1 Conditional probability

If two events A and B are **not independent**, knowledge that one event occurred, say event B, may influence the likelihood that A occurs. To express this mathematically, one defines the concept of **conditional probability** as follows.

1.1 Definition

If A and B are two events in the sample space Ω, and $P(B) > 0$, then the probability that A occurs given that B occurred is called the **conditional probability of** A **given** B, written $P(A|B)$, and given by

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}} \tag{1}$$

Note 1: If two events A and B are mutually exclusive, $A \cap B = \emptyset$ so $P(A \cap B) = 0$, and according to the equality above, $P(A|B) = 0$ as one would expect.

Note 2: The definition can be rewritten as

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

which is sometimes called the multiplication rule. This formula can be quite convenient when computing $P(A \cap B)$ directly is difficult. You will encounter several such situations in problems we will solve in this course.

1.2 Examples

- Imagine the experiment consists in rolling a dice once, and consider the events

  $$A = \{\text{4 appears}\} \quad , \quad B = \{\text{The outcome is even}\}$$

  $A \subset B$, so the knowledge that B occurs clearly changes the likelihood that A occurs. Here, we have $P(A|B) = \frac{1}{3}$, which can be seen directly with a simple counting argument.

  One may also verify:

  $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

- Let us return to our day of birth experiment, and consider the events

  $$A = \{\text{The day of the week has 6 letters}\} \quad , \quad B = \{\text{The day of the week starts with an `S'}\}$$

  The probability that the day of the week has 6 letters given that the day of the week starts with an $'S'$ is directly calculated as

  $$P(A|B) = \frac{1}{2}$$

  One can verify that

  $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{7}}{\frac{2}{7}} = \frac{1}{2}$$

- Given three people randomly chosen in the population, let us compute the probability that their birthday (viewed now as day of the year) are all different. For this calculation, we will assume that any day of the year is equiprobable, which is a simplification which does not exactly match reality [1].

[1] Indeed, the number of births has been observed to vary in a noticeable way depending on the time of the year, and depending on geographical, historical, and sociological factors, as you can see in the following interesting research article: *Human birth seasonality: latitudinal gradient and interplay with childhood disease dynamics* by Micaela Martinez-Bakker, Kevin M. Bakker, Aaron A. King and Pejman Rohani, *Proceedings of the Royal Society B* **281**:20132438 (2014)

(i) We start with two people: whatever the birthday of the first person is, there are 364 days to "choose" from for the birthday of the second person not to match the first person's. Ignoring leap years for simplicity, we may write

$$P(B_2) = \frac{364}{365} \approx 0.99726$$

where B_2 is the event "The two persons do not have their birthdays on the same day".

(ii) We are now ready for the situation with three people, and that is where conditional probabilities come in handy.

The event $B_3 =$"The three randomly chosen people all have different birthdays" may be viewed as the intersection of B_2 with the event $A_3 =$"The third person has a birthday that does not coincide with either of the birthdays of the first two persons". We can then write

$$P(B_3) = P(A_3 \cap B_2) = P(A_3|B_2)P(B_2)$$

Now,

$$P(A_3|B_2) = \frac{363}{365}$$

so

$$P(A_3 \cap B_2) = \frac{363 \times 364}{365^2} \approx 0.99180$$

(iii) Let us push the envelope: what is the formula for N people? Well, we can continue with the same approach, and write

$$\begin{aligned}
P(B_N) &= P(A_N|B_{N-1})P(B_{N-1}) \\
&= \frac{365 - (N-1)}{365} P(B_{N-1}) \\
&= \frac{366 - N}{365} P(A_{N-1}|B_{N-2})P(B_{N-2}) \\
&= \frac{366 - N}{365} \frac{365 - (N-2)}{365} P(B_{N-2}) \\
&= \dots \\
&= \overbrace{\frac{366 - N}{365} \frac{367 - N}{365} \cdots \frac{363}{365}}^{N-2 \text{ terms}} P(B_2) \\
&= \frac{(366 - N)(367 - N) \dots 363 \times 364}{365^{N-1}}
\end{aligned}$$

With a calculator and some patience, you can then verify that

$$P(B_{22}) \approx 0.5243 \qquad \text{and} \qquad P(B_{23}) \approx 0.4927$$

In other words, in a group of 23 randomly chosen people, the probability that two people have their birthday on the same day is greater than $\frac{1}{2}$! And that result is even pessimistic, as it assumes that every day of the year is equiprobable, as we already discussed.

