1 Multivariate normal distributions

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

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

(1)

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

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

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

(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 \( y' = (y'_1, y'_2) \) and \( t' = (t'_1, t'_2) \). then \( y_1 \) and \( y_2 \) are independent if and only if

\[
M_y(t) = M_{y_1}(t_1) M_{y_2}(t_2).
\]
2 Properties of the Multivariate normal distribution

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

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

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

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

\[
M_z(t) = E(e^{t\mathbf{a}'\mathbf{y}}) = E(e^{(t\mathbf{a})'\mathbf{y}}) \\
= e^{(t\mathbf{a})'\mu + (t\mathbf{a})'\Sigma(t\mathbf{a})/2} \\
= e^{(\mathbf{a}'\mu)t + (\mathbf{a}'\Sigma\mathbf{a})t^2/2}.
\]

Comparing with the moment-generating function of univariate normal, it is clear that \( z = \mathbf{a}'\mathbf{y} \) is univariate normal with mean \( \mathbf{a}'\mu \) and variance \( \mathbf{a}'\Sigma\mathbf{a} \).
(ii) The moment-generating function of \( Ay \) is
\[
M_Z(t) = e^{t' (A\mu) + t' (A\Sigma A') t / 2}.
\]

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

**Theorem 2.2** If \( y \) is \( N_p(\mu, \Sigma) \), then any \( r \times 1 \) subvector of \( y \) 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 \( y \) be partitioned as \( y' = (y'_1, y'_2) \), where \( y'_1 \) is the \( r \times 1 \) subvector of interest. Let \( \mu \) and \( \Sigma \) be partitioned accordingly:
\[
y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.
\]
Define \( A = (I_r, O) \), then \( Ay = y_1 \). Hence, \( y_1 \) is distributed as \( N_r(\mu_1, \Sigma_{11}) \). □

**Corollary 2.1** If \( y \) is \( N_p(\mu, \Sigma) \), then any individual variable \( y_i \) in \( y \) is distributed as \( N(\mu_i, \sigma_{ii}^2) \).
Theorem 2.3  If \( v = \begin{pmatrix} y \\ x \end{pmatrix} \) is \( \mathcal{N}_p(\mu, \Sigma) \), then \( y \) and \( x \) are independent if and only if \( \Sigma_{yx} = O \).

**Proof:** (i) If \( y \) and \( x \) are independent, \( \Sigma_{yx} = O \).

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

\[
\Sigma = \begin{pmatrix} \Sigma_{yy} & O \\ O & \Sigma_{xx} \end{pmatrix}.
\]

The exponent of the moment-generating function becomes

\[
t'\mu + \frac{1}{2}t'\Sigma t = t'_y \mu_y + t'_x \mu_x + \frac{1}{2}t'_y \Sigma_{yy} t_y + \frac{1}{2}t'_x \Sigma_{xx} t_x \\
= M_y(t)M_x(t).
\]

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

Corollary 2.2  If \( y \) is \( \mathcal{N}_p(\mu, \Sigma) \), then any two individual variables \( y_i \) and \( y_j \) are independent if \( \sigma_{i,j} = 0 \).
Corollary 2.3 If $y$ is $N_p(\mu, \Sigma)$ and if $\text{cov}(Ay, By) = A\Sigma B = O$, then $Ay$ and $By$ are independent.

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

$$E(y|x) = \mu_y + \Sigma_{yx} \Sigma_{xx}^{-1} (x - \mu_x),$$

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

Corollary 2.4  If \( \mathbf{v} = (y, \mathbf{x}')' \) with

\[
\begin{align*}
\mathbf{\mu} &= \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \\
\Sigma &= \begin{pmatrix} \sigma_y^2 & \sigma'_yx \\ \sigma'yx & \Sigma_{xx} \end{pmatrix},
\end{align*}
\]

then \( y|\mathbf{x} \) is normal with

\[
\begin{align*}
E(y|bx) &= \mu_y + \sigma'_yx \Sigma_{xx}^{-1}(x - \mu_x), \\
var(y|\mathbf{x}) &= \sigma_y^2 - \sigma'_yx \Sigma_{xx}^{-1} \sigma'yx.
\end{align*}
\]

Since \( \Sigma_{xx}^{-1} \) is positive definite, the corollary implies that

\[
var(y|\mathbf{x}) \leq var(y).
\]
3 Partial Correlation

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

\[
\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}.
\]

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

\[
\rho_{ij \cdot rs...q} = \frac{\sigma_{ij \cdot rs...q}}{\sqrt{\sigma_{ii \cdot rs...q} \sigma_{jj \cdot rs...q}}},
\]

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

\[
\mathbf{P}_{y \cdot x} = \mathbf{D}_{y \cdot x}^{-1} \mathbf{\Sigma}_{y \cdot x} \mathbf{D}_{y \cdot x}^{-1},
\]

where \( \mathbf{\Sigma}_{y \cdot x} = \mathbf{\Sigma}_{yy} - \mathbf{\Sigma}_{yx} \mathbf{\Sigma}_{xx}^{-1} \mathbf{\Sigma}_{xy} \) and \( \mathbf{D}_{y \cdot x} = [\text{diag}(\mathbf{\Sigma}_{y \cdot x})]^{1/2} \).
Unless $y$ and $x$ are independent ($\Sigma_{yx} = O$), the partial correlation $\rho_{ij \cdot rs \ldots q}$ is different from the usual correlation $\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii} \sigma_{jj}}$. 