Lecture - 22

February 27, 2012

Multivariate Normal

Consider the random vector $Z = \big(Z_1, \cdots, Z_n\big)^{'}$

Definition 0.1. *$\underline{Z}$ is said to follow multivariate normal if any linear combination of $\underline{Z}$ follow normal distribution.*

We now try to develop another definition in the terms of moment generating function as it can be used to prove different properties of multivariate normal distribution.

Suppose, components of random vector $\underline{Z}$ follow $N(0,1)$. Then the density of $\underline{Z}$ is

$$\begin{aligned} f_Z(z) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2} \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\Sigma_{i=1}^{n} z_i^2\right] \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\underline{z}^{'}\underline{z}\right] \end{aligned}$$

for $\underline{Z} \in R^n$. Because the $Z_i^{\prime}s$ have mean 0 and variance 1, and are uncorrelated, the mean and covariance matrix of $\underline{Z}$ are

$$E(\underline{Z}) = 0 \quad \text{and} \quad cov(\underline{Z}) = I_n$$

. Where I_n denotes the identity matrix of order n.

Recall that m.g.f. of Z_i is $\exp\left(\frac{t_i^2}{2}\right)$. Hence, the m.g.f. $\underline{Z}$ is

$$\begin{aligned} M_{\underline{Z}}(t) &= E\Bigg(\exp\left(t^{'}\underline{Z}\right)\Bigg) \\ &= E\Bigg(\prod_{i=1}^{n}\exp\left(t_i Z_i\right)\Bigg) \\ &= \prod_{i=1}^{n} E\Bigg(\exp\left(t_i Z_i\right)\Bigg) \\ &= \exp\left[\frac{1}{2}\Sigma_{i=1}^{n} t_i^2\right] \\ &= \exp\left[\frac{1}{2}t_i^{'}t\right] \end{aligned}$$

for all $t \in R^n$. We say that $\underline{Z}$ has a multivariate normal distribution with mean vector 0 and covariance matrix I_n. we abbreviate this by saying that Z has an $N_n(0, I_n)$ distribution.

For general case, suppose Σ is an $n \times n$ symmetric and positive semi-definite matrix (p.s.d.). Then from linear algebra, we can always decompose Σ as $\Sigma = \Gamma^{'} \wedge \Gamma$. Where $\wedge$ is the diagonal matrix i.e. $\wedge = diag(\lambda_1, \lambda_2. \cdots, \lambda_n)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are the eigen values of Σ, and the columns of $\Gamma^{'}$, $v_1, v_2, \cdots, v_n$ are the corresponding eigenvectors. This decomposition of Σ. The matrix Γ is orthogonal, i.e. $\Gamma^{-1} = \Gamma^{'}$, and hence, $\Gamma\Gamma^{'} = I$. We can write the decomposition is another way i.e.

$$\Sigma = \Gamma^{'} \wedge \Gamma = \Sigma_{i=1}^{n}\lambda_i \underline{v_i v_i^{'}}$$

As $\lambda_i^{'}s$ are non-negative, we can define the diagonal matrix $\wedge^{1/2} = diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n})$.

Then the orthogonality of Γ implies

$$\Sigma = \Gamma^{'} \wedge^{1/2} \Gamma\Gamma^{'} \wedge^{1/2} \Gamma$$

Define the square root of the p.s.d. matrix Σ as

$$\Sigma^{1/2} = \Gamma^{'} \wedge^{1/2} \Gamma$$

. Note that $\Sigma^{1/2}$ is symmetric and p.s.d. Suppose, Σ is p.d. i.e. all of its eigen values are strictly positive. Then it is easy to show that

$$\left(\Sigma^{1/2}\right)^{-1} = \Sigma^{-1/2} = \Gamma' \wedge^{-1/2} \Gamma$$

