihood given the current responsibilities.

Step 4: Check for Convergence

After each E-step and M-step, we check if the algorithm has converged. The convergence criterion can be based on one of the following:

- The change in log-likelihood is below a small threshold,
- The parameters (means, variances, and weights) stop changing significantly between iterations,
- A pre-set maximum number of iterations has been reached.

If the algorithm has not converged, return to the E-step and repeat the process.



Mathematical Objective

The EM algorithm seeks to maximize the log-likelihood function:

$$\log L(\theta) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2) \right)$$

where $\theta = \{\mu_k, \sigma_k, \pi_k\}$ represents the parameters of the model.

Since directly maximizing this log-likelihood is difficult due to the latent variables (which Gaussian generated each data point), the EM algorithm iteratively maximizes a lower bound on this log-likelihood.

Implementation

The result obtained by setting the value $k = 6$ in python is

Summary

The EM algorithm can be summarized in the following steps:

1. Initialize parameters μ_k, Σ_k, π_k,

2. Perform the E-step: calculate responsibilities γ_{ik},

3. Perform the M-step: update parameters μ_k, σ_k, π_k,

4. Repeat the E-step and M-step until convergence.

.

3.3 Properties

For $E[X], var[X]$, let's generate the MGF for X. It is given by

$$
\begin{aligned}
M_X(t) &= E(e^{Xt}) \\
&= \int_{-\infty}^{+\infty} e^{xt} \cdot P[X = x] \\
&= \int_{-\infty}^{+\infty} e^{xt} \cdot \sum_{i=1}^{k} p_i P[X_i = x] \\
&= \sum_{i=1}^{k} (p_i \int_{-\infty}^{+\infty} e^{xt} \cdot P[X_i = x])
\end{aligned}
$$

Since X_i is a Gaussian distribution, we can use that $M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$

$$
\begin{aligned}
M_X(t) &= \sum_{i=1}^{k} p_i \cdot e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}} \\
&= \sum_{i=1}^{k} p_i M_{X_i}(t)
\end{aligned}
$$

To obtain $E(X)$ and $E(X^2)$, we need to find $M'_X(t)$ and $M''_X(t)$ at $t = 0$

$$
\begin{aligned}
M'_X(0) &= \sum_{i=1}^{k} p_i M'_{X_i}(0) \\
&= \sum_{i=1}^{k} p_i \mu_i \\
M''_X(0) &= \sum_{i=1}^{k} p_i M''_{X_i}(0) \\
&= \sum_{i=1}^{k} p_i (\sigma_i^2 + \mu_i^2)
\end{aligned}
$$

And since $E[X] = M'_X(0)$, $E[X^2] = M''_X(0)$ and $Var[X] = E[X^2] - E[X]^2$,

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{k} p_i \mu_i \\
Var[X] &= \sum_{i=1}^{k} p_i (\sigma_i^2 + \mu_i^2) - (\sum_{i=1}^{k} p_i \mu_i)^2 \\
&= \sum_{i=1}^{k} p_i \sigma_i^2 + \sum_{i=1}^{k} p_i \mu_i^2 - (\sum_{i=1}^{k} p_i \mu_i)^2 \\
M_X(t) &= \sum_{i=1}^{k} p_i \cdot e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}
\end{aligned}
$$

3.4 Difference from Weighted sum of Gaussians

Let Y be defined as

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

where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, a Gaussian distribtuion

Although the above random variable X and the Z defined (as the weighted sum of Gaussians) look the same, they differ sublty.

In the literal sense, X becoming x is the probability that one Gaussian equals x but in Z, any one Gaussian need not be x but the sum of their weights is. Let's analyse the mean, variance to check the difference.

Let's construct the MGF for Z.

$$\begin{aligned} M_Z(t) &= E(e^{Zt}) \\ &= E(e^{t\sum_{i=1}^{k} p_i X_i}) \\ &= E(\prod_{i=1}^{k} e^{tp_i X_i}) \end{aligned}$$

Since the events or Gaussians are independent, we can take the product outside

$$\begin{aligned} M_Z(t) &= \prod_{i=1}^{k} E(e^{tp_i X_i}) \\ &= \prod_{i=1}^{k} E(e^{(tp_i) X_i}) \end{aligned}$$

