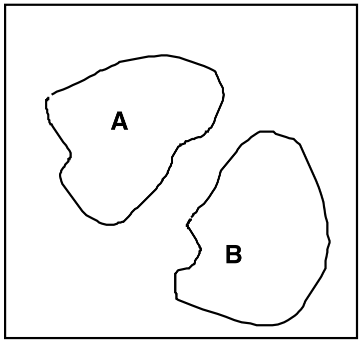
Probability

Chapter 5

The *probability* of an event is its true relative frequency, the proportion of times the event would occur if we repeated the same process over and over again.

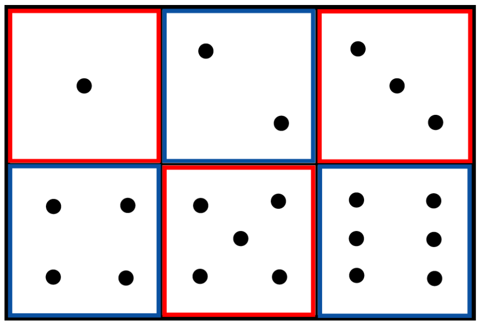
Two events are *mutually exclusive* if they cannot both be true.

A and B are mutually exclusive



"Venn diagram"

Mutually exclusive



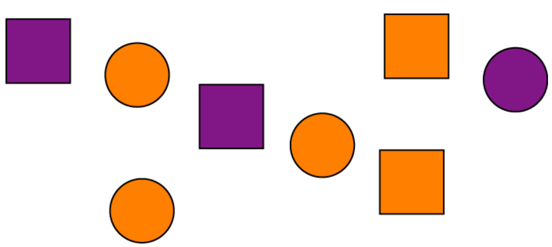
Mutually exclusive

Pr(*A* and *B*) = 0

Not mutually exclusive

Pr(*A* and *B*) ≠ 0

Pr(*purple* AND *square*) ≠ 0



For example

Event A: First child is female
Event B: Second child is female

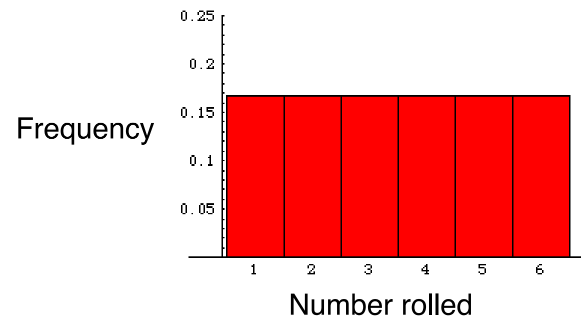
P(A) = 0.48
P(B) = 0.48

But P(*A* and *B*) ≠ 0, so these events are NOT mutually exclusive.

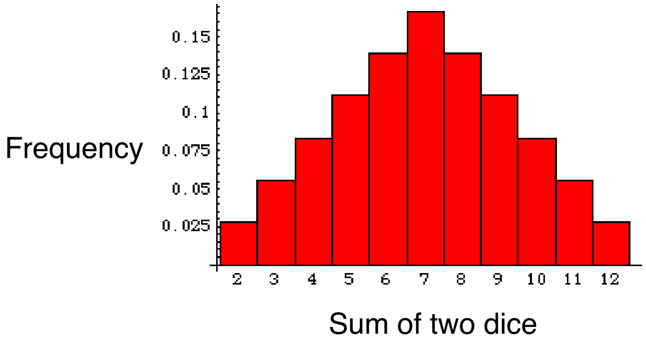
Probability distribution

A *probability distribution* describes the true relative frequency of all possible values of a random variable.

