Joint Distribution Example

Toss 4 fair coins, and consider the two r.v.'s:

- X = number of H's (successes) in the first 3 coin tosses
- Y = number of H's (successes) in the last 2 coin tosses

It is simplest for us to obtain the joint distribution of these two discrete r.v.'s from the sample space description of this game.

elementary outcome	X	Y
0000	0	0
0001	0	1
0010	1	1
0011	1	2
0100	1	0
0101	1	1
0110	2	1
0111	2	2
1000	1	0
1001	1	1
1010	2	1
1011	2	2
1100	2	0
1101	2	1
1110	3	1
1111	3	2

We obtain the joint probability mass function for (X, Y) below. The entries are $P(X = x, Y = y)$ for appropriate x, y.

$y\backslash x$	0	1	2	3	
0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	0	
1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	
2	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	

Marginal Distributions

Suppose (X, Y) has bivariate pmf $p_{X,Y}$. What is the distribution of X? Since X is discrete we want to find it pmf, that is the function p_X so that $p_X(x) = P(X = x)$ for all possible x.

Consider the event $\{X = x\}$. It is also the event

$$\begin{aligned} \{X = x\} &= \cup_y \{(X, Y) = (x, y)\} \\ &= \cup_{y: p_{X,Y}(x,y)>0} \{(X, Y) = (x, y)\} \end{aligned}$$

This last piece writes the event as a countable union of disjoint events, but more importantly simple events for which we have the probabilities of these in terms of the pmf $p_{X,Y}$. Thus

$$\begin{aligned} p_X(x) &= P(\{X = x\}) \\ &= P(\cup_{y: p_{X,Y}(x,y)>0} \{(X, Y) = (x, y)\}) \\ &= \sum_{y: p_{X,Y}(x,y)>0} P(\{(X, Y) = (x, y)\}) \\ &= \sum_{y: p_{X,Y}(x,y)>0} p_{X,Y}(x, y) \\ &= \sum_y p_{X,Y}(x, y) \end{aligned}$$

Thus using the Axioms of Probability we can find the distribution of X.

This is also called the *marginal* distribution of X, so that it reminds us that it is derived from a higher dimensional *joint* distribution.

One can similarly derive a formula for the distribution of Y, also referred to as the *marginal* distribution of Y. More generally for a trivatiate random vector (X_1, X_2, X_3) one has 3 different bivariate marginal distributions and also 3 different univariate marginal distributions. This notion extends to n dimensional random vectors. We will use this notion to study order statistics and sampling distributions of various statistics or estimators.

At home the student should find the marginal probability mass functions (pmf's) of both X and Y.

Are X and Y (statistically) independent? If they were then

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for all } x, y \ . \tag{1}$$

Notice that if we take $(x, y) = (0, 2)$ then $P(X = 0, Y = 2) = 0$ but $P(X = 0) = \frac{1}{8}$ and $P(Y = 2) = \frac{1}{4}$. Thus (1) is not satisfied for all x, y.

Aside : The r.v.s X, Y both are determined from the same set of 4 independent coin tosses, and have one coin in common, so it is not too surprising that they are dependent. However it may be that r.v.s are functions of the same underlying set of independent r.v.s and may still be statistically independent. Here is a a simple example, based on tossing two independent fair coins.

Pr	elementary outcome	W	Z
$\frac{1}{4}$	00	0	0
$\frac{1}{4}$	01	0	1
$\frac{1}{4}$	10	1	1
$\frac{1}{4}$	11	1	0

The r.v.s W, Z are both functions of (determined by) the two coin tossing r.v.s. The student should verify that W, Z are independent.

CDF

One actually needs to draw a 3D picture for this, but for the purposes of this handout we only tabulate the CDF at these selected points. In the rectangles bounded by the *jumps* the CDF is piecewise constant. Some continuous 3D plots of bivariate CDFs are shown in the text. Cdfs in dimension 1 are quite useful, and useful in dimensions larger than 1, but in this level of intermediate probability plays a lesser role. However it will be important in particular to see why the conditions for checking independence, especially for multi-variate continuous random variables can be written in an equivalent way in terms of pdfs. In particular this will mean that the notion of independence of r.v.s is the same whether we are dealing with continuous, discrete or r.v.s that are neither of these (partly continuous and partly discrete).

The second important place where the multivariate continuous CDF is important is to obtain, for invertible differentiable transformations, the change of variables formula that requires the notion of a Jacobian. Jacobian will not apply otherwise.

Continue the example from above.

This table only gives the values of the CDF at the jump points (in $\mathcal{R}^2$), the function being locally constant except at these jump points.

