Statistics: probability

S. Maset

Dipartimento di Matematica e Geoscienze, Università di Trieste

PEM 2018-2019

Outline

Introduction

- Suppose we want to obtain information about the winner of a ballot between two candidates A and B in some political election.

  Assume that a representative sample of 100 voters has been (randomly) chosen. If 62 of them declare to vote for A, can we conclude that the winner of the ballot will be A?

  Of course, we cannot conclude this. In fact, the proportion of A's supporters in the sample can be greater than the proportion of A's supporters in the population of voters (the electorate).

  But, if we know that it is very difficult to have 62 voters for A in a representative sample of 100 voters, when A has less than 50% of the voters in the electorate, then we can be quite confident that A will be the winner of the ballot.

Obtaining information on the winner of the election starting from a sample of voters is what we have called **Inferential Statistics**.

As a general rule, in order to draw valid inferences about a population from a sample, one needs to know how likely are certain events regarding the population.

For example, in the ballot situation, we need to know how likely is the event "there are 62 voters for A in a sample of 100 voters", when A has less than 50% of the voters in the electorate.

The determination of the likelihood, or possibility, that an event will occur is the subject of study of **Probability Theory**, that is presented now.

We come back to this question regarding the ballot later on.

- The word "probability" is a commonly used term that relates to the possibility that a particular event will occur when some experiment is performed.

  For example, in the experiment given by the flip of a normal coin, we say that the event "Head appears" has probability 50%.

  The basic concepts of Probability Theory are those of **experiment**, **event** and **probability** that now we introduce.

Experiments with outcomes not known a priori

- An **experiment** is a process that produces an **outcome**, not known a priori.

  The set of all possible outcomes is called the **sample space** of the experiment.

  In the following, we denote the sample space by Ω.

The name "sample space" comes from the fact that when we try to learn about something regarding a population by examining an its sample of a given size *n*, first of all we have to accomplish the experiment of selecting such a sample (the outcome of the experiment) from the set of all possible sample of size *n* in the population. In this particular experiment, it does make sense to call "sample space" the set of all possible outcomes.

In the ballot example, we select a sample of 100 voters from the set of all possible samples of 100 voters in the electorate.

- Now, we present some examples of experiments.

Example.

- *Experiment: first birth in Italy in 2019.*
- *Outcome: gender of the newborn.*
- *Sample space:*

$$\Omega = \{girl, boy\}.$$

Example.

- *Experiment: flip two coins.*
- *Outcome: the pair of the faces shown. It is understood that there are a first coin and a second coin: the pair outcome has the face shown by the first coin as first component and the face shown by the second coin as second component.*
- *Sample space:*

$$\Omega = \{H, T\}^2 = \{(H, H), (H, T), (T, H), (T, T)\},$$

where H stands for Head and T stands for Tail.

Example.

- *Experiment: roll two six-sided dice.*
- *Outcome: the pair of scores obtained. It is understood that there are a first die and a second die: the pair outcome has the score of the first die as first component and the score of the second die as second component.*
- *Sample space:*

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 .$$

Example.

- *Experiment: 100m run at the Olympic Games with* 8 *runners denoted by* $1, 2, 3, 4, 5, 6, 7, 8$ *(their starting lane).*
- *Outcome: order of arrival.*
- *Sample space:*

$$\Omega = \textit{set of the permutations of } 1, 2, 3, 4, 5, 6, 7, 8.$$

- In the previous four examples, the sample space was finite. Here are other two examples where it is infinite. In the first example Ω is infinite but countable and in the second example Ω is infinite but uncountable.

  Example with Ω infinite but countable.

  - *Experiment: flip a coin until Head appears.*
  - *Outcome: the sequence of the faces shown.*
  - *Sample space:*

$$\Omega = \left\{ \underbrace{(T, T, \ldots, T, H)}_{n \text{ components}} : n \in \{1, 2, 3, \ldots\} \right\} \cup \{(T, T, T, \ldots)\}$$

Example with Ω infinite but uncountable

- *Experiment: next falling of a meteor on the Earth.*
- *Outcome: geographic coordinates (longitude and latitude) of the impact point.*
- *Sample space:*

$$\Omega = (-\pi, \pi] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

- Observe that for the same experiment we could consider different outcomes.

  For example:

  - *in the experiment of the two dice, instead of the pair of scores, we can consider the sum of the scores as the outcome;*
  - *in the 100m run experiment, instead of the order of arrival, we can consider only the winner as the outcome;*
  - *in the falling meteor experiment, we can consider the geographic coordinates of the impact point plus the diameter of the crater as the outcome.*

  Exercise. Describe the new sample spaces in these three examples.

Events

- An **event** relative to the experiment is a statement that says something about the outcome.

  Example. In the experiment of flipping two coins, an event is the statement

  "the faces shown by the two coins are equal".

An event is identified with a subset of Ω: precisely, the statement $A(\omega)$, that says something about the outcome $\omega \in \Omega$, is identified with the subset of Ω

$$\{\omega \in \Omega : A(\omega)\}.$$

In the experiment of the two coins, the statement

$A(\omega) =$ *"the faces shown by the two coins are equal"* $=$ "$\omega_1 = \omega_2$ "

is identified with

$$\{\omega \in \Omega : A(\omega)\} = \{\omega \in \Omega : \omega_1 = \omega_2\} = \{(H, H), (T, T)\}.$$

Here are other examples of events in the previous experiments.

In the experiment of the first birth in Italy in 2019, events are

$$\text{"the newborn is a boy"} = \{boy\}$$
$$\text{"the newborn is a girl"} = \{girl\}.$$

In the experiment of the two coins, an event is

$$\text{"the first coin shows Head"} = \{(H, H), (H, T)\}.$$

In the experiment of the dice, an event is

$$\text{"the sum of the scores is } 7\text{"}$$
$$= \left\{(i, j) \in \{1, 2, 3, 4, 5, 6\}^2 : i + j = 7\right\}$$
$$= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

In the experiment of the 100m run, an event is

"the runner in lane 4 *wins the run"* $= \{\omega = \omega_1 \ldots \omega_8 \in \Omega : \omega_1 = 4\}$.

In the experiment of flip a coin until Head appears, an event is

"Head appears within four flips" $= \{H, (T, H), (T, T, H), (T, T, T, H)\}$

In the experiment of the meteor falling, events are

"the meteor falls in the austral hemisphere" $= (-\pi, \pi] \times [-\frac{\pi}{2}, 0]$

"the meteor falls in Italy"

$= \{(\lambda, \phi) \in (-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] : (\lambda, \phi)$ *is in Italy*$\}$.

- There are two particular events:
  - the event Ω, namely a statement $A(\omega)$ true for any outcome $\omega \in \Omega$;
  - the event $\emptyset$, namely a statement $A(\omega)$ false for any outcome $\omega \in \Omega$.

In the example of the two dice:

  - *the event "the score of the second die is less than 7" is Ω;*
  - *the event "the sum of scores is 1" is $\emptyset$.*

- It is known that statements can be combined by logical operations: starting with the statements $A(\omega)$ and $B(\omega)$, one can obtain the new statements

$$\text{not } A(\omega), \quad A(\omega) \text{ and } B(\omega), \quad A(\omega) \text{ or } B(\omega).$$

When we interpret the events as subsets of Ω, these logical operations between events correspond to set theory operations:

Logical operation	Set theory operation
not $A(\omega)$	$A^c = \Omega \setminus A$
$A(\omega)$ and $B(\omega)$	$A \cap B$
$A(\omega)$ or $B(\omega)$	$A \cup B$

Exercise. In the dice experiment, consider the events

$$A = \text{"both scores are odd"}, \quad B = \text{"the sum of the scores is 6"}.$$

Find "not A", "A and B" and "not (A or B)" as subsets of Ω, by listing their elements.

- We say that the events $A(\omega)$ and $B(\omega)$ are **disjoint** if

$$\forall \omega \in \Omega : \text{"}A(\omega) \text{ and } B(\omega)\text{" is false}$$

and this corresponds to

$$A \cap B = \emptyset$$

when the events are interpreted as subsets of Ω.

Examples.

In the dice experiment, the events

"Both scores are odd" and "The sum of scores is odd"

are disjoint and this corresponds to

$$\{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\}$$
$$\cap\{(1,2),(1,4),(1,6),(2,1),(2,3),(2,5),(3,2),(3,4),(3,6),$$
$$(4,1),(4,3),(4,5),(5,2),(5,4),(5,6),(6,1),(6,3),(6,5)\}$$
$$=\emptyset.$$

In the falling meteor experiment, the events

"The meteor falls in Italy" and "The meteor falls in the austral hemisphere"

are disjoint and this corresponds to

$$\left\{(\lambda,\phi)\in(-\pi,\pi]\times\left[-\frac{\pi}{2},\frac{\pi}{2}\right]:(\lambda,\phi)\ \textit{is in Italy}\right\}\cap(-\pi,\pi]\times\left[-\frac{\pi}{2},0\right]=\emptyset.$$

- The relation of implication between the events $A(\omega)$ and $B(\omega)$, i.e.

$$\forall \omega \in \Omega : A(\omega) \Rightarrow B(\omega)$$

corresponds to the relation of inclusion

$$A \subseteq B$$

when the events are interpreted as subsets of Ω.

Examples.

In the dice experiment, the implication

"Both scores are odd" $\Rightarrow$ *"The sum of scores is even"*

corresponds to the inclusion

$$\{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\}$$
$$\subseteq \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),$$
$$(4,2),(4,4),(4,6),(5,1),(5,3),(5,5),(6,2),(6,4),(6,6)\}.$$

In the falling meteor experiment, the implication

"The meteor falls in Italy" $\Rightarrow$ *"The meteor falls in the boreal hemisphere"*

corresponds to the inclusion

$$\left\{(\lambda,\phi) \in (-\pi,\pi] \times \left[-\frac{\pi}{2},\frac{\pi}{2}\right] : (\lambda,\phi) \text{ is in Italy}\right\} \subseteq (-\pi,\pi] \times \left[0,\frac{\pi}{2}\right].$$

- Exercise. In the four examples with Ω finite, how many events are there?

  Exercise. Let $A(\omega)$ and $B(\omega)$ events. Show that

  $$\forall \omega \in \Omega : \text{"} A(\omega) \ \text{ and not } B(\omega) \text{" is false}$$

  is equivalent to $A \subseteq B$, when the events are interpretated as subsets of Ω.

Probability

- Consider an experiment with sample space Ω. First, we suppose that Ω is finite.

  A **measure of probability** for the experiment associates at each event relative to the experiment a number in $[0, 1]$ that measures the possibility that the event occurs: the larger the number, the larger the possibility ; 0 denotes that the event surely does not occur and 1 denotes that the event surely does occur.

Formally, a measure of probability for the experiment is a function

$$\mathbb{P} : \text{set of the subsets of } \Omega = \text{set of the events} \rightarrow \mathbb{R}$$

satisfying the following three properties:

1) **positivity**: for any event A, $\mathbb{P}(A) \geq 0$;
2) **normalization**: $\mathbb{P}(\Omega) = 1$;
3) **additivity**: for any events A, B that are disjoint, i.e. $A \cap B = \emptyset$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Given a measure of probability $\mathbb{P}$ for the experiment and an event A, the number $\mathbb{P}(A)$ is called the **probability** of A.

The additivity 3) for two events implies the additivity for an arbitrary number of events:

- for any events $A_1, A_2, \ldots, A_n$ that are disjoint, i.e.

  $$A_i \cap A_j = \emptyset \text{ for } i, j \in \{1, 2, \ldots, n\} \text{ with } i \neq j,$$

  we have

  $$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n).$$

In fact

$$\begin{aligned}
&\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \\
&= \mathbb{P}(A_1 \cup (A_2 \cup \cdots \cup A_n)) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2 \cup \cdots \cup A_n) \\
&\quad \text{by 3): } A_1 \cap (A_2 \cup \cdots \cup A_n) = A_1 \cap A_2 \cup \cdots \cup A_1 \cap A_n = \emptyset \cup \cdots \cup \emptyset = \emptyset \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2 \cup (A_3 \cup \cdots \cup A_n)) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3 \cup \cdots \cup A_n) \\
&\quad \text{by 3): } A_2 \cap (A_3 \cup \cdots \cup A_n) = A_2 \cap A_3 \cup \cdots \cup A_2 \cap A_n = \emptyset \cup \cdots \cup \emptyset = \emptyset \\
&= \ldots \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n).
\end{aligned}$$

- Assume now the general situation where Ω is an arbitrary set, which can be finite, infinite countable or infinite uncountable.

  In this general situation we have to replace, in the above definition of measure of probability, the additivity property 3) with

  - for any sequence of events $A_1, A_2, A_3, \ldots$ that are disjoint, i.e.

    $$A_i \cap A_j = \emptyset \text{ for } i, j \in \{1, 2, 3, \ldots\} \text{ with } i \neq j$$

    we have

    $$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

  Unlike the case with Ω finite, where there are only a finite number of events (subsets of Ω) and so the previous additivity property for an arbitrary finite number of events is sufficient, in the case of Ω infinite there are an infinite number of events and so a new additivity property for an infinite number of events is required.

- When Ω is infinite uncountable another difficulty appears.

  Consider the experiment of the falling meteor, where the infinite uncountable sample space of the geographic coordinates

  $$\Omega = (-\pi, \pi] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

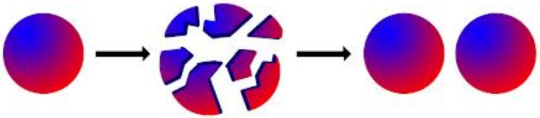
  is the surface S of the Earth, considered spherical.

  It is reasonable to assign the following measure of probability: for any event B, i.e. for any B subset of S,

  $$\mathbb{P}(B) = \frac{\text{area}(B)}{\text{area}(S)}.$$

  Observe that B is the event "$\omega \in B$", i.e. "the meteor falls in B", and so $\mathbb{P}(B)$ is the probability that the meteor falls in B.

However, the **Banach-Tarski paradox** holds : S can be disassembled in a finite number of disjoint subsets $B_1, B_2, \ldots, B_n$ of S and, under suitable rotations and translations, they can be reassembled in two identical copies of S.



Since the area is invariant under rotations and translations and the area of the union of disjoint sets is the sum of the areas, we conclude that

$$\text{area}(S) = \text{area}(B_1) + \text{area}(B_2) + \cdots + \text{area}(B_n) = 2 \cdot \text{area}(S).$$

Absurd! This means that such subsets $B_1, B_2, \ldots, B_n$ of S cannot have an area and, as a consequence, a probability.

Conclusion: when Ω is infinite uncountable, we cannot pretend that every subset of Ω is an event, where for an event we mean a subset of Ω to which can be assigned a probability.

In the general case, where Ω is any set, first to assign a measure of probability, we have to introduce a suitable family of subset of Ω, called a σ-algebra of subsets of Ω.

The events are then the elements of the σ-algebra.

A **σ-algebra** of subsets of Ω is a family $\mathcal{F}$ of subsets of Ω such that

- $\Omega \in \mathcal{F}$;
- $\mathcal{F}$ is closed with respect to the complementation: for any $A \in \mathcal{F}$, we have $A^c \in \mathcal{F}$;
- $\mathcal{F}$ is closed with respect to the countable union: for any sequence $A_1, A_2, A_3, \ldots \in \mathcal{F}$ we have

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Given a $\sigma-$algebra $\mathcal{F}$ of subset of Ω, the previous properties imply these other properties:

- $\emptyset \in \mathcal{F}$.

  Proof: $\emptyset = \Omega^c$ and $\Omega \in \mathcal{F} \Rightarrow \Omega^c = \emptyset \in \mathcal{F}$.

