$$= \frac{4}{\pi} - 1$$
$$= 0.2732$$

and it follows that

$$\sigma^2 = 0.2732 - (0.4413)^2 = 0.0785$$

and $\sigma = \sqrt{0.0785} = 0.2802$.

The following is another theorem that is of importance in work connected with standard deviations or variances.

THEOREM 7. If X has the variance σ^2, then

$$\text{var}(aX + b) = a^2\sigma^2$$

The proof of this theorem will be left to the reader, but let us point out the following corollaries: For $a = 1$, we find that the addition of a constant to the values of a random variable, resulting in a shift of all the values of X to the left or to the right, in no way affects the spread of its distribution; for $b = 0$, we find that if the values of a random variable are multiplied by a constant, the variance is multiplied by the square of that constant, resulting in a corresponding change in the spread of the distribution.

4 Chebyshev's Theorem

To demonstrate how σ or σ^2 is indicative of the spread or dispersion of the distribution of a random variable, let us now prove the following theorem, called **Chebyshev's theorem** after the nineteenth-century Russian mathematician P. L. Chebyshev. We shall prove it here only for the continuous case, leaving the discrete case as an exercise.

THEOREM 8. (*Chebyshev's Theorem*) If μ and σ are the mean and the standard deviation of a random variable X, then for any positive constant k the probability is *at least* $1 - \frac{1}{k^2}$ that X will take on a value within k standard deviations of the mean; symbolically,

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \quad \sigma \neq 0$$

Proof According to Definitions 4 and 5, we write

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)\, dx$$

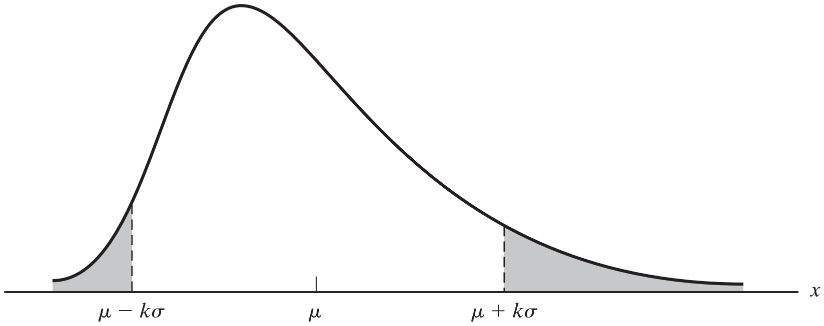

Figure 2. Diagram for proof of Chebyshev's theorem.

Then, dividing the integral into three parts as shown in Figure 2, we get

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 \cdot f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 \cdot f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$$

Since the integrand $(x-\mu)^2 \cdot f(x)$ is nonnegative, we can form the inequality

$$\sigma^2 \geqq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 \cdot f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$$

by deleting the second integral. Therefore, since $(x-\mu)^2 \geqq k^2\sigma^2$ for $x \leqq \mu - k\sigma$ or $x \geqq \mu + k\sigma$ it follows that

$$\sigma^2 \geqq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 \cdot f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 \cdot f(x)\,dx$$

and hence that

$$\frac{1}{k^2} \geqq \int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx$$

provided $\sigma^2 \neq 0$. Since the sum of the two integrals on the right-hand side is the probability that X will take on a value less than or equal to $\mu - k\sigma$ or greater than or equal to $\mu + k\sigma$, we have thus shown that

$$P(|X-\mu| \geqq k\sigma) \leqq \frac{1}{k^2}$$

and it follows that

$$P(|X-\mu| < k\sigma) \geqq 1 - \frac{1}{k^2}$$

For instance, the probability is at least $1 - \frac{1}{2^2} = \frac{3}{4}$ that a random variable X will take on a value within two standard deviations of the mean, the probability is at least $1 - \frac{1}{3^2} = \frac{8}{9}$ that it will take on a value within three standard deviations of the mean, and the probability is at least $1 - \frac{1}{5^2} = \frac{24}{25}$ that it will take on a value within five standard deviations of the mean. It is in this sense that σ controls the spread or dispersion of the distribution of a random variable. Clearly, the probability given by Chebyshev's theorem is only a lower bound; whether the probability that a given random variable will take on a value within k standard deviations of the mean is actually greater than $1 - \frac{1}{k^2}$ and, if so, by how much we cannot say, but Chebyshev's theorem assures us that this probability cannot be less than $1 - \frac{1}{k^2}$. Only when the distribution of a random variable is known can we calculate the exact probability.

