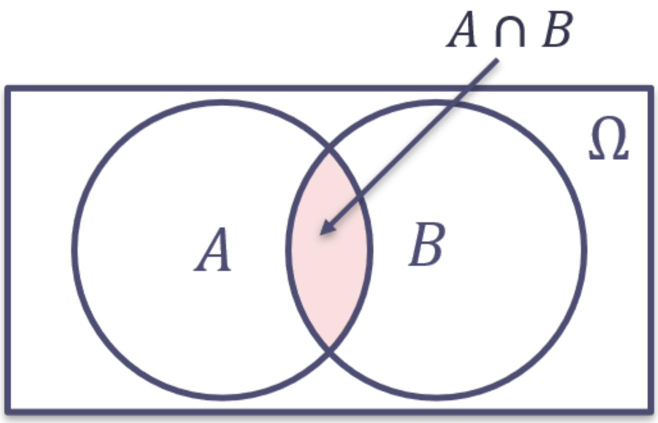
Conditional Probabilities

Recap

Conditional Probability:

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



Bayes' Formula:

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

Mutual Independence

The events $A_1, ..., A_n$ are **mutually independent**, if $\forall k \in \{1, ..., n\}$ and $\forall (i_1, ..., i_k) \in \mathbb{N}^k$ such that $1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq n$, we have:

$$P(A_{i_1} \cap ... \cap A_{i_k}) = P(A_{i_1}) \times ... \times P(A_{i_k})$$

i.e., for the case of 3 events A, B, C, the following conditions must be satisfied:

1. pairwise independence:

- $P(A \cap B) = P(A) \times P(B)$
- $P(A \cap C) = P(A) \times P(C)$
- $P(B \cap C) = P(B) \times P(C)$

2. $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

Remark: events can be pairwise independent without being mutually independent

Exercise 1

A family has two children (consecutively). What is the probability that they are both boys, knowing that the first is a boy? What is the probability that they are both boys knowing that at least one of the two is a boy?

Exercise 2

Let A, B and C be the events corresponding to the toss of two balanced and distinguishable coins as follows:

- $A =$ "*The first coin landed on heads*",
- $B =$ "*The second coin landed on heads*",
- $C =$ "*The two coins landed on different faces*".

Show that A, B and C are pairwise independent. Are they (mutually) independent?

Exercise 3

A warehouse is equipped with an alarm device. When there is an attempted burglary, the device is triggered with a probability equal to 0.99. When there is no attempted burglary, the device is triggered anyway by error during a day with a probability equal to 0.01. Assuming that an attempted burglary during a day occurs with a probability equal to 0.001, what is the probability that an alarm triggered on a given day is triggered by an attempted burglary?

We will denote D the event "*the alarm is triggered*" and T the event "*there is an attempted burglary*", all considered during the given day.

Exercise 4

We are interested in the transfer of votes in the context of the second round of a two-round majority election.

Suppose there were 4 candidates in the first round, A, B, C and D. The scores of the candidates in the first round are 35% for A, 25% for B, 20% for C and 20% for D among the votes cast. 23% of voters abstained.

Candidates C and D were eliminated in the first round of the election. Among the electorate of C from the first round,

- 75% will vote for A in the second round,
- 20% will vote for B in the second round,
- 5% will abstain.

Among the electorate of D from the first round,

- 50% will vote for A in the second round,
- 15% will vote for B in the second round,
- 35% will abstain.

We assume that people who voted for A and for B in the first round will not change their mind and will not abstain in the second round. We also assume that people who abstained in the first round will not go vote in the second.

1. If I take a ballot for A in the second round, what is the probability that it was deposited by someone who voted for C in the first round?
2. Who will win the election?
3. What will be the abstention rate in the second round?

Simpson's Paradox

In a high school, the following results were observed on an exam:

Track	Liberal Arts		Science		Total	
Gender	F	B	F	B	F	B
Success	20	1	200	600	220	601
Failure	180	19	100	400	280	419
Success rate	$1/10 > 1/20$		$2/3 > 3/5$		$0.44 < 0.59$	

Are girls better than boys at the exam?

Comments

This paradox describes a situation where a phenomenon observed (the success rate on an exam) in several groups can be reversed when the data is considered as a whole. This result is related to elements that are not taken into account (e.g., the presence of non-independent variables or differences in group sizes, etc.).

We can reorganize the table as follows:

	F	B
Liberal Arts	1/10 (20/200)	1/20 (1/19)
Science	2/3 (200/300)	3/5 (600/1000)
Total	0.44 (220/500)	0.59 (601/1020)

There are differences between groups that are not taken into account.

For example, note that the group sizes (the number of representatives of each group) are very different: * 200 F vs. 20 B (liberal arts) * 300 F vs. 1,000 B (science) * 500 F vs 1,020 B (total) The groups by track are not balanced at the F:B level.

We note that the total number of girls is 2 times smaller than the number of boys, which plays significantly in the percentage calculation.

In addition, the total result for boys is dominated by the "Science" track (1,000 vs 19), and a bit less for girls (300 vs. 200).

Looking at the rates by track, we also note that students encountered more difficulty in the Liberal Arts track.

We thus note that context is important to qualify the notion of success.