1.4.0 Conditional Probability

In this section, we discuss one of the most fundamental concepts in probability theory. Here is the question: as you obtain additional information, how should you update probabilities of events? For example, suppose that in a certain city, 23 percent of the days are rainy. Thus, if you pick a random day, the probability that it rains that day is 23 percent:

$$P(R) = 0.23, \text{where } R \text{ is the event that it rains on the randomly chosen day.}$$

Now suppose that I pick a random day, but I also tell you that it is cloudy on the chosen day. Now that you have this extra piece of information, how do you update the chance that it rains on that day? In other words, what is the probability that it rains **given that** it is cloudy? If C is the event that it is cloudy, then we write this as $P(R|C)$, the *conditional probability of R given that C has occurred*. It is reasonable to assume that in this example, $P(R|C)$ should be larger than the original $P(R)$, which is called the **prior probability** of R. But what exactly should $P(R|C)$ be? Before providing a general formula, let's look at a simple example.

Example 1.15
I roll a fair die. Let A be the event that the outcome is an odd number, i.e., $A = \{1, 3, 5\}$. Also let B be the event that the outcome is less than or equal to 3, i.e., $B = \{1, 2, 3\}$. What is the probability of A, $P(A)$? What is the probability of A given B, $P(A|B)$?

> **Solution**

This is a finite sample space, so

$$P(A) = \frac{|A|}{|S|} = \frac{|\{1, 3, 5\}|}{6} = \frac{1}{2}.$$

Now, let's find the conditional probability of A given that B occurred. If we know B has occurred, the outcome must be among $\{1, 2, 3\}$. For A to also happen the outcome must be in $A \cap B = \{1, 3\}$. Since all die rolls are equally likely, we argue that $P(A|B)$ must be equal to

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{2}{3}.$$

Now let's see how we can generalize the above example. We can rewrite the calculation by dividing the numerator and denominator by $|S|$ in the following way

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

Although the above calculation has been done for a finite sample space with equally likely outcomes, it turns out the resulting formula is quite general and can be applied in any setting. Below, we formally provide the formula and then explain the intuition behind it.

> If A and B are two events in a sample space S, then the **conditional probability of A given B** is defined as
>
> $$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0.$$

Here is the intuition behind the formula. When we know that B has occurred, every outcome that is outside B should be discarded. Thus, *our sample space is reduced to the set B*, Figure 1.21. Now the only way that A can happen is when the outcome belongs to the set $A \cap B$. We divide $P(A \cap B)$ by $P(B)$, so that the conditional probability of the new sample space becomes 1, i.e., $P(B|B) = \frac{P(B \cap B)}{P(B)} = 1$.

Note that conditional probability of $P(A|B)$ is undefined when $P(B) = 0$. That is okay because if $P(B) = 0$, it means that the event B never occurs so it does not make sense to talk about the probability of A given B.
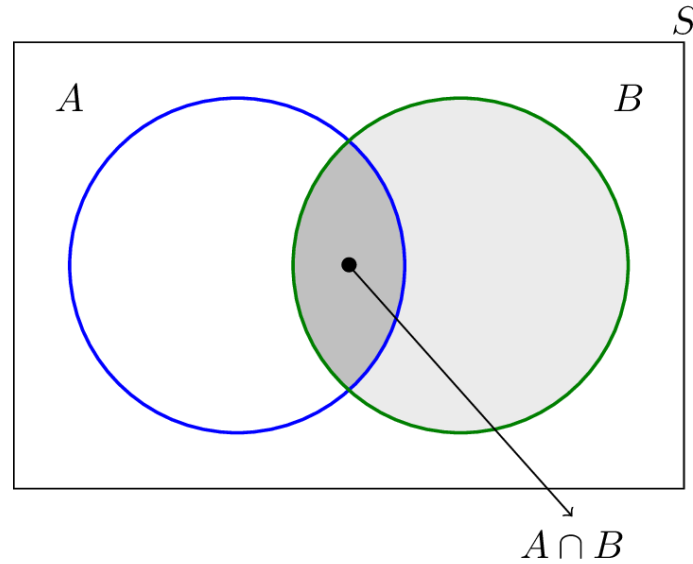
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Fig. 1.21 - Venn diagram for conditional probability, $P(A|B)$.

It is important to note that conditional probability itself is a probability measure, so it satisfies probability axioms. In particular,

- Axiom 1: For any event A, $P(A|B) \geq 0$.
- Axiom 2: Conditional probability of B given B is 1, i.e., $P(B|B) = 1$.
- Axiom 3: If $A_1, A_2, A_3, \cdots$ are disjoint events, then
  $P(A_1 \cup A_2 \cup A_3 \cdots |B) = P(A_1|B) + P(A_2|B) + P(A_3|B) + \cdots$.