1.3 Bayes' formula

Consider two events A and B in a sample space Ω, such that $P(B) > 0$. We just saw that

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

Now,

$$A \cap B = B \cap A$$

so

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

We just proved Bayes' formula:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \tag{2}$$

<u>Elementary illustrations of the formula</u>

- In the example of the dice we treated in page 1, $P(B|A) = 1$ so the formula leads to

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

  as before.

- For the example of the day of the week, $P(B|A) = \frac{1}{3}$, so the formula gives

  $$P(A|B) = \frac{\frac{1}{3} \cdot \frac{3}{7}}{\frac{2}{7}} = \frac{1}{2}$$

  recovering the result we got previously.

It is often the case that $P(B|A)$ and $P(A|B)$ are not equally easy/hard to compute. Bayes' formula can be a convenient way to go from one to the other.

1.4 Law of total probability

If $B_1, \ldots, B_N$ are **mutually disjoint events** such that $\bigcup_{i=1}^{N} B_i = \Omega$, and $P(B_i) > 0$ for all $i = 1, 2, \ldots, N$, then

$$\boxed{P(A) = \sum_{i=1}^{N} P(A|B_i)P(B_i)} \tag{3}$$

This is called the **law of total probability**.

Proof:

$$\begin{aligned} P(A) &= P(\bigcup_{i=1}^{N} (A \cap B_i)) \\ &= \sum_{i=1}^{N} P(A \cap B_i) \qquad \text{(additive property of probability)} \\ &= \sum_{i=1}^{N} P(A|B_i)P(B_i) \qquad \square \end{aligned}$$

A graphical illustration of the situation for the case $N = 4$ is given in Figure 1. It can help to understand the first equality in our proof, for example.

The law of total probability is often combined with Bayes' formula as follows. For any $i = 1, 2, \ldots, N$,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

Thus,

$$\boxed{P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{N} P(A|B_j)P(B_j)}}$$

This result is sometimes called **Bayes' theorem**.

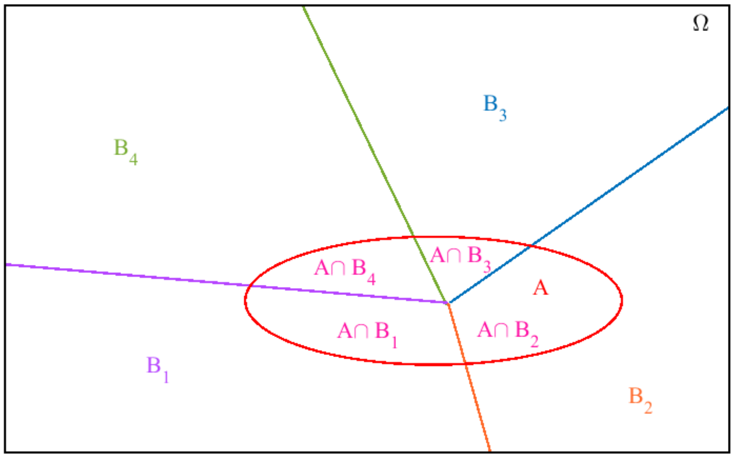


Figure 1: Illustration of the set configuration for the law of total probability, given in Eq.(3), with $N = 4$.

Example 1: A football team wins $\frac{2}{3}$ of its games when it scores first, and $\frac{2}{10}$ of its games when the opposing team scores first. If the team scores first $\frac{1}{3}$ of the time, what fraction of the games does it win?

Let $W =$"The team wins" and $SF =$"The team scores first". We use the law of total probability to write

$$\begin{aligned} P(W) &= P(W|SF)P(SF) + P(W|SF^{\mathrm{C}})P(SF^{\mathrm{C}}) \\ &= \frac{2}{3}\frac{1}{3} + \frac{2}{10}\frac{2}{3} = \frac{2}{9} + \frac{4}{30} = \frac{20+12}{90} = \frac{16}{45} \approx 35.6\% \end{aligned}$$

Example 2: Blood tests for a disease are never 100% reliable. Consider a test and a disease with the following properties.

