9 Conditional Probability

Suppose A and B are events in a sample space S. If we are told that B has occurred what effect does this have on the probability that A has occurred? We can use conditional probability to answer this question.

Definition If A and B are events and $\mathbb{P}(B) \neq 0$ then the *conditional probability of A given B* is

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

This is usually denoted by $\mathbb{P}(A|B)$.

Note that the definition does not assume that A happens after B. One way of thinking of this is to imagine the experiment is performed secretly and the fact that B occurred is revealed to you (without the full outcome being revealed). The conditional probability of A given B is the new probability of A in these circumstances.

This is an important definition but it is often the source of great confusion; make sure you understand it and can use it. Also, don't confuse $\mathbb{P}(A|B)$ with $\mathbb{P}(A \setminus B)$.

One use of conditional probability is to calculate the probability of the intersection of several events. We saw in the last section that if the events are mutually independent then the probability of their intersection is just the product of their probabilities. If the events are not mutually independent then we need to use conditional probability. We will need to consider conditional probabilities like $\mathbb{P}(A_3|A_1 \cap A_2)$ i.e. the conditional probability of A_3 given that both A_1 and A_2 occur.

Theorem 9.1. *Let $A_1, A_2, \ldots, A_n$ be events in a sample space S. Suppose that $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$. Then*

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_3|A_1 \cap A_2) \times \ldots \times \mathbb{P}(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

We use induction on n to prove this. The idea is that we first prove that the statement is true for $n = 2$ (*the base case of the induction*). We then assume that $n \geq 3$ and that the statement of the theorem is true whenever we have less than n events (*the induction hypothesis*). We then use this assumption to prove that the statement of the theorem is true when we have

n events (*the inductive step*). This is an important proof technique. You will meet it again in many other modules.

Proof of Theorem 9.1

Base case. Suppose $n = 2$. Using the definition of conditional probability we have

$$\mathbb{P}(A_1)\mathbb{P}(A_2|A_1) = \mathbb{P}(A_1)\frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)} = \mathbb{P}(A_1 \cap A_2)$$

since $A_2 \cap A_1 = A_1 \cap A_2$. Thus the statement of the theorem is true when $n = 2$.

Induction Hypothesis Suppose that $n \geq 3$ and that the statement of the theorem is true whenever we have less than n events. In particular the statement of the theorem is true when we have $n-1$ events. Since $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{n-2}) \geq \mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$ we may deduce that

$$\begin{aligned}&\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) = \\ &\quad \mathbb{P}(A_1) \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_3|A_1 \cap A_2) \times \ldots \times \mathbb{P}(A_{n-1}|A_1 \cap A_2 \cap \ldots \cap A_{n-2}).\end{aligned}$$

Inductive Step Let $B = A_1 \cap A_2 \cap \cdots \cap A_{n-1}$. Then

$$\begin{aligned}&\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \\ &\quad = \mathbb{P}(B \cap A_n) \\ &\quad = \mathbb{P}(B) \times \mathbb{P}(A_n|B) \\ &\quad = \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \times \mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \\ &\quad = \mathbb{P}(A_1) \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_3|A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})\end{aligned}$$

where the second equality uses the base case to evaluate $\mathbb{P}(B \cap A_n)$, the third equality is a direct substitution for B, and the fourth equality uses the induction hypothesis to evaluate $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1})$. Thus the truth of of the theorem for less than n events implies the theorem is also true when we have n events.

We may now deduce that the theorem is true for all values of n. •

We saw examples in lectures of how this theorem can be used to calculate the probability of the intersection of several events.

The Theorem of Total Probability

In this subsection we will see how conditional probability can be used to calculate the probability of an event when the sample space has been 'partitioned' into several events.

Definition The events $E_1, E_2, \ldots, E_n$ *partition* S if they are non-empty pairwise disjoint sets and their union is the whole of S.

Theorem 9.2 (The Theorem of Total Probability). *Suppose that $E_1, E_2, \ldots, E_m$ are events which partition the sample space S, and that $\mathbb{P}(E_i) > 0$ for all $1 \leq i \leq m$. Let A be an event in S. Then*

$$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(A|E_i)\mathbb{P}(E_i).$$

Proof Let $A_i = A \cap E_i$ for $1 \leq i \leq m$. We have $\mathbb{P}(A|E_i) = \mathbb{P}(A \cap E_i)/\mathbb{P}(E_i) = \mathbb{P}(A_i)/\mathbb{P}(E_i)$ so $\mathbb{P}(A_i) = \mathbb{P}(A|E_i)\,\mathbb{P}(E_i)$. The fact that $E_1, E_2, \ldots E_m$ partition S implies that $A_1, A_2, \ldots A_m$ partition A. Hence by Kolmogorov's third axiom:

$$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(A_i) = \sum_{i=1}^{m} \mathbb{P}(A|E_i)\mathbb{P}(E_i).$$

•

There is also an analogue of the theorem of total probability for conditional probabilities.