We can substitute the MGFs for individual Gaussians

$$\begin{aligned} M_Z(t) &= \prod_{i=1}^{k} (e^{\mu_i(tp_i) + \frac{\sigma_i^2 (tp_i)^2}{2}}) \\ &= e^{\sum_{i=1}^{k} \mu_i(tp_i) + \frac{\sigma_i^2 (tp_i)^2}{2}} \end{aligned}$$

From the above expression, we can extract the mean and variance as

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

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

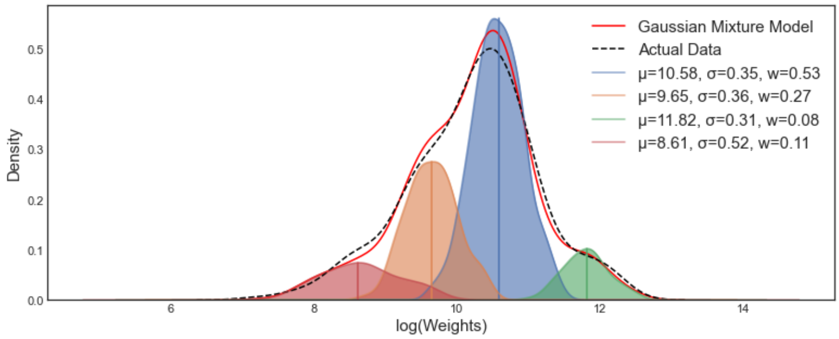
Since the MGF and PDF uniquely define each other (which is proved in the next task), we can say that the MGF of Z is same as $Z \sim \mathcal{N}(\sum_{i=1}^{k} p_i \mu_i, \sum_{i=1}^{k} p_i^2 \sigma_i^2)$. We can concude that Z **is also another Gaussian distribution**.

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

$$\boxed{M_Z(t) = e^{t\mu' + \frac{t^2 \sigma'^2}{2}}}$$

where $\mu' = \sum_{i=1}^{k} \mu_i p_i$ and $\sigma' = \sum_{i=1}^{k} p_i^2 \sigma_i^2$

The difference between Z and X is that Z is a Gaussian distribution, constructed by adding Gaussians and contains only one peak (as a Gaussian would). X is mixture of Gaussian distributions and has multiple peaks.



an example of a GMM

Uniqueness of the Moment Generating Function

To prove that MGF and PDF uniquely determine each other, we have to prove that

$$M_X(t) = M_Y(t) \iff f_X(u) = f_Y(u)$$

Finite and Discrete:

For a discrete random variable, a normal definition of PDF doesn't work. This requires a new type of function 'dirac delta function'. This function is zero everywhere but explodes to near infinity or a very large value. It can be written as

$$\lim_{a\to 0} \delta_a = \frac{1}{|a|\sqrt{\pi}} e^{-x^2/a^2}$$

For this function, altough the value at 0 explodes, the integral is always 1.

For the Finite discrete cases, let's assume the possible outcomes are $x_1, x_2, \ldots x_n$ with probabilities $p_1, p_2, \ldots p_n$. The probability density function can we expressed as the weighted sums of the above dirac delta functions.

$$f_X(u) = \lim_{a\to 0} \sum_{i=1}^{n} p_i \cdot \frac{1}{|a|\sqrt{\pi}} e^{-(u-x_i)^2/a^2}$$

With this, the MGF is expressed as

$$M(t) = \lim_{a\to 0} \sum_{i=1}^{n} e^{ut} \cdot p_i \cdot \frac{1}{|a|\sqrt{\pi}} e^{-(u-x_i)^2/a^2}$$

We can expand $M_X(t)$ and $M_Y(t)$ as

$$M_X(t) = \sum_{i=1}^{n} f_X(u) e^{ut} \tag{4.1}$$

$$M_Y(t) = \sum_{i=1}^{n} f_Y(u) e^{ut} \tag{4.2}$$

(1/2) prooiving $M_X(t) = M_Y(t) \implies f_X(u) = f_Y(u)$:

by subtracting the above two equations 4.1 and 4.2 and since $M_X(t) = M_Y(t)$,

$$0 = \sum_{i=1}^{n} (f_X(u) - f_Y(u)) e^{ut}$$

Since X and Y are finite, $f_X(u)$ and $f_Y(u)$ are non-zero at finite points. Let the points be $u_1, u_2, \ldots u_k$ and $f_X(u) - f_Y(u) = f(u)$

$$f(u_1) \cdot e^{u_1 t} + f(u_2) \cdot e^{u_2 t} + \dots + f(u_n) \cdot e^{u_n t} = 0$$

now differentiating w.r.t t for p times,

$$\begin{aligned} f(u_1) \cdot e^{u_1 t} + f(u_2) \cdot e^{u_2 t} + \dots + f(u_p) \cdot e^{u_p t} &= 0 \\ f(u_1) \cdot u_1 e^{u_1 t} + f(u_2) \cdot u_2 e^{u_2 t} + \dots + f(u_p) \cdot u_p e^{u_p t} &= 0 \\ &\vdots \\ f(u_1) \cdot u_1^p e^{u_1 t} + f(u_2) \cdot u_2^p e^{u_2 t} + \dots + f(u_p) \cdot u_p^p e^{u_p t} &= 0 \end{aligned}$$

