15 Joint Distribution of Random Variables

Sometimes it is useful to consider more than one random variable at the same time, or to write a random variable as a combination of other random variables. In this section we develop some of this theory in the discrete case. All random variables mentioned are assumed to be discrete. Much of what follows is also true for continuous random variables (but with sums replaced by integrals and probability mass functions replaced by probability density functions).

Suppose we have two discrete random variables X and Y defined on the same sample space. Then the function

$$(x, y) \mapsto \mathbb{P}(X = x, Y = y)$$

is called the *joint probability mass function* of X and Y, or more simple the *joint distribution* of X and Y. (We use $\mathbb{P}(X = x, Y = y)$ to denote the probability of the event that $X = x$ *and* $Y = y$.) The joint distribution of X and Y can be presented as a table. The example we had in lectures gave the following:

		R			
		0	1	2	3
	0	0	3/35	6/35	1/35
B	1	2/35	12/35	6/35	0
	2	2/35	3/35	0	0

Here, for example, the top right entry means that $\mathbb{P}(R = 3, B = 0) = 1/35$.

Proposition 15.1 (Properties of a joint distribution)**.** *Suppose X and Y are discrete random variables. Then:*

(a) $\sum_x \sum_y \mathbb{P}(X = x, Y = y) = 1,$

(b) $\mathbb{P}(X = k) = \sum_y \mathbb{P}(X = k, Y = y)$ *for all* $k \in \text{Range}(X)$,

(c) $\mathbb{P}(Y = \ell) = \sum_x \mathbb{P}(X = x, Y = \ell)$ *for all* $\ell \in \text{Range}(Y)$,

where $\sum_x$ and $\sum_y$ run over all elements in the respective ranges of X and Y.

Proposition 15.1(a) is useful for checking that we have calculated the joint distribution correctly. Proposition 15.1(b) and (c) tell us how to calculate the probability mass functions of X and Y from their joint distribution.

Suppose X, Y are discrete random variables and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a real-valued function of two variables. Then $g(X, Y)$ is another random variable. Its expectation is given by

$$\mathbb{E}(g(X,Y)) = \sum_x \sum_y g(x,y)\mathbb{P}(X = x, Y = y).$$

Proposition 15.2. *Suppose X and Y are discrete random variables. Then*

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

More generally we have

Proposition 15.3. *Suppose $X_1, X_2, \ldots, X_n$ are discrete random variables and $c_1, c_2, \ldots, c_n \in \mathbb{R}$ are constants. Then*

$$E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \mathbb{E}(X_i).$$

Definition Two discrete random variables X and Y are *independent* if for all x and y we have

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

More generally, we say that the discrete random variables $X_1, X_2, \ldots, X_n$ are *independent* if for all $x_1, x_2, \ldots, x_n$ we have

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\ldots\mathbb{P}(X_n = x_n).$$

It is worth noting (because it is a frequent misconception) that we do *not* require the random variables to be independent in Propositions 15.2 and 15.3 However, if they are independent then we can say more.

Proposition 15.4. *Suppose X and Y are independent discrete random variables. Then:*

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

(b)) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

More generally we have

Proposition 15.5. *Suppose $X_1, X_2, \ldots, X_n$ are independent discrete random variables and $c_1, c_2, \ldots, c_n \in \mathbb{R}$ are constants. Then*

$$\mathrm{Var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 \mathrm{Var}(X_i).$$

The converses of Propositions 15.4 and 15.5 are false. For example, it is possible to have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ even when X and Y are not independent.

