1 Introduction

Mathematical structures like Euclidean geometry or algebraic fields are defined by a set of axioms. “Mathematical reality” is then developed through the introduction of concepts and the proofs of theorems. These axioms are inspired, in the instances introduced above, by our intuitive understanding, for example, of the nature of parallel lines or the real numbers. Probability is a branch of mathematics based on three axioms inspired originally by calculating chances from card and dice games.

Statistics, in its role as a facilitator of science, begins with the collection of data. From this collection, we are asked to make inference on the state of nature, that is to determine the conditions that are likely to produce these data. Probability, in undertaking the task of investigating differing states of nature, takes the complementary perspective. It begins by examining random phenomena, i.e., those whose exact outcomes are uncertain. Consequently, in order to determine the “scientific reality” behind the data, we must spend some time working with the concepts in the theory of probability to investigate properties of the possible states of nature to assess which are most useful in making inference from data.

We will motivate the axioms of probability through the case of equally likely outcomes for some simple games of chance and look at some of the direct consequences of the axioms. In order to extend our ability to use the axioms, we will learn counting techniques, e.g., permutations and combinations, based on the multiplication principle.

A probability model has two essential pieces of its description.

• $\Omega$, the sample space, the set of possible outcomes.
  
  - An event is a collection of outcomes. We can give explicitly define and event via its outcomes,
    
    $$A = \{\omega_1, \omega_2, \ldots, \omega_n\}$$
  
  or with a description
    
    $$A = \{\omega : \omega \text{ has property } P\}.$$  

In either case, $A$ is subset of the sample space, $A \subset \Omega$.

• $P$, the probability assigns a number to each event.

Thus, a probability is a function. We are familiar with functions in which both the domain and range are subsets of the real numbers. The domain of a probability function is the collection of all possible outcomes. The range is still a number. We will see soon which numbers we will accept as probabilities of events.
## 2 Set Theory - Probability Theory Dictionary

<table>
<thead>
<tr>
<th>Event Language</th>
<th>Set Language</th>
<th>Set Notation</th>
<th>Venn Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample space</td>
<td>universal set</td>
<td>$\Omega$</td>
<td></td>
</tr>
<tr>
<td>event</td>
<td>subset</td>
<td>$A, B, C, \cdots$</td>
<td></td>
</tr>
<tr>
<td>outcome</td>
<td>element</td>
<td>$\omega$</td>
<td></td>
</tr>
<tr>
<td>impossible event</td>
<td>empty set</td>
<td>$\emptyset$</td>
<td></td>
</tr>
<tr>
<td>not $A$</td>
<td>$A$ complement</td>
<td>$A^c$</td>
<td></td>
</tr>
<tr>
<td>$A$ or $B$</td>
<td>$A$ union $B$</td>
<td>$A \cup B$</td>
<td></td>
</tr>
<tr>
<td>$A$ and $B$</td>
<td>$A$ intersect $B$</td>
<td>$A \cap B$</td>
<td></td>
</tr>
<tr>
<td>$A$ and $B$ are mutually exclusive</td>
<td>$A$ and $B$ are disjoint</td>
<td>$A \cap B = \emptyset$</td>
<td></td>
</tr>
<tr>
<td>if $A$ then $B$</td>
<td>$A$ is a subset of $B$</td>
<td>$A \subset B$</td>
<td></td>
</tr>
</tbody>
</table>
3 Equally Likely Outcomes

The essential relationship between events and the probability are described through the three axioms of probability. These axioms can be motivated through the first uses of probability, namely the case of equal likely outcomes.

If \( \Omega \) is a finite sample space, then if each outcome is equally likely, we define the probability of \( A \) as the fraction of outcomes that are in \( A \).

\[
P(A) = \frac{\#(A)}{\#(\Omega)}.
\]

Thus, computing \( P(A) \) means counting the number of outcomes in the event \( A \) and the number of outcomes in the sample space \( \Omega \) and dividing.

Exercise 1. 1. Toss a coin. Then

\[
\Omega = \{ H, T \}
\]
\[
A = \{ H \}
\]
\[
P(\text{heads}) = \frac{\#(A)}{\#(\Omega)} = \quad .
\]

2. Toss a coin three times.

\[
\Omega = \{ HHH, HHT, HTH, HTT, THH, THT, TTH, TTT \}
\]
\[
P(\text{toss at least two heads in a row}) = \frac{\#(A)}{\#(\Omega)} = \quad .
\]

3. Roll two dice.

\[
\Omega = \{ (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \}
\]
\[
P(\text{sum is 7}) = \frac{\#(A)}{\#(\Omega)} = \quad .
\]

Because we always have \( 0 \leq \#(A) \leq \#(\Omega) \), we always have

\[
P(A) \geq 0 \quad (1)
\]

and

\[
P(\Omega) = 1 \quad (2)
\]

This gives us 2 of the three axioms, The third will require more development.

