3.5 The Multivariate Normal Distribution

In this section we present the multivariate normal distribution. In the first part of the section, we introduce the bivariate normal distribution, leaving most of the proofs to the later section, Section 3.5.2.

3.5.1 Bivariate Normal Distribution

We say that (X, Y) follows a bivariate normal distribution if its pdf is given by

$$f(x, y) \quad \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\ q\ 2}, \quad -\infty \quad x \quad \infty, \quad -\infty \quad y \quad \infty, \tag{3.5.1}$$

where

$$q \quad \frac{1}{1-\rho^2}\left[\left(\frac{x - \ _1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x - \ _1}{\sigma_1}\right)\left(\frac{y - \ _2}{\sigma_2}\right) \quad \left(\frac{y - \ _2}{\sigma_2}\right)^2\right], \tag{3.5.2}$$

and $-\infty \quad _i \quad \infty$, $\sigma_i \quad 0$, for $i \quad 1, 2$, and ρ satisfies $\rho^2 \quad 1$. Clearly, this function is positive everywhere in R^2. As we show in Section 3.5.2, it is a pdf with the mgf given by:

$$M_{(X,Y)}(t_1, t_2) \quad \exp\left\{t_1 \ _1 \quad t_2 \ _2 \quad \frac{1}{2}(t_1^2\sigma_1^2 \quad 2t_1t_2\rho\sigma_1\sigma_2 \quad t_2^2\sigma_2^2)\right\}. \tag{3.5.3}$$

Thus, the mgf of X is

$$M_X(t_1) \quad M_{(X,Y)}(t_1, 0) \quad \exp\left\{t_1 \ _1 \quad \frac{1}{2}t_1^2\sigma_1^2\right\}$$

hence, X has a $N(\ _1, \sigma_1^2)$ distribution. In the same way, Y has a $N(\ _2, \sigma_2^2)$ distribution. Thus $\ _1$ and $\ _2$ are the respective means of X and Y and σ_1^2 and σ_2^2 are the respective variances of X and Y. For the parameter ρ, Exercise 3.5.3 shows that

$$E(XY) \quad \frac{\partial^2 M_{(X,Y)}}{\partial t_1 \partial t_2}(0, 0) \quad \rho\sigma_1\sigma_2 \quad \ _1 \ _2. \tag{3.5.4}$$

Hence, $\mathrm{cov}(X, Y) \quad \rho\sigma_1\sigma_2$ and thus, as the notation suggests, ρ is the correlation coefficient between X and Y. We know by Theorem 2.5.2 that if X and Y are independent then $\rho \quad 0$. Further, from expression (3.5.3), if $\rho \quad 0$ then the joint mgf of (X, Y) factors into the product of the marginal mgfs and, hence, X and Y are independent random variables. Thus if (X, Y) has a bivariate normal distribution, then X and Y are independent if and only if they are uncorrelated.

The bivariate normal pdf, (3.5.1), is mound shaped over R^2 and peaks at its mean $(\ _1, \ _2)$ see Exercise 3.5.4. For a given $c \quad 0$, the points of equal probability (or density) are given by $\{(x, y) : f(x, y) \quad c\}$. It follows with some algebra that these sets are ellipses. In general for multivariate distributions, we call these sets contours of the pdfs. Hence, the contours of bivariate normal distributions are elliptical. If X and Y are independent then these contours are circular. The interested reader can consult a book on multivariate statistics for discussions on the geometry of the ellipses. For example, if $\sigma_1 \quad \sigma_2$ and $\rho > 0$, the main axis of the ellipse goes through the mean at a 45^{o} angle see Johnson and Wichern (2008) for discussion.

Figure 3.5.1 displays a three-dimensional plot of the bivariate normal pdf with $(\; _1, \; _2) \quad (0,0)$, $\sigma_1 \quad \sigma_2 \quad 1$, and $\rho \quad 0.5$. For location, the peak is at $(\; _1, \; _2)$ $(0,0)$. The elliptical contours are apparent. Locate the main axis. For a region A in the plane, P $(X,Y) \in A$ is the volume under the surface over A. In general such probabilities are calculated by numerical integration methods.

