

CS 547 Lecture 6: Axioms of Probability

Daniel Myers

The goal of probability theory is to reason about the outcomes of experiments. Here, *experiment* is an extremely general term that encompasses pretty much any observation we might care to make about the world.

The basic mathematical construction we're interested in is the *probability space*, which has three components:

- the sample space, Ω , which is the set of all possible outcomes of the experiment
- the collection of events, \mathcal{F} , the set of outcomes of the experiment we wish to reason about
- the probability measure, P , which assigns probabilities to the events in \mathcal{F}

An individual event, $E \in \mathcal{F}$, is a subset of the sample space. If $\omega \in \Omega$ is the outcome of an experiment and $\omega \in E$, we say that event E has *occurred*.

It's intuitive to define $P(E)$, the probability of event E , as the fraction of occurrence of E . That is, if we perform our experiment N times, and N_E of those trials result in event E occurring, we could say

$$P(E) = \lim_{N \rightarrow \infty} \frac{N_E}{N}.$$

This definition is intuitively reasonable, but makes several technical assumptions about the existence of the limit and the fact that the ratio converges to a single value for all possible sequences of experimental outcomes.

Rather than make those kinds of assumptions, modern probability theorists prefer to define a more basic set of *axioms* (assumptions) about the nature of the probability measure P . Using these axioms, we can then derive more complex results, including the intuitive definition of probability as a fraction of occurrence.

The three Axioms of Probability are:

1. $0 \leq P(E) \leq 1$
2. $P(\Omega) = 1$
3. If E and F are mutually exclusive events, $P(E \cup F) = P(E) + P(F)$

The first axiom states that the probability of an event is a number between 0 and 1. This is in keeping with our intuitive definition of probability as a fraction of occurrence.

The second axiom states that the event described by the entire sample space has probability of 1. If ω is the outcome of an experimental trial, then $\omega \in \Omega$, by the definition of the sample space, so the event described by Ω must occur on every trial. Intuitively, we can say that every experiment we perform must yield some kind of result, and that result must be in the sample space, so the "event" described by the sample space is so general that it must always occur.

The third axiom describes how to combine the probabilities of mutually exclusive (also called *disjoint*) events. Intuitively, if E and F cannot occur at the same time, then the fraction of time that E or F occurs is the sum of the fractions of time that E and F occur individually.

Random Variables

Very often, we're less interested in the outcome of an experiment than we are in the value of some function calculated from the outcome. For example, we may roll two dice and add their faces, or flip coins and count the number of heads that occur. In games of chance, the "events" may correspond to the fall of the dice, the spin of a wheel, or the play of cards, but we're primarily interested in how much money is won or lost on each play. In computer systems, we may be interested in measuring the latency of a request, which depends in a complex way on the state of the system while the request is being processed.

A *random variable* (RV) is a function that associates a number with the outcome of an experiment. More formally, it is real-valued function defined on the sample space. For the most part, we'll prefer to reason about random variables without worrying about the details of the underlying probability space.

Finally, recall that we wanted to develop the theory of probability to reason about the outcomes of experiments. The Axioms of Probability gave us the tools we needed to do this. Because the value taken by a random variable depends on the outcome of an experiment, *we can also use probability to reason about random variables*. The set of probabilities associated with the values of a random variable is called the *distribution* of the variable. Describing the distribution of a RV totally summarizes its behavior and allows us to draw inferences about it.