Probability distribution for the outcome of a roll of a die



Probability distribution for the sum of a roll of two dice



The addition principle

The *addition principle*: If two events *A* and *B* are mutually exclusive, then

Pr[*A* OR *B*] = Pr[*A*] + Pr[*B*]

The probability of a range

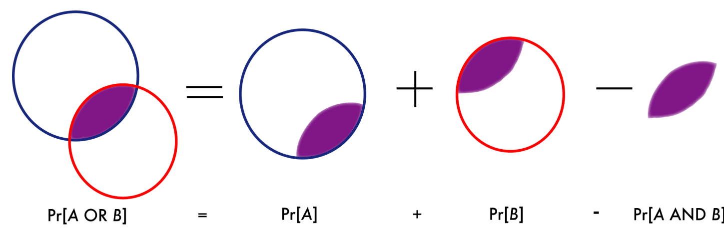
Pr[*sum of two dice* ≥ 8] = Pr[8] + Pr[9] + Pr[10] + . . .

The probabilities of all possibilities add to 1.

Probability of *Not*

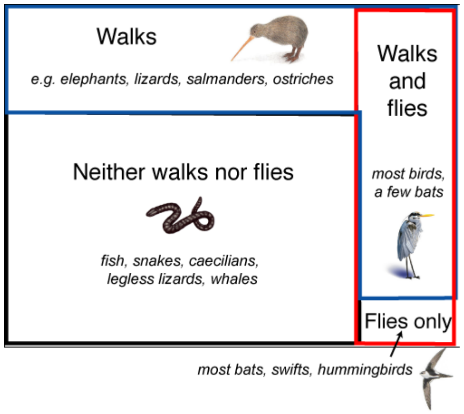
Pr[NOT *rolling a 2*] = 1 – Pr[*Rolling a 2*] = 5/6

General Addition Principle



General addition principle

Pr[*A* OR *B*] = Pr[*A*] + Pr[*B*] - Pr[*A* AND *B*].



Independence

Two events are *independent* if the occurrence of one gives no information about whether the second will occur.

Multiplication principle

The *multiplication principle*: If two events *A* and *B* are independent, then

Pr[*A* AND *B*] = Pr[*A*] × Pr[*B*]

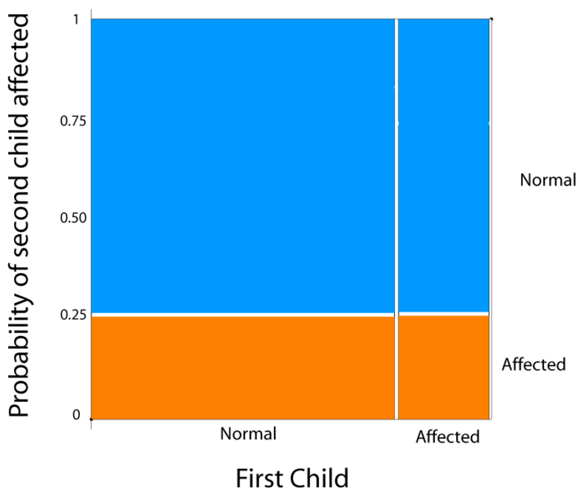
Offspring of two "carriers":

Pr[congenital nightblindness]=0.25

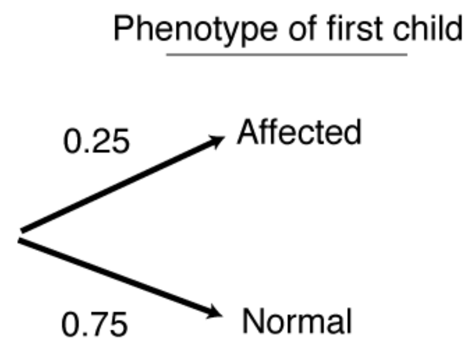
What is the probability that two kids from this family both have nightblindness?

