Lecture 9 / Week 6

Multivariate Normal Distribution

OUTLINE

1) Expectation and Variance of Random Vectors
2) Definition of MND
3) Properties of MND

Expectation and Variance of Random Vectors

In this section we will introduce vector and matrix notation. From now on, a vector is always defined as a **column** vector. The superscript T, will be used to denote the transpose of the vector which will be row vector. Hence,

$$\mathbf{v}= \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ v_m \end{bmatrix}_{(mx1)} \qquad \mathbf{v}^T = (v_1 \ v_2 \ . \ . \ . \ . \ v_m)_{(1\times m)}$$

$$\mathbf{X}= \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{bmatrix}_{(nx1)} \qquad \mathbf{E}(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ . \\ . \\ E(X_n) \end{bmatrix}_{(nx1)}$$

where X is a random vector and E(X) is its expected value. (again a vector) Then we can also define the variance which will turn out to be a matrix.

$$\mathbf{var}(\mathbf{X})_{(n\times n)} = \mathbf{E}((\mathbf{X}-\mathbf{E}(\mathbf{X}))_{(nx1)}.(\mathbf{X}-\mathbf{E}(\mathbf{X})^T)_{(1\times n)}$$

To see how how this matrix looks like we will analyse a simple case with n=2;

Example Let **X** be $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. Then the matrix will be

$$\mathbf{E}\left(\begin{pmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \end{pmatrix}.((X_1 - E(X_1)) \quad (X_2 - E(X_2)))\right) =$$

$$\mathbf{E}\begin{pmatrix} (X_1 - E(X_1))^2 & ((X_1 - E(X_1)) \quad (X_2 - E(X_2))) \\ ((X_1 - E(X_1)) \quad (X_2 - E(X_2))) & (X_2 - E(X_2))^2 \end{pmatrix} =$$

$$\begin{pmatrix} var(X_1) & cov(X_1, X_2) \\ cov(X_1, X_2) & var(X_2) \end{pmatrix} = \sum$$

where $\sum_{ii} = var(X_i)$ and $\sum_{ij} = cov(X_i, X_j)$. This $\sum$ which quite often used in econometrics is called the $\sum$ =**variance covariance matrix**.

The following is a numerical example:

Example Let **X** be $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. Let $var(X_1) = 2$, $var(X_2) = 1$, $cov(X_1, X_2) = 1$. Then **var(X)**=$\sum = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

Note also that if X_1 and X_2 are independent then $cov(X_1, X_2) = 0$ which amounts to saying that **var(X)** is a **diagonal matrix**; i.e.

$$\sum = \begin{pmatrix} var(X_1) & 0 \\ 0 & var(X_2) \end{pmatrix}$$

In general if $X_1, X_2......X_n$ are independent then

$$\sum = \begin{pmatrix} var(X_1) & 0 & 0 & 0 \\ 0 & var(X_2) & 0 & 0 \\ 0 & 0 & var(X_3) & 0 \\ 0 & 0 & 0 & var(X_4) \end{pmatrix}$$

One should be cautious, because even though indepence implies no covariance (no correlation) between two random variables, the reverse does not hold, i.e no covariance does not imply indepence. In other words, indepence is a stronger condition than covariance. This idea can be better understood given the following example;

Example Consider the following table$X\backslash Y$

X\Y	**-1**	**0**	**1**	**Mar. Prob.**
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
Mar. Prob.	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

This table should be read as follows;

$P(X = 0, Y = -1) = P(X = 0, Y = 1) = P(X = -1, Y = 0) = P(X = 1, Y = 0) = \frac{1}{4}$.

$P(X = -1, Y = -1) = P(X = 0, Y = 0) = P(X = 1, Y = -1) = P(X = 1, Y = 1) = 0$.

$P(X = 0) = \frac{1}{2}, P(X = -1) = P(X = 1) = \frac{1}{4}$.

$P(Y = 0) = \frac{1}{2}, P(Y = -1) = P(Y = 1) = \frac{1}{4}$.

Given this information we can see that X and Y are not independent since for instance

$$\begin{aligned} P(X &= -1, Y = -1) = 0 \\ P(X &= -1).P(Y = -1) = \frac{1}{4}.\frac{1}{4} = \frac{1}{16} \end{aligned}$$

On the other hand we can also calculate the expectation using the information given in the table:

$$\begin{aligned}
E(X) &= 0.\frac{1}{2}+(-1).\frac{1}{4}+1.\frac{1}{4}=0 \\
E(Y) &= 0.\frac{1}{2}+(-1).\frac{1}{4}+1.\frac{1}{4}=0 \\
E(X.Y) &= E(g(X,Y)) \\
g(X,Y) &= X.Y \\
E(X.Y) &= \sum_{x,y} x.y.p_{X,Y}(x,y)=0.
\end{aligned}$$

Proposition Given the random vector $X_{(n\times 1)}$ and the vector $a_{(n\times 1)}$ and the scalar b.

$$E(a^T.X+b)=a.E(X)+b$$

Proof Given the linearity of the expectation we have
$E(a_1.X_1+a_2.X_2+......a_n.X_n+b)=a_1.E(X_1)+a_2.E(X_2)+....+a_nE(X_n)+b = a^T.E(X)+b$

In general we have given the random vector $\mathrm{X}_{(n\times 1)}$, the matrix $A_{(m\times n)}$ and the vector $b_{(m\times 1)}$, we have

$$E(A.X+b)=A.E(X)+b$$

Proof Exercise.

