



Why

A set of outcomes may be finite or infinite. For now, we consider finite samples spaces. To talk about the uncertain outcomes, we assign credibility to each outcome according to our intuition of proportion.¹

Definition

Suppose Ω is a finite set. A $p : \Omega \rightarrow \mathbf{R}$ on Ω is a *probability distribution* on Ω if it is nonnegative (i.e., $p(\omega) \geq 0$ for every $\omega \in \Omega$) and

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

The *probability* of an outcome $\omega \in \Omega$ under the distribution p is the result $p(\omega)$.

Interpretations

There are two usual meanings of the word “probability”. The first, is its intuitive interpretation as frequency—the fraction of times that an outcome ω will occur if we are able to repeat the scenario producing the outcomes many times. This is the so-called *frequentist* viewpoint.

The trouble is that some scenarios are not “repeatable” (e.g., whether it will rain or not *tomorrow*). Thus, it is sometimes natural to think of probabilities as *beliefs* or *degrees of belief* which are updated according to particular rules. This is the so-called *Bayesian viewpoint*.

This second interpretation matches the English etymology: the word probability has its roots in the English word probable, which has the Middle English sense “worthy of belief”. The probability of an outcome models how worthy of belief it is, relative to other outcomes. In the case of flipping a coin, or rolling a die, we may assert that all outcomes are equally worthy of belief.

¹Future editions may drop the dependence on real numbers, and use intuition of repeated trials to introduce *rational* probability distributions.

If a first outcome has a larger probability than a second outcome, we call the first *more probable* than the second. Similarly, we call the second outcome *less probable* than the first outcome.

Examples

Probabilities for flipping a coin. Suppose we model flipping a coin, as before, with the sample space $\{0, 1\}$. We may model both heads and tails as equally worthy of belief. Thus we would like to pick two nonnegative numbers $p(1)$ and $p(2)$ so that they are non-negative and $p(1) + p(2) = 1$. Consequently, we model outcome $p(0) = p(1) = 1/2$. We often refer to this particular model as a *fair coin*. Neither heads nor tails is *more* or *less* probable.

Example: die

When rolling a die, all six sides are equally worthy of belief. Thus we say that the outcomes 1, 2, 3, 4, 5, 6 each have probability 1/6. Prior to the roll, each is equally credible. We speak of modeling a *fair die*. No outcome is *more* or *less* probable than any other.

Other terminology

Other terminology for probability distribution includes *distribution*, *probability mass function*, *pmf*, *proportion distribution*, and *probabilities*.

Simple consequences of definition

Proposition 1. *Suppose $p : A \rightarrow \mathbf{R}$ is a distribution. For all $a \in A$, $p(a) \leq 1$.*

Proof. Let $a \in A$. By definition, $p(a) \geq 0$. Second, since p is normalized, $p(a) \leq \sum_{b \in A} p(b) = 1$. \square

Consequently, the range of p is contained in $[0, 1]$. For this reason, we often introduce a distribution with the notation $p : A \rightarrow [0, 1]$; notice the

codomain here is the interval $[0, 1]$ and *not* the real numbers \mathbf{R} .

