5.12 The Bivariate Normal Distribution

The first multivariate continuous distribution for which we have a name is a generalization of the normal distribution to two coordinates. There is more structure to the bivariate normal distribution than just a pair of normal marginal distributions.

Definition of the Bivariate Normal Distribution

Suppose that Z_1 and Z_2 are independent random variables, each of which has a standard normal distribution. Then the joint p.d.f. $g(z_1, z_2)$ of Z_1 and Z_2 is specified for all values of z_1 and z_2 by the equation

$$g(z_1, z_2) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(z_1^2 + z_2^2)\right]. \tag{5.12.1}$$

For constants μ_1, μ_2, σ_1, σ_2, and ρ such that $-\infty < \mu_i < \infty$ $(i = 1, 2)$, $\sigma_i > 0$ $(i = 1, 2)$, and $-1 < \rho < 1$, we shall now define two new random variables X_1 and X_2 as follows:

$$\begin{aligned} X_1 &= \sigma_1 Z_1 + \mu_1, \\ X_2 &= \sigma_2 \left[\rho Z_1 + (1-\rho^2)^{1/2} Z_2\right] + \mu_2. \end{aligned} \tag{5.12.2}$$

We shall derive the joint p.d.f. $f(x_1, x_2)$ of X_1 and X_2.

The transformation from Z_1 and Z_2 to X_1 and X_2 is a linear transformation; and it will be found that the determinant Δ of the matrix of coefficients of Z_1 and Z_2 has the value $\Delta = (1-\rho^2)^{1/2}\sigma_1\sigma_2$. Therefore, as discussed in Section 3.9, the Jacobian J of the inverse transformation from X_1 and X_2 to Z_1 and Z_2 is

$$J = \frac{1}{\Delta} = \frac{1}{(1-\rho^2)^{1/2}\sigma_1\sigma_2}. \tag{5.12.3}$$

Since $J > 0$, the value of $|J|$ is equal to the value of J itself. If the relations (5.12.2) are solved for Z_1 and Z_2 in terms of X_1 and X_2, then the joint p.d.f. $f(x_1, x_2)$ can be obtained by replacing z_1 and z_2 in Eq. (5.12.1) by their expressions in terms of x_1 and x_2, and then multiplying by $|J|$. It can be shown that the result is, for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$,

$$\begin{aligned} f(x_1, x_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_1\sigma_2} \exp\Bigg\{ -\frac{1}{2(1-\rho^2)} \Bigg[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \\ - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \Bigg] \Bigg\}. \end{aligned} \tag{5.12.4}$$

When the joint p.d.f. of two random variables X_1 and X_2 is of the form in Eq. (5.12.4), it is said that X_1 and X_2 have a *bivariate normal distribution*. The means and the variances of the bivariate normal distribution specified by Eq. (5.12.4) are easily derived from the definitions in Eq. (5.12.2). Because Z_1 and Z_2 are independent and each has mean 0 and variance 1, it follows that $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $\text{Var}(X_1) = \sigma_1^2$, and $\text{Var}(X_2) = \sigma_2^2$. Furthermore, it can be shown by using Eq. (5.12.2) that $\text{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$. Therefore, the correlation of X_1 and X_2 is simply ρ. In summary, if X_1 and X_2 have a bivariate normal distribution for which the p.d.f. is specified by Eq. (5.12.4), then

$$E(X_i) = \mu_i \quad \text{and} \quad \text{Var}(X_i) = \sigma_i^2 \quad \text{for } i = 1, 2.$$

Also,

$$\rho(X_1, X_2) = \rho.$$

It has been convenient for us to introduce the bivariate normal distribution as the joint distribution of certain linear combinations of independent random variables having standard normal distributions. It should be emphasized, however, that the bivariate normal distribution arises directly and naturally in many practical problems. For example, for many populations the joint distribution of two physical characteristics such as the heights and the weights of the individuals in the population will be approximately a bivariate normal distribution. For other populations, the joint distribution of the scores of the individuals in the population on two related tests will be approximately a bivariate normal distribution.

