3.20. In northern climates, roads must be cleared of snow quickly following a storm. One measure of storm severity is x_1 = its duration in hours, while the effectiveness of snow removal can be quantified by x_2 = the number of hours crews, men, and machine, spend to clear snow. Here are the results for 25 incidents in Wisconsin.

Table 3.2 Snow Data

x_1	x_2	x_1	x_2	x_1	x_2
12.5	13.7	9.0	24.4	3.5	26.1
14.5	16.5	6.5	18.2	8.0	14.5
8.0	17.4	10.5	22.0	17.5	42.3
9.0	11.0	10.0	32.5	10.5	17.5
19.5	23.6	4.5	18.7	12.0	21.8
8.0	13.2	7.0	15.8	6.0	10.4
9.0	32.1	8.5	15.6	13.0	25.6
7.0	12.3	6.5	12.0		
7.0	11.8	8.0	12.8		

(a) Find the sample mean and variance of the difference $x_2 - x_1$ by first obtaining the summary statistics.

(b) Obtain the mean and variance by first obtaining the individual values $x_{j2} - x_{j1}$, for $j = 1, 2, \ldots, 25$ and then calculating the mean and variance. Compare these values with those obtained in part a.

References

1. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley, 2003.
2. Eaton, M., and M. Perlman. "The Non-Singularity of Generalized Sample Covariance Matrices." *Annals of Statistics*, **1** (1973), 710–717.

Chapter 4

THE MULTIVARIATE NORMAL DISTRIBUTION

4.1 Introduction

A generalization of the familiar bell-shaped normal density to several dimensions plays a fundamental role in multivariate analysis. In fact, most of the techniques encountered in this book are based on the assumption that the data were generated from a *multivariate* normal distribution. While real data are never *exactly* multivariate normal, the normal density is often a useful approximation to the "true" population distribution.

One advantage of the multivariate normal distribution stems from the fact that it is mathematically tractable and "nice" results can be obtained. This is frequently not the case for other data-generating distributions. Of course, mathematical attractiveness per se is of little use to the practitioner. It turns out, however, that normal distributions are useful in practice for two reasons: First, the normal distribution serves as a bona fide population model in some instances; second, the sampling distributions of many multivariate statistics are approximately normal, regardless of the form of the parent population, because of a *central limit* effect.

To summarize, many real-world problems fall naturally within the framework of normal theory. The importance of the normal distribution rests on its dual role as both population model for certain natural phenomena and approximate sampling distribution for many statistics.

4.2 The Multivariate Normal Density and Its Properties

The multivariate normal density is a generalization of the univariate normal density to $p \geq 2$ dimensions. Recall that the univariate normal distribution, with mean μ and variance σ^2, has the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \qquad -\infty < x < \infty \qquad (4\text{-}1)$$

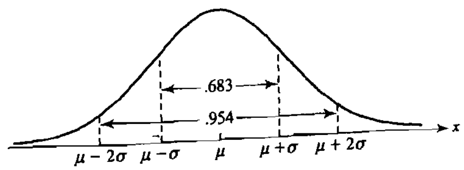

Figure 4.1 A normal density with mean μ and variance σ^2 and selected areas under the curve.

A plot of this function yields the familiar bell-shaped curve shown in Figure 4.1. Also shown in the figure are approximate areas under the curve within ± 1 standard deviations and ± 2 standard deviations of the mean. These areas represent probabilities, and thus, for the normal random variable X,

$$P(\mu - \sigma \le X \le \mu + \sigma) \doteq .68$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) \doteq .95$$

It is convenient to denote the normal density function with mean μ and variance σ^2 by $N(\mu, \sigma^2)$. Therefore, $N(10, 4)$ refers to the function in (4-1) with $\mu = 10$ and $\sigma = 2$. This notation will be extended to the multivariate case later.

