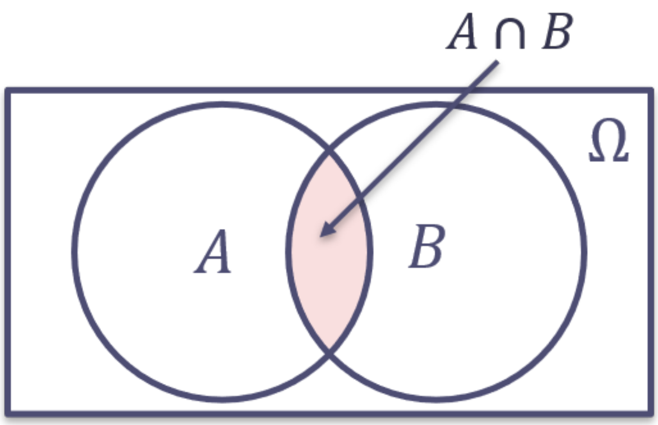
Conditional Probabilities

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?

Solution

We consider that the possible outcomes are: $\Omega = \{BB, BG, GB, GG\}$, where BG means that the 1st child is a boy (B) and the 2nd child is a girl (G).

We can define the following table of possible options:

child 1 \ child 2	G	B
G	GG	GB
B	BG	BB

Remark: as we did for dice

Two boys knowing that the 1st is a boy Knowing that the 1st child among 2 is a boy, the possibilities are: BG and BB. Therefore, the probability of having two boys is one in two: $P(BB) = 1/2 = 0.5$.

child 1 \ child 2	G	B
G	GG	GB
B	BG	BB

Another way to represent:

We can represent the possibilities in the form of the following tree (B = boy, G = girl):

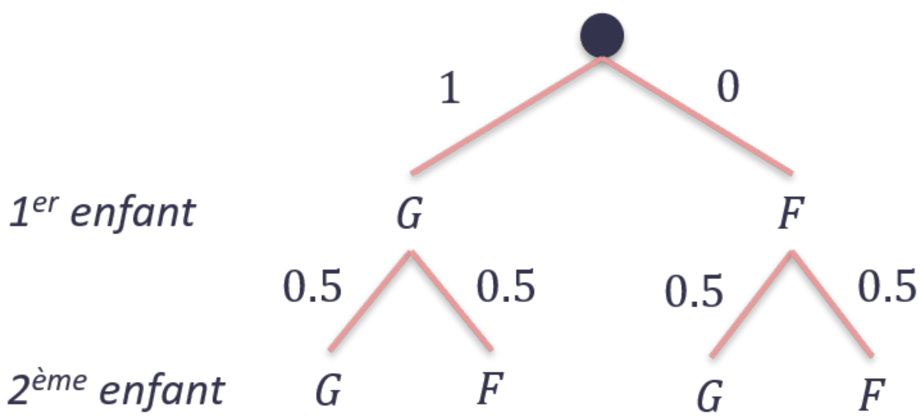


Figure 1: Fig.2. Case 1: Graph with probabilities

When we know that the 1st child is a boy, the probability on this branch is 1.

For the 2nd child, the remaining choice is between B (boy) or G (girl) with the same probability of 1/2.

The desired case is BB:

Let's multiply the probabilities on the branch leading to BB: $P(BB|1st = B) = 0.5 \times 1 = 0.5$

Two boys knowing that at least one of the two is a boy Knowing that at least 1 is a boy, the possibilities for the two children are: BB, BG, GB. Therefore, $P(BB) = 1/3$.

child 1 \ child 2	G	B
G	GG	GB
B	BG	BB

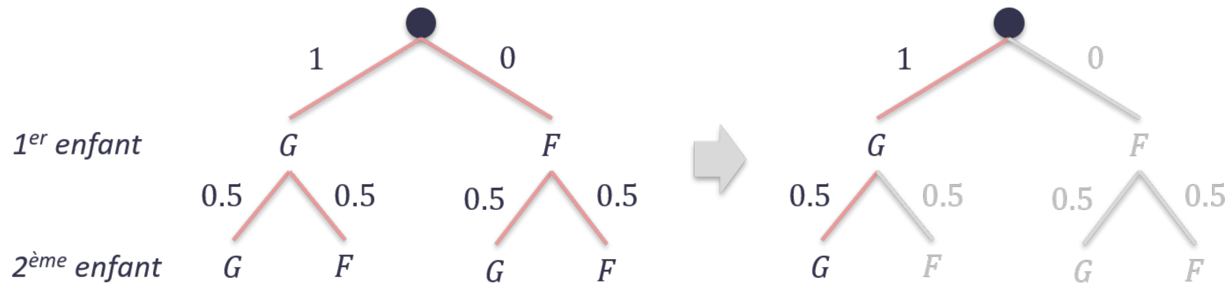


Figure 2: Fig.3. Case 1: Graph with probabilities for BB

By writing Bayes' formula:

$$P(BB| \geq 1B) = \frac{P(\geq 1B|BB) \cdot P(BB)}{P(\geq 1B)}$$

Note that $P(\geq 1B|BB) = 1$ (the probability of having at least 1 boy knowing that there are two).

