

5.1: Joint Distributions of Discrete Random Variables

In this chapter we consider two or more random variables defined on the same sample space and discuss how to model the probability distribution of the random variables *jointly*. We will begin with the discrete case by looking at the joint probability mass function for two discrete random variables. In the following section, we will consider continuous random variables.

Definition 5.1.1

If discrete random variables X and Y are defined on the same sample space S , then their **joint probability mass function (joint pmf)** is given by

$$p(x, y) = P(X = x \text{ and } Y = y),$$

where (x, y) is a pair of possible values for the pair of random variables (X, Y) , and $p(x, y)$ satisfies the following conditions:

- $0 \leq p(x, y) \leq 1$
- $\sum_{(x,y)} p(x, y) = 1$
- $P((X, Y) \in A) = \sum_{(x,y) \in A} p(x, y)$

Note that conditions #1 and #2 in Definition 5.1.1 are required for $p(x, y)$ to be a valid joint pmf, while the third condition tells us how to use the joint pmf to find probabilities for the pair of random variables (X, Y) .

In the discrete case, we can obtain the **joint cumulative distribution function (joint cdf)** of X and Y by *summing* the joint pmf:

$$F(x, y) = P(X \leq x \text{ and } Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j),$$

where x_i denotes possible values of X and y_j denotes possible values of Y . From the joint pmf, we can also obtain the individual probability distributions of X and Y *separately* as shown in the next definition.

Definition 5.1.2

Suppose that discrete random variables X and Y have joint pmf $p(x, y)$. Let $x_1, x_2, \dots, x_i, \dots$ denote the possible values of X , and let $y_1, y_2, \dots, y_j, \dots$ denote the possible values of Y . The **marginal probability mass functions (marginal pmf's)** of X and Y are respectively given by the following:

$$p_X(x) = \sum_j p(x, y_j) \quad (\text{fix a value of } X \text{ and sum over possible values of } Y)$$

$$p_Y(y) = \sum_i p(x_i, y) \quad (\text{fix a value of } Y \text{ and sum over possible values of } X)$$

[Link to Video: Overview of Definitions 5.1.1 & 5.1.2](#)

Example 5.1.1

Consider again the probability experiment of [Example 3.3.2](#), where we toss a fair coin three times and record the sequence of heads (h) and tails (t). Again, we let random variable X denote the number of heads obtained. We also let random variable Y denote the winnings earned in a single play of a game with the following rules, based on the outcomes of the probability experiment (this is the same as [Example 3.6.2](#)):

- player wins \$1 if first h occurs on the first toss
- player wins \$2 if first h occurs on the second toss

- player wins \$3 if first h occurs on the third toss
- player loses \$1 if no h occur

Note that the possible values of X are $x = 0, 1, 2, 3$, and the possible values of Y are $y = -1, 1, 2, 3$. We represent the joint pmf using a table:

Table 1: joint pmf of X and Y

$p(x, y)$	X			
Y	0	1	2	3
-1	1/8	0	0	0
1	0	1/8	2/8	1/8
2	0	1/8	1/8	0
3	0	1/8	0	0

The values in Table 1 give the values of $p(x, y)$. For example, consider $p(0, -1)$:

$$p(0, -1) = P(X = 0 \text{ and } Y = -1) = P(ttt) = \frac{1}{8}.$$

Since the outcomes are equally likely, the values of $p(x, y)$ are found by counting the number of outcomes in the sample space S that result in the specified values of the random variables, and then dividing by 8, the total number of outcomes in S . The sample space is given below, color coded to help explain the values of $p(x, y)$:

$$S = \{ttt, htt, tht, tth, hht, hth, thh, hhh\}$$

Given the joint pmf, we can now find the marginal pmf's. Note that the marginal pmf for X is found by computing sums of the *columns* in Table 1, and the marginal pmf for Y corresponds to the *row* sums. (Note that we found the pmf for X in [Example 3.3.2](#) as well, it is a binomial random variable. We also found the pmf for Y in [Example 3.6.2](#).)