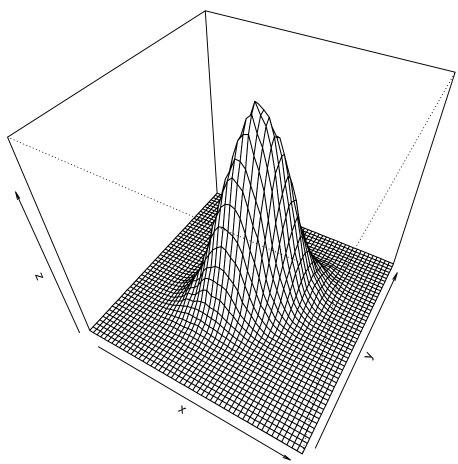

Figure 3.5.1: A sketch of the surface of a bivariate normal distribution with mean $(0,0)$, $\sigma_1 \quad \sigma_2 \quad 1$, and $\rho \quad 0.5$.

In the next section, we extend the discussion to the general multivariate case however, Remark 3.5.1, below, returns to the bivariate case and can be read with minor knowledge of vector and matrices.

3.5.2 *Multivariate Normal Distribution, General Case

In this section we generalize the bivariate normal distribution to the n-dimensional multivariate normal distribution. As with Section 3.4 on the normal distribution, the derivation of the distribution is simplified by first discussing the standardized variable case and then proceeding to the general case. Also, in this section, vector and matrix notation are used.

Consider the random vector $\mathbf{Z} \quad (Z_1, \ldots, Z_n)'$, where $Z_1, \ldots, Z_n$ are iid $N(0,1)$ random variables. Then the density of $\mathbf{Z}$ is

$$
\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) \quad & \prod_{i\ \ 1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z_i^2\right\} \quad \left(\frac{1}{2\pi}\right)^{n\ \ 2} \exp\left\{-\frac{1}{2}\sum_{i\ \ 1}^{n} z_i^2\right\} \\
& \left(\frac{1}{2\pi}\right)^{n\ \ 2} \exp\left\{-\frac{1}{2}\mathbf{z}'\mathbf{z}\right\},
\end{aligned}
\tag{3.5.5}
$$

for $\mathbf{z} \in R^n$. Because the Z_is have mean 0, have variance 1, and are uncorrelated, the mean and covariance matrix of $\mathbf{Z}$ are

$$
E[\mathbf{Z}] \quad \mathbf{0} \text{ and } \operatorname{Cov}[\mathbf{Z}] \quad \mathbf{I}_n, \tag{3.5.6}
$$

where $\mathbf{I}_n$ denotes the identity matrix of order n. Recall that the mgf of Z_i evaluated at t_i is $\exp\{t_i^2\ \ 2\}$. Hence, because the Z_is are independent, the mgf of $\mathbf{Z}$ is

$$
\begin{aligned}
M_{\mathbf{Z}}(\mathbf{t}) \quad E[\exp\{\mathbf{t}'\mathbf{Z}\}] \quad & E\left[\prod_{i\ \ 1}^{n} \exp\{t_i Z_i\}\right] \quad \prod_{i\ \ 1}^{n} E[\exp\{t_i Z_i\}] \\
& \exp\left\{\frac{1}{2}\sum_{i\ \ 1}^{n} t_i^2\right\} \quad \exp\left\{\frac{1}{2}\mathbf{t}'\mathbf{t}\right\},
\end{aligned}
\tag{3.5.7}
$$

for all $\mathbf{t} \in R^n$. We say that $\mathbf{Z}$ has a **multivariate normal distribution** with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_n$. We abbreviate this by saying that $\mathbf{Z}$ has an $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution.