$P(\geq 1B)$ corresponds to the following cases: BB, BG, GB among 4 possible (recall $\Omega = \{BB, BG, GB, GG\}$). Therefore, $P(\geq 1B) = 3/4$.

$P(BB) = 1/4$ because it corresponds to one case among 4 possible.

Therefore:

$$P(BB| \geq 1B) = \frac{P(\geq 1B|BB) \cdot P(BB)}{P(\geq 1B)} = \frac{1 \cdot 1/4}{3/4} = 1/3$$

Graphically, we could represent the solution as follows:

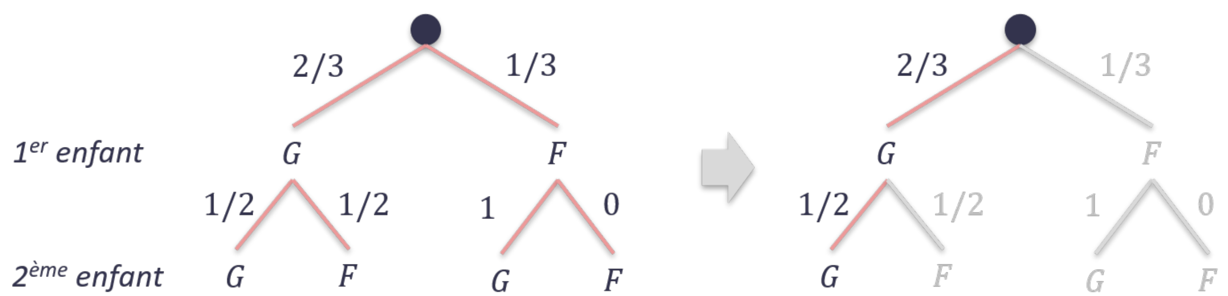


Figure 3: Fig.4. Case 2: Graph with probabilities

The probability 2/3 on the B branch for the 1st child comes from the fact that this branch gives 2 possibilities out of 3 to satisfy the condition "at least 1 boy".

Let's multiply the probabilities of the branch leading to BB: $P(BB| \geq 1B) = 2/3 \cdot 1/2 = 1/3$.

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?

Solution

We can represent the possible outcomes in the form of the following table:

coin 1 \ coin 2	Heads (H)	Tails (T)
Heads (H)	HH	HT
Tails (T)	TH	TT

We can thus define the sample space as follows: $\Omega = \{HH, HT, TH, TT\}$, where HT corresponds to the case where the 1st coin landed on heads (H), and the 2nd on tails (T).

Note that the number of possible results (outcomes) is 4.

We can rewrite the events in set notation as follows:

- $A =$ "$The\ first\ coin\ landed\ on\ heads$" $= \{HH, HT\}$,
- $B =$ "$The\ second\ coin\ landed\ on\ heads$" $= \{TH, HH\}$,
- $C =$ "$The\ two\ coins\ landed\ on\ different\ faces$" $= \{HT, TH\}$.

Let's calculate the probabilities of each event:

- $P(A) = 2/4 = 1/2$
- $P(B) = 2/4 = 1/2$
- $P(C) = 2/4 = 1/2$

Pairwise Independence To verify pairwise independence, we can use the conditional probability formula. Thus, two events A and B are independent if

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

.

Let's calculate the probabilities of intersections:

- $A \cap B = \{HH\} \Rightarrow P(A \cap B) = 1/4$
- $A \cap C = \{HT\} \Rightarrow P(A \cap C) = 1/4$
- $B \cap C = \{TH\} \Rightarrow P(B \cap C) = 1/4$

Let's apply the formula:

- $1/4 = P(A \cap B) = P(A) \cdot P(B) = 1/2 \times 1/2 = 1/4$
- $1/4 = P(A \cap C) = P(A) \cdot P(C) = 1/2 \times 1/2 = 1/4$
- $1/4 = P(B \cap C) = P(B) \cdot P(C) = 1/2 \times 1/2 = 1/4$

Thus we can deduce the pairwise independence of these three events.

Mutual Independence For mutual independence, 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. $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

Condition 1 is satisfied (see above) because events A, B and C are pairwise independent. It remains to verify condition 2.

Let's calculate the probability of the intersection $P(A \cap B \cap C)$:

$$A \cap B \cap C = \emptyset \Rightarrow P(A \cap B \cap C) = 0$$

$$0 = P(A \cap B \cap C) \neq P(A) \times P(B) \times P(C) = 1/2 \times 1/2 \times 1/2 = 1/8$$

Therefore, events A, B and C are not mutually independent.

Remark: Note that if the coin were not balanced (we get "heads" with probability p, where $p \neq 1/2$, C would not have been independent of either A or B.

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.

