Gør tanke til handling

VIA University College

Probability theory 2

NOTE: This slideshow is a continuation of "Probability theory 1"

Conditional probability

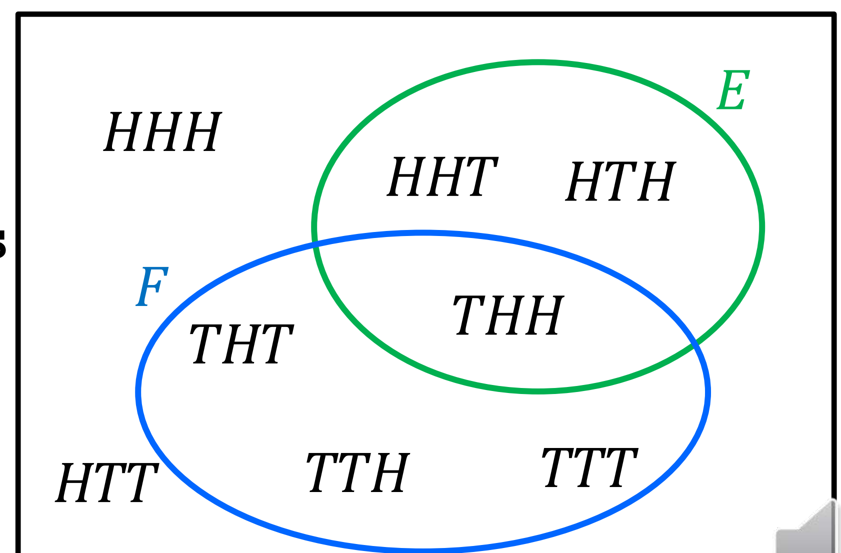
The probability of an event A happening *given* that another event, B, happens is written as $P(A|B)$ – this is read "the probability of A given B". This can be calculated as

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

<u>Example:</u>

Consider the events E and F from the last slide. The probability of getting tails exactly once (E) given that the first flip is tails (F) is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$
$$= \frac{1/8}{4/8} = \frac{1}{4}$$



Definition: dependent and independent events

Two events are said to be **independent** if the probability for one of the events is the same whether or not the other event happens.

This can be expressed in terms of conditional probabilities: The events A and B are independent if and only if

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

or, equivalently, if $P(B|A) = P(B)$.

Conversely, if $P(A|B) \neq P(A)$, the events A and B are said to be **dependent.**

Example of dependent events

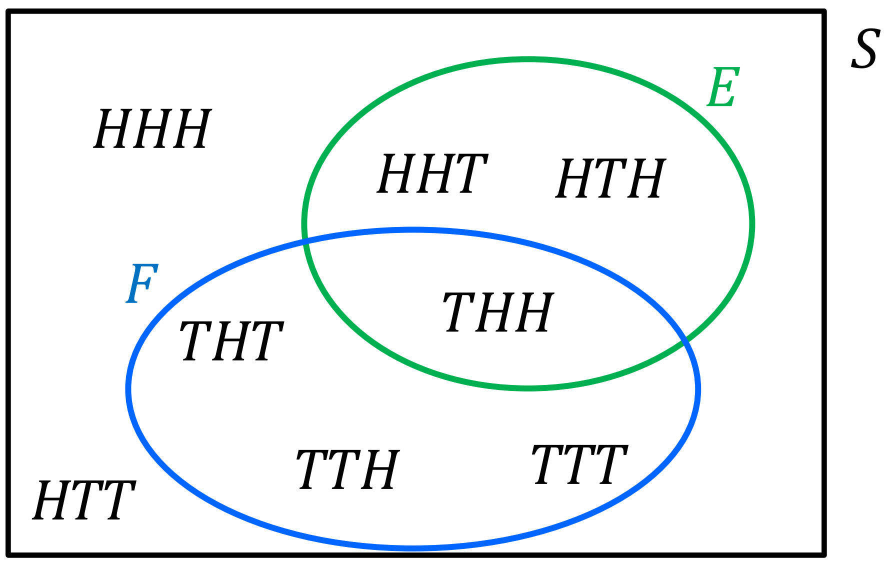
Consider the events

E: getting tails exactly once

F: first flip is tails

We calculate $P(E) = \frac{3}{8}$ and $P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/8}{4/8} = \frac{1}{4}$.

Since $P(E) \neq P(E|F)$, the probability for E depends on whether or not F happens. So E and F are dependent events.