For the general case, suppose $\mathbf{\Sigma}$ is an $n \times n$, symmetric, and positive semi-definite matrix. Then from linear algebra, we can always decompose $\mathbf{\Sigma}$ as

$$
\mathbf{\Sigma} \quad \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}, \tag{3.5.8}
$$

where $\mathbf{\Lambda}$ is the diagonal matrix $\mathbf{\Lambda} \quad \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are the eigenvalues of $\mathbf{\Sigma}$, and the columns of $\mathbf{\Gamma}'$, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, are the corresponding eigenvectors. This decomposition is called the **spectral decomposition** of $\mathbf{\Sigma}$. The matrix $\mathbf{\Gamma}$ is orthogonal, i.e., $\mathbf{\Gamma}^{\ \ 1} \quad \mathbf{\Gamma}'$, and, hence, $\mathbf{\Gamma}\mathbf{\Gamma}' \quad \mathbf{I}$. As Exercise 3.5.19 shows, we can write the spectral decomposition in another way, as

$$
\mathbf{\Sigma} \quad \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} \quad \sum_{i\ \ 1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i'. \tag{3.5.9}
$$

Because the λ_is are nonnegative, we can define the diagonal matrix $\mathbf{\Lambda}^{1\ \ 2}$ $\operatorname{diag}\{\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}\}$. Then the orthogonality of $\mathbf{\Gamma}$ implies

$$
\mathbf{\Sigma} \quad [\mathbf{\Gamma}'\mathbf{\Lambda}^{1\ \ 2}\mathbf{\Gamma}][\mathbf{\Gamma}'\mathbf{\Lambda}^{1\ \ 2}\mathbf{\Gamma}].
$$

We define the matrix product in brackets as the **square root** of the positive semi-definite matrix $\mathbf{\Sigma}$ and write it as

$$
\mathbf{\Sigma}^{1\ \ 2} \quad \mathbf{\Gamma}'\mathbf{\Lambda}^{1\ \ 2}\mathbf{\Gamma}. \tag{3.5.10}
$$

Note that $\Sigma^{1\ 2}$ is symmetric and positive semi-definite. Suppose Σ is positive definite that is, all of its eigenvalues are strictly positive. Based on this, it is then easy to show that

$$\left(\Sigma^{1\ 2}\right)^{\ 1} \quad \Gamma'\Lambda^{\ 1\ 2}\Gamma \tag{3.5.11}$$

see Exercise 3.5.13. We write the left side of this equation as $\Sigma^{\ 1\ 2}$. These matrices enjoy many additional properties of the law of exponents for numbers see, for example, Arnold (1981). Here, though, all we need are the properties given above.

Suppose $\mathbf{Z}$ has a $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution. Let Σ be a positive semi-definite, symmetric matrix and let be an $n \times 1$ vector of constants. Define the random vector $\mathbf{X}$ by

$$\mathbf{X} \quad \Sigma^{1\ 2}\mathbf{Z} \quad . \tag{3.5.12}$$

By (3.5.6) and Theorem 2.6.3, we immediately have

$$E\,\mathbf{X} \qquad \text{and } \operatorname{Cov}\mathbf{X} \quad \Sigma^{1\ 2}\Sigma^{1\ 2} \quad \Sigma. \tag{3.5.13}$$

Further, the mgf of $\mathbf{X}$ is given by

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) \quad E\,\exp\{\mathbf{t}'\mathbf{X}\} \qquad & E\left[\exp\{\mathbf{t}'\Sigma^{1\ 2}\mathbf{Z} \quad \mathbf{t}' \quad \}\right] \\ & \exp\{\mathbf{t}' \quad \}E\left[\exp\left\{\left(\Sigma^{1\ 2}\mathbf{t}\right)'\mathbf{Z}\right\}\right] \\ & \exp\{\mathbf{t}' \quad \}\exp\left\{(1\ \ 2)\left(\Sigma^{1\ 2}\mathbf{t}\right)'\Sigma^{1\ 2}\mathbf{t}\right\} \\ & \exp\{\mathbf{t}' \quad \}\exp\{(1\ \ 2)\mathbf{t}'\Sigma\mathbf{t}\}. \end{aligned} \tag{3.5.14}$$

