



Why

To talk about uncertain outcomes, we assign credibility to each outcome according to our intuition of proportion.¹

Definition

A *probability distribution* (or *distribution*, *probability mass function*, *pmf*, *proportion distribution*) is a real-valued function from a finite set of outcomes which is non-negative and normalized. A real-valued function on a finite set is *normalized* if the sum of its results is 1.

The *probability of an outcome* is the result of the outcome under the distribution. The probability of an outcome is meant to indicate the credibility of an outcome. 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.

When an outcome has a higher probability than a second outcome we say that the first is *more probable* than the second. Similarly we say that the first outcome is *less probable* than the second outcome if it has higher probability than the second.

Notation

Let A be a set and let $p : A \rightarrow \mathbf{R}$ be a distribution. Then

$$p(a) \geq 0 \text{ for all } a \in A \text{ and } \sum_{a \in A} p(a) = 1.$$

Simple consequences of definition

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

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

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} .

Example: coin

When flipping a coin, both a heads and a tails are equally worthy of belief. Thus we say that the outcome 0 (tails) has probability $1/2$ and likewise for the outcome 1 (heads). (Recall that we associate the outcome tails to the number 0 and heads to the number 1.) We speak of modeling 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.

