The multiple integral in (4.21) is equal to 1 because the multivariate normal density in (4.9) integrates to 1 for any mean vector, including $\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{t}$. □

Corollary 1. The moment generating function for $\mathbf{y} - \boldsymbol{\mu}$ is

$$M_{\mathbf{y}-\boldsymbol{\mu}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2}. \tag{4.22}$$

□

We now list two important properties of moment generating functions.

1. If two random vectors have the same moment generating function, they have the same density.
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 $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2')$ and $\mathbf{t}' = (\mathbf{t}_1', \mathbf{t}_2')$, then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent if and only if

$$M_{\mathbf{y}}(\mathbf{t}) = M_{\mathbf{y}_1}(\mathbf{t}_1)M_{\mathbf{y}_2}(\mathbf{t}_2). \tag{4.23}$$

4.4 PROPERTIES OF THE MULTIVARIATE NORMAL DISTRIBUTION

We first consider the distribution of linear functions of multivariate normal random variables.

Theorem 4.4a. Let the $p \times 1$ random vector $\mathbf{y}$ be $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $\mathbf{a}$ be any $p \times 1$ vector of constants, and let $\mathbf{A}$ be any $k \times p$ matrix of constants with rank $k \leq p$. Then

(i) $z = \mathbf{a}'\mathbf{y}$ is $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$
(ii) $\mathbf{z} = \mathbf{A}\mathbf{y}$ is $N_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

PROOF

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

$$\begin{aligned} M_z(t) &= E(e^{tz}) = E(e^{t\mathbf{a}'\mathbf{y}}) = E(e^{(t\mathbf{a})'\mathbf{y}}) \\ &= e^{(t\mathbf{a})'\boldsymbol{\mu} + (t\mathbf{a})'\boldsymbol{\Sigma}(t\mathbf{a})/2} \qquad \text{[by (4.18)]} \\ &= e^{(\mathbf{a}'\boldsymbol{\mu})t + (\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})t^2/2}. \end{aligned} \tag{4.24}$$

On comparing (4.24) with (4.11), it is clear that $z = \mathbf{a}'\mathbf{y}$ is univariate normal with mean $\mathbf{a}'\boldsymbol{\mu}$ and variance $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$.

(ii) The moment generating function for $\mathbf{z} = \mathbf{A}\mathbf{y}$ is given by

$$M_{\mathbf{z}}(\mathbf{t}) = E(e^{\mathbf{t}'\mathbf{z}}) = E(e^{\mathbf{t}'\mathbf{A}\mathbf{y}}),$$

which becomes

$$M_{\mathbf{z}}(\mathbf{t}) = e^{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu}) + \mathbf{t}'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')\mathbf{t}/2} \tag{4.25}$$

(see Problem 4.7). By Corollary 1 to Theorem 2.6b, the covariance matrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ is positive definite. Thus, by (4.18) and (4.25), the $k \times 1$ random vector $\mathbf{z} = \mathbf{A}\mathbf{y}$ is distributed as the k-variate normal $N_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. □

Corollary 1. If $\mathbf{b}$ is any $k \times 1$ vector of constants, then

$$\mathbf{z} = \mathbf{A}\mathbf{y} + \mathbf{b} \quad \text{is} \quad N_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}').$$ □

The marginal distributions of multivariate normal variables are also normal, as shown in the following theorem.

Theorem 4.4b. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any $r \times 1$ subvector of $\mathbf{y}$ has an r-variate normal distribution with the same means, variances, and covariances as in the original p-variate normal distribution.

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

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

Define $\mathbf{A} = (\mathbf{I}_r, \mathbf{O})$, where $\mathbf{I}_r$ is an $r \times r$ identity matrix and $\mathbf{O}$ is an $r \times (p - r)$ matrix of 0s. Then $\mathbf{A}\mathbf{y} = \mathbf{y}_1$, and by Theorem 4.4a (ii), $\mathbf{y}_1$ is distributed as $N_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$. □

Corollary 1. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any individual variable y_i in $\mathbf{y}$ is distributed as $N(\mu_i, \sigma_{ii})$. □

For the next two theorems, we use the notation of Section 3.5, in which the random vector $\mathbf{v}$ is partitioned into two subvectors denoted by $\mathbf{y}$ and $\mathbf{x}$, where $\mathbf{y}$ is $p \times 1$ and $\mathbf{x}$

is $q \times 1$, with a corresponding partitioning of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ [see (3.32) and (3.33)]:

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

By (3.15), if two random variables y_i and y_j are independent, then $\sigma_{ij} = 0$. The converse of this is not true, as illustrated in Example 3.2. By extension, if two random vectors $\mathbf{y}$ and $\mathbf{x}$ are independent (i.e., each y_i is independent of each x_j), then $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$ (the covariance of each y_i with each x_j is 0). The converse is not true in general, but it is true for multivariate normal random vectors.