Example 5.12.1 **Anthropometry of Flea Beetles.** Lubischew (1962) reports the measurements of several physical features of a variety of species of flea beetle. The investigation was concerned with whether some combination of easily obtained measurements could be used to distinguish the different species. Figure 5.8 shows a scatterplot of measurements of the first joint in the first tarsus versus the second joint in the first tarsus for a sample of 31 from the species *Chaetocnema heikertingeri*. The plot also includes three ellipses that correspond to a fitted bivariate normal distribution. The ellipses were chosen to contain 25%, 50%, and 75% of the probability of the fitted bivariate normal distribution. The correlation of the fitted distribution is 0.64. ◀

Marginal and Conditional Distributions

Marginal Distributions. We shall continue to assume that the random variables X_1 and X_2 have a bivariate normal distribution, and their joint p.d.f. is specified by Eq. (5.12.4). In the study of the properties of this distribution, it will be convenient to represent X_1 and X_2 as in Eq. (5.12.2), where Z_1 and Z_2 are independent random variables with standard normal distributions. In particular, since both X_1 and X_2 are linear combinations of Z_1 and Z_2, it follows from this representation and from Corollary 5.6.1 that the marginal distributions of both X_1 and X_2 are also normal distributions. Thus, for $i = 1, 2$, the marginal distribution of X_i is a normal distribution with mean μ_i and variance σ_i^2.

Independence and Correlation. If X_1 and X_2 are uncorrelated, then $\rho = 0$. In this case, it can be seen from Eq. (5.12.4) that the joint p.d.f. $f(x_1, x_2)$ factors into the product of the marginal p.d.f. of X_1 and the marginal p.d.f. of X_2. Hence, X_1 and X_2 are independent, and the following result has been established:

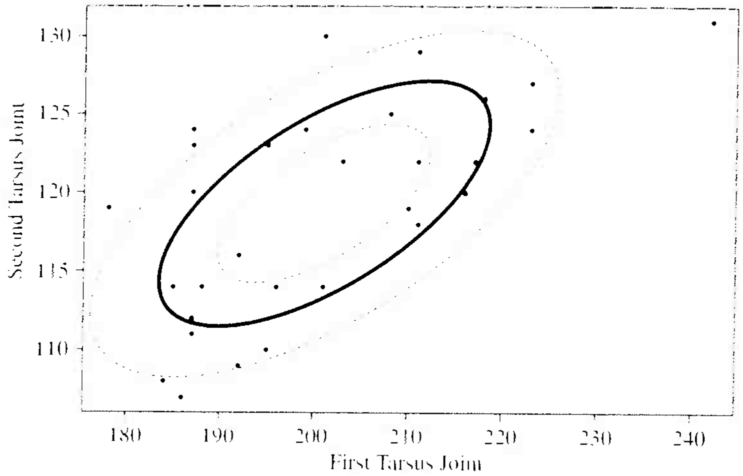

Figure 5.8 Scatterplot of flea beetle data with 25%, 50%, and 75% bivariate normal ellipses for Example 5.12.1.

Two random variables X_1 and X_2 that have a bivariate normal distribution are independent if and only if they are uncorrelated.

We have already seen in Section 4.6 that two random variables X_1 and X_2 with an arbitrary joint distribution can be uncorrelated without being independent.

Conditional Distributions. The conditional distribution of X_2 given that $X_1 = x_1$ can also be derived from the representation in Eq. (5.12.2). If $X_1 = x_1$, then $Z_1 = (x_1 - \mu_1)/\sigma_1$. Therefore, the conditional distribution of X_2 given that $X_1 = x_1$ is the same as the conditional distribution of

$$(1-\rho^2)^{1/2}\sigma_2 Z_2 + \mu_2 + \rho\sigma_2\left(\frac{x_1-\mu_1}{\sigma_1}\right). \tag{5.12.5}$$

Because Z_2 has a standard normal distribution and is independent of X_1, it follows from (5.12.5) that the conditional distribution of X_2 given that $X_1 = x_1$ is a normal distribution, for which the mean is

$$E(X_2|x_1) = \mu_2 + \rho\sigma_2\left(\frac{x_1-\mu_1}{\sigma_1}\right). \tag{5.12.6}$$

and the variance is $(1-\rho^2)\sigma_2^2$.

