Lecture 10 / Week 6

Properties of Multivariate Normal Distribution

Example Assume we have the random vector $X \sim N_n(\mu, \sum)$ and $Y_{(n\times 1)} = AX + b$, where A is a (n×n) nonsingular matrix and b is a (n×1) vector. Find the probability distribution of Y.

We know that X has multivariate normal distribution, hence its density function is the following

$$f_X(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}.\det\Sigma^{-\frac{1}{2}}e^{-\frac{1}{2}\{(\mathbf{x}-\boldsymbol{\mu})^T.\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}}$$

To find the density function of Y, we have to do the following transformation

$$\begin{aligned} Y &= g(X) \Leftrightarrow X = g^{-1}(Y) \\ X_{(n\times 1)} &= A^{-1}_{(n\times n)}.(Y-b)_{(n\times 1)} = g^{-1}(Y) \end{aligned}$$

Then we can apply the transformation formula

$$f_Y(y) = f_X(g^{-1}(y)).|\det J(y)|$$

First of all, notice that $J(y) = A^{-1}$. We plug it into the formula

$$\begin{aligned} f_Y(\mathbf{y}) &= (2\pi)^{-\frac{n}{2}}.\det\Sigma^{-\frac{1}{2}}e^{-\frac{1}{2}\{(A^{-1}.(\mathbf{y}-\mathbf{b})-\boldsymbol{\mu})^T.\Sigma^{-1}(A^{-1}.(\mathbf{y}-\mathbf{b})-\boldsymbol{\mu})\}}.|\det A^{-1}| = \\ &= (2\pi)^{-\frac{n}{2}}.|A|^{-1}.|\Sigma|^{-\frac{1}{2}}.e^{-\frac{1}{2}\{(\mathbf{A}^{-1}\mathbf{y}-\mathbf{A}^{-1}\mathbf{b}-\boldsymbol{\mu})^T.\Sigma^{-1}(\mathbf{A}^{-1}\mathbf{y}-\mathbf{A}^{-1}\mathbf{b}-\boldsymbol{\mu})\}} \\ \text{notice that } |\det A^{-1}| &= |A|^{-1}, \det\Sigma^{-\frac{1}{2}} = |\Sigma|^{-\frac{1}{2}} \\ |A|^{-1} &= |A^2|^{-\frac{1}{2}} = |A.A^T|^{-\frac{1}{2}} \Rightarrow |A.A^T|^{-\frac{1}{2}}.|\Sigma|^{-\frac{1}{2}} = |A.\Sigma.A^T|^{-\frac{1}{2}} \end{aligned}$$

$= (2\pi)^{-\frac{n}{2}}.|A.\Sigma.A^T|^{-\frac{1}{2}}.e^{-\frac{1}{2}\{(\mathbf{A}^{-1}\mathbf{y}-\mathbf{A}^{-1}\mathbf{b}-\mathbf{A}^{-1}.A.\boldsymbol{\mu})^T.\Sigma^{-1}(\mathbf{A}^{-1}\mathbf{y}-\mathbf{A}^{-1}\mathbf{b}-\mathbf{A}^{-1}.\mathbf{A}\boldsymbol{\mu})\}}$
notice that we multiplied μ with $A^{-1}.A = \mathbf{I}_{(n\times n)}$
$= (2\pi)^{-\frac{n}{2}}.|A.\Sigma.A^T|^{-\frac{1}{2}}.e^{-\frac{1}{2}\{(\mathbf{y}-\mathbf{b}-A.\boldsymbol{\mu})^T(\mathbf{A}^{-1})^T.\Sigma^{-1}\mathbf{A}^{-1}.(\mathbf{y}-\mathbf{b}-\mathbf{A}\boldsymbol{\mu})\}}$
we take out $\mathbf{A}^{-1}$.(Note transpose gets out as transpose)
$= (2\pi)^{-\frac{n}{2}}.|A.\Sigma.A^T|^{-\frac{1}{2}}.e^{-\frac{1}{2}\{(\mathbf{y}-\mathbf{b}-A.\boldsymbol{\mu})^T(A.\Sigma.A^T)^{-1}.(\mathbf{y}-\mathbf{b}-\mathbf{A}\boldsymbol{\mu})\}}$.

$$= (2\pi)^{-\frac{n}{2}}.|A.\Sigma.A^T|^{-\frac{1}{2}}.e^{-\frac{1}{2}\{(\mathbf{y}-(\mathbf{b}+A.\boldsymbol{\mu}))^T(A.\Sigma.A^T)^{-1}.(\mathbf{y}-(\mathbf{b}+\mathbf{A}\boldsymbol{\mu}))\}}$$

Thus we have shown that $Y \sim N_n(A.\boldsymbol{\mu}+\mathbf{b}, A.\Sigma.A^T)$.

