4 Outcomes, Events, and Random Variables

Probability models can be applied to any situation in which there are multiple potential outcomes and there is uncertainty about which outcome will occur.

4.1 Outcomes

- Due to the wide variety of types of random phenomena, an **outcome** can be virtually anything
- In particular, an outcome does *not* have to be a number.
- The **sample space**, denoted Ω (the uppercase Greek letter "omega"), is the set of all possible outcomes of a random phenomenon. An **outcome**, denoted ω (the lowercase Greek letter "omega"), is an element of the sample space: $\omega \in \Omega$.
- Mathematically, the sample space Ω is a *set* containing all possible outcomes, while an individual outcome ω is a *point* or *element* in Ω.
- A random phenomenon is modeled by a *single* sample space, with respect to which all objects (events, random variables) are defined. Whenever possible, a sample space outcome should be defined to provide the maximum amount of information about the outcome of random phenomenon.
- In practice we rarely enumerate the sample space as we'll for some of the examples in this class. Nonetheless, there is always some underlying sample space corresponding to all possible outcomes of the random phenomenon.

Example 4.1. Roll a four-sided die twice, and record the result of each roll in sequence as an ordered pair. For example, the outcome $(3, 1)$ represents a 3 on the first roll and a 1 on the second; this is not the same outcome as $(1, 3)$.

1. How many possible outcomes are there? Identify the sample space.

2. We might be interested in the sum of the two dice. Explain why it is still advantageous to define the sample space as in the previous part, rather than as $\Omega = \{2, \dots, 8\}$.

- **Multiplication principle for counting** Suppose that stage 1 of a process can be completed in any one of n_1 ways. Further, suppose that for each way of completing the stage 1, stage 2 can be completed in any one of n_2 ways. Then the two-stage process can be completed in any one of $n_1 \times n_2$ ways.
- This rule extends naturally to a ℓ-stage process, which can then be completed in any one of $n_1 \times n_2 \times n_3 \times \cdots \times n_\ell$ ways.
- It is not important whether there is a "first" or "second" stage. What is important is that there are distinct stages, each with its own number of "choices".

Example 4.2. The "matching problem" is one well known probability problem. The general setup involves n distincts objects labeled $1, \dots, n$ which are placed in n distinct boxes labeled $1, \dots, n$, with exactly one object placed in each box.

1. Consider the matching problem with $n = 3$. Label the objects 1, 2, 3, and the spots 1, 2, 3, with spot 1 the correct spot for object 1, etc. Specify an appropriate definition of an outcome, determine the number of outcomes, and specify the sample space.

2. For a general n, how many possible outcomes are there?

Example 4.3. Regina and Cady plan to meet for lunch They will definitely arrive between noon and 1, but their exact arrival times are uncertain. Rather than dealing with clock time, it is helpful to represent noon as time 0 and measure time as minutes after noon, including fractions of a minute, so that arrival times take values in the continuous interval [0, 60].

Specify an appropriate definition of an outcome and draw a picture representing the sample space.

4.2 Events

- An event is something that could happen or might be true.
- An *event* is a collection of outcomes that satisfy some criteria.
- Mathematically, an **event** A is a *subset* of the sample space: $A \subseteq \Omega$.
- Events are typically denoted with capital letters near the start of the alphabet, with or without subscripts (e.g. A, B, C, A_1, A_2). Events can be composed from others using basic set operations like unions ($A \cup B$), intersections ($A \cap B$), and complements (A^c).
  - Read A^c as "not A".
  - Read $A \cap B$ as "A and B"
  - Read $A \cup B$ as "A or B". Note that unions ($\cup$, "or") are always inclusive. $A \cup B$ occurs if A occurs but B does not, B occurs but A does not, or both A and B occur.
- A collection of events $A_1, A_2, ...$ are **disjoint** (a.k.a. mutually exclusive) if $A_i \cap A_j = \emptyset$ for all $i \neq j$. That is, multiple events are disjoint if none of the events have any outcomes in common.
- If the sample space outcomes are represented by rows in a spreadsheet, then an event is a subset of rows that satisfies some criteria

Example 4.4. Matching problem. For $n = 3$, objects labeled 1, 2, 3, are placed at random in spots labeled 1, 2, 3, with spot 1 the correct spot for object 1, etc. Using the sample space from Example 4.2, identify the following events.

1. B, the event that no objects are put in the correct spot.

2. In words, what does B^c represent?

3. A, the event that all objects are put in the correct spot.

4. C, the event that exactly 2 objects are put in the correct spot.

5. A_3, the event that object 3 is put (correctly) in spot 3.

6. For a general n let A the event that all objects are put in the correct spot, let B the event that no objects are put in the correct spot, and let A_i be the event that object i is put (correctly) in spot i, $i = 1, \ldots, n$. What is the relationship between A and $A_1, \ldots, A_n$? What is the relationship between B and $A_1, \ldots, A_n$?

