Assignment

Aryan Jain - EE22BTECH11011*

Question: Suppose that $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}, \mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_n}$ are independent and identically distributed random vectors each having $N_p(\boldsymbol{\mu}, \Sigma)$ distributions, where Σ is non-singular, $p > 1$ and $n > 1$. If $\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X_i}$ and $\mathbf{Y} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y_i}$, then which one of the following statements is true?

(a) There exists $c > 0$ such that $c(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})$ has χ^2-distribution with p degrees of freedom.
(b) There exists $c > 0$ such that $c(\mathbf{X} - \mathbf{Y})^T \Sigma^{-1} (\mathbf{X} - \mathbf{Y})$ has χ^2-distribution with $(p - 1)$ degrees of freedom.
(c) There exists $c > 0$ such that $c\sum_{i=1}^{n} (\mathbf{X_i} - \mathbf{X})^T \Sigma^{-1} (\mathbf{X_i} - \mathbf{X})$ has χ^2-distribution with p degrees of freedom.
(d) There exists $c > 0$ such that $c\sum_{i=1}^{n} (\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})^T \Sigma^{-1} (\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})$ has χ^2-distribution with p degrees of freedom.

GATE ST Paper 2023

Solution:
We are given that,

$$\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}, \mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_n} \sim N_p(\boldsymbol{\mu}, \Sigma) \tag{1}$$

Also,

$$\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X_i} \tag{2}$$

$$\mathbf{Y} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y_i} \tag{3}$$

The mean of $\mathbf{X}$ is given by:

$$\boldsymbol{\mu}_{\mathbf{X}} = E(\mathbf{X}) \tag{4}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{X_i}) \tag{5}$$

$$= \boldsymbol{\mu} \tag{6}$$

Similarly,

$$\boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu} \tag{7}$$

The covariance of $\mathbf{X}$ is given by:

$$\begin{aligned}
\Sigma_{\mathbf{X}} &= E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\right] && (8)\\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X_i}-\boldsymbol{\mu}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X_i}-\boldsymbol{\mu}\right)^T\right] && (9)\\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n}(\mathbf{X_i}-\boldsymbol{\mu})(\mathbf{X_i}-\boldsymbol{\mu})^T\right] && (10)\\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n}E\left(\mathbf{X_i}^2+\boldsymbol{\mu}^2-2\boldsymbol{\mu}\mathbf{X_i}\right)\right] && (11)\\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n}E\left(\mathbf{X_i}^2\right)+\sum_{i=1}^{n}E\left(\boldsymbol{\mu}^2\right)-2\boldsymbol{\mu}\sum_{i=1}^{n}E(\mathbf{X_i})\right] && (12)\\
&= \frac{1}{n^2}\left[n\Sigma+n\boldsymbol{\mu}^2+n\boldsymbol{\mu}^2-2\boldsymbol{\mu}^2\right] \quad \left[\because E\left(\mathbf{X_i}^2\right)=\Sigma_{\mathbf{X_i}}+E(\mathbf{X_i})^2\right] && (13)\\
&= \frac{\Sigma}{n} && (14)
\end{aligned}$$

Similarly,

$$\Sigma_{\mathbf{Y}} = \frac{\Sigma}{n} \tag{15}$$

(a) To check option (A):
let us say,

$$\mathbf{A} = c(\mathbf{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) \tag{16}$$

$$\tag{17}$$

And,

$$\begin{aligned}
\Sigma^{-1} &= \mathbf{F}^T\mathbf{F} && (18)\\
\mathbf{y} &= \mathbf{F}(\mathbf{X}-\boldsymbol{\mu}) && (19)\\
\implies \mathbf{A} &= c\mathbf{y}^T\overline{\mathbf{y}} && (20)\\
&= c\|\mathbf{y}\|^2 && (21)
\end{aligned}$$

Equation (21) shows that $\mathbf{A}$ can have χ^2-distribution.
To confirm that we will find the mean and covariance-matrix of $\overline{\mathbf{y}}$.

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{y}} &= E(\mathbf{y}) && (22)\\
&= E(\mathbf{F}(\mathbf{X}-\boldsymbol{\mu})) && (23)\\
&= \mathbf{F}[E(\mathbf{X})-E(\boldsymbol{\mu})] && (24)\\
&= \mathbf{F}[\boldsymbol{\mu}-\boldsymbol{\mu}] \quad from(6) && (25)\\
&= 0 && (26)
\end{aligned}$$