- $\mathcal{F}$ is closed with respect to the countable intersection: for any sequence $A_1, A_2, A_3, \ldots \in \mathcal{F}$, we have

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

  Proof: by the De Morgan's laws

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c.$$

  and

$$\begin{aligned} A_1, A_2, A_3, \ldots \in \mathcal{F} \quad &\Rightarrow \quad A_1^c, A_2^c, A_3^c, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F} \\ &\Rightarrow \quad \bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{F}. \end{aligned}$$

- $\mathcal{F}$ is closed with respect to the finite union: for any finite sequence $A_1, A_2, \ldots, A_n \in \mathcal{F}$, we have

$$\bigcup_{i=1}^{n} A_i \in \mathcal{F}.$$

Proof: by setting $A_{n+1} = A_{n+2} = A_{n+3} = \cdots = \emptyset$, we have $A_1, A_2, A_3, \ldots \in \mathcal{F}$ and

$$A_1, A_2, A_3, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

- $\mathcal{F}$ is closed with respect to the finite intersection: for any finite sequence $A_1, A_2, \ldots, A_n \in \mathcal{F}$, we have

$$\bigcap_{i=1}^{n} A_i \in \mathcal{F}.$$

Proof: by setting $A_{n+1} = A_{n+2} = A_{n+3} = \cdots = \Omega$, we have $A_1, A_2, A_3, \ldots \in \mathcal{F}$ and

$$A_1, A_2, A_3, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

- $\mathcal{F}$ is closed with respect to the set difference, i.e. for any $A, B \in \mathcal{F}$, we have
$$A \setminus B = \{\omega \in A : \omega \notin B\} \in \mathcal{F}.$$

Proof: we have $A \setminus B = A \cap B^c$ and

$$A, B \in \mathcal{F} \Rightarrow A, B^c \in \mathcal{F} \Rightarrow A \setminus B = A \cap B^c \in \mathcal{F}.$$

Now, we are ready to introduce the notions of event and measure of probability in the general case, where Ω is any set.

Definition

Given an experiment of sample space Ω, we define **events** of the experiment the elements of a some prefixed $\sigma-$algebra $\mathcal{F}$ of subsets of Ω and we define a **measure of probability** for the experiment a function

$$\mathbb{P} : \mathcal{F} = \text{set of events} \to \mathbb{R}$$

satisfying the following three properties:

1) **positivity**: for any event A, we have $\mathbb{P}(A) \geq 0$;
2) **normalization**: $\mathbb{P}(\Omega) = 1$;
3) **countable additivity**: for any sequence of events $A_1, A_2, A_3, \ldots$ that are disjoint, we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Borel subsets

- We say that:
  - Ω is **discrete** if it is finite or infinite countable;
  - Ω is **continuous** if
    $$\Omega = I_1 \times I_2 \times \cdots \times I_d \subseteq \mathbb{R}^d,$$
    where $I_1, I_2, \ldots, I_d$ are intervals of $\mathbb{R}$ not reduced to points.

Note that if Ω is continuous, then it is infinite uncountable.

When Ω is discrete, we consider as events all the subset of Ω, i.e. we use the $\sigma-$algebra

$$\mathcal{F} = \text{family of all subsets of } \Omega.$$

When Ω is continuous, we use the $\sigma-$algebra

$$\mathcal{F} = \text{family of the Borel subsets of } \Omega.$$

- The **Borel subsets** of $\Omega = I_1 \times I_2 \times \cdots \times I_d$ are the subsets of Ω that can be obtained starting from the **closed boxes**

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d],$$

where $[a_k, b_k] \subseteq I_k$, $k \in \{1, \ldots d\}$, by repeated applications of the operations of complementation

$$A \mapsto A^c$$

and countable union

$$A_1, A_2, A_3, \ldots \mapsto \bigcup_{i=1}^{\infty} A_i.$$

In other words, if A is a Borel subset, then A^c is a Borel subset by definition; if $A_1, A_2, A_3, \ldots$ are Borel subsets, then $\bigcup_{i=1}^{\infty} A_i$ is a Borel subset by definition.

The family of the Borel subsets is clearly a $\sigma-$algebra of subsets of Ω.

In fact:

- The family contains $\emptyset$ and so it contains also $\Omega = \emptyset^c$, in fact $\emptyset$ is a particular closed box: if $a_k > b_k$ for some $k \in \{1, \ldots, d\}$, then $[a_k, b_k] = \emptyset$ and so

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d] = \emptyset.$$

- by definition the family is closed with respect to complementation and countable union.

Starting from Borel subsets, we can obtain new Borel subsets, not only by complementation and countable union, but also by countable intersection, finite union, finite intersection and set difference.

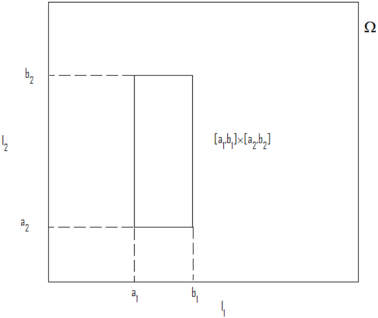
In fact, these last four operations can be reproduced by using complementation and countable union:

$$A_1, A_2, A_3, \ldots \mapsto \bigcap_{i=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c$$

$$A_1, A_2, \ldots A_n, A_{n+1} = \emptyset, A_{n+2} = \emptyset \ldots \mapsto \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i$$

$$A_1, A_2, \ldots A_n, A_{n+1} = \Omega, A_{n+2} = \Omega, \ldots \mapsto \bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{\infty} A_i$$

$$A, B \mapsto A \setminus B = A \cap B^c.$$



- Consider the bidimensional case $d = 2$. We have $\Omega = I_1 \times I_2$ and the closed boxes $[a_1, b_1] \times [a_2, b_2]$ with $[a_1, b_1] \subseteq I_1$ and $[a_2, b_2] \subseteq I_2$.

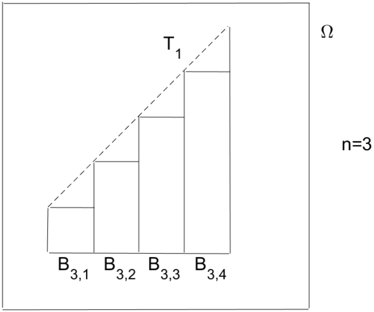
  Particular closed boxes are:

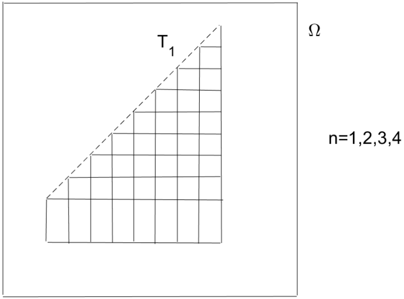
  - the empty set: for $a_1 > b_1$ or $a_2 > b_2$;
  - points: for $a_1 = b_1$ and $a_2 = b_2$;
  - horizontal or vertical closed segments: for $a_2 = b_2$ or $a_1 = b_1$.

Horizontal or vertical open segments are Borel subsets: they are obtained by subtracting to the horizontal or vertical closed segments (that are Borel subsets) the end points (that are Borel subsets).

A closed box (that is a Borel subset) minus some of the vertex points (that are Borel subsets) or some of the horizontal or vertical open segments sides of the box (that are Borel subsets) is a Borel subset.

For example, an open box $(a_1, b_1) \times (a_2 \times b_2)$ is a Borel subset, since it is the closed box $[a_1, b_1] \times [a_2 \times b_2]$ minus the four vertex points and the four open segment sides.



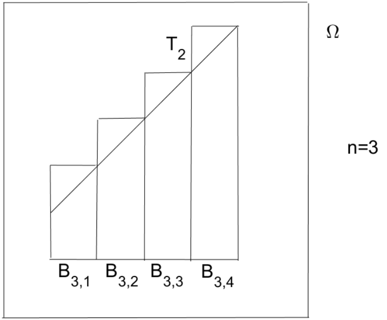


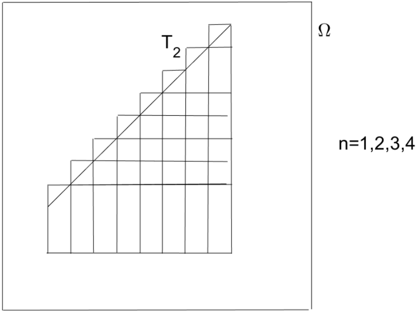
The trapezium T_1, that does not include the slanting upper border, is a Borel subset: we have

$$T_1 = \bigcup_{n=1}^{\infty} (B_{n,1} \cup B_{n,2} \cup \cdots \cup B_{n,2^{n-1}}),$$

where $B_{n,1}, B_{n,2}, \ldots, B_{n,2^{n-1}}$ are closed boxes, whose basis length is $\frac{b-a}{2^{n-1}}$, without the vertex points on the slanting upper border.

Exercise. If the vertex points on the slanting upper border are not excluded, do we obtain T_1 with the border as the countable union?





The trapezium T_2, that includes the slanting upper border, is a Borel subset: we have

$$T_2 = \bigcap_{n=1}^{\infty} (B_{n,1} \cup B_{n,2} \cup \cdots \cup B_{n,2^{n-1}}),$$

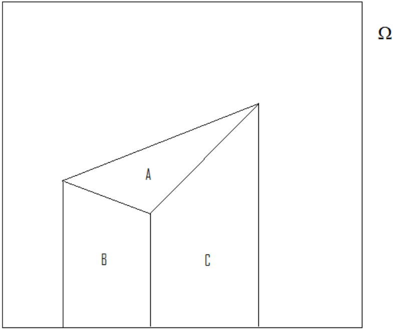
where $B_{n,1}, B_{n,2}, \ldots, B_{n,2^{n-1}}$ are closed boxes, whose basis length is $\frac{b-a}{2^{n-1}}$.

The slanting segment $T_2 \setminus T_1$ is a Borel subset. So, any segment, horizontal or vertical or slanting, is a Borel subset.

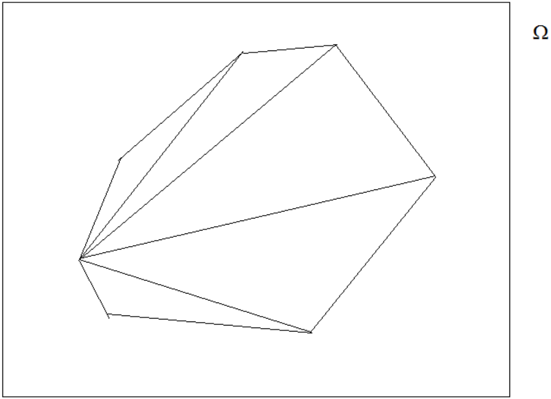
Triangles are Borel subsets:

$$A = (A \cup B \cup C) \setminus (B \cup C)$$

where $A \cup B \cup C$, B and C are trapeziums.

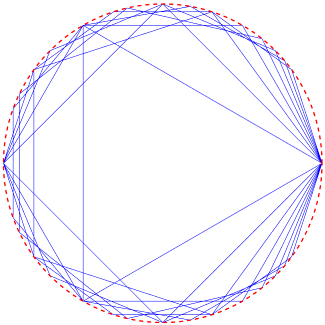


Polygons are Borel subsets: they are finite unions of triangles.



Curved geometric figures are Borel subsets: they are countable union of polygons or countable intersection of polygons.

For example, an open circle is the countable union of open regular polygons inscribed on it.



Similarly, a closed circle is the countable intersection of closed regular polygons circumscribed to it.

- In the general $d-$dimensional case, the Borel subsets of $\Omega = I_1 \times I_2 \times \cdots \times I_d$ are the subsets of Ω to which we can assign a measure:
  - for $d = 1$, measure means length;
  - for $d = 2$, measure means area;
  - for $d = 3$, measure means volume.
  - for a general d, the measure $\mathrm{measure}(A)$ of a Borel subset A of Ω, which is a $d-$dimensional area or volume of A, is given by

$$\mathrm{measure}(A) = \int_{\boldsymbol{x}\in A} d\boldsymbol{x}.$$

More about this later.

Interpretation of the probability

- In the previous definition of a measure of probability for an experiment, we have introduced the concept of probability in an axiomatic manner, giving the properties to be satisfied by what is called a measure of probability.

  But, what is the real meaning of the concept of probability?

  There are two interpretations of the probability:

  - the frequentist interpretation;
  - the bayesian interpretation.

  The adjective "bayesian" comes from the English statistician, philosopher and Presbyterian minister Thomas Bayes (1701-1761).

- In the **frequentist interpretation of the probability**, the probability $\mathbb{P}(A)$ of an event A is defined in the following manner.

  Imagine to repeat the experiment a very large number n of times, all repetitions of the experiment carried out in the same conditions.

  Let $\omega_1^{\text{obs}}, \omega_2^{\text{obs}}, \ldots, \omega_n^{\text{obs}}$ be the observed outcomes in these n ripetitions. We set,

$$\begin{aligned}\mathbb{P}(A) &= \textbf{Long Time Relative Frequency of A} \\ &:= \frac{\left|\left\{i \in \{1, 2, \ldots, n\} : A\left(\omega_i^{\text{obs}}\right)\right\}\right|}{n} \\ &= \frac{\text{number of times that } A \text{ occurs}}{n}.\end{aligned}$$

To avoid the problem of how large has to be n, we can define

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{\text{number of times that } A \text{ occurs}}{n},$$

but now we have to postulate the existence of such a limit.

Exercise. Verify that this definition of probability satisfies the three properties of positivity, normalization and additivity in the definition of a measure di probability.

Example: experiment of flipping a coin with $\Omega = \{H, T\}$. *In the frequentist interpretation, say that the event*

$$\text{"H appears"} = \{H\}$$

has probability $\frac{1}{2}$ *means:*

- *if we repeat n times with n large the experiment of flipping that coin, the relative frequency of the occurrences of H is close to* $\frac{1}{2}$ *and tends to* $\frac{1}{2}$ *as* $n \to \infty$.

- The frequentist interpretation of the probability can be successfully applied to the experiments of flipping a coin or rolling a die, but it is quite difficult to apply it at the Olympic Games 100m run, since it is a unique situation that cannot be repeated in the same conditions (as any sport situation).

  In this situation the bayesian interpretation works better.

  In the **bayesian interpretation of the probability**, the probability $\mathbb{P}(A)$ of an event A is interpreted as a "degree of belief" in the statement A, rather than a Long Time Relative Frequency of A.

  Unlike the frequentist interpretation, the bayesian interpretation does not define what is a probability, but it tries to give a meaning to the probability introduced in the axiomatic manner.

Example: experiment of flipping a coin with $\Omega = \{H, T\}$*. In the bayesian interpretation, say that the event*

$$A = \text{"Head appears"} = \{H\}$$

has probability $\frac{1}{2}$ *means:*

> *"The degree of belief" that H occurs is the same as the "degree of belief" that T occurs.*

In the bayesian intepretation, the probability "degree of belief" $\mathbb{P}(A) = \frac{m}{n}$ of an event A, where m and n are positive integer with $m \leq n$, has the following meaning: to say that A occurs is like to say that a red ball is picked when a ball is randomly selected from an urn with n balls, where m are red and $m - n$ are black.

Implicit in the bayesian interpretation of the probability is the fact that the measure of probability depends on the available information.

Example: in the experiment of the 100m run at the Olympic Games:

- *if one does not know who are the runners, she/he can assign "degree of belief"* $\frac{1}{8}$ *to the event "the runner at lane* 4 *wins";*
- *but, she/he will change this "degree of belief" if it is known that the runner at lane 4 is Usain Bolt.*

In the bayesian interpretation, the measure of probability for the experiment is subjective: each person assigns own probabilities to the events, probabilities that are based on own information as well as personal convictions.

On the other hand, in the frequentist interpretation it is objective: all people assign the same probabilities, probabilities that are based on repetitions of the experiment.

- Exercise. In the experiment of rolling a die with $\Omega = \{1, \dots, 6\}$, what does it mean in the two interpretations say that the events "the score is 1", $\dots$, "the score is 6" have all probability $\frac{1}{6}$?

  Exercise. In the following cases, say which interpretation of the probability is more appropriate (explain the meaning of the statements to someone that does not know probability).

  - The probability to select a jolly from a normal deck of playcards is $\frac{2}{54}$.
  - The p. that Juventus will win the 2018-2019 Champions League is 50%.
  - The p. that the patient will recover from the disease is 80%.
  - The p. that tomorrow Sap's stocks will increase their value over 5% is 90%.
  - The p. that next car passing along this street is red is 5%.
  - The p. that tomorrow will rain is 90%.
  - The p. that the International Space Station will have a malfunction within one month is 10%.
  - The p. that the life time of the lamp will be less then one year is 10%.
  - The p. that the defendant is guilty is 99%.

Properties of the probability

- Let $\mathbb{P}$ be a measure of probability for an experiment with sample space Ω and $\sigma-$algebra of the events $\mathcal{F}$.

  Starting from the three properties in the definition of a measure of probability, we can prove these other nine properties of $\mathbb{P}$.

- 4) $\mathbb{P}(\emptyset) = 0$.