Toss a coin 4 times.

\[
A = \{\text{exactly 3 heads}\} = \{ \text{HHHT, HHTH, HTHH, THHH} \}
\]
\[
\#(\Omega) = 16 \quad \#(A) = 4
\]
\[ P(A) = \frac{4}{16} = \frac{1}{4} \]

\[ B = \{\text{exactly 4 heads}\} = \{\text{HHHH}\} \quad \#(B) = 1 \]

\[ P(B) = \frac{1}{16} \]

Now let’s define the set \( C = \{\text{at least three heads}\} \). If you are asked to supply the probability of \( C \), your intuition is likely to give you an immediate answer.

\[ P(C) = \ldots \]

Let’s have a look at this intuition. The events \( A \) and \( B \) have no outcomes in common. We say that the two events are **disjoint** or **mutually exclusive** and write \( A \cap B = \emptyset \). In this situation,

\[ \#(A \cup B) = \#(A) + \#(B). \]

If we take this **addition principle** and divide by \( \#(\Omega) \), then we obtain the following identity:

If \( A \cap B = \emptyset \), then

\[ P(A \cup B) = P(A) + P(B). \tag{3} \]

Using this property, we see that

\[ P(\{\text{at least 3 heads}\}) = P(\{\text{exactly 3 heads}\}) + P(\{\text{exactly 4 heads}\}) = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}. \]

We are saying that any function \( P \) that accepts events as its domain and returns numbers as its range and satisfies (1), (2), and (3) can be called a **probability**.

If we iterate the procedure in Axiom 3, we can also state that if the events, \( A_1, A_2, \cdots, A_n \), are mutually exclusive, then

\[ P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \tag{3'} \]

This is a sufficient definition for a probability if the sample space \( \Omega \) is finite. Thus, statements (1), (2), and (3’) give us the complete axioms of probability for a finite sample space.

## 4 Consequences of the Axioms

Other properties that we associate with a probability can be derived from the axioms.

1. **The Complement Rule.** Because \( A \) and its complement \( A^c = \{\omega; \omega \notin A\} \) are mutually exclusive

\[ P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1 \]

or

\[ P(A^c) = 1 - P(A). \]

For example, if we toss a biased coin. We may want to say that \( P(\{\text{heads}\}) = p \) where \( p \) is not necessarily equal to 1/2. By necessity,

\[ P(\{\text{tails}\}) = 1 - p. \]

Toss a coin 4 times.

\[ P(\{\text{fewer than 3 heads}\}) = 1 - P(\{\text{at least 3 heads}\}) = 1 - \frac{5}{16} = \frac{11}{16}. \]
2. **The Difference Rule** Write $B \setminus A$ to denote the outcomes that are in $B$ but not in $A$. If $A \subset B$, then

$$P(B \setminus A) = P(B) - P(A).$$

Because $P(B \setminus A) \geq 0$, we have the following:

3. **Monotonicity Rule** If $A \subset B$, then $P(A) \leq P(B)$

We already know that for any event $A$, $P(A) \geq 0$. The monotonicity rule adds to this the fact that

$$P(A) \leq P(\Omega) = 1.$$

This, the range of a probability are the numbers between 0 and 1.

4. **The Inclusion-Exclusion Rule.** For any two events $A$ and $B$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (4)$$

($P(A) + P(B)$ accounts for the outcomes in $A \cap B$ twice, so remove $P(A \cap B)$.)

Deal two cards.

$A = \{\text{ace on the second card}\}$, \quad $B = \{\text{ace on the first card}\}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\text{at least one ace}) = \frac{1}{13} + \frac{1}{13} - ?$$

To complete this computation, we will need to compute $P(A \cap B) = P(\text{both cards are aces})$.

5. **The Bonferroni Inequality.** For any two events $A$ and $B$,

$$P(A \cup B) \leq P(A) + P(B).$$
6. **Continuity Property.** Use the symbol $\subset$ to denote “contains in”. If events satisfy

$$B_1 \subset B_2 \subset \cdots \text{ and } B = \bigcup_{i=1}^{\infty} B_i$$

Then, by the monotonicity rule, $P(B_i)$ is an increasing sequence satisfying

$$P(B) = \lim_{i \to \infty} P(B_i).$$

(5)

Similarly, use the symbol $\supset$ to denote “contains”. If events satisfy

$$C_1 \supset C_2 \supset \cdots \text{ and } C = \bigcap_{i=1}^{\infty} C_i$$

Again, by the monotonicity rule, $P(C_i)$ is an decreasing sequence satisfying

$$P(C) = \lim_{i \to \infty} P(C_i).$$

(6)

**Exercise 2 (odds).** The statement of $a:b$ odds for an event $A$ indicates that

$$\frac{P(A)}{P(A^c)} = \frac{a}{b}$$

Show that

$$P(A) = \frac{a}{a+b}.$$

So, for example, 1:2 odds means $P(A) = 1/3$ and 5:3 odds means $P(A) = 5/8$.