Theorem 9.3. *Suppose that $E_1, E_2, \ldots, E_m$ are events which partition the sample space S. Let A and B be an events in S with $\mathbb{P}(B \cap E_i) > 0$ for all $1 \leq i \leq m$. Then*

$$\mathbb{P}(A|B) = \sum_{i=1}^{m} \mathbb{P}(A|B \cap E_i)\mathbb{P}(E_i|B).$$

Proof The fact that $E_1, E_2, \ldots E_m$ partition S implies that

$$A \cap B \cap E_1, A \cap B \cap E_2, \ldots, A \cap B \cap E_m$$

partition $A \cap B$. Hence by Kolmogorov's third axiom:

$$\mathbb{P}(A \cap B) = \sum_{i=1}^{m} \mathbb{P}(A \cap B \cap E_i).$$

Hence

$$\begin{aligned}
\mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{m} \frac{\mathbb{P}(A \cap B \cap E_i)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{m} \frac{\mathbb{P}(A \cap B \cap E_i)}{\mathbb{P}(B \cap E_i)} \times \frac{\mathbb{P}(B \cap E_i)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{m} \mathbb{P}(A|B \cap E_i)\, \mathbb{P}(E_i|B)
\end{aligned}$$

•

Special cases of Theorems 9.2 and 9.3 occur when E is an event with both $\mathbb{P}(E) > 0$ and $\mathbb{P}(E^c) > 0$ i.e. $0 < \mathbb{P}(E) < 1$. Then E, E^c partitions S so for any event A we have

$$\mathbb{P}(A) = \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)\mathbb{P}(E^c).$$

Furthermore, if B is an event such that $\mathbb{P}(E|B) > 0$ and $\mathbb{P}(E^c|B) > 0$, then

$$\mathbb{P}(A|B) = \mathbb{P}(A|B \cap E)\mathbb{P}(E|B) + \mathbb{P}(A|B \cap E^c)\mathbb{P}(E^c|B).$$

Example I have two coins in a bag. One is fair and the other has a head on both sides. I select a coin at random from the bag, then throw it twice (without determining which coin it is).
Question 1 What is the probability that the first throw is a head?
Question 2 What is the probability that the second throw is a head, given that the first throw is a head?

Answers Let F be the event that the fair coin is chosen, H_1 be the event that the first throw is a head, and H_2 be the event that the second throw is a head. By Theorem 9.2

$$\mathbb{P}(H_1) = \mathbb{P}(H_1|F)\mathbb{P}(F) + \mathbb{P}(H_1|F^c)\mathbb{P}(F^c) = \frac{1}{2} \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{3}{4}$$

This answers Question 1. By Theorem 9.3 we also have

$$\mathbb{P}(H_2|H_1) = \mathbb{P}(H_2|F \cap H_1)\mathbb{P}(F|H_1) + \mathbb{P}(H_2|F^c \cap H_1)\mathbb{P}(F^c|H_1) \qquad (1)$$

We need to evaluate the terms on the right hand side of this equation. Since $\mathbb{P}(H_1|F \cap H_1)$ is the probability that I get a head on the second throw given that I chose the fair coin and got a head on the first throw, we have $\mathbb{P}(H_1|F \cap H_1) = 1/2$ (the assumption that I got a head on the first throw is irrelevent). Similarly $\mathbb{P}(H_2|F^c \cap H_1) = 1$. It remains to calculate $\mathbb{P}(F|H_1)$ and $\mathbb{P}(F^c|H_1)$. It is not obvious what these are, but it is clear that $\mathbb{P}(H_1|F) = 1/2$ and $\mathbb{P}(H_1)|F^c) = 1$. We can use these values to calculate $\mathbb{P}(F|H_1)$ and $\mathbb{P}(F^c|H_1)$ as follows. We have

$$\mathbb{P}(F|H_1) = \frac{\mathbb{P}(F \cap H_1)}{\mathbb{P}(H_1)} = \frac{\mathbb{P}(F \cap H_1)}{\mathbb{P}(F)} \times \frac{\mathbb{P}(F)}{\mathbb{P}(H_1)} = \mathbb{P}(H_1|F) \times \frac{\mathbb{P}(F)}{\mathbb{P}(H_1)} = \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{3}{4}} = \frac{1}{3}$$

using the answer to Question 1. We can show similarly that $\mathbb{P}(F^c|H_1) = 2/3$, or more simply use $\mathbb{P}(F^c|H_1) = 1 - \mathbb{P}(F|H_1)$. Substituting into (1) we obtain

$$\mathbb{P}(H_2|H_1) = \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{5}{6}$$

Bayes Theorem

We saw in the above example that we could use $\mathbb{P}(H_1|F)$ to calculate $\mathbb{P}(F|H_1)$. The following theorem formalizes this idea.

Theorem 9.4 (Bayes' theorem)**.** *Suppose A and B are events with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Then*

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

Proof We have

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

•