



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 Ω . Given an event $E \subset \Omega$, the *probability* (or *chance*) of E under p is 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 Ω 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 $P(\{\omega\})$, which is just $p(\omega)$. It is therefore easy to see that from P we can compute p , and vice versa.

Examples

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

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

Rolling two dice. We consider the usual model of rolling two die at once (see **Outcome Probabilities**). We take $\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$. In other words, Ω is $\{1, 2, 3, 4, 5, 6\}^2$. Suppose we model a distribution on outcomes $p : \Omega \rightarrow [0, 1]$ by defining $p(\omega) = 1/36$ for each $\omega \in \Omega$. We use the set $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$ for the event corresponding to the statement that the sum of the two numbers is 5. In other words,

$$A = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = 5\}$$

The probability of A is

$$P(A) = p((1, 4)) + p((2, 3)) + p((3, 2)) + p((4, 1)) = 4/36 = 1/9.$$

Suppose we modify the statement so that $B = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = 12\}$. We have $P(B) = 1/36$. So we have modeled that the sum of the number of the pips on the two die being 12 as less probable than the event that the sum of the number of pips being 5.

Flipping a coin three times. We model flipping a coin three times with the outcome space $\Omega = \{0, 1\}^3$. We interpret $(\omega_1, \omega_2, \omega_3) \in \Omega$ so that ω_1 is the outcome of the first flip—heads is 1 and tails is 0. Suppose we model each outcome as equally probable, and so put a distribution $p : \Omega \rightarrow [0, 1]$ on Ω satisfying $p(\omega) = 1/8$ for every $\omega \in \Omega$. We want to consider all outcomes in which we see two heads. Our model is the event $A \subset \Omega$ defined by

$$A = \{(1, 1, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}$$

Under our chosen distribution, $P(A) = 1/2$.

Flipping a coin n times. We model flipping a coin n times with a sample space $\Omega = \{0, 1\}^n$. Here, we agree to interpret $(\omega_1, \dots, \omega_n) \in \Omega$ so that ω_i is 1 if the coin lands heads on the i th toss and 0 if it lands tails; $i = 1, \dots, n$. The size of Ω is 2^n , since $|\{0, 1\}| = 2$. Suppose we choose a distribution $p : \Omega \rightarrow [0, 1]$ so that

$$p(\omega) = \frac{1}{2^n}$$

Now consider the event H_k defined by

$$H_k = \{\omega \in \Omega \mid |\{i \mid \omega_i = 1\}| = k\}.$$

so that it contains all outcomes having a total of k heads. Then

$$P(H_k) = \frac{|H_k|}{2^n} = \frac{\binom{n}{k}}{2^n}$$

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. $P(A) + P(B) = P(A \cup B) + P(A \cap B)$ for all $A, B \subset \Omega$ and $A \cap B = \emptyset$.

The last statement (3) follows from the more general identity—known as the *inclusion-exclusion formula*—

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

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

Disjoint events. Two events A and B are *disjoint* or *mutually exclusive* if $A \cap B = \emptyset$. Likewise, a list of events A_1, \dots, A_n are *disjoint* or *mutually exclusive* if $A_i \cap A_j = \emptyset$ for all $i \neq j$, $i, j \in \{1, \dots, n\}$. A direct

consequence of (3) above is

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$$

Probability by cases. Suppose A_1, \dots, A_n partition Ω . 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**). A simple consequence is that for any A, B

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

since $A, \Omega - A$ partition Ω .

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 $P(A \cup B)$.

