



Why

A set of outcomes may be finite or infinite. For now, we consider finite sample 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 function $p : \Omega \rightarrow \mathbf{R}$ 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.² This is the so-called *Bayesian viewpoint*.

This second interpretation matches the English etymology: the word *probabilly* 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.

²Future editions may elaborate on the justification for these rules, according to Keynes and Jaynes.

If a first outcome has a larger probability than a second outcome, we call the first *more probable* (or *more likely*) than the second. Similarly, we call the second outcome *less probable* (or *less likely*) 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 define $p(0) = p(1) = 1/2$. We often refer to this particular model as a *fair coin*. Neither heads nor tails is *more* or *less* probable.

Probabilities for rolling a die. Suppose we model rolling a die, as before, with the sample space $\{1, 2, 3, 4, 5, 6\}$. We may model each side of the die as equally likely to face up. Thus we want numbers $p(1), p(2), p(3), p(4), p(5), p(6)$ so that

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6)$$

Consequently, we choose $p(\omega) = 1/6$ for each $\omega \in \Omega$. A Bayesian interpretation is that, prior to the roll, each outcome is *equally credible*. A frequentist interpretation is that, after rolling the die several (hundreds, say) of times, the ratio of times the die lands with six pips facing up to the number of rolls is a rational number close to $1/6$ —and likewise with the other number of pips. We often speak of this particular model as *fair die*. As with the fair coin, 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*.³

Simple consequences on the range of a distribution

Proposition 1. *Suppose $p : \Omega \rightarrow \mathbf{R}$ is a distribution. Then $p(\omega) \leq 1$ for all $\omega \in \Omega$.*

Proof. Let $\omega \in \Omega$. We claim

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

This holds because $p(t) \geq 0$ for every $t \in \Omega$. □

Consequently, the **range** of p is contained in $[0, 1]$. For this reason, we often introduce a distribution on the finite sample space Ω with the notation $p : \Omega \rightarrow [0, 1]$, to remind ourselves that $\text{range}(p) \subset [0, 1]$.

³Many authors reserve the term *probability mass function* for the case in which $\Omega = \mathbf{R}$.

