Ch3. Multivariate Normal Distribution

1 Multivariate normal distributions

Definition 1.1 *A random vector $\boldsymbol{x} = (x_1, \cdots, x_p)'$ is said to have a p-variate normal distribution if its probability density function can be written as*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}, \tag{1}$$

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$ and $\Sigma > 0$. We will denote a p-variate normal by $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$.

Theorem 1.1 *If $\boldsymbol{y}$ is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, its moment-generating function is given by*

$$M_y(\boldsymbol{t}) = E(e^{\boldsymbol{t}'\boldsymbol{y}}) = e^{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}}. \tag{2}$$

Two important properties of moment-generating functions:

1. If two random vectors have the same moment-generating function, if and only if they have the same density. This property is called the uniqueness of the moment-generating function.

2. Two random vectors are independent if and only if their joint moment-generating function factors into the product of their two separate moment-generating functions; that is, if $\boldsymbol{y}' = (\boldsymbol{y}_1', \boldsymbol{y}_2')$ and $\boldsymbol{t}' = (\boldsymbol{t}_1', \boldsymbol{t}_2')$. then $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are independent if and only if

$$M_y(\boldsymbol{t}) = M_{\boldsymbol{y}_1}(\boldsymbol{t}_1) M_{\boldsymbol{y}_2}(\boldsymbol{t}_2).$$

2 Properties of the Multivariate normal distribution

Theorem 2.1 *Let the $p \times 1$ random vector $\boldsymbol{y}$ be $N_p(\mu, \Sigma)$, let $\boldsymbol{a}$ be any $p \times 1$ vector of constants, and let $\boldsymbol{A}$ be any $k \times p$ matrix of constants with rank $k \leq p$. Then*

(i) $z = \boldsymbol{a}'\boldsymbol{y}$ *is* $N(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a})$.

(ii) $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{y}$ *is* $N_k(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}')$.

PROOF: (i) The moment-generating function for $z = \boldsymbol{a}'\boldsymbol{y}$ is given by

$$\begin{aligned} M_z(\boldsymbol{t}) &= E(e^{t\boldsymbol{a}'\boldsymbol{y}}) = E(e^{(t\boldsymbol{a})'\boldsymbol{y}}) \\ &= e^{(t\boldsymbol{a})'\boldsymbol{\mu}+(t\boldsymbol{a})'\Sigma(t\boldsymbol{a})/2} \qquad by(2) \\ &= e^{(\boldsymbol{a}'\boldsymbol{\mu})t+(\boldsymbol{a}'\Sigma\boldsymbol{a})t^2/2}. \end{aligned}$$

Comparing with the moment-generating function of univariate normal, it is clear that $z = \boldsymbol{a}'\boldsymbol{y}$ is univariate normal with mean $\boldsymbol{a}'\boldsymbol{\mu}$ and variance $\boldsymbol{a}'\Sigma\boldsymbol{a}$.

(ii) The moment-generating function of $\boldsymbol{Ay}$ is

$$M_{\boldsymbol{z}}(\boldsymbol{t}) = e^{\boldsymbol{t}'(\boldsymbol{A\mu})+\boldsymbol{t}'(\boldsymbol{A}\Sigma\boldsymbol{A}')\boldsymbol{t}/2}.$$

Sine $\boldsymbol{A}$ is a row-full-rank matrix, $\boldsymbol{A}\Sigma\boldsymbol{A}'$ is positive definite. Hence, $\boldsymbol{Ay}$ is distributed as the k-variate normal $N_k(\boldsymbol{A\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}')$. □

Theorem 2.2 *If $\boldsymbol{y}$ is $N_p(\boldsymbol{\mu}, \Sigma)$, then any $r \times 1$ subvector of $\boldsymbol{y}$ has an r-variate normal distribution with the same means, variance, and covariances as in the original p-variate normal distribution.*

PROOF: Without loss of generality, let $\boldsymbol{y}$ be partitioned as $\boldsymbol{y}' = (\boldsymbol{y}_1', \boldsymbol{y}_2')$, where $\boldsymbol{y}_1$ is the $r \times 1$ subvector of interest. Let $\boldsymbol{\mu}$ and Σ be partitioned accordingly:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Define $\boldsymbol{A} = (\boldsymbol{I}_r, \boldsymbol{O})$, then $\boldsymbol{Ay} = \boldsymbol{y}_1$. Hence, $\boldsymbol{y}_1$ is distributed as $N_r(\boldsymbol{\mu}_1, \Sigma_{11})$. □

Corollary 2.1 *If $\boldsymbol{y}$ is $N_p(\boldsymbol{\mu}, \Sigma)$, then any individual variable y_i in $\boldsymbol{y}$ is distributed as $N(\mu_i, \sigma_{ii}^2)$.*

Theorem 2.3 *If* $\boldsymbol{v} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix}$ *is* $N_p(\boldsymbol{\mu}, \Sigma)$*, then* $\boldsymbol{y}$ *and* $\boldsymbol{x}$ *are independent if and only if* $\Sigma_{\boldsymbol{yx}} = \boldsymbol{O}$.

