PROBABILITY
(MATH 361)

Richard Sowers

Department of Mathematics
University of Illinois at Urbana-Champaign
Urbana, IL 61801
r-sowers@uiuc.edu

October 18, 1999
Chapter 1

Axioms, Definitions, and Consequences

1.1. Axioms

Let’s begin with some almost self-evident examples, from which we will develop some axioms of probability.

Example 1. Toss a fair coin. The possible outcomes are heads and tails, which we will henceforth denote as H and T. The probability of heads is \( p \) (a parameter in \([0, 1]\)) and the probability of tails is \( 1 - p \). We also have that the probability of heads or tails is 1, and the probability that neither heads nor tails occur is 0.

Example 2. Toss a fair die. The possible outcomes are 1, 2, 3, 4, 5, and 6. The probability of throwing a 1 is 1/6, as is the probability of throwing a 2, as is the probability of throwing a 3, etc. We can also see that the probability that we throw an even number (i.e., 2, 4, or 6) is 1/2, while the probability that we throw a number strictly greater than 1 (i.e., a 2, 3, 4, 5 or 6) is 5/6. It is fairly easy to see that there are \( 2^6 = 64 \) probabilities that we can compute\(^1\). More specifically, we can compute the probability of any subset of \( \{1, 2, 3, 4, 5, 6\} \)

Definition (Power Set). If \( \Omega \) is any set, we denote by \( 2^\Omega \) the collection of all subsets of \( \Omega \).

---

\(^1\) We can count the number of subsets of \( \{1, 2, 3, 4, 5, 6\} \) by looking at 6-digit binary numbers. Each digit corresponds to a different number between 1 and 6, and if the digit is a one, then that number is included in the subset of interest; otherwise it is not. For example, \( \{2, 4, 6\} \) would correspond to 010101, while \( \{5, 6\} \) would correspond to 000011.
Thus the collection of all subsets in Example 2 is $2^{\{1,2,3,4,5,6\}}$.

**Example 3.** Toss two fair coins. We denote by \( H \) a heads and by \( T \) a tails. The possible outcomes are \((H,H), (H,T), (T,H), \text{ and } (T,T)\). The stipulation that we are tossing two fair coins implies that the probability of \((H,H)\) is 1/4, as is the probability of \((H,T), (T,H), \text{ and } (T,T)\). There are in total $2^4 = 16$ probabilities we can compute.

**Definition (Cartesian Products).** If \( \Omega_1 \) and \( \Omega_2 \) are sets, then \( \Omega_1 \times \Omega_2 \) is the collection of all ordered pairs \((\omega_1,\omega_2)\), where \( \omega_1 \in \Omega_1 \) and \( \omega_2 \in \Omega_2 \).

Thus the possible outcomes of two tosses of a coin is \( \{H,T\}^2 \).

What do all of these examples have in common?

**Definition (Event Space).** A set \( \Omega \) of possible outcomes of an experiment is called the event space.

In Example 1, \( \Omega = \{H,T\} \) (heads and tails). In Example 2, we have that \( \Omega = \{1,2,3,4,5,6\} \). In Example 3, \( \Omega = \{H,T\}^2 \).

The point of notation is that it often provides a symbolic representation of complicated things which is consistent with simpler things.

**Example 4.** Toss an infinite sequence of coins. The collection of possible outcomes is \( \{H,T\}^\mathbb{N} \), where \( \mathbb{N} = \{1,2,\ldots\} \). In other words, for the first toss, we have a copy of \( \{H,T\} \), for the second toss we also have a copy of \( \{H,T\} \), and so forth.

Once we have a collection of possible outcomes \( \Omega \) of an experiment, we can assign likelihoods or probabilities to different subsets of \( \Omega \). We need to be careful here, both to be technically correct and to have a sufficiently robust theory.

