University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Multivariate normal distribution and distribution theory in multiple regression

We say that a random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$ follows the multivariate normal distribution if its probability density function is given by

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})}, \quad (1)$$

and we write, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Moment generating function
If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ using the transformation $\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$ we find that $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$.

Theorem 1
Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and let $\mathbf{A}$ be an $m \times n$ matrix of rank m and $\mathbf{c}$ be an $m \times 1$ vector. Then $\mathbf{AY} + \mathbf{c} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Apply theorem 1 in multiple regression:

Theorem 2
Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Sub-vectors of $\mathbf{Y}$ follow the multivariate normal distribution and linear combinations of $Y_1, Y_2, \ldots, Y_n$ follow the univariate normal distribution. For example, suppose $\mathbf{Y}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are partitioned as follows $\mathbf{Y} = \begin{pmatrix} \mathbf{Q_1} \\ \mathbf{Q_2} \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, where $\mathbf{Q_1}$ is $p \times 1$. It follows that $\mathbf{Q_1} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{Q_2} \sim N_{n-p}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. For a linear combination of $Y_1, Y_2, \ldots, Y_n$, i.e. $a_1Y_1 + a_2Y_2 + \ldots + a_nY_n = \mathbf{a}'\mathbf{Y}$, it follows that, $\mathbf{a}'\mathbf{Y} \sim N(\mathbf{a}'\boldsymbol{\mu}, \sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}})$.

Example

Let $\mathbf{Y} = \left(\begin{array}{c} Y_1 \\ Y_2 \\ \hline Y_3 \\ Y_4 \\ Y_5 \end{array}\right)$, $\boldsymbol{\mu} = \left(\begin{array}{c} \mu_1 \\ \mu_2 \\ \hline \mu_3 \\ \mu_4 \\ \mu_5 \end{array}\right)$, $\boldsymbol{\Sigma} = \left(\begin{array}{cc|ccc} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \hline \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 \end{array}\right)$, then if $\mathbf{Q_1} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$,

it follows that $\mathbf{Q_1} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right]$.

Apply theorem 2 in multiple regression:

Theorem 3 - Statistical independence
Suppose $\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ are partitioned as in theorem 2. We say that $\mathbf{Q_1}, \mathbf{Q_2}$ are statistically independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Application:
Suppose $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and define the following two vectors $\mathbf{Q_1} = \mathbf{AY}$ and $\mathbf{Q_2} = \mathbf{BY}$. Then, $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are independent if $cov(\mathbf{Q_1}, \mathbf{Q_2}) = \mathbf{A\Sigma B'} = \mathbf{0}$. We stack the two vectors as follows: $\mathbf{Q} = \begin{pmatrix} \mathbf{Q_1} \\ \mathbf{Q_2} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{Y} = \mathbf{LY}$. Therefore using theorem 1 we find that $\mathbf{Q} \sim N(\mathbf{L}\boldsymbol{\mu}, \mathbf{L\Sigma L'})$ or $\mathbf{Q} \sim N\left[\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \boldsymbol{\mu}, \begin{pmatrix} \mathbf{A\Sigma A'} & \mathbf{A\Sigma B'} \\ \mathbf{B\Sigma A'} & \mathbf{B\Sigma B'} \end{pmatrix}\right]$, and we conclude that $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are independent if and only if $\mathbf{A\Sigma B'} = \mathbf{0}$. Here, we can just simply find the covariance between the vectors $\mathbf{Q_1}$ and $\mathbf{Q_2}$ and if it is $\mathbf{0}$ then we conclude that $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are independent. In general $\text{cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A\Sigma B'}$. Why?

Apply theorem 3 in multiple regression:

Conditional distributions from multivariate normal:

Consider the bivariate normal distribution (see page 1). From theorem 1 it follows that $Y_1 \sim N(\mu_1, \sigma_1)$. This is also called the marginal probability distribution of Y_1. We want to find the conditional distribution of Y_2 given Y_1.

From the conditional probability law, $f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1 Y_2}(y_1, y_2)}{f_{Y_1}(y_1)}$, and after substituting the bivariate density and the marginal density it can be shown that the conditional probability density function of Y_2 given Y_1 is given by

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{1}{\sqrt{\sigma_2^2(1-\rho)^2}\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{Y_2 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)}{\sigma_2^2(1-\rho^2)}\right)\right].$$

We recognize that this is a normal probability density function with mean $\mu_{Y_2|Y_1} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)$ and variance $\sigma^2_{Y_2|Y_1} = \sigma_2^2(1-\rho^2)$.

In general:

Suppose that $\mathbf{Y}$, $\boldsymbol{\mu}$, and $\mathbf{\Sigma}$ are partitioned as follows $\mathbf{Y} = \begin{pmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$, and $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$. It can be shown that the conditional distribution of $\mathbf{Y_1}$ given $\mathbf{Y_2}$ is also multivariate normal, $\mathbf{Y_1}|\mathbf{Y_2} \sim N(\boldsymbol{\mu}_{1|2}, \mathbf{\Sigma}_{1|2})$, where $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{Y_2} - \boldsymbol{\mu}_2)$, and $\mathbf{\Sigma}_{1|2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.