This leads to the following definition:

Definition 3.5.1 (Multivariate Normal). We say an n-dimensional random vector $\mathbf{X}$ has a **multivariate normal distribution** if its mgf is

$$M_{\mathbf{X}}(\mathbf{t}) \quad \exp\{\mathbf{t}' \qquad (1\ \ 2)\mathbf{t}'\Sigma\mathbf{t}\}, \quad \text{for all } \mathbf{t} \in R^n. \tag{3.5.15}$$

where Σ is a symmetric, positive semi-definite matrix and $\ \in R^n$. We abbreviate this by saying that $\mathbf{X}$ has a $N_n(\ ,\Sigma)$ distribution.

Note that our definition is for positive semi-definite matrices Σ. Usually Σ is positive definite, in which case we can further obtain the density of $\mathbf{X}$. If Σ is positive definite, then so is $\Sigma^{1\ 2}$ and, as discussed above, its inverse is given by expression (3.5.11). Thus the transformation between $\mathbf{X}$ and $\mathbf{Z}$, (3.5.12), is one-to-one with the inverse transformation

$$\mathbf{Z} \quad \Sigma^{\ 1\ 2}(\mathbf{X} - \quad)$$

and the Jacobian $|\Sigma^{\ 1\ 2}| \quad |\Sigma|^{\ 1\ 2}$. Hence, upon simplification, the pdf of $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) \quad \frac{1}{(2\pi)^{n\ 2}|\Sigma|^{1\ 2}}\exp\left\{-\frac{1}{2}(\mathbf{x} - \quad)'\Sigma^{\ 1}(\mathbf{x} - \quad)\right\}, \quad \text{for } \mathbf{x} \in R^n. \tag{3.5.16}$$

In Section 3.5.1, we discussed the contours of the bivariate normal distribution. We now extend that discussion to the general case, adding probabilities to the contours. Let $\mathbf{X}$ have a $N_n(\quad ,\Sigma)$ distribution. In the n-dimensional case, the contours of constant probability for the pdf of $\mathbf{X}$, (3.5.16), are the ellipsoids

$$(\mathbf{x} - \quad)'\Sigma^{\ \ 1}(\mathbf{x} - \quad) \quad c^2,$$

for $c \quad 0$. Define the random variable $Y \quad (\mathbf{X} - \quad)'\Sigma^{\ \ 1}(\mathbf{X} - \quad)$. Then using expression (3.5.12), we have

$$Y \quad \mathbf{Z}'\Sigma^{1 \ 2}\Sigma^{\ \ 1}\Sigma^{1 \ 2}\mathbf{Z} \quad \mathbf{Z}'\mathbf{Z} \quad \sum_{i \ \ 1}^{n} Z_i^2.$$

Since $Z_1, \ldots, Z_n$ are iid $N(0,1)$, Y has χ^2-distribution with n degrees of freedom. Denote the cdf of Y by $F_{\chi^2_n}$. Then we have

$$P\left[(\mathbf{X} - \quad)'\Sigma^{\ \ 1}(\mathbf{X} - \quad) \leq c^2\right] \quad P(Y \leq c^2) \quad F_{\chi^2_n}(c^2). \tag{3.5.17}$$

These probabilities are often used to label the contour plots see Exercise 3.5.5. For reference, we summarize the above proof in the following theorem. Note that this theorem is a generalization of the univariate result given in Theorem 3.4.1.

Theorem 3.5.1. *Suppose $\mathbf{X}$ has a $N_n(\quad ,\Sigma)$ distribution, where Σ is positive definite. Then the random variable $Y \quad (\mathbf{X} - \quad)'\Sigma^{\ \ 1}(\mathbf{X} - \quad)$ has a $\chi^2(n)$ distribution.*

The following two theorems are very useful. The first says that a linear transformation of a multivariate normal random vector has a multivariate normal distribution.