The term

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu) \tag{4-2}$$

in the exponent of the univariate normal density function measures the square of the distance from x to μ in standard deviation units. This can be generalized for a $p \times 1$ vector $\mathbf{x}$ of observations on several variables as

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \tag{4-3}$$

The $p \times 1$ vector $\boldsymbol{\mu}$ represents the expected value of the random vector $\mathbf{X}$, and the $p \times p$ matrix $\boldsymbol{\Sigma}$ is the variance–covariance matrix of $\mathbf{X}$. [See (2-30) and (2-31).] We shall assume that the symmetric matrix $\boldsymbol{\Sigma}$ is positive definite, so the expression in (4-3) is the square of the generalized distance from $\mathbf{x}$ to $\boldsymbol{\mu}$.

The multivariate normal density is obtained by replacing the univariate distance in (4-2) by the multivariate generalized distance of (4-3) in the density function of (4-1). When this replacement is made, the univariate normalizing constant $(2\pi)^{-1/2}(\sigma^2)^{-1/2}$ must be changed to a more general constant that makes the *volume* under the surface of the multivariate density function unity for any p. This is necessary because, in the multivariate case, probabilities are represented by volumes under the surface over regions defined by intervals of the x_i values. It can be shown (see [1]) that this constant is $(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}$, and consequently, a p-dimensional normal density for the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ has the form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2} \tag{4-4}$$

where $-\infty < x_i < \infty$, $i = 1, 2, \ldots, p$. We shall denote this p-dimensional normal density by $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which is analogous to the normal density in the univariate case.

Example 4.1 (Bivariate normal density) Let us evaluate the $p = 2$-variate normal density in terms of the individual parameters $\mu_1 = E(X_1)$, $\mu_2 = E(X_2)$, $\sigma_{11} = \text{Var}(X_1)$, $\sigma_{22} = \text{Var}(X_2)$, and $\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = \text{Corr}(X_1, X_2)$.

Using Result 2A.8, we find that the inverse of the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

is

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

Introducing the correlation coefficient ρ_{12} by writing $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$, we obtain $\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho_{12}^2)$, and the squared distance becomes

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

$$= [x_1 - \mu_1, x_2 - \mu_2]\frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$= \frac{\sigma_{22}(x_1 - \mu_1)^2 + \sigma_{11}(x_2 - \mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)}$$

$$= \frac{1}{1 - \rho_{12}^2}\left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)\right] \tag{4-5}$$

The last expression is written in terms of the standardized values $(x_1 - \mu_1)/\sqrt{\sigma_{11}}$ and $(x_2 - \mu_2)/\sqrt{\sigma_{22}}$.

Next, since $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho_{12}^2)$, we can substitute for $\boldsymbol{\Sigma}^{-1}$ and $|\boldsymbol{\Sigma}|$ in (4-4) to get the expression for the bivariate ($p = 2$) normal density involving the individual parameters $\mu_1, \mu_2, \sigma_{11}, \sigma_{22}$, and ρ_{12}:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)}} \times \exp\left\{-\frac{1}{2(1 - \rho_{12}^2)}\left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\} \tag{4-6}$$

The expression in (4-6) is somewhat unwieldy, and the compact general form in (4-4) is more informative in many ways. On the other hand, the expression in (4-6) is useful for discussing certain properties of the normal distribution. For example, if the random variables X_1 and X_2 are uncorrelated, so that $\rho_{12} = 0$, the joint density can be written as the product of two univariate normal densities each of the form of (4-1).

That is, $f(x_1, x_2) = f(x_1)f(x_2)$ and X_1 and X_2 are independent. [See (2-28).] This result is true in general. (See Result 4.5.)

