Chapter 6

Correlation and covariance

6.1 Two discrete random variables

Suppose we have a problem involving a pair of random variables. For example, In general, given two discrete random variables X, Y, the probability that both events occur may be related. To quantify these double events, we define a *joint probability mass* function:

$$f_{XY}(x, y) \equiv P(X = x \text{ and } Y = y) \qquad . \tag{6.1}$$

The corresponding *joint probability distribution* is defined

$$F_{XY}(x, y) \equiv P(X \leq x \text{ and } Y \leq y) \qquad . \tag{6.2}$$

6.1.1 Independent Events

We are in a position to give an authoritative definition of independent events. Recall that if the events A and B are independent:

$$P(A \cap B) = P(A)P(B) \qquad . \tag{6.3}$$

The discrete random variables X, Y are *independent* if, and only if,

$$f_{XY}(x, y) = f_X(x) f_Y(y) \text{ for all } \qquad x,\ y \qquad . \tag{6.4}$$

Suppose that the variables take on the discrete set of values:

$$X \in \{x_1, x_2, \ldots, x_i, \ldots, x_m\} \qquad , \qquad Y \in \{y_1, y_2, \ldots, y_j, \ldots, y_n\}$$

Then the total probability relation is the double series (in long or shorthand version):

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{XY}(x_i, y_j) = \sum_{i,j} f_{XY}(x_i, y_j) = 1 \qquad .$$

If X, Y are independent, then it is easily shown that:

$$\mathbb{E}\left(XY\right) = \mathbb{E}\left(X\right) \mathbb{E}\left(Y\right). \tag{6.5}$$

Proof

Start with the definition:

$$\mathbb{E}\left(XY\right) \equiv \sum_{i,j} x_i y_j f_{XY}(x_i, y_i) \quad ,$$

then, if they are independent, the joint mass can be factorised as follows:

$$\mathbb{E}\left(XY\right) = \sum_{i,j} x_i y_j f_X(x_i) f_Y(y_j)$$

The double sum can be evaluated as:

$$\mathbb{E}(XY) = \sum_i x_i f_X(x_i) \left(\sum_j y_j f_Y(y_j) \right) = \sum_i x_i f_X(x_i) \left(\mathbb{E}(Y)\right)$$

Since $\mathbb{E}(Y)$ is just a number, a constant factor common to all terms, it can be extracted so that:

$$\mathbb{E}(XY) = \mathbb{E}(Y) . \sum_i x_i f_X(x_i) = \mathbb{E}(Y)\,\mathbb{E}(X) \qquad .$$

6.1.2 Marginal probability

We define the *marginal probability mass functions* as follows:

$$f_X(x) = P(X = x) = \sum_j P(X = x, Y = y_j) = \sum_j f_{X,Y}(x, y_j) \quad , \tag{6.6}$$

and

$$f_Y(y) = \sum_i f_{X,Y}(x_i, y) \qquad . \tag{6.7}$$

6.1.3 Inclusion-Exclusion

Suppose we are interested in whether one or either event occurred. Then

$$P(X = x \text{ or } Y = y) \quad , \tag{6.8}$$

would be written in this notation of *marginal* and *joint* masses as:

$$f_X(x) + f_Y(y) - f_{XY}(x, y) \quad . \tag{6.9}$$

It then follows that, for two *disjoint events*:

$$f_{XY}(x, y) = 0 \tag{6.10}$$

and hence, for *mutually exclusive* discrete random variables:

$$P(X = x \text{ or } Y = y) = f_X(x) + f_Y(y) \quad . \tag{6.11}$$

6.1.4 Correlation

In general the expectation of a function of the pair of random variables X, Y is defined:

$$\mathbb{E}(g(X, Y)) \equiv \sum_{i,j} g(x_i, y_j) f_{XY}(x_i, y_j) \qquad . \tag{6.12}$$

If it is the case that

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y) \tag{6.13}$$

then X, Y are said to be *uncorrelated*. Note that if two variables are independent, this implies they are uncorrelated. The converse is *not* necessarily true.