Theorem 3.5.2. *Suppose $\mathbf{X}$ has a $N_n(\quad ,\Sigma)$ distribution. Let $\mathbf{Y} \quad \mathbf{A}\mathbf{X} \quad \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b} \in R^m$. Then $\mathbf{Y}$ has a $N_m(\mathbf{A} \quad \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}')$ distribution.*

Proof: From (3.5.15), for $\mathbf{t} \in R^m$, the mgf of $\mathbf{Y}$ is

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) \quad & E\left[\exp\{\mathbf{t}'\mathbf{Y}\}\right] \\
& E\left[\exp\{\mathbf{t}'(\mathbf{A}\mathbf{X} \quad \mathbf{b})\}\right] \\
& \exp\{\mathbf{t}'\mathbf{b}\}E\left[\exp\{(\mathbf{A}'\mathbf{t})'\mathbf{X}\}\right] \\
& \exp\{\mathbf{t}'\mathbf{b}\}\exp\{(\mathbf{A}'\mathbf{t})' \qquad (1 \ 2)(\mathbf{A}'\mathbf{t})'\Sigma(\mathbf{A}'\mathbf{t})\} \\
& \exp\{\mathbf{t}'(\mathbf{A} \qquad \mathbf{b}) \quad (1 \ 2)\mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}\},
\end{aligned}$$

which is the mgf of an $N_m(\mathbf{A} \quad \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}')$ distribution. ■

A simple corollary to this theorem gives marginal distributions of a multivariate normal random variable. Let $\mathbf{X}_1$ be any subvector of $\mathbf{X}$, say of dimension m n. Because we can always rearrange means and correlations, there is no loss in generality in writing $\mathbf{X}$ as

$$\mathbf{X} \quad \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \tag{3.5.18}$$

where X_2 is of dimension $p \quad n - m$. In the same way, partition the mean and covariance matrix of X that is,

$$\left[\begin{array}{c} {}_1 \\ {}_2 \end{array}\right] \text{ and } \Sigma \quad \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right] \tag{3.5.19}$$

with the same dimensions as in expression (3.5.18). Note, for instance, that Σ_{11} is the covariance matrix of X_1 and Σ_{12} contains all the covariances between the components of X_1 and X_2. Now define A to be the matrix

$$A \quad I_m \vdots O_{mp},$$

where O_{mp} is an $m \times p$ matrix of zeroes. Then $X_1 \quad AX$. Hence, applying Theorem 3.5.2 to this transformation, along with some matrix algebra, we have the following corollary:

Corollary 3.5.1. Suppose X has a $N_n(\ ,\Sigma)$ distribution, partitioned as in expressions (3.5.18) and (3.5.19). Then X_1 has a $N_m(\ _1, \Sigma_{11})$ distribution.

This is a useful result because it says that any marginal distribution of X is also normal and, further, its mean and covariance matrix are those associated with that partial vector.

Recall in Section 2.5, Theorem 2.5.2, that if two random variables are independent then their covariance is 0. In general, the converse is not true. However, as the following theorem shows, it is true for the multivariate normal distribution.

Theorem 3.5.3. Suppose X has a $N_n(\ ,\Sigma)$ distribution, partitioned as in the expressions (3.5.18) and (3.5.19). Then X_1 and X_2 are independent if and only if $\Sigma_{12} \quad O$.