Proof of 4). Let $A_1 = \emptyset, A_2 = \emptyset, A_3 = \emptyset, \ldots$, so that $A_1, A_2, A_3, \ldots$ are disjoint and $\bigcup_{i=1}^{\infty} A_i = \emptyset$. By the countable additivity, we have

$$\mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset).$$

The equation

$$\mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$$

holds only if $\mathbb{P}(\emptyset) = 0$. Exercise. Why? What happens if $\mathbb{P}(\emptyset) > 0$?

- 5) **Finite additivity**: for any events $A_1, A_2, \ldots, A_n$ that are disjoint, we have

$$\mathbb{P}\left(A_1 \cup A_2 \cup \cdots \cup A_n\right) = \mathbb{P}\left(A_1\right) + \mathbb{P}\left(A_2\right) + \cdots + \mathbb{P}\left(A_n\right).$$

Proof of 5). Let $A_{n+1} = \emptyset, A_{n+2} = \emptyset, A_{n+3} = \emptyset, \ldots$, so that $A_1, A_2, A_3, \ldots$ are disjoint and

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots \cup A_n.$$

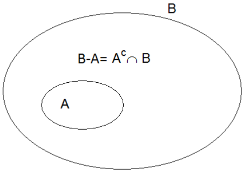
By the countable additivity and 4), we have

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right) \\ &= \sum_{i=1}^{n} \mathbb{P}\left(A_i\right) + \sum_{i=n+1}^{\infty} \underbrace{\mathbb{P}\left(\emptyset\right)}_{=0} \\ &= \mathbb{P}\left(A_1\right) + \mathbb{P}\left(A_2\right) + \cdots + \mathbb{P}\left(A_n\right). \end{aligned}$$

Exercise. Prove that, in case of Ω finite, the countable additivity and the finite additivity are equivalent. As a consequence we have that, in case of Ω finite, the additivity (finite additivity with two events) is equivalent to the countable additivity. Why?

- 6) **Difference property**: for any events A and B with $A \subseteq B$, we have
$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

Proof of 6). A and $B \setminus A = A^c \cap B$ are disjoint with $A \cup (B \setminus A) = B$.



By the finite additivity 5), we have $\mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(B)$.

- 7) For any event A, we have

$$\mathbb{P}\left(A^c\right) = 1 - \mathbb{P}\left(A\right).$$

Proof of 7). Take $B = \Omega$ in the difference property 6). We have $B \setminus A = \Omega \setminus A = A^c$ and, by normalization, $\mathbb{P}(B) = \mathbb{P}(\Omega) = 1$ and so

$$\mathbb{P}(A^c) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A).$$

- 8) For any event A, we have $\mathbb{P}(A) \leq 1$.

Proof of 8). By positivity we have $\mathbb{P}(A^c) \geq 0$ and so, by 7),

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \leq 1.$$

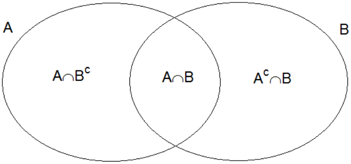
- 9) For any events A and B with $A \subseteq B$, we have $\mathbb{P}(A) \leq \mathbb{P}(B)$.

  Proof of 9). By positivity we have $\mathbb{P}(B \setminus A) \geq 0$ and so, by the difference property 6),

$$0 \leq \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

- 10) **Inclusion-exclusion principle**: for any events A and B, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$



Proof of 10). $A \cap B^c$, $A \cap B$ and $A^c \cap B$ are disjoint. We have

$$(A \cap B^c) \cup (A \cap B) = A \text{ and then } \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) = \mathbb{P}(A)$$

$$(A \cap B) \cup (A^c \cap B) = B \text{ and then } \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \mathbb{P}(B)$$

$$(A \cap B^c) \cup (A \cap B) \cup (A \cap B^c) = A \cup B$$

$$\text{and then } \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \mathbb{P}(A \cup B).$$

by using the finite additivity 5). Thus

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).
\end{aligned}$$

- 11) **Upper monotone convergence**: for any sequence $A_1, A_2, A_3, \dots$ of events such that

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \quad (\text{ and then } \mathbb{P}(A_1) \leq \mathbb{P}(A_2) \leq \mathbb{P}(A_3) \leq \dots)$$

we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i) = \sup_{i\in\{1,2,3,\dots\}} \mathbb{P}(A_i).$$

12) **Lower monotone convergence**: For any sequence $A_1, A_2, A_3, \dots$ of events such that

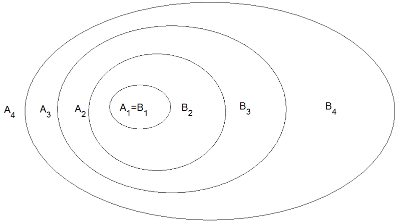
$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots \quad (\text{ and then } \mathbb{P}(A_1) \geq \mathbb{P}(A_2) \geq \mathbb{P}(A_3) \geq \dots)$$

we have

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i) = \inf_{i\in\{1,2,3,\dots\}} \mathbb{P}(A_i).$$

Proof of 11). Define

$$B_1 = A_1 \text{ and } B_{k+1} = A_{k+1} \setminus \bigcup_{i=1}^{k} A_i = A_{k+1} \setminus A_k,\ k \in \{1, 2, 3, \ldots\}.$$



The events $B_1, B_2, B_3, \ldots$ are disjoint with

$$\bigcup_{i=1}^{k} B_i = A_k,\ k \in \{1, 2, 3, \ldots\}, \text{ and } \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

Thus, by countable and finite additivity we have

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{k\to\infty} \sum_{i=1}^{k} \mathbb{P}(B_i) = \lim_{k\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{k} B_i\right) \\ &= \lim_{k\to\infty} \mathbb{P}(A_k).\end{aligned}$$

Proof of 12). We have, by 7),

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \mathbb{P}\left(\left(\bigcap_{i=1}^{\infty} A_i\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i^c\right).$$

By $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, we obtain

$$A_1^c \subseteq A_2^c \subseteq A_3^c \subseteq \ldots.$$

Thus, by the upper monotone convergence 11) and 7), we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i^c\right) = \lim_{i\to\infty} \mathbb{P}(A_i^c) = \lim_{i\to\infty} (1 - \mathbb{P}(A_i)) = 1 - \lim_{i\to\infty} \mathbb{P}(A_i)$$

and so

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i^c\right) = 1 - \left(1 - \lim_{i\to\infty} \mathbb{P}(A_i)\right) = \lim_{i\to\infty} \mathbb{P}(A_i).$$

Probability as area

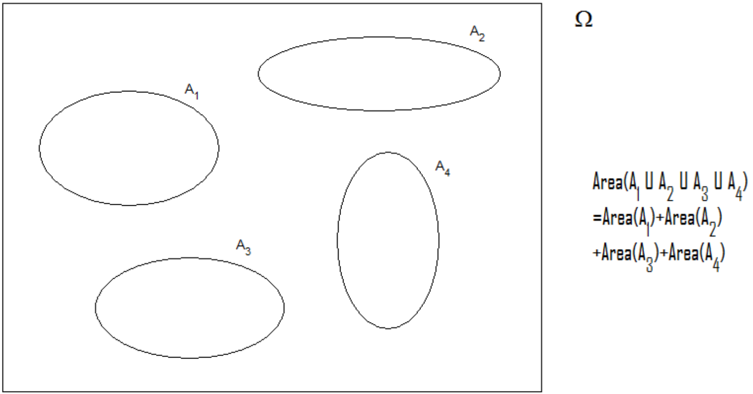
- Previously, we have seen that in case of a continuous sample space $\Omega \subseteq \mathbb{R}^d$, the events are the Borel subsets: this means that the subset of Ω to which we assign a probability are exactly those to which we can assign a measure, i.e. a $d-$dimensional area or volume.

  Indeed, for a general sample space, the probability of an event can be thought as an "area" (or "volume") of the event, normalized in order to have "area" of the sample space equal to 1.

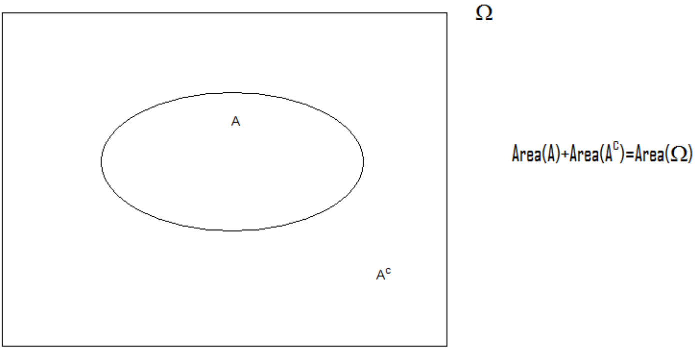
  In fact, a measure of probability has the same properties as the area (or the volume) has.

  Here are some examples of this.

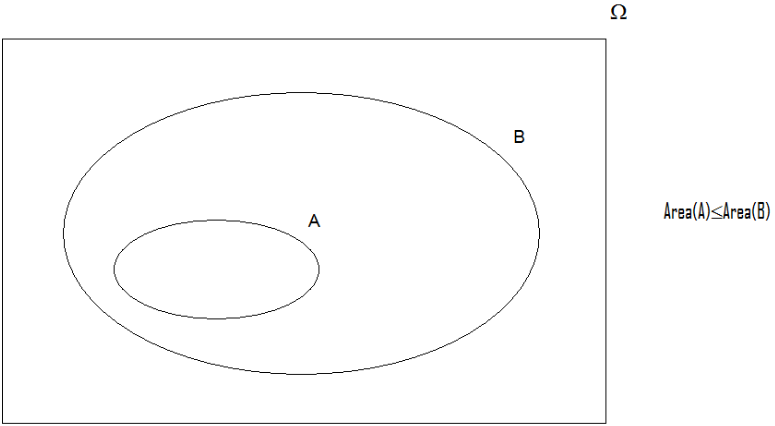
$$A_1, A_2, A_3, A_4 \text{ disjoint } \Rightarrow \mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \mathbb{P}(A_4).$$



$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega).$$



$$A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B).$$



- Indeed, the abstraction of the notion of "area" or "volume" is the notion of "measure".

  Given a set Ω and a $\sigma-$algebra $\mathcal{F}$ of subsets of Ω, we define a **measure** for the subsets in $\mathcal{F}$ a function

  $$m : \mathcal{F} \to \mathbb{R}$$

  satisfying the properties of positivity and countable additivity.

  So, a measure of probability is a measure with the normalization $m(\Omega) = 1$.

  A measure satisfies all the previous properties of probability from 4) to 12): in 7) and 8) substitute 1 with $\mathbb{P}(\Omega)$.

Probability for experiments with discrete sample space

- Consider an experiment with a discrete sample space Ω.

  The events

  $$A(\omega) = \text{"the outcome } x \text{ is obtained"} = \text{"}\omega = x\text{"} = \{x\}\,,\ x \in \Omega,$$

  are called **elementary events.**

- In case of a discrete sample space, we can generate a measure of probability in the following way.

  Given a function $p : \Omega \to [0, +\infty)$ such that

  $$\sum_{x \in \Omega} p(x) = 1,$$

  we assign the probabilities of the elementary events:

  $$\mathbb{P}(x) = p(x),\ x \in \Omega,$$

  where $\mathbb{P}(x)$ stands for $\mathbb{P}(\{x\})$.

  Once we have assigned a probability to each elementary event, for each event $A \subseteq \Omega$, we have, by finite or countable additivity,

  $$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{x \in A} \{x\}\right) = \sum_{x \in A} \mathbb{P}(x) = \sum_{x \in A} p(x).$$

Exercise. Above we have proved that if $\mathbb{P}$ is a measure of probability such that

$$\mathbb{P}(x) = p(x), \; x \in A, \tag{1}$$

then, for any event $A \subseteq \Omega$,

$$\mathbb{P}(A) = \sum_{x \in A} p(x). \tag{2}$$

In other words, we have proved that, if there exists, there is a unique measure of probability $\mathbb{P}$ such that (1) holds and it is given by (2).
To complete our discussion, prove that the function
$\mathbb{P}$: family of all subsets of $\Omega \to \mathbb{R}$ given by

$$A \mapsto \mathbb{P}(A) = \sum_{x \in A} p(x), \; A \subseteq \Omega,$$

is a measure of probability. This shows that indeed there exists a measure of probability such that (1) holds.

- In case of Ω finite and

$$p(x) = \frac{1}{|\Omega|}, \; x \in \Omega,$$

i.e. all the outcomes have the same probability to be obtained, we have

$$\mathbb{P}(A) = \sum_{x \in A} p(x) = \sum_{x \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}, \quad A \subseteq \Omega.$$

This is what is called **classical probability**: the probability of an event is simply the ratio

$$\frac{\text{number of favorable outcomes}}{\text{number of total outcomes}}$$

where for a favorable outcome we mean an outcome for which the event is true.

- *Now, we present measures of probability for the previous experiments with discrete sample space.*

  In the experiment of the first birth in 2019 in Italy, it is reasonable to assume

$$\mathbb{P}(girl) = \mathbb{P}(boy) = \frac{1}{|\Omega|} = \frac{1}{2}.$$

  In the experiment of flipping two coins, it is reasonable to assume

$$\mathbb{P}((H,H)) = \mathbb{P}((H,T)) = \mathbb{P}((T,H)) = \mathbb{P}((T,T)) = \frac{1}{|\Omega|} = \frac{1}{4},$$

  for two regular coins.

  In the experiment of rolling two dice, it is reasonable to assume

$$\mathbb{P}((i,j)) = \frac{1}{|\Omega|} = \frac{1}{36}, \ (i,j) \in \Omega = \{1,2,3,4,5,6\}^2,$$

  for two regular dice.

In the experiment of the 100m run at the Olympic Games, it is reasonable assume

$$\mathbb{P}(x) = \frac{1}{|\Omega|} = \frac{1}{8!}, \; x \text{ permutation of } 1,2,3,4,5,6,7,8,$$

if one does not know who are the runners, but it is unreasonable if one knows that Usain Bolt is one of the runners.

In the experiment of flipping a regular coin until Head appears, it is reasonable to assume

$$\mathbb{P}\left(\left(\underbrace{T,\ldots,T}_{n-1 \text{ times}},H\right)\right) = \frac{1}{2^n}, \; n \in \{1,2,3,\ldots\},$$

$$\mathbb{P}((T,T,T,\ldots)) = 0.$$

- Exercise. Prove that in case of Ω discrete and infinite, the probabilities $p(x)$, $x \in \Omega$, of the elementary events cannot be all equal.

  Exercise. In the experiment of the 100m run at the Olympic Games, assume

  $$\mathbb{P}(x) = \frac{1}{|\Omega|} = \frac{1}{8!}, \; x \text{ permutation of } 1, 2, 3, 4, 5, 6, 7, 8.$$

  What is the probability of the event "the runner at lane 4 wins"?

  Exercise. Prove that in the experiment of flipping a regular coin until Head appears, we have

  $$\sum_{x \in \Omega} p(x) = 1.$$

  Moreover, explain why the probabilities of the elementary events given for this experiment are reasonable. Finally, compute the probability that Head appears after an even number of flips.

Some problems of probability

- Now, we present some problems of computation of probabilities for experiments such that:
  - the sample space is finite;
  - all the outcomes have the same probability to be obtained and so we are in the context of classical probability.

- *Problem 1. Assume that the engineering freshmen at the University of Trieste are 420: 144 are smokers and 276 are not. If a freshman is randomly selected, what is the probability that she/he is a smoker?*

  Experiment: selection of an engineering freshman.

  Outcome: the selected engineering freshman.

  Sample space: set of the engineering freshmen.

  Measure of probability: the fact that the engineering freshman is "randomly selected" means that all the elementary events have the same probability

$$\frac{1}{|\Omega|} = \frac{1}{420}.$$

Therefore, the event

$$A = \text{"the selected freshman is a smoker"}$$
$$= \{\omega \in \Omega : \omega \text{ is a smoker}\}$$

has probability

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{144}{420} = 0.3429 = 34.29\%.$$

- *Problem 2. Suppose that two regular dice are rolled. What is the probability that the sum of the scores is s, where $s \in \{2, \ldots, 12\}$?*

  Experiment: roll of the two dice.

  Outcome: the pair of the scores.