And,

$$\begin{aligned}
\Sigma_{\mathbf{y}} &= E\left[\left(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{y}}\right)\left(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{y}}\right)^T\right] && (27)\\
&= E\left[(\mathbf{F}(\mathbf{X}-\boldsymbol{\mu}))(\mathbf{F}(\mathbf{X}-\boldsymbol{\mu}))^T\right] && (28)\\
&= E\left[\mathbf{F}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\mathbf{F}^T\right] && (29)\\
&= \mathbf{F}\left[E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\right]\right]\mathbf{F}^T && (30)\\
&= \mathbf{F}\Sigma\mathbf{F}^T && (31)
\end{aligned}$$

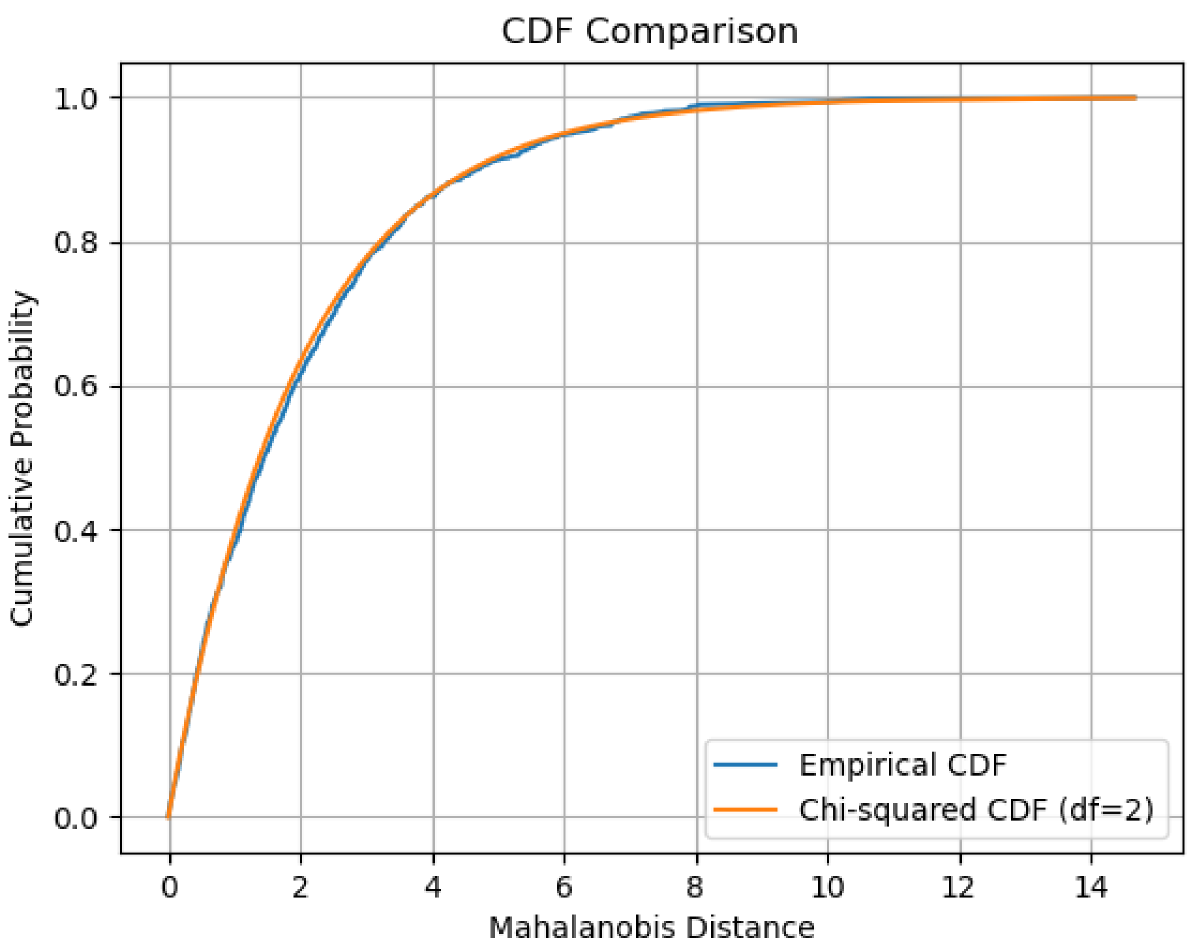
since,

$$\Sigma^{-1} = \mathbf{F}^T\mathbf{F} \tag{32}$$
$$\Sigma\Sigma^{-1} = \Sigma\mathbf{F}^T\mathbf{F} \tag{33}$$
$$\mathbf{I} = \Sigma\mathbf{F}^T\mathbf{F} \tag{34}$$
$$\mathbf{I}\mathbf{F}^{-1} = \Sigma\mathbf{F}^T \tag{35}$$
$$\mathbf{F}\mathbf{F}^{-1} = \mathbf{F}\Sigma\mathbf{F}^T \tag{36}$$
$$\mathbf{I} = \mathbf{F}\Sigma\mathbf{F}^T \tag{37}$$

So using (37),

$$\Sigma_{\mathbf{y}} = \mathbf{I} \tag{38}$$

Hence, For $c = 1$ $\mathbf{A}$ has χ^2-distribution with p degrees of freedom.
So option (A) is correct.