Two bivariate distributions with $\sigma_{11} = \sigma_{22}$ are shown in Figure 4.2. In Figure 4.2(a), X_1 and X_2 are independent ($\rho_{12} = 0$). In Figure 4.2(b), $\rho_{12} = .75$. Notice how the presence of correlation causes the probability to concentrate along a line. ■

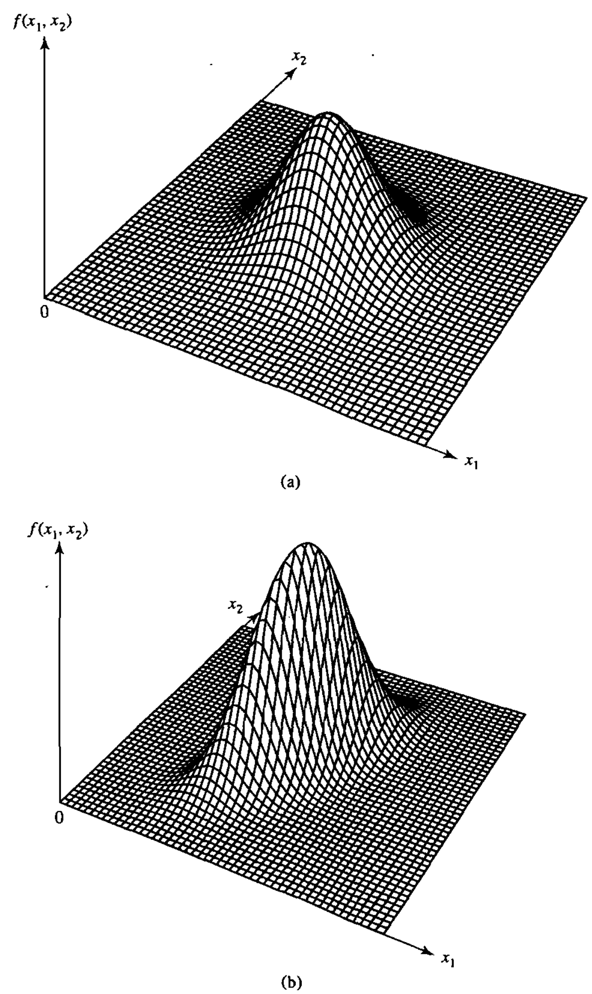

Figure 4.2 Two bivariate normal distributions. (a) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = 0$. (b) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = .75$.

From the expression in (4-4) for the density of a p-dimensional normal variable, it should be clear that the paths of $\mathbf{x}$ values yielding a constant height for the density are ellipsoids. That is, the multivariate normal density is constant on surfaces where the square of the distance $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ is constant. These paths are called *contours*:

$$\textit{Constant probability density contour} = \{\text{all } \mathbf{x} \text{ such that } (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2\}$$
$$= \text{surface of an ellipsoid centered at } \boldsymbol{\mu}$$

The axes of each ellipsoid of constant density are in the direction of the eigenvectors of $\boldsymbol{\Sigma}^{-1}$, and their lengths are proportional to the reciprocals of the square roots of the eigenvalues of $\boldsymbol{\Sigma}^{-1}$. Fortunately, we can avoid the calculation of $\boldsymbol{\Sigma}^{-1}$ when determining the axes, since these ellipsoids are also determined by the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$. We state the correspondence formally for later reference.

Result 4.1. If $\boldsymbol{\Sigma}$ is positive definite, so that $\boldsymbol{\Sigma}^{-1}$ exists, then

$$\boldsymbol{\Sigma}\mathbf{e} = \lambda\mathbf{e} \quad \text{implies} \quad \boldsymbol{\Sigma}^{-1}\mathbf{e} = \left(\frac{1}{\lambda}\right)\mathbf{e}$$

so $(\lambda, \mathbf{e})$ is an eigenvalue–eigenvector pair for $\boldsymbol{\Sigma}$ corresponding to the pair $(1/\lambda, \mathbf{e})$ for $\boldsymbol{\Sigma}^{-1}$. Also, $\boldsymbol{\Sigma}^{-1}$ is positive definite.

