- Probability can be expressed mathematically as a numerical index with a range between zero (an absolute impossibility) to unity (an absolute certainty).
- Most events have a probability index strictly between 0 and 1, which means that each event has *at least* two possible outcomes: favourable outcome or success, and unfavourable outcome or failure.

$$P(success) = \frac{the\ number\ of\ successes}{the\ number\ of\ possible\ outcomes}$$

$$P(failure) = \frac{the\ number\ of\ failures}{the\ number\ of\ possible\ outcomes}$$

- If s is the number of times success can occur, and f is the number of times failure can occur, then

$$P(success) = p = \frac{s}{s+f}$$

$$P(failure) = q = \frac{f}{s+f}$$

and $$p + q = 1$$

- If we throw a coin, the probability of getting a head will be equal to the probability of getting a tail. In a single throw, $s = f = 1$, and therefore the probability of getting a head (or a tail) is 0.5.

Conditional probability

- Let *A* be an event in the world and *B* be another event. Suppose that events *A* and *B* are not mutually exclusive, but occur conditionally on the occurrence of the other. The probability that event *A* will occur if event *B* occurs is called the **conditional probability**. Conditional probability is denoted mathematically as $p(A|B)$ in which the vertical bar represents *GIVEN* and the complete probability expression is interpreted as *"Conditional probability of event A occurring given that event B has occurred"*.

$$p\left(A\middle|B\right) = \frac{the\ number\ of\ times\ A\ and\ B\ can\ occur}{the\ number\ of\ times\ B\ can\ occur}$$

- The number of times *A* and *B* can occur, or the probability that both *A* and *B* will occur, is called the **joint probability** of *A* and *B*. It is represented mathematically as $p(A \cap B)$. The number of ways *B* can occur is the probability of *B*, $p(B)$, and thus

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

- Similarly, the conditional probability of event *B* occurring given that event *A* has occurred equals

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

Hence, $p(B \cap A) = p(B|A) \times p(A)$

and $p(A \cap B) = p(B|A) \times p(A)$

Substituting the last equation into the equation

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

yields the **Bayesian rule**:

Bayesian rule

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

where:

$p(A|B)$ is the conditional probability that event A occurs given that event B has occurred;

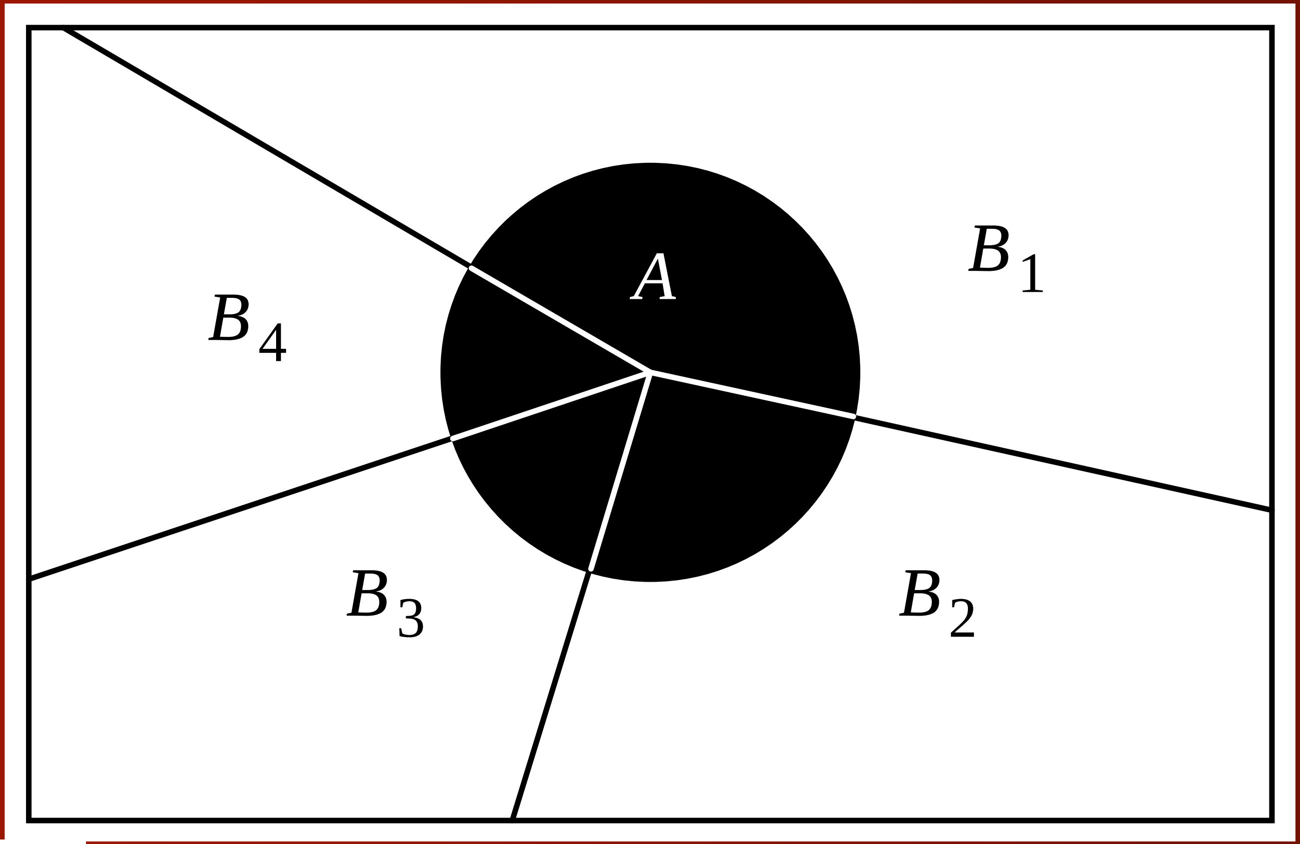
$p(B|A)$ is the conditional probability of event B occurring given that event A has occurred;