In fact, all rules that we have learned so far can be extended to conditional probability. For example, the formulas given in Example 1.10 can be rewritten:

Example 1.16
For three events, A, B, and C, with $P(C) > 0$, we have

- $P(A^c|C) = 1 - P(A|C)$;
- $P(\emptyset|C) = 0$;
- $P(A|C) \leq 1$;
- $P(A - B|C) = P(A|C) - P(A \cap B|C)$;
- $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$;
- if $A \subset B$ then $P(A|C) \leq P(B|C)$.

Let's look at some special cases of conditional probability:

- When A and B are disjoint: In this case $A \cap B = \emptyset$, so

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(\emptyset)}{P(B)} \\
&= 0.
\end{aligned}
$$

  This makes sense. In particular, since A and B are disjoint they cannot both occur at the same time. Thus, given that B has occurred, the probability of A must be zero.

- When B is a subset of A: If $B \subset A$, then whenever B happens, A also happens. Thus, given that B occurred, we expect that probability of A be one. In this case $A \cap B = B$, so

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(B)}{P(B)} \\
&= 1.
\end{aligned}
$$

- When A is a subset of B: In this case $A \cap B = A$, so

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(A)}{P(B)}.
\end{aligned}
$$

Example 1.17

I roll a fair die twice and obtain two numbers $X_1 =$ result of the first roll and $X_2 =$ result of the second roll. Given that I know $X_1 + X_2 = 7$, what is the probability that $X_1 = 4$ or $X_2 = 4$?

Solution

Let A be the event that $X_1 = 4$ or $X_2 = 4$ and B be the event that $X_1 + X_2 = 7$. We are interested in $P(A|B)$, so we can use

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

We note that

$$A = \{(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (1,4), (2,4), (3,4), (4,4), (5,4), (6,4)\},$$

$$B = \{(6,1), (5,2), (4,3), (3,4), (2,5), (1,6)\},$$

$$A \cap B = \{(4,3), (3,4)\}.$$

We conclude

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

$$= \frac{\frac{2}{36}}{\frac{6}{36}}$$

$$= \frac{1}{3}.$$

Let's look at a famous probability problem, called the two-child problem. Many versions of this problem have been discussed [1] in the literature and we will review a few of them in this chapter. We suggest that you try to guess the answers before solving the problem using probability formulas.

Example 1.18
Consider a family that has two children. We are interested in the children's genders. Our sample space is $S = \{(G,G), (G,B), (B,G), (B,B)\}$. Also assume that all four possible outcomes are equally likely.

a. What is the probability that both children are girls given that the first child is a girl?
b. We ask the father: "Do you have at least one daughter?" He responds "Yes!" Given this extra information, what is the probability that both children are girls? In other words, what is the probability that both children are girls given that we know at least one of them is a girl?

Let A be the event that both children are girls, i.e., $A = \{(G,G)\}$. Let B be the event that the first child is a girl, i.e., $B = \{(G,G),(G,B)\}$. Finally, let C be the event that at least one of the children is a girl, i.e., $C = \{(G,G),(G,B),(B,G)\}$. Since the outcomes are equally likely, we can write

$$P(A) = \frac{1}{4},$$

$$P(B) = \frac{2}{4} = \frac{1}{2},$$

$$P(C) = \frac{3}{4}.$$

a. What is the probability that both children are girls given that the first child is a girl? This is $P(A|B)$, thus we can write

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{P(A)}{P(B)} \qquad (\text{since } A \subset B)$$
$$= \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

b. What is the probability that both children are girls given that we know at least one of them is a girl? This is $P(A|C)$, thus we can write

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$
$$= \frac{P(A)}{P(C)} \qquad (\text{since } A \subset C)$$
$$= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Discussion: Asked to guess the answers in the above example, many people would guess that both $P(A|B)$ and $P(A|C)$ should be 50 percent. However, as we see $P(A|B)$ is 50 percent, while $P(A|C)$ is only 33 percent. This is an example where the answers might seem counterintuitive. To understand the results of this problem, it is helpful to

note that the event B is a subset of the event C. In fact, it is strictly smaller: it does not include the element (B, G), while C has that element. Thus the set C has more outcomes that are not in A than B, which means that $P(A|C)$ should be smaller than $P(A|B)$.