Table 2: marginal pmf's for X and Y

x	$p_X(x)$	y	$p_Y(y)$
0	1/8	-1	1/8
1	3/8	1	1/2
2	3/8	2	1/4
3	1/8	3	1/8

Finally, we can find the joint cdf for X and Y by summing over values of the joint frequency function. For example, consider $F(1, 1)$:

$$F(1, 1) = P(X \leq 1 \text{ and } Y \leq 1) = \sum_{x \leq 1} \sum_{y \leq 1} p(x, y) = p(0, -1) + p(0, 1) + p(-1, 1) + p(1, 1) = \frac{1}{4}$$

Again, we can represent the joint cdf using a table:

Table 3: joint cdf of X and Y

$F(x, y)$	X			
Y	0	1	2	3
-1	1/8	1/8	1/8	1/8
1	1/8	1/4	1/2	5/8
2	1/8	3/8	3/4	7/8
3	1/8	1/2	7/8	1

[Link to Video: Walkthrough of Example 5.1.1](#)

Expectations of Functions of Jointly Distributed Discrete Random Variables

We now look at taking the expectation of jointly distributed discrete random variables. Because expected values are defined for a single quantity, we will actually define the expected value of a combination of the pair of random variables, i.e., we look at the expected value of a function applied to (X, Y) .

Theorem 5.1.1

Suppose that X and Y are jointly distributed discrete random variables with joint pmf $p(x, y)$.

If $g(X, Y)$ is a function of these two random variables, then its expected value is given by the following:

$$E[g(X, Y)] = \sum_{(x, y)} g(x, y) p(x, y).$$

Example 5.1.2

Consider again the discrete random variables we defined in [Example 5.1.1](#) with joint pmf given in [Table 1](#). We will find the expected value of three different functions applied to (X, Y) .

1. First, we define $g(x, y) = xy$, and compute the expected value of XY :

$$\begin{aligned}
 E[XY] &= \sum_{(x, y)} xy \cdot p(x, y) = (0)(-1) \left(\frac{1}{8} \right) \\
 &\quad + (1)(1) \left(\frac{1}{8} \right) + (2)(1) \left(\frac{2}{8} \right) + (3)(1) \left(\frac{1}{8} \right) \\
 &\quad + (1)(2) \left(\frac{1}{8} \right) + (2)(2) \left(\frac{1}{8} \right) \\
 &\quad + (1)(3) \left(\frac{1}{8} \right) \\
 &= \frac{17}{8} = 2.125
 \end{aligned}$$

2. Next, we define $g(x) = x$, and compute the expected value of X :

$$\begin{aligned}
 E[X] &= \sum_{(x,y)} \sum x \cdot p(x,y) = (0) \left(\frac{1}{8}\right) \\
 &\quad + (1) \left(\frac{1}{8}\right) + (2) \left(\frac{2}{8}\right) + (3) \left(\frac{1}{8}\right) \\
 &\quad + (1) \left(\frac{1}{8}\right) + (2) \left(\frac{1}{8}\right) \\
 &\quad + (1) \left(\frac{1}{8}\right) \\
 &= \frac{12}{8} = 1.5
 \end{aligned}$$

Recall that $X \sim \text{binomial}(n = 3, p = 0.5)$, and that the expected value of a binomial random variable is given by np . Thus, we can verify the expected value of X that we calculated above using Theorem 5.1.1 using this fact for binomial distributions: $E[X] = np = 3(0.5) = 1.5$.

3. Lastly, we define $g(x, y) = y$, and calculate the expected value of Y :

$$\begin{aligned}
 E[Y] &= \sum_{(x,y)} \sum y \cdot p(x,y) = (-1) \left(\frac{1}{8}\right) \\
 &\quad + (1) \left(\frac{1}{8}\right) + (1) \left(\frac{2}{8}\right) + (1) \left(\frac{1}{8}\right) \\
 &\quad + (2) \left(\frac{1}{8}\right) + (2) \left(\frac{1}{8}\right) \\
 &\quad + (3) \left(\frac{1}{8}\right) \\
 &= \frac{10}{8} = 1.25
 \end{aligned}$$

Again, we can verify this result by reviewing the calculations done in [Example 3.6.2](#).

