2.1 Introduction

In this chapter, we introduce the concept of probability, and learn some techniques for computing probabilities. Intuitively, the probability of a given event should be the *long run frequency* of that event when the experiment is performed repeatedly, under the same conditions. But how do we know that such a limiting frequency exists?

Instead of assuming that such a limiting frequency exists, we assume some simpler, more self-evident axioms. Working from these basic axioms, we will eventually be able to *prove* that the limiting frequencies described above do exist in some sense – such results are called *laws of large numbers*, and we will prove two such laws towards the end of the course.

2.2 Sample Spaces and Events

Before we introduce probability, we need some terminology related to experiments and their outcomes. We consider experiments in which

- the outcome is not predictable with certainty, but
- the *set* of of all possible outcomes is known.

The set of all possible outcomes is known as the *sample space* of the experiment; it is usually denoted by S.

Example 2.2.1. Describe the sample space S of the given experiment.

(a) Rolling a six-sided die.

$$S = \{1, 2, 3, 4, 5, 6\}$$

(b) Flipping two coins.

$$S = \{HH, HT, TH, TT\}$$

(c) A race among 8 swimmers. (Assume that we only care about the order of finish.)

S is the set of all $8!$ permutations of the eight swimmers.

(d) Counting the number of customers that enter a store on a given day.

$S = \{0, 1, 2, 3, \ldots\} = \mathbb{N}_0$ (All nonnegative integers.)

(e) Measuring (in seconds) the length of time between successive bikes crossing a sensor on a bike path.

$S = (0, \infty)$, the set of all positive real numbers

A subset E of the sample space S is known as an *event*.

- In other words, an event is a set of possible outcomes of an experiment.
- If the outcome of the experiment is in E, then we say that the event E has *occurred*.

Example 2.2.2. Consider the experiment of rolling two six-sided dice.

(a) Write the sample space S.

$S = \{ (a,b) : 1 \le a, b \le 6 \text{ and } a,b \in \mathbb{N} \}$

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
⚁	(2,1)	(2,2)				
⚂	(3,1)					
⚃						
⚄						
⚅						

(b) Write the following events as a subset of S.

- The event that the same number appears on both dice.

  $\{(1,1), (2,2), \dots, (6,6)\} = \{(a,a) : 1 \le a \le 6\}$

- The event that the numbers on the dice sum to 4.

  $\{(a,b) : a+b = 4\} = \{(1,3), (2,2), (3,1)\}.$

(c) Describe the following events in words.

- $E = \{(1,1), (1,2), (2,1), (2,2)\}$

  The event that the number on both dice is at most 2.

- $F = \{(5,6), (6,5), (6,6)\}$

  The event that the numbers on the dice sum to at least 11.

Example 2.2.3. Consider the experiment of measuring the length of time (in hours) that a phone battery lasts on a full charge.

(a) Write the sample space.

$S = (0, \infty)$

(b) Describe the event $E = (10, 12)$ in words.

The phone dies at some point between 10 and 12 hours.

(c) Decribe the event $F = [24, \infty)$ in words.

The phone battery lasts at least a day.

It is often useful to write events in terms of other events. For example, we may be interested in when at least one of two given events occurs, or when one event occurs while another does not occur. We can use the familiar notions of union, intersection, and complement of sets to describe such events.

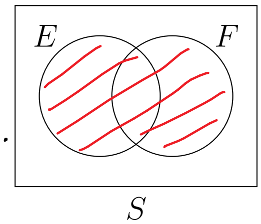
Notation 2.2.4. For events E and F of a sample space S, we let

union

- $E \cup F = \{ s \in S : s \in E \text{ or } s \in F \}$

  In words, the event $E \cup F$ occurs if and only if

  either E occurs or F occurs (or both).



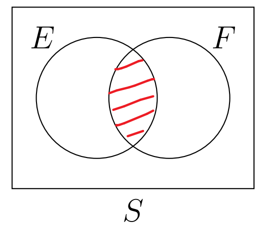
book writes EF

intersection

- $E \cap F = \{ s \in S : s \in E \text{ and } s \in F \}$

  In words, the event $E \cap F$ occurs if and only if

  both E and F occur.

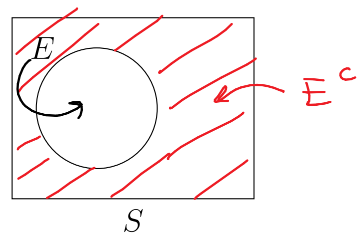


complement

- $E^c = \{ s \in S : s \notin E \}$

  In words, the event E^c occurs if and only if

  E does not occur.