Proof: First note that $\Sigma_{21} \quad \Sigma'_{12}$. The joint mgf of X_1 and X_2 is given by

$$M_{X_1,X_2}(t_1,t_2) \quad \exp\left\{t'_1 \ {}_1 \quad t'_2 \ {}_2 \quad \frac{1}{2}(t'_1\Sigma_{11}t_1 \quad t'_2\Sigma_{22}t_2 \quad t'_2\Sigma_{21}t_1 \quad t'_1\Sigma_{12}t_2)\right\} \tag{3.5.20}$$

where $t' \quad (t'_1, t'_2)$ is partitioned the same as . By Corollary 3.5.1, X_1 has a $N_m(\ _1, \Sigma_{11})$ distribution and X_2 has a $N_p(\ _2, \Sigma_{22})$ distribution. Hence, the product of their marginal mgfs is

$$M_{X_1}(t_1)M_{X_2}(t_2) \quad \exp\left\{t'_1 \ {}_1 \quad t'_2 \ {}_2 \quad \frac{1}{2}(t'_1\Sigma_{11}t_1 \quad t'_2\Sigma_{22}t_2)\right\}. \tag{3.5.21}$$

By (2.6.6) of Section 2.6, X_1 and X_2 are independent if and only if the expressions (3.5.20) and (3.5.21) are the same. If $\Sigma_{12} \quad O'$ and, hence, $\Sigma_{21} \quad O$, then the expressions are the same and X_1 and X_2 are independent. If X_1 and X_2 are independent, then the covariances between their components are all 0 i.e., $\Sigma_{12} \quad O'$ and $\Sigma_{21} \quad O$. ∎

Corollary 3.5.1 showed that the marginal distributions of a multivariate normal are themselves normal. This is true for conditional distributions, too. As the following proof shows, we can combine the results of Theorems 3.5.2 and 3.5.3 to obtain the following theorem.

Theorem 3.5.4. Suppose $\mathbf{X}$ has a $N_n(\ \ ,\Sigma)$ distribution, which is partitioned as in expressions (3.5.18) and (3.5.19). Assume that Σ is positive definite. Then the conditional distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ is

$$N_m(\ _1\quad \Sigma_{12}\Sigma_{22}^{\ 1}(\mathbf{X}_2 -\quad _2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{\ 1}\Sigma_{21}). \tag{3.5.22}$$

Proof: Consider first the joint distribution of the random vector $\mathbf{W}\quad \mathbf{X}_1 - \Sigma_{12}\Sigma_{22}^{\ 1}\mathbf{X}_2$ and $\mathbf{X}_2$. This distribution is obtained from the transformation

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix} \quad \begin{bmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{\ 1} \\ \mathbf{O} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

Because this is a linear transformation, it follows from Theorem 3.5.2 that the joint distribution is multivariate normal, with $E\,\mathbf{W}\qquad _1 - \Sigma_{12}\Sigma_{22}^{\ 1}\ \ _2$, $E\,\mathbf{X}_2\qquad _2$, and covariance matrix

$$\begin{bmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{\ 1} \\ \mathbf{O} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{O}' \\ -\Sigma_{22}^{\ 1}\Sigma_{21} & \mathbf{I}_p \end{bmatrix}$$
$$\begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{\ 1}\Sigma_{21} & \mathbf{O}' \\ \mathbf{O} & \Sigma_{22} \end{bmatrix}.$$

Hence, by Theorem 3.5.3 the random vectors $\mathbf{W}$ and $\mathbf{X}_2$ are independent. Thus the conditional distribution of $\mathbf{W} \mid \mathbf{X}_2$ is the same as the marginal distribution of $\mathbf{W}$ that is,

$$\mathbf{W} \mid \mathbf{X}_2 \text{ is } N_m(\ _1 - \Sigma_{12}\Sigma_{22}^{\ 1}\ \ _2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{\ 1}\Sigma_{21}).$$

Further, because of this independence, $\mathbf{W}\quad \Sigma_{12}\Sigma_{22}^{\ 1}\mathbf{X}_2$ given $\mathbf{X}_2$ is distributed as

$$N_m(\ _1 - \Sigma_{12}\Sigma_{22}^{\ 1}\ \ _2\quad \Sigma_{12}\Sigma_{22}^{\ 1}\mathbf{X}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{\ 1}\Sigma_{21}), \tag{3.5.23}$$

which is the desired result. ∎

In the following remark, we return to the bivariate normal using the above general notation.