$p(A)$ is the probability of event A occurring;

$p(B)$ is the probability of event B occurring.

The joint probability

$$\sum_{i=1}^{n} p(A \cap B_i) = \sum_{i=1}^{n} p(A|B_i) \times p(B_i)$$



If the occurrence of event *A* depends on only two mutually exclusive events, *B* and NOT *B*, we obtain:

$$p(A) = p(A|B) \times p(B) + p(A|\neg B) \times p(\neg B)$$

where $\neg$ is the logical function NOT.

Similarly,

$$p(B) = p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)$$

Substituting this equation into the Bayesian rule yields:

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}$$

Bayesian reasoning

Suppose all rules in the knowledge base are represented in the following form:

IF	*E* is true
THEN	*H* is true {with probability *p*}

This rule implies that if event *E* occurs, then the probability that event *H* will occur is *p*.

In expert systems, *H* usually represents a hypothesis and *E* denotes evidence to support this hypothesis.

The Bayesian rule expressed in terms of hypotheses and evidence looks like this:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}$$

where:

$p(H)$ is the prior probability of hypothesis H being true;

$p(E|H)$ is the probability that hypothesis H being true will result in evidence E;

$p(\neg H)$ is the prior probability of hypothesis H being false;

$p(E|\neg H)$ is the probability of finding evidence E even when hypothesis H is false.

- In expert systems, the probabilities required to solve a problem are provided by experts. An expert determines the **prior probabilities** for possible hypotheses $p(H)$ and $p(\neg H)$, and also the **conditional probabilities** for observing evidence E if hypothesis H is true, $p(E|H)$, and if hypothesis H is false, $p(E|\neg H)$.
- Users provide information about the evidence observed and the expert system computes $p(H|E)$ for hypothesis H in light of the user-supplied evidence E. Probability $p(H|E)$ is called the **posterior probability** of hypothesis H upon observing evidence E.

- We can take into account both multiple hypotheses $H_1, H_2, \ldots, H_m$ and multiple evidences $E_1, E_2, \ldots, E_n$. The hypotheses as well as the evidences must be mutually exclusive and exhaustive.
- Single evidence E and multiple hypotheses follow:

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E|H_k) \times p(H_k)}$$

- Multiple evidences and multiple hypotheses follow:

$$p(H_i|E_1\, E_2 \ldots E_n) = \frac{p(E_1\, E_2 \ldots E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1\, E_2 \ldots E_n|H_k) \times p(H_k)}$$

- This requires to obtain the conditional probabilities of all possible combinations of evidences for all hypotheses, and thus places an enormous burden on the expert.
- Therefore, in expert systems, conditional independence among different evidences assumed. Thus, instead of the unworkable *equation*, we attain:

$$p\left(H_i \middle| E_1\, E_2 \ldots E_n\right) = \frac{p\left(E_1 \middle| H_i\right) \times p\left(E_2 \middle| H_i\right) \times \ldots \times p\left(E_n \middle| H_i\right) \times p\left(H_i\right)}{\sum_{k=1}^{m} p\left(E_1 \middle| H_k\right) \times p\left(E_2 \middle| H_k\right) \times \ldots \times p\left(E_n \middle| H_k\right) \times p\left(H_k\right)}$$

Ranking potentially true hypotheses

Let us consider a simple example.

Suppose an expert, given three conditionally independent evidences E_1, E_2 and E_3, creates three mutually exclusive and exhaustive hypotheses H_1, H_2 and H_3, and provides prior probabilities for these hypotheses – $p(H_1)$, $p(H_2)$ and $p(H_3)$, respectively. The expert also determines the conditional probabilities of observing each evidence for all possible hypotheses.