Similarly, for events $E_1, E_2, \ldots, E_n$, we let

- $\displaystyle\bigcup_{i=1}^{n} E_i = E_1 \cup E_2 \cup \cdots \cup E_n = \{ s \in S : s \in E_i \text{ for some } i \in \{1, 2, \ldots, n\} \}$

  In words, the event $\displaystyle\bigcup_{i=1}^{n} E_i$ occurs if and only if at least one of the events E_i occurs.

- $\displaystyle\bigcap_{i=1}^{n} E_i = E_1 \cap E_2 \cap \cdots \cap E_n = \{ s \in S : s \in E_i \text{ for } \underline{\text{all}}\ i \in \{1, 2, \ldots, n\} \}$

  In words, the event $\displaystyle\bigcap_{i=1}^{n} E_i$ occurs if and only if all of the events E_i occur.

Note: For an infinite sequence of events $E_1, E_2, \ldots$, we can replace n with ∞.

Example 2.2.5. Consider the experiment of rolling two dice.

- Let E be the event that the number on the first die is 3.
- Let F be the event that the numbers on the dice sum to 5.
- Let G be the event that the same number appears on both dice.

Write the followings events as a subset of the sample space S.

(a) $E = \{(3,1),(3,2),(3,3),\ldots,(3,6)\}$

(b) $F = \{(4,1),(3,2),(2,3),(1,4)\}$

(c) $E \cup F = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(2,3),(1,4)\}$

(d) $E \cap F = \{(3,2)\}$

(e) $E \cap F^c = \{(3,1),(3,3),(3,4),(3,5),(3,6)\}$

(f) $E^c \cap F = \{(4,1),(2,3),(1,4)\}$

(g) $E \cap G = \{(3,3)\}$

(h) $F \cap G = \emptyset$

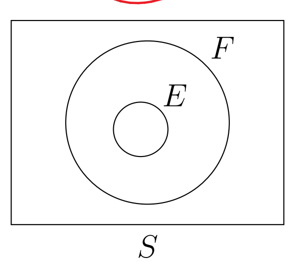
Notation 2.2.6. The event containing no outcomes is denoted by: $\emptyset$ the "null event"

- If $E \cap F = \emptyset$, then E and F are said to be mutually exclusive.
- More generally, events $E_1, E_2, E_3, \ldots$ are said to be *mutually exclusive* if

$$E_i \cap E_j = \emptyset \quad \text{for all} \quad i \neq j$$

For example, if we roll two dice, and for all $i \in \{2, 3, \ldots, 12\}$, we let E_i be the event that the sum of the numbers on the dice is exactly i, then $E_2, E_3, \ldots, E_{12}$ are mutually exclusive.

Notation 2.2.7. If every outcome in E is also in F, then we write: $E \subseteq F$ book writes $E \subset F$



- In this case, we say that
  - E is *contained* in F; or that
  - E is a *subset* of F.
- If $E \subseteq F$ and E occurs, then it must be the case that...

  F occurs.

For example, if we roll two dice, and for all $i \in \{2, 3, \ldots, 12\}$, we let F_i be the event that the sum of the numbers on the dice is at most i, then we have:

$$F_2 \subseteq F_3 \subseteq F_4 \subseteq \cdots \subseteq F_{12}$$

Example 2.2.8. Consider the experiment consisting of a sequence of coin flips, and let E_i be the event that the ith flip results in heads. Describe the following events in words.

(a) $E_1 \cup E_2$
This is the event that at least one of the first two flips comes up heads.

(b) $E_1 \cap E_2$
This is the event that both of the first two flips come up heads.

(c) $E_1 \cap E_2^c$ This is the event that the first flip comes up heads and the second flip comes up tails.

(d) $(E_1 \cup E_2)^c = E_1^c \cap E_2^c$ This is the event that both of the first two flips come up tails.

(e) $\bigcup_{i=1}^{10} E_i$ This is the event that at least one of the first ten flips comes up heads.

(f) $\bigcap_{i=1}^{\infty} E_i$ This is the event that all of the flips come up heads... for ever...

We have the following laws for the operations of union, intersection, and complement of events.