The conditional distribution of X_1 given that $X_2 = x_2$ cannot be derived so easily from Eq. (5.12.2) because of the different ways in which Z_1 and Z_2 enter Eq. (5.12.2). However, it is seen from Eq. (5.12.4) that the joint p.d.f. $f(x_1, x_2)$ is symmetric in the two variables $(x_1 - \mu_1)/\sigma_1$ and $(x_2 - \mu_2)/\sigma_2$. Therefore, it follows that the conditional distribution of X_1 given that $X_2 = x_2$ can be found from the conditional distribution of X_2 given that $X_1 = x_1$ (this distribution has just been derived) simply by interchanging

x_1 and x_2, interchanging μ_1 and μ_2, and interchanging σ_1 and σ_2. Thus, the conditional distribution of X_1 given that $X_2 = x_2$ must be a normal distribution, for which the mean is

$$E(X_1|x_2) = \mu_1 + \rho\sigma_1\left(\frac{x_2 - \mu_2}{\sigma_2}\right), \tag{5.12.7}$$

and the variance is $(1 - \rho^2)\sigma_1^2$.

We have now shown that each marginal distribution and each conditional distribution of a bivariate normal distribution is a univariate normal distribution.

Some particular features of the conditional distribution of X_2 given that $X_1 = x_1$ should be noted. If $\rho \neq 0$, then $E(X_2|x_1)$ is a linear function of the given value x_1. If $\rho > 0$, the slope of this linear function is positive. If $\rho < 0$, the slope of the function is negative. However, the variance of the conditional distribution of X_2 given that $X_1 = x_1$ is $(1 - \rho^2)\sigma_2^2$, and its value does not depend on the given value x_1. Furthermore, this variance of the conditional distribution of X_2 is smaller than the variance σ_2^2 of the marginal distribution of X_2.

Example 5.12.2 **Predicting a Person's Weight.** Let X_1 denote the height of a person selected at random from a certain population, and let X_2 denote the weight of the person. Suppose that these random variables have a bivariate normal distribution for which the p.d.f. is specified by Eq. (5.12.4) and that the person's weight X_2 must be predicted. We shall compare the smallest M.S.E. that can be attained if the person's height X_1 is known when her weight must be predicted with the smallest M.S.E. that can be attained if her height is not known.

If the person's height is not known, then the best prediction of her weight is the mean $E(X_2) = \mu_2$; and the M.S.E. of this prediction is the variance σ_2^2. If it is known that the person's height is x_1, then the best prediction is the mean $E(X_2|x_1)$ of the conditional distribution of X_2 given that $X_1 = x_1$; and the M.S.E. of this prediction is the variance $(1 - \rho^2)\sigma_2^2$ of that conditional distribution. Hence, when the value of X_1 is known, the M.S.E. is reduced from σ_2^2 to $(1 - \rho^2)\sigma_2^2$. ◀

Since the variance of the conditional distribution in Example 5.12.2 is $(1 - \rho^2)\sigma_2^2$, regardless of the known height x_1 of the person, it follows that the difficulty of predicting the person's weight is the same for a tall person, a short person, or a person of medium height. Furthermore, since the variance $(1 - \rho^2)\sigma_2^2$ decreases as $|\rho|$ increases, it follows that it is easier to predict a person's weight from her height when the person is selected from a population in which height and weight are highly correlated.

Example 5.12.3 **Determining a Marginal Distribution.** Suppose that a random variable X has a normal distribution with mean μ and variance σ^2, and that for every number x, the conditional distribution of another random variable Y given that $X = x$ is a normal distribution with mean x and variance τ^2. We shall determine the marginal distribution of Y.

We know that the marginal distribution of X is a normal distribution, and the conditional distribution of Y given that $X = x$ is a normal distribution, for which the mean is a linear function of x and the variance is constant. It follows that the joint distribution of

X and Y must be a bivariate normal distribution (see Exercise 14). Hence, the marginal distribution of Y is also a normal distribution. The mean and the variance of Y must be determined.