Now we will see the counterpart of the previous result for var(Y)=$\sum$, where $Y=a^T.X$. Then we have

$$\begin{aligned}
var(Y) &= E((a^T.X-E(a^T.X)).(a^T.X-E(a^T.X))^T)= \\
&= E(a^T.(X-E(X)).(X-E(X))^T.a)= \\
&= a^T.E((X-E(X)).(X-E(X))^T).a= \\
&= a^T_{(1\times n)}.var(X)_{(n\times n)}.a_{(n\times 1)}. \text{ (Note it's a positive scalar.)}
\end{aligned}$$

We have the assumption that $E(Y)=E(a^T.X)=0$, then the above result simplifies to

$$var(Y)=var(a^T.X)=E((a^T.X).(a^T.X)^T)=E(a^T.X.X^T.a)=a^T.E(X.X^T).a$$

Proposition Let $\mathrm{X}_{(n\times 1)}$ a random vector, $A_{(m\times n)}$ a matrix and $b_{(m\times 1)}$ another vector. Then

$$var(Y)_{(m\times m)}=A_{(m\times n)}.var(X)_{(n\times n)}.A^T_{(n\times m)}$$

Proof Exercise.

Definition If $\sum$ is a variance covariance matrix, then for every vector a

$$\begin{aligned} a^T.\sum.a &\geq 0 \Rightarrow \sum \text{ is nonnegative definite.} \\ a^T.\sum.a &> 0 \Rightarrow \sum \text{ is positive definite.} \\ \sum \text{ is positive definite.} &\Rightarrow \text{ The inverse of } \sum \text{ exists.} \\ \text{The inverse of } \sum \text{ exists.} &\Leftrightarrow \det \sum \neq 0. \end{aligned}$$

recall that $\sum^{-1}\sum = \mathbf{I}$ (identity matrix.) $\sum -$ the variance covariance matrix is non-negative definite and symmetric matrix. (i.e. $\sum^T = \sum$).

Definition Let X and Y be random vectors. Then

$$cov(X,Y) = E((X - E(X)).(Y - E(Y))^T)$$

Example Assume we have $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$. Then

$$\begin{aligned} cov(X,Y) &= \mathbf{E}\left(\begin{pmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \end{pmatrix}_{(2\times1)} .((Y_1 - E(Y_1)) \quad (Y_2 - E(Y_2)) \quad (Y_2 - E(Y_2)))_{(1\times3)} \right) = \\ &= \begin{pmatrix} cov(X_1, Y_1) & cov(X_1, Y_2) & cov(X_1, Y_3) \\ cov(X_{2,Y_1}) & cov(X_2, Y_2) & cov(X_2, Y_3) \end{pmatrix}_{(2\times3)} \end{aligned}$$

Question What is the relationship between var(X_1),var(X_2) and cov(X_1,X_2) if we have two random vectors such as $X_{1(k\times1)}$ and $X_{2((n-k)\times1)}$ such that we have $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}_{(n\times1)}$.

Proposition In the above case we will have

$$var\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}_{(n\times n)} = \begin{pmatrix} \mathbf{var(X}_1)_{(k\times k)} & cov(X_1, X_2)_{(k\times(n-k))} \\ cov(X_1, X_2)^T_{((n-k)\times k)} & \mathbf{var(X}_1)_{((n-k)\times(n-k))} \end{pmatrix}$$

Proof Exercise.

Example Assume that we have $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$, then $\begin{pmatrix} X \\ Y \end{pmatrix}$ will be $\begin{pmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$. The following term will be its variance

$$var \begin{pmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \mathbf{var(X_1)} & \mathbf{cov(X_1, X_2)} & cov(X_1, Y_1) & cov(X_1, Y_2) & cov(X_1, Y_3) \\ \mathbf{cov(X_1, X_2)} & \mathbf{var(X_2)} & cov(X_2, Y_1) & cov(X_2, Y_2) & cov(X_2, Y_3) \\ cov(X_1, Y_1) & cov(X_2, Y_1) & \mathbf{var(Y_1)} & \mathbf{cov(Y_1, Y_2)} & \mathbf{cov(Y_1, Y_3)} \\ cov(X_1, Y_2) & cov(X_2, Y_2) & \mathbf{cov(Y_1, Y_2)} & \mathbf{var(Y_2)} & \mathbf{cov(Y_2, Y_3)} \\ cov(X_1, Y_3) & cov(X_2, Y_3) & \mathbf{cov(Y_1, Y_3)} & \mathbf{cov(Y_2, Y_3)} & \mathbf{var(Y_3)} \end{pmatrix}$$

to simplfy the notattion generally $\sum = \begin{pmatrix} \sum_{11} & \sum_{21} \\ \sum_{12} & \sum_{22} \end{pmatrix}$ is used.

