



## Why

How should we modify probabilities, given that we know some aspect of the outcome (i.e., that some event has occurred)? In other words, and roughly speaking, how does knowledge about one aspect of an outcome give us knowledge about another aspect.

## Definition

Suppose  $\Omega$  is a finite set of outcomes with a probability even function  $P$ . Suppose  $A, B \subset \Omega$  are two events and  $P(B) \neq 0$ . The *conditional probability* of  $A$  *given*  $B$  is the ratio of the probability of  $A \cap B$  to the probability of  $B$ . Other language includes the *conditional probability of  $A$  given that  $B$  has happened*.

The frequentist interpretation is straightforward. We collect many outcomes, and  $P(B)$  is the fraction of times that the event  $B$  occurs.  $P(A \cap B)$  is the fraction of times both  $A$  and  $B$  occurs. The conditional probability of  $A$  given  $B$  is the fraction of times that  $A$  happened *among the outcomes in which  $B$  happened*.

## Notation

In a slightly slippery but universally standard notation, we denote the conditional probability of  $A$  given  $B$  by  $P(A \mid B)$ . In other words, we define

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

for all  $A, B \subset \Omega$ , whenever  $P(B) \neq 0$ .

Notice that for any two events,  $P(A \cap B) = P(A)P(B \mid A)$ . Pleasantly, this equation makes sense even if  $P(A) = 0$ , since then  $P(A \cap B) \leq P(A) = 0$  means  $P(A \cap B)$  is also 0—irrespective of our definition of  $P(B \mid A)$  in this case. (Here we have swapped the order of  $A$  and  $B$ ).

For example, we can express the law of total probability (see Event

Probabilities) as

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i),$$

where  $A_1, \dots, A_n$  partition  $\Omega$  and  $B \subset \Omega$  with  $P(B) > 0$ . A particular nice example,

$$P(B) = P(B | A)P(A) + P(B | \Omega - A)P(\Omega - A)$$

### Examples

*Rolling a die.* As usual, model rolling a die with the sample space  $\Omega = \{1, \dots, 6\}$  and distribution  $p : \Omega \rightarrow [0, 1]$  defined by  $p(\omega) = 1/6$  for all  $\omega \in \Omega$ . Take the two events,  $A = \{6\}$  and  $B = \{4, 5, 6\}$ , which we interpret as “the number of pips is 6” and “the number of pips is at least 4”, respectively. Then

$$P(A | B) = \frac{1/6}{1/2} = \frac{1}{3}$$

This has a nice interpretation, given that we know the number of pips face up is either 4 or 5 or 6, the chance that it is 6 is 1 in 3.

### Conditional probability measure

It happens that  $P(\cdot | B)$  is itself a probability event function on  $\Omega$ . We therefore refer to  $P(\cdot | B)$  as a *conditional probability function* or *conditional probability measure*.

To see this, suppose  $B \subset \Omega$  and  $P(B) > 0$ . Then (i)  $P(A | B) \geq 0$  for all  $A \subset \Omega$ , since  $P(A \cap B) \geq 0$ . Moreover, (ii)

$$P(\Omega | B) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$$

Similarly,  $P(\emptyset | B) = P(\emptyset \cap B)/P(B) = 0/P(B) = 0$ .

Finally, (iii) if  $A \cap C = \emptyset$ , then

$$\begin{aligned} P(A \cap C \mid B) &= \frac{P((A \cap C) \cap B)}{P(B)} \\ &\stackrel{(a)}{=} \frac{P(A \cap B) + P(C \cap B)}{P(B)} \\ &= P(A \mid B) + P(C \mid B). \end{aligned}$$

where (a) follows since  $(A \cap C) \cap B = (A \cap B) \cup (C \cap B)$  and the sets  $A \cap B$ ,  $C \cap B$  are disjoint.

### Induced conditional distribution

Since  $P(\cdot \mid B)$  is a probability event function, we expect it to have a corresponding distribution. Denote the distribution of  $P$  by  $p : \Omega \rightarrow [0, 1]$ . In other words,  $p$  satisfies  $p(\omega) = P(\{\omega\})$  as usual. Now define  $q : \Omega \rightarrow \mathbf{R}$  by

$$q(\omega) = \begin{cases} \frac{p(\omega)}{P(B)} & \text{if } \omega \in B \\ 0 & \text{otherwise.} \end{cases}$$

In this case the event probability function induced by  $A$  is  $P(\cdot \mid B)$ . We call  $q$  the *conditional distribution* induced by *conditioning on* the event  $B$ .

### Finite intersections

As a simple repeated application of our definition, suppose  $A_1, \dots, A_n \subset \Omega$ . Then

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2 \mid A_1) \cdots P(A_n \mid A_1 \cap \dots \cap A_{n-1})$$

Many authors call this the *chain rule*. The order of the  $A_i$  is inconsequential.