Let T be the event "The test is positive".

Let D be the event "The person has the disease".

The blood test is such that:

$$P(T|D) = 0.99 \qquad P(T^{\mathrm{C}}|D^{\mathrm{C}}) = 0.99$$

and the disease is such that for the population which is considered,

$$P(D) = 0.01$$

What is the probability that someone who tests positive actually has the disease?

Observe that $P(T|D^{\mathrm{C}}) = 1 - 0.99 = 0.01$, so the probability is certainly not 1. Let us compute that probability, using Bayes' theorem.

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^{\mathrm{C}})P(D^{\mathrm{C}})} \\ &\frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.01 \cdot 0.99} = \frac{1}{2} \end{aligned}$$

We see that because the disease is so rare, we find that 50% of those who test positive do not have the disease, even though the test is fairly reliable.

2 Independence

2.1 Definition

Two events are said to be **independent** if

$$\underline{P(A \cap B) = P(A)P(B)}$$

2.2 Immediate consequences

- Say A and B are independent, and $P(B) > 0$. Then

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

  Conversely, if $P(A|B) = P(A)$, $P(A \cap B) = P(A)P(B)$. We thus proved the following property:

$$\text{If } P(B) > 0 \;, \qquad \boxed{P(A \cap B) = P(A)P(B) \;\Leftrightarrow\; P(A|B) = P(A)}$$

  The right-hand side of the equivalence conveys the concept of independence quite intuitively: if the occurence of B does not affect the occurence of A, one would expect $P(A|B) = P(A)$.

- We know that

$$P(A^{\mathrm{C}}|B) = 1 - P(A|B)$$

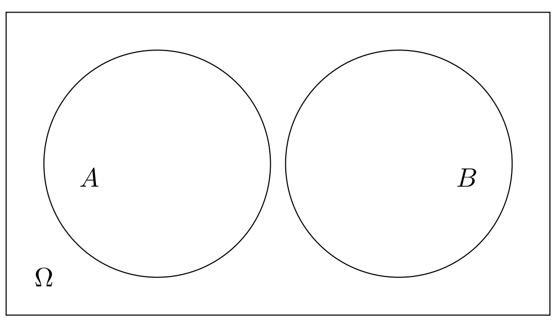
  If A and B are independent,

$$P(A^{\mathrm{C}}|B) = 1 - P(A) = P(A^{\mathrm{C}})$$

  In other words, if A and B are independent, A^{C} and B are also independent.

2.3 Independence vs. mutually exclusive

One should not confuse the concepts of independent events and disjoint events. They are very different in nature. To see this, consider the events A and B as shown in the figure below:



Disjoint events A and B in a sample space Ω.

$P(A \cap B) = 0$, which is the signature of *disjoint* events.
However, the events are NOT independent, since

$$P(A) \neq 0 \;\;,\;\; P(B) \neq 0 \;\Rightarrow\; P(A)P(B) \neq 0$$

We thus have

$$P(A)P(B) \neq P(A \cap B)$$

Another way to view this is that we have

$$P(A|B) = 0 \neq P(A)$$

2.4 Examples

- Consider an unfair coin such that $P(\{H\}) = 0.9$, $P(\{T\}) = 0.1$, which is tossed (fairly) three times in a row. Consider the events:

  A="H is obtained in the first coin toss"

  B="H is obtained in the second coin toss"

  C="T is obtained in the second coin toss"

  D="$3H$ are obtained in a row"

  $P(A) = 0.9$

  $P(B) = \underbrace{0.9 \cdot 0.9}_{\text{first two tosses are HH}} + \underbrace{0.1 \cdot 0.9}_{\text{first two tosses are TH}} = 0.9 = P(B|A)$

  The events A and B are independent.

  $P(C) = \underbrace{0.9 \cdot 0.1}_{\text{first two tosses are HT}} + \underbrace{0.1 \cdot 0.1}_{\text{first two tosses are TT}} = 0.1 = P(C|A)$

  The events A and C are independent.

