University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Multivariate normal distribution

One of the most important distributions in statistical inference is the multivariate normal distribution. The probability density function of the multivariate normal distribution, its moment generating function, and its properties are discussed here.

Probability density function

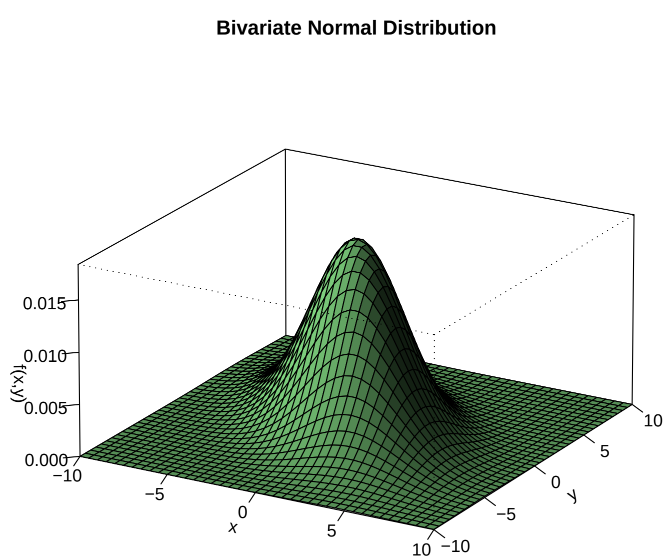
We say that a random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$ follows the multivariate normal distribution if its probability density function is given by

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

and we write, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\mathbf{Y} = (Y_1, Y_2)$ then we have a bivariate normal distribution and its probability density function can be expressed as

$$\begin{aligned} f(y_1, y_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \; exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right]\right] \end{aligned}$$

Here, we have $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$. The previous expression can be obtained by finding the inverse of $\boldsymbol{\Sigma}$ and substituting it into (1). Here is the bivariate normal pdf.



Moment generating function

A useful tool in statistical theory is the moment generating function. The joint moment generating function is defined as $M_{\mathbf{Y}}(\mathbf{t}) = Ee^{\mathbf{t}'\mathbf{Y}} = Ee^{\sum_{i=1}^{n} y_i t_i},$, where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ and $\mathbf{t} = (t_1, t_2, \ldots, t_n)'$. We will find the joint moment generating function of $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- Suppose $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$. Since $Z_1, Z_2, \ldots, Z_n$ are independent the joint moment generating function of $\mathbf{Z}$ is $M_{\mathbf{Z}}(\mathbf{t}) = e^{\frac{1}{2}\mathbf{t}'\mathbf{t}}$.

- If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ show that $\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$ follows $N(\mathbf{0}, \mathbf{I})$.

- Aside note: What is $\boldsymbol{\Sigma}^{-\frac{1}{2}}$?

- To find the joint moment generating function of $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ we use the transformation $\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$ to get $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$.

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 $\mathbf{AY} + \mathbf{c} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Proof

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.

Proof

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 $\mathbf{Y_1}$ is $p \times 1$. We will show that $\mathbf{Q_1} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{Q_2} \sim N_{n-p}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. The result follows directly by using the previous theorem with $\mathbf{A} = (\mathbf{I_p}, \mathbf{0})$. For a linear combination of $Y_1, Y_2, \ldots, Y_n$, i.e. $a_1Y_1 + a_2Y_2 + \ldots + a_nY_n = \mathbf{a}'\mathbf{Y}$, the matrix $\mathbf{A}$ of theorem 1 is a vector and therefore, $\mathbf{a}'\mathbf{Y} \sim N(\mathbf{a}'\boldsymbol{\mu}, \sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}})$.

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

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 $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. We can show this using the joint moment generating function of $\mathbf{Y}$. Recall that the exponent of the joint moment generating function of the multivariate normal distribution is $\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}$ which after partitioning $\mathbf{t}$ conformably (according to the partitioning of $\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$) can be expressed as $\mathbf{t_1}'\boldsymbol{\mu}_1 + \mathbf{t_2}'\boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t_1}'\boldsymbol{\Sigma}_{11}\mathbf{t_1} + \frac{1}{2}\mathbf{t_2}'\boldsymbol{\Sigma}_{22}\mathbf{t_2} + \mathbf{t_1}'\boldsymbol{\Sigma}_{12}\mathbf{t_2}$. When $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, the joint moment generating function can be expressed as the product of the two marginal moment generating functions of $\mathbf{Q_1}$ and $\mathbf{Q_2}$, i.e. $M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{Q_1}}(\mathbf{t_1})M_{\mathbf{Q_2}}(\mathbf{t_2})$, therefore, $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are independent.