Commutative Laws $E \cup F = F \cup E$ $\qquad E \cap F = F \cap E$

same

same

Associative Laws $(E \cup F) \cup G = E \cup (F \cup G)$ $= E \cup F \cup G$ $\qquad (E \cap F) \cap G = E \cap (F \cap G)$ $E \cap F \cap G$

Distributive Laws $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ $\qquad (E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

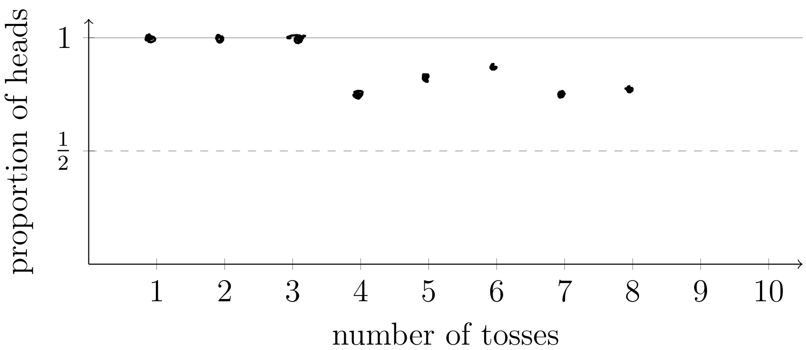
DeMorgan's Laws

$(E \cup F)^c = E^c \cap F^c$ $\qquad (E \cap F)^c = E^c \cup F^c$

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} (E_i^c) \qquad \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} (E_i^c)$$

2.3 Axioms of Probability

Intuitively, we often think of the probability of an event in terms of its *long run frequency*. For example, the probability that a coin toss lands on heads is $\frac{1}{2}$, because if we were to flip a coin repeatedly, we'd expect the proportion of tosses that land on heads to approach 1/2 in the long run.



Let's try to make this more precise.

- Suppose that we perform an experiment repeatedly (ad infinitum) under exactly the same conditions.

- For each event E, let $m_n(E)$ be the number of times that E occurs in the first n repetitions of the experiment.

- Then we'd like to define $P(E)$, the *probability* that the event E occurs, as

$$P(E) = \lim_{n\to\infty} \frac{m_n(E)}{n}$$

But there is an issue – how do we know that this limiting frequency exists?!

- In our example, do we really know that the proportion of heads obtained in the first n flips will approach 1/2 as n approaches ∞?

- If ten different people all start flipping a coin repeatedly, will they really all observe the same limiting proportion of heads?

One approach would simply be to assume that such limiting frequencies do exist, and build up our theory based on this assumption. However, this assumption seems rather complicated and unwieldy – imagine trying to use this assumption to prove precise statements about probability.

So instead of assuming that such limiting frequencies exist, we make some simpler, more self-evident assumptions, which we take as *axioms*. Assuming only these simple axioms, we will gradually build up the theory of probability. Towards the end of the course, we will prove two so-called *laws of large numbers*, which state that the limiting frequencies discussed above do exist in some sense!

Kolmogorov's axioms

Consider an experiment with sample space S. $\neq \emptyset$ For each event E of the sample space, assume that a number $P(E)$ is defined in such a way that the following three axioms are satisfied.

$E_i \cap E_j = \emptyset$ for all $i \neq j$

Axiom 1. $0 \leq P(E) \leq 1$

Axiom 2. $P(S) = 1$

Axiom 3. For any infinite sequence of mutually exclusive events $E_1, E_2, E_3, \ldots,$ we have

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Then we say that $P(E)$ is the *probability* of the event E.

Note that P is a function from the set of events to the set of real numbers. Throughout the remainder of the course, the letter P will denote a function that satisfies these three axioms.

Proposition 2.3.1. $P(\emptyset) = 0$.

Proof. Consider the sequence $E_1 = S$, $E_2 = \emptyset$, $E_3 = \emptyset, \ldots$ Then the events $E_1, E_2, \ldots$ are mutually exclusive, and $\bigcup_{i=1}^{\infty} E_i = S$, so by Axiom 3, we have

$$P(S) = \sum_{i=1}^{\infty} P(E_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

$$\Longrightarrow \quad 0 = \sum_{i=2}^{\infty} P(\emptyset) \quad \Longrightarrow \quad 0 = P(\emptyset)$$

$\square$

Proposition 2.3.2. *If $E_1, E_2, \ldots, E_n$ is any finite sequence of mutually exclusive events, then*

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i)$$

Feel free to say "by Axiom 3" when you apply this.

