Chapter 3

Random Variables and Distribution Functions

3.1 Discrete and Continuous Random Variables

In Chapter 1 of this text we explored the concepts of outcomes, sample spaces and events. In particular, we saw that an event is a subset of the sample space and we further defined an event space as a partition of all the outcomes in a sample space into a set of mutually exclusive and collectively exhaustive events. Each outcome belongs to one subset, i.e., one event, and only one subset/event, of the event space. In Chapter 2, we learned how to count the number of outcomes that constitute events. This provided us with the wherewithal to associate a value to the probability of a given event occurring. These counting techniques proved to be particularly useful when the sample space consisted of more outcomes than could be conveniently enumerated.

Now in Chapter 3 we move away from working directly with the individual outcomes in a sample space and will instead focus on broader groupings of these outcomes, specifically, as introduced in Chapter 1, the events that constitute event spaces. In many probability experiments this is all we need: when throwing two dice, it may be that our only interest is the number of spots that appear, not a detailed description of all possible outcomes; when tossing a bunch of coins, we are possibly only concerned about the number of heads that appear. As we shall see below, grouping outcomes by the number of spots that can appear has the effect of partitioning the sample space — each possible number of spots from 2 through 12 defines an event, namely the event that consists of all outcomes that produce that number of spots. The same is true when grouping coins according to the number of heads obtained — for each different number of heads that can be observed, there corresponds an event in the event space. The lessons learned in Chapter 2 permit us to handle much more complex scenarios.

This is where random variables come into play. A random variable partitions the outcomes in a sample space into an event space. All the outcomes in a given event of that event space are mapped by the random variable into the same real number. For
example, when a random variable is used to count the number of spots that appear when
two dice are thrown, the sample space is partitioned by the random variable into 11
events: the outcome \((i, j)\), where \(i\) is the number of spots shown by the first die and \(j\) the
number shown by the second, is mapped into the integer \(i + j\). The set of outcomes that
are mapped by the random variable into the real number \(k \in \{2, 3, \ldots, 12\}\) constitutes
an event in the event space. It is the event that consists of all outcomes in which the
sum of spots is equal to \(k\).

It follows then, that random variables, frequently abbreviated as RV, define func-
tions, with domain and range, rather than variables. Simply put, a random variable is
a function whose domain is a sample space \(\Omega\) and whose range is a subset of the real
numbers, \(\mathbb{R}\). Each elementary event \(\omega \in \Omega\) is mapped by the random variable into \(\mathbb{R}\). Random variables, usually denoted by uppercase latin letters, assign real numbers to
the outcomes in their sample space.

**Example 3.1** Consider a probability experiment that consists of throwing two dice. There
are 36 possible outcomes which may be represented by the pairs \((i, j)\); \(i = 1, 2, \ldots, 6\), \(j = 1, 2, \ldots, 6\). These are the 36 elements of the sample space. It is possible to associate many
random variables having this domain. One possibility is the random variable \(X\), defined
explicitly by enumeration as

\[
\begin{align*}
X(1, 1) &= 2; \\
X(1, 2) &= X(2, 1) = 3; \\
X(1, 3) &= X(2, 2) = X(3, 1) = 4; \\
X(1, 4) &= X(2, 3) = X(3, 2) = X(4, 1) = 5; \\
X(1, 5) &= X(2, 4) = X(3, 3) = X(4, 2) = X(5, 1) = 6; \\
X(1, 6) &= X(2, 5) = X(3, 4) = X(4, 3) = X(5, 2) = X(6, 1) = 7; \\
X(2, 6) &= X(3, 5) = X(4, 4) = X(5, 3) = X(6, 2) = 8; \\
X(3, 6) &= X(4, 5) = X(5, 4) = X(6, 3) = 9; \\
X(4, 6) &= X(5, 5) = X(6, 4) = 10; \\
X(5, 6) &= X(6, 5) = 11; \\
X(6, 6) &= 12.
\end{align*}
\]

Thus the random variable \(X\) maps the outcome \((1, 1)\) into the real number \(2\); the
outcomes \((1, 2)\) and \((2, 1)\) are both mapped into the real number \(3\), and so on. It is
apparent that the random variable \(X\) represents the sum of the spots obtained when two
dice are thrown simultaneously. As indicated in Figure 3.1, the domain of this random
variable is the 36 elements of the sample space and its range is the set of integers from 2
through 12. The event space created by the random variable \(X\) consists of 11 events that
are mutually exclusive and collectively exhaustive. These events are distinguished according
to the particular value taken by the function. The function (random variable) \(X\) assigns the
value 4 to the event \(\{(1, 3), (2, 2), (3, 1)\}\); the value 5 to the event \(\{(5, 6), (6, 5)\}\), and so
on.
Different random variables may be defined on the same sample space. They have
the same domain, but their range can be different.

Example 3.2 Consider a random variable $Y$ whose domain is the same sample space as
that of $X$ but which is defined as the result obtained when the number of spots shown by
the second die is subtracted from the number shown by the first. In this case we have

\[
\begin{align*}
Y(1,1) &= Y(2,2) = Y(3,3) = Y(4,4) = Y(5,5) = Y(6,6) = 0; \\
Y(2,1) &= Y(3,2) = Y(4,3) = Y(5,4) = Y(6,5) = 1; \\
Y(1,2) &= Y(2,3) = Y(3,4) = Y(4,5) = Y(5,6) = -1; \\
Y(3,1) &= Y(4,2) = Y(5,3) = Y(6,4) = 2; \\
Y(1,3) &= Y(2,4) = Y(3,5) = Y(4,6) = -2; \\
Y(4,1) &= Y(5,2) = Y(6,3) = 3; \\
Y(1,4) &= Y(2,5) = Y(3,6) = -3; \\
Y(5,1) &= Y(6,2) = 4; \\
Y(1,5) &= Y(2,6) = -4; \\
Y(6,1) &= 5; \\
Y(1,6) &= -5.
\end{align*}
\]

Observe that the random variable $Y$ has the same domain as $X$ but the range of $Y$
is the set of integers between $-5$ and $+5$ inclusive. As illustrated in Figure 3.2, $Y$ also
partitions the sample space into 11 subsets, but not the same 11 as those obtained with
$X$.

Figure 3.2 tries to capture the two different partitions on the same diagram: the
partition due to the random variable $X$ is displayed horizontally and its 11 events are
enclosed by dashed lines; the partition due to $Y$ is displayed vertically and its 11 events
are enclosed by dotted lines.

Yet a third random variable, $Z$, may be defined as the *absolute* value of the difference
between the spots obtained on the two dice. Again, $Z$ has the same domain as $X$ and $Y$,
but now its range is the set of integers between 0 and 5 inclusive. This random variable