Proof. For $\boldsymbol{\Sigma}$ positive definite and $\mathbf{e} \neq \mathbf{0}$ an eigenvector, we have $0 < \mathbf{e}'\boldsymbol{\Sigma}\mathbf{e} = \mathbf{e}'(\boldsymbol{\Sigma}\mathbf{e}) = \mathbf{e}'(\lambda\mathbf{e}) = \lambda\mathbf{e}'\mathbf{e} = \lambda$. Moreover, $\mathbf{e} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\mathbf{e}) = \boldsymbol{\Sigma}^{-1}(\lambda\mathbf{e})$, or $\mathbf{e} = \lambda\boldsymbol{\Sigma}^{-1}\mathbf{e}$, and division by $\lambda > 0$ gives $\boldsymbol{\Sigma}^{-1}\mathbf{e} = (1/\lambda)\mathbf{e}$. Thus, $(1/\lambda, \mathbf{e})$ is an eigenvalue–eigenvector pair for $\boldsymbol{\Sigma}^{-1}$. Also, for any $p \times 1$ $\mathbf{x}$, by (2-21)

$$\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} = \mathbf{x}'\left(\sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)\mathbf{e}_i\mathbf{e}_i'\right)\mathbf{x}$$
$$= \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)(\mathbf{x}'\mathbf{e}_i)^2 \geq 0$$

since each term $\lambda_i^{-1}(\mathbf{x}'\mathbf{e}_i)^2$ is nonnegative. In addition, $\mathbf{x}'\mathbf{e}_i = 0$ for all i only if $\mathbf{x} = \mathbf{0}$. So $\mathbf{x} \neq \mathbf{0}$ implies that $\sum_{i=1}^{p}(1/\lambda_i)(\mathbf{x}'\mathbf{e}_i)^2 > 0$, and it follows that $\boldsymbol{\Sigma}^{-1}$ is positive definite. ■

The following summarizes these concepts:

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such the that

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2 \tag{4-7}$$

These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\,\mathbf{e}_i$, where $\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i\mathbf{e}_i$ for $i = 1, 2, \ldots, p$.

A contour of constant density for a bivariate normal distribution with $\sigma_{11} = \sigma_{22}$ is obtained in the following example.

Example 4.2 (Contours of the bivariate normal density) We shall obtain the axes of constant probability density contours for a bivariate normal distribution when $\sigma_{11} = \sigma_{22}$. From (4-7), these axes are given by the eigenvalues and eigenvectors of $\mathbf{\Sigma}$. Here $|\mathbf{\Sigma} - \lambda\mathbf{I}| = 0$ becomes

$$0 = \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix} = (\sigma_{11} - \lambda)^2 - \sigma_{12}^2$$

$$= (\lambda - \sigma_{11} - \sigma_{12})(\lambda - \sigma_{11} + \sigma_{12})$$

Consequently, the eigenvalues are $\lambda_1 = \sigma_{11} + \sigma_{12}$ and $\lambda_2 = \sigma_{11} - \sigma_{12}$. The eigenvector $\mathbf{e}_1$ is determined from

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = (\sigma_{11} + \sigma_{12}) \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

or

$$\sigma_{11}e_1 + \sigma_{12}e_2 = (\sigma_{11} + \sigma_{12})e_1$$

$$\sigma_{12}e_1 + \sigma_{11}e_2 = (\sigma_{11} + \sigma_{12})e_2$$

These equations imply that $e_1 = e_2$, and after normalization, the first eigenvalue–eigenvector pair is

$$\lambda_1 = \sigma_{11} + \sigma_{12}, \qquad \mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Similarly, $\lambda_2 = \sigma_{11} - \sigma_{12}$ yields the eigenvector $\mathbf{e}_2' = [1/\sqrt{2}, -1/\sqrt{2}]$.