**Example 5.** Toss an unfair die, so that the event space is \( \Omega = \{1,2,3,4,5,6\} \). Assume that we know only that

\[
\begin{align*}
\mathbb{P}\{1,2,3,4\} &= \frac{9}{12} & \mathbb{P}\{5\} &= \frac{1}{12} & \mathbb{P}\{6\} &= \frac{2}{12} \\
\end{align*}
\]

(1.1)

While we can’t find \( \mathbb{P}\{1\}, \mathbb{P}\{2\}, \mathbb{P}\{3\}, \text{ or } \mathbb{P}\{4\} \) from this information, we still can answer a number of questions. In addition to the information we have been given, it seems reasonable that

\[
\begin{align*}
\mathbb{P}\{\emptyset\} &= 0 & \mathbb{P}\{5,6\} &= \frac{3}{12} & \mathbb{P}\{1,2,3,4,6\} &= \frac{11}{12} \\
\mathbb{P}\{1,2,3,4,5\} &= \frac{10}{12} & \mathbb{P}\{1,2,3,4,5,6\} &= 1
\end{align*}
\]
From this example, we see that there can be some situations where we need to be careful about what probabilistic questions we can fruitfully ask. This becomes a bit too complicated for these notes, but in the interest of intellectual honesty, let’s make a definition.

**Definition (Sigma-algebra).** Consider an event space \( \Omega \). A subset \( \mathcal{A} \) of \( 2^\Omega \) is called a *sigma-algebra* of subsets of \( \Omega \) if

1. \( \emptyset \in \mathcal{A} \)
2. If \( A \in \mathcal{A} \), then \( A^c \in \mathcal{A} \),
3. If \( A_1, A_2, \ldots \) sets in \( \mathcal{A} \), then \( \bigcup_n A_n \) is a set in \( \mathcal{A} \).

In Example 4, we had that \( \mathcal{A} \) consisted of the following sets: \( \emptyset \), \{1, 2, 3, 4\}, \{5\}, \{6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \{5, 6\}, and \{1, 2, 3, 4, 5, 6\}. Of course \( 2^\Omega \) is also a sigma-algebra. The point of this definition is that if a collection of subsets of an event space is not a sigma-algebra, then it doesn’t have enough structure to do all of the calculations we might imagine.

Note that we thus far have defined
1. An event space
2. A sigma-algebra of subsets of the event space.

Let’s now look at the probabilities themselves. We see that the following are natural:

1. \( P(\emptyset) = 0 \). Clearly the probability that *nothing* happens is zero.
2. If \( P(A) \) is well-defined, then \( P(A) \geq 0 \). For a toss of a die, it would be unreasonable to have that \( P\{\text{even}\} = -0.2 \).
3. If \( P(A_1), P(A_2), P(A_3) \), etc. are well-defined and the \( A_n \)’s are disjoint (i.e., \( A_i \cap A_j = \emptyset \) if \( i \neq j \)), then

\[
P(\bigcup_n A_n) = \sum_n P(A_n).
\]

In Example 2, it would be unreasonable to have \( P\{5\} = P\{6\} = 1/6 \), and \( P\{5, 6\} = 3/6 \). Since \( \{5, 6\} = \{5\} \cup \{6\} \) and \( \{5\} \cap \{6\} = \emptyset \), we should rather have that \( P\{5, 6\} = P\{5\} + P\{6\} = 2/6 \). In fact, the requirement that the die is fair, i.e., that

\[
P\{1\} = P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} = 1/6
\]

is enough to specify \( P(A) \) for *any* \( A \in 2^{\{1,2,3,4,5,6\}} \). Another interpretation of this requirement is that you can add probabilities *as long as you don’t overcount*.

Note that the sigma-algebra structure already is useful. Suppose that we can compute \( P(A) \) for *any* \( A \) in a certain sigma-algebra \( \mathcal{A} \) of subsets of \( \Omega \). If \( A \in \mathcal{A} \), then \( A^c \in \mathcal{A} \), so we automatically know that it is reasonable
1.1. Axioms

to try to compute $\mathbb{P}(A^c)$. If we have a disjoint countable collection of $A_n$’s in $\mathcal{A}$, then we automatically know that it is reasonable to try to compute $\mathbb{P}(\bigcup_n A_n)$.

Let’s make the following definition

**Definition (Probability Measure).** Fix an event space $\Omega$ and a sigma-algebra $\mathcal{A}$ of subsets of $\Omega$. A collection $\mathbb{P}(A)$’s where $A \in \mathcal{A}$ is a measure if

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A) > 0$ for all $A \in \mathcal{A}$.
3. If $A_1$, $A_2$, $A_3$, etc. are all in $\mathcal{A}$ and are all disjoint (i.e., $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n).$$

If in addition

$$\mathbb{P}(\Omega) = 1,$$

then $\mathbb{P}$ is a probability measure.