Proof. Consider the sequence starting with $E_1, E_2, \ldots, E_n$, followed by $E_{n+1} = \emptyset$, $E_{n+2} = \emptyset, \ldots$ Then the events $E_1, E_2, \ldots$ are mutually exclusive, so by Axiom 3, we have

$$P\left(\bigcup_{i=1}^{n} E_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(E_i) + \sum_{i=n+1}^{\infty} P(\emptyset)$$

$$= \sum_{i=1}^{n} P(E_i).$$

$\square$

One might wonder where the function P comes from in practical applications. We often simply define P using the nature of the situation, making sure to check that the three axioms are satisfied. This means that our definition of P often reflects our intuitive notion of probability as a long run frequency.

Example 2.3.3. If a fair six-sided die is rolled, then the sample space is $S = \{1, 2, 3, 4, 5, 6\}$, and P is defined by

$$P(\{1\}) = \frac{1}{6} = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\})$$

Note that it suffices to define P on all of the *elementary events* (those events containing a single outcome). The probability of the other events is completely determined by Axiom 3. For example, we have

- $P(\{1, 2\}) = P(\{1\} \cup \{2\}) \overset{\text{by Axiom 3}}{=} P(\{1\}) + P(\{2\}) = \frac{2}{6} = \frac{1}{3}$
- $P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$

Remark: In general, if the sample space S is finite or countably infinite, then it suffices to define P on the elementary events in such a way that $P(\{s\}) \geq 0$ for all $s \in S$, and $\sum_{s \in S} P(\{s\}) = 1$.

Example 2.3.4. Suppose that a number in $(0, 1)$ is selected uniformly at random, i.e., all numbers between 0 and 1 are equally likely to be chosen.

- The sample space is: $(0,1)$
- For any event $E \subseteq (0, 1)$, we define P by:

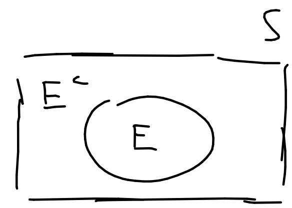
$$\int_E 1 \cdot dx$$

- For example, we have:
  - $\frac{1}{2}$ $P((0, 1/2)) = \int_0^{\frac{1}{2}} 1 \cdot dx = \left[x \right|_0^{\frac{1}{2}} = \frac{1}{2} - 0 = \frac{1}{2}$
  - $\frac{1}{3}$ $P((1/3, 2/3)) = \int_{\frac{1}{3}}^{2/3} dx = \left[x \right|_{\frac{1}{3}}^{2/3} = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$
- Think about why the three axioms hold!

Remark: The situation is more complicated when the sample space is uncountably infinite. In fact, we have already hidden an issue that arises when it comes to uncountably infinite sample spaces. We have assumed that $P(E)$ is defined for all of the events E of the sample space, i.e., for all subsets of S. But when dealing with uncountably infinite sample spaces, one can only define $P(E)$ for events E that are *measurable*. (And in order to understand what this means, one needs to study some *measure theory*!) Happily, we don't really need to worry about this issue, because all events of interest to us will be measurable.

2.4 Some Simple Propositions

Proposition 2.4.1. $P(E^c) = 1 - P(E)$.



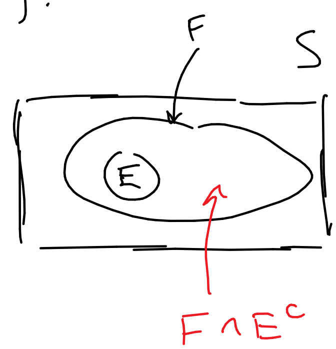
Proof. Note that E and E^c are mutually exclusive and $E \cup E^c = S$. Therefore, by Axioms 2 and 3, we have

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

$$\Rightarrow \quad P(E^c) = 1 - P(E).$$

□

Proposition 2.4.2. *If $E \subseteq F$, then* $P(E) \leq P(F)$.



Proof. Note that

$$F = E \cup (F \cap E^c),$$

and that E and $F \cap E^c$ are mutually exclusive.