When the covariance σ_{12} (or correlation ρ_{12}) is positive, $\lambda_1 = \sigma_{11} + \sigma_{12}$ is the *largest* eigenvalue, and its associated eigenvector $\mathbf{e}_1' = [1/\sqrt{2}, 1/\sqrt{2}]$ lies along the 45° line through the point $\boldsymbol{\mu}' = [\mu_1, \mu_2]$. This is true for any positive value of the covariance (correlation). Since the axes of the constant-density ellipses are given by $\pm c\sqrt{\lambda_1}\,\mathbf{e}_1$ and $\pm c\sqrt{\lambda_2}\,\mathbf{e}_2$ [see (4-7)], and the eigenvectors each have length unity, the major axis will be associated with the largest eigenvalue. For positively correlated normal random variables, then, the *major* axis of the constant-density ellipses will be along the 45° line through $\boldsymbol{\mu}$. (See Figure 4.3.)

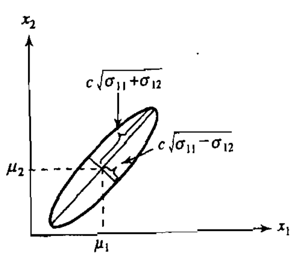

Figure 4.3 A constant-density contour for a bivariate normal distribution with $\sigma_{11} = \sigma_{22}$ and $\sigma_{12} > 0$ (or $\rho_{12} > 0$).

When the covariance (correlation) is negative, $\lambda_2 = \sigma_{11} - \sigma_{12}$ will be the largest eigenvalue, and the major axes of the constant-density ellipses will lie along a line at right angles to the 45° line through $\boldsymbol{\mu}$. (These results are true only for $\sigma_{11} = \sigma_{22}$.)

To summarize, the axes of the ellipses of constant density for a bivariate normal distribution with $\sigma_{11} = \sigma_{22}$ are determined by

$$\pm c\sqrt{\sigma_{11} + \sigma_{12}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad \pm c\sqrt{\sigma_{11} - \sigma_{12}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

■

We show in Result 4.7 that the choice $c^2 = \chi_p^2(\alpha)$, where $\chi_p^2(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p degrees of freedom, leads to contours that contain $(1 - \alpha) \times 100\%$ of the probability. Specifically, the following is true for a p-dimensional normal distribution:

The solid ellipsoid of $\mathbf{x}$ values satisfying

$$(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \le \chi_p^2(\alpha) \tag{4-8}$$

has probability $1 - \alpha$.

The constant-density contours containing 50% and 90% of the probability under the bivariate normal surfaces in Figure 4.2 are pictured in Figure 4.4.

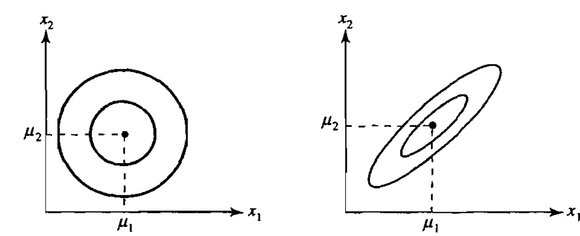

Figure 4.4 The 50% and 90% contours for the bivariate normal distributions in Figure 4.2.

The p-variate normal density in (4-4) has a maximum value when the squared distance in (4-3) is zero—that is, when $\mathbf{x} = \boldsymbol{\mu}$. Thus, $\boldsymbol{\mu}$ is the point of maximum density, or *mode*, as well as the expected value of $\mathbf{X}$, or *mean*. The fact that $\boldsymbol{\mu}$ is the mean of the multivariate normal distribution follows from the symmetry exhibited by the constant-density contours: These contours are centered, or balanced, at $\boldsymbol{\mu}$.

Additional Properties of the Multivariate Normal Distribution

Certain properties of the normal distribution will be needed repeatedly in our explanations of statistical models and methods. These properties make it possible to manipulate normal distributions easily and, as we suggested in Section 4.1, are partly responsible for the popularity of the normal distribution. The key properties, which we shall soon discuss in some mathematical detail, can be stated rather simply.

The following are true for a random vector $\mathbf{X}$ having a multivariate normal distribution:

1. Linear combinations of the components of $\mathbf{X}$ are normally distributed.
2. All subsets of the components of $\mathbf{X}$ have a (multivariate) normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are (multivariate) normal.