Suppose that $\Omega$ is finite, as in Examples 1 through 3 above. If we specify $\mathbb{P}\{\omega\}$ for each $\omega \in \Omega$ (as we did in Examples 1 through 3), then we can compute $\mathbb{P}(A)$ for any $A \in 2^\Omega$. If $\Omega$ is infinite, as in Example 5, then things become a bit more complicated. We will comment on this in a moment.

Let’s see what the three rules of probability can get us. Henceforth, we will assume that all relevant sets are in an appropriate sigma-algebra of subsets of an event space on which a probability measure $\mathbb{P}$ is defined.

**Theorem 1.** We have that

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

**Proof.** Consider Example 2. If we know that $\mathbb{P}\{1\} = 1/6$, then since $\{1\}$ and $\{2, 3, 4, 5, 6\}$ are disjoint and $\{1\} \cup \{2, 3, 4, 5, 6\} = \Omega$,

$$1/6 + \mathbb{P}\{2, 3, 4, 5, 6\} = \mathbb{P}\{1\} + \mathbb{P}\{2, 3, 4, 5, 6\} = \mathbb{P}(\Omega) = 1.$$

Thus $\mathbb{P}\{2, 3, 4, 5, 6\} = 5/6$. More generally, we have that $A$ and $A^c$ are disjoint, $A \cup A^c = \Omega$, and $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$. \qed

The next result simply says that if you add two probabilities, you should remove the amount by which you have overcounted.

**Theorem 2.** We have that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$
Proof. Again, consider Example 2. Assume that we know that \(P\{2, 3, 4\} = 3/6, P\{4, 5, 6\} = 3/6,\) and \(P\{4\} = 1/6.\) Thus \(P\{2, 3\} = 2/6\) and \(P\{5, 6\} = 2/6,\) and \(P\{2, 3, 4, 5, 6\} = 2/6 + 2/6 + 1/6.\) In general, \(A\) and \(B \setminus A\) are disjoint and \(A \cup (B \setminus A) = A \cup B,\) so
\[
P(A \cup B) = P(A) + P(B \setminus A). \quad (1.2)
\]
Similarly, \(B\) and \(A \setminus B\) are disjoint and \(B \cup (A \setminus B) = A \cup B,\) so
\[
P(A \cup B) = P(B) + P(A \setminus B). \quad (1.3)
\]
Finally, \(A \setminus B, B \setminus A,\) and \(A \cap B\) are disjoint and \((A \setminus B) \cup (B \setminus A) \cup (A \cap B) = A \cup B,\) so
\[
P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B). \quad (1.4)
\]
Combine (1.2), (1.3), and (1.4) to get that
\[
P(A \cup B) = (P(A \cup B) - P(A)) + (P(A \cup B) - P(B)) + BP(A \cap B).
\]
This easily leads to the claimed result. \(\square\)

Theorem 3. If \(A \subset B,\) then
\[
P(A) \leq P(B).
\]

Proof. Again, consider Example 2. Note that \(\{2, 3, 4, 5\} = \{2, 3, 4\} \cup \{5\},\) where \(\{2, 3, 4\}\) and \(\{5\}\) are disjoint. Thus \(P\{2, 3, 4, 5\} = P\{2, 3, 4\} + P\{5\},\) and since \(P\{5\} \geq 0,\) we see that \(P\{2, 3, 4\} \leq P\{2, 3, 4, 5\}.\) More generally, we can write that \(B = A \cup (B \setminus A),\) where \(A\) and \(B \setminus A\) are disjoint. Thus \(P(B) = P(A) + P(B \setminus A),\) and since \(P(B \setminus A) \geq 0,\) we get the result.