Another significant result is that, if X and Y are uncorrelated then:

$$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) \qquad . \tag{6.14}$$

More generally, given two constants, a and b, and X and Y *uncorrelated*, we obtain:

$$\begin{aligned}\text{var}\left(aX+bY\right) &= \mathbb{E}\left((aX+bY)^2\right)-\left(\mathbb{E}\left(aX+bY\right)\right)^2\\ &= \mathbb{E}\left(a^2X^2+2abXY+b^2Y^2\right)-\left(a\mathbb{E}\left(X\right)+b\mathbb{E}\left(Y\right)\right)^2\\ &= a^2\mathbb{E}\left(X^2\right)+2ab\mathbb{E}\left(XY\right)+b^2\mathbb{E}\left(Y^2\right)-a^2\mathbb{E}\left(X\right)^2-2ab\mathbb{E}\left(X\right)\mathbb{E}\left(Y\right)-b^2\mathbb{E}\left(Y\right)^2\\ &= a^2(\mathbb{E}\left(X^2\right)-\mathbb{E}\left(X\right)^2)+b^2(\mathbb{E}\left(Y^2\right)-\mathbb{E}\left(Y\right)^2)\end{aligned}$$

That is:

$$\text{var}\left(aX+bY\right)=a^2\text{var}\left(X\right)+b^2\text{var}\left(Y\right) \quad . \tag{6.15}$$

The *covariance* of two discrete random variables is defined as:

$$\text{cov}(X,Y)\equiv\mathbb{E}\left(XY\right)-\mathbb{E}\left(X\right).\mathbb{E}\left(Y\right)=\mathbb{E}\left((X-\mathbb{E}\left(X\right))(Y-\mathbb{E}\left(Y\right))\right) \quad . \tag{6.16}$$

Clearly if X, Y are uncorrelated then:

$$\text{cov}(X,Y)=0. \tag{6.17}$$

A measure of correlation is given by the *correlation coefficient*, $\rho(X,Y)$. This is defined as:

$$\rho(X,Y)=\frac{\text{cov}(X,Y)}{\sqrt{\text{var}\left(X\right)\text{var}\left(Y\right)}} \tag{6.18}$$

where, clearly, $\rho(X,Y)=0$, describes two uncorrelated variables. The larger the value of $\rho(X,Y)$, the more correlated the variables are. In particular, if the correlation is perfect, for example, $Y=aX+b$, with $a>0$,then, $\rho(X,Y)=1$. Similary X and Y are said to be perfectly *anti-correlated* when $a<0$, in which case: $\rho(X,Y)=-1$.

6.2 Cauchy-Schwarz inequality

The *Cauchy-Schwarz inequality* states that, for any pair of random variables:

$$\left(\mathbb{E}\left(XY\right)\right)^2\leq\mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right) \tag{6.19}$$

Proof:

Let $Z=\alpha X+Y$, where α is an arbitrary real constant. Then clearly: $Z^2=(\alpha X+Y)^2\geq 0$. Thus: $\mathbb{E}\left(Z^2\right)\geq 0$. That is

$$\mathbb{E}\left(\alpha^2X^2+2\alpha XY+Y^2\right)\geq 0 \quad .$$

Therefore:

$$\alpha^2\mathbb{E}\left(X^2\right)+2\alpha\mathbb{E}\left(XY\right)+\mathbb{E}\left(Y^2\right)\geq 0 \quad .$$

This holds for any (all) real α. The requirement that a quadratic expression $(a\alpha^2+b\alpha+c)$ is non-negative means that: $a>0, b^2-4ac\leq 0$. This means that there are no real roots for the zero, or the roots are repeated, which in turn implies that:

$$\left[2\mathbb{E}\left(XY\right)\right]^2-4\mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)\leq 0 \quad .$$

and thus:

$$\left(\mathbb{E}\left(XY\right)\right)^2\leq\mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right) \tag{6.20}$$

6.2.1 Limits of correlation coefficient

We can calculate the limits of the correlation coefficient using the Cauchy-Schwarz inequality. Making the replacements: $X \to X - \mathbb{E}(X)$, and $Y \to Y - \mathbb{E}(Y)$, gives the expression:

$$[\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]^2 \leq \mathbb{E}\left((X - \mathbb{E}(X))^2\right)\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) \qquad .$$

Simplifying both sides we find:

$$[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)]^2 \leq \mathbb{E}\left((X - \mathbb{E}(X))^2\right)\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) \qquad .$$

which reduces to:

$$[\text{cov}(X, Y)]^2 \leq \text{var}(X)\text{var}(Y) \qquad .$$

imposing a bound on the covariance, which in turn, implies that the correlation coefficient is bounded by the limits:

$$-1 \leq \rho(X, Y) \leq +1 \qquad , \tag{6.21}$$

with perfect correlation corresponding to the upper limit $\rho = 1$, and perfect anti-correlation to the lower limit $\rho = -1$.