2.2.1 Conditional Probability

Let’s go back to the example of students in CSE312 liking donuts and ice cream. Recall we defined event $A$ as liking ice cream and event $B$ as liking donuts. Then, remember we had 36 students that only like ice cream ($A \cap B^C$), 7 students that like donuts and ice cream ($A \cap B$), and 13 students that only like donuts ($B \cap A^C$). Let’s also say that we have 14 students that don’t like either ($A^C \cap B^C$). That leaves us with the following picture, which makes up the whole sample space:

Now, what if we asked the question, what’s the probability that someone likes ice cream given that they like donuts? We can approach this with the knowledge that 20 of the students like donuts (13 who don’t like ice cream and 7 who do). What this question is getting at, is given the knowledge that someone likes donuts, what’s the chance that they also like ice cream. Well, 7 of the 20 who like donuts like ice cream, so we are left with the probability $\frac{7}{20}$. We write this as $P(A \mid B)$ (read the probability of $A$ given $B$) and in this case we have the following:

\[
P(A \mid B) = \frac{7}{20} = \frac{|A \cap B|}{|B|} \quad \text{[we got 7 and 20 from the cardinality of these sets]}
\]

\[
= \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{P(A \cap B)}{P(B)} \quad \text{[divide top and bottom by cardinality of } \Omega]\]

\[
= \frac{P(A \cap B)}{P(B)} \quad \text{[if we have equally likely outcomes]}
\]

This intuition (which worked only in the special case equally likely outcomes), leads us to the definition of
conditional probability:

**Definition 2.2.1.1: Conditional Probability**

The **conditional probability** of event $A$ given that event $B$ happened is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

An equivalent and useful formula we can derive (by multiplying both sides by the denominator, $P(B)$, and switching the sides of the equation is:

$$P(A \cap B) = P(A \mid B) P(B)$$

Let’s consider an important question: does $P(A \mid B) = P(B \mid A)$? No!

This is a common misconception we can show with some examples. Given the above example with ice cream, we showed already $P(A \mid B) = \frac{7}{20}$, but $P(B \mid A) = \frac{7}{36}$, and these are not equal.

Consider another example where $W$ is the event that you are wet and $S$ is the event you are swimming. Then, the probability you are wet given you are swimming, $P(W \mid S) = 1$, as if you are swimming you are certainly wet. But, the probability you are swimming given you are wet, $P(S \mid W) \neq 1$, because there are numerous other reasons you could be wet that don’t involve swimming (being in the rain, showering, etc.).

### 2.2.2 Bayes Theorem

This brings us to Bayes Theorem:

**Theorem 2.2.2.1: Bayes Theorem**

Let $A, B$ be events with nonzero probability. Then,

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

Note that in the above $P(A)$ is called the **prior**, which is our belief without knowing anything about event $B$. $P(A \mid B)$ is called the **posterior**, our belief after learning that event $B$ occurred.

This theorem is important because it allows to reverse the conditioning! Notice that both $P(A \mid B)$ and $P(B \mid A)$ appear in this equation. So if we know $P(A)$ and $P(B)$ and can more easily calculate one of $P(A \mid B)$ or $P(B \mid A)$, we can use **Bayes Theorem** to derive the other.

**Proof.** Recall the definition of conditional probability:

$$P(A \cap B) = P(A \mid B) P(B) \quad (2.2.1)$$

Swapping $A$ and $B$ we can also get that:

$$P(B \cap A) = P(B \mid A) P(A) \quad (2.2.2)$$
But, because $A \cap B = B \cap A$ (since these are the outcomes in both events $A$ and $B$, and the order of intersection does not matter), $P(A \cap B) = P(B \cap A)$, so (2.2.1) and (2.2.2) are equal and we have (by setting the right-hand sides equal):

$$P(A \mid B) P(B) = P(B \mid A) P(A)$$

We can divide both sides by $P(B)$ and get Bayes Theorem:

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

2.2.3 Law of Total Probability

<table>
<thead>
<tr>
<th>Definition 2.2.3.1: Partitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-empty events $E_1, \ldots, E_n$ partition the sample space $\Omega$ if they are:</td>
</tr>
<tr>
<td>• <strong>(Exhaustive)</strong> $E_1 \cup E_2 \cup \cdots \cup E_n = \bigcup_{i=1}^n E_i = \Omega$; that is, they cover the entire of the sample space.</td>
</tr>
<tr>
<td>• <strong>(Pairwise Mutually Exclusive)</strong> For all $i \neq j$, $E_i \cap E_j = \emptyset$; that is, none of them overlap.</td>
</tr>
</tbody>
</table>

Note that for any event $E$, $E$ and $E^C$ always form a partition of $\Omega$.