Pr[(*first child has nightblindness*) AND

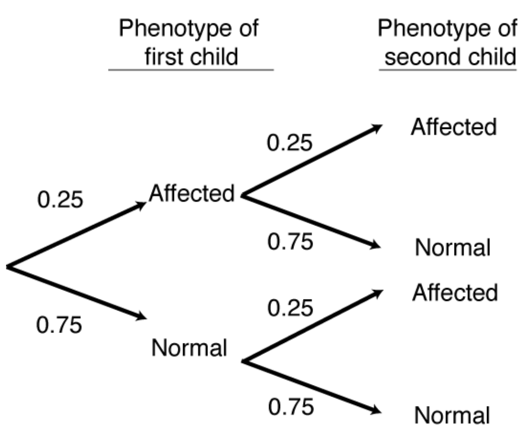
(*second child has nightblindness*)] = 0.25 × 0.25 = 0.0625.



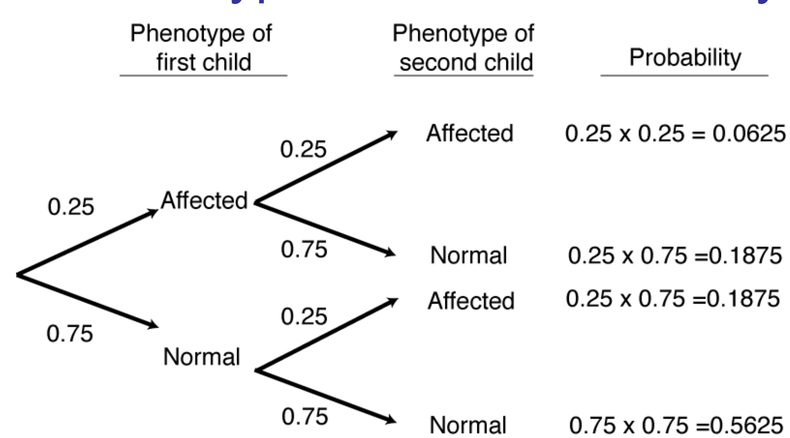
Probability trees



Phenotypes in two-child family



Phenotypes in two-child family



Short summary

The probability of A *OR* B involves addition.

Pr(*A* or *B*) = Pr(*A*) + Pr(*B*) if the two are mutually exclusive.

The probability of A *AND* B involves multiplication

Pr(*A* and *B*) = Pr(*A*) Pr(*B*) if the two are independent

Dependent events

Variables are not always independent.

The probability of one event may depend on the outcome of another event

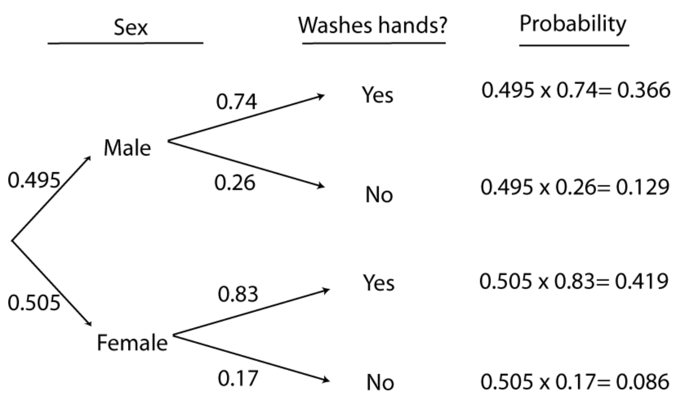
Washing hands

Hand washing after using the restroom

- Pr[male] = 0.495
- Pr[male washes his hands] = 0.74
- Pr[female washes her hands] = 0.83

Data from a press release from American Society of Microbiology, 15 Sept. 2003.

Hand washing



Are sex and hand washing independent?