The prior and conditional probabilities

Probability	*Hypothesis*		
	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1 \mid H_i)$	0.3	0.8	0.5
$p(E_2 \mid H_i)$	0.9	0.0	0.7
$p(E_3 \mid H_i)$	0.6	0.7	0.9

Assume that we first observe evidence E_3. The expert system computes the posterior probabilities for all hypotheses as

$$p\left(H_i \middle| E_3\right) = \frac{p\left(E_3 \middle| H_i\right) \times p\left(H_i\right)}{\sum_{k=1}^{3} p\left(E_3 \middle| H_k\right) \times p\left(H_k\right)}, \qquad i = 1, 2, 3$$

Thus,

$$p\left(H_1 \middle| E_3\right) = \frac{0.6 \cdot 0.40}{0.6 \cdot 0.40 + 0.7 \cdot 0.35 + 0.9 \cdot 0.25} = 0.34$$

$$p\left(H_2 \middle| E_3\right) = \frac{0.7 \cdot 0.35}{0.6 \cdot 0.40 + 0.7 \cdot 0.35 + 0.9 \cdot 0.25} = 0.34$$

$$p\left(H_3 \middle| E_3\right) = \frac{0.9 \cdot 0.25}{0.6 \cdot 0.40 + 0.7 \cdot 0.35 + 0.9 \cdot 0.25} = 0.32$$

After evidence E_3 is observed, belief in hypothesis H_2 increases and becomes equal to belief in hypothesis H_1. Belief in hypothesis H_3 also increases and even nearly reaches beliefs in hypotheses H_1 and H_2.

Suppose now that we observe evidence E_1. The posterior probabilities are calculated as

$$p\left(H_i \middle| E_1 E_3\right) = \frac{p\left(E_1 \middle| H_i\right) \times p\left(E_3 \middle| H_i\right) \times p\left(H_i\right)}{\sum_{k=1}^{3} p\left(E_1 \middle| H_k\right) \times p\left(E_3 \middle| H_k\right) \times p\left(H_k\right)}, \qquad i = 1, 2, 3$$

Hence,

$$p\left(H_1 \middle| E_1 E_3\right) = \frac{0.3 \cdot 0.6 \cdot 0.40}{0.3 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.9 \cdot 0.25} = 0.19$$

$$p\left(H_2 \middle| E_1 E_3\right) = \frac{0.8 \cdot 0.7 \cdot 0.35}{0.3 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.9 \cdot 0.25} = 0.52$$

$$p\left(H_3 \middle| E_1 E_3\right) = \frac{0.5 \cdot 0.9 \cdot 0.25}{0.3 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.9 \cdot 0.25} = 0.29$$

Hypothesis H_2 has now become the most likely one.

After observing evidence E_2, the final posterior probabilities for all hypotheses are calculated:

$$p(H_i|E_1E_2E_3) = \frac{p(E_1|H_i)\times p(E_2|H_i)\times p(E_3|H_i)\times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k)\times p(E_2|H_k)\times p(E_3|H_k)\times p(H_k)}, \qquad i = 1, 2, 3$$

$$p(H_1|E_1E_2E_3) = \frac{0.3\cdot 0.9\cdot 0.6\cdot 0.40}{0.3\cdot 0.9\cdot 0.6\cdot 0.40 + 0.8\cdot 0.0\cdot 0.7\cdot 0.35 + 0.5\cdot 0.7\cdot 0.9\cdot 0.25} = 0.45$$

$$p(H_2|E_1E_2E_3) = \frac{0.8\cdot 0.0\cdot 0.7\cdot 0.35}{0.3\cdot 0.9\cdot 0.6\cdot 0.40 + 0.8\cdot 0.0\cdot 0.7\cdot 0.35 + 0.5\cdot 0.7\cdot 0.9\cdot 0.25} = 0$$

$$p(H_3|E_1E_2E_3) = \frac{0.5\cdot 0.7\cdot 0.9\cdot 0.25}{0.3\cdot 0.9\cdot 0.6\cdot 0.40 + 0.8\cdot 0.0\cdot 0.7\cdot 0.35 + 0.5\cdot 0.7\cdot 0.9\cdot 0.25} = 0.55$$

Although the initial ranking was H_1, H_2 and H_3, only hypotheses H_1 and H_3 remain under consideration after all evidences (E_1, E_2 and E_3) were observed.

Bias of the Bayesian method

- The framework for Bayesian reasoning requires probability values as primary inputs. The assessment of these values usually involves human judgement. However, psychological research shows that humans cannot elicit probability values consistent with the Bayesian rules.
- This suggests that the conditional probabilities may be inconsistent with the prior probabilities given by the expert.

- The Bayesian method is likely to be the most appropriate if reliable statistical data exists, the knowledge engineer is able to lead, and the expert is available for serious decision-analytical conversations.
- In the absence of any of the specified conditions, the Bayesian approach might be too arbitrary and even biased to produce meaningful results.
- The Bayesian belief propagation is of exponential complexity, and thus is impractical for large knowledge bases.