$\therefore$ By Axiom 3, we have

$$\begin{aligned} P(F) = P(E \cup (F \cap E^c)) &= P(E) + P(F \cap E^c) \\ &\geq P(E) + 0 \\ &= P(E) \end{aligned}$$

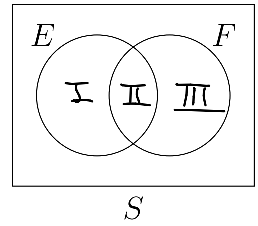
(By Axiom 2)

$\therefore \quad P(F) \geq P(E)$ □

Proposition 2.4.3. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Proof. Note that events I, II, and III are mutually exclusive, and

$$E \cup F = \mathrm{I} \cup \mathrm{II} \cup \mathrm{III}, \quad E = \mathrm{I} \cup \mathrm{II} \quad \text{and} \quad F = \mathrm{II} \cup \mathrm{III}$$



So by Axiom 3, we have

$$P(E \cup F) = P(\mathrm{I} \cup \mathrm{II} \cup \mathrm{III}) = P(\mathrm{I}) + P(\mathrm{II}) + P(\mathrm{III})$$

$$P(E) = P(\mathrm{I}) + P(\mathrm{II})$$

$$P(F) = P(\mathrm{II}) + P(\mathrm{III})$$

$$\mathrm{I} = E \cap F^c$$

$$\mathrm{II} = E \cap F$$

$$\mathrm{III} = E^c \cap F$$

$$\therefore \quad P(E) + P(F) = \underbrace{P(\mathrm{I}) + P(\mathrm{II}) + P(\mathrm{III})} + P(\mathrm{II})$$

$$= P(E \cup F) + P(E \cap F),$$

which gives $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. □

Example 2.4.4. The probability that a disc golfer gets...

- a birdie on Hole 1 is 0.3;
- a birdie on Hole 2 is 0.5; and
- a birdie on both Hole 1 and Hole 2 is 0.2.

What is the probability that they get a birdie on neither Hole 1 nor Hole 2?

For $i \in \{1, 2\}$, let E_i be the event that the disc golfer gets a birdie on hole i. We know that

$$P(E_1) = 0.3, \quad P(E_2) = 0.5 \quad \text{and} \quad P(E_1 \cap E_2) = 0.2$$

We want $P(E_1^c \cap E_2^c) = P((E_1 \cup E_2)^c) = 1 - P(E_1 \cup E_2) = 0.4$

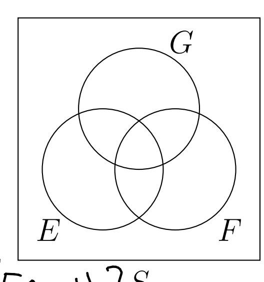
By Prop. 2.4.3, $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = 0.3 + 0.5 - 0.2 = 0.6$

We can apply Proposition 2.4.3 repeatedly to find an expression for the probability of the union of three events:

$$P((E \cup F) \cup G) = P(E \cup F) + P(G) - P((E \cup F) \cap G)$$

$$= P(E) + P(F) - P(E \cap F) + P(G)$$

$$- P((E \cap G) \cup (F \cap G))$$



$$= P(E) + P(F) + P(G) - P(E \cap F)$$

$$- [P(E \cap G) + P(F \cap G) - P((E \cap G) \cap (F \cap G))]$$

$$= P(E) + P(F) + P(G)$$

$$- P(E \cap F) - P(E \cap G) - P(F \cap G)$$

$$+ P(E \cap F \cap G)$$

In general, we have the following result, which can be proven by mathematical induction.

Theorem 2.4.5 (The principle of inclusion-exclusion).

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2})$$
$$+ \ldots + (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$
$$+ \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$
$$= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$

- Summing only the terms involving one event gives an upper bound on the probability of the union:

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) \leq \sum_{i=1}^{n} P(E_i)$$

- Summing the terms involving at most two events gives a lower bound:

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) \geq \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2})$$

- Summing the terms involving at most three events gives an upper bound:

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) \leq \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1 < i_2 < i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3})$$

- And so on – the truncated sums continue to alternate between upper and lower bounds.

The first of these inequalities is used frequently enough that we give it a name.

Proposition 2.4.6 (Boole's inequality). $P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq P(E_1) + P(E_2) + \ldots + P(E_n)$

Example 2.4.7. You wake up early to go for an overnight hiking trip to Trophy Mountain. The probability that it will rain during the day today is 0.2, the probability that it will rain overnight is 0.3, and the probability that it will rain tomorrow is 0.4. What can you say about the probability that you will get rained on?

E_1 E_3 E_2

By Boole's Inequality,

$$P(E_1 \cup E_2 \cup E_3) \leq P(E_1) + P(E_2) + P(E_3)$$
$$= 0.2 + 0.3 + 0.4$$
$$= 0.9$$

2.5 Sample Spaces Having Equally Likely Outcomes

In many experiments, it is natural to assume that all outcomes in the sample space are equally likely to occur.