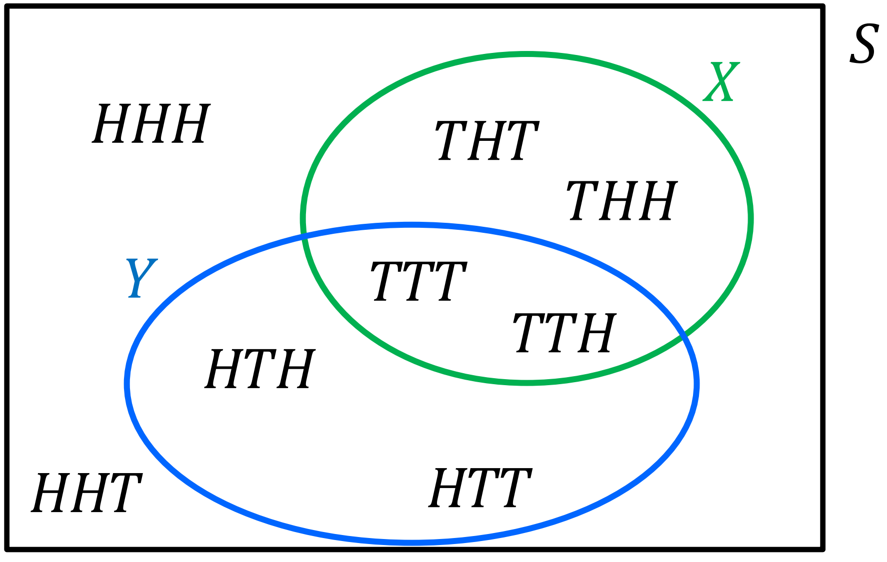
Example of independent events

Consider the events

$$X\text{: first flip is tails}$$
$$Y\text{: second flip is tails}$$

We calculate $P(X) = \frac{4}{8} = \frac{1}{2}$ and $P(X|Y) = \frac{P(X\cap Y)}{P(Y)} = \frac{2/8}{4/8} = \frac{1}{2}$.

Since $P(X) = P(X|Y)$, the probability for X does not depend on whether or not Y happens. So X and Y are independent events.



Probability of intersection

It follows from the rule for conditional probability that the probability for the intersection between A and B can be calculated as

$$P(A \cap B) = P(A|B) \cdot P(B).$$

If A and B are independent events, $P(A|B) = P(A)$, so **for independent events**, the probability for A and B is given by

$$P(A \cap B) = P(A) \cdot P(B).$$

Probability of intersection - example

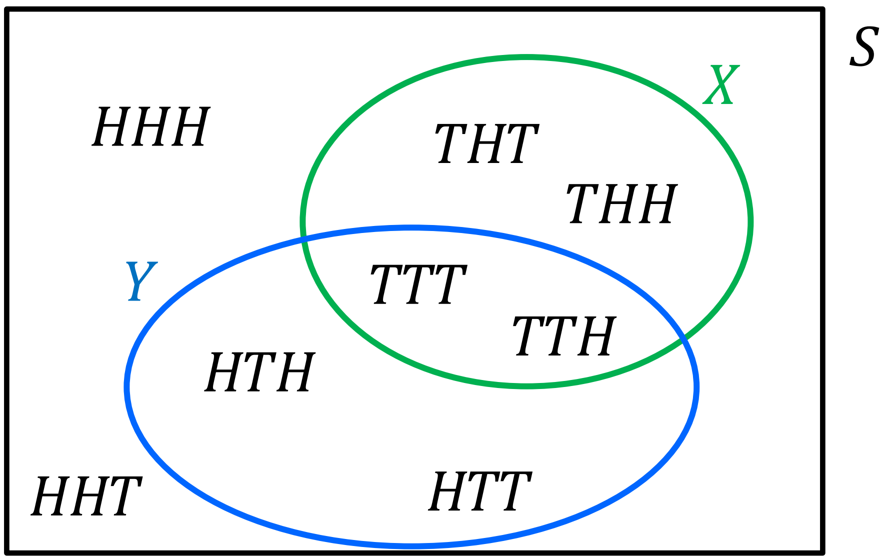
Consider the events

$$X\text{: first flip is tails}$$
$$Y\text{: second flip is tails}$$

As we saw previously, $P(X|Y) = P(X)$, so X and Y are independent.