The third theorem shows us what can happen in the case of Example 5. What is the probability of the sequence \((H, H, H, H, \ldots)\)? For any \(n \in \mathbb{N},\)
\[
(H, H, H, H, \ldots) \in \underbrace{\{H, H, H, \ldots\}}_{\text{n times}}
\]
In other words, the outcome of flipping only heads is a subset of flipping heads on only the first \(n\) tries. According to Theorem 3,
\[
P\{(H, T, H, T, \ldots)\} \leq P\{(H, H, H, H, \ldots)\} = \frac{1}{2^n}.
\]
Since \(n\) was arbitrary, we have that
\[
P\{(H, H, H, H, \ldots)\} = 0.
\]
Similarly, we see that for any point \(\omega \in \Omega\) (i.e., an infinite sequence of heads or tails), we must have that \(P\{\omega\} = 0.\) Nevertheless, we can still see that this model is a realistic one. In the case of Example 5, one must use honest measure theory to deal with details.

Henceforth, we will assume that if the event space is finite, the relevant sigma-algebra is \(2^\Omega.\)
1.2. Independence and Conditional Probability

Let’s now pass to independence. Again, let’s start with an example.

Example 6. Again consider two coins as in Example 3. Set

\[ A \overset{\text{def}}{=} \{ \text{heads on the first toss} \} = \{ (H, H), (H, T) \} = \{ H \} \times \{ H, T \} \]

and set

\[ B \overset{\text{def}}{=} \{ \text{heads on the second toss} \} = \{ (H, H), (H, H) \} = \{ H, T \} \times \{ H \}. \]

It is easy to see that \( \mathbb{P}(A) = \mathbb{P}(B) = 1/2 \). Note that \( A \cap B = \{ (H, H) \} \), and that \( \mathbb{P}(A \cap B) = 1/4 \). Thus

\[ \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \]

On the other hand, we can consider Example 2. Set \( A = \{ \text{even number} \} = \{ 2, 4, 6 \} \) and let \( B = \{ \text{strictly greater than 1} \} = \{ 2, 3, 4, 5, 6 \} \). Here \( A \) and \( B \) are in some way related; if we know that we throw a number strictly greater than 1, we have extra information about whether the number is even or not. Note that \( \mathbb{P}(A) = 3/6, \mathbb{P}(B) = 5/6, A \cap B = \{ 2, 4, 6 \} \), and \( \mathbb{P}(A \cap B) = 3/6 \). Here,

\[ \mathbb{P}(A \cap B) = 3/6 \neq (3/6)(5/6) = \mathbb{P}(A)\mathbb{P}(B). \]

Let’s make the following definition.

Definition (Independence). We say that two sets \( A \) and \( B \) are independent if

\[ \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \]

If \( A \) and \( B \) are not independent, we say that they are dependent.

Warning. In the case of Example 6, we note that

\[ A \cap B \neq \emptyset. \]

We thus see that if two sets are independent, they are in general not disjoint. Independence is not the same as disjointedness.

We also note that independence is useful in characterizing probability laws.
Example 7. Again consider two coins as in Example 3, but assume that the coins are possibly unfair; in that the probability of heads for either coin is \( p \in [0, 1] \). We will use the sets \( A \) and \( B \) of Example 6. The specification of \( p \) implies that \( P(A) = P(B) = p \), while \( P(A^c) = P(B^c) = 1 - p \). The specification of independence implies that \( P\{\{H, H\}\} = P(A)P(B) = p^2 \). Similarly, \( P\{\{H, T\}\} = P(A)P(B^c) = p(1 - p) \), \( P\{\{T, H\}\} = P(A^c)P(B) = (1 - p)p \), and \( P\{\{T, T\}\} = P(A^c)P(B^c) = (1 - p)^2 \). From this we can find the probability of any subset of \( \Omega \).

Roughly, independence characterizes the property that \( A \) gives us no indication of whether or not \( B \) has occurred. If \( A \) and \( B \) are dependent, can we somehow characterize the amount of information \( B \) gives us about \( A \), or vice-versa?

Example 8. Back to Example 2. Assume that we know that \( B = \{2, 3, 4, 5, 6\} \) occurs. What is the likelihood that \( A = \{2, 4, 6\} \) occurs? It seems natural to reducing the event space to \( B \), and require that each element of \( B \) be equally likely (i.e., of probability 1/5). On this reduced space, the probability of \( A \) should be 3/5. We note that 3/5 can be gotten in another way. We have that \( P(B) = 5/6 \), and \( P(A \cap B) = P\{2, 4, 6\} = 3/6 \), and we see that

\[
\frac{3}{5} = \frac{3/6}{5/6} = \frac{P(A \cap B)}{P(B)}.
\]

We shall take (1.5) as our cue for the idea of conditional probability

Definition (Conditional Probability). If \( P(B) > 0 \), then

\[
P(A|B) \overset{\text{def}}{=} \frac{P(A \cap B)}{P(B)}. \tag{1.6}
\]

The relation of conditional probability to independence is as follows.