Now we are going to use the above decomposition. Let Z have a $N_n(0, I_n)$ distribution. Σ be a p.s.d., symmetric matrix and let μ be an $n \times 1$ vector of constants. Define, the random vector X by

$$\underline{X} = \Sigma^{1/2}\underline{Z} + \underline{\mu}$$

We immediately have

$$E(\underline{X}) = \underline{\mu} \quad \text{and} \quad Cov(\underline{X}) = \Sigma^{1/2}\Sigma^{1/2} = \Sigma$$

Further m.g.f. of $\underline{X}$ is given by

$$\begin{aligned} M_{\underline{X}}(t) &= E\Bigg(\exp\left(\underline{t}'\underline{X}\right)\Bigg) \qquad (1) \\ &= E\Bigg(\exp\left(\underline{t}'\Sigma^{1/2}\underline{Z} + \underline{\mu}\right)\Bigg) \\ &= e^{\underline{\mu}}E\Bigg(\exp\left(\underline{t}'\underline{Z}\right)\Bigg) \\ &= e^{\underline{\mu}+\frac{1}{2}\underline{t}'\Sigma\underline{t}} \end{aligned}$$

This leads to be following definition :

Definition 0.2. *We say an n-dimensional random vector X has multivariate normal if its m.g.f. is*

$$M_{\underline{X}}(\underline{t}) = e^{\underline{t}'\mu+\frac{1}{2}\underline{t}'\Sigma\underline{t}}$$

for all $\underline{t} \in R^n$ and Σ is a symmetric positive semi-definite matrix and $\underline{\mu} \in R^n$. We abbreviate this by saying that X has a $N_n(\mu, \Sigma)$ distribution.

Note that our definition is for positive semi-definite matrix. usually Σ is positive definite, in which case, we can further obtain the density of X. If Σ is p.d. so is $\Sigma^{1/2}$ and as we density its inverse can be given by

$$\Sigma^{-1/2} = \Gamma^{'} \wedge^{-1/2} \Gamma$$

Thus the transformation from X and Z, is one-to-one with the inverse transformation

$$Z = \Sigma^{-1/2}(X - \mu)$$

with jacobian $|\Sigma|^{-1/2} = |\Sigma|^{-1/2}$, Hence, the p.d.f. of X is given by

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^{'}\Sigma^{-1}(\underline{x}-\underline{\mu})} \;;\quad \underline{x} \in R^n \;,\; \underline{\mu} \in R^n$$

Property 1 :

Suppose X has a $N_n(\mu, \Sigma)$ distribution, Let $Y = AX + b$, where A is a $m \times n$ matrix and $b \in R^n$. Then Y has a $N_m(A\mu + b, A\Sigma A^{'})$ distribution.

Proof. For $t \in R^n$, the m.g.f. of Y

$$\begin{aligned}
M_Y(\underline{t}) &= E\big(e^{\underline{t}^{'}\underline{Y}}\big) \\
&= E\big(e^{\underline{t}^{'}(A\underline{X}+\underline{b})}\big) \\
&= e^{\underline{t}^{'}\underline{b}} E\big(e^{\underline{t^*}^{'}\underline{X}}\big) \quad [t^* = A^{'}\underline{t}] \\
&= e^{\underline{t}^{'}\underline{b}} e^{\underline{t}^{'}\underline{\mu} + \underline{t^*}^{'}\Sigma\underline{t^*}} \\
&= e^{\underline{t}^{'}(A\underline{\mu}+\underline{b}) + \underline{t}^{'}(A\Sigma A)\underline{t}}
\end{aligned}$$

which m.g.f. of an $N_m(A\mu + b, A\Sigma A^{'})$ distribution. ∎

Property 2 :

Suppose, X has a $N_n(\mu, \Sigma)$ distribution; partitioned as follows :

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Then X_1 has a $N_m(\mu_1, \Sigma_{11})$ distribution.

Proof. $\quad X_1^{m\times 1} = A^{m\times n}\underline{X}^{n\times 1}$ where $A = [I_m, \vdots, O_{m\times\overline{n-m}}]$

Therefore, applying the previous property the result follows. ∎

Remark :

This is a useful result, because it says that any marginal distribution of $\underline{X}$ is also normal and further its mean and covariance matrix are those associated with that partial vector.