[Link to Video: Walkthrough of Example 5.1.2](#)

Independent Random Variables

In some cases, the probability distribution of one random variable will not be affected by the distribution of another random variable defined on the same sample space. In those cases, the joint distribution functions have a very simple form, and we refer to the random variables as independent.

Definition 5.1.3

Discrete random variables X_1, X_2, \dots, X_n are **independent** if the joint pmf factors into a product of the marginal pmf's:

$$p(x_1, x_2, \dots, x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdots p_{X_n}(x_n). \quad (5.1.1)$$

It is equivalent to check that this condition holds for the cumulative distribution functions.

Recall the definition of independent *events* ([Definition 2.3.2](#)): A and B are independent events if $P(A \cap B) = P(A) P(B)$. This is the basis for the definition of independent random variables because we can write the pmf's in Equation 5.1.1 in terms of events as follows:

$$p(x, y) = P(X = x \text{ and } Y = y) = P(\{X = x\} \cap \{Y = y\}) = P(X = x)P(Y = y) = p_X(x)p_Y(y)$$

In the above, we use the idea that if X and Y are independent, then the event that X takes on a given value x is independent of the event that Y takes the value y .

Example 5.1.3

Consider yet again the discrete random variables defined in [Example 5.1.1](#). According to the definition, X and Y are independent if

$$p(x, y) = p_X(x) \cdot p_Y(y),$$

for all pairs (x, y) . Recall that the joint pmf for (X, Y) is given in [Table 1](#) and that the marginal pmf's for X and Y are given in [Table 2](#). Note that, for $(x, y) = (0, -1)$, we have the following

$$p(0, -1) = \frac{1}{8}, \quad p_X(0) = \frac{1}{8}, \quad p_Y(-1) = \frac{1}{8} \quad \Rightarrow \quad p(0, -1) \neq p_X(0) \cdot p_Y(-1).$$

Thus, X and Y are **not** independent, or in other words, X and Y are **dependent**. This should make sense given the definition of X and Y . The winnings earned depend on the number of heads obtained. So the probabilities assigned to the values of Y will be affected by the values of X .

We also have the following very useful theorem about the expected value of a product of independent random variables, which is simply given by the product of the expected values for the individual random variables.

Theorem 5.1.2

If X and Y are independent random variables, then $E[XY] = E[X] E[Y]$.

Proof

Assume X and Y are independent random variables. If we let $p(x, y)$ denote the joint pmf of (X, Y) , then, by [Definition 5.1.3](#), $p(x, y) = p_X(x)p_Y(y)$, for all pairs (x, y) . Using this fact and [Theorem 5.1.1](#), we have

$$\begin{aligned} E[XY] &= \sum_{(x,y)} xy \cdot p(x, y) = \sum_{(x,y)} xy \cdot p_X(x)p_Y(y) \\ &= \sum_x \sum_y xyp(x)p_Y(y) = \sum_x xp_X(x) \left(\sum_y p_Y(y) \right) = \sum_x xp_X(x)E[Y] \\ &= E[Y] \sum_x xp_X(x) = E[Y] E[X]. \end{aligned}$$

Theorem 5.1.2 can be used to show that two random variables are **not** independent: if $E[XY] \neq E[X] E[Y]$, then X and Y **cannot** be independent. However, beware using Theorem 5.1.2 to show that random variables are independent. Note that Theorem 5.1.2 **assumes** that X and Y are independent and then the property about the expected value follows. The other direction does not hold. In other words, if $E[XY] = E[X] E[Y]$, then X and Y **may or may not** be independent.

[Link to Video: Independent Random Variables](#)

This page titled [5.1: Joint Distributions of Discrete Random Variables](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Kristin Kuter](#).