$y \backslash x$	0	1	2	3
2	$\frac{2}{16}$	$\frac{8}{16}$	$\frac{14}{16}$	$\frac{16}{16}$
1	$\frac{2}{16}$	$\frac{7}{16}$	$\frac{11}{16}$	$\frac{13}{16}$
0	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{4}{16}$

In general probabilities of rectangles can be obtained from the bivariate CDF. For a rectangle $(a_1, b_1] \times (a_2, b_2]$ (the student should sketch this) we

$$P((X,Y) \in (a_1,b_1] \times (a_2,b_2]) = F(a_2,b_2) - F(a_2,b_1) - F(a_1,b_2) + F(a_1,b_1) \ .$$

Conditional distributions:

Examples

What is the conditional pmf of Y given $X = 1$? Recall by the definition of conditional probability that

$$P(Y = y \mid X = 1) = \frac{P(\{Y = y\} \cap \{X = 1\})}{P(X = 1)} = \frac{P(X = 1, Y = y)}{P(X = 1)} \ .$$

Therefore

$$P(Y = y \mid X = 1) = \begin{cases} \frac{2}{6} & \text{if} \quad y = 0 \\ \frac{3}{6} & \text{if} \quad y = 1 \\ \frac{1}{6} & \text{if} \quad y = 2 \\ 0 & \quad\ \text{for all other } y \end{cases}$$

Similarly the student can verify that the conditional pmf of Y given that $X = 2$ is given by

$$P(Y = y \mid X = 2) = \begin{cases} \frac{1}{6} & \text{if} \quad y = 0 \\ \frac{3}{6} & \text{if} \quad y = 1 \\ \frac{2}{6} & \text{if} \quad y = 2 \\ 0 & \quad\ \text{for all other } y \end{cases}$$

Remark : *If* X and Y were (statistically) independent then since $P(X = 1) > 0$ and $P(X = 2) > 0$, then we would also have the conditional distribution of Y given $X = 1$ must be the same as the conditional distribution of Y given $X = 2$; that is

$$P(Y = y|X = 1) = P(Y = y|X = 2) \text{ for all } y\ .$$

This is due to the fact that *if* X and Y were independent then for all x such that $P(X = x) > 0$ (so the conditional distribution is defined), then we have $P(Y = y|X = x) = P(Y = y)$ for all y, that is the conditional distribution must be equal to the marginal distribution.

From the above calculation we see that these two conditional distributions are not equal and hence we can conclude that X and Y are not (statistically) independent, that is they are dependent.

Suppose that $W = X + Y$. Find the distribution of W.

First we need to decide if W is continuous or discrete, or neither. W is discrete since it can only take on a finite number of values. In fact it can take on at most $4 \times 3 = 12$ different values. (Aside : it actually takes on fewer distinct values that 12).

Thus we can specify the distribution of W in either of 2 equivalent forms, either its CDF or its pmf (or frequency function). We will determine its pmf. This means that we must specify a function, say p_W that gives the pmf.

Recall

$$\begin{aligned} p_W(w) &= P(W = w) \\ &= P((X,Y) \in \{(x,y) : x + y = w\}) \\ &= \sum_{(x,y):x+y=w} P(X = x, Y = y) \end{aligned}$$

We find

$$\begin{aligned} P(W=0) &= P((X,Y) = (0,0)) = \frac{1}{16} \\ P(W=1) &= P((X,Y) = (0,1)) + P((X,Y) = (1,0)) \\ &= \frac{1}{16} + \frac{2}{16} = \frac{3}{16} \\ P(W=2) &= P((X,Y) = (0,2)) + P((X,Y) = (1,1)) + P((X,Y) = (2,0)) \\ &= 0 + \frac{3}{16} + \frac{1}{16} = \frac{1}{4} \end{aligned}$$

and so on.

There is no real shortcut here, we must find for each w the set $\{(x,y) : x + y = w\}$. In a general transformation case $W = h(X,Y)$ we will have to find the sets $\{(x,y) : h(x,y) = w\}$ for all w.

Completing the above calculation we find

$$p_W(w) = \begin{cases} \frac{1}{16} & \text{if} \quad w = 0 \\ \frac{3}{16} & \text{if} \quad w = 1 \\ \frac{4}{16} & \text{if} \quad w = 2 \\ \frac{4}{16} & \text{if} \quad w = 3 \\ \frac{3}{16} & \text{if} \quad w = 4 \\ \frac{1}{16} & \text{if} \quad w = 5 \\ 0 & \text{otherwise} \end{cases}$$

For a general function g, as long as $g(x,y)$ is defined for all (x,y) in the support of $p_{X,Y}$ then one can obtain the distribution of $W = g(X,Y)$. In our example a function that will not work is

$g(x, y) = x/y$. In this case since $p_{X,Y}(1, 0) > 0$ then $W = 1/0$ would happen with positive probability. This would mean that W is not real valued and hence is not a random variable.

The student should try to see how the Axioms of Probability will give a formula for the distribution of $W = g(X, Y)$, that is the distribution in terms of the given pmf $p_{X,Y}$.