  Sample space:

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 .$$

  Measure of probability: the fact that the dice are "regular" means that all the elementary events have the same probability

$$\frac{1}{|\Omega|} = \frac{1}{36}.$$

Therefore, the event

$$\begin{aligned} A &= \text{"the sum of the scores is } s\text{"} \\ &= \{(i,j) \in \Omega : i + j = s\} \\ &= \{(i, s-i) : 1 \leq i \leq 6 \text{ and } 1 \leq s - i \leq 6\} \\ &= \{(i, s-i) : 1 \leq i \leq 6 \text{ and } s - 6 \leq i \leq s - 1\} \\ &= \{(i, s-i) : \max\{1, s-6\} \leq i \leq \min\{6, s-1\}\} \end{aligned}$$

has probability

$$\begin{aligned} \mathbb{P}(A) &= \frac{|A|}{|\Omega|} = \frac{\min\{6, s-1\} - \max\{1, s-6\} + 1}{36} \\ &= \begin{cases} \frac{s-1-1+1}{36} & \text{if } s - 1 \leq 6 \text{ (i.e. } s - 6 \leq 1\text{)} \\ \frac{6-(s-6)+1}{36} & \text{if } s - 1 > 6 \text{ (i.e. } s - 6 > 1\text{)} \end{cases} \\ &= \begin{cases} \frac{s-1}{36} & \text{if } s \leq 7 \\ \frac{13-s}{36} & \text{if } s > 7. \end{cases} \end{aligned}$$

So, we have the following table of probabilities

s	$\mathbb{P}$ (" The sum of the scores is s ")
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

where it appears that the most probable sum of scores is 7 *with probability* $\frac{1}{6}$.

Exercise. What is the probability that the sum of the scores is larger than 6*?*

- *Problem 3. Given 10 married couples, the women are grouped and the men are grouped. Then a woman and a man are randomly and independently selected. What is the probability that the selected woman and man are married to each other?*

  Experiment: selection of a woman in the group of the women and of a man in the group of men.

  Outcome: the pair $(\textit{selected woman}, \textit{selected man})$.

  Sample space:

  $$\Omega = \textit{set of the women} \times \textit{set of the men}.$$

  Measure of probability: the fact that a woman and a man are "randomly and independently selected" means that all the elementary events have the same probability

  $$\frac{1}{|\Omega|} = \frac{1}{|\textit{set of the women}| \cdot |\textit{set of the men}|} = \frac{1}{10 \cdot 10} = \frac{1}{100}.$$

Therefore, the event

$$A = \text{"the selected woman and man are married to each other"}$$
$$= \{(w, m) \in \Omega : w \text{ and } m \text{ are married}\}$$

has probability

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{10}{100} = \frac{1}{10}.$$

Exercise. A person is said mature if she/he is 50 *or more years old. Assume that there are* 4 *mature women and* 7 *mature men. What is the probability that in the selected pair the woman is mature and the man is not.*

- *Problem 4. A shop accepts, as credit cards, American Express or VISA. A total of 22 percent of customers have an American Express card, 58 percent have a VISA card and 14 percent have both. What is the probability that the next customer entering the shop, i.e. a randomly selected customer, can pay by credit card in this shop?*

  Experiment: selection of a customer.

  Outcome: the selected customer.

  Sample space: the set of customers.

  Measure of probability: all the elementary events have the same probability $\frac{1}{|\Omega|}$.

The events

$$\begin{aligned} A &= \text{"the selected customer has an Amexican Express card"} \\ &= \{\omega \in \Omega : \omega \text{ has an Amexican Express card}\} \end{aligned}$$

$$\begin{aligned} B &= \text{"the selected customer has a VISA card"} \\ &= \{\omega \in \Omega : \omega \text{ has a VISA card}\} \end{aligned}$$

$$\begin{aligned} A \cap B &= \text{"the selected customer has both cards"} \\ &= \{\omega \in \Omega : \omega \text{ has both cards}\} \end{aligned}$$

have probabilities

$$\begin{aligned} \mathbb{P}(A) &= \frac{|A|}{|\Omega|} = 22\% \\ \mathbb{P}(B) &= \frac{|B|}{|\Omega|} = 58\% \\ \mathbb{P}(A \cap B) &= \frac{|A \cap B|}{|\Omega|} = 14\%. \end{aligned}$$

Therefore, the event

$$\begin{aligned} A \cup B &= \text{"the selected customer can pay by credit card in this shop"} \\ &= \{\omega \in \Omega : \omega \text{ has an American Express card or a VISA card}\} \end{aligned}$$

has probability

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= 22\% + 58\% - 14\% = 66\%. \end{aligned}$$

Exercise. What is the probability that the selected customer has a Visa card and not an American Express card.

Exercise. In a wardrobe, the trousers have zip or buttons and some of them are jeans. Over all trousers, the percentage of jeans is $100p\%$, $p \in [0, 1]$, *and the percentage of trousers with zip is* $100q\%$, $q \in [0, 1]$. *Trousers have zip or are jeans. What is the probability that a randomly selected trousers is a jeans with zip?*

- *Problem 5. Consider the following table of earning in US in 1989*

Table 4.3 Earnings of Workers by Sex, 1989

Earnings group (in $1000)	Number		Distribution (percent)	
	Women	Men	Women	Men
<5	427,000	548,000	1.4	1.1
5–10	440,000	358,000	1.4	.7
10–15	1,274,000	889,000	4.1	1.8
15–20	1,982,000	1,454,000	6.3	2.9
20–30	6,291,000	5,081,000	20.1	10.2
30–40	6,555,000	6,386,000	20.9	12.9
40–50	5,169,000	6,648,000	16.5	13.4
50–100	8,255,000	20,984,000	26.3	42.1
>100	947,000	7,377,000	3.0	14.9
Total	31,340,000	49,678,000	100.0	100.0

Source: Department of Commerce, Bureau of the Census.

Consider the experiment where a worker is randomly selected.

What is the probability of the event

$$A = \text{" the selected worker is a woman"?}$$

We have

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{31340K}{31340K + 49678K} = 0.3868.$$

What is the probability of the event

$$B = \text{" the selected worker is a man"?}$$

We have $B = A^c$ *and*

$$\mathbb{P}(B) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - 0.3868 = 0.6132.$$

What is the probability of the event

$$C = \text{" the selected worker is a woman and she earns over 50,000 dollars" ?}$$

We have

$$\mathbb{P}(C) = \frac{|C|}{|\Omega|} = \frac{8255K + 947K}{31340K + 49678K} = 0.1136.$$

Exercise. Compute the probability of the events

$$D = \text{" the selected worker is a man and he earns over 50,000 dollars"}$$

and

$$E = \text{" the selected worker earns over 50,000 dollars" .}$$

Exercise. Suppose that the column "Numbers" in the previous table is erased, except for the totals. Compute the probability that the selected worker is in a given earning group with percentages $100p\%$ *(women) and* $100q\%$ *(men) in the column "Distribution (percent)".*

Table 4.3 Earnings of Workers by Sex, 1989

Earnings group (in \$1000)	Number		Distribution (percent)	
	Women	Men	Women	Men
<5			1.4	1.1
5–10			1.4	.7
10–15			4.1	1.8
15–20			6.3	2.9
20–30			20.1	10.2
30–40			20.9	12.9
40–50			16.5	13.4
50–100			26.3	42.1
>100			3.0	14.9
Total	31,340,000	49,678,000	100.0	100.0

Source: Department of Commerce, Bureau of the Census.

Probability for experiments with continuous sample space

- Consider an experiment with a continuous sample space

$$\Omega = I_1 \times I_2 \times \cdots \times I_d \subseteq \mathbb{R}^d,$$

where $I_1, I_2, \cdots, I_d$ are intervals of $\mathbb{R}$ not reduced to points.

In case of a continuous sample space, we can generate a measure of probability in the following way.

Given an integrable function $p : \Omega \to [0, +\infty)$ such that

$$\int_{x\in\Omega} p(x)dx = \int_{x\in I_1\times I_2\times\ldots\times I_d} p(x)dx$$
$$= \int_{I_1}\int_{I_2}\cdots\int_{I_d} p(x_1, x_2, \ldots, x_d)dx_d\cdots dx_2dx_1 = 1$$

we assign to a closed box

$$[a_1, b_1]\times[a_2, b_2]\times\cdots\times[a_d, b_d]\subseteq\Omega,$$

where $[a_k, b_k]\subseteq I_k$, $k\in\{1,\ldots,d\}$, the probability

$$\mathbb{P}\left([a_1, b_1]\times[a_2, b_2]\times\cdots\times[a_d, b_d]\right)$$

$$= \int_{x\in[a_1,b_1]\times[a_2,b_2]\times\cdots\times[a_d,b_d]} p(x)dx$$

$$= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_d}^{b_d} p(x_1, x_2, \ldots, x_d)dx_d\cdots dx_2dx_1.$$

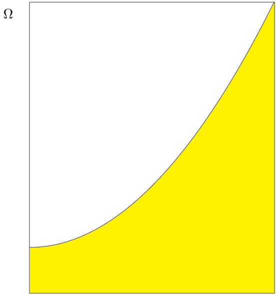
The closed boxes play the same role as the elementary events in case of a discrete sample space.

Once we have assigned the probabilities of the closed boxes, by using the properties of a measure of probability we can assign a probability to all Borel subsets in an incremental manner, i.e. in the same manner with which the Borel subsets are constructed,

In fact, we can proceed as follows.

- Given disjoint Borel subsets $A_1, A_2, \ldots, A_n$ whose probabilities are assigned, we can assign a probability to the new Borel subset $\bigcup_{i=1}^n A_i$ by the finite additivity property: $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i)$.
- Given disjoint Borel subsets $A_1, A_2, A_3, \ldots$ whose probabilities are assigned, we can assign a probability to the new Borel subset $\bigcup_{i=1}^\infty A_i$ by the countable additivity property: $\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mathbb{P}(A_i)$.
- Given Borel subsets A and B with $A \subseteq B$ whose probabilities are assigned, we can assign a probability to the new Borel subset $B \setminus A$ by the difference property: $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.
- Given Borel subsets $A_1, A_2, A_3, \ldots$ such that $A_1 \subseteq A_2 \subseteq A_3, \ldots$ whose probabilities are assigned, we can assign a probability of the new Borel subset $\bigcup_{i=1}^\infty A_i$ by the upper convergence property: $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i)$. An analogous consideration holds for the lower convergence property.

Example:

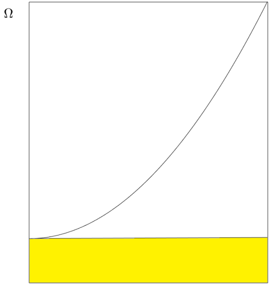


A=

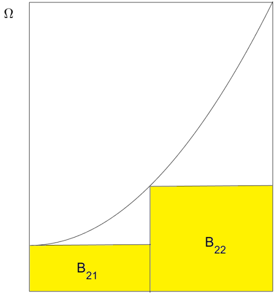
$A = \bigcup_{i=1}^{\infty} A_i$

$P(A) = \lim_{i \to \infty} P(A_i)$

$A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$

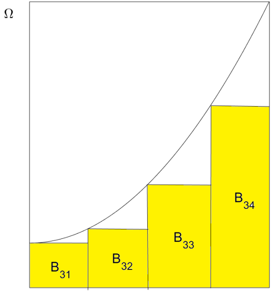


A_1=



A_2=

$P(A_2) = P(B_{21}) + P(B_{22})$



A_3=

$P(A_3) = P(B_{31}) + \ldots + P(B_{34})$

In this manner, we assign to each Borel subset A of Ω the probability

$$\mathbb{P}(A) = \int_{x \in A} p(x) dx.$$

In other words, what we have done is to define and construct the integral of p on arbitrary Borel subsets of Ω starting from the integrals of p on closed boxes

$$\int_{x \in [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]} p(x) dx$$
$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_d}^{b_d} p(x_1, x_2, \ldots, x_d) dx_d \cdots dx_2 dx_1,$$

which can be easily computed by nested one-dimensional integrals on intervals.

Exercise. Above we have proved that if $\mathbb{P}$ is a measure di probability such that

$$\mathbb{P}\left([a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]\right) = \int_{x \in [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]} p(x) dx \qquad (3)$$

for any closed box

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d] \subseteq \Omega,$$

then, for any Borel subset A of Ω,

$$\mathbb{P}(A) = \int_{x \in A} p(x) dx. \qquad (4)$$

In other words, we have proved that, if there exists, there is a unique measure of probability $\mathbb{P}$ such that (3) holds and it is given by (4).
To finish our discussion, prove that the function

$$A \text{ Borel subset of } \Omega \mapsto \mathbb{P}(A) = \int_{x \in A} p(x) dx$$

is a measure of probability. This shows that indeed there exists a measure of probability such that (3) holds.

- Observe that if a Borel subset A of Ω has dimension smaller than the dimension d of Ω, then

$$\mathbb{P}(A) = \int_{x \in A} p(x)dx = 0.$$

Then:

  - For $d = 2$, points and curves included in Ω have probability zero.
  - For $d = 3$, points, curves and surfaces in Ω have probability zero.

- In case of a continuous sample space Ω, the elementary events

$$"\omega = x" = \{x\}, x \in \Omega,$$

have probability zero since they are points.

But, we can say that the infinitesimal closed box

$$[x_1, x_1 + dx_1] \times [x_2, x_2 + dx_2] \times \cdots \times [x_d, x_d + dx_d]$$

with vertex $x = (x_1, x_2, \ldots, x_d)$ has infinitesimal probability

$$\int_{y \in [x_1, x_1+dx_1] \times [x_2, x_2+dx_2] \times \cdots \times [x_d, x_d+dx_d]} p(y) dy$$

$$= p(x) \int_{y \in [x_1, x_1+dx_1] \times [x_2, x_2+dx_2] \times \cdots \times [x_d, x_d+dx_d]} dy$$

p has constant value $p(x)$ on the infinitesimal closed box

$$= p(x) \int_{x_1}^{x_1+dx_1} \int_{x_2}^{x_2+dx_2} \cdots \int_{x_d}^{x_d+dx_d} dy_d \cdots dy_2 dy_1$$

$$= p(x) dx_d \cdots dx_2 dx_1 .$$

Therefore, for a Borel subset A of Ω, we can interpreted the probability

$$\mathbb{P}(A) = \int_{x \in A} p(x) dx$$

as the sum of the infinitesimal probabilities

$$p(x)dx = p(x)dx_d \cdots dx_2 dx_1$$

of all infinitesimal closed boxes

$$[x_1, x + dx_1] \times [x_2, x_2 + dx_2] \times \cdots \times [x_d, x_d + dx_d]$$
$$x = (x_1, x_2, \ldots, x_d) \in A,$$

which completely cover A.

Thus, similarly to the case of a discrete sample space, we can say that the probability of an event is the sum of the probabilities of the elementary events that constitute it: but for a continuous sample space this is an infinite sum of infinitesimal probabilities.

- When $I_1, I_2, \ldots, I_d$ are bounded intervals of $\mathbb{R}$, i.e. Ω is a bounded subset of $\mathbb{R}^d$, and

$$\begin{aligned} p(x) &= \frac{1}{\text{measure}(\Omega)} \\ &= \frac{1}{\text{length}(I_1) \cdot \text{length}(I_2) \cdot \ldots \cdot \text{length}(I_d)}, \quad x \in \Omega, \end{aligned}$$

we have, for each Borel subset A of Ω,

$$\begin{aligned} \mathbb{P}(A) &= \int_{x \in A} \frac{1}{\text{measure}(\Omega)} dx = \frac{1}{\text{measure}(\Omega)} \cdot \int_{x \in A} dx \\ &= \frac{\text{measure}(A)}{\text{measure}(\Omega)}. \end{aligned}$$

This is the notion of **classical probability** for a continuous sample space.

In the theoretical experiment of a point randomly selected in Ω, with Ω bounded subset of $\mathbb{R}^d$, "randomly selected" just means

$$p(x) = \frac{1}{\text{measure}(\Omega)}, \; x \in \Omega.$$

We say "theoretical experiment" because points are only abstract notions that do not exists in the real world.

Practical instances of this experiment:

- The random number generators in the computers approximate the random selection of a real number in $[0, 1]$.
- A needle is thrown on the floor. The angle of the needle with a prefixed direction is a randomly selected number in $(-\pi, \pi]$.
- A blinded player kicks a ball in an empty room. The point where the ball comes to rest after many bounces on the walls of the room is a randomly selected point on the floor.
- A fish is swimming in a pool (or a fly is flying in a room). The position of the fish at a given time is a randomly selected point in the pool.