Therefore, the probability for the first flip *and* the second flip being tails is

$$P(X \cap Y) = P(X) \cdot P(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$



Law of total probability

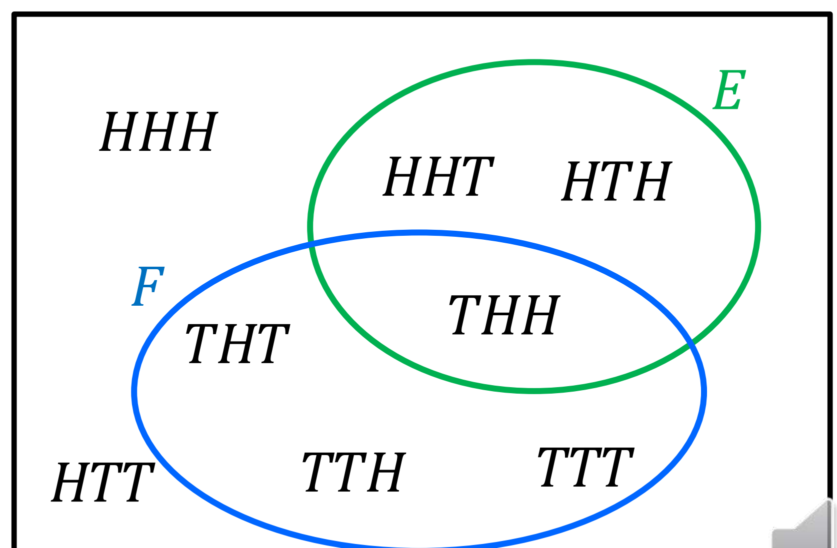
Assume that you know what the probability of an event A is given that B has happened, and you also know the probability for A given that B has *not* happened. From this, you can compute the total probability for A using **the law of total probability:**

$$P(A) = P(A|B) \cdot P(B) + P(A|B') \cdot P(B')$$

<u>Example:</u>

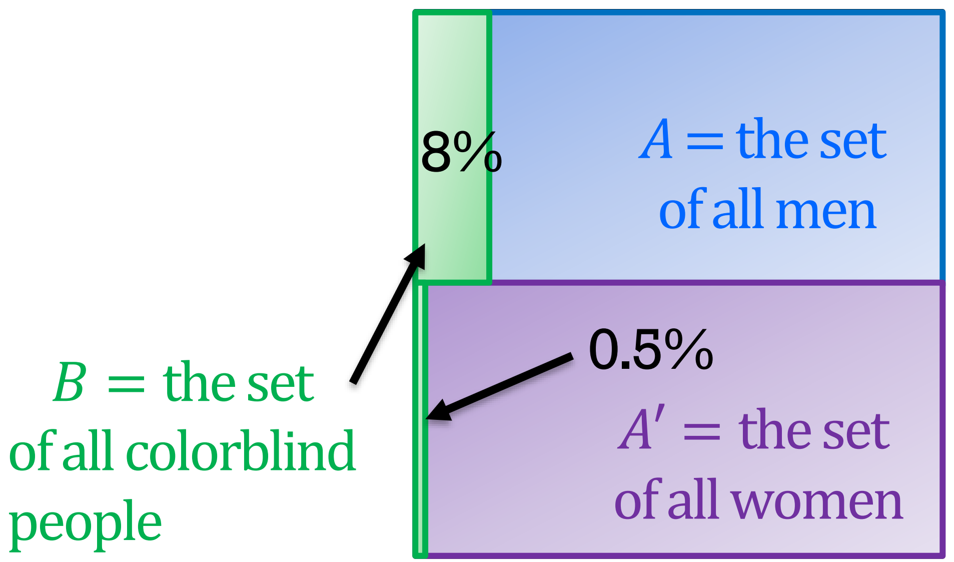
In the coin flipping example, we can see that $P(E|F) = \frac{1}{4}$, $P(E|F') = \frac{2}{4}$, $P(F) = \frac{4}{8}$ and $P(F') = \frac{4}{8}$. From this, we can compute

$$P(E) = P(E|F)P(F) + P(E|F')P(F')$$
$$= \frac{1}{4} \cdot \frac{4}{8} + \frac{2}{4} \cdot \frac{4}{8} = \frac{3}{8}.$$



Law of total probability – example

As illustrated in the Venn diagram below, the probability of being colorblind is 8% for men and 0.5% for women. Assuming that there are as many men as woman, what is the probability that a random person is colorblind ("CB")?



$$P(\text{CB}) = P(\text{CB}|\,\text{man}) \cdot P(\text{man}) + P(\text{CB}|\,\text{not man}) \cdot P(\text{not man})$$
$$= 0.08 \cdot 0.5 + 0.005 \cdot 0.5 = 0.0435 = 4.35\%$$

Bayes' theorem

From the rules for conditional probability, we can deduce a very useful rule:

$$P(A \cap B) = P(B|A) \cdot P(A)$$
$$P(A \cap B) = P(A|B) \cdot P(B)$$
$$\Downarrow$$
$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$
$$\Downarrow$$

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

This is called **Bayes' theorem,** and is probably the most famous formula in all of probablity theory!

