BIVARIATE DISTRIBUTIONS

The probabilities $f(x_i)$ of the values $\{x_1, x_2, \ldots, x_n\}$ assumed by a discrete random variable x are such that

$$f(x_i) \geq 0 \quad \text{for all} \quad i \qquad \text{and} \quad \sum_i f(x_i) = 1.$$

For example:

x	$f(x)$
-1	0.25
1	0.75

If x and y take values $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$, respectively, and if $f(x_i, y_j)$, is their joint probability mass function, then

$$f(x_i, y_j) \geq 0 \quad \text{for all} \quad i \qquad \text{and} \quad \sum_i \sum_j f(x_i, y_j) = 1.$$

An example is the following table, which contains the values of $f(x_i, y_j)$:

	y		
x	-1	0	1
-1	0.04	0.01	0.20
1	0.12	0.03	0.60

The marginal function of x gives the probabilities of the values of x_i regardless of the values of y_j with which they are associated. It is defined by

$$f(x_i) = \sum_j f(x_i, y_j); \quad i = 1, \ldots, n.$$

It follows that

$$f(x_i) \geq 0, \quad \text{and} \quad \sum_i f(x_i) = \sum_i \sum_j f(x_i, y_j) = 1,$$

The marginal function $f(y_j)$ is defined likewise. The bivariate table above gives

$$\begin{aligned} f(x = -1) &= 0.04 + 0.01 + 0.20 = 0.25, \\ f(x = 1) &= 0.12 + 0.03 + 0.60 = 0.75, \\ f(y = -1) &= 0.04 + 0.12 = 0.16, \\ f(y = 0) &= 0.01 + 0.03 = 0.04, \\ f(y = 1) &= 0.20 + 0.60 = 0.80. \end{aligned}$$

The conditional function $f(x_i|y_j)$ gives the probability of the values of x_i when $y = y_j$:

$$f(x_i|y_j) = \frac{f(x_i, y_j)}{f(y_j)}.$$

Observe that

$$\sum_i f(x_i|y_j) = \frac{\sum_i f(x_i, y_j)}{f(y_j)} = \frac{f(y_j)}{f(y_j)} = 1.$$

An example based on the bivariate table is as follows:

$f(x|y)$

x	$f(x\|y = -1)$	$f(x\|y = 0)$	$f(x\|y = 1)$
-1	$0.25 = \frac{0.04}{0.16}$	$0.25 = \frac{0.01}{0.04}$	$0.25 = \frac{0.20}{0.80}$
1	$0.75 = \frac{0.12}{0.16}$	$0.75 = \frac{0.03}{0.04}$	$0.75 = \frac{0.60}{0.80}$

Indpdendence. We may say that x is independent of y if and only if the conditional distribution of x is the same for all values of y.

The conditional functions of x are the same for all values of y if and only all are equal to the marginal are function of x.

Proof. Suppose that $f^*(x_i) = f(x|y_1) = \cdots = f(x|y_m)$. Then

$$f(x_i) = \sum_j f(x_i|y_j)f(y_j) = f^*(x_i)\sum_j f(y_j) = f^*(x_i).$$

Conversely, if the conditionals are all equal to the marginal, then they must be equal to each other.

If $f(x_i|y_j) = f(x_i)$ and $f(y_j|x_i) = f(y_j)$, then, equivalently,

$$f(x_i, y_j) = f(x_i|y_j)f(y_j) = f(y_j|x_i)f(x_i) = f(x_i)f(y_j),$$

and $f(x_i, y_j) = f(x_i)f(y_j)$ also signifies independence.

Density functions. For the continuous case, consider a space $\mathcal{R}^2$, which is the set of all pairs $(x, y); -\infty < x, y < \infty$ that correspond to the co-ordinates of the points in a plane.

A probability measure $P(\mathcal{A})$ is defined over $\mathcal{R}^2$, which gives the probability that (x, y) falls in any $\mathcal{A} \subset \mathcal{R}^2$.

If $\mathcal{A} = \{a < x \leq b, a < x \leq b\}$, which is a rectangle in the plane, then

$$P(\mathcal{A}) = \int_{y=c}^{d} \left\{ \int_{x=a}^{b} f(x, y)dx \right\} dy.$$

Example. Let (x, y) be a random vector with a p.d.f of

$$f(x, y) = \frac{1}{8}(6 - x - y); \quad 0 \le x \le 2; \quad 2 \le y \le 4.$$

It needs to be confirmed that this does integrate to unity over the specified range of (x, y). There is

$$\begin{aligned}\frac{1}{8}\int_{x=0}^{2}\int_{y=2}^{4}(6 - x - y)dydx &= \frac{1}{8}\int_{x=0}^{2}\left[6y - xy - \frac{y^2}{2}\right]_2^4 dx\\ &= \frac{1}{8}\int_{x=0}^{2}(6 - 2x)\,dx = \frac{1}{8}\left[6x - x^2\right]_0^2 = \frac{8}{8} = 1.\end{aligned}$$

Moments of a bivariate distribution. Let (x, y) have the p.d.f. $f(x, y)$. The expected value of x is defined by

$$E(x) = \int_x\int_y xf(x, y)dydx = \int_x xf(x)dx,$$

if x is continuous, and by

$$E(x) = \sum_x\sum_y xf(x, y) = \sum_x xf(x), \quad \text{if } x \text{ is discrete.}$$

Joint moments of x and y can also be defined. For example, there is

$$\int_x\int_y (x - a)^r(y - b)^s f(x, y)dydx,$$

where r, s are integers and a, b are fixed data.