Application Propositions 15.3 and 15.5 give us an alternative way to calculate the expectation and variance of a binomial random variable. Suppose $X \sim \mathrm{Bin}(n,p)$. Then X is the number of successes in n independent Bernoulli(p) trials. For every $1 \leq i \leq n$ we define a random variable

$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial results in success} \\ 0 & \text{if the } i\text{th trial results in failure} \end{cases}$$

Then $X = X_1 + X_2 + \cdots + X_n$. Also, for every i we have $X_i \sim \mathrm{Bernoulli}(p)$ and so $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \cdots = \mathbb{E}(X_n) = p$. Proposition 15.3 now tells us that

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = np.$$

Furthermore, since the X_i are independent random variables and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_n) = p(1-p)$, Proposition 15.5 implies that

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = np(1-p).$$

The X_i in this argument are sometimes called *indicator variables.*

Definition Suppose X and Y are discrete random variables with $\mathbb{E}(X) = \mu_X$ and $\mathbb{E}(Y) = \mu_Y$. Then the *covariance of X and Y* is

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left([X-\mu_X][Y-\mu_Y]\right) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)\mathbb{P}(X=x, Y=y).$$

The *correlation coefficient of* X *and* Y is

$$\mathrm{corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

$\mathrm{Cov}(X,Y)$ and $\mathrm{corr}(X,Y)$ measure how far X and Y are from being independent. For example, if $\mathrm{Cov}(X,Y) > 0$ then

$$\mathbb{P}(X \geq \mu_X \,|\, Y \geq \mu_Y) \; > \; \mathbb{P}(X \geq \mu_X),$$

and if $\mathrm{Cov}(X,Y) < 0$ then

$$\mathbb{P}(X \geq \mu_X \,|\, Y \geq \mu_Y) \; < \; \mathbb{P}(X \geq \mu_X).$$

Proposition 15.6. *Suppose* X *and* Y *are discrete random variables. Then*

(a) $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

(b) If X *and* Y *are independent then* $\mathrm{Cov}(X,Y) = 0$.

Proposition 15.7. *Suppose* X *and* Y *are discrete random variables.*

(a) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$.

(b) If $a, b, c, d \in \mathbb{R}$ *are constants then*

$$\mathrm{Cov}(aX+b, cY+d) = ac\mathrm{Cov}(X,Y),$$

and

$$\mathrm{corr}(aX+b, cY+d) = \begin{cases} \mathrm{corr}(X,Y) & \text{if } ac > 0 \\ -\mathrm{corr}(X,Y) & \text{if } ac < 0 \end{cases}$$

(c) $-1 \leq \mathrm{corr}(X,Y) \leq 1$.

Parts (a), (b) were proved in lectures. The point of part (b) is that if we decide to scale one or both of the random variables by a linear transformation (for example measuring it in a different unit) then the covariance may change but the correlation coefficient will not (as long as $ac > 0$).

Proof of (c) (not examinable). We will need the following elementary result about quadratic polynomials.

Claim. *Suppose* $p, q, r \in \mathbb{R}$. *If* $pz^2 + 2qz + r \geq 0$ *for all* $z \in \mathbb{R}$ *then* $q^2 \leq pr$.

Proof We can solve the equation $pz^2 + qz + r = 0$ to obtain $z = (-q \pm \sqrt{q^2 - pr})/p$. Hence, if $q^2 > pr$, then $pz^2 + 2qz + r$ has two real roots. This would imply that $pz^2 + 2qz + r < 0$ for some $z \in \mathbb{R}$ and contradict the hypothesis of the claim. Thus we must have $q^2 \leq pr$.

We can now prove (c). Choose $z \in \mathbb{R}$. Parts (a) and (b) imply that

$$\begin{aligned}\mathrm{Var}(zX + Y) &= \mathrm{Var}(zX) + \mathrm{Var}(Y) + 2\mathrm{Cov}(zX, Y)\\ &= z^2\mathrm{Var}(X) + \mathrm{Var}(Y) + 2z\mathrm{Cov}(X, Y).\end{aligned}$$

Since variance cannot be negative we have

$$z^2\mathrm{Var}(X) + \mathrm{Var}(Y) + 2z\mathrm{Cov}(X, Y) \geq 0$$

for all $z \in \mathbb{R}$. We can now apply the Claim with $p = \mathrm{Var}(X)$, $q = \mathrm{Cov}(X, Y)$ and $r = \mathrm{Var}(Y)$. This gives $\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ and hence

$$\frac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} \leq 1.$$

Taking square roots we obtain $-1 \leq \mathrm{corr}(X, Y) \leq 1$.