EXAMPLE 12

If the probability density of X is given by

$$f(x) = \begin{cases} 630x^4(1-x)^4 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability that it will take on a value within two standard deviations of the mean and compare this probability with the lower bound provided by Chebyshev's theorem.

Solution

Straightforward integration shows that $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{44}$, so that $\sigma = \sqrt{1/44}$ or approximately 0.15. Thus, the probability that X will take on a value within two standard deviations of the mean is the probability that it will take on a value between 0.20 and 0.80, that is,

$$\begin{aligned} P(0.20 < X < 0.80) &= \int_{0.20}^{0.80} 630x^4(1-x)^4\, dx \\ &= 0.96 \end{aligned}$$

Observe that the statement "the probability is 0.96" is a much stronger statement than "the probability is at least 0.75," which is provided by Chebyshev's theorem.

---

5 Moment-Generating Functions

Although the moments of most distributions can be determined directly by evaluating the necessary integrals or sums, an alternative procedure sometimes provides considerable simplifications. This technique utilizes **moment-generating functions**.

DEFINITION 6. MOMENT GENERATING FUNCTION. *The **moment generating function** of a random variable* X*, where it exists, is given by*

$$M_X(t) = E(e^{tX}) = \sum_x e^{tX} \cdot f(x)$$

when X *is discrete, and*

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)dx$$

when X *is continuous.*

Exercises

41. Prove that $\text{cov}(X, Y) = \text{cov}(Y, X)$ for both discrete and continuous random variables X and Y.

42. If X and Y have the joint probability distribution $f(x, y) = \frac{1}{4}$ for $x = -3$ and $y = -5, x = -1$ and $y = -1, x = 1$ and $y = 1$, and $x = 3$ and $y = 5$, find $\text{cov}(X, Y)$.

43. This has been intentionally omitted for this edition.

44. This has been intentionally omitted for this edition.

45. This has been intentionally omitted for this edition.

46. If X and Y have the joint probability distribution $f(-1, 0) = 0, f(-1, 1) = \frac{1}{4}, f(0, 0) = \frac{1}{6}, f(0, 1) = 0, f(1, 0) = \frac{1}{12}$, and $f(1, 1) = \frac{1}{2}$, show that

(a) $\text{cov}(X, Y) = 0$;

(b) the two random variables are not independent.

47. If the probability density of X is given by

$$f(x) = \begin{cases} 1+x & \text{for } -1 < x \leqq 0 \\ 1-x & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and $U = X$ and $V = X^2$, show that

(a) $\text{cov}(U, V) = 0$;

(b) U and V are dependent.

48. For k random variables $X_1, X_2, \ldots, X_k$, the values of their **joint moment-generating function** are given by

$$E\left(e^{t_1X_1+t_2X_2+\cdots+t_kX_k}\right)$$

(a) Show for either the discrete case or the continuous case that the partial derivative of the joint moment-generating function with respect to t_i at $t_1 = t_2 = \cdots = t_k = 0$ is $E(X_i)$.

(b) Show for either the discrete case or the continuous case that the second partial derivative of the joint moment-generating function with respect to t_i and $t_j, i \neq j$, at $t_1 = t_2 = \cdots = t_k = 0$ is $E(X_iX_j)$.