Exercise. In the experiment of kicking the ball, assume that the room is squared and determine the probability that the ball stops in the circle inscribed in the square.

Exercise. In the experiment of the swimming fish, assume a rectangular pool of sides 50 m and 20 m and determine the probability that the fish is within 1 m from the pool's border.

Exercise. Consider a random number generator.

- What is the probability that the selected number is equal to 0.5?
- What is the probability that the selected number is between 0.2 and 0.7?
- What is the probability that the selected number has second digit (after the point) 0?
- What is the probability that the selected number has second digit k, $k \in \{0, 1, \dots, 9\}$?
- What is the probability that the selected number has l-th digit k, $l \in \{1, 2, 3, \dots\}$ and $k \in \{0, 1, \dots, 9\}$?
- Do the last point by considering the base B representation, instead of the base 10 representation.
- What is the probability that the selected number is irrational?

- In the experiment of the meteor, the geographic coordinates of the impact point are not a randomly selected point in $\Omega = (-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$.

  If this was true, then fall in $(-\pi, \pi] \times [0, 1^\circ]$ and fall in $(-\pi, \pi] \times [89^\circ, 90^\circ]$ had the same probabilities, but the first subset has a much larger area on the Earth's surface.

  The problem is that the impact point lies on the spherical Earth's surface, not on the rectangle Ω which is the Mercator projection of the Earth's surface found in atlases.

  For this experiment, we take

$$p(\lambda, \phi) = \frac{1}{4\pi} \cos \phi, \ (\lambda, \phi) \in \Omega = (-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}],$$

  and then, for a Borel subset A of Ω, we have

$$\mathbb{P}(A) = \frac{\text{Area}(A)}{\text{Area of the Earth's surface}},$$

  where $\text{Area}(A)$ is the area of A on the Earth's surface. The impact point is a randomly selected point on the Earth's surface, not on Ω.

Exercise. In the experiment of the falling meteor:

- by looking at the infinitesimal area on the Earth's surface of the infinitesimal closed box $[\lambda, \lambda + d\lambda] \times [\phi, \phi + d\phi]$ of vertex the point of geographic coordinates (λ, ϕ), find the function p with which

$$\mathbb{P}(A) = \frac{\text{Area}(A)}{\text{Area of the Earth's surface}}$$

  holds, where Area(A) is the area of the Borel subset A of Ω on the Earth's surface;

- check that $\int_{x \in \Omega} p(x) dx = 1$;
- determine the probabilities of the closed boxes;
- determine the probability that the meteor falls in Italy.

Conditional probability

- Consider an experiment and a measure of probability $\mathbb{P}$ for this experiment. Suppose that the experiment, or part of it, has been accomplished but the outcome is still unknown.

  Assume that some additional information relative to the unknown outcome becomes now available and that this information permits to say that a certain event surely occurs.

  This new knowledge modifies the measure of probability $\mathbb{P}$.

  Example: consider a person randomly selected from the population of a given town. The probability of the event

  "the selected person has got a Ferrari"

  is small but it changes if additional information permits to say that the event

  "the selected person is a tycoon"

  surely occurs.

Example: consider the experiment with the dice. The events

$$B = \text{"the sum of the scores is } 10\text{"},\ C = \text{"both scores are even"},$$
$$D = \text{"both scores are odd"}.$$

have probabilities

$$\mathbb{P}(B) = \frac{13-10}{36} = \frac{3}{36} = \frac{1}{12},\ \mathbb{P}(C) = \mathbb{P}(D) = \frac{9}{36} = \frac{1}{4}.$$

Now, assume that the event

$$\begin{aligned} A &= \text{"the score of the first die is } 4\text{"} \\ &= \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\} \end{aligned}$$

occurs, and then the possible outcomes are only those in the subset A of Ω. The previous probabilities becomes

$$\mathbb{P}_{\text{new}}(B) = \frac{1}{6},\ \mathbb{P}_{\text{new}}(C) = \frac{3}{6} = \frac{1}{2},\ \mathbb{P}_{\text{new}}(D) = 0.$$

- Now, we formalize the change of the probabilities when we known that a given event surely occurs.

Definition

Let $\mathbb{P}$ be a measure of probability for an experiment and let A be an event relative to this experiment such that $\mathbb{P}(A) \neq 0$. The **conditional measure of probability given** A is defined as

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

for any event B relative to the experiment. $\mathbb{P}(B|A)$ is said the **conditional probability of** B **given** A.

When it is known that the event A occurs, the probability $\mathbb{P}(B)$ of the event B changes to

$$\mathbb{P}_{\text{new}}(B) = \mathbb{P}(B|A).$$

In the example of the dice, by applying the definition of conditional probability we find the new probabilities already computed:

$$\mathbb{P}_{\text{new}}(B) = \mathbb{P}(B|A)$$
$$= \mathbb{P}(\textit{"the sum of the scores is } 10\textit{"}|\textit{"the score of the first die is } 4\textit{"})$$
$$= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}((4,6))}{\mathbb{P}(A)} = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6}$$

$$\mathbb{P}_{\text{new}}(C) = \mathbb{P}(C|A)$$
$$= \mathbb{P}(\textit{"both scores are even"}|\textit{"the score of the first die is } 4\textit{"})$$
$$= \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\{(4,2),(4,4),(4,6)\}}{\mathbb{P}(A)} = \frac{\frac{3}{36}}{\frac{6}{36}} = \frac{1}{2}$$

$$\mathbb{P}_{\text{new}}(D) = \mathbb{P}(D|A)$$
$$= \mathbb{P}(\textit{"both scores are odd"}|\textit{"the score of the first die is } 4\textit{"})$$
$$= \frac{\mathbb{P}(A \cap D)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(A)} = 0.$$

- The function

$$\mathbb{P}_{\text{new}}(\cdot) = \mathbb{P}(\cdot|A) : \mathcal{F} = \text{ set of events } \to \mathbb{R}$$
$$\mathbb{P}_{\text{new}}(B) = \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \quad B \in \mathcal{F}, \text{ i.e. } B \text{ is an event,}$$

is a (new) measure of probability for the experiment: it satisfies the three conditions required in the definition of a measure of probability, as now we show.

Positivity: for any event B, we have

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \geq 0.$$

Normalization:

$$\mathbb{P}(\Omega|A) = \frac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

Countable additivity: for a sequence $B_1, B_2, B_3, \ldots$ of disjoint events we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i | A\right) = \frac{\mathbb{P}\left(A \cap \bigcup_{i=1}^{\infty} B_i\right)}{\mathbb{P}(A)} = \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} A \cap B_i\right)}{\mathbb{P}(A)}$$

$$= \frac{\sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} \quad \text{since } A \cap B_1, A \cap B_2, A \cap B_3, \ldots \quad \text{are disjoint}$$

$$= \sum_{i=1}^{\infty} \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \sum_{i=1}^{\infty} \mathbb{P}(B_i | A).$$

Note that:

- if B and A are disjoint, then
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(A)} = 0;$$
- if B is included in A, then
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)}.$$

The function

$$\mathbb{P}(\cdot|A) : \mathcal{G} \to \mathbb{R} \text{ with } \mathcal{G} := \text{"set of events included in } A\text{"} = \{B \in \mathcal{F} : B \subseteq A\}$$
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)}{\mathbb{P}(A)}, \quad B \in \mathcal{G} \text{ , i.e. } B \text{ is an event included in } A, \tag{5}$$

is a measure of probability for the experiment with the new sample space A, instead of Ω.

Exercise. Prove that $\mathcal{G}$ is a $\sigma-$algebra of subsets of A and that (5) is a measure of probability.

- We can justify this definition of conditional probability in three different ways.

  First way (classical probability). Consider an experiment with a finite sample space Ω and assume that all the elementary events have the same probability $\frac{1}{|\Omega|}$ (as in the example of the dice).

  Suppose that the event A occurs. Basing on this, we want to assign a new measure of probability, denoted by $\mathbb{P}_{\rm new}$ to the events: for any event B, we have

$$\begin{aligned} \mathbb{P}_{\rm new}(B) &= \sum_{x\in B} \mathbb{P}_{\rm new}(x) = \sum_{x\in A\cap B} \mathbb{P}_{\rm new}(x) + \sum_{x\in A^c\cap B} \underbrace{\mathbb{P}_{\rm new}(x)}_{=0 \text{ since } x\notin A} \\ &= \sum_{x\in A\cap B} \mathbb{P}_{\rm new}(x)\,. \end{aligned} \tag{6}$$

Moreover

$$1 = \mathbb{P}_{\rm new}(A) = \sum_{x\in A} \mathbb{P}_{\rm new}(x)\,. \tag{7}$$

Since all the elementary events in A have the same probability $\mathbb{P}$, it is reasonable assume that they have also the same probability $\mathbb{P}_{\text{new}}$, since both measure of probabilities refer to the same experiment. Let c be this common new probability. Then, by (7),

$$1 = \sum_{x \in A} \mathbb{P}_{\text{new}}(x) = \sum_{x \in A} c = |A| c$$

and then

$$\mathbb{P}_{\text{new}}(x) = c = \frac{1}{|A|}, \ x \in A.$$

Therefore, by (6) we have

$$\mathbb{P}_{\text{new}}(B) = \sum_{x \in A \cap B} \mathbb{P}_{\text{new}}(x) = \sum_{x \in A \cap B} \frac{1}{|A|} = \frac{|A \cap B|}{|A|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Second way (frequentist interpretation). Repeat the experiment a very large number n of times and let $\omega_1^{\text{obs}}, \omega_2^{\text{obs}}, \ldots, \omega_n^{\text{obs}}$ be the observed outcomes.

Let B be an event, we have

$$\begin{aligned} \mathbb{P}(B) &= \text{Long Time Relative Frequency of } B \\ &= \lim_{n\to\infty} \frac{\left|\left\{i \in \{1, 2 \ldots, n\} : B\left(\omega_i^{\text{obs}}\right\}\right|}{n}. \end{aligned}$$

Assume that the event A occurs. We want to assign a new measure of probability $\mathbb{P}_{\text{new}}$ to the events.

Since A occurs, it is reasonable to set, for any event B,

$$\begin{aligned} \mathbb{P}_{\text{new}}(B) &= \text{Long Time Relative Frequency of } B \text{ when } A \text{ occurs} \\ &= \lim_{n\to\infty} \frac{\left|\left\{i \in \{1, 2, \ldots, n\} : A\left(\omega_i^{\text{obs}}\right) \text{ and } B\left(\omega_i^{\text{obs}}\right)\right\}\right|}{\left|\left\{i \in \{1, 2, \ldots, n\} : A\left(\omega_i^{\text{obs}}\right)\right\}\right|}. \end{aligned}$$

In fact, since A is the new set of possible outcomes, we have to exclude repetitions of the experiment with outcomes not in A.

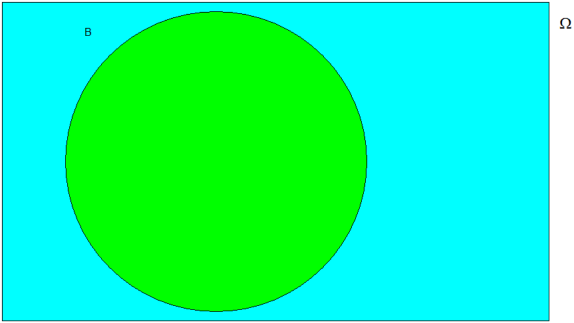
We have

$$
\begin{aligned}
\mathbb{P}_{\text{new}}(B) &= \lim_{n\to\infty} \frac{\left|\left\{ i \in \{1,2,\ldots,n\} : A\left(\omega_i^{\text{obs}}\right) \text{ and } B\left(\omega_i^{\text{obs}}\right) \right\}\right|}{\left|\left\{ i \in \{1,2,\ldots,n\} : A\left(\omega_i^{\text{obs}}\right) \right\}\right|} \\
&= \lim_{n\to\infty} \frac{\frac{\left|\left\{ i \in \{1,2,\ldots,n\} : A\left(\omega_i^{\text{obs}}\right) \text{ and } B\left(\omega_i^{\text{obs}}\right) \right\}\right|}{n}}{\frac{\left|\left\{ i \in \{1,2,\ldots,n\} : A\left(\omega_i^{\text{obs}}\right) \right\}\right|}{n}} \\
&= \frac{\lim_{n\to\infty} \frac{\left|\left\{ i \in \{1,2,\ldots,n\} : A\left(\omega_i^{\text{obs}}\right) \text{ and } B\left(\omega_i^{\text{obs}}\right) \right\}\right|}{n}}{\lim_{n\to\infty} \frac{\left|\left\{ i \in \{1,2,\ldots,n\} : A\left(\omega_i^{\text{obs}}\right) \right\}\right|}{n}} \\
&= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.
\end{aligned}
$$

Third way (Probability as "area"). The probability of an event can thought as the "area" of the event, normalized in order to have "area" of the sample space Ω equal to 1:

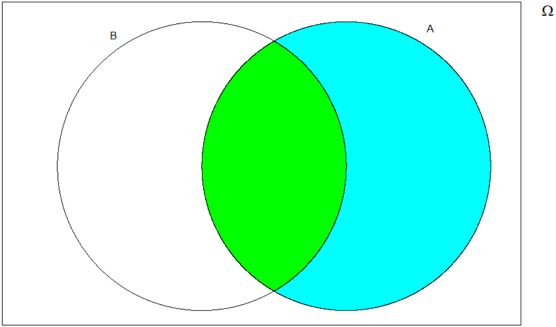
$$\mathbb{P}(B) = \frac{\text{Area}(B)}{\text{Area}(\Omega)}.$$

Thus, in a measure of probability, we compare the "area" of the event B with the "area" of the sample space Ω.



Assume that the event A occurs. Now, we have a new measure of probability $\mathbb{P}_{\text{new}}$, where we compare the "area" of the event B, which is restricted to $A \cap B$ since A occurs, to the "area" of the new set of possible outcomes A:

$$\mathbb{P}_{\text{new}}(B) = \frac{\text{Area}(A \cap B)}{\text{Area}(A)} = \frac{\frac{\text{Area}(A \cap B)}{\text{Area}(\Omega)}}{\frac{\text{Area}(A)}{\text{Area}(\Omega)}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$



- If Ω is finite and all the elementary events have the same probability $\frac{1}{|\Omega|}$ (classical probability), then

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

In fact

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

Exercise. Find a formula for the conditional probability $\mathbb{P}(B|A)$ in the context of classical probability for continuous sample space.

Now, we present two problems of computation of conditional probabilities where Ω is finite and all the elementary events have the same probability.

Problem 1. Assume that someone, say Peter, has two children. What is the probability that Peter has two daughters given that he has a daughter?

Since the other child is a boy or a girl, it seems that the probability is $\frac{1}{2}$, but this is not the right answer.

Experiment: the births of the two children of Peter.

Outcome: $(\textit{gender of the first child}, \ \textit{gender of the second child})$.

Sample space: $\Omega = \{(g, g), (g, b), (b, g), (b, b)\}$.

Measure of probability: the elementary events have probabilities

$$\mathbb{P}((g, g)) = \mathbb{P}((g, b)) = \mathbb{P}((b, g)) = \mathbb{P}((b, b)) = \frac{1}{4}.$$

Consider the events

$$A = \text{"Peter has a daughter"} = \{(g,g),(g,b),(b,g)\}$$

and

$$B = \text{"Peter has two daughters"} = \{(g,g)\}.$$

We have

$$\mathbb{P}(B|A) = \frac{|A \cap B|}{|A|} = \frac{|B|}{|A|} = \frac{1}{3}.$$

The answer is $\frac{1}{3}$, not $\frac{1}{2}$, because the daughter cited in the event A can be the first or the second child of Peter.

Of course, if A was "Peter has a daughter as first child", i.e. $A = \{(g,g),(g,b)\}$, then $\mathbb{P}(B|A)$ was $\frac{1}{2}$.

Exercise. Assume that Peter has n children. What is the probability that Peter has n daughters given that he has $n - 1$ *daughters?*

Exercise. Assume that Peter has three children. What is the probability that Peter has three sons given that he has a son and this son is not the youngest children of Peter.

Problem 2. In table below are listed the number of students enrolled in an american college, categorized by sex and age.