Theorem 3
Using theorem 1 and the statement about statistical independence above, we prove the following theorem. Suppose $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and define the following two vectors $\mathbf{W_1} = \mathbf{AY}$ and $\mathbf{W_2} = \mathbf{BY}$. Then, $\mathbf{W_1}$ and $\mathbf{W_2}$ are independent if $cov(\mathbf{W_1}, \mathbf{W_2}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$.

Proof
We stack the two vectors as follows: $\mathbf{W} = \begin{pmatrix} \mathbf{W_1} \\ \mathbf{W_2} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{Y} = \mathbf{LY}$. Therefore using theorem 1 we find that $\mathbf{W} \sim N(\mathbf{L}\boldsymbol{\mu}, \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}')$ or

$\mathbf{W} \sim N\left[\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}\boldsymbol{\mu}, \begin{pmatrix} \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' & \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' \\ \mathbf{B}\boldsymbol{\Sigma}\mathbf{A}' & \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}' \end{pmatrix}\right]$, and we conclude that $\mathbf{W_1}$ and $\mathbf{W_2}$ are independent if and only if $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$.

Conditional probability density functions for 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, $f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1Y_2}(y_1,y_2)}{f_{Y_1}(y_1)}$, 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 $\boldsymbol{\Sigma}$ are partitioned as follows $\mathbf{Y} = \begin{pmatrix} \mathbf{Q_1} \\ \mathbf{Q_2} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, and $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. It can be shown that the conditional distribution of $\mathbf{Q_1}$ given $\mathbf{Q_2}$ is also multivariate normal, $\mathbf{Q_1}|\mathbf{Q_2} \sim N(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$, where $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Q_2} - \boldsymbol{\mu}_2)$, and $\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Proof:
Let $\mathbf{U} = \mathbf{Q_1} - \mathbf{\Sigma_{12}\Sigma_{22}^{-1}Q_2}$ and $\mathbf{V} = \mathbf{Q_2}$.

Example 1

Suppose the prices (in \$), Y_1, Y_2, Y_3, Y_4 of objects A, B, C, and D are jointly normally distributed as $\mathbf{Y} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where, $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 3 \\ 6 \\ 4 \end{pmatrix}$, and $\boldsymbol{\Sigma} = \begin{pmatrix} 3 & 2 & 3 & 3 \\ 2 & 5 & 5 & 4 \\ 3 & 5 & 9 & 5 \\ 3 & 4 & 5 & 6 \end{pmatrix}$.

Answer the following questions:

a. Suppose a person wants to buy three of product A, four of product B, and one of product C. Find the probability that the person will spend more than \$30.

b. Find the moment generating function of Y_1.

c. Find the joint moment generating function of (Y_1, Y_3).

d. Find the correlation coefficient between Y_3 and Y_4.

Example 2

Suppose $\mathbf{Y} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$, and $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Find the joint distribution of $Q_1 = Y_1 + Y_2 + Y_3$ and $Q_2 = Y_1 - Y_2$.

Example 3

Answer the following questions:

a. Let $\mathbf{X} \sim N_n(\mu \mathbf{1}, \boldsymbol{\Sigma})$, where $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, and $\boldsymbol{\Sigma}$ is the variance covariance matrix of $\mathbf{X}$. Let

$\boldsymbol{\Sigma} = (1-\rho)\mathbf{I} + \rho \mathbf{J}$, with $\rho > -\frac{1}{n-1}$, $\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $\mathbf{J} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix}$.

Therefore, when $\rho = 0$ we have $\mathbf{X} \sim N_n(\mu \mathbf{1}, \mathbf{I})$, and in this case we showed in class that $\bar{X}$ and $\sum_{i=1}^{n} (X_i - \bar{X})^2$ are independent. Are they independent when $\rho \neq 0$?

b. Suppose $\boldsymbol{\epsilon} \sim N_3(\mathbf{0}, \sigma^2 \mathbf{I_3})$ and that $Y_0 \sim N(0, \sigma^2)$, independently of the ϵ_i's. Therefore the vector $\begin{pmatrix} Y_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$, is multivariate normal. Define $Y_i = \rho Y_{i-1} + \epsilon_i$ for $i = 1, 2, 3$. Express Y_1, Y_2, Y_3 in terms of ρ, Y_0, and the ϵ_i's.

c. Refer to part (b). Find the covariance matrix of $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$,

d. What is the distribution of $\mathbf{Y}$?