When $t = 0$, the above equations become a system of liner equations

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ u_1 & u_2 & \dots & u_p \\ \vdots & & & \\ u_1^p & u_2^p & \dots & u_p^p \end{bmatrix}_{(p+1)\times p} \begin{bmatrix} f(u_1) \\ f(u_2) \\ \vdots \\ f(u_p) \end{bmatrix}_{p\times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(p+1)\times 1}$$

Since, the u_i are distinct, the solution to this set of equations is

$$\begin{aligned} f(u_1) = f(u_2) = \dots = f(u_p) &= 0 \\ \implies f(u) = f_X(u) = f_Y(u) &= 0 \end{aligned}$$

(2/2) prooiving $f_X(t) = f_Y(t) \implies M_X(u) = M_Y(u)$:

by subtracting the above two equations 4.3 and 4.4 and since $f_X(u) = f_Y(u)$,

$$\begin{aligned} M_X(t) - M_Y(t) &= \sum_{i=1}^{n} (0) \cdot e^{ut} \ \forall t \in \mathcal{R} \\ M_X(t) - M_Y(t) &= 0 \end{aligned}$$

Continuous

We can expand $M_X(t)$ and $M_Y(t)$ as

$$M_X(t) = \int_{-\infty}^{\infty} f_X(u) e^{ut} dt \tag{4.3}$$

$$M_Y(t) = \int_{-\infty}^{\infty} f_Y(u) e^{ut} dt \tag{4.4}$$

(1/2) prooiving $M_X(t) = M_Y(t) \implies f_X(u) = f_Y(u)$:

by subtracting the above two equations 4.3 and 4.4 and since $M_X(t) = M_Y(t)$,

$$0 = \int_{-\infty}^{\infty} (f_X(u) - f_Y(u)) e^{ut} dt \ \forall t \in \mathcal{R}$$

This equation is lot like similar to a 'Laplace transform' (substitute t with s) and since the Laplace is unique, we can say that "$f_X(u) - f_Y(u) = 0$". To give a rough proof,

Theorem 1. *if $g(x)$ is a continous function in $[0, 1]$ and $\int_0^1 g(x) x^n dx = 0$ for all the values of $n = 0, 1, 2, \dots$ then $g(x) \equiv 0$*

proof

By the Weierstrass approximation theorem we can find a polynomial $P(\epsilon, x)$ which is very close to the function $g(x)$ i.e $|P(\epsilon, x) - g(x)|_\infty < \epsilon$. This generally states that we can create a polynomial which is as close to the actual function $g(x)$. Now

$$\int_0^1 g(x)P(\epsilon, x)dx = 0$$

Since P is a polynomial with integer coefficients and if expanded, we get the above general formula $\int_0^1 g(x)x^n dx = 0$. Now since $P(\epsilon, x)$ is close enough to $g(x)$,

$$\int_0^1 g(x)g(x)dx = 0$$

Thus we can conclude that $g(x) \equiv 0$.

Theorem 2. *for the integral $\int_0^\infty f(x)e^{-xs}dx$, with the change of varibles, this can be written as $\int_0^1 x^s g(-\ln x)dx$*

From the above two lemmas, we can convert our equation and conclude that '$f_X(u) - f_Y(u) = 0$'

prooiving $f_X(t) = f_Y(t) \implies M_X(u) = M_Y(u)$:

by subtracting the above two equations 4.3 and 4.4 and since $f_X(u) = f_Y(u)$,

$$M_X(t) - M_Y(t) = \int_{-\infty}^{\infty} (0) \cdot e^{ut} dt \ \forall t \in \mathcal{R}$$
$$M_X(t) - M_Y(t) = 0$$

comparing X and Z:

as formulated above,

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

Since these MGFs are not the same, we can conclude that X and Z have different probability Density Functions