Exercise Show that given $g^{-1}(Y) = X = A^{-1}.(Y-b)$, then $J(y) = A^{-1}$.

Proof Recall that $J_{ij}(y) = \frac{\partial g_i^{-1}}{\partial y_j}$. Since $X_{(n\times 1)}$ and $Y_{(n\times 1)}$, then i=n, j=n, so we will have an $n\times n$ matrix. (A^{-1} being a $n\times n$ matrix satisfies this.) Let's see with n=2 case. Then we will have $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} . \begin{pmatrix} Y_1 - b_1 \\ Y_2 - b_2 \end{pmatrix} = \begin{pmatrix} a_{11}(Y_1 - b_1) + a_{12}.(Y_2 - b_2) \\ a_{21}(Y_1 - b_1) + a_{22}.(Y_2 - b_2) \end{pmatrix}$. Then

$$\frac{\partial g_i^{-1}}{\partial y_j} = \frac{\partial X_i}{\partial Y_j} = \begin{pmatrix} \frac{\partial X_1}{\partial Y_1} & \frac{\partial X_2}{\partial Y_1} \\ \frac{\partial X_1}{\partial Y_2} & \frac{\partial X_2}{\partial Y_2} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = A^{-1}.$$

Properties

Theorem Let $X \sim N_n(\mu, \sum)$ and $Y = AX + b$ with $A_{(n\times n)}$ non-singular matrix. Then $Y \sim N_n(A.\boldsymbol{\mu} + \mathbf{b}, A.\Sigma.A^T)$.

Proof We have just proved in the above example.

Assume we have $Y \sim N_n(\mu, \sum)$ and also rank$(\sum) = K$. We wonder if we can define a normal distribution once we have a singular $\sum(\equiv n > K)$. Recall that

$$\sum = \begin{pmatrix} \sum_{11} & \sum_{21} \\ \sum_{12} & \sum_{22} \end{pmatrix}$$

We relax the assumption that the whole $\sum$ is nonsingular, but we will only assume that $\sum_{11}$ is a $(k \times k)$ nonsingular matrix, i.e. $\sum$ itself can be singular or not. Then we claim that the vector $Y = \begin{pmatrix} Y_{1_{(n\times 1)}} \\ Y_{2_{(n-k\times 1)}} \end{pmatrix}_{(n\times 1)} \sim N_n\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sum_{11} & \sum_{21} \\ \sum_{12} & \sum_{22} \end{pmatrix}\right)$ if $Y_1 \sim N_k\left((\mu_1), (\sum_{11})\right)$ where $\sum_{11}$ is nonsingular and $Y_2 = AY_1 + b$ for some A and b.

Example $\sum = \begin{pmatrix} \mathbf{1} & \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{1} & 1 \\ 1 & 1 & 2 \end{pmatrix}$, $\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. We can see that $\sum$ is singular, since the third row is a linear combination of the first two rows. (det$\sum = 0$). In fact rank$(\sum) = 2$. Then we can split the Y vector into $Y = \begin{pmatrix} Y_{1(2\times 1)} \\ Y_{2_{(1\times 1)}} \end{pmatrix}_{(3\times 1)}$ and define $Y_1 \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ and $Y_2 = A_{(1\times 2)}.Y_{1_{(2\times 1)}} + b$. Notice that $\sum_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is nonsingular.(n=K=2), where $E(Y_2) = 0, var(Y_2) = 2$.

Theorem Let $X \sim N_n(\mu, \sum)$ and $Y = AX + b$ and $A_{(m\times n)}$. (notice not a sqaure matrix.) and b be vector of size m, s.t

$$Y_{(m\times 1)} = A_{(m\times n)}.X_{(n\times 1)} + b_{(m\times 1)}$$

Then $Y \sim N_m(A\boldsymbol{\mu} + \mathbf{b},\ A.\Sigma.A^T)$.

The above result holds whatever dimension and whatever rank of A.

Example $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $Y_{(3\times 1)} = A_{(3\times 2)}X_{(2\times 1)} + b_{(3\times 1)}$. Then $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$ has a normal distribution. Note that variance covariance matrix of Y is singular. (rank$(A_{(3\times 2)}.\Sigma_{(2\times 2)}.A^T_{(2\times 3)})_{(3\times 3)} = 2$, because rank$(\Sigma_{(2\times 2)}) = 2$).

