



## Why

Since one and only one outcome occurs, given a distribution on outcomes, we define the probability of a set of outcomes as the sum of their probabilities.

## Definition

Suppose  $p$  is a distribution on a *finite* set of outcomes  $\Omega$ . Given an event  $E \subset \Omega$ , the *probability of  $E$  under  $p$*  as the sum of the probabilities of the outcomes in  $E$ . The frequentist interpretation is clear—the probability of an event is the proportion of times any of its outcomes will occur in the long run.

## Notation

It is common to define a function  $P : \mathcal{P}(\Omega) \rightarrow \mathbf{R}$  by

$$P(A) = \sum_{a \in A} p(a) \quad \text{for all } A \subset \Omega$$

We call this function  $P$  the *event probability function* (or the *probability measure*) associated with  $p$ . Since it depends on the sample space  $\Omega$  and the distribution  $p$ , we occasionally denote this dependence by  $P_{\Omega,p}$  or  $P_p$ .

It is tempting, and therefore common to write  $P(\omega)$  when  $\omega \in \Omega$  and one intends to denote  $P(\{\omega\})$ . Of course, this corresponds with  $p(\omega)$ . It is therefore easy to see that from  $P$  we can compute  $p$ , and vice versa.

## Examples

*Rolling a die.* We consider the usual fair die model (see Outcome Probabilities). Here we have  $\Omega = \{1, \dots, 6\}$  and a distribution  $p : \Omega \rightarrow [0, 1]$  defined by

$$p(\omega) = 1/6 \quad \text{for all } \omega \in \Omega$$

Given the model, the probability of the event  $E = \{2, 4, 6\}$  is

$$P(E) = \sum_{\omega \in E} p(\omega) = p(2) + p(4) + p(6) = 1/2.$$

## Properties of event probabilities

The properties of  $p$  ensure that  $P$  satisfies

1.  $P(A) \geq 0$  for all  $A \subset \Omega$ ;
2.  $P(\Omega) = 1$  (and  $P(\emptyset) = 0$ );
3.  $P(A) + P(B)$  for all  $A, B \subset \Omega$  and  $A \cap B = \emptyset$ .

The last statement (3) follows from the more general identity

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

for  $A, B \subset \Omega$ , by using  $\mathbf{P}(\emptyset) = 0$  of (2) above. These three conditions are sometimes called the *axioms of probability for finite sets*.

Do all such  $P$  satisfying (1)-(3) have a corresponding underlying probability distribution? The answer is easily seen to be yes. Suppose  $f : \mathcal{P}(\Omega) \rightarrow \mathbf{R}$  satisfies (1)-(3). Define  $q : \Omega \rightarrow \mathbf{R}$  by  $q(\omega) = f(\{\omega\})$ . If  $f$  satisfies the axioms, then  $q$  is a probability distribution. For this reason we call any function satisfying (i)-(iii) an *event probability function* (or a *(finite) probability measure*).

## Other basic consequences

*Probability by cases.* Suppose  $A_1, \dots, A_n$  partition  $\Omega$ . Then for any  $B \subset \Omega$ ,

$$P(B) = \sum_{i=1}^n P(A_i \cap B).$$

Some authors call this the *law of total probability*. This is easy to see by using the distributive laws of set algebra (see **Set Unions and Intersections**).

*Monotonicity.* If  $A \subseteq B$ , then  $P(A) \leq P(B)$ . This is easy to see by splitting  $B$  into  $A \cap B$  and  $B - A$ , and applying (1) and (3).

*Subadditivity.* For  $A, B \subset \Omega$ ,  $P(A \cup B) \leq P(A) + P(B)$ . This is easy to see from the more general identity in (3) above. This is sometimes referred to as a *union bound*, in reference to *bounding* the quantity  $P(A \cup B)$ .