The mean of Y is

$$E(Y) = E[E(Y|X)] = E(X) = \mu.$$

Furthermore, by Exercise 11 of Section 4.7,

$$\begin{aligned}\text{Var}(Y) &= E[\text{Var}(Y|X)] + \text{Var}[E(Y|X)] \\ &= E(\tau^2) + \text{Var}(X) \\ &= \tau^2 + \sigma^2.\end{aligned}$$

Hence, the distribution of Y is a normal distribution with mean μ and variance $\tau^2 + \sigma^2$. ◀

Linear Combinations

Suppose again that two random variables X_1 and X_2 have a bivariate normal distribution, for which the p.d.f. is specified by Eq. (5.12.4). Now consider the random variable $Y = a_1X_1 + a_2X_2 + b$, where a_1, a_2, and b are arbitrary given constants. Both X_1 and X_2 can be represented, as in Eq. (5.12.2), as linear combinations of independent and normally distributed random variables Z_1 and Z_2. Since Y is a linear combination of X_1 and X_2, it follows that Y can also be represented as a linear combination of Z_1 and Z_2. Therefore, by Corollary 5.6.1, the distribution of Y will also be a normal distribution. Thus, the following important property has been established.

If two random variables X_1 and X_2 have a bivariate normal distribution, then each linear combination $Y = a_1X_1 + a_2X_2 + b$ will have a normal distribution.

The mean and variance of Y are as follows:

$$\begin{aligned}E(Y) &= a_1E(X_1) + a_2E(X_2) + b \\ &= a_1\mu_1 + a_2\mu_2 + b\end{aligned}$$

and

$$\begin{aligned}\text{Var}(Y) &= a_1^2\,\text{Var}(X_1) + a_2^2\,\text{Var}(X_2) + 2a_1a_2\,\text{Cov}(X_1, X_2) \\ &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2. \qquad (5.12.8)\end{aligned}$$

Example 5.12.4 **Heights of Husbands and Wives.** Suppose that a married couple is selected at random from a certain population of married couples, and that the joint distribution of the height of the wife and the height of her husband is a bivariate normal distribution. Suppose that the heights of the wives have a mean of 66.8 inches and a standard deviation of 2 inches, the heights of the husbands have a mean of 70 inches and a standard deviation of 2 inches, and the correlation between these two heights is 0.68. We shall determine the probability that the wife will be taller than her husband.

If we let X denote the height of the wife, and let Y denote the height of her husband, then we must determine the value of $\Pr(X - Y > 0)$. Since X and Y have a bivariate normal distribution, it follows that the distribution of $X - Y$ will be a normal distribution, for which the mean is

$$E(X - Y) = 66.8 - 70 = -3.2$$

and the variance is

$$\begin{aligned}\text{Var}(X - Y) &= \text{Var}(X) + \text{Var}(Y) - 2\,\text{Cov}(X, Y)\\ &= 4 + 4 - 2(0.68)(2)(2) = 2.56.\end{aligned}$$

Hence, the standard deviation of $X - Y$ is 1.6.

The random variable $Z = (X - Y + 3.2)/(1.6)$ will have a standard normal distribution. It can be found from the table given at the end of this book that

$$\begin{aligned}\Pr(X - Y > 0) &= \Pr(Z > 2) = 1 - \Phi(2)\\ &= 0.0227.\end{aligned}$$

Therefore, the probability that the wife will be taller than her husband is 0.0227. ◂

Summary

If a random vector (X, Y) has a bivariate normal distribution, then every linear combination $aX + bY + c$ has a normal distribution. In particular, the marginal distributions of X and Y are normal. Also, the conditional distribution of X given $Y = y$ is normal with the conditional mean being a linear function of y and the conditional variance being constant in y. (Similarly, for the conditional distribution of Y given $X = x$.) A more thorough treatment of the bivariate normal distribution and higher-dimensional generalizations can be found in the book by D. F. Morrison (1990).

EXERCISES

1. Consider again the joint distribution of heights of husbands and wives in Example 5.12.4. Find the 0.95 quantile of the conditional distribution of the height of the wife given that the height of the husband is 72 inches.

2. Suppose that two different tests A and B are to be given to a student chosen at random from a certain population. Suppose also that the mean score on test A is 85, and the standard deviation is 10; the mean score on test B is 90, and the standard deviation is 16; the scores on the two tests have a bivariate normal distribution; and the correlation of the two scores is 0.8. If the student's score on test A is 80, what is the probability that her score on test B will be higher than 90?

3. Consider again the two tests A and B described in Exercise 2. If a student is chosen at random, what is the probability that the sum of her scores on the two tests will be greater than 200?