(c) If two random variables have the joint density given by

$$f(x, y) = \begin{cases} e^{-x-y} & \text{for } x > 0,\ y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find their joint moment-generating function and use it to determine the values of $E(XY)$, $E(X)$, $E(Y)$, and $\text{cov}(X, Y)$.

49. If X_1, X_2, and X_3 are independent and have the means 4, 9, and 3 and the variances 3, 7, and 5, find the mean and the variance of

(a) $Y = 2X_1 - 3X_2 + 4X_3$;

(b) $Z = X_1 + 2X_2 - X_3$.

50. Repeat both parts of Exercise 49, dropping the assumption of independence and using instead the information that $\text{cov}(X_1, X_2) = 1, \text{cov}(X_2, X_3) = -2$, and $\text{cov}(X_1, X_3) = -3$.

51. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{3}(x+y) & \text{for } 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find the variance of $W = 3X + 4Y - 5$.

52. Prove Theorem 15.

53. Express $\text{var}(X+Y), \text{var}(X-Y)$, and $\text{cov}(X+Y, X-Y)$ in terms of the variances and covariance of X and Y.

54. If $\text{var}(X_1) = 5, \text{var}(X_2) = 4, \text{var}(X_3) = 7, \text{cov}(X_1, X_2) = 3, \text{cov}(X_1, X_3) = -2$, and X_2 and X_3 are independent, find the covariance of $Y_1 = X_1 - 2X_2 + 3X_3$ and $Y_2 = -2X_1 + 3X_2 + 4X_3$.

55. With reference to Exercise 49, find $\text{cov}(Y, Z)$.

56. This question has been intentionally omitted for this edition.

57. This question has been intentionally omitted for this edition.

58. This question has been intentionally omitted for this edition.

59. This question has been intentionally omitted for this edition.

60. (a) Show that the conditional distribution function of the continuous random variable X, given $a < X \leqq b$, is given by

$$F(x|a < X \leqq b) = \begin{cases} 0 & \text{for } x \leqq a \\ \dfrac{F(x) - F(a)}{F(b) - F(a)} & \text{for } a < x \leqq b \\ 1 & \text{for } x > b \end{cases}$$

(b) Differentiate the result of part (a) with respect to x to find the conditional probability density of X given $a < X \leqq b$, and show that

$$E[u(X)|a < X \leqq b] = \frac{\displaystyle\int_a^b u(x)f(x)\,dx}{\displaystyle\int_a^b f(x)\,dx}$$

40. With reference to Exercise 39, show that for normal distributions $\kappa_2 = \sigma^2$ and all other cumulants are zero.

41. Show that if X is a random variable having the Poisson distribution with the parameter λ and $\lambda \to \infty$, then the moment-generating function of

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

that is, that of a standardized Poisson random variable, approaches the moment-generating function of the standard normal distribution.

42. Show that when $\alpha \to \infty$ and β remains constant, the moment-generating function of a standardized gamma random variable approaches the moment-generating function of the standard normal distribution.

7 The Bivariate Normal Distribution

Among multivariate densities, of special importance is the **multivariate normal distribution**, which is a generalization of the normal distribution in one variable. As it is best (indeed, virtually necessary) to present this distribution in matrix notation, we shall give here only the **bivariate** case; discussions of the general case are listed among the references at the end of this chapter.

DEFINITION 8. BIVARIATE NORMAL DISTRIBUTION. *A pair of random variables X and Y have a* ***bivariate normal distribution*** *and they are referred to as jointly normally distributed random variables if and only if their joint probability density is given by*

$$f(x, y) = \frac{e^{-\frac{1}{2(1-\rho)^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

for $-\infty < x < \infty$ *and* $-\infty < y < \infty$, *where* $\sigma_1 > 0$, $\sigma_2 > 0$, *and* $-1 < \rho < 1$.