(b) To check option (B):
Let us say,

$$\mathbf{B} = c(\mathbf{X} - \mathbf{Y})^T\Sigma^{-1}(\mathbf{X} - \mathbf{Y}) \tag{39}$$

And,

$$\Sigma^{-1} = \mathbf{F}^T\mathbf{F} \tag{40}$$
$$\mathbf{y} = \mathbf{F}(\mathbf{X} - \mathbf{Y}) \tag{41}$$
$$\implies \mathbf{B} = c\mathbf{y}^T\overline{\mathbf{y}} \tag{42}$$
$$= c||\mathbf{y}||^2 \tag{43}$$

Equation (43) shows that **B** can have χ^2-distribution.
To confirm that we will find the mean and covariance-matrix of $\overline{\mathbf{y}}$.

$$\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}) \tag{44}$$
$$= E[F(\mathbf{X} - \mathbf{Y})] \tag{45}$$
$$= F[E(\mathbf{X}) - E(\mathbf{Y})] \tag{46}$$
$$= F[\boldsymbol{\mu} - \boldsymbol{\mu}] \tag{47}$$
$$= 0 \tag{48}$$

And,

$$\Sigma_{\mathbf{y}} = E\left[\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)^T\right] \tag{49}$$
$$= E\left[(\mathbf{F}(\mathbf{X} - \mathbf{Y}))(\mathbf{F}(\mathbf{X} - \mathbf{Y}))^T\right] \tag{50}$$
$$= E\left[\mathbf{F}(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})^T\mathbf{F}^T\right] \tag{51}$$
$$= \mathbf{F}\left[E\left[(\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})^T\right]\right]\mathbf{F}^T \tag{52}$$
$$= \mathbf{F}\left[E\left[||\mathbf{X} - \mathbf{Y}||^2\right]\right]\mathbf{F}^T \tag{53}$$
$$= \mathbf{F}\left[E\left(\mathbf{X}^2\right) + E\left(\mathbf{Y}^2\right) - E(2\mathbf{X}\mathbf{Y})\right]\mathbf{F}^T \tag{54}$$
$$= \mathbf{F}\left[\frac{\Sigma}{n} + \boldsymbol{\mu}^2 + \frac{\Sigma}{n} + \boldsymbol{\mu}^2 - 2\boldsymbol{\mu}^2\right]\mathbf{F}^T \quad \left[\because E\left(\mathbf{X}^2\right) = \Sigma_{\mathbf{X}} + E(\mathbf{X})^2\right] \tag{55}$$
$$= \frac{2}{n}\mathbf{F}\Sigma\mathbf{F}^T \tag{56}$$
$$= \frac{2}{n}\mathbf{I} \tag{57}$$

Hence, for $c = \frac{n}{2}$, **B** has χ^2-distribution with p degrees of freedom.
So option (B) is incorrect.

(c) To check option (C):
let us say,

$$\mathbf{C} = c\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{X})^T\Sigma^{-1}(\mathbf{X_i} - \mathbf{X}) \tag{58}$$

And,

$$\Sigma^{-1} = \mathbf{F}^T\mathbf{F} \tag{59}$$
$$\mathbf{y} = \mathbf{F}\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{X})\right) \tag{60}$$
$$\implies \mathbf{C} = c\mathbf{y}^T\overline{\mathbf{y}} \tag{61}$$
$$= c||\mathbf{y}||^2 \tag{62}$$

Equation (62) shows that $\mathbf{C}$ can have χ^2-distribution.
To confirm that we will find the mean and covariance-matrix of $\overline{\mathbf{y}}$.

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{y}} &= E(\mathbf{y}) && (63)\\
&= E\left[\mathbf{F}\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{X})\right)\right] && (64)\\
&= \mathbf{F}\left[\sum_{i=1}^{n}(E(\mathbf{X_i}) - E(\mathbf{X}))\right] && (65)\\
&= \mathbf{F}\left[E(X_1) - E(X) + E(X_2) - E(X) + \ldots + E(X_n) - E(X)\right] && (66)\\
&= 0 && (67)
\end{aligned}$$