Multivariate Normal Distribution

Recall that in the univariate case we have the normal distribution $N(\mu, \sigma^2)$ of a random variable X that has the following properties

$$\begin{aligned} E(X) &= \mu \\ var(X) &= \sigma^2 \\ f_X(x) &= \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}\frac{(x-\mu)}{\sigma^2}^2} \end{aligned}$$

Now we will analyze the analogous case for the random vector $X = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{pmatrix}$.

Recall that we have an absolutely continuous distribution. Since the density function of a random vector is the joint density of its components $X_1, X_2......X_n$. Then we will have

$$\begin{aligned} E(X) &= \mu_{(n\times 1)} \\ var(X) &= \sum_{(n\times n)} \\ f_{X_{(n\times 1)}}(\mathbf{x}_{(n\times 1)}) &= (2\pi)^{-\frac{n}{2}}.\det \Sigma^{-\frac{1}{2}} e^{-\frac{1}{2}\{(\mathbf{x}-\boldsymbol{\mu})^T.\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}} \end{aligned}$$

where $\sum$ is not singular$\Rightarrow$ $\det \sum \neq 0 \Rightarrow \exists \sum^{-1}$.

Example Suppose we have $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $E(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mu$, $var(X) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \sum$. Hence a random vector with normal distribution $N_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix})$
Then we can find $\det\Sigma$ and Σ^{-1}

$$\begin{aligned} det\Sigma &= 1 \\ \Sigma^{-1} &= \frac{1}{1}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \end{aligned}$$

we can always check $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} . \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \mathbf{I}$. Then plugging in these results into our formula

$$\begin{aligned} f(x_1, x_2) &= (2\pi)^{-\frac{2}{2}}.1^{-\frac{1}{2}}.\exp^{-\frac{1}{2}\{\begin{matrix} x_1 & x_2 \end{matrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\}} \\ f(x_1, x_2) &= \frac{1}{2\pi}.\exp^{-\frac{1}{2}\{x_1^2 - 2x_1x_2 + 2x_2^2\}} \end{aligned}$$

Since we have found the density function then we can calculate for instance the following probability by integrating;

$$P(0 < X_1 < 1, X_2 > 2) = \int_0^1 \left(\int_2^\infty \frac{1}{2\pi}.\exp^{-\frac{1}{2}\{x_1^2 - 2x_1x_2 + 2x_2^2\}} dx_2 \right) dx_1$$

$N_n(\mu, \sum)$ is the usual notation for multivariate normal distributions.

In the following part, we will analyze the analogous case for standard normal distribution:
Recall that in the univariate case we have $N(0, 1)$.

$$\begin{aligned} E(X) &= 0 \\ var(X) &= 1 \\ f_X(x) &= \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}x^2} \end{aligned}$$

The multivariate standard normal distribution will be denoted by $N_n(0, \mathbf{I}_n)$, where

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1) \end{pmatrix}_{(n \times n)}$$

note that $\det \mathbf{I}_n = 1$ and $\mathbf{I}_n^{-1} = \mathbf{I}_n$. Keeping this in mind we can write the density

$$\begin{aligned}
f_X(\mathbf{x}) &= (2\pi)^{-\frac{n}{2}} \exp^{-\frac{1}{2}\{\mathbf{x}^T \mathbf{I}_n \mathbf{x}\}} = (2\pi)^{-\frac{n}{2}} . \exp^{-\frac{1}{2}\{\mathbf{x}^T \mathbf{x}\}} = \\
&= (2\pi)^{-\frac{n}{2}} . \exp^{-\frac{1}{2}\Sigma_{j=1}^n x_j^2} = \\
\text{since } \mathbf{x}^T\mathbf{x} &= (\quad x_1 \; x_2 . \; . \; . \; x_n). \begin{pmatrix} x_1 \\ x_2 \\ . \\ x_n \end{pmatrix} = x_1^2 + x_2^2 + ... + x_n^2 \\
&= (2\pi)^{-\frac{n}{2}} . \Pi_{j=1}^n \exp^{-\frac{1}{2}x_j^2} = \\
&= \Pi_{j=1}^n \left(\frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}x_j^2} \right)
\end{aligned}$$

notice that each term of the product is the density of the univariate standard normal distribution, hence we have $X_1, X_2....X_n$ are i.i.d (independently and identically distributed) with $N(0,1)$.

$$f_{X_1,....X_n}(x_1....x_n) = \prod_{j=1}^{n} \varphi(x_j)$$

where φ is the density of the univariate standard normal distribution. So we have actually proved the following proposition:

Proposition The random vector $X = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{pmatrix}$ has $N_n(0, \mathbf{I}_n)$ distribution if and only if $X_1, X_2....X_n$ are i.i.d.