Example: Using Bayes' theorem

"A woman has a mammography test (X-ray examination) to look for breast cancer. The test has an accuracy of 90% (that is, a healthy individual well get a positive result 10% of the time, and a sick individual will get a negative result 10% of the time.) Furthermore, we know that 0.5% of all women have breast cancer. If the result of the test is positive, what is the probability that the woman has breast cancer?"

From the text above, we can deduce the following probabilities (where "pos" denotes the event that the test is positive, "neg" denotes the event that the test is negative, and "cancer"/"not cancer" denotes whether she has cancer or not):

$$P(\text{pos}|\text{cancer}) = 0.9, P(\text{neg}|\text{not cancer}) = 0.9, P(\text{neg}|\text{cancer}) = 0.1, \; P(\text{pos}|\text{not cancer}) = 0.1.$$

We can then use Bayes' theorem and the law of total probability to calculate the probability that she has cancer, given that the test was positive:

Law of total probability

$$P(\text{cancer}|\text{pos}) = \frac{P(\text{pos}|\text{cancer}) \cdot P(\text{cancer})}{P(\text{pos})} = \frac{P(\text{pos}|\text{cancer}) \cdot P(\text{cancer})}{P(\text{pos}|\text{cancer}) \cdot P(\text{cancer}) + P(\text{pos}|\text{not cancer}) \cdot P(\text{not cancer})}$$

Bayes' theorem

$$= \frac{0.9 \cdot 0.005}{0.9 \cdot 0.005 + 0.1 \cdot 0.995} = 0.043 = 4.3\%$$

Contingency table

A much used tool in statistics and probability theory is a **contingency table** (*contingency* means "a possible event or circumstance").

A contingency table shows how a set of items are distributed over two variables – in the example below, the "items" are *people*, and the "variables" are *gender* (or *sex*) and *handedness.*

In the next slides, we will see some examples of how such tables can be used to answer various questions in probability theory.

Handed-ness / Sex	Right handed	Left handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	100

Example of a contingency table from Wikipedia

Using a contingency table - examples

Let A denote the event that a person is male, and let B denote the event that a person is left handed. Use the contingency table to determine $P(A \cup B), P(A \cap B)$ and $P(B|A)$ for the people in the table.

Handed-ness / Sex	Right handed	Left handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	100

- The event $A \cup B$ corresponds to people who are men *or* who are left handed. This group consists of the cell marked with yellow in the table. In total, $52 + 4 = 56$ people belongs to the event $A \cup B$ out of 100, so $P(A \cup B) = \frac{56}{100} = 0.56$.
- The event $A \cap B$ corresponds to people who are men *and* who are left handed - marked with blue in the table. 9 people belongs to this event out of 100, so $P(A \cap B) = \frac{9}{100} = 0.09$.
- To determine $P(B|A)$ – the probability that a person is left handed given that he is a man, we focus on the row marked with green and see that there are 52 men of which 9 are left handed. This tells us that that $P(B|A) = \frac{9}{52} = 0.173$.

Joint probability table

A contingency table can easily be turned in to a table containing probabilities (called a **joint probability table**), by dividing the numbers in each cell by the total number of items included in the table. This is illustrated below, using a (made-up) table of 2000 women who were tested for breast cancer.

Test result	Has cancer: Yes	No	Total
Pos.	9	199	208
Neg.	1	1791	1792
Total	10	1990	2000

Test result	Has cancer: Yes	No	Total
Pos.	0.45%	9.95%	10.4%
Neg.	0.05%	89.55%	89.6%
Total	0.5%	99.5%	100%

$$\frac{1791}{2000} = 89.55\%$$

Using a joint probability table - example

Earlier we used Bayes' theorem and the law of total probability to answer the question "what is the probability that a woman has breast cancer, given that she had a positive mammography test?"

Test result	Has cancer	Yes	No	Total
	Pos.	0.45%	9.95%	10.4%
	Neg.	0.05%	89.55%	89.6%
	Total	0.5%	99.5%	100%

We could also answer this question by using a joint probability table:

Use the law of conditional probability, $P(A|B) = \frac{P(A \cap B)}{P(B)}$, with A denoting the event that the woman has cancer and B denoting the event that the test is positive. From the table you can see that $P(\text{cancer} \cap \text{pos}) = 0.45\%$ and $P(\text{pos}) = 10.4\%$, so

$$P(\text{cancer}\,|\text{pos}) = \frac{P(\text{cancer } \cap \text{ pos})}{P(\text{pos})} = \frac{0.45\%}{10.4\%} = 4.3\%$$