And,

$$\begin{aligned}
\Sigma_{\mathbf{y}} &= E\left[\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)^T\right] && (68)\\
&= \mathbf{F}E\left[\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{X})\right)\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{X})\right)^T\right]\mathbf{F}^T && (69)\\
&= \mathbf{F}E\left[\left(\sum_{i=1}^{n}\mathbf{X_i} - n\mathbf{X}\right)\left(\sum_{i=1}^{n}\mathbf{X_i} - n\mathbf{X}\right)^T\right]\mathbf{F}^T && (70)\\
&= \mathbf{F}E\left[(n\mathbf{X} - n\mathbf{X})(n\mathbf{X} - n\mathbf{X})^T\right]\mathbf{F}^T && (71)\\
&= \mathbf{F}\mathbf{0}\mathbf{F}^T && (72)\\
&= \mathbf{0} && (73)
\end{aligned}$$

Hence, There is no value of $c > 0$ for which $\mathbf{C}$ have χ^2-distribution.
So option (C) is incorrect.

(d) To check option (D):
let us say,

$$\mathbf{D} = c\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})^T\Sigma^{-1}(\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y}) \qquad (74)$$

And,

$$\begin{aligned}
\Sigma^{-1} &= \mathbf{F}^T\mathbf{F} && (75)\\
\mathbf{y} &= \mathbf{F}\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})\right) && (76)\\
\implies \mathbf{C} &= c\mathbf{y}^T\overline{\mathbf{y}} && (77)\\
&= c\|\mathbf{y}\|^2 && (78)
\end{aligned}$$

Equation (78) shows that $\mathbf{D}$ can have χ^2-distribution.
To confirm that we will find the mean and covariance-matrix of $\overline{\mathbf{y}}$.

$$\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y}) \tag{79}$$
$$= E\left(\mathbf{F}\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})\right)\right) \tag{80}$$
$$= \mathbf{F}E\left[\sum_{i=1}^{n}\mathbf{X_i} - \sum_{i=1}^{n}\mathbf{Y_i} - n\mathbf{X} + n\mathbf{Y}\right] \tag{81}$$
$$= \mathbf{F}\left[\sum_{i=1}^{n}E(\mathbf{X_i}) - \sum_{i=1}^{n}E(\mathbf{Y_i}) - nE(\mathbf{X}) + nE(\mathbf{Y})\right] \tag{82}$$
$$= \mathbf{F}[n\boldsymbol{\mu} - n\boldsymbol{\mu} - n\boldsymbol{\mu} + n\boldsymbol{\mu}] \tag{83}$$
$$= 0 \tag{84}$$

And,

$$\Sigma_{\mathbf{y}} = E\left[\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)^T\right] \tag{85}$$
$$= \mathbf{F}E\left[\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})\right)\left(\sum_{i=1}^{n}(\mathbf{X_i} - \mathbf{Y_i} - \mathbf{X} + \mathbf{Y})\right)^T\right]\mathbf{F}^T \tag{86}$$
$$= \mathbf{F}E\left[\left(\sum_{i=1}^{n}\mathbf{X_i} - \sum_{i=1}^{n}\mathbf{Y_i} - n\mathbf{X} + n\mathbf{Y}\right)\left(\sum_{i=1}^{n}\mathbf{X_i} - \sum_{i=1}^{n}\mathbf{Y_i} - n\mathbf{X} + n\mathbf{Y}\right)^T\right]\mathbf{F}^T \tag{87}$$
$$= \mathbf{F}E\left[(n\mathbf{X} - n\mathbf{Y} - n\mathbf{X} + n\mathbf{Y})(n\mathbf{X} - n\mathbf{Y} - n\mathbf{X} + n\mathbf{Y})^T\right]\mathbf{F}^T \tag{88}$$
$$= \mathbf{F}\mathbf{0}\mathbf{F}^T \tag{89}$$
$$= \mathbf{0} \tag{90}$$

Hence, There is no value of $c > 0$ for which $\mathbf{D}$ have χ^2-distribution.
So option (D) is incorrect.

Steps for simulation:

1) Firstly in the the file "gauss.c", I have generated 1000 random vectors with dimension 2 using Box-Muller method and listed the data in the file "randomvectors.dat".
2) Then in the file "distance.c", using the random vectors generated in the first step, I found the value of $c(\mathbf{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})$ distribution which will give us 1×1 matrix.
3) So as we have generated 1000 random vectors in first step, we will have 1000 values of the distribution.
4) Then I have listed the values of the distribution $c(\mathbf{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})$ that I got in the file "mahalanobisdistances.dat".
5) Now in the file "cdf.py", I have plotted the cdf of the distribution $c(\mathbf{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})$ and also plotted the theoretical cdf plot of a χ^2 distribution.

The variables that are used in the simulation are:

Variable	Definition
$\mathbf{X}$	random vector
p	dimension of vector
n	number of vectors
$\boldsymbol{\mu}$	mean
$\boldsymbol{\Sigma}$	Covariance