### Examples

Two example partitions can be seen in the image below:

Now, suppose we have some event $F$ which intersects with various events that form a partition of $\Omega$. This is illustrated by the picture below:
Notice that $F$ is composed of its intersection with each of $E_1$, $E_2$, and $E_3$. This means that we can write the following:

$$P(F) = P(F \cap E_1) + P(F \cap E_2) + P(F \cap E_3)$$

Note that $F$ and $E_4$ do not intersect, so $F \cap E_4 = \emptyset$. For completion, we can include $E_4$ in the above equation, because $P(F \cap E_4) = 0$. So, in all we have:

$$P(F) = P(F \cap E_1) + P(F \cap E_2) + P(F \cap E_3) + P(F \cap E_4)$$

This leads us to the law of total probability.

### Theorem 2.2.3.1: Law of Total Probability (LTP)

If events $E_1, \ldots, E_n$ partition $\Omega$, then for any event $F$

$$P(F) = P(F \cap E_1) + \cdots + P(F \cap E_n) = \sum_{i=1}^{n} P(F \cap E_i)$$

Using the definition of conditional probability, $P(F \cap E_i) = P(F \mid E_i) P(E_i)$, we can replace each of the terms above and get the (typically) more useful formula:

$$P(F) = P(F \mid E_1) P(E_1) + \cdots + P(F \mid E_n) P(E_n) = \sum_{i=1}^{n} P(F \mid E_i) P(E_i)$$

That is, to compute the probability of an event $F$ overall; suppose we have $n$ disjoint cases $E_1, \ldots, E_n$ for which we can (easily) compute the probability of $F$ in each of these cases ($P(F \mid E_i)$). Then, take the weighted average of these probabilities, using the probabilities $P(E_i)$ as weights (the probability of being in each case).
Examples

Let’s consider an example in which we are trying to determine the probability that we fail chemistry. Let’s call the event $F$ failing, and consider the three events $E_1$ for getting the Mean Teacher, $E_2$ for getting the Nice Teacher, and $E_3$ for getting the Hard Teacher which partition the sample space.

<table>
<thead>
<tr>
<th>Probability of Teaching You $P(E_i)$</th>
<th>Mean Teacher $E_1$</th>
<th>Nice Teacher $E_2$</th>
<th>Hard Teacher $E_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of Failing You $P(F \mid E_i)$</td>
<td>$1\over8$</td>
<td>$0\over8$</td>
<td>$1\over2$</td>
</tr>
</tbody>
</table>

Then, using the Law of Total Probability (LTP), we have the following:

$$P(F) = \sum_{i=1}^{n} P(F \mid E_i)P(E_i) = P(F \mid E_1)P(E_1) + P(F \mid E_2)P(E_2) + P(F \mid E_3)P(E_3)$$

$$= 1\cdot\frac{6}{8} + 0\cdot\frac{1}{8} + \frac{1}{2}\cdot\frac{1}{8} = \frac{13}{16}$$

Notice to get the probability of failing, what we did was: consider the probability of failing in each of the 3 cases, and take a weighted average of using the probability of each case. This is exactly what the law of total probability lets us do!