To study this joint distribution, let us first show that the parameters μ_1, μ_2, σ_1, and σ_2 are, respectively, the means and the standard deviations of the two random variables X and Y. To begin with, we integrate on y from $-\infty$ to ∞, getting

$$g(x) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]} dy$$

for the marginal density of X. Then, temporarily making the substitution $u = \frac{x-\mu_1}{\sigma_1}$ to simplify the notation and changing the variable of integration by letting $v = \frac{y-\mu_2}{\sigma_2}$, we obtain

$$g(x) = \frac{e^{-\frac{1}{2(1-\rho^2)}\mu^2}}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}(v^2 - 2\rho uv)} dv$$

After completing the square by letting

$$v^2 - 2\rho uv = (v - \rho u)^2 - \rho^2 u^2$$

and collecting terms, this becomes

$$g(x) = \frac{e^{-\frac{1}{2}u^2}}{\sigma_1\sqrt{2\pi}} \left\{ \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{v-\rho u}{\sqrt{1-\rho^2}}\right)^2} dv \right\}$$

Finally, identifying the quantity in parentheses as the integral of a normal density from $-\infty$ to ∞, and hence equaling 1, we get

$$g(x) = \frac{e^{-\frac{1}{2}u^2}}{\sigma_1\sqrt{2\pi}} = \frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}$$

for $-\infty < x < \infty$. It follows by inspection that the marginal density of X is a normal distribution with the mean μ_1 and the standard deviation σ_1 and, by symmetry, that the marginal density of Y is a normal distribution with the mean μ_2 and the standard deviation σ_2.

As far as the parameter ρ is concerned, where ρ is the lowercase Greek letter *rho*, it is called the **correlation coefficient**, and the necessary integration will show that $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$. Thus, the parameter ρ measures how the two random variables X and Y vary together.

When we deal with a pair of random variables having a bivariate normal distribution, their conditional densities are also of importance; let us prove the following theorem.

THEOREM 9. If X and Y have a bivariate normal distribution, the conditional density of Y given $X = x$ is a normal distribution with the mean

$$\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

and the variance

$$\sigma^2_{Y|x} = \sigma_2^2(1 - \rho^2)$$

and the conditional density of X given $Y = y$ is a normal distribution with the mean

$$\mu_{X|y} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

and the variance

$$\sigma^2_{X|y} = \sigma_1^2(1 - \rho^2)$$

Proof Writing $w(y|x) = \dfrac{f(x, y)}{g(x)}$ in accordance with the definition of conditional density and letting $u = \dfrac{x - \mu_1}{\sigma_1}$ and $v = \dfrac{y - \mu_2}{\sigma_2}$ to simplify the notation, we get

$$\begin{aligned}
w(y|x) &= \frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[u^2 - 2\rho uv + v^2]}}{\dfrac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}u^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[v^2 - 2\rho uv + \rho^2u^2]} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left[\frac{v - \rho u}{\sqrt{1-\rho^2}}\right]^2}
\end{aligned}$$

Then, expressing this result in terms of the original variables, we obtain

$$w(y|x) = \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[\frac{y-\left\{\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}}{\sigma_2\sqrt{1-\rho^2}}\right]^2}$$

for $-\infty < y < \infty$, and it can be seen by inspection that this is a normal density with the mean $\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$ and the variance $\sigma^2_{Y|x} = \sigma^2_2(1 - \rho^2)$. The corresponding results for the conditional density of X given $Y = y$ follow by symmetry.

The bivariate normal distribution has many important properties, some statistical and some purely mathematical. Among the former, there is the following property, which the reader will be asked to prove in Exercise 43.

THEOREM 10. If two random variables have a bivariate normal distribution, they are independent if and only if $\rho = 0$.

In this connection, if $\rho = 0$, the random variables are said to be **uncorrelated**.

