Conditional Probability

Prerequisites

You should be familiar with the definition of the probability and with probability trees.

Definition of probability

As a result of an experiment involving one or more trials a sample space is created comprising possible outcomes. It is assumed that every outcome of sample space is equally likely. An event is a subset of this sample space. Let *S* be a sample space and let the number of possible outcomes in sample space be *n*. Let *A* be an event comprising a set of possible outcomes (sample points) of *S*, and let the number of sample points of *A* be *r*. Then the probability that an event *A* shall occur, denoted $P(A)$, is $P(A) = \frac{r}{n} = \frac{\text{number of outcomes in which the event occurs}}{\text{total number of possible outcomes}}$.

Example (1)

A bag contains 12 sweets. 5 of these sweets are green, 4 are red and 3 are blue. In an experiment 2 sweets are chosen at random without replacement in two successive trials.

(*a*) Draw a probability tree to model this experiment.

(*b*) Find the probability that two sweets of different colours are chosen.

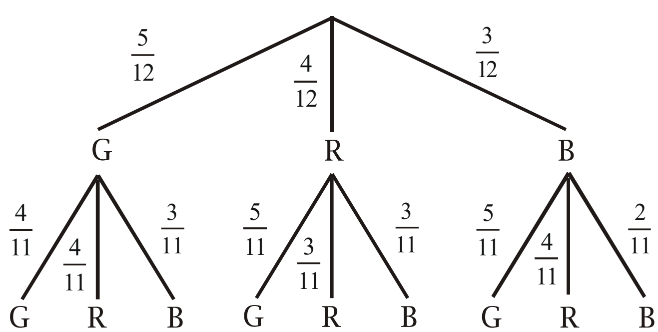
(*c*) What is the probability that a red sweet was chosen in the first trial if we already know that 1 red sweet was chosen at both trials?

(*d*) Describe in general how in this experiment the probability of choosing the second sweet affected by the choice of the first.

(*e*) How would the answers to (*c*) and (*d*) differ if the experiment had been conducted *with replacement*?

Solution

(*a*)



(*b*) $P(\text{Two different colours}) = P(\text{G, other}) + P(\text{R, other}) + P(\text{B, other})$

$$= \left(\frac{5}{12}\times\frac{7}{11}\right) + \left(\frac{4}{12}\times\frac{8}{11}\right) + \left(\frac{3}{12}\times\frac{7}{11}\right)$$

$$= \frac{2}{3}$$

(*c*) It is certain that if 2 red sweets were chosen then at the first trial a red sweet was chosen. In this case the probability that the first sweet chosen was red, given that both sweets were red, is 1.

(*d*) In this experiment, if you choose a red sweet on the first trial then the probability of choosing a red sweet at the second trial decreases. It decreases from $\frac{4}{12}$ to $\frac{3}{11}$. The same applies to choosing a green sweet. Because the sweet is not replaced taking any sweet out of the bag diminishes the chances that the next sweet will be the same colour.

(*e*) If the sweets had been replaced then the answers to the last two questions would be different. For instance, the probability of choosing a red sweet would always be the same, namely $\frac{4}{12}$. The fact that two red sweets were chosen would not tell you that the first sweet must have been red. The probabilities for choosing sweets would then be unaffected by the outcome of any of the trials.

Two trials are independent if the actual outcome of the first trial does not affect the probability of the possible outcomes of the second trial and vice-versa. In example (1) we see that in a probability problem *without replacement* the outcome of the second trial is not independent of the outcome of the first. In a problem with replacement each trial is independent of the other.