Consider the reverse question, what is the probability that we had the Hard Teacher given that we failed? We can answer this question with Bayes Theorem. We want to solve for $P(E_3 \mid F)$ and so Bayes theorem says the following:

$$P(E_3 \mid F) = \frac{P(F \mid E_3)P(E_3)}{P(F)}$$

$$= \frac{1\cdot\frac{1}{8}}{\frac{13}{16}}$$

$$= \frac{1}{13}$$

2.2.4 Bayes Theorem with the Law of Total Probability

Oftentimes, the denominator in Bayes Theorem is hard, so we must compute it using the LTP. Here, we just combine two powerful formulae: Bayes Theorem and the Law of Total Probability:
Theorem 2.2.4.1: Bayes Theorem with the Law of Total Probability

Let events \(E_1, \ldots, E_n\) partition the sample space \(\Omega\), and let \(F\) be another event. Then:

\[
P(E_1 \mid F) = \frac{P(F \mid E_1) P(E_1)}{P(F)} \quad \text{[by Bayes theorem]}
\]

\[
= \frac{P(F \mid E_1) P(E_1)}{\sum_{i=1}^{n} P(F \mid E_i) P(E_i)} \quad \text{[by the law of total probability]}
\]

In particular, in the case of a simple partition of \(\Omega\) into \(E\) and \(E^C\), if \(E\) is an event with nonzero probability, then:

\[
P(E \mid F) = \frac{P(F \mid E) P(E)}{P(F)} = \frac{P(F \mid E) P(E)}{P(F \mid E) P(E) + P(F \mid E^C) P(E^C)} \quad \text{[by the law of total probability]}
\]

### 2.2.5 Exercises

1. We'll investigate two slightly different questions whose answers don't seem that they should be different, but are. Suppose a family has two children (whom at birth, were each equally likely to be male or female). Let's say a telemarketer calls home and one of the two children picks up.

(a) If the child who responded was male, and says “Let me get my older sibling”, what is the probability that both children are male?

(b) If the child who responded was male, and says “Let me get my other sibling”, what is the probability that both children are male?

**Solution:** There are four equally likely outcomes, MM, MF, FM, and FF (where M represents male and F represents female). Let \(A\) be the event both children are male.

(a) In this part, we’re given that the younger sibling is male. So we can rule out 2 of the 4 outcomes above and we’re left with MF and MM. Out of these two, in one of these cases we get MM, and so our desired probability is \(1/2\).

More formally, let this event be \(B\), which happens with probability \(2/4\) (2 out of 4 equally likely outcomes). Then, \(P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{2/4} = \frac{1}{2}\), since \(P(A \cap B)\) is the probability both children are male, which happens in 1 out of 4 equally likely scenarios. This is because the older sibling’s sex is independent of the younger sibling’s, so knowing the younger sibling is male doesn’t change the probability of the older sibling being male (which is what we computed just now).

(b) In this part, we’re given that at least one sibling is male. That is, out of the 4 outcomes, we can only rule out the FF option. Out of the remaining options MM, MF, and FM, only one has both siblings being male. Hence, the probability desired is \(1/3\). You can do a similar more formal argument like we did above!

See how a slight wording change changed the answer?

2. Suppose we have four fair die: one with three sides, one with four sides, one with five sides, and one with six sides (The numbering of an \(n\)-sided die is \(1, 2, \ldots, n\)). We pick one of the four die, each with equal probability, and roll the same die three times. We get all 4’s. What is the probability we chose
the 5-sided die to begin with?

**Solution:** Let $D_i$ be the event we rolled the $i$-sided die, for $i = 3, 4, 5, 6$. Notice that these $D_3, D_4, D_5, D_6$ partition the sample space.

\[
P(D_5|444) = \frac{P(444|D_5)P(D_5)}{P(444)} \quad \text{[by bayes theorem]} \]

\[
= \frac{P(444|D_5)P(D_5)}{P(444|D_3)P(D_3) + P(444|D_4)P(D_4) + P(444|D_5)P(D_5) + P(444|D_6)P(D_6)} \quad \text{[by ltp]} \\
= \frac{\frac{1}{5^3} \cdot \frac{1}{4}}{\frac{1}{3^3} \cdot \frac{1}{3} + \frac{1}{4^3} \cdot \frac{1}{4} + \frac{1}{5^3} \cdot \frac{1}{4} + \frac{1}{6^3} \cdot \frac{1}{4}} \\
= \frac{1}{64} + \frac{1}{125} + \frac{1}{216} \\
= \frac{1728}{6103} \approx 0.2831
\]

Note that we compute $P(444|D_i)$ by noting there’s only one outcome where we get (4,4,4) out of the $i^3$ equally likely outcomes. This is true except when $i = 3$, where it’s not possible to roll all 4’s.