Theorem. Assume that \( P(A) > 0 \) and \( P(B) > 0 \). Then the following are equivalent:

1. \( A \) and \( B \) are independent
2. \( P(A|B) = P(A) \)
3. \( P(B|A) = P(B) \)

Proof. Assume that \( A \) and \( B \) are independent. Then

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).
\]
1.2. Independence and Conditional Probability

Assume now that $\mathbb{P}(A|B) = \mathbb{P}(A)$. Then by rearranging (1.6), we have that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B).$$

(1.7)

Thus $A$ and $B$ are independent. This proves that the second condition is equivalent to independence. Switch $A$ and $B$ to see that the third condition is equivalent to independence. \hfill \square

This result formalizes the idea that if $A$ and $B$ are independent, then $A$ gives us no information about $B$ and vice-versa.

Let’s emphasize (1.7) in its own right.

**Theorem.** If $\mathbb{P}(A) > 0$, then

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B).$$

**Proof.** Rearrange (1.6). \hfill \square

Let’s consider the following problem.

**Example 9.** Suppose that the probability of having a disease is 0.2; we shall write this as $\mathbb{P}(D) = 0.2$. Suppose that if you have the disease, you test positive with probability 0.9; we shall write this as $\mathbb{P}(+|D) = 0.9$. If you don’t have the disease, you test positive with probability 0.3; i.e., $\mathbb{P}(+|D^c) = 0.3$. If you test positive, what is the probability that you have the disease?

We are interested in finding

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(D \cap +)}{\mathbb{P}(+)}$$

(whereas we are given, amongst other things, $\mathbb{P}(+|D^c)$). Note from the above theorem that

$$\mathbb{P}(D \cap +) = \mathbb{P}(+ \cap D) = \mathbb{P}(+|D)\mathbb{P}(D) = (0.9)(0.2).$$

This calculation shows us how to compute $\mathbb{P}(+)$. Since $+ = (+ \cap D) \cup (+ \cap D^c)$ where $(+ \cap D)$ and $(+ \cap D^c)$ are disjoint; we have that $\mathbb{P}(+) = \mathbb{P}(+ \cap D) + \mathbb{P}(+ \cap D^c)$. It is easy to see that $\mathbb{P}(D^c) = 1 - \mathbb{P}(D) = 1 - 0.2 = 0.8$. Thus

$$\mathbb{P}(+ \cap D^c) = \mathbb{P}(+|D^c)\mathbb{P}(D^c) = (0.3)(0.8),$$

and so $\mathbb{P}(+) = (0.9)(0.2) + (0.3)(0.8)$. Hence

$$\mathbb{P}(D|+) = \frac{(0.9)(0.2)}{(0.9)(0.2) + (0.3)(0.8)}.$$

From this we get something known impressively as Bayes’ rule.
**Theorem (Bayes’ Rule).** If \( P(B) > 0 \), then

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.
\]

Rather than trying to remember Bayes’ rule, I suggest that you remember its proof, which requires computing the numerator and denominator of \( P(A|B) \).
Some Tricks

2.1. Tricks

A collection of mathematical tricks is next. We will be brief.

First, let’s discuss how we can count the number of possible outcomes of a complex experiment.

**Theorem (Compound Experiments).** Assume that we have a sequence $E_1, E_2 \ldots E_n$ of experiments. Assume that experiment $E_i$ has $m_i$ possible outcomes. Then the number of ways that the sequence $E_1, E_2 \ldots E_n$ of experiments has $\prod_{i=1}^{n} n_i = (n_1 \cdot n_2 \cdot n_3 \ldots n_m)$ possible outcomes.

**Proof.** The result is clearly true if $n = 2$. If it is true for some $n$, then we can group $E_1, E_2 \ldots E_n$ as one experiment with $\prod_{i=1}^{n} m_i$ outcomes. By using the $n = 2$ case, we see that the sequence $E_1, E_2 \ldots E_{n+1}$ of experiments can occur in $(\prod_{i=1}^{n} m_i) \times m_{n+1}$ ways. The result thus follows by induction.