Table 4.4 Enrollment

Sex and age	
Total	12,544
Male	5,881
14 to 17 years old	91
18 and 19 years old	1,309
20 and 21 years old	1,089
22 to 24 years old	1,080
25 to 29 years old	1,016
30 to 34 years old	613
35 years old and over	684
Female	6,663
14 to 17 years old	119
18 and 19 years old	1,455
20 and 21 years old	1,135
22 to 24 years old	968
25 to 29 years old	931
30 to 34 years old	716
35 years old and over	1,339

Consider the experiment where a student enrolled in this college is randomly selected.

What is the conditional probability that the selected student is over 30, given that this student is a man?

We have

$$\mathbb{P}\left(\text{"the student is over 30"} \middle| \text{"the student is a man"}\right) = \frac{|\{\textit{over 30}\} \cap \{\textit{men}\}|}{|\{\textit{men}\}|} = \frac{613 + 684}{5881} = 22.05\%.$$

What is the conditional probability that the selected student is a woman, given that this student is under 20?

We have

$$\mathbb{P}\left(\text{"the student is a woman"} \middle| \text{"the student is under 20"}\right) = \frac{|\{\textit{women}\} \cap \{\textit{under } 20\}|}{|\{\textit{under } 20\}|} = \frac{119 + 1455}{119 + 1455 + 91 + 1309} = 52.93\%.$$

Exercise. What is the conditional probability that the selected student is a man, given that this student is under 20?

- Exercise. In the previous problem of the earning in US, what is the conditional probability that the selected worker is a woman, given that she/he earns over 50,000 dollars?

  Exercise. In the experiment of rolling two dice, what is the conditional probability that the sum of the scores is s, where $s \in \{2, 3, \ldots, 12\}$, given that the sum of the scores is even?

  Exercise. In the experiment of the 100 m run with all arrival orders having the same probability, what is the conditional probability that the runner 4 is at least fourth, given that the runner 3 is at least third?

  Exercise. In the experiment of flipping a coin until Head appears, what is the conditional probability that Head appears within 10 flips, given it appears after an even number of flips?

Exercise. In the experiment of the meteor falling, what is the conditional probability that the meteor falls in Italy given that it falls in Europe?

Exercise. Let A and B disjoint events. Shows that $\mathbb{P}(B|A) \leq \mathbb{P}(A|B)$ if and only if $\mathbb{P}(B) \leq \mathbb{P}(A)$. Does this equivalence still hold when A and B are disjoint?

Assign probabilities by conditional probability

- We can rewrite the conditional probability equation as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \tag{8}$$

This formula is useful in the following situation.

We have seen that, in case of Ω discrete, the measure of probability is determined once that the probabilities of the elementary events are assigned.

Often, it is no so easy assign directly such probabilities in the situation where the sample space is some cartesian product (or some subset of a cartesian product), as $\Omega = \{H, T\}^2$ for the experiment of the two coins or $\Omega = \{1, \ldots, 6\}^2$ for the experiment of the two dice.

In such a situation, (8) helps to assign the probabilities of the elementary events, as it is now explained.

Let $\Omega \subseteq E \times F$, where E and F are discrete sets and so Ω is also a discrete set.

The probability of the elementary event $\{(e, f)\}$, with $(e, f) \in \Omega$, can be assigned by

$$\begin{aligned}
&\mathbb{P}((e, f)) \\
&= \mathbb{P}(\{\omega \in \Omega : \omega_1 = e\} \cap \{\omega \in \Omega : \omega_2 = f\}) \\
&= \mathbb{P}\left(\{\omega \in \Omega : \omega_1 = e\}\right) \cdot \mathbb{P}\left(\{\omega \in \Omega : \omega_2 = f\} \,|\, \{\omega \in \Omega : \omega_1 = e\}\right)
\end{aligned}$$

if the probabilities

$$\mathbb{P}\left(\{\omega \in \Omega : \omega_1 = e\}\right)$$

and

$$\mathbb{P}\left(\{\omega \in \Omega : \omega_2 = f\} \,|\, \{\omega \in \Omega : \omega_1 = e\}\right)$$

are easy to assign.

Example. In the experiment of the two coins, we have

$$\mathbb{P}\left(\{\omega \in \Omega : \omega_1 = e\}\right) = \mathbb{P}\left(\textit{the first coin shows } e\right) = \frac{1}{2}$$

and

$$\begin{aligned}&\mathbb{P}\left(\{\omega \in \Omega : \omega_2 = f\} \mid \{\omega \in \Omega : \omega_1 = e\}\right) \\ &= \mathbb{P}\left(\textit{the second coin shows } f \middle| \textit{the first coin shows } e\right) = \frac{1}{2}.\end{aligned}$$

since the face shown by the first coin does not affect the face shown by the second coin.

Thus, for any $(e, f) \in \Omega = \{H, T\}^2$*, we have*

$$\mathbb{P}((e, f)) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Example. In the experiment of the two dice, we have

$$\mathbb{P}\left(\{\omega \in \Omega : \omega_1 = e\}\right) = \mathbb{P}\left(\textit{the first die scores } e\right) = \frac{1}{6}$$

and

$$\mathbb{P}\left(\{\omega \in \Omega : \omega_2 = f\} \,|\, \{\omega \in \Omega : \omega_1 = e\}\right)$$
$$= \mathbb{P}\left(\textit{the second die scores } f | \textit{the first die scores } e\right) = \frac{1}{6}.$$

since the score of the first die does not affect the score of the second die.

Thus, for any $(e, f) \in \Omega = \{1, \ldots, 6\}^2$, *we have*

$$\mathbb{P}((e, f)) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Example. Consider the experiment where two playcards are randomly chosen from a deck D of n cards.

The outcome is the pair $(\textit{first chosen card}, \textit{second chosen card})$.

The sample space is

$$\Omega = \left\{ \omega \in D^2 : \omega_1 \neq \omega_2 \right\}.$$

Here it is understood that when the second card is chosen, the first card chosen is not re-inserted in the deck.

Given a pair $(e, f) \in \Omega$, the probability of the elementary event $\{(e, f)\}$, i.e. the event

"the chosen cards are e and f"

is

$$\begin{aligned} &\mathbb{P}\,(\textit{the first chosen card is } e) \\ &\cdot \mathbb{P}\,(\textit{the second chosen card is } f | \textit{the first chosen card is } e) \\ &= \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n\,(n-1)}. \end{aligned}$$

Exercise. Assume to have a standard deck of 52 *playcards. Determine the probabilities of the following events:*

- *Both chosen cards are Hearts.*
- *The chosen cards have the same value.*
- *One chosen card is the Queen of Spades and the other is the King of Clubs.*

Exercise. Consider the previous experiment with a standard deck of 52 playcards, but now suppose that if the first chosen card is a Jack, a Queen or a King , then all the Jacks, Queens and Kings are removed from the deck before to choose the second card. Determine the probability of the elementary events.

- The generalization of equation

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

to many events $A_1, A_2, \dots, A_n$ is

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \quad \cdots \quad \cdot \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

In fact

$$\begin{aligned}
&\mathbb{P}\left(A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1} \cap A_n\right) \\
&= \mathbb{P}\left(A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1}\right) \cdot \mathbb{P}\left(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1}\right) \\
&= \mathbb{P}\left(A_1 \cap A_2 \cap \dots \cap A_{n-2}\right) \cdot \mathbb{P}\left(A_{n-1}|A_1 \cap A_2 \cap \dots \cap A_{n-2}\right) \\
&\qquad \cdot \mathbb{P}\left(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1}\right) \\
&= \dots \\
&= \mathbb{P}\left(A_1\right) \cdot \mathbb{P}\left(A_2|A_1\right) \cdot \quad \cdots \quad \cdot \ \mathbb{P}\left(A_{n-1}|A_1 \cap A_2 \cap \dots \cap A_{n-2}\right) \\
&\qquad \cdot \mathbb{P}\left(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1}\right)
\end{aligned}$$

Now, we present an example where this formula is used.

Example. Suppose that three people are randomly chosen from a group of 4 women and 6 men.

What is the probability that all of them are women?

What is the probability that one is a woman and the other two are men?

The outcome of the experiment is the triple

$$(\textit{first chosen person}, \textit{second chosen person}, \textit{third chosen person}) .$$

The sample space is

$$\Omega = \left\{\omega \in E^3 : \omega_1, \omega_2 \textit{ and } \omega_3 \textit{ are distinct}\right\} ,$$

where E is the set of the women and men.

We have

$$
\begin{aligned}
&\mathbb{P}(\textit{"all women"}) \\
&= \mathbb{P}(\textit{"the first is woman"}) \\
&\quad \cdot \mathbb{P}(\textit{"the second is woman"}|\textit{"the first is woman"}) \\
&\quad \cdot \mathbb{P}(\textit{"the third is woman"}|\textit{"the first and the second are women"}) \\
&= \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{30}.
\end{aligned}
$$

We have

$$
\begin{aligned}
&\mathbb{P}(\textit{"a woman and two men"}) \\
&= \mathbb{P}(\textit{"the first is woman and the other two are men"}) \\
&\quad + \mathbb{P}(\textit{"the second is woman and the other two are men"}) \\
&\quad + \mathbb{P}(\textit{"the third is woman and the other two are men"})
\end{aligned}
$$

with

$$\mathbb{P}\left(\textit{"the first is woman and the other two are men"}\right)$$
$$= \mathbb{P}\left(\textit{"the first is woman"}\right)$$
$$\cdot \mathbb{P}\left(\textit{"the second is man"}|\textit{"the first is woman"}\right)$$
$$\cdot \mathbb{P}\left(\textit{"the third is man"}|\textit{"the first is woman and the second is man"}\right)$$
$$= \frac{4}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} = \frac{1}{6}$$

$$\mathbb{P}\left(\textit{"the second is woman and the other two are men"}\right)$$
$$= \mathbb{P}\left(\textit{"the first is man"}\right)$$
$$\cdot \mathbb{P}\left(\textit{"the second is woman"}|\textit{"the first is man"}\right)$$
$$\cdot \mathbb{P}\left(\textit{"the third is man"}|\textit{"the first is man and the second is woman"}\right)$$
$$= \frac{6}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{1}{6}$$

and

$$\begin{aligned}
&\mathbb{P}\left(\textit{"the third is woman and the other two are men"}\right)\\
&=\mathbb{P}\left(\textit{"the first is man"}\right)\\
&\cdot\mathbb{P}\left(\textit{"the second is man"}|\textit{"the first is man"}\right)\\
&\cdot\mathbb{P}\left(\textit{"the third is woman"}|\textit{"the first is man and the second is man"}\right)\\
&=\frac{6}{10}\cdot\frac{5}{9}\cdot\frac{4}{8}=\frac{1}{6}.
\end{aligned}$$

Thus

$$\mathbb{P}(\textit{"a woman and two men"})=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}.$$

Exercise. What is the probability that all are men and the probability that one is man and two are women.

Exercise. Determine the probabilities of the elementary events for this experiment.

- Exercise. What is the probability of obtain n times, one after the other, "Red" at the roulette game?

  Exercise. Explain why the probabilities of the elementary events in the experiment of flipping a coin until H appears are just those previously introduced.

  Exercise. In the experiment where k playcards are randomly chosen from a deck of n playcards, determine the probability of the elementary events.

  Exercise. In the experiment where four playcards are randomly chosen from a standard deck of 52 playcards, determine the probabilities of the following events:

  - The chosen cards are all Hearts.
  - The chosen cards have all different values.
  - The chosen cards are the four Queens.

Exercise (The birthdays problem). Consider the experiment where n people are randomly selected from an huge population. What is the probability that the birthdays of the selected people are all different? What is the probability that there are (at least) two people with the same birthday? Find the minimum n for which it is easier to have two people with the same birthday than not to. Do not consider leap years.

- Here is another example where we use conditional probabilities for assigning probabilities to the elementary events.

  The Monty Hall problem. The 70's american show "Let's Make a Deal" with Monty Hall (Maurice Halprin) as host

  contained a game, described below, where a player had the possibility of winning a car.

There are three closed doors: behind one door there is a car and behind the other two there are goats.

The game consists of three stages.

1. *The host invites the player to choose one of the closed door without open it.*
2. *Then, the host, which knows where is the car, opens one of the two doors not chosen by the player revealing a goat.*
3. *Finally, the host asks to the player if she/he wants to change the previous selected door with the other closed door or keep it. After the answer and the possible change, the player opens her/him door and wins what is behind, a car or a goat.*

Now, the question is:

- *in order to win the car, what is the best choice for the player at the stage 3: change the door or keep it?*

It seems that it is the same to change or to keep: in any case, by changing or by keeping, we have two closed doors with a car behind one door and a goat behind the other.

Observe that in "Let's Make a Deal" there was not the stage 3, for reasons that will be clear soon. The stage 2 was done only for the show, to keep high the tension.

Now, we give a probabilistic structure to the game.

Experiment: the selection of the three doors at the stages 1, 2 and 3.

Outcome: the triple

$$(\textit{door selected in 1, door selected in 2, door selected in 3})\,.$$

Sample space: let C *be the door with car behind and let* $G1$ *and* $G2$ *be the other two doors with the goats behind, we have*

$$\begin{aligned}\Omega \;=\; &\{(C, G1, C), (C, G1, G2), (C, G2, C), (C, G2, G1),\\ &(G1, G2, G1), (G1, G2, C),\\ &(G2, G1, G2), (G2, G1, C)\ \}.\end{aligned}$$

Measure of probability: if the player at the stage 3 changes the door, we have

$$\mathbb{P}\left((C, G1, C)\right) = \mathbb{P}\left(''\omega_1 = C''\right) \cdot \mathbb{P}\left(''\omega_2 = G1'' | ''\omega_1 = C''\right) \cdot$$
$$\cdot \mathbb{P}\left(''\omega_3 = C'' | ''\omega_1 = C'' \cap ''\omega_2 = G1''\right) = \frac{1}{3} \cdot \frac{1}{2} \cdot 0 = 0$$
$$\mathbb{P}\left((C, G1, G2)\right) = \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{6}$$
$$\mathbb{P}\left((C, G2, C)\right) = \frac{1}{3} \cdot \frac{1}{2} \cdot 0 = 0$$
$$\mathbb{P}\left((C, G2, G1)\right) = \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{6}$$
$$\mathbb{P}\left((G1, G2, G1)\right) = \frac{1}{3} \cdot 1 \cdot 0 = 0$$
$$\mathbb{P}\left((G1, G2, C)\right) = \frac{1}{3} \cdot 1 \cdot 1 = \frac{1}{3}$$
$$\mathbb{P}\left((G2, G1, G2)\right) = \frac{1}{3} \cdot 1 \cdot 0 = 0$$
$$\mathbb{P}\left((G2, G1, C)\right) = \frac{1}{3} \cdot 1 \cdot 1 = \frac{1}{3}.$$

The probability of winning the car, i.e. the prob. of $''\omega_3 = C''$*, is* $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$.

Measure of probability: if the player at the stage 3 keeps the door, we have

$$
\begin{aligned}
\mathbb{P}\left((C, G1, C)\right) &= \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{6} \\
\mathbb{P}\left((C, G1, G2)\right) &= \frac{1}{3} \cdot \frac{1}{2} \cdot 0 = 0 \\
\mathbb{P}\left((C, G2, C)\right) &= \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{6} \\
\mathbb{P}\left((C, G2, G1)\right) &= \frac{1}{3} \cdot \frac{1}{2} \cdot 0 = 0 \\
\mathbb{P}\left((G1, G2, G1)\right) &= \frac{1}{3} \cdot 1 \cdot 1 = \frac{1}{3} \\
\mathbb{P}\left((G1, G2, C)\right) &= \frac{1}{3} \cdot 1 \cdot 0 = 0 \\
\mathbb{P}\left((G2, G1, G2)\right) &= \frac{1}{3} \cdot 1 \cdot 1 = \frac{1}{3} \\
\mathbb{P}\left((G2, G1, C)\right) &= \frac{1}{3} \cdot 1 \cdot 0 = 0.
\end{aligned}
$$

The probability of winning the car is $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Conclusion:

- *it is better to change at the stage 3.*

This is clear also by the following simple argument:

- *If the player keeps the door, then she/he wins the car if and only if at the stage 1 it is selected the door with the car behind and this has probability* $\frac{1}{3}$.
- *If the player changes the door, then she/he wins the car if and only if at the stage 1 it is selected a door with a goat behind and this has probability* $\frac{2}{3}$.

Exercise. What happens in the situation where the host does not know what is behind the doors and so he can also open at the stage 2 the door with the car behind?