Example 4.5. Using the sample space from Example 4.3, identify the following events using pictures.

1. Identify A, the event that Regina arrives after Cady.

2. Identify B, the event that either Regina or Cady arrives before 12:30.

3. Identify C, the event that they arrive within 15 minutes of each other.

4. Identify D, the event that Regina arrives before 12:24.

4.3 Random variables

- Roughly, a *random variable* assigns a number measuring some quantity of interest to each outcome of a random phenomenon.
- Mathematically, a **random variable (RV)** X is a *function* that takes an outcome in the sample space as input and returns a real number as output
- The random variable itself is typically denoted with a capital letter (X); possible values of that random variable are denoted with lower case letters (x).
  - Think of the capital letter X as a label standing in for a formula like "the number of heads in 4 flips of a coin" and
  - x as a dummy variable standing in for a particular value like 3.
- **Discrete random variables** take at most countably many possible values (e.g., $0, 1, 2, ...$). They are often counting variables (e.g., the number of Heads in 10 coin flips).
- **Continuous random variables** can take any real value in some interval (e.g., $[0, 1]$, $[0, \infty)$, $(-\infty, \infty)$.). That is, continuous random variables can take uncountably many different values. Continuous random variables are often measurement variables (e.g., height, weight, income).
- A function of a random variable is also a random variable: if X is a RV then so is $g(X)$
- Sums and products, etc., of random variables *defined on the same sample space* are random variables. IfX and Y are RVs defined on the same sample space then so are $X + Y$, $X - Y$, XY
- If the sample space outcomes are represented by rows in a spreadsheet, then random variables correspond to columns.
- Expressions like $X = x$ or $\{X = x\}$ represent *events*: for which outcomes is the value of the random variable X equal to the value x

Example 4.6. Roll a four-sided die twice, and record the result of each roll in sequence. Recall the sample space from Example 4.1. Let X be the sum of the two dice, and let Y be the larger of the two rolls (or the common value if both rolls are the same).

Outcome	X	Y
(1, 1)		
(1, 2)		
(1, 3)		
(1, 4)		
(2, 1)		
(2, 2)		
(2, 3)		
(2, 4)		
(3, 1)		
(3, 2)		
(3, 3)		
(3, 4)		
(4, 1)		
(4, 2)		
(4, 3)		
(4, 4)		

1. Construct a table identifying the values of X and Y for each outcome in the sample space.

2. Identify the possible values of X.

3. Identify the possible values of Y.

4. Identify the possible values of the pair (X, Y).

5. Identify $\{Y = 1\}$.

6. Identify $\{Y = 2\}$.

7. Identify $\{Y = 3\}$.

8. Identify $\{Y = 4\}$.

9. Identify $\{X \leq 4\}$.

10. Identify $\{X = 4, Y = 3\}$

Example 4.7. Matching problem. For $n = 3$ objects labeled 1, 2, 3, are placed at random in spots labeled 1, 2, 3, with spot 1 the correct spot for object 1, etc. Recall the sample space from Example 4.2. Let the random variable X count the number of objects that are put in the correct spot. Let I_1 be equal to 1 if object 1 is placed (correctly) in spot 1, and define I_2, I_3 similarly.

Outcome	X	I_1	I_2	I_3
123				
132				
213				
231				
312				
321				

1. Construct a table identifying the value of X, I_1, I_2, I_3 for each outcome in the sample space.
2. Identify the possible values of X.

3. What is the relationship between I_3 and event A_3 from Example 4.4?

4. How can you express X in terms of I_1, I_2, I_3?

- The **indicator (a.k.a., Bernoulli) random variable** corresponding to event A is equal to 1 if A occurs and 0 otherwise
- Indicator random variables can be used for incremental counting and are often useful in problems involving "find the expected number of"

Example 4.8. Regina and Cady will definitely arrive between noon and 1, but their exact arrival times are uncertain. Recall the sample space from Example 4.3. Let R be the random variable representing Regina's arrival time (minutes after noon), and Y for Cady.

1. What does the random variable $T = \min(R, Y)$ represent? What are the possible values of T?

2. What does the random variable $W = |R - Y|$ represent? What are the possible values of W?

3. Let N be the number of people (out of 2) who arrive before 12:30. How can you represent N in terms of R and Y. (Hint: use indicators.)

4. Identify each of the random variables above as discrete or continuous.

5. Interpret each of the following in words and draw a picture representing it.

   a. $\{R > Y\}$.

   b. $\{T < 30\}$.

   c. $\{W < 15\}$.

d. $\{R < 24\}$.