Mathematical treatment

See also: Probability axioms

Consider an experiment that can produce a number of results. The collection of all results is called the sample space of the experiment. The power set of the sample space is formed by considering all different collections of possible results. For example, rolling a die can produce six possible results. One collection of possible results gives an odd number on the die. Thus, the subset {1,3,5} is an element of the power set of the sample space of die rolls. These collections are called "events." In this case, {1,3,5} is the event that the die falls on some odd number. If the results that actually occur fall in a given event, the event is said to have occurred.

A probability is a way of assigning every event a value between zero and one, with the requirement that the event made up of all possible results (in our example, the event {1,2,3,4,5,6}) is assigned a value of one. To qualify as a probability, the assignment of values must satisfy the requirement that if you look at a collection of mutually exclusive events (events with no common results, e.g., the events {1,6}, {3}, and {2,4} are all mutually exclusive), the probability that at least one of the events will occur is given by the sum of the probabilities of all the individual events.[17]

The probability of an event *A* is written as P(*A*), p(*A*) or Pr(*A*).[18] This mathematical definition of probability can extend to infinite sample spaces, and even uncountable sample spaces, using the concept of a measure.

The *opposite* or *complement* of an event *A* is the event [not *A*] (that is, the event of *A* not occurring); its probability is given by P(not *A*) = 1 - P(*A*).[19] As an example, the chance of not rolling a six on a six-sided die is 1 – (chance of rolling a six) $= 1 - \tfrac{1}{6} = \tfrac{5}{6}$. See Complementary event for a more complete treatment.

If two events *A* and *B* occur on a single performance of an experiment, this is called the intersection or joint probability of *A* and *B*, denoted as $P(A \cap B)$.

Independent probability

If two events, *A* and *B* are independent then the joint probability is

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B),$$

for example, if two coins are flipped the chance of both being heads is $\tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}$.[20]

Mutually exclusive

If either event *A* or event *B* or both events occur on a single performance of an experiment this is called the union of the events *A* and *B* denoted as $P(A \cup B)$. If two events are mutually exclusive then the probability of either occurring is

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B).$$

For example, the chance of rolling a 1 or 2 on a six-sided die is
$P(1 \text{ or } 2) = P(1) + P(2) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{1}{3}.$

Not mutually exclusive

If the events are not mutually exclusive then

$$\mathrm{P}\left(A \text{ or } B\right) = \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right) - \mathrm{P}\left(A \text{ and } B\right).$$

For example, when drawing a single card at random from a regular deck of cards, the chance of getting a heart or a face card (J,Q,K) (or one that is both) is $\frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{11}{26}$, because of the 52 cards of a deck 13 are hearts, 12 are face cards, and 3 are both: here the possibilities included in the "3 that are both" are included in each of the "13 hearts" and the "12 face cards" but should only be counted once.

Conditional probability

Conditional probability is the probability of some event *A*, given the occurrence of some other event *B*. Conditional probability is written $\mathrm{P}(A \mid B)$, and is read "the probability of *A*, given *B*". It is defined by[21]

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

If $\mathrm{P}(B) = 0$ then $\mathrm{P}(A \mid B)$ is formally undefined by this expression. However, it is possible to define a conditional probability for some zero-probability events using a σ-algebra of such events (such as those arising from a continuous random variable).[*citation needed*]

For example, in a bag of 2 red balls and 2 blue balls (4 balls in total), the probability of taking a red ball is $1/2$; however, when taking a second ball, the probability of it being either a red ball or a blue ball depends on the ball previously taken, such as, if a red ball was taken, the probability of picking a red ball again would be $1/3$ since only 1 red and 2 blue balls would have been remaining.

Inverse probability

In probability theory and applications, **Bayes' rule** relates the odds of event A_1 to event A_2, before (prior to) and after (posterior to) conditioning on another event B. The odds on A_1 to event A_2 is simply the ratio of the probabilities of the two events. When arbitrarily many events A are of interest, not just two, the rule can be rephrased as **posterior is proportional to prior times likelihood**, $P(A|B) \propto P(A)P(B|A)$ where the proportionality symbol means that the left hand side is proportional to (i.e., equals a constant times) the right hand side as A varies, for fixed or given B (Lee, 2012; Bertsch McGrayne, 2012). In this form it goes back to Laplace (1774) and to Cournot (1843); see Fienberg (2005). See Inverse probability and Bayes' rule.

Summary of probabilities

Summary of probabilities

Event	Probability
A	$P(A) \in [0,1]$
not A	$P(A^c) = 1 - P(A)$
A or B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cup B) = P(A) + P(B) \qquad \text{if A and B are mutually exclusive}$
A and B	$P(A \cap B) = P(A\|B)P(B) = P(B\|A)P(A)$ $P(A \cap B) = P(A)P(B) \qquad \text{if A and B are independent}$
A given B	$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B\|A)P(A)}{P(B)}$

Bayes Theorem
Intuition
Suppose a man told you he had had a nice conversation with someone on the train. Not knowing anything about this conversation, the probability that he was speaking to a woman is 50% (assuming the train had an equal number of men and women and the speaker was as likely to strike up a conversation with a man as with a woman). Now suppose he also told you that his conversational partner had long hair. It is now more likely he was speaking to a woman, since women are more likely to have long hair than men. Bayes' theorem can be used to calculate the probability that the person was a woman.

To see how this is done, let W represent the event that the conversation was held with a woman, and L denote the event that the conversation was held with a long-haired person. It can be assumed that women constitute half the population for this example. So, not knowing anything else, the probability that W occurs is $P(W) = 0.5$.

Suppose it is also known that 75% of women have long hair, which we denote as $P(L|W) = 0.75$ (read: the probability of event L given event W is 0.75). Likewise, suppose it is known that 15% of men have long hair, or $P(L|M) = 0.15$, where M is the complementary event of W, i.e., the event that the conversation was held with a man (assuming that every human is either a man or a woman).

Our goal is to calculate the probability that the conversation was held with a woman, given the fact that the person had long hair, or, in our notation, $P(W|L)$. Using the formula for Bayes' theorem, we have:

$$P(W|L) = \frac{P(L|W)P(W)}{P(L)} = \frac{P(L|W)P(W)}{P(L|W)P(W) + P(L|M)P(M)}$$

where we have used the law of total probability. The numeric answer can be obtained by substituting the above values into this formula. This yields

$$P(W|L) = \frac{0.75 \cdot 0.50}{0.75 \cdot 0.50 + 0.15 \cdot 0.50} = \frac{5}{6} \approx 0.83,$$

i.e., the probability that the conversation was held with a woman, given that the person had long hair, is about 83%.