Exercise. What is the probability of winning the car if at the stage 3 the player changes the door with probability p and keeps it with probability $q = 1 - p$? Consider both the cases where the host knows what is behind the doors and does not know what is behind.

Exercise. What happens in the situation where there are n doors with k doors with a car behind and the other $n - k$ doors with a goat behind? Consider both the cases where the host knows what is behind the doors and does not know what is behind.

Independence of events

Definition

Let $\mathbb{P}$ be a measure of probability for an experiment and let A and B be events relative to this experiment. The following three facts are equivalent:

a) $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

b) $\mathbb{P}(B|A) = \mathbb{P}(B)$ if $\mathbb{P}(A) \neq 0$

c) $\mathbb{P}(A|B) = \mathbb{P}(A)$ if $\mathbb{P}(B) \neq 0$.

If one between a), b) and c) is satisfied (and so all a), b) and c) are satisfied), we say that A and B are **independent**.

In other words, two events are independent if the knowledge that one of them occurs does not modify the probability of the other.

Now, we prove the equivalence of a), b) and c).

a) $\Rightarrow$ *b*). If a) holds, then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)\,\mathbb{P}(B)}{\mathbb{P}(A)} = \mathbb{P}(B) \text{ if } \mathbb{P}(A) \neq 0.$$

b) $\Rightarrow$ *a*). If b) holds, then

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) = \mathbb{P}(A) \cdot \mathbb{P}(B) \text{ if } \mathbb{P}(A) \neq 0.$$

If $\mathbb{P}(A) = 0$, then, since $A \cap B \subseteq A$, we have

$$0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(A) = 0$$

and so $\mathbb{P}(A \cap B) = 0$ and we conclude

$$\mathbb{P}(A \cap B) = 0 = 0 \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Exercise. Prove *a*) $\Rightarrow$ *c*), *c*) $\Rightarrow$ *a*), *b*) $\Rightarrow$ *c*) and *c*) $\Rightarrow$ *b*).

Example. Consider the experiment of the dice and the events

$$A = \text{"the score of the first die is } 3\text{"}$$
$$B = \text{"the sum of the scores is } 8\text{"}$$
$$C = \text{"the sum of the scores is } 7\text{".}$$

A and B are not independent:

$$\mathbb{P}(A \cap B) = \mathbb{P}((3,5)) = \frac{1}{36},\ \mathbb{P}(A) = \frac{1}{6},\ \mathbb{P}(B) = \frac{13-8}{36} = \frac{5}{36}$$
$$\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\,\mathbb{P}(B)\,.$$

A and C are independent:

$$\mathbb{P}(A \cap C) = \mathbb{P}((3,4)) = \frac{1}{36},\ \mathbb{P}(A) = \frac{1}{6},\ \mathbb{P}(C) = \frac{7-1}{36} = \frac{6}{36}$$
$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C)\,.$$

We can understand this also noting that

$$\mathbb{P}(A) = \mathbb{P}(\textit{"the score of the first die is } 3\textit{"}) = \frac{1}{6}$$

$$\begin{aligned}
&\mathbb{P}(A|B) \\
&= \mathbb{P}(\textit{"the score of the first die is } 3\textit{"} | \textit{"the sum of the scores is } 8\textit{"}) \\
&= \frac{1}{5}
\end{aligned}$$

since "the sum of the scores is 8*"*$= \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$.

$$\begin{aligned}
&\mathbb{P}(A|C) \\
&= \mathbb{P}(\textit{"the score of the first die is } 3\textit{"} | \textit{"the sum of the scores is } 7\textit{"}) \\
&= \frac{1}{6}
\end{aligned}$$

since "the sum of the scores is 7*"*$= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

Then, A and B are not independent and A and C are independent.

- Let $\Omega = E \times F$, with E and F finite. Assume that all the elementary events $\{(e,f)\}$, where $(e,f) \in \Omega$, have the same probability

$$\mathbb{P}\left((e,f)\right) = \frac{1}{|\Omega|} = \frac{1}{|E| \cdot |F|}.$$

Then, for any $U \subseteq E$ and $V \subseteq F$, the events "$\omega_1 \in U$" and "$\omega_2 \in V$" are independent and

$$\mathbb{P}(\text{"}\omega_1 \in U\text{"}) = \frac{|U|}{|E|} \text{ and } \mathbb{P}(\text{"}\omega_2 \in V\text{"}) = \frac{|V|}{|F|}.$$

In fact

$$\mathbb{P}\left(\text{"}\omega_1 \in U\text{"}\right) = \mathbb{P}\left(U \times F\right) = \frac{|U \times F|}{|\Omega|} = \frac{|U| \cdot |F|}{|E| \cdot |F|} = \frac{|U|}{|E|}$$

$$\mathbb{P}\left(\text{"}\omega_2 \in V\text{"}\right) = \mathbb{P}\left(E \times V\right) = \frac{|E \times V|}{|\Omega|} = \frac{|E| \cdot |V|}{|E| \cdot |F|} = \frac{|V|}{|F|}$$

$$\mathbb{P}\left(\text{"}\omega_1 \in U\text{"} \cap \text{"}\omega_2 \in V\text{"}\right) = \mathbb{P}\left(U \times V\right) = \frac{|U \times V|}{|\Omega|} = \frac{|U| \cdot |V|}{|E| \cdot |F|}$$

$$= \mathbb{P}(\text{"}\omega_1 \in U\text{"}) \cdot \mathbb{P}(\text{"}\omega_2 \in V\text{"}).$$

Examples:

- *in the experiment of the coins, the events "the first coin shows a face in E" and "the second coin shows a face in F", where $E, F \subseteq \{H, T\}$, are independent.*
- *In the experiment of the dice, the events "the score of the first die is in U" and "the score of the second die is in V", where $U, V \subseteq E = F = \{1, \ldots, 6\}$, are independent.*

So, in the example of the dice, the events "the score of the first die is odd" and "the score of the second die is smaller than four" are independent.

Example. Consider once again the table

Table 4.4 Enrollment

Sex and age	
Total	12,544
Male	5,881
14 to 17 years old	91
18 and 19 years old	1,309
20 and 21 years old	1,089
22 to 24 years old	1,080
25 to 29 years old	1,016
30 to 34 years old	613
35 years old and over	684
Female	6,663
14 to 17 years old	119
18 and 19 years old	1,455
20 and 21 years old	1,135
22 to 24 years old	968
25 to 29 years old	931
30 to 34 years old	716
35 years old and over	1,339

of the students enrolled in an american college.

Suppose that a male student is randomly selected and, independently, also a female student is randomly selected.

We want to find the probability that the students selected are both between 22 and 24 years old.

The sample space is

$$\Omega = \{men\} \times \{women\}.$$

The fact that the male student and the female student are "randomly and independently selected" means that, for any $(m, w) \in \Omega$*, we have*

$$\begin{aligned} \mathbb{P}\left(\text{"the selected man is } m\text{"}\right) &= \frac{1}{|\{men\}|} \\ \mathbb{P}\left(\text{"the selected woman is } w\text{"}\right) &= \frac{1}{|\{women\}|} \end{aligned}$$

and the events

"the selected man is m" and "the selected woman is w"

are independent.

The probabilities of the elementary events are all equal:

$$\mathbb{P}((m,w)) = \mathbb{P}(\text{"the selected man is } m\text{"} \cap \text{"the selected woman is } w\text{"})$$
$$= \mathbb{P}(\text{"the selected man is } m\text{"}) \cdot \mathbb{P}(\text{"the selected woman is } w\text{"})$$
$$= \frac{1}{|\{men\}|} \cdot \frac{1}{|\{women\}|}, \ (m,w) \in \Omega.$$

Hence, the events

$$A = \text{"the selected man is between 22 and 24"}$$
$$= \text{"}\omega_1 \in U\text{"}, \ U = \{\text{men between } 22 \text{ and } 24\}\text{"}$$
$$B = \text{"the selected woman is between 22 and 24"}$$
$$= \text{"}\omega_2 \in \ V\text{"}, \ V = \{\text{women between } 22 \text{ and } 24\}$$

are independent and

$$\mathbb{P}(A) = \frac{|U|}{|\{men\}|} \ \text{ and } \ \mathbb{P}(B) = \frac{|V|}{|\{women\}|}.$$

So

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{|U|}{|\{men\}|} \cdot \frac{|V|}{|\{women\}|} = \frac{1080}{5881} \cdot \frac{968}{6663} = 2.67\%.$$

Exercise. Find the prob. that both students are in the same age group.

- Exercise. Consider the experiment of flipping two coins. Are the events

  "both faces are equal" and "the first coin shows H"

  independent?

  Exercise. Let A and B events. Prove that if A and B are independent, then B and A are independent.

  Exercise. Let A and B events. Prove that if A and B are independent, then A^c and B are independent, A and B^c are independent and A^c and B^c are independent.

  Exercise. Prove that if one of two events is $\emptyset$ or Ω, the events are independent.

Independence of many events

- We can generalize the notion of independence to a finite sequence $A_1, A_2, \ldots, A_n$, or to an infinite sequence $A_1, A_2, A_3, \ldots$, of events in the following way.

  Let I be the set of indices for the sequence: we have $I = \{1, 2, \ldots, n\}$ for the finite sequence and $I = \{1, 2, 3, \ldots\}$ for the infinite sequence.

  The events of the sequence are called **independent** if, for any positive integer k such that $2 \leq k \leq |I| =$ "number of indices" and for any indices $i_1, \ldots, i_{k-1}, i_k \in I$ distinct such that $\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_{k-1}}\right) \neq 0$, we have

  $$\mathbb{P}\left(A_{i_k} | A_{i_1} \cap \cdots \cap A_{i_{k-1}}\right) = \mathbb{P}\left(A_{i_k}\right).$$

  In other words, the events are independent if the knowledge that some of them occur does not modify the probability of the others.

In case of $n = 3$ events A_1, A_2, A_3, this definition becomes:

$$k = 2: \quad \begin{array}{l} \mathbb{P}(A_1|A_2) = \mathbb{P}(A_1|A_3) = \mathbb{P}(A_1) \\ \mathbb{P}(A_2|A_1) = \mathbb{P}(A_2|A_3) = \mathbb{P}(A_2) \\ \mathbb{P}(A_3|A_1) = \mathbb{P}(A_3|A_2) = \mathbb{P}(A_3) \end{array}$$

and

$$k = 3: \quad \begin{array}{l} \mathbb{P}(A_1|A_2 \cap A_3) = \mathbb{P}(A_1) \\ \mathbb{P}(A_2|A_1 \cap A_3) = \mathbb{P}(A_2) \\ \mathbb{P}(A_3|A_1 \cap A_2) = \mathbb{P}(A_3). \end{array}$$

- Equivalently, the independence of the events can be expressed by saying that, for any positive integer k such that $2 \leq k \leq |I|$ and for any indices $i_1, i_2, \ldots, i_k \in I$ distinct, we have

$$\mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = \mathbb{P}\left(A_{i_1}\right) \cdot \mathbb{P}\left(A_{i_2}\right) \cdot \quad \cdots \quad \cdot \mathbb{P}\left(A_{i_k}\right).$$

In fact, if this is true, then, the previous definition of independence is true: for $i_1, \ldots, i_{k-1}, i_k \in I$ distinct such that $\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_{k-1}}\right) \neq 0$, we have

$$\begin{aligned}\mathbb{P}\left(A_{i_k} | A_{i_1} \cap \cdots \cap A_{i_{k-1}}\right) &= \frac{\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_{k-1}} \cap A_{i_k}\right)}{\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_{k-1}}\right)} \\ &= \frac{\mathbb{P}\left(A_{i_1}\right) \cdots \mathbb{P}\left(A_{i_{k-1}}\right) \mathbb{P}\left(A_{i_k}\right)}{\mathbb{P}\left(A_{i_1}\right) \cdots \mathbb{P}\left(A_{i_{k-1}}\right)} = \mathbb{P}\left(A_{i_k}\right).\end{aligned}$$

Viceversa, if the events are independent with respect to the previous definition, then for $i_1, i_2, \ldots, i_k \in I$ distinct, we have

$$\begin{aligned}\mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) &= \mathbb{P}\left(A_{i_1}\right) \mathbb{P}\left(A_{i_2} | A_{i_1}\right) \cdots \mathbb{P}\left(A_{i_k} | A_{i_1} \cap A_{i_2} \cdots \cap A_{i_{k-1}}\right) \\ &= \mathbb{P}\left(A_{i_1}\right) \mathbb{P}\left(A_{i_2}\right) \cdots \mathbb{P}\left(A_{i_k}\right).\end{aligned}$$

This is valid when $\mathbb{P}(A_{i_1}), \mathbb{P}(A_{i_1} \cap A_{i_2}), \ldots, \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_{k-1}}\right)$ are all not zero. Assume that this is not true. Since

$$\mathbb{P}(A_{i_1}) \geq \mathbb{P}(A_{i_1} \cap A_{i_2}) \geq \cdots \geq \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_{k-1}}\right)$$

holds, there is an index $s \in \{1, \ldots, k-1\}$ such that

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}) = 0 \text{ and } \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_{s-1}}\right) \neq 0.$$

Since

$$\mathbb{P}\left(A_{i_s} | A_{i_1} \cap \cdots \cap A_{i_{s-1}}\right) = \mathbb{P}(A_{i_s})$$

we obtain

$$\begin{aligned}0 &= \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}) = \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_{s-1}}\right) \mathbb{P}\left(A_{i_s} | A_{i_1} \cap \cdots \cap A_{i_{s-1}}\right) \\ &= \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_{s-1}}\right) \mathbb{P}(A_{i_s})\end{aligned}$$

and then $\mathbb{P}(A_{i_s}) = 0$ and so

$$\mathbb{P}(A_{i_1}) \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}) = 0.$$

On the other hand, by

$$0 = \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}) \geq \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \geq 0,$$

we obtain

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = 0 = \mathbb{P}(A_{i_1}) \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

This equivalent expression of the independence in case of $n = 3$ events A_1, A_2, A_3 becomes:

$$k = 2: \qquad \begin{aligned} \mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(A_1)\,\mathbb{P}(A_2) \\ \mathbb{P}(A_1 \cap A_3) &= \mathbb{P}(A_1)\,\mathbb{P}(A_3) \\ \mathbb{P}(A_2 \cap A_3) &= \mathbb{P}(A_2)\,\mathbb{P}(A_3) \end{aligned}$$

and

$$k = 3: \quad \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3).$$

- The definition of independence for more than two events cannot be reduced to say that they are independent two by two. This is explained by the following exercise.

  Exercise. Consider the experiment of flipping two coins and the events

$$A = \text{"both faces are equal"}$$
$$B = \text{"the first coin shows } H\text{"}$$
$$C = \text{"the second coin shows } H\text{".}$$

  Are the following pairs of events

$$A, B$$
$$A, C$$
$$B, C$$

  independent? Is the following triple of events

$$A, B, C$$

  independent?

Factorized form of the function p and independence

Let $\Omega = E_1 \times E_2 \times \cdots \times E_d$, where E_i, $i \in \{1, 2, \ldots, d\}$, is a discrete set (so also Ω is a discrete set).

Theorem

If the function $p : \Omega \to [0, +\infty)$ such that $\sum_{x \in \Omega} p(x) = 1$, with which we define the probabilities of the elementary events, can be factorized as

$$p(x) = p_1(x_1)\, p_2(x_2) \cdots p_d(x_d), \; x = (x_1, x_2, \ldots, x_d) \in \Omega,$$

where $p_i : E_i \to [0, +\infty)$, $i \in \{1, \ldots, d\}$, is such that

$$\sum_{y \in E_i} p_i(y) = 1,$$

then, for any $J_1 \subseteq E_1, J_2 \subseteq E_2, \ldots, J_d \subseteq E_d$, the events

$$"\omega_1 \in J_1", "\omega_2 \in J_2", \ldots, "\omega_d \in J_d"$$

are independent.

Proof.