  The events B and C are not independent, since $C = B^{\mathrm{C}}$.

  $P(D) = 0.9^3 = 0.729$.

  $P(A \cap D) = P(D) = 0.9^3 \neq P(A)P(D) = 0.9^4$. The events A and D are not independent.

- We roll a dice twice, and consider the events:

  E_1="The sum of the outcomes is odd"

  E_2="The outcome of the first dice is 1"

  Let us count.

  $$|\Omega| = 36$$

  The event E_1 is the set

  $$E_1 = \{(1,2),(1,4),(1,6),(2,1),(2,3),(2,5),(3,2),(3,4),(3,6),(4,1),(4,3),(4,5),(5,2),(5,4),(5,6),(6,1),(6,3),(6,5)\}$$

  $$|E_1| = 18$$

  The event E_2 is the set

  $$E_2 = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}$$

  $$|E_2| = 6$$

  Now,

  $$E_1 \cap E_2 = \{(1,2),(1,4),(1,6)\}$$

  $$|E_1 \cap E_2| = 3$$

  We thus have

  $$P(E_1 \cap E_2) = \frac{3}{36} = \frac{1}{12} \qquad , \qquad P(E_1)P(E_2) = \frac{18}{36} \cdot \frac{6}{36} = \frac{1}{12}$$

  $P(E_1 \cap P(E_2) = P(E_1)P(E_2)$, so the events E_1 and E_2 are independent.

Note: The second example is interesting, in the sense that one could have imagined that the probability of E_1 is modified by the knowledge that E_2 occurred (although elementary mathematical reasoning explains why it is not).

When events are the result of physically independent experiments (e.g. two events each associated with a separate coin toss), we expect these events to be independent mathematically, and this is indeed true.

However, it can also happen on certain occasions that events which are associated with the same physical process are mathematically independent, as we have seen in the example we just covered.

For this reason, the mathematical/probabilistic concept of independence is sometimes more precisely called "stochastic independence" or "statistic independence". In this course, we will still use the word independence by itself, for conciseness.

2.5 Independence for multiple events

Let us imagine we consider 3 events A, B, and C. How does one mathematically express the fact that the three events are independent?

Intuitively, it makes sense that the three events should be independent pairwise:

$$P(A \cap B) = P(A)P(B) \quad , \quad P(A \cap C) = P(A)P(C) \quad , \quad P(B \cap C) = P(B)P(C)$$

It turns out that these three conditions are not sufficient, as can be seen from the following standard example.

Consider two successive tosses of a fair coin.
Let A="Heads on toss 1"
B="Heads on toss 2"
C="The two tosses are equal"
We have

$$P(A) = \frac{1}{2} \quad , \quad P(B) = \frac{1}{2} \quad , \quad P(C) = \frac{1}{2}$$

and

$$P(A \cap B) = \frac{1}{4} = P(A)P(B) \ , \ P(A \cap C) = \frac{1}{4} = P(A)P(C) \ , \ P(B \cap C) = \frac{1}{4} = P(B)P(C)$$

Yet

$$P(A \cap B \cap C) = P(A \cap B) = \frac{1}{4} \neq P(A)P(B)P(C)$$

The last result shows that A, B, and C are not independent, as one intuitively understands: if A and B occur, then C MUST occur.

Three events A, B, and C are independent if and only if:

$$\boxed{P(A \cap B) = P(A)P(B) \quad , \quad P(A \cap C) = P(A)P(C) \quad , \quad P(B \cap C) = P(B)P(C) \quad , \quad P(A \cap B \cap C) = P(A)P(B)P(C)}$$

More generally, N events A_1, A_2,…,A_N are called independent if

$$\boxed{\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) && \text{for all } i = 1, \ldots, N \ , \ j = 1, \ldots, N \text{ such that } i \neq j \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) && \text{for all } i = 1, \ldots, N \ , \ j = 1, \ldots, N \ , \ k = 1, \ldots, N \text{ such that } i \neq j \neq k \\ &\vdots && \vdots \\ P(A_1 \cap A_2 \cap \ldots \cap A_N) &= P(A_1)P(A_2) \ldots P(A_N) \end{aligned}}$$