It is often useful to think of probability as percentages. For example, to better understand the results of this problem, let us imagine that there are 4000 families that have two children. Since the outcomes $(G, G), (G, B), (B, G)$, and (B, B) are equally likely, we will have roughly 1000 families associated with each outcome as shown in Figure 1.22. To find probability $P(A|C)$, we are performing the following experiment: we choose a random family from the families with at least one daughter. These are the families shown in the box. From these families, there are 1000 families with two girls and there are 2000 families with exactly one girl. Thus, the probability of choosing a family with two girls is $\frac{1}{3}$.
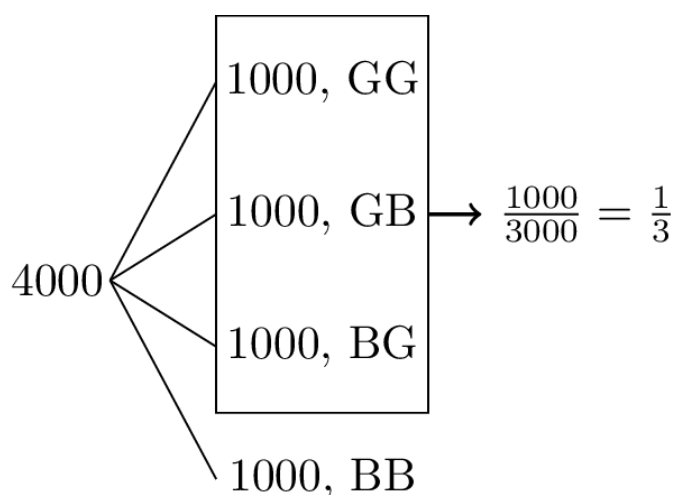

Fig.1.22 - An example to help the understanding of $P(A|C)$ in Example 1.18.

Chain rule for conditional probability:

Let us write the formula for conditional probability in the following format

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B) \tag{1.5}$$

This format is particularly useful in situations when we know the conditional probability, but we are interested in the probability of the intersection. We can interpret this formula using a tree diagram such as the one shown in Figure 1.23. In this figure, we obtain the probability at each point by multiplying probabilities on the branches leading to that point. This type of diagram can be very useful for some problems.
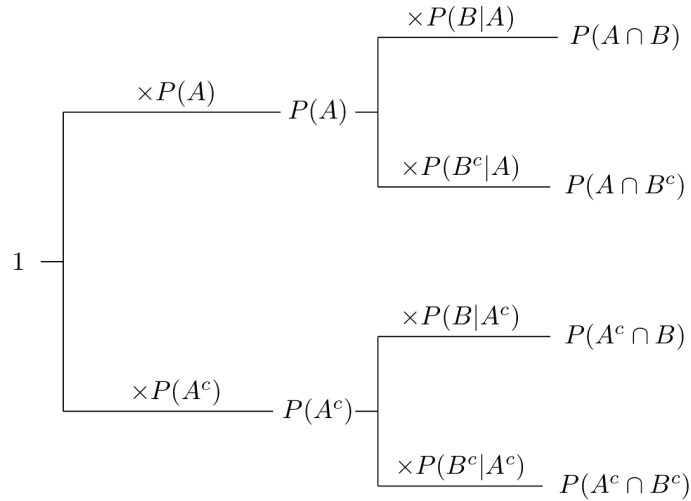
Fig.1.23 - A tree diagram.

Now we can extend this formula to three or more events:

$$P(A \cap B \cap C) = P\big(A \cap (B \cap C)\big) = P(A)P(B \cap C|A) \tag{1.6}$$

From Equation 1.5,

$$P(B \cap C) = P(B)P(C|B).$$

Conditioning both sides on A, we obtain

$$P(B \cap C|A) = P(B|A)P(C|A, B) \tag{1.7}$$

Combining Equation 1.6 and 1.7 we obtain the following chain rule:

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A, B).$$

The point here is understanding how you can derive these formulas and trying to have intuition about them rather than memorizing them. You can extend the tree in Figure 1.22 to this case. Here the tree will have eight leaves. A general statement of the chain rule for n events is as follows:

Chain rule for conditional probability:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1) \cdots P(A_n|A_{n-1}A_{n-2} \cdots A_1)$$

Example 1.19

In a factory there are 100 units of a certain product, 5 of which are defective. We pick three units from the 100 units at random. What is the probability that none of them are defective?

Let us define A_i as the event that the ith chosen unit is not defective, for $i = 1, 2, 3$. We are interested in $P(A_1 \cap A_2 \cap A_3)$. Note that

$$P(A_1) = \frac{95}{100}.$$

Given that the first chosen item was good, the second item will be chosen from 94 good units and 5 defective units, thus

$$P(A_2 | A_1) = \frac{94}{99}.$$

Given that the first and second chosen items were okay, the third item will be chosen from 93 good units and 5 defective units, thus

$$P(A_3 | A_2, A_1) = \frac{93}{98}.$$

Thus, we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1)$$
$$= \frac{95}{100} \frac{94}{99} \frac{93}{98}$$
$$= 0.8560$$

As we will see later on, another way to solve this problem is to use counting arguments.