PROOF: (i) If $\boldsymbol{y}$ and $\boldsymbol{x}$ are independent, $\Sigma_{\boldsymbol{yx}} = \boldsymbol{O}$.

(ii) If $\Sigma_{yx} = \boldsymbol{O}$, then

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \boldsymbol{O} \\ \boldsymbol{O} & \Sigma_{xx} \end{pmatrix}.$$

The exponent of the moment-generating function becomes

$$\begin{aligned} \boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t} &= \boldsymbol{t}'_y\boldsymbol{\mu}_y + \boldsymbol{t}'_x\boldsymbol{\mu}_x + \frac{1}{2}\boldsymbol{t}'_y\Sigma_{yy}\boldsymbol{t}_y + \frac{1}{2}\boldsymbol{t}'_x\Sigma_{xx}\boldsymbol{t}_x \\ &= M_y(\boldsymbol{t})M_x(\boldsymbol{t}). \end{aligned}$$

Hence, $\boldsymbol{y}$ and $\boldsymbol{x}$ are independent if $\Sigma_{\boldsymbol{yx}} = \boldsymbol{O}$. □

Corollary 2.2 *If* $\boldsymbol{y}$ *is* $N_p(\boldsymbol{\mu}, \Sigma)$*, then any two individual variables* y_i *and* y_j *are independent if* $\sigma_{ij} = 0$.

Corollary 2.3 *If $\boldsymbol{y}$ is $N_p(\boldsymbol{\mu}, \Sigma)$ and if $cov(\boldsymbol{Ay}, \boldsymbol{By}) = \boldsymbol{A}\Sigma\boldsymbol{B} = \boldsymbol{O}$, then $\boldsymbol{Ay}$ and $\boldsymbol{By}$ are independent.*

Theorem 2.4 *If $\boldsymbol{y}$ and $\boldsymbol{x}$ are jointly multivariate normal with $\Sigma_{yx} \neq \boldsymbol{O}$, then the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$, $f(\boldsymbol{y}|\boldsymbol{x})$, is multivariate normal with mean vector and covariance matrix,*

$$E(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x),$$

$$cov(\boldsymbol{y}|\boldsymbol{x}) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

PROOF: See Rencher and Schaalje (2008, pp.95). □

Corollary 2.4 *If* $\boldsymbol{v} = (y, \boldsymbol{x}')'$ *with*

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \mu_{\boldsymbol{x}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \Sigma_{xx} \end{pmatrix},$$

then $y|\boldsymbol{x}$ *is normal with*

$$E(y|bx) = \mu_y + \boldsymbol{\sigma}'_{yx}\Sigma_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x),$$
$$var(y|\boldsymbol{x}) = \sigma_y^2 - \boldsymbol{\sigma}'_{yx}\Sigma_{xx}^{-1}\boldsymbol{\sigma}_{yx}.$$

Since Σ_{xx}^{-1} is positive definite, the corollary implies that

$$var(y|\boldsymbol{x}) \le var(y).$$

3 Partial Correlation

Let $\boldsymbol{v}$ be $N_{p+q}(\boldsymbol{\mu}, \Sigma)$ and let $\boldsymbol{v}$, $\boldsymbol{\mu}$ and Σ be partitioned as follows.

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}.$$

The covariance of y_i and y_j in the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ will be denoted by $\sigma_{ij \cdot rs \ldots q}$, where y_i and y_j are two of the variables in $\boldsymbol{y}$ and $y_r, y_s, \ldots, y_q$ are all variables in $\boldsymbol{x}$. Thus $\sigma_{ij \cdot rs \ldots q}$ is the (ij)th element of $cov(\boldsymbol{y}|\boldsymbol{x}) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$. The **partial correlation coefficient** $\rho_{ij \cdot rs \ldots q}$ is defined to be the correlation between y_i and y_j in the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$, that is,

$$\rho_{ij \cdot rs \ldots q} = \frac{\sigma_{ij \cdot rs \ldots q}}{\sqrt{\sigma_{ii \cdot rs \ldots q}\sigma_{jj \cdot rs \ldots q}}}.$$

The matrix form is $\boldsymbol{P}_{y \cdot x} = (\rho_{ij \cdot rs \ldots q})$, where

$$\boldsymbol{P}_{y \cdot x} = \boldsymbol{D}_{y \cdot x}^{-1}\Sigma_{y \cdot x}\boldsymbol{D}_{y \cdot x}^{-1},$$

where $\Sigma_{y \cdot x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$ and $\boldsymbol{D}_{y \cdot x} = [diag(\Sigma_{y \cdot x})]^{1/2}$.

Unless $\boldsymbol{y}$ and $\boldsymbol{x}$ are independent ($\Sigma_{yx} = \boldsymbol{O}$), the partial correlation $\rho_{ij \cdot rs \ldots q}$ is different from the usual correlation $\rho_{ij} = \sigma_{ij} \big/ \sqrt{\sigma_{ii}\sigma_{jj}}$.