Example 2.5.1. If two six-sided dice are rolled, what is the probability that the numbers appearing on the dice will sum to 5?

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Each of these 36 outcomes is equally likely to occur.

This probability is

$$\frac{4}{36} = \frac{1}{9}$$

In general, consider an experiment whose sample space S is a finite set, say

$$S = \{1, 2, \ldots, N\},$$

and suppose that all outcomes in S are equally likely to occur, i.e., that

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}).$$

Then by Axioms 2 and 3, we have

$$\begin{aligned} 1 = P(S) = P(\{1\} \cup \{2\} \cup \cdots \cup \{N\}) &= P(\{1\}) + P(\{2\}) + \cdots + P(\{N\}) \\ &= \sum_{i=1}^{N} P(\{i\}) \\ &= N\,P(\{1\}) \end{aligned}$$

$$\Rightarrow \quad P(\{1\}) = \frac{1}{N} = P(\{2\}) = \cdots = P(\{N\})$$

The following proposition is obtained from the above observation and Axiom 3.

Proposition 2.5.2. *If all outcomes of an experiment with a finite sample space are equally likely to occur, then for any event E, we have*

$$P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}$$

Example 2.5.3. A bag contains five red balls and seven blue balls, and three balls are drawn from the bag randomly without replacement.

(a) What is the probability that all three balls are blue?

↳ Let E be the event that all 3 balls are blue.

$$P(E) = \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S}$$

$$= \frac{\binom{7}{3}}{\binom{12}{3}}$$

The outcomes are unordered selections of 3 balls.

$$= \frac{\frac{7!}{3!\,4!}}{\frac{12!}{3!\,9!}} =$$

(b) What is the probability that at least two of the balls are red?

↳ Let F be this event.

$$P(F) = \frac{\#\text{ of outcomes in } F}{\#\text{ of outcomes in } S}$$

$$= \frac{\binom{5}{3} + \binom{5}{2}\cdot\binom{7}{1}}{\binom{12}{3}} =$$

(c) What is the probability that the first ball is red, and the other two are blue?

↳ Call this the event G.

We now consider the outcome to be an ordered selection of 3 balls. We assume that all outcomes are equally likely.

$$P(G) = \frac{\#\text{ of outcomes in } G}{\#\text{ of outcomes in } S}$$

$$= \frac{5\cdot 7\cdot 6}{12\cdot 11\cdot 10}$$

We saw in the previous example that when the experiment consists of a random selection of k items from a set of n items, we can let the outcome of the experiment be either of the following.

- The unordered selection of the k items. From this perspective, the number of possible outcomes is: $\binom{n}{k}$

  Some authors use this notation

- The ordered selection of the k items. From this perspective, the number of possible outcomes is: $n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot (n-k+1) = \frac{n!}{(n-k)!} = (n)_k$

Our choice will depend on the event(s) that we are interested in, and our personal preference.

Example 2.5.4. Suppose that four people are randomly selected from among seven pairs of university dorm roommates. Let E be the event that no two of the four chosen people are roommates.

(a) Find $P(E)$ by viewing the outcome as an unordered selection.

$$P(E) = \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S}$$

$$= \frac{\binom{7}{4} \cdot 2^4}{\binom{14}{4}}$$

choose 4 of the 7 rooms

then choose one of the two people in each room.

$$= \frac{80}{143}$$

(b) Find $P(E)$ by viewing the outcome as an ordered selection.

$$P(E) = \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S}$$

$$= \frac{14 \cdot 12 \cdot 10 \cdot 8}{14 \cdot 13 \cdot 12 \cdot 11}$$

$$= \frac{80}{143}$$

Card Games

A standard deck of cards consists of 52 cards.

- There are four suits: *diamonds* and *hearts* are red, while *clubs* and *spades* are black.
- There are thirteen cards in each suit, ranked in value from lowest to highest as follows:

  A(ce), 2, 3, 4, 5, 6, 7, 8, 9, 10, J(ack), Q(ueen), K(ing), and A(ce) again!

  Bewildering Note: Aces are ranked lowest in some games, highest in some other games, and can play either role in yet other games. In poker, Aces can serve as either the highest card or the lowest card, but not both simultaneously.

Example 2.5.5. A 5-card poker hand is dealt randomly from a standard deck of 52 cards.