**Definition.** For any integer $n \in \mathbb{N}$, we define $n! \overset{\text{def}}{=} 1 \cdot 2 \cdot 3 \ldots n = \prod_{i=1}^{n} i$. For convenience, we define $0! \overset{\text{def}}{=} 1$. For any $n \in \mathbb{N}$ and any $k \in \{0, 1 \ldots n\}$, we define

$$(n)_k \overset{\text{def}}{=} \frac{n!}{(n-k)!} \quad \text{and} \quad \binom{n}{k} \overset{\text{def}}{=} \frac{n!}{k!(n-k)!}.$$

We will be able to efficiently count using the symbols $(n)_k$ and $\binom{n}{k}$. To understand the problems we are about to discuss, it is crucial to remember the difference between a sequence and a set. A **sequence** is an **ordered** collection of numbers. A **set** is a **non-ordered** collection of (nonrepeating) numbers.
**Example 1.** If we toss two coins, then we have either H or T for the first toss, and either H or T for the second value. If we get H on the first toss and T on the second, we can write this as \((H, T)\). The parentheses imply that order is important, just like \((2, 3)\), which is usually interpreted as a point on the plane, is not equal to \((3, 2)\), when this too is interpreted as a point on the plane. On the other hand, if we pick the set (or rather hand) \(\{A_\bigodot, 5_\bigodot, J_\spadesuit, K_\spadesuit\}\) from a deck, then we don’t care about the order of the elements; \(\{A_\bigodot, 5_\bigodot, J_\spadesuit, K_\spadesuit\} = \{5_\bigodot, A_\bigodot, J_\spadesuit, K_\spadesuit\}\).

We will also need to understand the difference between *sampling with replacement* and *sampling without replacement*.

**Example 2.** Suppose that we choose two cards from a deck of cards. We can do this in two ways. After choosing the first card, we can either put it back (sampling with replacement), or not (sampling without replacement). If we put it back, the first sample can occur in 52 ways, and so can the second, so the two cards can be chosen in \((52)^2 = 2704\) ways. If we don’t put the card back, we can choose the first card in 52 ways. Once such a card is taken from the deck, there are 51 ways to choose the second, so there are \(52 \times 51 = 2652\) ways if we don’t replace the first card. In other words, if we twice sample the deck with replacement, we have \((52)^2\) possible outcomes. If we sample the deck without replacement, we have \(52 \times 51 = (52)_2\) possible outcomes.

We have the following result.

**Theorem.** Suppose we have a box of \(n\) elements. The number of ways that we can select a sequence of \(k\) elements (i.e., make \(k\) selections from the box) with replacement is \(n^k\). The number of ways that we can select a sequence of \(k\) elements (i.e., make \(k\) selections from the box) without replacement is \((n)_k\).

**Proof.** In Example 2, if we sampled twice with replacement, we got \((52)^2\); in general, we will get \(n^k\) by the same reasoning as in Example 2. In Example 2, if we sampled twice without replacement, we got

\[
52 \times 51 = \frac{52!}{50!}
\]

possible outcomes. If we have \(n\) objects and sample \(k\) times without replacement, then on the first sample we have \(n\) possible outcomes, on the second we have \(n - 1\) possible outcomes, and so forth. We end up with

\[
\prod_{i=0}^{k-1} (n - i) = \prod_{i=n-k+1}^{n} i = \frac{n!}{\prod_{i=1}^{n-k} i} = \frac{n!}{(n-k)!}
\]
possible outcomes. \hfill \Box

The second part of this result leads to our next result.

**Theorem.** Suppose we have a box of \( n \) elements. The number of ways that we can select a set of \( k \) elements (i.e., make \( k \) selections from the box) without replacement is \( \binom{n}{k} \).