Or another example would be $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$, $Y = A_{(2\times 3)}X_{(3\times 1)} + b_{(2\times 1)}$, then $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2(A\mu + b, (A.\Sigma.A^T)_{(2\times 2)})$.

Another interesting case is when we have $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{bmatrix}_{(nx1)}$ and

$$A = (I_k \mid 0) = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & 0 & 0 \end{pmatrix}_{(k\times n)}.$$

$$\text{Then } Y = A.X = (I_k \mid 0). \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ . \\ \mathbf{X}_k \\ X_{k+1} \\ . \\ X_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_k \end{bmatrix} \sim N_k\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_k \end{bmatrix}, var\begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_k \end{bmatrix}\right).$$

Theorem Let X_1 and X_2 be random vectors of dimensions k and $(n-k)$, respectively. If $\begin{pmatrix} \mathbf{X}_{1_{(n\times 1)}} \\ \mathbf{X}_{2_{(n-k\times 1)}} \end{pmatrix}_{(n\times 1)} \sim N_n\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sum_{11} & \sum_{21} \\ \sum_{12} & \sum_{22} \end{pmatrix}\right)$. Then $X_1 \sim N_k\left((\mu_1), (\sum_{11})\right)$ and $X_2 \sim N_{n-k}\left((\mu_2), (\sum_{22})\right)$.

Example $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3$ and then $X_1 \sim N$, $X_2 \sim N$, $X_3 \sim N$. $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2$, $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N_2$, $\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} \sim N_2$.

This theorem asserts that every subvector of a random vector with a normal distribution has a normal distribution.

Now we will state two important properties of multivariate normal distribution:

1. As we have shown in the first example, normality is preserved by linear transformation: $Y = AX + b$.

2. Even though in general no covariance does not apply indepence (recall yesterday's example), in normal distribution case $cov(X_1, X_2) = 0 \Rightarrow$ *indepence* of X_1 and X_2 .(i.e when $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_n$. This property will be formalized and proved in the following theorem.

Theorem Let X_1 and X_2 be random vectors such that $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_n$.(*i.e* it has multivariate normal distribution.). Then X_1 and X_2 are indepent if and only if $cov(X_1, X_2) = 0$.

Proof The theorem says $cov(\mathbf{X}_1, \mathbf{X}_2) = 0 \Leftrightarrow$ *indepence* of X_1 and X_2 under normal distribution, so we have to prove both directions, but "$\Leftarrow$ " is already shown yesterday and it holds in general regardless of the underlying distribution. So it left to prove that under multivariate normal distribution "$\Rightarrow$ " holds.

"$\Rightarrow$ " Suppose $cov(X_1, X_2) = 0$, under normality assumption we have

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}_{(n \times 1)} \sim N_n \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sum_{11} & 0 \\ 0 & \sum_{22} \end{pmatrix} \right)$$

Recall the properties of diagonal matrices:

$$\left| \begin{pmatrix} \sum_{11} & 0 \\ 0 & \sum_{22} \end{pmatrix} \right| = \left| \sum_{11} \right| \cdot \left| \sum_{22} \right|$$

$$\begin{pmatrix} \sum_{11} & 0 \\ 0 & \sum_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \sum_{11}^{-1} & 0 \\ 0 & \sum_{22}^{-1} \end{pmatrix}$$

Be cautious because the above properties do not hold in general, only if we have diagonal matrix.(i.e off-diagonal elements are 0's.)

Then we can write the joint density of X_1 and X_2 in the following way,

$$f_{X_1,X_2}(\mathbf{x}_1, \mathbf{x}_2) = (2\pi)^{-\frac{n}{2}} . |\Sigma_{11}|^{-\frac{1}{2}} . |\Sigma_{22}|^{-\frac{1}{2}} \exp^{\left\{ -\frac{1}{2} \begin{pmatrix} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ (\mathbf{x}_2 - \boldsymbol{\mu}_2) \end{pmatrix}^T . \begin{pmatrix} \sum_{11}^{-1} & 0 \\ 0 & \sum_{22}^{-1} \end{pmatrix} \begin{pmatrix} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ (\mathbf{x}_2 - \boldsymbol{\mu}_2) \end{pmatrix} \right\}}$$

