



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

Given a distribution $p : \Omega \rightarrow \mathbf{R}$, the *probability of an event* $A \subset \Omega$ is $\sum_{a \in A} p(a)$, the sum of probabilities of its outcomes.

Notation

Define $\mathbf{P} : \mathcal{P}(\Omega) \rightarrow \mathbf{R}$ by

$$\mathbf{P}(A) = \sum_{a \in A} p(a).$$

We call \mathbf{P} the *event probability function* (or *probability measure*) of (or induced by) p .

Example: die

Define $p : \{1, \dots, 6\} \rightarrow \mathbf{R}$ by $p(\omega) = 1/6$ for $\omega = 1, \dots, 6$. Define the event $E = \{2, 4, 6\}$. Then

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

Probability measures

Notice that for all $A \subset \Omega$, (i) $\mathbf{P}(A) \geq 0$. In particular, (ii) $\mathbf{P}(\Omega) = 1$ (and $\mathbf{P}(\emptyset) = 0$). For all $A, B \subset \Omega$, $\mathbf{P}(A \cup B) =$

$\mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$. In particular, if $A \cap B = \emptyset$, (iii) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

Conversely, suppose $f : \mathcal{P}(\Omega) \rightarrow \mathbf{R}$ satisfies (i), (ii), (iii). These three conditions are sometimes called the *axioms of probability* (for finite sets). Define $p : \Omega \rightarrow \mathbf{R}$ by

$$p(\omega) = f(\{\omega\}).$$

In case f satisfies the axioms, p is a probability distribution (nonnegative and sums to one). For this reason we call f satisfying (i)-(iii) an *event probability function* (or *probability measure*). In the case that we think of a probability event function \mathbf{P} as induced by a distribution p , we write \mathbf{P}_p .

We conclude that p and \mathbf{P} are two perspectives. We can think of elementary events (outcomes) and define their probabilities individually in a way that they sum to one and are nonnegative. Or we can think of the compound events, and define their probabilities in a way consistent with (i)-(iii).

Probability by cases

Let \mathbf{P} be a probability event function. Suppose A_1, \dots, A_n partition Ω . Then for any $B \subset \Omega$,

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

Some authors call this the *law of total probability*.