The covariance of x and y is defined by

$$C(x,y) = \int_x \int_y \{x - E(x)\}\{y - E(y)\} f(x,y) dy dx.$$

If x, y are statistically independent, with $f(x,y) = f(x)f(y)$, then their joint moments can be expressed as the products of their separate moments.

For example, if x, y are independent then

$$\begin{aligned} &E\{[x - E(x)]^2 [y - E(y)]^2\} \\ &= \int_x [x - E(x)]^2 f(x) dx \int_y [y - E(y)]^2 f(y) dy = V(x)V(y). \end{aligned}$$

When x, y are independent, the covariance is

$$\begin{aligned} C(x,y) &= E\{[x - E(x)][y - E(y)]\} \\ &= \{[E(x) - E(x)][E(y) - E(y)]\} = 0. \end{aligned}$$

This can be expressed using the expectations operator:

$$\begin{aligned} C(x,y) &= E\{[x - E(x)][y - E(y)]\} \\ &= E[xy - E(x)y - xE(y) + E(x)E(y)] \\ &= E(xy) - E(x)E(y) - E(x)E(y) + E(x)E(y) \\ &= E(xy) - E(x)E(y). \end{aligned}$$

Since $E(xy) = E(x)E(y)$ if x, y are independent, it follows that $C(x,y) = 0$. Also, $C(x,x) = E\{[x - E(x)]^2\} = V(x)$.

Now consider the variance of the sum $x + y$. This is

$$\begin{aligned} V(x+y) &= E\left\{[(x+y) - E(x+y)]^2\right\} \\ &= E\left\{[\{x - E(x)\} + \{y - E(y)\}]^2\right\} \\ &= E\Big\{[x - E(x)]^2\} + E\{[y - E(y)]^2\} + 2[x - E(x)][y - E(y)]\Big\} \\ &= V(x) + V(y) + 2C(x,y). \end{aligned}$$

If x, y are independent, then $C(x,y) = 0$ and $V(x+y) = V(x) + V(y)$. Note that

If x, y are independent, then the covariance is $C(x,y) = 0$. However, the condition $C(x,y) = 0$ does not, in general, imply that x, y are independent.

If x, y are normally distributed, then $C(x,y) = 0$ does imply their independence.

The correlation coefficient. To measure the relatedness of x and y, we use the correlation coefficient, defined by

$$\begin{aligned} \mathrm{Corr}(x,y) &= \frac{C(x,y)}{\sqrt{V(x)V(y)}} \\ &= \frac{E\{[x - E(x)][y - E(y)]\})}{\sqrt{E\{[x - E(x)]^2\}E\{[y - E(y)]^2\}}}. \end{aligned}$$

There is $-1 \leq \mathrm{Corr}(x,y) \leq 1$.

If $\mathrm{Corr}(x,y) = 1$, then x, y lie on a straight line of positive slope. If $\mathrm{Corr}(x,y) = -1$, then the line has negative slope.

If $\mathrm{Corr}(x,y) = 0$, then there is no linear relationship between x and y.

REGRESSION AND CONDITIONAL EXPECTATIONS

Linear conditional expectations. If x, y are correlated, then, the conditional expectation $E(y|x)$ provides a better prediction than $E(y)$. Assume that

$$E(y|x) = \alpha + x\beta. \tag{i}$$

which is described as a linear regression. If the prediction error is $\varepsilon = y - E(y|x)$, then

$$y = E(y|x) + \varepsilon = \alpha + x\beta + \varepsilon.$$

The object is to express α and β as functions of the moments of the joint probability distribution of x and y.

Multiplying equation (i) throughout by $f(x)$, and by integrating with respect to x gives

$$E(y) = \alpha + \beta E(x), \tag{ii}$$

whence

$$\alpha = E(y) - \beta E(x). \tag{iii}$$

Thus the regression line passes through $\{E(x), E(y)\}$, which is the expected value of the joint distribution.

By putting (iii) into (i), we find that

$$E(y|x) = E(y) + \beta\{x - E(x)\},$$

which shows how the conditional expectation of y differs from the unconditional expectation in proportion to the error of predicting x by its expected value.