These statements are reproduced mathematically in the results that follow. Many of these results are illustrated with examples. The proofs that are included should help improve your understanding of matrix manipulations and also lead you to an appreciation for the manner in which the results successively build on themselves.

Result 4.2 can be taken as a working definition of the normal distribution. With this in hand, the subsequent properties are almost immediate. Our partial proof of Result 4.2 indicates how the linear combination definition of a normal density relates to the multivariate density in (4-4).

Result 4.2. If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any linear combination of variables $\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$. Also, if $\mathbf{a}'\mathbf{X}$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for every $\mathbf{a}$, then $\mathbf{X}$ must be $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Proof. The expected value and variance of $\mathbf{a}'\mathbf{X}$ follow from (2-43). Proving that $\mathbf{a}'\mathbf{X}$ is normally distributed if $\mathbf{X}$ is multivariate normal is more difficult. You can find a proof in [1]. The second part of result 4.2 is also demonstrated in [1]. ■

Example 4.3 (The distribution of a linear combination of the components of a normal random vector) Consider the linear combination $\mathbf{a}'\mathbf{X}$ of a multivariate normal random vector determined by the choice $\mathbf{a}' = [1, 0, \ldots, 0]$. Since

$$\mathbf{a}'\mathbf{X} = [1, 0, \ldots, 0]\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = X_1$$

and

$$\mathbf{a}'\boldsymbol{\mu} = [1, 0, \ldots, 0]\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_1$$

we have

$$\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = [1, 0, \ldots, 0]\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{11}$$

and it follows from Result 4.2 that X_1 is distributed as $N(\mu_1, \sigma_{11})$. More generally, the marginal distribution of any component X_i of $\mathbf{X}$ is $N(\mu_i, \sigma_{ii})$. ■

The next result considers several linear combinations of a multivariate normal vector $\mathbf{X}$.

Result 4.3. If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the q linear combinations

$$\underset{(q\times p)(p\times 1)}{\mathbf{A}\;\mathbf{X}} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. Also, $\underset{(p\times 1)}{\mathbf{X}} + \underset{(p\times 1)}{\mathbf{d}}$, where $\mathbf{d}$ is a vector of constants, is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$.

Proof. The expected value $E(\mathbf{AX})$ and the covariance matrix of $\mathbf{AX}$ follow from (2-45). Any linear combination $\mathbf{b}'(\mathbf{AX})$ is a linear combination of $\mathbf{X}$, of the form $\mathbf{a}'\mathbf{X}$ with $\mathbf{a} = \mathbf{A}'\mathbf{b}$. Thus, the conclusion concerning $\mathbf{AX}$ follows directly from Result 4.2.

The second part of the result can be obtained by considering $\mathbf{a}'(\mathbf{X} + \mathbf{d}) = \mathbf{a}'\mathbf{X} + (\mathbf{a}'\mathbf{d})$, where $\mathbf{a}'\mathbf{X}$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$. It is known from the univariate case that adding a constant $\mathbf{a}'\mathbf{d}$ to the random variable $\mathbf{a}'\mathbf{X}$ leaves the variance unchanged and translates the mean to $\mathbf{a}'\boldsymbol{\mu} + \mathbf{a}'\mathbf{d} = \mathbf{a}'(\boldsymbol{\mu} + \mathbf{d})$. Since $\mathbf{a}$ was arbitrary, $\mathbf{X} + \mathbf{d}$ is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$. ■

Example 4.4 (The distribution of two linear combinations of the components of a normal random vector) For $\mathbf{X}$ distributed as $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, find the distribution of

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{AX}$$

By Result 4.3, the distribution of $\mathbf{AX}$ is multivariate normal with mean

$$\mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

and covariance matrix

$$\mathbf{A\Sigma A}' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} & \sigma_{13} - \sigma_{23} \\ \sigma_{12} - \sigma_{13} & \sigma_{22} - \sigma_{23} & \sigma_{23} - \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} \\ \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

Alternatively, the mean vector $\mathbf{A}\boldsymbol{\mu}$ and covariance matrix $\mathbf{A\Sigma A}'$ may be verified by direct calculation of the means and covariances of the two random variables $Y_1 = X_1 - X_2$ and $Y_2 = X_2 - X_3$. ■

We have mentioned that all subsets of a multivariate normal random vector $\mathbf{X}$ are themselves normally distributed. We state this property formally as Result 4.4.