**Proof.** Let’s first do the proof with \( n = 10 \) and \( k = 6 \); if we let \( \gamma \) be the number of ways that we can select a set of 6 elements from a box of 10 elements, we want to show that \( \gamma = \binom{10}{6} \). Let’s look at a different problem, namely that of ordering the 10 elements; i.e., selecting a sequence of 10 elements (without replacement) from our box of 10 elements. Of course by the previous result, we know that we can do this in \( 10! \) ways. However, let’s solve this problem in another way. Let’s select a set of 6 elements and color them green; we will let these be the first 6 elements. We can select the elements which we will color green in \( \gamma \) ways. Once we have colored them green, we can order them in \( 6! \) ways. We can then order the remaining (i.e., non-green) elements in \( 4! \) ways and put them after the 6 elements. Thus we have

\[
10! = \gamma \times 6! \times 4!
\]

and so we must have that

\[
\gamma = \frac{10!}{6!4!} = \binom{10}{6}.
\]

The general case is similar. Now let \( \gamma \) be the number of ways we can select a set of \( k \) elements from \( n \) elements. We look at an auxiliary problem of ordering the \( n \) elements; we can do this in \( n! \) ways. We can also select a subset of \( k \) elements which we will reserve for the first \( k \) slots; we can make this selection in \( \gamma \) ways. Having made the selection, we order them in \( k! \) ways and then order the remaining \( n - k \) elements, which we can do in \( (n - k)! \) ways, for the last slots. Thus we have

\[
n! = \gamma k!(n - k)!
\]

and hence \( \gamma = \binom{n}{k} \). \hfill \Box

**Example 3.** Consider the set \( \{ A_\diamondsuit, 5_\heartsuit, J_\spadesuit, K_\spadesuit \} \), and suppose that we want to select a sequence of 2 elements (without replacement) from this set of 4 elements. Then there are \( (4)_2 = 4 \times 3 = 12 \) ways to do this, namely

\[
\begin{align*}
(A_\diamondsuit, 5_\heartsuit) & \quad (5_\heartsuit, A_\diamondsuit) & \quad (J_\spadesuit, A_\diamondsuit) & \quad (K_\spadesuit, A_\diamondsuit) \\
(A_\diamondsuit, J_\spadesuit) & \quad (5_\heartsuit, J_\spadesuit) & \quad (J_\spadesuit, 5_\heartsuit) & \quad (K_\spadesuit, 5_\heartsuit) \\
(A_\diamondsuit, K_\spadesuit) & \quad (5_\heartsuit, K_\spadesuit) & \quad (J_\spadesuit, K_\spadesuit) & \quad (K_\spadesuit, J_\spadesuit)
\end{align*}
\]
If we were only interested in selecting a set of 2 elements, we would realize that

\[(A_\varnothing, 5_\varnothing) = (5_\varnothing, A_\varnothing)\]
\[(A_\varnothing, J_\varnothing) = (J_\varnothing, A_\varnothing)\]
\[(A_\varnothing, K_\varnothing) = (K_\varnothing, A_\varnothing)\]
\[(5_\varnothing, J_\varnothing) = (J_\varnothing, 5_\varnothing)\]
\[(5_\varnothing, K_\varnothing) = (K_\varnothing, 5_\varnothing)\]
\[(J_\varnothing, K_\varnothing) = (K_\varnothing, J_\varnothing)\]

Thus we can break the different sequences in (1) into pairs; i.e., we are overcounting by a factor of 2. This factor of 2 = \(2!\) comes from the different ways we can order each of the six sets \(\{A_\varnothing, 5_\varnothing\}, \{A_\varnothing, J_\varnothing\}, \{A_\varnothing, K_\varnothing\}, \{5_\varnothing, J_\varnothing\}, \{5_\varnothing, K_\varnothing\}, \text{and} \{J_\varnothing, K_\varnothing\}\).

Let’s next consider geometric sums.

**Theorem (Geometric Summation).** For any \(a \neq 1\), we have that

\[
\sum_{i=0}^{n} \alpha^i = \frac{1 - \alpha^{n+1}}{1 - \alpha}.
\]

If \(|\alpha| < 1\), then

\[
\sum_{i=0}^{\infty} \alpha^i = \frac{1}{1 - \alpha}.
\] (2.1)

**Proof.** For each \(n \in \mathbb{N}\), define

\[
S_n \triangleq \sum_{i=0}^{n} \alpha^i.
\]

Then

\[
S_{n+1} = S_n + \alpha^{n+1} \quad \text{and} \quad S_{n+1} = \alpha S_n + 1.
\]

Solve for \(S_n\). Let \(n\) tend to infinity to get (2.1).

We can put this to good use.