Solution

$D =$ "*the alarm is triggered*"

$T =$ "*there is an attempted burglary*"

We know the following probabilities:

1. When there is an attempted burglary, the device is triggered with a probability equal to 0.99: $P(D|T) = 0.99$
2. When there is no attempted burglary, the device is triggered with a probability equal to 0.01: $P(D|\bar{T}) = 0.01$
3. An attempted burglary during a day occurs with a probability equal to 0.001: $P(T) = 0.001$

We are looking for: $P(T|D) =?$

We can use Bayes' formula:

$$P(T|D) = \frac{P(D|T) \cdot P(T)}{P(D)}$$

We therefore need to find the total probability of event D, $P(D)$.

To find it, we can represent the case in the form of the following tree:

Knowing that $P(\bar{A}) = 1 - P(A)$, we can find $P(\bar{T}) = 1 - P(T) = 1 - 0.001 = 0.999$.

The total probability of D is then given by:

$$P(D) = P(D|T) \cdot P(T) + P(D|\bar{T}) \cdot P(\bar{T}) = \frac{99}{100} \cdot \frac{1}{1000} + \frac{1}{100} \cdot \frac{999}{1000} = \frac{99}{10000} + \frac{999}{10000} = \frac{1098}{10000}$$

Let's return to Bayes' formula:

$$P(T|D) = \frac{P(D|T) \cdot P(T)}{P(D)} = \frac{\frac{1}{1000} \cdot \frac{99}{100}}{\frac{1098}{10000}} = \frac{\frac{99}{10000}}{\frac{1098}{10000}} = \frac{99}{1098} = 0.09016393 \approx 0.09$$

Figure 4: Fig.5. Alarm triggering

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,

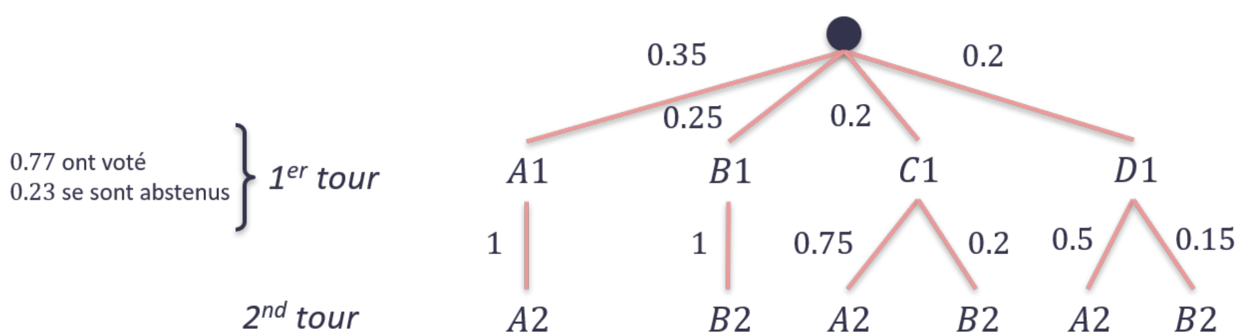
- 50% will vote for A in the second round,
- 15% will vote for B in the second round,
- 35% will abstain.

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?

Solution

We can represent the data in the form of the following tree:



The numbers 1 or 2 are used to distinguish votes in the 1st and 2nd rounds, i.e., $A1$ designates votes for candidate A in the first round and $A2$ in the second round.

Probability that the vote for A in the 2nd round comes from people who voted C in the 1st round We are looking for $P(C1|A2)$.

To answer this question, we can use Bayes' formula:

$$P(C1|A2) = \frac{P(A2|C1) \cdot P(C1)}{P(A2)}$$

Let's calculate the total probability of $A2$:

$$P(A2) = P(A2|A1) \cdot P(A1) + P(A2|C1) \cdot P(C1) + P(A2|D1) \cdot P(D1) = 1 \cdot 0.35 + 0.75 \cdot 0.2 + 0.5 \cdot 0.2 = 0.35 + 0.15 + 0.1 = 0.6$$

Let's return to Bayes' formula:

$$P(C1|A2) = \frac{P(A2|C1) \cdot P(C1)}{P(A2)} = \frac{0.75 \cdot 0.2}{0.6} = \frac{0.15}{0.6} = 1/4 = 0.25$$

Who will win the election? The choice in the second round is between 2 candidates: A and B.

According to the previous calculation, $P(A2) = 0.6$ which represents the majority of votes. Thus, we can deduce that candidate A will win.

The abstention rate in the second round We know that in the 1st round the abstention rate is 23% and we also assume that people who abstained in the first round will not go vote in the second. (We must include these 23% in the calculation.)

That is, there are 77% of the population who participated in the second round.

Among these 77% of the population, we know that 5% of those who voted C in the 1st round and 35% of those who voted D in the 1st round will abstain.

Thus the abstention rate in the second round can be calculated as follows:

$$rate = 0.23 + 0.77(P(abstention2|C1) \cdot P(C1) + P(abstention2|D1) \cdot P(D1)) = 0.23 + 0.77(0.05 \cdot 0.2 + 0.35 \cdot 0.2) =$$

$$= 0.23 + 0.77(0.01 + 0.07) = 0.23 + 0.77 \cdot 0.08 = 0.23 + 0.0616 = 0.2916$$

The abstention rate in the second round is 29.16%.

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.