Note that we use the previous result that X_1 and X_2 are normally distributed. Then multiplying out the term in the power of the exponential we have

$$f_{X_1,X_2}(\mathbf{x}_1, \mathbf{x}_2) = (2\pi)^{-\frac{n}{2}} . |\Sigma_{11}|^{-\frac{1}{2}} . |\Sigma_{22}|^{-\frac{1}{2}} \exp^{-\frac{1}{2}\left\{ ((\mathbf{x}_1 - \boldsymbol{\mu}_1))^T . \sum_{11}^{-1} .(\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T . \sum_{22}^{-1} .(\mathbf{x}_2 - \boldsymbol{\mu}_2) \right\}}$$

Also notice that $((\mathbf{x}_1 - \boldsymbol{\mu}_1))^T . \sum_{11}^{-1} .(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ and $(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T . \sum_{22}^{-1} .(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ are scalars. ($A.\Sigma.A^T = matrix, A^T.\Sigma.A = scalar.$) Now we can split it into a product of two terms which will turn out to be the densities of both vectors X_1 and X_2, which completes the proof, so it follows

$= (2\pi)^{-\frac{k}{2}} . |\Sigma_{11}|^{-\frac{1}{2}} . \exp^{-\frac{1}{2}\left\{((\mathbf{x}_1 - \boldsymbol{\mu}_1))^T . \sum_{11}^{-1} .(\mathbf{x}_1 - \boldsymbol{\mu}_1)\right\}} .(2\pi)^{-\frac{n-k}{2}} . |\Sigma_{22}|^{-\frac{1}{2}} \exp^{\left\{(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T . \sum_{22}^{-1} .(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right\}}$

$f_{X_1,X_2}(\mathbf{x}_1, \mathbf{x}_2) = f_{X_1}(\mathbf{x}_1).f_{X_2}(\mathbf{x}_2). \Rightarrow X_1 and\ X_2$ are independent. QED.

Keep in mind this result because it has nice consequences that will be useful in most of the applications.

Exercise Show that $\begin{pmatrix} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ (\mathbf{x}_2 - \boldsymbol{\mu}_2) \end{pmatrix}^T . \begin{pmatrix} \sum_{11}^{-1} & 0 \\ 0 & \sum_{22}^{-1} \end{pmatrix} . \begin{pmatrix} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ (\mathbf{x}_2 - \boldsymbol{\mu}_2) \end{pmatrix} = \left\{((\mathbf{x}_1 - \boldsymbol{\mu}_1))^T . \sum_{11}^{-1} .(\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T . \sum_{22}^{-1} .(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right\}$.

Theorem Let $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$ and $\mathbf{Y} = B_{(k\times n)} X_{(n\times 1)}$, $\mathbf{Z} = C_{((l\times n)} X_{(n\times 1)}$. Then Y and Z are independent if and only if $BC^T = 0$.

Proof We know that $\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}_{((k+l)\times 1)} = \begin{pmatrix} \mathbf{B} \\ \mathbf{C} \end{pmatrix}_{((k+l)\times n)} .\mathbf{X}$. Then $\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{B}\mu \\ \mathbf{C}\mu \end{pmatrix}, \left(\begin{pmatrix} \mathbf{B} \\ \mathbf{C} \end{pmatrix} .\mathbf{I}_n \begin{pmatrix} \mathbf{B}^T & \mathbf{C}^T \end{pmatrix}\right)\right)$.

By multiplying the variance term we find

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{B}\mu \\ \mathbf{C}\mu \end{pmatrix}, \begin{pmatrix} \mathbf{B}.\mathbf{B}^T & \mathbf{B}.\mathbf{C}^T \\ \mathbf{C}.\mathbf{B}^T & \mathbf{C}.\mathbf{C}^T \end{pmatrix}\right)$$

Notice that this theorem uses both properties of the multivariate normal distribution. $\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$ is normally distributed because of the first property and the second property tells us that Y and Z are independent if and only if $cov(\mathbf{Y}, \mathbf{Z}) = 0 \Leftrightarrow \mathbf{B}.\mathbf{C}^T = 0 = \mathbf{C}.\mathbf{B}^T$.

We might also be interested if the above theorem holds in case of quadratic transformation.

Let $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$ and let C be a nonsingular, symmetric matrix such that $\mathbf{Y} = B_{(k\times n)} X_{(n\times 1)}$ and $\mathbf{Z} = X^T_{(1\times n)}.C_{(n\times n)} X_{(n\times 1)}$. Note that the first one is a linear and the second one is a quadratic transformation. (Intuitively, we can think in terms of scalars y=b.x vs. z=c.x^2). Then we have

$$\mathbf{Z} = X^T.C.X = X^T.C.C^{-1}.C.X = (C.X)^T.C^{-1}.C.X = g(C.X)$$

the last equality tells us that the quadratic transformation is a linear function. Then if $B^T.C = B.C = 0 \Rightarrow Y$ and $C.X$ are independent$\Rightarrow Y$ and

$g(C.X)$ are independent$\Rightarrow$ Y and Z are independent. The following theorem formalizes this:

Theorem Let $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$ and let C be symmetric matrix such that $\mathbf{Y} = B_{(k\times n)}X_{(n\times 1)}$ and $\mathbf{Z} = X^T_{(1\times n)}.C_{(n\times n)}X_{(n\times 1)}$. Then if $B.C = 0$, then Y and Z are independent.

