Lecture 8 Independent Events

BIO210 Biostatistics

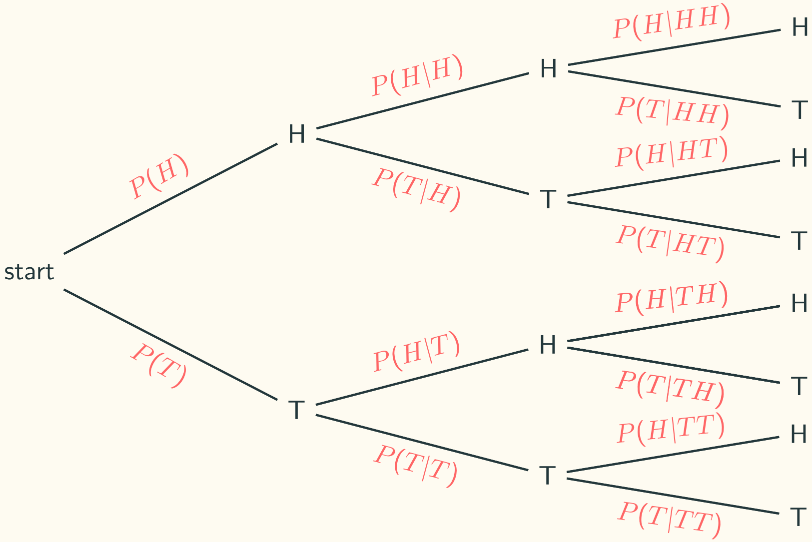
Xi Chen

Spring, 2022

School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Coin Flip Example



$P(HTT) = ?$

$P(\{\text{two heads}\}) = ?$

Independence of Two Events

Definition 1

Events A and B are independent if $P(B|A) = P(B), P(A) \neq 0$

Meaning of Definition 1: the occurrence of A provides no information about the occurrence of B.

Independence of Two Events

Definition 2

Events A and B are independent if $P(A \cap B) = P(A) \cdot P(B)$

Advantages of Definition 2:

- Symmetric with respect to A and B.
- $P(A)$ or $P(B)$ can be 0

Independence

Experiment (Lecture 4): keep flipping a coin until we obtain a head for the first time and stop. Let n be the number of flips.

Sample space: $\Omega = \{$H, TH, TTH, TTTH, ...$\}$

$$P(n) = \frac{1}{2^n},\, n = 1, 2, 3, 4, \ldots$$

Probabilistic model:

$$P(H) = p$$

$$P(T) = 1 - p$$

$$P(TH) = (1-p)p$$

$$P(TTH) = (1-p)(1-p)p$$

$$\vdots$$

$$P(\underbrace{TTT\ldots TTT}_{n-1 \text{ tails}}H) = (1-p)^{n-1}p$$

Independent of a collection of events

Intuitive definition

Information on some of the events does not provide any information about probabilities of the remaining events:

$$P[(A \cap B \cap C \cap D)|(E \cap F)] = P(A \cap B \cap C \cap D)$$

Independent of a collection of events

Mathematics definition

Events $A_1, A_2, A_3, ..., A_n$ are called independent if and only if:

$$P(A_i \cap A_j \cap ... \cap A_q) = P(A_i) \cdot P(A_j) \cdots P(A_q)$$

for any distinct indices $i, j, ..., q$ chosen from $\{1, 2, ..., n\}$

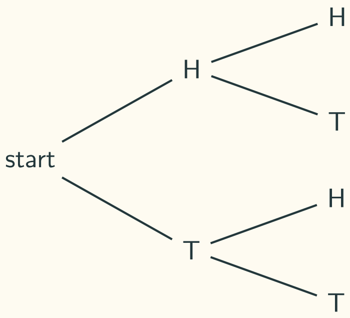
Independent of a collection of events

According to the definition, for a collection of events $\{A_1, A_2, A_3\}$ to be independent, they need to satisfy all the following conditions:

- $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$
- Pairwise independence:
  - $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$
  - $P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$
  - $P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$

Independent of a collection of events

Example 1: two independent coin (fair) flips.



$A = \{$the first is H$\}$

$B = \{$the second is H$\}$

$C = \{$the first and the second give the same result $\}$

$P(A \cap B) = ?$

$P(A \cap C) = ?$

$P(B \cap C) = ?$

$P(C | A \cap B) = ?$

Pairwise independence does not imply independence.

Independent of a collection of events

Example 2: flipping a fair coin 4 times

Sample Space $\Omega = \{$HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, THHT, THTH, TTHH, HTTH, TTTH, TTHT, THTT, HTTT, TTTT$\}$