Theorem 4.4c. If $\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}$ is $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{y}$ and $\mathbf{x}$ are independent if $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$.

Proof. Suppose $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$. Then

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

and the exponent of the moment generating function in (4.18) becomes

$$\begin{aligned} \mathbf{t}'\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} &= (\mathbf{t}'_y, \mathbf{t}'_x)\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix} + \tfrac{1}{2}(\mathbf{t}'_y, \mathbf{t}'_x)\begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}\begin{pmatrix} \mathbf{t}_y \\ \mathbf{t}_x \end{pmatrix} \\ &= \mathbf{t}'_y\boldsymbol{\mu}_y + \mathbf{t}'_x\boldsymbol{\mu}_x + \tfrac{1}{2}\mathbf{t}'_y\boldsymbol{\Sigma}_{yy}\mathbf{t}_y + \tfrac{1}{2}\mathbf{t}'_x\boldsymbol{\Sigma}_{xx}\mathbf{t}_x. \end{aligned}$$

The moment generating function can then be written as

$$M_{\mathbf{v}}(\mathbf{t}) = e^{\mathbf{t}'_y\boldsymbol{\mu}_y + \mathbf{t}'_y\boldsymbol{\Sigma}_{yy}\mathbf{t}_y/2} e^{\mathbf{t}'_x\boldsymbol{\mu}_x + \mathbf{t}'_x\boldsymbol{\Sigma}_{xx}\mathbf{t}_x/2},$$

which is the product of the moment generating functions of $\mathbf{y}$ and $\mathbf{x}$. Hence, by (4.23), $\mathbf{y}$ and $\mathbf{x}$ are independent. □

Corollary 1. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any two individual variables y_i and y_j are independent if $\sigma_{ij} = 0$.

Corollary 2. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\text{cov}(\mathbf{Ay}, \mathbf{By}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{O}$, then $\mathbf{Ay}$ and $\mathbf{By}$ are independent. □

The relationship between subvectors $\mathbf{y}$ and $\mathbf{x}$ when they are not independent ($\boldsymbol{\Sigma}_{yx} \neq \mathbf{O}$) is given in the following theorem.

Theorem 4.4d. If $\mathbf{y}$ and $\mathbf{x}$ are jointly multivariate normal with $\boldsymbol{\Sigma}_{yx} \neq \mathbf{O}$, then the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$, $f(\mathbf{y}|x)$, is multivariate normal with mean vector and covariance matrix

$$E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \tag{4.26}$$

$$\mathrm{cov}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}. \tag{4.27}$$

PROOF. By an extension of (3.18), the conditional density of $\mathbf{y}$ given $\mathbf{x}$ is

$$f(\mathbf{y}|\mathbf{x}) = \frac{g(\mathbf{y}, \mathbf{x})}{h(\mathbf{x})}, \tag{4.28}$$

where $g(\mathbf{y}, \mathbf{x})$ is the joint density of $\mathbf{y}$ and $\mathbf{x}$, and $h(\mathbf{x})$ is the marginal density of $\mathbf{x}$. The proof can be carried out by directly evaluating the ratio on the right hand side of (4.28), using results (2.50) and (2.71) (see Problem 4.13). For variety, we use an alternative approach that avoids working explicitly with $g(\mathbf{y}, \mathbf{x})$ and $h(\mathbf{x})$ and the resulting partitioned matrix formulas.

Consider the function

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{u} \end{pmatrix} = \mathbf{A}\left[\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}\right], \tag{4.29}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} \\ \mathbf{O} & \mathbf{I} \end{pmatrix}.$$

To be conformal, the identity matrix in $\mathbf{A}_1$ is $p \times p$ while the identity in $\mathbf{A}_2$ is $q \times q$. Simplifying and rearranging (4.29), we obtain $\mathbf{w} = \mathbf{y} - [\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)]$ and $\mathbf{u} = \mathbf{x} - \boldsymbol{\mu}_x$. Using the multivariate change-of-variable technique [referred to in (4.6], the joint density of $(\mathbf{w}, \mathbf{u})$ is

$$p(\mathbf{w}, \mathbf{u}) = g(\mathbf{y}, \mathbf{x})|\mathbf{A}^{-1}| = g(\mathbf{y}, \mathbf{x})$$

[employing Theorem 2.9a (ii) and (vi)]. Similarly, the marginal density of $\mathbf{u}$ is

$$q(\mathbf{u}) = h(\mathbf{x})|\mathbf{I}^{-1}| = h(\mathbf{x}).$$

Using (3.45), it also turns out that

$$\mathrm{cov}(\mathbf{w}, \mathbf{u}) = \mathbf{A}_1\boldsymbol{\Sigma}\mathbf{A}_2 = \boldsymbol{\Sigma}_{yx} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xx} = \mathbf{O} \tag{4.30}$$