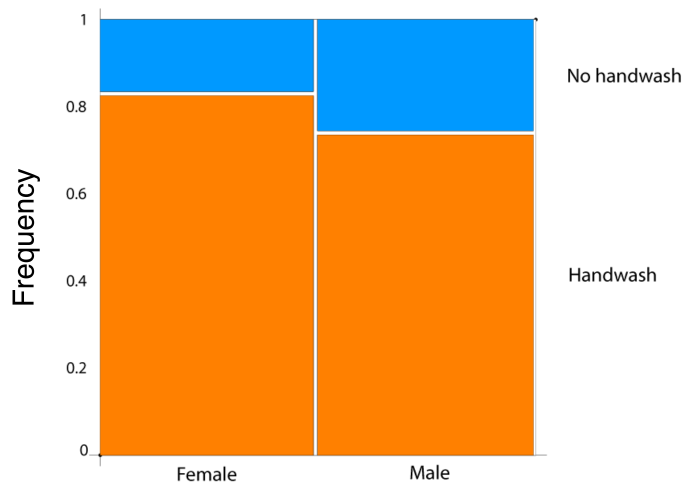
Pr(male) = 0.495

Pr(hand washing) = 0.366 + 0.419 = 0.785

Pr(male AND hand washing) = 0.366 ≠

Pr(male) × Pr(hand washing) = 0.495 × 0.785 = 0.389

So these two events are NOT independent.



Conditional probability

The conditional probability of an event is the probability of that event occurring *given that* a condition is met.

Pr[X|Y]

Pr(X | Y) means the probability of X if Y is true.

It is read as "the probability of X given Y."

Pr(hand washing | male) = 0.74.

The Law of Total Probability

$$\Pr[X] = \sum_{\text{All values of } Y} \Pr[X \mid Y] \Pr[Y]$$

The probability of hand washing is

Pr[hand washing] =
Pr(hand washing | male) Pr(male) +
Pr(hand washing | female) Pr(female)

= 0.74 (0.495) + 0.83 (0.505) = 0.785

The general multiplication rule

Pr[A AND B] = Pr[A] Pr[B | A].

The general multiplication rule

Pr[*A* AND *B*] = Pr[*A*] Pr[*B* | *A*].

Pr[*A* AND *B*] = Pr[*B*] Pr[*A* | *B*].

Therefore

Pr[*A*] Pr[*B* | *A*] = Pr[*B*] Pr[*A* | *B*].

Bayes' theorem

$$\Pr\left[A \mid B\right] = \frac{\Pr\left[B \mid A\right] \Pr[A]}{\Pr[B]}$$

Example: Bayes' theorem

Using data collected in 1975, the probability that women had cervical cancer was 0.0001.

The probability that a biopsy would correctly identify these women as having cancer was 0.90.

The probabilities of a "false positive" (the test saying there was cancer when there was not) was 0.001.

What is the probability that a woman with a positive result actually has cancer?

Pr[cancer | positive result] = ???

Pr[cancer | positive result] = $\frac{\text{Pr[positive result | cancer] Pr[cancer]}}{\text{Pr[positive result]}}$

Pr[cancer] = 0.0001

Pr[no cancer] = 1-0.0001 = 0.9999

Pr[positive result | cancer]=0.9

Pr[positive result |no cancer] = 0.001

Pr[positive result] = ???

Pr[positive result] =

Pr[positive result | cancer] Pr[cancer] +

Pr[positive result | no cancer] Pr[no cancer]

= (0.9)(0.0001) + (0.001)(0.9999)

= 0.0010899

$$\Pr[\text{cancer} \mid \text{positive result}] = \frac{\Pr[\text{positive result} \mid \text{cancer}]\ \Pr[\text{cancer}]}{\Pr[\text{positive result}]}$$

Pr[cancer] = 0.0001

Pr[no cancer] = 1-0.0001 = 0.9999

Pr[positive result | cancer]=0.9

Pr[positive result |no cancer] = 0.001

Pr[positive result] = 0.0010899

Answer

$$\Pr[\text{cancer} \mid \text{positive result}] = \frac{(0.9)(0.0001)}{0.0010899} = 0.0826$$

Pr[cancer] = 0.0001

Pr[no cancer] = 1-0.0001 = 0.9999

Pr[positive result | cancer]=0.9

Pr[positive result |no cancer] = 0.001

Pr[positive result] = 0.0010899