# Chapter 5

# Multivariate Distributions

- **Learning objectives:**
  - understand the definition of bivariate and multivariate distributions in both the discrete and continuous setting
  - be able to compute and use marginal distributions in both the discrete and continuous setting
  - be able to compute and use conditional distributions in both the discrete and continuous setting
  - be able to compute expected value and variance in the bivariate setting
  - be able to compute the covariance between two random variables

In the last two chapters, our focus was on the distribution of a single random variable. In many situations, however, we are considering several random variables simultaneously. Some basic examples include the following:

- we randomly choose an individual and measure $X$ = height and $Y$ = weight
- we randomly choose a patient and measure $X$ = blood pressure, $Y$ = weight, and $Z$ = heart rate
- three individuals each toss 10 coins and $X_i$ = number of heads obtained by person $i$ for $i = 1, 2, 3$

In each of the above examples, two or more values are being randomly selected jointly. Our goal is to determine the distribution of pairs or triples of values that could occur.

We start with the case in which the random variables are discrete because it is much easier to visualize than the continuous case.

## 5.1 Discrete Bivariate and Multivariate Distributions

**Definition 5.1.1**

Let $X$ and $Y$ be discrete random variables that occur jointly. That is, there is a single random action that produces a pair of values $(x, y)$. The **joint pmf for $X$ and $Y$** is defined as follows:

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

This is called the joint pmf because it can be used to answer questions that are asked "jointly" about $X$ and $Y$.

This definition can be extended in the obvious way for cases in which more than two random variables are being considered. For instance, if $X$, $Y$, and $Z$ are discrete random variables that occur jointly, their joint pmf is given by

$$f_{X,Y,Z}(x,y,z) = P(X = x \cap Y = y \cap Z = z).$$

As was the case in Chapter 3 when we considered a single random variable, the joint pmf can be given in tabular or functional format. Throughout this section and those that follow, we will sometimes refer to the four joint pmfs listed below and labeled as $A$, $B$, $C$, or $D$.

**Distribution $A$:**

| $X \backslash Y$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0.01 | 0.06 | 0.08 |
| 1 | 0.05 | 0.02 | 0.03 |
| 2 | 0.06 | 0.11 | 0.04 |
| 3 | 0.20 | 0.07 | 0.05 |
| 4 | 0.07 | 0.12 | 0.03 |

**Distribution $B$:** the joint pmf of the discrete random variables $X$ and $Y$ is given by

$$f_{X,Y}(x,y) = \frac{x^2}{21y} \quad \text{for} \quad x = 1,2,3 \quad \text{and} \quad y = 1,2$$

**Distribution $C$:**

$Z = 2$

| $X \backslash Y$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 0 | 0.01 | 0.02 | 0.01 | 0.03 |
| 1 | 0.02 | 0.03 | 0.01 | 0.02 |
| 2 | 0.03 | 0.04 | 0.02 | 0.01 |
| 3 | 0.04 | 0.05 | 0.02 | 0.01 |

$Z = 4$

| $X \backslash Y$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 0 | 0.03 | 0.02 | 0.03 | 0.01 |
| 1 | 0.05 | 0.04 | 0.07 | 0.02 |
| 2 | 0.04 | 0.06 | 0.02 | 0.03 |
| 3 | 0.01 | 0.08 | 0.08 | 0.04 |

**Distribution $D$:** the joint pmf of the discrete random variables $X$, $Y$, and $Z$ is given by

$$f_{X,Y,Z}(x,y,z) = \frac{1}{504}(x + y^2 + z^3) \quad \text{for} \quad x = 1,2,3, \quad y = 1,2,3 \quad \text{and} \quad z = 1,2,3$$

**Definition 5.1.2**

The **support** of the joint random variables $X$ and $Y$ is the set of ordered pairs $(x,y)$ for which $f_{X,Y}(x,y) > 0$.

The support is defined similarly for more than two jointly distributed random variables.