Also, we have shown that for two random variables having a bivariate normal distribution the two marginal densities are normal, but the converse is not necessarily true. In other words, the marginal distributions may both be normal without the joint distribution being a bivariate normal distribution. For instance, if the bivariate density of X and Y is given by

$$f^*(x, y) = \begin{cases} 2f(x, y) & \text{inside squares 2 and 4 of Figure 10} \\ 0 & \text{inside squares 1 and 3 of Figure 10} \\ f(x, y) & \text{elsewhere} \end{cases}$$

where $f(x, y)$ is the value of the bivariate normal density with $\mu_1 = 0, \mu_2 = 0$, and $\rho = 0$ at (x, y), it is easy to see that the marginal densities of X and Y are normal even though their joint density is not a bivariate normal distribution.

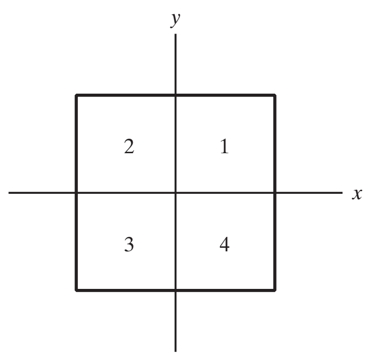

Figure 10. Sample space for the bivariate density given by $f^*(x, y)$.

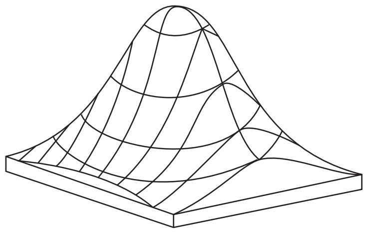

Figure 11. Bivariate normal surface.

Many interesting properties of the bivariate normal density are obtained by studying the **bivariate normal surface**, pictured in Figure 11, whose equation is $z = f(x, y)$, where $f(x, y)$ is the value of the bivariate normal density at (x, y). As the reader will be asked to verify in some of the exercises that follow, the bivariate normal surface has a maximum at (μ_1, μ_2), any plane parallel to the z-axis intersects the surface in a curve having the shape of a normal distribution, and any plane parallel to the xy-plane that intersects the surface intersects it in an ellipse called a **contour of constant probability density**. When $\rho = 0$ and $\sigma_1 = \sigma_2$, the contours of constant probability density are circles, and it is customary to refer to the corresponding joint density as a **circular normal distribution**.

Exercises

43. To prove Theorem 10, show that if X and Y have a bivariate normal distribution, then
(a) their independence implies that $\rho = 0$;
(b) $\rho = 0$ implies that they are independent.

44. Show that any plane perpendicular to the xy-plane intersects the bivariate normal surface in a curve having the shape of a normal distribution.

45. If the exponent of e of a bivariate normal density is

$$\frac{-1}{102}[(x+2)^2 - 2.8(x+2)(y-1) + 4(y-1)^2]$$

find
(a) $\mu_1, \mu_2, \sigma_1, \sigma_2$, and ρ;
(b) $\mu_{Y|x}$ and $\sigma^2_{Y|x}$.

46. If the exponent of e of a bivariate normal density is

$$\frac{-1}{54}(x^2 + 4y^2 + 2xy + 2x + 8y + 4)$$

find σ_1, σ_2, and ρ, given that $\mu_1 = 0$ and $\mu_2 = -1$.

47. If X and Y have the bivariate normal distribution with $\mu_1 = 2, \mu_2 = 5, \sigma_1 = 3, \sigma_2 = 6$, and $\rho = \frac{2}{3}$, find $\mu_{Y|1}$ and $\sigma_{Y|1}$.

48. If X and Y have a bivariate normal distribution and $U = X + Y$ and $V = X - Y$, find an expression for the correlation coefficient of U and V.

49. If X and Y have a bivariate normal distribution, it can be shown that their joint moment-generating function is given by

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= E(e^{t_1 X + t_2 Y}) \\ &= e^{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)} \end{aligned}$$

Verify that
(a) the first partial derivative of this function with respect to t_1 at $t_1 = 0$ and $t_2 = 0$ is μ_1;
(b) the second partial derivative with respect to t_1 at $t_1 = 0$ and $t_2 = 0$ is $\sigma_1^2 + \mu_1^2$;
(c) the second partial derivative with respect to t_1 and t_2 at $t_1 = 0$ and $t_2 = 0$ is $\rho\sigma_1\sigma_2 + \mu_1\mu_2$.