A$=\{$ HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, THHT $\}$

B$=\{$ THHT, THTH, TTHH, HTTH, TTTH, TTHT, THTT, HTTT $\}$

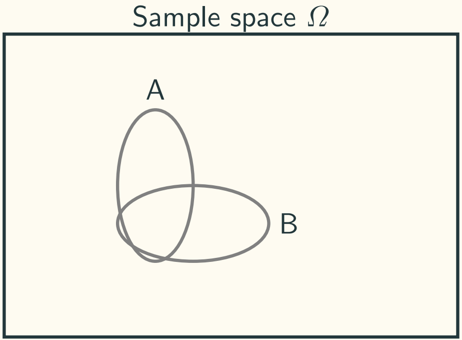
C$=\{$ THHT, THTH, TTHH, HTTH $\}$

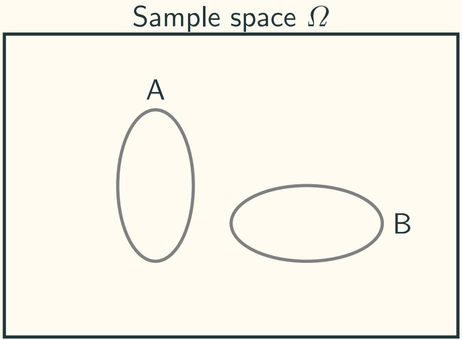
$P(A \cap B \cap C) =?$

$P(B \cap C) =?$

Simple multiplication does not imply independence.

Independent Events





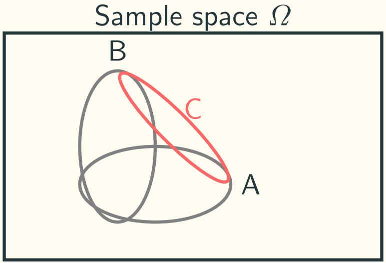
- Venn diagram is not sufficient to display independent events.
- Do not confuse independent events with disjoint events.

Conditional Independence

Definition

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

Events A and B are independent in the following Venn diagram:



Having independence in the original model does not imply independence in the conditional model.

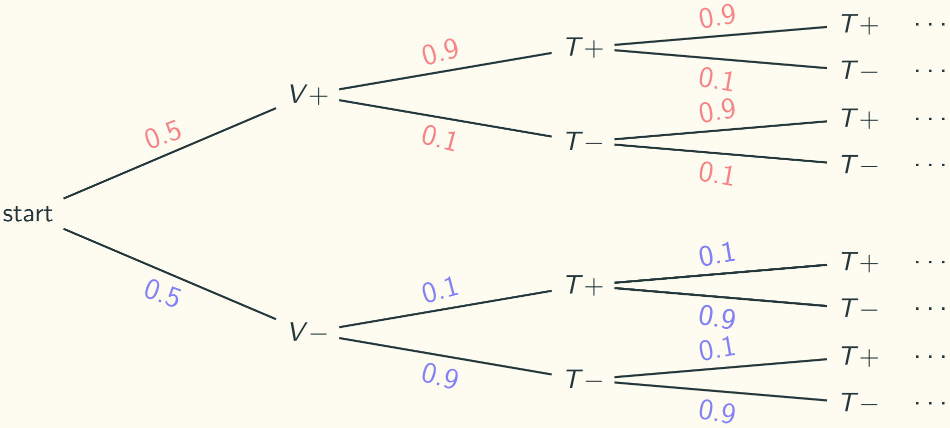
Conditional independence

Example of conditional independence - a virus detection kit:

- If a person carries the virus, the kit has 90% of the chance of showing a positive result.
- If a person dow not carry the virus, the kit has 90% of the chance of showing a negative result.
- The virus is common and non-harmful. In general, 50% of the whole population carry the virus without any symptoms or illness.
- We have a random person called Li Lei. He gets tested by the kit repeatedly. Let event A $=$ { the 11th test is positive } and event B $=$ { the first 10 tests are all positive }

Questions: are events A and B independent? Does the answer depend on if we know Li Lei carries the virus or not?

Conditional independence



Conditional independence

- A = { the 11th test is positive }
- B = { the first 10 tests are all positive }

- We know Li Lei carries the virus.
  - $P(A) = ?$
  - $P(A|B) = ?$

- We know Li Lei does NOT carries the virus.
  - $P(A) = ?$
  - $P(A|B) = ?$

- We don't know if Li Lei carries the virus or not.
  - $P(A) = ?$
  - $P(A|B) = ?$

Having independence in the conditional model does not imply independence in the original model.

The Gambler's Fallacy

The Gambler's Fallacy in RPG Games

The Sally Clark Case

- Sudden infant death syndrome (SIDS) is the sudden unexplained death of a child of less than one year of age.
- Clark's first son died in December 1996 within a few weeks of his birth.
- Her second son died in similar circumstances in January 1998.
- She was convicted in November 1999. The convictions were overturned in January 2003.
- As a result of her case, the Attorney-General ordered a review of hundreds of other cases, and two other women had their convictions overturned.

The Sally Clark Case

The CESDI Report

Groups	SIDS incidence in this group
Overall population	363 in 472,823
Anybody smokes in the household	1 in 737
Nobody smokes in the household	1 in 5041
No waged income in the household	1 in 486
At least one waged income in the household	1 in 2,088
Mother $<$ 27 years and parity	1 in 567
Mother $>$ 26 years and parity	1 in 1882
None of these factors	1 in 8,543
One of these factors	1 in 1,616
Two of these factors	1 in 596
All three of these factors	1 in 214

The Sally Clark Case

Professor Sir Roy Meadow, a highly respected expert in field of child abuse at the time of the trial:

"you have to multiply 1 in 8,543 times 1 in 8,543 and I think it gives that in the penultimate paragraph, it points out that it's approximately a chance of 1 in 73 million "

The Sally Clark Case: one of the great miscarriages of justice in modern British legal history