Remark 3.5.1 (Continuation of the Bivariate Normal). Suppose (X, Y) has a $N_2(\ \ ,\Sigma)$ distribution, where

$$\begin{bmatrix} \ _1 \\ \ _2 \end{bmatrix} \text{ and } \Sigma\quad \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \tag{3.5.24}$$

Substituting $\rho\sigma_1\sigma_2$ for σ_{12} in Σ, it is easy to see that the determinant of Σ is $\sigma_1^2\sigma_2^2(1-\rho^2)$. Recall that $\rho^2 \le 1$. For the remainder of this remark, assume that $\rho^2\quad 1$. In this case, Σ is invertible (it is also positive definite). Further, since Σ is a 2×2 matrix, its inverse can easily be determined to be

$$\Sigma^{\ 1}\quad \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}. \tag{3.5.25}$$

This shows the equivalence of the bivariate normal pdf notation, (3.5.1), and the general multivariate normal distribution with n 2 pdf notation, (3.5.16).

To simplify the conditional normal distribution (3.5.22) for the bivariate case, consider once more the bivariate normal distribution that was given in Section 3.5.1. For this case, reversing the roles so that Y X_1 and X X_2, expression (3.5.22) shows that the conditional distribution of Y given X x is

$$N\left[\ {}_2\quad \rho\frac{\sigma_2}{\sigma_1}(x - \ {}_1), \sigma_2^2(1-\rho^2)\right]. \tag{3.5.26}$$

Thus, with a bivariate normal distribution, the conditional mean of Y, given that X x, is linear in x and is given by

$$E(Y|x) \quad {}_2\quad \rho\frac{\sigma_2}{\sigma_1}(x - \ {}_1).$$

Although the mean of the conditional distribution of Y, given X x, depends upon x (unless ρ 0), the variance $\sigma_2^2(1-\rho^2)$ is the same for all real values of x. Thus, by way of example, given that X x, the conditional probability that Y is within $(2.576)\sigma_2\sqrt{1-\rho^2}$ units of the conditional mean is 0.99, whatever the value of x may be. In this sense, most of the probability for the distribution of X and Y lies in the band

$${}_2\quad \rho\frac{\sigma_2}{\sigma_1}(x - \ {}_1) \pm 2.576\sigma_2\sqrt{1-\rho^2}$$

about the graph of the linear conditional mean. For every fixed positive σ_2, the width of this band depends upon ρ. Because the band is narrow when ρ^2 is nearly 1, we see that ρ does measure the intensity of the concentration of the probability for X and Y about the linear conditional mean. We alluded to this fact in the remark of Section 2.5.

In a similar manner we can show that the conditional distribution of X, given Y y, is the normal distribution

$$N\left[\ {}_1\quad \rho\frac{\sigma_1}{\sigma_2}(y - \ {}_2),\ \sigma_1^2(1-\rho^2)\right]. \ \blacksquare$$

Example 3.5.1. Let us assume that in a certain population of married couples the height X_1 of the husband and the height X_2 of the wife have a bivariate normal distribution with parameters ${}_1$ 5.8 feet, ${}_2$ 5.3 feet, σ_1 σ_2 0.2 foot, and ρ 0.6. The conditional pdf of X_2, given X_1 6.3, is normal, with mean 5.3 (0.6)(6.3 − 5.8) 5.6 and standard deviation $(0.2)\sqrt{(1-0.36)}$ 0.16. Accordingly, given that the height of the husband is 6.3 feet, the probability that his wife has a height between 5.28 and 5.92 feet is

$$P(5.28 \quad X_2 \quad 5.92|X_1 \quad 6.3) = \Phi(2) - \Phi(-2) \quad 0.954.$$

The interval (5.28, 5.92) could be thought of as a 95.4 prediction interval for the wife's height, given X_1 6.3. ■