**Example 5.1.1**
Determine $P(X = 2 \cap Y = 1)$ using distribution $A$.

$$f(2,1) = 0.11$$

We see that when the joint pmf is given in tabular form, we can answer any joint probability question with little effort just by finding the appropriate cell of the table. Recall that the functional notation for the joint pmf is $f(x,y) = f_{X,Y}(x,y) = P(X = x \cap Y = y)$. We use the subscripts in our notation in cases where we feel the need to clarify which random variables are being considered. If no confusion will arise, we sometimes omit the subscripts.

**Example 5.1.2**
Determine $P(X = 3, Y = 2)$ using distribution $A$.

$$f(3,2) = 0.05$$

**Example 5.1.3**
Determine $f_{X,Y}(0,2)$ and $f_{X,Y}(3,3)$ using distribution $A$.

$f_{X,Y}(0,2) = 0.08$

$f_{X,Y}(3,3) = 0$ since $Y = 3$ is not possible and so $(3,3)$ is not in support.

**Example 5.1.4**
Determine $P(X = 3, Y = 2)$ using distribution $B$.

$$f(3,2) = \frac{3^2}{21(2)} = \frac{9}{42}$$

**Example 5.1.5**
Determine $f_{X,Y}(1,2)$ and $f_{X,Y}(3,1)$ using distribution $B$.

$$f(1,2) = \frac{1^2}{21(2)} = \frac{1}{42}$$

$$f(3,1) = \frac{3^2}{21(1)} = \frac{9}{21}$$

**Example 5.1.6**
Determine $P(X = 3, Y = 1, Z = 2)$ using distribution $C$.

$$f(3,1,2) = 0.04$$

**Example 5.1.7**
Determine $f_{X,Y,Z}(1,3,4)$ using distribution $C$.

$$f(1,3,4) = 0.07$$

**Example 5.1.8**
Determine $P(X = 1, Y = 1, Z = 1)$ using distribution $D$.

$$f(1,1,1) = \frac{1}{504}(1 + 1^2 + 1^3) = \frac{3}{504}$$

**Example 5.1.9**
Determine $f_{X,Y,Z}(1,2,3)$ using distribution $D$.

$$f(1,2,3) = \frac{1}{504}(1 + 2^2 + 3^3) = \frac{32}{504}$$

**Example 5.1.10**

Determine the value of $c$ that makes the table below a valid joint pmf:

| $X \backslash Y$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | $c$ | 0.06 | 0.08 |
| 1 | 0.05 | 0.02 | 0.03 |
| 2 | 0.06 | 0.03 | 0.04 |
| 3 | 0.12 | 0.07 | 0.05 |
| 4 | 0.07 | 0.12 | 0.03 |
| | 0.30+c | 0.30 | 0.23 |

Need $\sum_y \sum_x f(x,y) = 1 \Rightarrow \boxed{c = 0.17}$

**Example 5.1.11**

Determine the value of $c$ that makes the function below a valid joint pmf:

$$f_{X,Y}(x,y) = c(2x + y^2) \text{ for } x = 1,2,3 \text{ and } y = 1,2,3$$

| x | y | $c(2x+y^2)$ |
|---|---|---|
| 1 | 1 | 3c |
| 1 | 2 | 6c |
| 1 | 3 | 11c |
| 2 | 1 | 5c |
| 2 | 2 | 8c |
| 2 | 3 | 13c |
| 3 | 1 | 7c |
| 3 | 2 | 10c |
| 3 | 3 | 15c |

$\Sigma = 78c \overset{\text{Must}}{=} 1$

$\Rightarrow \boxed{c = \frac{1}{78}}$

## 5.2 Discrete Marginal Distributions

In the previous section, we saw how to use a joint pmf to answer joint probability questions about about $X$ and $Y$ (or about $X$, $Y$, and $Z$ in the three variable case). In this setting, we are often interested in answering probability questions involving just one of the random variables. Such questions require the marginal distributions.