8 The Theory in Practice

In many of the applications of statistics it is assumed that the data are approximately normally distributed. Thus, it is important to make sure that the assumption

Regression and Correlation

1 Introduction

A major objective of many statistical investigations is to establish relationships that make it possible to predict one or more variables in terms of others. Thus, studies are made to predict the potential sales of a new product in terms of its price, a patient's weight in terms of the number of weeks he or she has been on a diet, family expenditures on entertainment in terms of family income, the per capita consumption of certain foods in terms of their nutritional values and the amount of money spent advertising them on television, and so forth.

Although it is, of course, desirable to be able to predict one quantity exactly in terms of others, this is seldom possible, and in most instances we have to be satisfied with predicting averages or expected values. Thus, we may not be able to predict exactly how much money Mr. Brown will make 10 years after graduating from college, but, given suitable data, we can predict the average income of a college graduate in terms of the number of years he has been out of college. Similarly, we can at best predict the average yield of a given variety of wheat in terms of data on the rainfall in July, and we can at best predict the average performance of students starting college in terms of their I.Q.'s.

Formally, if we are given the joint distribution of two random variables X and Y, and X is known to take on the value x, the basic problem of **bivariate regression** is that of determining the conditional mean $\mu_{Y|x}$, that is, the "average" value of Y for the given value of X. The term "regression," as it is used here, dates back to Francis Galton, who employed it to indicate certain relationships in the theory of heredity. In problems involving more than two random variables, that is, in **multiple regression**, we are concerned with quantities such as $\mu_{Z|x,y}$, the mean of Z for given values of X and Y, $\mu_{X_4|x_1,\, x_2,\, x_3}$, the mean of X_4 for given values of X_1, X_2, and X_3, and so on.

Definition 1. Bivariate Regression; Regression Equation. *If* f(x, y) *is the value of the joint density of two random variables* X *and* Y, ***bivariate regression*** *consists of determining the conditional density of* Y, *given* X = x *and then evaluating the integral*

$$\mu_{Y|x} = E(Y|x) = \int_{-\infty}^{\infty} y \cdot w(y|x)dy$$

From Chapter 14 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

The resulting equation is called the ***regression equation of*** Y ***on*** X*. Alternately, the* ***regression equation of*** X ***on*** Y *is given by*

$$\mu_{X|y} = E(X|y) = \int_{-\infty}^{\infty} x \cdot f(x|y)dy$$

In the discrete case, when we are dealing with probability distributions instead of probability densities, the integrals in the two regression equations given in Definition 1 are simply replaced by sums. When we do not know the joint probability density or distribution of the two random variables, or at least not all its parameters, the determination of $\mu_{Y|x}$ or $\mu_{X|y}$ becomes a problem of estimation based on sample data; this is an entirely different problem, which we shall discuss in Sections 3 and 4.

EXAMPLE 1

Given the two random variables X and Y that have the joint density

$$f(x, y) = \begin{cases} x \cdot e^{-x(1+y)} & \text{for } x > 0 \text{ and } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the regression equation of Y on X and sketch the regression curve.

Solution

Integrating out y, we find that the marginal density of X is given by

$$g(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and hence the conditional density of Y given $X = x$ is given by

$$w(y|x) = \frac{f(x, y)}{g(x)} = \frac{x \cdot e^{-x(1+y)}}{e^{-x}} = x \cdot e^{-xy}$$

for $y > 0$ and $w(y|x) = 0$ elsewhere, which we recognize as an exponential density with $\theta = \frac{1}{x}$. Hence, by evaluating

$$\mu_{Y|x} = \int_{0}^{\infty} y \cdot x \cdot e^{-xy}\, dy$$

or by referring to the corollary of a theorem given here "The mean and the variance of the exponential distribution are given by $\mu = \theta$ and $\sigma^2 = \theta^2$," we find that the regression equation of Y on X is given by

$$\mu_{Y|x} = \frac{1}{x}$$

The corresponding regression curve is shown in Figure 1.