(a) A *straight flush* is a hand consisting of 5 cards of consecutive value of the same suit. What is the probability that one is dealt a straight flush?

e.g. $5\diamondsuit\ 6\diamondsuit\ 7\diamondsuit\ 8\diamondsuit\ 9\diamondsuit$

Let the outcome be an unordered selection of five cards, and it seems reasonable to assume that all outcomes are equally likely.

$$P(\text{straight flush}) = \frac{\#\text{ of outcomes that result in a straight flush}}{\#\text{ of outcomes in } S}$$

$$= \frac{4 \cdot 10}{\binom{52}{5}}$$

choice of suit

choice of the lowest card in the hand.

(b) A *straight* is a hand consisting of 5 cards of consecutive value, not all of the same suit. What is the probability that one is dealt a straight?

e.g. $5\diamondsuit\ 6\heartsuit\ 7\heartsuit\ 8\spadesuit\ 9\diamondsuit$

Again, let the outcome be an unordered selection of five cards.

$$P(\text{straight}) = \frac{\#\text{ of outcomes that result in a straight}}{\#\text{ of outcomes in } S}$$

$$= \frac{10 \cdot 4^5 - 4 \cdot 10}{\binom{52}{5}}$$

choice of the lowest card

choice of suit for each of the five cards

# of ways to get a straight flush.

Example 2.5.6. In the game of bridge, the entire deck of 52 cards is dealt out to 4 players.

(a) What is the probability that each player receives exactly one ace?

We let the outcome be a division of 52 cards into 4 distinct groups. So the total number of outcomes in S is $\binom{52}{13,13,13,13}$.

Let A be the event that each player receives exactly one ace. Assuming that all outcomes are equally likely, we have

$$P(A) = \frac{\#\text{ of outcomes in } A}{\#\text{ of outcomes in } S} = \frac{\binom{4}{1,1,1,1}\binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}} = \frac{4! \cdot \frac{48!}{(12!)^4}}{\frac{52!}{(13!)^4}}$$

(b) What is the probability that one player receives all 4 aces?

call this event B

\# of ways to distribute aces

$$P(B) = \frac{4 \cdot \binom{48}{13,13,13,9}}{\binom{52}{13,13,13,13}}$$

\# of ways to distribute non-aces

$$=$$

Example 2.5.7. A room contains n random people. For simplicity, we assume that no one in the room is born on February 29.

(a) What is the probability that at least two of the people have the same birthday?

Let the outcome be an ordered list of birthdays, and we'll assume that all outcomes are equally likely.
Let E be the event that no two people have the same birthday. Then we want $P(E^c) = 1 - P(E)$.

We have

$$P(E) = \frac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365-n+1)}{365^n}$$

$$= \frac{\frac{365!}{(365-n)!}}{365^n}$$

$$= \frac{365!}{(365-n)!\, 365^n}$$

So the probability that at least two people have the same birthday is $P(E^c) = 1 - \dfrac{365!}{(365-n)!\, 365^n}$

(b) How large does n need to be so that this probability is greater than $\frac{1}{2}$?

We used Sage (Google "Sage Cell") to find that n only needs to be 23 or more.

Example 2.5.8. Suppose that N people at a party throw their hats into the center of the room. The hats are mixed up, and then each person randomly selects a hat. What is the probability that no one selects their own hat?

Solution: We label the people and their hats with the numbers 1 through N. We regard the outcome as a vector of N numbers, where the ith element is the number of the hat drawn by the ith person. For example, the outcome $(1, 2, 3, \ldots, N)$ is the one where everyone gets their own hat back.

- The total number of (equally likely) outcomes is: $N!$
- Let E_i be the event that person i selects their own hat
- Then the event that at least one person selects their own hat is: $E_1 \cup E_2 \cup \ldots \cup E_N = \bigcup_{i=1}^{N} E_i$
- So the event that no one selects their own hat is: $(E_1 \cup E_2 \cup \ldots \cup E_n)^c$

By the inclusion-exclusion principle, we have

$$P\left(\bigcup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} P(E_i) - \sum_{i_1<i_2} P(E_{i_1} \cap E_{i_2}) + \ldots + (-1)^{n+1} \sum_{i_1<i_2<\ldots<i_n} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_n}) + \ldots + (-1)^{N+1} P(E_1 \cap E_2 \cap \ldots \cap E_N)$$

Now consider the event $E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}$.