Result 4.4. All subsets of $\mathbf{X}$ are normally distributed. If we respectively partition $\mathbf{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{\Sigma}$ as

$$\underset{(p\times 1)}{\mathbf{X}} = \begin{bmatrix} \underset{(q\times 1)}{\mathbf{X}_1} \\ \cdots \\ \underset{((p-q)\times 1)}{\mathbf{X}_2} \end{bmatrix} \qquad \underset{(p\times 1)}{\boldsymbol{\mu}} = \begin{bmatrix} \underset{(q\times 1)}{\boldsymbol{\mu}_1} \\ \cdots \\ \underset{((p-q)\times 1)}{\boldsymbol{\mu}_2} \end{bmatrix}$$

and

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \left[\begin{array}{c|c} \underset{(q\times q)}{\boldsymbol{\Sigma}_{11}} & \underset{(q\times(p-q))}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{((p-q)\times q)}{\boldsymbol{\Sigma}_{21}} & \underset{((p-q)\times(p-q))}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

then $\mathbf{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Proof. Set $\underset{(q\times p)}{\mathbf{A}} = \left[\underset{(q\times q)}{\mathbf{I}} \,\vdots\, \underset{(q\times(p-q))}{\mathbf{0}}\right]$ in Result 4.3, and the conclusion follows. To apply Result 4.4 to an *arbitrary* subset of the components of $\mathbf{X}$, we simply relabel the subset of interest as $\mathbf{X}_1$ and select the corresponding component means and covariances as $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_{11}$, respectively. ■

Example 4.5 (The distribution of a subset of a normal random vector)

If $\mathbf{X}$ is distributed as $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$. We set

$$\mathbf{X}_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \quad \boldsymbol{\mu}_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

and note that with this assignment, $\mathbf{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ can respectively be rearranged and partitioned as

$$\mathbf{X} = \begin{bmatrix} X_2 \\ X_4 \\ \hline X_1 \\ X_3 \\ X_5 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \hline \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \left[\begin{array}{cc|ccc} \sigma_{22} & \sigma_{24} & \sigma_{12} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \hline \sigma_{12} & \sigma_{14} & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{23} & \sigma_{34} & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \sigma_{25} & \sigma_{45} & \sigma_{15} & \sigma_{35} & \sigma_{55} \end{array}\right]$$

or

$$\mathbf{X} = \begin{bmatrix} \underset{(2\times 1)}{\mathbf{X}_1} \\ \cdots \\ \underset{(3\times 1)}{\mathbf{X}_2} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \underset{(2\times 1)}{\boldsymbol{\mu}_1} \\ \cdots \\ \underset{(3\times 1)}{\boldsymbol{\mu}_2} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \left[\begin{array}{c|c} \underset{(2\times 2)}{\boldsymbol{\Sigma}_{11}} & \underset{(2\times 3)}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{(3\times 2)}{\boldsymbol{\Sigma}_{21}} & \underset{(3\times 3)}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

Thus, from Result 4.4, for

$$\mathbf{X}_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$$

we have the distribution

$$N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) = N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}\right)$$

It is clear from this example that the normal distribution for any subset can be expressed by simply selecting the appropriate means and covariances from the original $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The formal process of relabeling and partitioning is unnecessary. ■

We are now in a position to state that zero correlation between normal random variables or sets of normal random variables is equivalent to statistical independence.

Result 4.5.