Remark

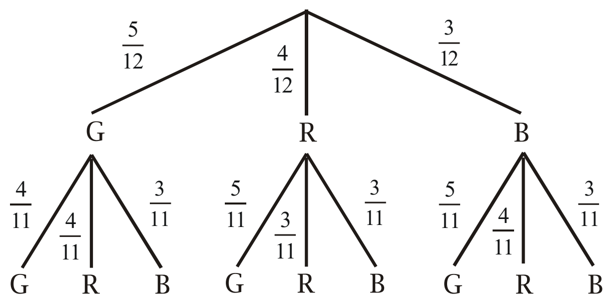
When we say that in the problem with replacement each trial is independent of the other we are assuming that at every trial all outcomes are equally likely. In a real life situation this might not be true. If, for example, when the sweet was replaced into the bag it was put back in such a way that it was very likely that the same sweet would be taken out again at the next trial, then every outcome ceases to be equally likely. So we are assuming in problems with replacement that, for instance, the bag of sweets is thoroughly mixed between trials.

Conditional probability

Example (1) illustrates the concept of *conditional probability*. In it we found that in the case of the two trials without replacement the probability that the first sweet chosen was red, *given that* both sweets was red was 1. We can write this as a conditional statement involving the word **"if"**.
If both sweets are red **then** the probability that the first sweet was red = 1.

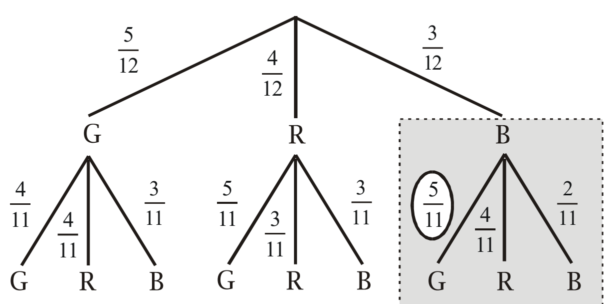
We call this conditional probability. The **if** part expresses the condition and the **then** part expresses the result of this condition. Let A and B be events, then the symbol $P(A|B)$ stands for the conditional probability of A occurring given that B has already taken place. Instinctively, we use our knowledge of conditional probabilities when we construct the tree diagram for a problem in probability.

Example (1) continued
The diagram shows the probability tree for example (1).



Let A denote the event that the second sweet chosen was green.
Let B denote the event that the first sweet chosen was blue.
Find $P(A|B)$.

Solution



We can read this solution directly from the tree diagram. The first trial has already taken place and we need to know only the probabilities of the second trial following the choice of a blue sweet for the first. From the tree this is $P(A|B) = P(\text{green}|\text{blue}) = \frac{5}{11}$.

However, problems in conditional probability can be more complicated than this, and we need another method of finding the conditional probability. The result we are looking for is the following.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A \textbf{ given } B) = \frac{P(A \textbf{ and } B)}{P(B)} \qquad \text{provided } P(B) \neq 0$$

First we will prove this formula, and then we will illustrate its use.

Proof

We have stated and used the formula

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A \textbf{ given } B) = \frac{P(A \textbf{ and } B)}{P(B)} \qquad \text{provided } P(B) \neq 0$$

Now we should justify it.

The definition of probability of an event E is

$$P(E) = \frac{n(E)}{n(S)} = \frac{\text{number of outcomes in which the event } E \text{ occurs}}{\text{total number of possible outcomes}}$$

Let $E = A|B$

$$P(A|B) = \frac{n(A|B)}{n(B)} \qquad P(B) \neq 0 \quad \Rightarrow \quad n(B) \neq 0$$

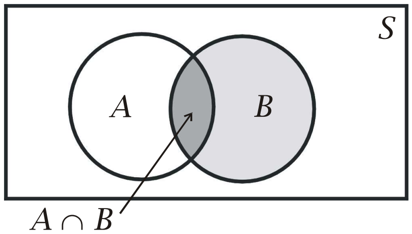
$$= \frac{\text{number of outcomes in which the event } A \text{ and } B \text{ occur}}{\text{total number of possible outcomes}}$$

$$= \frac{n(A \cap B)}{n(B)}$$

$$= \frac{\left(\frac{n(A \cap B)}{n(S)}\right)}{\left(\frac{n(B)}{n(S)}\right)} \qquad \text{Dividing top and bottom by } n(S)$$

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

This proof is illustrated by the following Venn diagram.



