THE BIVARIATE NORMAL DISTRIBUTION

Most of the results relating to the normal distrution can be obtained by examining the functional form of the bivariate distribution. Let x and y be the two variables. Let us denote their means by

$$(1) \qquad E(x) = \mu_x, \qquad E(y) = \mu_y,$$

their variances by

$$(2) \qquad V(x) = \sigma_x^2, \qquad V(y) = \sigma_y^2$$

and their covariance by

$$(3) \qquad C(x, y) = \rho\sigma_x\sigma_y.$$

Here

$$(4) \qquad \rho = \frac{C(x, y)}{\sqrt{V(x)V(y)}},$$

which is called the correlation coefficient of x and y, provides a measure of the relatedness of these variables.

The Cauchy–Schwarz inequality indicates that $-1 \leq \rho \leq 1$. If $\rho = 1$, then there is an exact positive linear relationship between the variables whereas, if $\rho = -1$, then there is an exact negative linear relationship, Neither of these extreme cases is admissible in the present context for, as we may see by examining the following formulae, they lead to the collapse of the bivariate distribution.

The bivariate distribution is specified by

$$(5) \qquad f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp Q(x, y),$$

where

$$(6) \quad Q = \frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$$

is a quadratic function of x and y.

The function can also be written as

$$(7) \qquad Q = \frac{-1}{2(1-\rho^2)}\left\{\left(\frac{y-\mu_y}{\sigma_y} - \rho\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}.$$

Thus we have

$$f(x,y) = f(y|x)f(x), \tag{8}$$

where

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right\}, \tag{9}$$

and

$$f(y|x) = \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(y-\mu_{y|x})^2}{2\sigma_y^2(1-\rho)^2}\right\}, \tag{10}$$

with

$$\mu_{y|x} = \mu_y + \frac{\rho\sigma_y^2}{\sigma_x^2}(x-\mu_x). \tag{11}$$

Equation (11) is the linear regression equation which specifies the value of $E(y|x) = \mu_{y|x}$ in terms of x; and it is simply the equation (12) in another notation. Equation (10) indicates that the variance of y about its conditional expectation is

$$V(y|x) = \sigma_y^2(1-\rho^2). \tag{12}$$

Since $(1-\rho^2) \leq 1$, it follows that variance of the conditional predictor $E(y|x)$ is less than that of the unconditional predictor $E(y)$ whenever $\rho \neq 0$—which is whenever there is a correlation between x and y. Moreover, as this correlation increases, the variance of the conditional predictor diminishes.

There is, of course, a perfect symmetry between the arguments x and y in the bivariate distribution. Thus, if we choose to factorise the joint probability density function as $f(x,y) = f(x|y)f(y)$, then, to obtain the relevant results, we need only interchange the x's and the y's in the formulae above.