Now multiply (i) by x and $f(x)$ and integrate with respect to x to give

$$E(xy) = \alpha E(x) + \beta E(x^2). \tag{iv}$$

Multiplying (ii) by $E(x)$ gives

$$E(x)E(y) = \alpha E(x) + \beta\big\{E(x)\big\}^2, \tag{v}$$

whence, on taking (v) from (iv), we get

$$E(xy) - E(x)E(y) = \beta\Big[E(x^2) - \big\{E(x)\big\}^2\Big],$$

which implies that

$$\begin{aligned}\beta &= \frac{E(xy) - E(x)E(y)}{E(x^2) - \big\{E(x)\big\}^2} \\ &= \frac{E\Big[\big\{x - E(x)\big\}\big\{y - E(y)\big\}\Big]}{E\Big[\big\{x - E(x)\big\}^2\Big]} \qquad \text{(vi)} \\ &= \frac{C(x,y)}{V(x)}.\end{aligned}$$

Thus, α and β are expressed in terms of $E(x)$, $E(y)$, $V(x)$ and $C(x,y)$, which are the moments of the joint distribution.

The prediction error $\varepsilon = y - E(y|x)$ is uncorrelated with x. This is shown by writing

$$E\Big[\big\{y - E(y|x)\big\}x\Big] = E(yx) - \alpha E(x) - \beta E(x^2) = 0, \tag{vii}$$

where the final equality comes from (iv).

The Cauchy–Schwarz inequality. This establishes the bounds on

$$\mathrm{Corr}(x,y) = C(x,y)/\sqrt{V(x)V(y)},$$

which is the coefficent of the correlation of x and y.

Let $\varepsilon = y - E(y|x)$, and consider

$$\begin{aligned} E(\varepsilon^2) =& E\left(\left[\{y - E(y)\} - \beta\{x - E(x)\}\right]^2\right) \\ &= V(y) - 2\beta C(x,y) + \beta^2 V(x) \geq 0. \end{aligned}$$

Setting $\beta = C(x,y)/V(x)$ gives

$$V(y) - 2\frac{\{C(x,y)\}^2}{V(x)} + \frac{\{C(x,y)\}^2}{V(x)} \geq 0,$$

whence

$$V(x)V(y) \geq \{C(x,y)\}^2.$$

It follows that $\{\mathrm{Corr}(x,y)\}^2 \leq 1$ and, therefore, that

$$-1 \leq \mathrm{Corr}(x,y) \leq 1.$$

Empirical Regressions. Given T observations on x and y, the sample moments can be calculated:

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t, \qquad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t,$$

$$S_x^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})x_t = \frac{1}{T}\sum_{t=1}^{T} x_t^2 - \bar{x}^2,$$

$$S_{xy} = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y}) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})y_t = \frac{1}{T}\sum_{t=1}^{T} x_t y_t - \bar{x}\bar{y},$$

Replacing the moments in the formulae (iii) and (vi) by the sample moments gives the following estimates of α and β:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},$$

$$\hat{\beta} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2}.$$

We can also estimate α and β by finding the values which minimise

$$S = \sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

$$= \sum_{t=1}^{T}(y_t - \alpha - x_t\beta)^2.$$

This is the sum of squares of the vertical distances, measured parallel to the y-axis, of the data points from an interpolated regression line.

Differentiating the function S with respect to α and setting the results to zero for a minimum gives

$$-2\sum(y_t - \alpha - \beta x_t) = 0, \quad \text{or, equivalently,}$$
$$\bar{y} - \alpha - \beta\bar{x} = 0.$$

This generates the estimating equation for α:

$$\alpha(\beta) = \bar{y} - \beta\bar{x}. \tag{viii}$$

Differentiating with respect to β and setting the result to zero gives

$$-2\sum x_t(y_t - \alpha - \beta x_t) = 0. \tag{ix}$$

On substituting for α from (vii) and eliminating the factor -2, this becomes

$$\sum x_t y_t - \sum x_t(\bar{y} - \beta\bar{x}) - \beta\sum x_t^2 = 0,$$

whence we get

$$\hat{\beta} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2}$$
$$= \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2}.$$

This expression is identical to the one derived by the method of moments. By putting $\hat{\beta}$ into the estimating equation for α under (viii), we derive the same estimate $\hat{\alpha}$ for the intercept parameter as the one obtained by the method of moments.

Equation (ix) is the empirical analogue of equation (vii) which expresses the condition that the prediction error is uncorrelated with the values of x.

The method of least squares does not provide an estimate of $\sigma^2 = E(\varepsilon_t^2)$. Instead , we invoke the method of moments.

Taking the regression residuals $e_t = y_t - \hat{\alpha} - \hat{\beta}x_t$ as estimates of the corresponding values of ε_t, we get an estimator in the form of

$$\tilde{\sigma}^2 = \frac{1}{T}\sum e_t^2.$$

In fact, this is a biased estimator with

$$E\left(T\tilde{\sigma}^2\right) = \left\{T - 2\right\}\sigma^2;$$

so it is common to adopt the unbiased estimator

$$\hat{\sigma}^2 = \frac{\sum e_t^2}{T-2}.$$