The probability $P(A|B)$ is represented in this diagram by the ratio of the number of sample points in $A \cap B$ (shaded darker grey) to the number of sample points in B (shaded light grey).

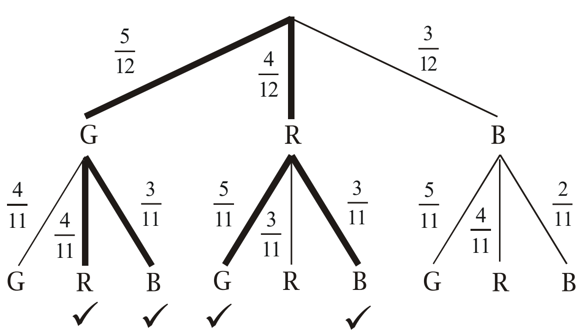
Let us now look at how the formula is used.

Example (1) continued
Let A denote the event that the first sweet chosen was either green or red.
Let B denote the event that both sweets chosen were different colours.
Use the result

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

to find $P(B|A)$.

Solution
Firstly, we must be careful about how we use the result. The result given is for $P(A|B)$ - A given B, but we are asked for $P(B|A)$ - B given A. Therefore, we must use the formula $P(B|A) = \frac{P(B \cap A)}{P(A)}$. From the tree we have $P(A) = \frac{9}{12} = \frac{3}{4}$. The event $B \cap A$ corresponds to the outcomes marked in the probability tree as follows.



From the tree we see that

$$P(B \cap A) = \left(\frac{5}{12} \times \frac{7}{11}\right) + \left(\frac{4}{12} \times \frac{8}{11}\right) = \frac{67}{132}$$

Hence $$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{67}{132}}{\frac{3}{4}} = \frac{67}{99}$$

Example (2)

In a simple model of the weather in October, each day is classified as either fine or rainy. The probability that a fine day is followed by a fine day is 0.75. The probability that a rainy day is followed by a fine day is 0.35. The probability that 10 October is fine is 0.6.

(*a*) Find the probabilities

(*i*) 11 October is rainy.

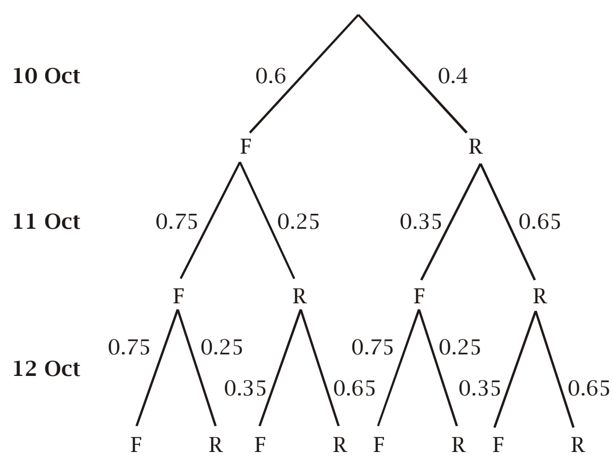
(*ii*) 12 October is fine.

(*b*) Find the conditional probabilities

(*i*) 12 October is fine, given that 10 October is rainy.

(*ii*) 10 October is rainy given that 12 October is fine.