5 **Counting**

In the case of equally likely outcomes, finding the probability of an event $A$ is the result of two counting problems - namely finding $\#(A)$, the number of outcomes in $A$ and finding $\#(\Omega)$, the number of outcomes in the probability space. These counting problems can become quite challenging and advanced mathematical techniques have been developed to address these problems. However, having some facility in counting is necessary to have a sufficiently rich number of examples to give meaning to the axioms of probability.

We begin this section on counting with the **multiplication principle**.

Suppose that two experiments are to be performed.

- Experiment 1 can have $n_1$ possible outcomes and
- for each outcome of experiment 1, experiment 2 has $n_2$ possible outcomes.

Then together there are $n_1 \times n_2$ possible outcomes.

**Example 3.** For a group of $n$ individuals, one is chosen to become the president and a second is chosen to become the treasurer. By the multiplication principle, if these position are held by different individuals, then this task can be accomplished in

$$n \times (n-1)$$

possible ways.

**Exercise 4.** Generalize the multiplication principle of counting to $k$ experiments.
Assume that we have a collection of \(n\) objects and we wish to make an ordered arrangement of \(k\) of these objects. Using the generalized principle of counting, the number of possible outcomes is
\[
n \times (n - 1) \times \cdots \times (n - k + 1).
\]
We will write this as \((n)_k\) and say \(n\) falling \(k\).

**Definition 5.** The ordered arrangement of all \(n\) objects is
\[
(n)_n = n \times (n - 1) \times \cdots \times 1 = n!,
\]
n factorial. We take \(0! = 1\).

**Exercise 6.**
\[
(n)_k = \frac{n!}{(n - k)!}.
\]

### 5.1 Combinations

Write
\[
\binom{n}{k}
\]
for the number of different groups of \(k\) objects that can be chosen from a collection of \(n\).

We will next find a formula for this number by counting the number of possible outcomes in two different ways. To introduce this with a concrete example, suppose 3 cities will be chosen out of 8 under consideration for a vacation. If we think of the vacation as visiting three cities in a particular order, for example, New York then Boston then Montreal.

Then we are looking at permutations. This results in
\[
8 \times 7 \times 6
\]
choices.

If we are just considering the 3 cities we visit, irrespective of order, then these unordered choices are combinations. The number of ways of doing this is written
\[
\binom{8}{3},
\]
a number that we do not yet know how to determine. After we have chosen the three cities, we will also have to also pick an order to see the cities and so using the multiplication principle, we have
\[
\binom{8}{3} \times 3 \times 2 \times 1
\]
possible vacations if the order of the cities is included in the choice.

These two strategies are counting the same possible outcomes and so must be equal.
\[
8 \times 7 \times 6 = \binom{8}{3} \times 3 \times 2 \times 1 \quad \text{or} \quad \binom{8}{3} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1}.
\]
Thus, we have a formula for \(\binom{8}{3}\). Let’s do this more generally.

**Theorem 7.**
\[
\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n - k)!}.
\]
The second equality follows from the previous exercise.
We will form an ordered arrangement of \( k \) objects from a collection of \( n \) by:

1. First choosing a group of \( k \) objects. The number of possible outcomes for this experiment is \( \binom{n}{k} \).

2. Then, arranging this \( k \) objects in order. The number of possible outcomes for this experiment is \( k! \).

So, by the multiplication principle,

\[
(n)_k = \binom{n}{k} \times k!.
\]

Now complete the argument by dividing both sides by \( k! \).

**Exercise 8** (binomial theorem).

\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.
\]

**Exercise 9.** \( \binom{n}{1} = \binom{n}{n-1} = n \). \( \binom{n}{k} = \binom{n}{n-k} \). Thus, we set \( \binom{n}{n} = \binom{n}{0} = 1 \)

The number of combinations is computed in R using \texttt{choose}. For example, \( \binom{8}{3} \)

\[
\texttt{> choose(8, 3)}
\]

\[
\texttt{[1] 56}
\]

**Theorem 10** (Pascal’s triangle).

\[
\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.
\]

To establish this identity, distinguish one of the \( n \) objects in the collection. Say that we are looking at a collection of \( n \) marbles, \( n-1 \) are blue and 1 is red.

1. For outcomes in which the red marble is chosen, we must choose \( k-1 \) marbles from the \( n-1 \) blue marbles. Thus, \( \binom{n-1}{k-1} \) different outcomes have the red marble.