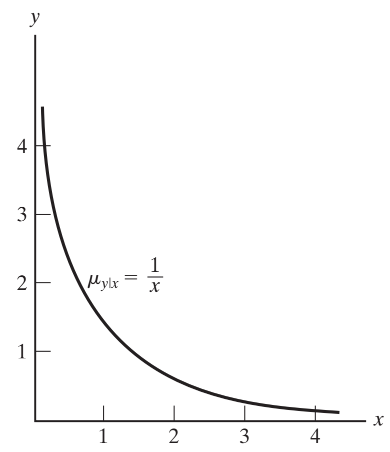

Figure 1. Regression curve of Example 1.

EXAMPLE 2

If X and Y have the multinomial distribution

$$f(x, y) = \binom{n}{x, y, n-x-y} \cdot \theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y}$$

for $x = 0, 1, 2, \ldots, n$, and $y = 0, 1, 2, \ldots, n$, with $x + y \leqq n$, find the regression equation of Y on X.

Solution

The marginal distribution of X is given by

$$\begin{aligned} g(x) &= \sum_{y=0}^{n-x} \binom{n}{x, y, n-x-y} \cdot \theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y} \\ &= \binom{n}{x} \theta_1^x (1-\theta_1)^{n-x} \end{aligned}$$

for $x = 0, 1, 2, \ldots, n$, which we recognize as a binomial distribution with the parameters n and θ_1. Hence,

$$w(y|x) = \frac{f(x, y)}{g(x)} = \frac{\binom{n-x}{y} \theta_2^y (1-\theta_1-\theta_2)^{n-x-y}}{(1-\theta_1)^{n-x}}$$

for $y = 0, 1, 2, \ldots, n-x$, and, rewriting this formula as

$$w(y|x) = \binom{n-x}{y} \left(\frac{\theta_2}{1-\theta_1}\right)^y \left(\frac{1-\theta_1-\theta_2}{1-\theta_1}\right)^{n-x-y}$$

we find by inspection that the conditional distribution of Y given $X = x$ is a binomial distribution with the parameters $n - x$ and $\dfrac{\theta_2}{1-\theta_1}$, so that the regression equation of Y on X is

$$\mu_{Y|x} = \frac{(n-x)\theta_2}{1-\theta_1}$$

With reference to the preceding example, if we let X be the number of times that an even number comes up in 30 rolls of a balanced die and Y be the number of times that the result is a 5, then the regression equation becomes

$$\mu_{Y|x} = \frac{(30-x)\frac{1}{6}}{1-\frac{1}{2}} = \frac{1}{3}(30-x)$$

This stands to reason, because there are three equally likely possibilities, 1, 3, or 5, for each of the $30 - x$ outcomes that are not even.

EXAMPLE 3

If the joint density of X_1, X_2, and X_3 is given by

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the regression equation of X_2 on X_1 and X_3.

Solution

The joint marginal density of X_1 and X_3 is given by

$$m(x_1, x_3) = \begin{cases} \left(x_1 + \frac{1}{2}\right)e^{-x_3} & \text{for } 0 < x_1 < 1, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore,

$$\mu_{X_2|x_1,x_3} = \int_{-\infty}^{\infty} x_2 \cdot \frac{f(x_1, x_2, x_3)}{m(x_1, x_3)}\, dx_2 = \int_0^1 \frac{x_2(x_1 + x_2)}{\left(x_1 + \frac{1}{2}\right)}\, dx_2$$

$$= \frac{x_1 + \frac{2}{3}}{2x_1 + 1}$$

Note that the conditional expectation obtained in the preceding example depends on x_1 but not on x_3. This could have been expected, since there is a pairwise independence between X_2 and X_3.