Solution



(*a*) (*i*) $P(\text{October 11 is rainy}) = P(\text{FR}) + P(\text{RR})$

$$= 0.6 \times 0.25 + 0.4 \times 0.65$$
$$= 0.15 + 0.26$$
$$= 0.41$$

(*ii*) $P(\text{October 12 is fine}) = P(\text{FFF}) + P(\text{FRF}) + P(\text{RFF}) + P(\text{RRF})$

$$= (0.6\times 0.75\times 0.75) + (0.6\times 0.25\times 0.35) + (0.4\times 0.35\times 0.75) + (0.4\times 0.65\times 0.35)$$

$$= 0.3375 + 0.0525 + 0.105 + 0.091$$

$$= 0.586$$

(*b*) (*i*) $$P(\text{12 fine}|\text{ 10 rainy}) = \frac{P(\text{12 fine and 10 rainy})}{P(\text{10 rainy})}$$

$$= \frac{P(\text{RFF}) + P(\text{RRF})}{P(\text{10 rainy})}$$

$$= \frac{(0.4\times 0.35\times 0.75) + (0.4\times 0.65\times 0.35)}{0.4} = 0.49$$

(*ii*) $$P(\text{10 rainy}|\text{ 12 fine}) = \frac{P(\text{10 rainy and 12 fine})}{P(\text{12 fine})} = \frac{0.196}{0.586} = 0.334 \quad \text{(3 s.f.)}$$

Bayes' theorem and the law of total probability

The relationship between the conditional probability of A given B and the conditional probability of B given A is

$$P(A|B)P(B) = P(B|A)P(A)$$

This is a simplified version of a result known as *Bayes' Theorem*, which we state subsequently. Firstly, we will prove the simplified result.

Proof of the simplified formula

The definition of conditional probability implies both

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \qquad (1)$$

and

$$P(B|A) = \frac{P(B\cap A)}{P(A)} \qquad (2)$$

From (1) $P(A\cap B) = P(A|B)P(B)$

and likewise from (2) $P(A\cap B) = P(B|A)P(A)$

On equating these

$$P(A|B)P(B) = P(B|A)P(A)$$

Example (3)

A collection of short stories is made of the work of three authors, David, Harry and Marvin. 45% of the stories are written by David, 25% are by Harry and the rest are by Marvin. The stories are either about science fiction or horror. The probability that a story by David will be science fiction is 0.5, and the respective probabilities for Harry and Marvin are 0.9 and 0.4. What is the probability that a horror story chosen at random will be written by David?

Solution

Let the following events be

$D =$ the story is by David

$H =$ the story is by Herbert

$M =$ the story is by Marvin

$X =$ the story is a horror story

$\bar{X} =$ the story is a science fiction story

We are given

$P(D) = 0.45$

$P(H) = 0.25$

$P(M) = 1 - 0.45 - 0.25 = 0.30$

$P(\bar{X}|D) = 0.5$

$P(\bar{X}|H) = 0.9$

$P(\bar{X}|M) = 0.4$

We need to find $P(D|X)$

We will use the formula

$P(D|X)P(X) = P(X|D)P(D)$

to "reverse the conditions". Specifically we will substitute into

$$P(D|X) = \frac{P(X|D) \times P(D)}{P(X)}.$$

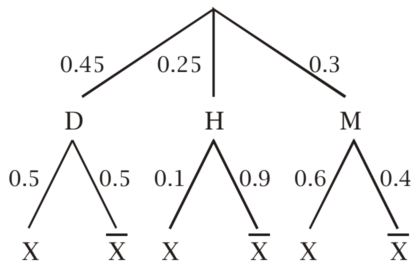
We already have

$P(X|D) = 1 - P(\bar{X}|D) = 1 - 0.5 = 0.5$ (1)

$P(D) = 0.45$ (2)

So require $P(X)$

$P(X)$ is the total probability that the story will be a horror story. To find this we construct a probability tree.