(see Problem 4.14). Thus, by Theorem 4.4c, $\mathbf{w}$ is independent of $\mathbf{u}$. Hence

$$p(\mathbf{w}, \mathbf{u}) = r(\mathbf{w})q(\mathbf{u}),$$

where $r(\mathbf{w})$ is the density of $\mathbf{w}$. Since $p(\mathbf{w}, \mathbf{u}) = g(\mathbf{y}, \mathbf{x})$ and $q(\mathbf{u}) = h(\mathbf{x})$, we also have

$$g(\mathbf{y}, \mathbf{x}) = r(\mathbf{w})h(\mathbf{x}),$$

and by (4.28),

$$r(\mathbf{w}) = \frac{g(\mathbf{y}, \mathbf{x})}{h(\mathbf{x})} = f(\mathbf{y}|\mathbf{x}).$$

Hence we obtain $f(\mathbf{y}|\mathbf{x})$ simply by finding $r(\mathbf{w})$. By Corollary 1 to Theorem 4.4a, $r(\mathbf{w})$ is the multivariate normal density with

$$\boldsymbol{\mu}_w = \mathbf{A}_1\left[\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}\right] = \mathbf{0}, \tag{4.31}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{ww} &= \mathbf{A}_1\boldsymbol{\Sigma}\mathbf{A}_1' \\
&= (\mathbf{I}, \ -\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1})\begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}\begin{pmatrix} \mathbf{I} \\ -\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \end{pmatrix} \\
&= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}.
\end{aligned} \tag{4.32}$$

Thus $r(\mathbf{w}) = r(\mathbf{y} - [\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)])$ is of the form $N_p(\mathbf{0}, \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})$. Equivalently, $\mathbf{y}|\mathbf{x}$ is $N_p[\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \ \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}]$. $\square$

Since $E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$ in (4.26) is a linear function of $\mathbf{x}$, any pair of variables y_i and y_j in a multivariate normal vector exhibits a linear trend $E(y_i|y_j) = \mu_i + (\sigma_{ij}/\sigma_{jj})(y_j - \mu_j)$. Thus the covariance σ_{ij} is related to the slope of the line representing the trend, and σ_{ij} is a useful measure of relationship between two normal variables. In the case of nonnormal variables that exhibit a curved trend, σ_{ij} may give a very misleading indication of the relationship, as illustrated in Example 3.2.

The conditional covariance matrix $\text{cov}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$ in (4.27) does not involve $\mathbf{x}$. For some nonnormal distributions, on the other hand, $\text{cov}(\mathbf{y}|\mathbf{x})$ is a function of $\mathbf{x}$.

If there is only one y, so that $\mathbf{v}$ is partitioned in the form $\mathbf{v} = (y, x_1, x_2, \ldots, x_q) = (y, \mathbf{x}')$, then $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have the form

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

where μ_y and σ_y^2 are the mean and variance of y, $\boldsymbol{\sigma}_{yx}' = (\sigma_{y1}, \sigma_{y2}, \ldots, \sigma_{yq})$ contains the covariances $\sigma_{yi} = \text{cov}(y, x_i)$, and $\boldsymbol{\Sigma}_{xx}$ contains the variances and covariances of the x variables. The conditional distribution is given in the following corollary to Theorem 4.4d.

Corollary 1. If $\mathbf{v} = (y, x_1, x_2, \ldots, x_q) = (y, \mathbf{x}')$, with

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

then $y|\mathbf{x}$ is normal with

$$E(y|\mathbf{x}) = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \tag{4.33}$$

$$\text{var}(y|\mathbf{x}) = \sigma_y^2 - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}. \tag{4.34}$$

□

In (4.34), $\boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} \geq 0$ because $\boldsymbol{\Sigma}_{xx}^{-1}$ is positive definite. Therefore

$$\text{var}(y|\mathbf{x}) \leq \text{var}(y). \tag{4.35}$$

Example 4.4a. To illustrate Theorems 4.4a–c, suppose that $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 3 \end{pmatrix}.$$

For $z = y_1 - 2y_2 + y_3 = (1, -2, 1)\mathbf{y} = \mathbf{a}'\mathbf{y}$, we have $\mathbf{a}'\boldsymbol{\mu} = 3$ and $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 19$. Hence by Theorem 4.4a(i), z is $N(3, 19)$.