2 Linear Regression

An important feature of Example 2 is that the regression equation is linear; that is, it is of the form

$$\mu_{Y|x} = \alpha + \beta x$$

where α and β are constants, called the **regression coefficients**. There are several reasons why linear regression equations are of special interest: First, they lend themselves readily to further mathematical treatment; then, they often provide good approximations to otherwise complicated regression equations; and, finally, in the case of the bivariate normal distribution, the regression equations are, in fact, linear.

To simplify the study of linear regression equations, let us express the regression coefficients α and β in terms of some of the lower moments of the joint distribution of X and Y, that is, in terms of $E(X) = \mu_1$, $E(Y) = \mu_2$, $\text{var}(X) = \sigma_1^2$, $\text{var}(Y) = \sigma_2^2$, and $\text{cov}(X, Y) = \sigma_{12}$. Then, also using the correlation coefficient

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

we can prove the following results.

THEOREM 1. If the regression of Y on X is linear, then

$$\mu_{Y|x} = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

and if the regression of X on Y is linear, then

$$\mu_{X|y} = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

Proof Since $\mu_{Y|x} = \alpha + \beta x$, it follows that

$$\int y \cdot w(y|x)\, dy = \alpha + \beta x$$

and if we multiply the expression on both sides of this equation by $g(x)$, the corresponding value of the marginal density of X, and integrate on x, we obtain

$$\iint y \cdot w(y|x) g(x)\, dy\, dx = \alpha \int g(x)\, dx + \beta \int x \cdot g(x)\, dx$$

or

$$\mu_2 = \alpha + \beta \mu_1$$

since $w(y|x)g(x) = f(x,y)$. If we had multiplied the equation for $\mu_{Y|x}$ on both sides by $x \cdot g(x)$ before integrating on x, we would have obtained

$$\iint xy \cdot f(x,y)\,dy\,dx = \alpha \int x \cdot g(x)\,dx + \beta \int x^2 \cdot g(x)\,dx$$

or

$$E(XY) = \alpha\mu_1 + \beta E(X^2)$$

Solving $\mu_2 = \alpha + \beta\mu_1$ and $E(XY) = \alpha\mu_1 + \beta E(X^2)$ for α and β and making use of the fact that $E(XY) = \sigma_{12} + \mu_1\mu_2$ and $E(X^2) = \sigma_1^2 + \mu_1^2$, we find that

$$\alpha = \mu_2 - \frac{\sigma_{12}}{\sigma_1^2} \cdot \mu_1 = \mu_2 - \rho\frac{\sigma_2}{\sigma_1} \cdot \mu_1$$

and

$$\beta = \frac{\sigma_{12}}{\sigma_1^2} = \rho\frac{\sigma_2}{\sigma_1}$$

This enables us to write the linear regression equation of Y on X as

$$\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

When the regression of X on Y is linear, similar steps lead to the equation

$$\mu_{X|y} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

It follows from Theorem 1 that if the regression equation is linear and $\rho = 0$, then $\mu_{Y|x}$ does not depend on x (or $\mu_{X|y}$ does not depend on y). When $\rho = 0$ and hence $\sigma_{12} = 0$, the two random variables X and Y are **uncorrelated**, and we can say that if two random variables are independent, they are also uncorrelated, but if two random variables are uncorrelated, they are not necessarily independent; the latter is again illustrated in Exercise 9.

The correlation coefficient and its estimates are of importance in many statistical investigations, and they will be discussed in some detail in Section 5. At this time, let us again point out that $-1 \leqq \rho \leqq +1$, as the reader will be asked to prove in Exercise 11, and the sign of ρ tells us directly whether the slope of a regression line is upward or downward.

3 The Method of Least Squares

In the preceding sections we have discussed the problem of regression only in connection with random variables having known joint distributions. In actual practice, there are many problems where a set of **paired data** gives the indication that the regression is linear, where we do not know the joint distribution of the random variables under consideration but, nevertheless, want to estimate the regression