Property 3 :

Suppose, X has a $N_n(\mu, \Sigma)$ distribution; partitioned as in the expressions like that of property-2, Then X_1 and X_2 are independent if and only if $\Sigma_{12} = 0$.

Proof. First note that $\Sigma_{21} = \Sigma_{12}'$. The joint m.g.f. of $\underline{X_1}$ and $\underline{X_2}$ is given by ,

$$M_{X_1,X_2} = \exp\left\{\underline{t_1}'\underline{\mu_1} + \underline{t_2}'\underline{\mu_2} + \frac{1}{2}(\underline{t_1}'\Sigma_{11}\underline{t_1} + \underline{t_1}'\Sigma_{22}\underline{t_2} + \underline{t_2}'\Sigma_{21}\underline{t_1} + \underline{t_1}'\Sigma_{12}\underline{t_2})\right\} \tag{2}$$

where,$\underline{t}' = (\underline{t_1}', \underline{t_2}')$ is partitioned the same as $\underline{\mu}$

Therefore by property 2, X_1 has a $N_m(\mu_1, \Sigma_{11})$ distribution and X_2 has a $N_m(\mu_2, \Sigma_{22})$ distribution. Hence the product of their marginal mgf's is

$$M_{X_1}(t_1)M_{X_2}(t_2) = \exp\left[\underline{t_1}'\underline{\mu_1} + \underline{t_2}'\underline{\mu_2} + \frac{1}{2}(\underline{t_1}'\Sigma_{11}\underline{t_1} + \underline{t_1}'\Sigma_{22}\underline{t_2})\right] \tag{3}$$

clearly, X_1 and X_2 are independent if and only if (2) and (3) are same. If $\Sigma_{12} = 0$ and hence $\Sigma_{21} = 0$, then the expression are the same and X_1 and X_2 are independent. If X_1 and X_2 are independent, then the covariance between their components are all 0; i.e. $\Sigma_{12} = 0$ and $\Sigma_{21} = 0$. ∎

Property 4 :

Suppose, X has a $N_n(\mu, \Sigma)$ distribution;which is partitioned as in the expressions described earlier. Assume that Σ is positive definite. Then the conditional distribution of $\underline{X_1}|\underline{X_2}$ is $N_m\big(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big)$.

Proof. Consider first the joint distribution of the random vector $\underline{W} = \underline{X_1} - \Sigma_{12}\Sigma_{22}^{-1}\underline{X_2}$ and $\underline{X_2}$. This distribution is obtained from the transformation

$$\begin{bmatrix} \underline{W} \\ \underline{X_2} \end{bmatrix} = \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_p \end{bmatrix} \begin{bmatrix} \underline{X_1} \\ \underline{X_2} \end{bmatrix}$$

As this is a linear transformation, it follows from property 1 that joint distribution is multivariate normal with

$$E(\underline{W}) = \underline{\mu_1} - \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu_2} \quad , \quad E(\underline{X_2}) = \underline{\mu_2}$$

and covariance matrix

$$\begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_p \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_m & 0 \\ -\Sigma_{12}\Sigma_{22}^{-1} & I_p \end{bmatrix}$$

$=$

$$\begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$$

Hence, the random vector $\underline{W}$ and $\underline{X_2}$ are independent. Thus the conditional distribution of $\underline{W}|\underline{X_2}$ is same as the marginal distribution of $\underline{W}$, i.e.
$\underline{W}|\underline{X_2}$ is $N_m\big(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big)$. Further, because of this independence, $\underline{W} - \Sigma_{12}\Sigma_{22}^{-1}\underline{X_2}$ given $\underline{X_2} = \underline{x_2}$ is distributed as
$N_m\big(\mu_1 - \mu_2\Sigma_{12}\Sigma_{22}^{-1} + \Sigma_{12}\Sigma_{22}^{-1}x_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big)$. ∎