The linear functions

$$z_1 = y_1 - y_2 + y_3, \quad z_2 = 3y_1 + y_2 - 2y_3$$

can be written as

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{Ay}.$$

Then by Theorem 3.6b(i) and Theorem 3.6d(i), we obtain

$$\mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, \quad \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{pmatrix} 14 & 4 \\ 4 & 29 \end{pmatrix},$$

and by Theorem 4.4a(ii), we have

$$\mathbf{z} \text{ is } N_2\left[\begin{pmatrix} 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 14 & 4 \\ 4 & 29 \end{pmatrix}\right].$$

To illustrate the marginal distributions in Theorem 4.4b, note that y_1 is $N(3, 4)$, y_3 is $N(2, 3)$, $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is $N_2\left[\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}\right]$, and $\begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$ is $N_2\left[\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}\right]$.

To illustrate Theorem 4.4c, we note that $\sigma_{12} = 0$, and therefore y_1 and y_2 are independent. □

Example 4.4b. To illustrate Theorem 4.4d, let the random vector $\mathbf{v}$ be $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 5 \\ -2 \\ 1 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{pmatrix}.$$

If $\mathbf{v}$ is partitioned as $\mathbf{v} = (y_1, y_2, x_1, x_2)'$, then $\boldsymbol{\mu}_y = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$, $\boldsymbol{\mu}_x = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$, $\boldsymbol{\Sigma}_{yy} = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{\Sigma}_{yx} = \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}$, and $\boldsymbol{\Sigma}_{xx} = \begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}$. By (4.26), we obtain

$$\begin{aligned}
E(\mathbf{y}|\mathbf{x}) &= \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \\
&= \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} x_1 + 2 \\ x_2 - 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \frac{1}{33}\begin{pmatrix} 30 & 27 \\ -1 & 9 \end{pmatrix}\begin{pmatrix} x_1 + 2 \\ x_2 - 1 \end{pmatrix} \\
&= \begin{pmatrix} 3 + \dfrac{10}{11}x_1 + \dfrac{9}{11}x_2 \\ \dfrac{14}{3} - \dfrac{1}{33}x_1 + \dfrac{3}{11}x_2 \end{pmatrix}.
\end{aligned}$$

By (4.27), we have

$$\begin{aligned}
\text{cov}(\mathbf{y}|\mathbf{x}) &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \\
&= \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} - \tfrac{1}{33}\begin{pmatrix} 171 & 24 \\ 24 & 19 \end{pmatrix} \\
&= \tfrac{1}{33}\begin{pmatrix} 126 & -24 \\ -24 & 14 \end{pmatrix}.
\end{aligned}$$

Thus

$$\mathbf{y}|\mathbf{x} \text{ is } N_2\left[\begin{pmatrix} 3 + \frac{10}{11}x_1 + \frac{9}{11}x_2 \\ \frac{14}{3} - \frac{1}{33}x_1 + \frac{3}{11}x_2 \end{pmatrix}, \frac{1}{33}\begin{pmatrix} 126 & -24 \\ -24 & 14 \end{pmatrix}\right].$$

□

Example 4.4c. To illustrate Corollary 1 to Theorem 4.4d, let $\mathbf{v}$ be $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are as given in Example 4.4b. If $\mathbf{v}$ is partitioned as $\mathbf{v} = (y, x_1, x_2, x_3)'$, then $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix} = \left(\begin{array}{c} 2 \\ 5 \\ \hline -2 \\ 1 \end{array}\right),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} = \left(\begin{array}{c|ccc} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ \hline 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{array}\right).$$

By (4.33), we have

$$\begin{aligned} E(y|x_1, x_2, x_3) &= \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \\ &= 2 + (0, 3, 3)\begin{pmatrix} 1 & -1 & 2 \\ -1 & 6 & -3 \\ 2 & -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} x_1 - 5 \\ x_2 + 2 \\ x_3 + 1 \end{pmatrix} \\ &= \frac{95}{7} - \frac{12}{7}x_1 + \frac{6}{7}x_2 + \frac{9}{7}x_3. \end{aligned}$$

By (4.34), we obtain

$$\begin{aligned} \text{var}(y|x_1, x_2, x_3) &= \sigma_y^2 - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} \\ &= 9 - (0, 3, 3)\begin{pmatrix} 1 & -1 & 2 \\ -1 & 6 & -3 \\ 2 & -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix} \\ &= 9 - \frac{45}{7} = \frac{18}{7}. \end{aligned}$$

Hence $y|x_1, x_2, x_3$ is $N(\frac{95}{7} - \frac{12}{7}x_1 + \frac{6}{7}x_2 + \frac{9}{7}x_3, \frac{18}{7})$. Note that $\text{var}(y|x_1, x_2, x_3) = \frac{18}{7}$ is less than $\text{var}(y) = 9$, which illustrates (4.35). □