Therefore

$$P(X) = (0.45 \times 0.5) + (0.25 \times 0.1) + (0.3 \times 0.6) = 0.43$$

Hence, on substituting we obtain

$$\begin{aligned} P(D|X) &= \frac{P(X|D) \times P(D)}{P(X)} \\ &= \frac{0.5 \times 0.45}{0.43} \\ &= 0.523 \ \text{(3 d.p.)} \end{aligned}$$

Bayes' Theorem is a generalised version of the result

$$P(A|B)P(B) = P(B|A)P(A)$$

In Example (3) we rearranged this formula at one point to obtain

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The probability $P(B)$ is the total probability that B will occur. Suppose B is an event that depends on certain other mutually exclusive and exhaustive events, $A_1, A_2, \ldots, A_n$, so that we have a series of conditional probabilities as follows.

$$P(B|A_1),\ P(B|A_2),\ \ldots, P(B|A_n)$$

Then the probability that B will occur is just the sum of these conditional probabilities

$$P(B) = P(B|A_1) + P(B|A_2) + \ldots + P(B|A_n) = \sum_{i=1}^{n} P(B|A_i)$$

This statement is known as the *law of total probability*. Substituting this formula into (1) above, we get

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

This is known as *Bayes' Theorem*. It looks more complicated than it really is, for the denominator just is the total probability that B will occur, and given this observation the formula may be simply stated as

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

which is the form that was used in the solution to Example (3). The introduction of the notation for a sum $\sum_{i=1}^{n}....$ and the subscripts may cause some trouble, so let us substitute $i=2$ into this formula and see what it says. So, with $i=2$, we get

$$P(A_j|B)=\frac{P(B|A_j)P(A_j)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)}$$

The denominator of the expression $P(A_j|B)=\frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n}P(B|A_i)P(A_i)}$ is

$$\sum_{i=1}^{n}P(B|A_i)P(A_i)=P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+...+P(B|A_n)P(A_n)$$

Applications of this result are also known as *reversing the conditions*. The following example, which is very similar to Example (3), also illustrates the case when $i=3$.

Example (4)

Three men, Ambrose, Benedict and Cuthbert, sing in a male choir. The probability that Ambrose will be asked by the musical director to sing the solo is 0.45; and the corresponding probabilities for Benedict and Cuthbert are 0.30 and 0.25 respectively. The probability that in a solo Ambrose will sing off key (that is, not in tune) is 0.25; the probability that Benedict will sing off key is 0.35 and the probability that Cuthbert will sing off key is 0.50. At a given performance the solo was sung off key. What is the probability that the solo singer was Ambrose?

Solution

Let A be the event that Ambrose sings the solo.
Let B be the event that Benedict sings the solo.
Let C be the event that Cuthbert sings the solo.
Let X be the event that the solo is sung off key.
Let $\bar{X}$ be the event that the solo is not sung off key (in other words, in tune).

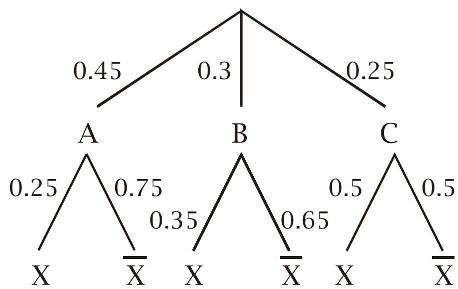
We are given

$P(A) = 0.45$ $\quad$ $P(X|A) = 0.25$

$P(B) = 0.30$ $\quad$ $P(X|B) = 0.35$

$P(C) = 0.25$ $\quad$ $P(X|A) = 0.50$

The tree diagram is



We are asked to find $P(A|X)$. Using Bayes' theorem to reverse the conditions this is

$$P(A|X) = \frac{P(X|A) \times P(A)}{P(X)}$$

The probability of X is given by the law of total probability as

$$P(X) = P(X|A) + P(X|B) + P(X|C)$$

$$= (0.25 \times 0.45) + (0.35 \times 0.3) + (0.5 \times 0.25)$$

$$= 0.3425$$

Hence

$$P(A|X) = \frac{P(X|A) \times P(A)}{P(X)} = \frac{0.25 \times 0.45}{0.3425} = 0.328 \text{ (3 s.f.)}$$