**Theorem.** We have that

\[
\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.
\]

**Proof.** Define

\[
\varphi(\alpha) \triangleq \sum_{i=0}^{n} \alpha^i
\]
for all $\alpha \neq 1$. Then
\[
\varphi'(\alpha) = \sum_{i=0}^{n} i\alpha^{i+1},
\]
and thus
\[
\sum_{i=0}^{n} i = \lim_{\alpha \to 1} \varphi'(\alpha) = \lim_{\alpha \to 1} \frac{-(n + 1)\alpha^{n}(1 - \alpha) + (1 - \alpha^{n+1})}{(1 - \alpha)^2}
\]
\[
= \lim_{\alpha \to 1} \frac{1 - (n + 1)\alpha^{n} + n\alpha^{n+1}}{(1 - \alpha)^2}
= \lim_{\alpha \to 1} \frac{-(n + 1)n\alpha^{n-1} + n(n + 1)\alpha^{n}}{-2(1 - \alpha)}
= \lim_{\alpha \to 1} \frac{(n + 1)n\alpha^{n-1}(1 - \alpha)}{2(1 - \alpha)} = \frac{n(n + 1)}{2}.
\]
\[
\square
\]

We also have

**Theorem.** For any $\alpha \neq 1$ and any nonnegative integers $n$ and $k$ with $n \leq k$, we have that
\[
\sum_{i=n}^{k} \alpha^{i} = \alpha^{n} \frac{1 - \alpha^{k-n+1}}{1 - \alpha}.
\]

If $|\alpha| < 1$, then
\[
\sum_{i=n}^{\infty} \alpha^{i} = \frac{\alpha^{n}}{1 - \alpha}.
\]

**Proof.** We have that $\sum_{i=n}^{k} \alpha^{i} = \alpha^{n} \sum_{i=0}^{k-n} \alpha^{i}$. Now use the geometric summation formula. $\square$

Let’s also remember the binomial theorem

**Theorem (Binomial Theorem).** For any $\alpha$ and $\beta$ in $\mathbb{R}$ and any positive integer $n$, we have that
\[
(\alpha + \beta)^{n} = \sum_{k=0}^{n} \binom{n}{k} \alpha^{k} \beta^{n-k}
\]

**Proof.** We use induction. The result is clearly true for $n = 1$;
\[
(\alpha + \beta)^{1} = \alpha + \beta = \binom{1}{0} \alpha^{0} \beta^{1} + \binom{1}{1} \alpha^{1} \beta^{0}.
\]
Assume the result is true for some \( n \). Then

\[
(\alpha + \beta)^{n+1} = \left\{ \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-k} \right\} (\alpha + \beta)
\]

\[
= \sum_{k=0}^{n} \binom{n}{k} (\alpha + \beta) \alpha^k \beta^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \alpha^{k+1} \beta^{n-k} + \alpha^k \beta^{n+1-k}
\]

\[
= \sum_{k=0}^{n} \binom{n}{k} \alpha^{k+1} \beta^{n-k} + \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n+1-k}
\]

\[
= \sum_{k=1}^{n+1} \binom{n}{k-1} \alpha^k \beta^{n+1-k} + \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n+1-k}
\]

\[
= \binom{n}{n} \alpha^{n+1} \beta^0 + \sum_{k=1}^{n} \left\{ \binom{n}{k-1} + \binom{n}{k} \right\} \alpha^k \beta^{n+1-k} + \binom{n}{0} \alpha^0 \beta^{n+1}.
\]

Note that

\[
\binom{n}{n} = 1 = \binom{n+1}{n+1} \quad \text{and} \quad \binom{n}{0} = 1 = \binom{n+1}{0}
\]

and that for any integer \( k \) between 1 and \( n \) (inclusive), we get that

\[
\binom{n}{k-1} + \binom{n}{k} = \frac{n!}{(k-1)!(n+1-k)!} + \frac{n!}{k!(n-k)!} = \frac{n!}{k!(n+1-k)!} \binom{k+(n+1-k)}{k} = \frac{n!}{k!(n+1-k)!} \binom{n+1}{k}.
\]

Thus

\[
(\alpha + \beta)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} \alpha^k \beta^{n+1-k},
\]

so we know that the result is true for \( n+1 \). This completes the inductive step and completes the proof. \( \square \)