2. If the red marble is not chosen, then we must choose \( k \) blue marbles. Thus, \( \binom{n-1}{k} \) outcomes do not have the red marbles.

3. These choices of groups of \( k \) marbles have no overlap. And so \( \binom{n}{k} \) is the sum of the values in 1 and 2.

To see this using the example above,

\[
\binom{8}{3} = \binom{7}{2} + \binom{7}{3}.
\]

Assume that one of the 8 cities on the list includes New York. Then of the \( \binom{8}{3} \) vacations, \( \binom{7}{2} \) include New York and \( \binom{7}{3} \) do not.

This gives us an iterative way to compute the values of \( \binom{n}{k} \). Build a table of values for \( n \) (vertically) and \( k \leq n \) (horizontally). Then, by the Pascal’s triangle formula, a given table entry is the sum of the number directly above it and the number above and one column to the left. We can get started by noting that \( \binom{n}{0} = \binom{n}{n} = 1 \).

8
Example 11. For the experiment on honey bee queen - if we rear 60 of the 90 queen eggs, the we have

\[ \text{choose}(90, 60) \]

more than \(10^{23}\) different possible simple random samples.

Example 12. Deal out three cards. There are

\[ \binom{52}{3} \]

possible outcomes. Let \(x\) be the number of hearts. Then we have chosen \(x\) hearts out of 13 and \(3 - x\) cards that are not hearts out of the remaining 39. Thus, by the multiplication principle there are

\[ \binom{13}{x} \cdot \binom{39}{3-x} \]

possible outcomes.

To find the probability of \(x\) hearts

\[ x <- c(0:3) \]
\[ \text{prob} <- \text{choose}(13, x) \times \text{choose}(39, 3-x) / \text{choose}(52, 3) \]
\[ \text{data.frame(x, prob)} \]
\[ x \quad \text{prob} \]
\[ 0 \quad 0.41352941 \]
\[ 1 \quad 0.43588235 \]
\[ 2 \quad 0.13764706 \]
\[ 3 \quad 0.01294118 \]

Notice that

\[ \text{sum(prob)} \]

[1] 1

Exercise 13. Deal out 5 cards. Let \(x\) be the number of fours. What values can \(x\) take? Find the probability of \(x\) fours for each possible value.

6 Answers to Selected Exercises

1. 1/2, 2. 3/8, 3. 6/36 = 1/6
7. If
\[ \frac{a}{b} = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)} \cdot \]
Then,
\[ a - aP(A) = bP(A), \quad a = (a + b)P(A), \quad P(A) = \frac{a}{a + b}. \]

9. Suppose that \( k \) experiments are to be performed and experiment \( i \) can have \( n_i \) possible outcomes irrespective of the outcomes on the other \( k - 1 \) experiments. Then together there are \( n_1 \times n_2 \times \cdots \times n_k \) possible outcomes.

12. \( (n)_k = n \times (n - 1) \times \cdots \times (n - k + 1) \times \frac{(n - k)!}{(n - k)!} = \frac{n \times (n - 1) \times \cdots \times (n - k + 1)(n - k)!}{(n - k)!} = \frac{n!}{(n - k)!} \).

14. Expansion of \((x + y)^n = (x + y) \times (x + y) \times \cdots \times (x + y)\) will result in \(2^n\) terms. Each of the terms is achieved by one choice of \( x \) or \( y \) from each of the factors in the product \((x + y)^n\). Each one of these terms will thus be a result in \( n \) factors - some of them \( x \) and the rest of them \( y \). For a given \( k \) from \( 0, 1, \ldots, n \), we will see choices that will result in \( k \) factors of \( x \) and \( n - k \) factors of \( y \), i.e., \( x^k y^{n-k} \). The number of such choices is the combination
\[ \binom{n}{k} \]
Add these terms together to obtain
\[ \binom{n}{k} x^k y^{n-k}. \]
Next adding these values over the possible choices for \( k \) results in
\[ (x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \]

15. The formulas are easy to work out. One way to consider \( \binom{n}{1} = \binom{n}{n-1} \) is to note that \( \binom{n}{1} \) is the number of ways to choose 1 out of a possible \( n \). This is the same as \( \binom{n}{n-1} \), the number of ways to exclude 1 out of a possible \( n \). A similar reasoning gives \( \binom{n}{k} = \binom{n}{n-k} \).

19. The possible values for \( x \) are 0, 1, 2, 3, and 4. When we have chosen \( x \) fours out of 4, we also have \( 5 - x \) cards that are not fours out of the remaining 48. Thus, by the multiplication principle, the probability of \( x \) fours is
\[ \binom{4}{x} \cdot \binom{52}{5-x} \cdot \binom{5}{5}. \]