Corollary If $(X_1, X_2.....X_n)$ is a random sample from $N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ and sample variation S^2 are independent. Note that $\bar{X} = B.X$ and $S^2 = X^T.C.X.$).

Theorem Let $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$ and let B, C be symmetric matrices such that $Y = \mathbf{X}^T_{(1\times n)}.\mathbf{B}_{(n\times n)}\mathbf{X}_{(n\times 1)}$ and $Z = \mathbf{X}^T_{(1\times n)}.\mathbf{C}_{(n\times n)}\mathbf{X}_{(n\times 1)}$. If $B.C = 0$, then Y and Z are independent.

Example Let $(X_1, X_2.....X_n) \sim^{i.i.d} N(0,1), \mathbf{B} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \mathbf{C} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{pmatrix}$. Then $\mathbf{B}.\mathbf{C} = \mathbf{0}_{(n\times n)}$.

$$
\begin{aligned}
Y &= \mathbf{X}^T_{(1\times n)}.\mathbf{B}_{(n\times n)}\mathbf{X}_{(n\times 1)} = \begin{pmatrix} X_1 & . & . & X_n \end{pmatrix} \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} X_1 \\ . \\ . \\ X_n \end{pmatrix} = \sum_{i=1}^{k} X_i^2 \\
Z &= \mathbf{X}^T_{(1\times n)}.\mathbf{C}_{(n\times n)}\mathbf{X}_{(n\times 1)} = \begin{pmatrix} X_1 & . & . & X_n \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{pmatrix} \begin{pmatrix} X_1 \\ . \\ . \\ X_n \end{pmatrix} = \sum_{i=k+1}^{n} X_i^2
\end{aligned}
$$

Notice from the summation indices that Y and Z are independent. Furthermore, in this special case where we had $(X_1, X_2.....X_n) \sim^{i.i.d} N(0,1)$ assumption, we have that

$$
\begin{aligned}
Y &= \mathbf{X}^T.\mathbf{B}.\mathbf{X} = \mathbf{X}^T.\begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.\mathbf{X} \\
Y &= \mathbf{X}^T.\mathbf{M}.\mathbf{X} \sim \mathcal{X}^2_k
\end{aligned}
$$

We also want to know under which conditions of $\mathbf{M}$, the random variable Y has a $\mathcal{X}^2$-distribution. The following theorem answers this question.

Theorem Let $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$ and let $\mathbf{M}$ be a symmetric *idempotent* matrix. (i.e.$\mathbf{M}^2 = \mathbf{M}$, also called *projection*, e.g identity matrix: $\mathbf{I}_n.\mathbf{I}_n = \mathbf{I}_n$). Let K be the rank of $\mathbf{M}$. Then $\mathbf{X}^T.\mathbf{M}.\mathbf{X} \sim \mathcal{X}^2_k$.

Proof Omitted.

Assume we have $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}_{(n\times 1)} \sim N_n\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sum_{XX} & \sum_{XY} \\ \sum_{YX} & \sum_{YY} \end{pmatrix}\right)$. What is the conditional distribution of Y given $X = x$?

The following theorem provides an answer to this question.

Theorem $Y \mid X = x \sim N\left(\mu_Y + \sum_{YX} \cdot \sum_{XX}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X), \sum_{YY} - \sum_{YX} \sum_{XX}^{-1} \sum_{XY}\right)$

Proof Only an outline of the proof will be sketched. Let $U = Y - \mu_Y - \sum_{YX} \sum_{XX}^{-1}(X - \mu_X)$. Then we have the following observations

$$\begin{aligned}
E(U) &= 0 \\
cov(U, X) &= 0 \Rightarrow U \text{ and } X \text{ are independent} \\
E(U \mid X) &= E(U) \\
var(U \mid X) &= var(U) \\
\begin{pmatrix} U \\ X \end{pmatrix} &\sim N \\
U \mid X &\sim N \rightarrow Y| \ X \sim N
\end{aligned}$$