Probability

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
Mutually exclusive

\[ \text{Pr}(A \text{ and } B) = 0 \]

Not mutually exclusive

\[ \text{Pr}(A \text{ and } B) \neq 0 \]
\[ \text{Pr}(\text{purple AND square}) \neq 0 \]

For example:

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

\[ P(A) = 0.48 \]
\[ P(B) = 0.48 \]

But \( P(A \text{ and } B) \neq 0 \), 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

The addition principle

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

$$\Pr[A \text{ OR } B] = \Pr[A] + \Pr[B]$$
The probability of a range

\[ \Pr[\text{Number of boys} \geq 6] = \Pr[6] + \Pr[7] + \Pr[8] \ldots \]

The probabilities of all possibilities add to 1.

Probability of Not

\[ \Pr[\text{NOT rolling a 2}] = 1 - \Pr[\text{Rolling a 2}] = \frac{5}{6} \]

General Addition Principle
General addition principle

\[ \Pr[A \text{ OR } B] = \Pr[A] + \Pr[B] - \Pr[A \text{ AND } B]. \]

Independence

Two events are \textit{independent} if the occurrence of one gives no information about whether the second will occur.

Multiplication principle

The \textit{multiplication principle}: If two events \( A \) and \( B \) are independent, then

\[ \Pr[A \text{ AND } B] = \Pr[A] \times \Pr[B] \]

Offspring of two "carriers":

\[ \Pr[\text{congenital nightblindness}] = 0.25 \]

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

\[ \Pr[ \text{(first child has nightblindness)} \text{ AND } \text{(second child has nightblindness)}] = 0.25 \times 0.25 = 0.0625. \]
Phenotypes in two-child family

<table>
<thead>
<tr>
<th>Phenotype of first child</th>
<th>Phenotype of second child</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Normal</td>
<td>Affected</td>
</tr>
<tr>
<td>Affected</td>
<td>Affected</td>
</tr>
</tbody>
</table>

Probability of second child affected

First Child

Normal

0.75

Affected

0.25

Phenotype of first child

0.25

Affected

0.75

Normal

0.75

Probability trees

Phenotype of first child

0.25

Affected

0.75

Normal

Phenotypes in two-child family

<table>
<thead>
<tr>
<th>Phenotype of first child</th>
<th>Phenotype of second child</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>0.75 x 0.75  = 0.5625</td>
</tr>
<tr>
<td>Normal</td>
<td>Affected</td>
<td>0.75 x 0.25 = 0.1875</td>
</tr>
<tr>
<td>Affected</td>
<td>Normal</td>
<td>0.25 x 0.75 = 0.1875</td>
</tr>
<tr>
<td>Affected</td>
<td>Affected</td>
<td>0.25 x 0.25 = 0.0625</td>
</tr>
</tbody>
</table>
The probability of $A \text{ OR } B$ involves addition.
$Pr(A \text{ or } B) = Pr(A) + Pr(B)$ if the two are mutually exclusive.

The probability of $A \text{ AND } B$ involves multiplication
$Pr(A \text{ and } B) = Pr(A) \cdot Pr(B)$ if the two are independent

**Short summary**

**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

- $Pr[\text{male}] = 0.495$
- $Pr[\text{male washes his hands}] = 0.74$
- $Pr[\text{female washes her hands}] = 0.83$
Hand washing

<table>
<thead>
<tr>
<th>Sex</th>
<th>Washes hands?</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Yes</td>
<td>0.495 x 0.74 = 0.366</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.495 x 0.26 = 0.129</td>
</tr>
<tr>
<td>Female</td>
<td>Yes</td>
<td>0.505 x 0.83 = 0.419</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0.505 x 0.17 = 0.086</td>
</tr>
</tbody>
</table>

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 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[X] = \sum_{All values of Y} Pr[X | Y] Pr[Y]
The general multiplication rule

\[
\Pr[A \text{ AND } B] = \Pr[A] \Pr[B | A].
\]

\[
\Pr[A \text{ AND } B] = \Pr[B] \Pr[A | B].
\]

Therefore

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

Bayes' theorem

\[
\Pr[A | B] = \frac{\Pr[B | A] \Pr[A]}{\Pr[B]}
\]

In class exercise

Using data collected in 1975, the probability of 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?

Using data collected in 1975, the probability of 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?

\[
\Pr[\text{cancer | positive result}] = ???
\]
Pr[cancer | positive result] = \frac{Pr[positive result | cancer] \cdot Pr[cancer]}{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[cancer | positive result] = \frac{Pr[positive result | cancer] \cdot Pr[cancer]}{Pr[positive result | cancer] \cdot Pr[cancer] + Pr[positive result | no cancer] \cdot Pr[no cancer]}

= \frac{(0.9)(0.0001) + (0.001)(0.9999)}{0.0010899}

= \frac{0.0010899}{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