(a) If $\underset{(q_1\times 1)}{\mathbf{X}_1}$ and $\underset{(q_2\times 1)}{\mathbf{X}_2}$ are independent, then $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

(b) If $\begin{bmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

(c) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, respectively, then $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$$

Proof. (See Exercise 4.14 for partial proofs based upon factoring the density function when $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.) ∎

Example 4.6 (The equivalence of zero covariance and independence for normal variables) Let $\underset{(3\times1)}{\mathbf{X}}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Are X_1 and X_2 independent? What about (X_1, X_2) and X_3?

Since X_1 and X_2 have covariance $\sigma_{12} = 1$, they are not independent. However, partitioning $\mathbf{X}$ and $\boldsymbol{\Sigma}$ as

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \hline X_3 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \left[\begin{array}{cc|c} 4 & 1 & 0 \\ 1 & 3 & 0 \\ \hline 0 & 0 & 2 \end{array}\right] = \begin{bmatrix} \underset{(2\times2)}{\boldsymbol{\Sigma}_{11}} & \underset{(2\times1)}{\boldsymbol{\Sigma}_{12}} \\ \underset{(1\times2)}{\boldsymbol{\Sigma}_{21}} & \underset{(1\times1)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}$$

we see that $\mathbf{X}_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and X_3 have covariance matrix $\boldsymbol{\Sigma}_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, (X_1, X_2) and X_3 are independent by Result 4.5. This implies X_3 is independent of X_1 and also of X_2. ∎

We pointed out in our discussion of the bivariate normal distribution that $\rho_{12} = 0$ (zero correlation) implied independence because the joint density function [see (4-6)] could then be written as the product of the marginal (normal) densities of X_1 and X_2. This fact, which we encouraged you to verify directly, is simply a special case of Result 4.5 with $q_1 = q_2 = 1$.

Result 4.6. Let $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, and $|\boldsymbol{\Sigma}_{22}| > 0$. Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

$$\text{Mean} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

and

$$\text{Covariance} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Note that the covariance does not depend on the value $\mathbf{x}_2$ of the conditioning variable.

Proof. We shall give an indirect proof. (See Exercise 4.13, which uses the densities directly.) Take

$$\underset{(p\times p)}{\mathbf{A}} = \begin{bmatrix} \underset{(q\times q)}{\mathbf{I}} & \underset{q\times(p-q)}{-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}} \\ \underset{(p-q)\times q}{\mathbf{0}} & \underset{(p-q)\times(p-q)}{\mathbf{I}} \end{bmatrix}$$

so

$$\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{A}\begin{bmatrix} \mathbf{X}_1 - \boldsymbol{\mu}_1 \\ \mathbf{X}_2 - \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \\ \mathbf{X}_2 - \boldsymbol{\mu}_2 \end{bmatrix}$$

is jointly normal with covariance matrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ given by

$$\begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{0}' \\ (-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1})' & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Since $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ and $\mathbf{X}_2 - \boldsymbol{\mu}_2$ have zero covariance, they are independent. Moreover, the quantity $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ has distribution $N_q(\mathbf{0}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$. Given that $\mathbf{X}_2 = \mathbf{x}_2$, $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ is a constant. Because $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ and $\mathbf{X}_2 - \boldsymbol{\mu}_2$ are independent, the conditional distribution of $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ is the same as the unconditional distribution of $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$. Since $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ is $N_q(\mathbf{0}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$, so is the random vector $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ when $\mathbf{X}_2$ has the particular value $\mathbf{x}_2$. Equivalently, given that $\mathbf{X}_2 = \mathbf{x}_2$, $\mathbf{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$. ∎

Example 4.7 (The conditional density of a bivariate normal distribution) The conditional density of X_1, given that $X_2 = x_2$ for any bivariate distribution, is defined by

$$f(x_1 | x_2) = \{\text{conditional density of } X_1 \text{ given that } X_2 = x_2\} = \frac{f(x_1, x_2)}{f(x_2)}$$

where $f(x_2)$ is the marginal distribution of X_2. If $f(x_1, x_2)$ is the bivariate normal density, show that $f(x_1 | x_2)$ is

$$N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right)$$