- In words, this is the event that people $i_1, i_2, \ldots, i_n$ select their own hat.
- The number of outcomes in $E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}$ is: $(N-n)!$
- Hence $P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}) = \dfrac{(N-n)!}{N!}$
- Further, the number of terms in $\sum_{i_1<i_2<\cdots<i_n} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n})$ is: $\binom{N}{n} = \dfrac{N!}{n!(N-n)!}$
- So we have $\sum_{i_1<i_2<\cdots<i_n} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}) = \dfrac{\cancel{(N-n)!}}{\cancel{N!}} \cdot \dfrac{\cancel{N!}}{n!\,\cancel{(N-n)!}} = \dfrac{1}{n!}$

Thus we have

$$P\left(\bigcup_{i=1}^{N} E_i\right) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{N+1} \cdot \frac{1}{N!}$$

and the desired probability is

$$1 - P\left(\bigcup_{i=1}^{N} E_i\right) = \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^N}{N!} = \sum_{n=0}^{N} \frac{(-1)^n}{n!}$$

Note: As N approaches infinity, this probability approaches:

Remember: $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} = e^{-1} = \frac{1}{e}$$

"derangements".

* This section is optional.

2.6 Probability as a Continuous Set Function

Definition 2.6.1. A sequence of events $E_1, E_2, \ldots$ is said to be *increasing* if

$$E_1 \subseteq E_2 \subseteq \ldots \subseteq E_i \subseteq \ldots$$

and *decreasing* if

$$E_1 \supseteq E_2 \supseteq \ldots \supseteq E_i \supseteq \ldots$$

Proposition 2.6.2. *If $E_1, E_2, \ldots$ is an increasing sequence of events, then*

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \lim_{i\to\infty} P(E_i)$$

If $E_1, E_2, \ldots$ is a decreasing sequence of events, then

$$P\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{i\to\infty} P(E_i)$$

Proof. The proof is in the textbook if you are interested.

Example 2.6.3. Suppose that a fair coin is flipped repeatedly, ad infinitum. What is the probability that all of the tosses land on heads?

Let E_i be the event that the first i tosses come up heads. Then

$$E_1 \supseteq E_2 \supseteq \ldots \supseteq E_i \supseteq \ldots$$

so we have a decreasing sequence of events.

Note that $P(E_i) = \dfrac{1}{2^i}$

← only one outcome with all i flips heads

← total # of outcomes after i flips

So by Proposition 2.6.2, we have

$$\underbrace{P\left(\bigcap_{i=1}^{\infty} E_i\right)}_{} = \lim_{i\to\infty} P(E_i) = \lim_{i\to\infty} \frac{1}{2^i} = 0$$

event that all of the infinitely many flips land heads!

Summary

- The *sample space* S of an experiment is the set of all possible outcomes.

  Note: We can sometimes view the sample space in more than one way, e.g., as a set of either ordered or unordered selections.

- An *event* is a subset of S.

- The *union* of events $E_1, E_2, \ldots, E_n$, denoted $E_1 \cup E_2 \cup \cdots \cup E_n$ or $\bigcup_{i=1}^{n} E_i$ is the set of all outcomes that belong to at least one of the E_i's.

- The *intersection* of events $E_1, E_2, \ldots, E_n$, denoted $E_1 \cap E_2 \cap \cdots \cap E_n$ or $\bigcap_{i=1}^{n} E_i$ is the set of all outcomes that belong to all of the E_i's.

- The *complement* of an event E, denoted E^c is the set of all outcomes that do not belong to E.

- The *null event*, denoted $\emptyset$ is the event containing no outcomes.

- Two events E and F are *mutually exclusive* if $E \cap F = \emptyset$

- Events $E_1, E_2, E_3, \ldots$ are *mutually exclusive* if $E_i \cap E_j = \emptyset$ for all $i \neq j$

For each event E of the sample space S, we assume that a number $P(E)$ is defined in such a way that the following three axioms are satisfied.

Axiom 1. $0 \leq P(E) \leq 1$

Axiom 2. $P(S) = 1$

Axiom 3. For any sequence of mutually exclusive events $E_1, E_2, E_3, \ldots$, we have $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$.

Some useful facts that follow from these axioms:

- $P(E^c) = 1 - P(E)$
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$
- $P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq P(E_1) + P(E_2) + \cdots + P(E_n)$
- If all outcomes of an experiment with a finite sample space are equally likely to occur, then for any event E, we have

$$P(E) = \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S}$$