For $i \in \{1, \ldots, d\}$, we have

$$\begin{aligned}
\mathbb{P}("\omega_i \in J_i") &= \mathbb{P}("\omega \in E_1 \times \cdots \times E_{i-1} \times J_i \times E_{i+1} \times \cdots \times E_d") \\
&= \sum_{x \in E_1 \times \cdots \times E_{i-1} \times J_i \times E_{i+1} \times \cdots \times E_d} p(x) \\
&= \sum_{x_1 \in E_1} \cdots \sum_{x_{i-1} \in E_{i-1}} \sum_{x_i \in J_i} \sum_{x_{i+1} \in E_{i+1}} \cdots \sum_{x_d \in E_d} p((x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d)) \\
&= \sum_{x_1 \in E_1} \cdots \sum_{x_{i-1} \in E_{i-1}} \sum_{x_i \in J_i} \sum_{x_{i+1} \in E_{i+1}} \cdots \sum_{x_d \in E_d} \\
&\qquad p_1(x_1) \cdots p_{i-1}(x_{i-1}) p_i(x_i) p_{i+1}(x_{i+1}) \cdots p_d(x_d) \\
&= \Big(\sum_{x_1 \in E_1} p_1(x_1)\Big) \cdots \Big(\sum_{x_{i-1} \in E_{i-1}} p_{i-1}(x_{i-1})\Big)\Big(\sum_{x_i \in J_i} p_i(x_i)\Big) \\
&\qquad \Big(\sum_{x_{i+1} \in E_{i+1}} p_{i+1}(x_{i+1})\Big) \cdots \Big(\sum_{x_d \in E_d} p_d(x_d)\Big) \\
&= 1 \cdots 1 \Big(\sum_{x_i \in J_i} p_i(x_i)\Big) 1 \cdots 1 \\
&= \sum_{x_i \in J_i} p_i(x_i).
\end{aligned}$$

Proof.

Now, we prove the independence of the events

$$"\omega_1 \in J_1", "\omega_2 \in J_2", \ldots, "\omega_d \in J_d".$$

For any indices $i_1, \ldots, i_k \in \{1, \ldots, d\}$ distinct, we have

$$\begin{aligned}
& \mathbb{P}\left("\omega_{i_1} \in J_{i_1} \cap \cdots \cap \omega_{i_k} \in J_{i_k}"\right) \\
= \ & \mathbb{P}\left("\omega \in E_1 \times \cdots \times E_{i_1-1} \times J_{i_1} \times E_{i_1+1} \times \cdots \times E_{i_k-1} \times J_{i_k} \times E_{i_k+1} \times \cdots \times E_d"\right) \\
= \ & \sum_{x \in E_1 \times \cdots \times E_{i_1-1} \times J_{i_1} \times E_{i_1+1} \times \cdots \times E_{i_k-1} \times J_{i_k} \times E_{i_k+1} \times \cdots \times E_d} p(x) \\
= \ & \sum_{x_1 \in E_1} \cdots \sum_{x_{i_1-1} \in E_{i_1-1}} \sum_{x_{i_1} \in J_{i_1}} \sum_{x_{i_1+1} \in E_{i_1+1}} \cdots \sum_{x_{i_k-1} \in E_{i_k-1}} \sum_{x_{i_k} \in J_{i_k}} \sum_{x_{i_k+1} \in E_{i_k+1}} \cdots \sum_{x_d \in E_d} \\
& \quad p\left(\left(x_1, \ldots, x_{i_1-1}, x_{i_1}, x_{i_1+1}, \ldots, x_{i_k-1}, x_{i_k}, x_{i_k+1}, \ldots, x_d\right)\right) \\
= \ & \sum_{x_1 \in E_1} \cdots \sum_{x_{i_1-1} \in E_{i_1-1}} \sum_{x_{i_1} \in J_{i_1}} \sum_{x_{i_1+1} \in E_{i_1+1}} \cdots \sum_{x_{i_k-1} \in E_{i_k-1}} \sum_{x_{i_k} \in J_{i_k}} \sum_{x_{i_k+1} \in E_{i_k+1}} \cdots \sum_{x_d \in E_d} \\
& \quad p_1(x_1) \cdots p_{i_1-1}(x_{i_1-1})\, p_{i_1}(x_{i_1})\, p_{i_1+1}(x_{i_1+1}) \cdots
\end{aligned}$$

Proof.

$$
\begin{aligned}
= \quad & \underbrace{\Big(\sum_{x_1\in E_1} p_1(x_1)\Big)}_{=1}\cdots \\
& \underbrace{\Big(\sum_{x_{i_1-1}\in E_{i_1-1}} p_{i_1-1}(x_{i_1-1})\Big)}_{=1}\Big(\sum_{x_{i_1}\in J_{i_1}} p_{i_1}(x_{i_1})\Big)\underbrace{\Big(\sum_{x_{i_1+1}\in E_{i_1+1}} p_{i_1+1}(x_{l_1+1})\Big)}_{=1}\cdots \\
& \underbrace{\Big(\sum_{x_{i_k-1}\in E_{i_k-1}} p_{i_k-1}(x_{k_1-1})\Big)}_{=1}\Big(\sum_{x_{i_k}\in J_{i_k}} p_{i_k}(x_{k_1})\Big)\underbrace{\Big(\sum_{x_{i_k+1}\in E_{i_k+1}} p_{i_k+1}(x_{k_1+1})\Big)}_{=1}\cdots \\
& \underbrace{\Big(\sum_{x_d\in E_d} p_d(x_d)\Big)}_{=1} \\
= \quad & \Big(\sum_{x_{i_1}\in J_{i_1}} p_{i_1}(x_{i_1})\Big)\cdots\Big(\sum_{x_{i_k}\in J_{i_k}} p_{i_k}(x_{k_1})\Big) = \mathbb{P}\left("\omega_{i_1}\in J_{i_1}"\right)\cdots\mathbb{P}\left("\omega_{i_k}\in J_{i_k}"\right).
\end{aligned}
$$

since $\mathbb{P}\left("\omega_i \in J_i"\right) = \sum_{x_i\in J_i} p_i(x_i)$, $i\in\{1,\ldots,d\}$, see first part of the proof. □

In the previous theorem, the requirement

$$\sum_{y \in E_i} p_i(y) = 1,\ i \in \{1, \ldots, d\}. \tag{9}$$

can be dropped.

In fact, if $p : \Omega \to [0, +\infty)$ such that $\sum_{x \in \Omega} p(x) = 1$ can be factorized as

$$p(x) = q_1(x_1)\, q_2(x_2) \cdots q_d(x_d),\ x = (x_1, x_2, \ldots, x_d) \in \Omega,$$

where $q_i : E_i \to [0, +\infty)$, $i \in \{1, \ldots, d\}$, is such that

$$\sum_{y \in E_i} q_i(y) > 0$$

then, by setting, for $i \in \{1, \ldots, d\}$,

$$p_i(y) = \frac{q_i(y)}{c_i},\ y \in E_i,$$

where

$$c_i = \sum_{y \in E_i} q_i(y),$$

we have (9). Exercise. Prove this.

Moreover, for $x \in \Omega$, we have

$$\begin{aligned} p(x) &= q_1(x_1)\, q_2(x_2) \cdots q_d(x_d) \\ &= c_1 p_1(x_1)\, c_2 p_2(x_2) \cdots c_d p_d(x_d) \\ &= c_1 c_2 \cdots c_d p_1(x_1)\, p_2(x_2) \cdots p_d(x_d)\,. \end{aligned}$$

Exercise. By looking at

$$\sum_{x \in \Omega} p(x) \text{ and } \sum_{x \in \Omega} c_1 c_2 \cdots c_d p_1(x_1)\, p_2(x_2) \cdots p_d(x_d)$$

conclude that

$$c_1 c_2 \cdots c_d = 1$$

and then

$$p(x) = p_1(x_1)\, p_2(x_2) \cdots p_d(x_d)\,,\ x \in \Omega.$$

Exercise. Consider a function $p : \Omega \to [0, +\infty)$ which can be factorized as

$$p(x) = p_1(x_1)\, p_2(x_2) \cdots p_d(x_d)\,,\ x = (x_1, x_2, \ldots, x_d) \in \Omega,$$

where $p_i : E_i \to [0, +\infty)$, $i \in \{1, \ldots, d\}$, is such that

$$\sum_{x \in E_i} p_i(x) = 1.$$

In which manner is this function different from the function p in the previous theorem? Show that

$$\sum_{x \in \Omega} p(x) = 1.$$

So, in the previous theorem, one, and only one, between

$$\sum_{x \in \Omega} p(x) = 1 \text{ and } \sum_{y \in E_i} p_i(y) = 1,\ i \in \{1, \ldots, d\}\,,$$

can be dropped.

Exercise. This exercise asks to prove a viceversa of the previous theorem. Prove that if, for any $x_1 \in E_1, x_2 \in E_2, \ldots, x_d \in E_d$, the events

$$"\omega_1 = x_1", "\omega_2 = x_2", \ldots, "\omega_d = x_d"$$

are independent, then the function p, with which we define the probabilities of the elementary events, has the factorized form given above with

$$p_i(y) = \mathbb{P}("\omega_i = y"), \; y \in E_i \text{ and } i \in \{1, 2, \ldots, d\}.$$

As a consequence, we have the following fact: if for any $x_1 \in E_1, x_2 \in E_2, \ldots, x_d \in E_d$, the events

$$"\omega_1 = x_1", "\omega_2 = x_2", \ldots, "\omega_d = x_d"$$

are independent, then , for any $J_1 \subseteq E_1, J_2 \subseteq E_2, \ldots, J_d \subseteq E_d$, the events

$$"\omega_1 \in J_1", "\omega_2 \in J_2", \ldots, "\omega_d \in J_d"$$

are independent.

- In the case where all elementary events have the same probability (classical probability), the function p satisfies the factorization condition of the previous property.

  In fact

$$\mathbb{P}(x) = \frac{1}{|\Omega|} = \frac{1}{|E_1| \cdot |E_2| \cdot \cdots \cdot |E_d|} = p_1(x_1)p_2(x_2)\cdots p_d(x_d),\ x \in \Omega,$$

  where

$$p_i(y) = \frac{1}{|E_i|},\ y \in E_i \text{ and } i \in \{1, \dots, d\}.$$

  So, the result on the independence of events for the components of the outcome applies in the case of classical probability.

  This fact has been already observed in the case $d = 2$, when we considered the independence of two events.

Example. Consider

- *the experiment of flipping d coins (or flipping d times a coin), where* $\Omega = \{H, T\}^d$ *and all the elementary events have the same probability* 2^{-d}*;*

  or

- *the experiment of rolling d dice (or rolling d times a die), where* $\Omega = \{1, \ldots, 6\}^d$ *and all the elementary events have the same probability* 6^{-d}*.*

Events for different coins or different dice are independent.

So, for the dice, the events "ω_i *is odd",* $i \in \{1, \ldots, d\}$*, are independent.*

- Now, we consider the case of a continuous sample space $\Omega = I_1 \times I_2 \times \cdots \times I_d$.

Theorem

If the integrable function $p : \Omega \to [0, +\infty)$ *such that* $\int\limits_{x\in\Omega} p(x) = 1$, *with which we define the probabilities of the closed boxes, can be factorized as*

$$p(x) = p_1(x_1)\, p_2(x_2) \cdots p_d(x_d)\,,\ x = (x_1, \ldots, x_d) \in \Omega,$$

where $p_i : I_i \to [0, +\infty)$, $i \in \{1, \ldots, d\}$, *is an integrable function such that*

$$\int\limits_{y\in I_i} p_i(y)\, dy = 1,$$

then, for any J_1 *Borel subset of* I_1, J_2 *Borel subset of* I_2, $\ldots$, J_d *Borel subset of* I_d, *the events*

$$"\omega_1 \in J_1", "\omega_2 \in J_2", \ldots, "\omega_d \in J_d"$$

are independent.

Exercise. Prove the previous theorem. Also show that the requirement

$$\int_{y\in I_i} p_i\left(y\right)dy = 1,\ i\in\{1,\ldots,d\},$$

can be dropped.

Exercise. Consider an integrable function $p:\Omega\to[0,+\infty)$ which can be factorized as

$$p\left(x\right) = p_1\left(x_1\right)p_2\left(x_2\right)\cdots p_d\left(x_d\right),\ x=\left(x_1,\ldots,x_d\right)\in\Omega,$$

where $p_i:I_i\to[0,+\infty)$, $i\in\{1,\ldots,d\}$, is an integrable functions such that

$$\int_{y\in I_i} p_i\left(y\right)dy = 1.$$

Show that

$$\int_{x\in\Omega} p(x) = 1.$$

Exercise. This exercise asks to prove a viceversa of the previous theorem. Prove that if, for any

$$[a_1, b_1] \subseteq I_1, [a_2, b_2] \subseteq I_2, \ldots, [a_d, b_d] \subseteq I_d$$

closed boxes, the events

$$"\omega_1 \in [a_1, b_1]", "\omega_2 \in [a_2, b_2]", \ldots, "\omega_d \in [a_d, b_d]"$$

are independent, then the function p with which we define the probabilities of the closed boxes has the factorized form given above with

$$p_i(y)dy = \mathbb{P}("\omega_i \in [y, y + dy]"), \; y \in I_i \text{ and } i \in \{1, 2, \ldots, d\}.$$

Exercise. Prove that the result on the independence of events for components of the outcome applies in the case of classical probability for continuous sample space.

Exercise. In the example of the falling meteor, prove that the events

$$\lambda \in J_1 \text{ and } \phi \in J_2,$$

are independent, where λ is the longitude of the impact point, ϕ is the latitude and $J_1 \subseteq (-\pi, \pi]$ and $J_2 \subseteq \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ are Borel subsets.

The Bernoulli process

- A **Bernoulli process** of length *n*, where *n* is a positive integer, is a composite experiment, where we repeat *n* times a basic experiment, called a **trial**, with two possible outcomes α and β.

  It is assumed that the outcome of each trial is independent of the outcomes of the other trials and that, in any trial, the outcome α is obtained with probability *p* and then β is obtained with probability $q := 1 - p$.

  The sample space of the Bernoulli process is $\Omega = \{\alpha, \beta\}^n$.

The probabilities of the elementary events are:

$$\begin{aligned}\mathbb{P}(x) &= \mathbb{P}\left(\bigcap_{i=1}^{n}\{\omega \in \Omega : \omega_i = x_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(\{\omega \in \Omega : \omega_i = x_i\}) \\ &= p^k q^{n-k},\ x = (x_1, \dots, x_n) \in \Omega,\end{aligned}$$

where k is the number of occurences of α in $x = (x_1, \dots, x_n)$.

In case of $p = q = \frac{1}{2}$, all the elementary events have the same probability

$$\mathbb{P}(x) = \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = 2^{-n},\ x \in \Omega.$$

Exercise. Write all the elements of Ω and their probabilities for a Bernoulli process of length $n = 2$. Moreover, show that the sum of the probabilities is 1.

Exercise. Do the same as in the previous exercise for $n = 3$.

Exercise. Show that the function

$$x \mapsto p^k q^{n-k}, \; x \in \Omega,$$

giving the probabilities of the elementary events for a general Bernoulli process of length n, with k the number of occurences of α in $x = (x_1, \dots, x_n)$, has the factorized form discuss in the previous section on the independence of events.

Exercise. By using the factorized form shown in the previous exercise, prove that the sum of the probabilities of the elementary events for a general Bernoulli process of length n is 1.

Practical examples of Bernoulli processes:

- *A regular coin is flipped n times (or n regular coins are flipped). The trials are the single throws with possible outcomes $\alpha = H$ and $\beta = T$. We have $p = q = \frac{1}{2}$.*
- *Consider a couple of parents that decides to have n children. The trials are the single births of the children with outcome the gender of the child: the possible outcomes are $\alpha =$ girl and $\beta =$ boy. We have $p = q = \frac{1}{2}$.*
- *Consider n commercial airplane flights. The trials are the single flights with possible outcomes $\alpha =$"the flight crashes with victims" and $\beta =$"the flight does not crash with victims". We can assume $p = 10^{-7}$.*
- *Consider a machine producing n pieces. The trials are the production of the single pieces with possible outcomes $\alpha =$"the piece is defective" and $\beta =$"the piece is not defective". We can assume $p = 1\%$ for a typical machine.*
- *Attempting n times the In Vitro Fertilization (IVF). The trials are the single attempts with possible outcomes $\alpha =$"the attempt has success and the woman becomes pregnant" and $\beta =$"the attempt has not success". We can assume $p = 20.5\%$ for a woman of age $40 - 42$ (a typical woman attempting IVF).*

- Now, we are interested in

$$\mathbb{P}\left(\text{"at least one outcome of the trials is } \alpha\text{"}\right).$$

We have

$$\mathbb{P}\left(\text{"at least one outcome of the trials is } \alpha\text{"}\right) = 1 - q^n.$$

In fact

$$\begin{aligned}
&\mathbb{P}\left(\text{"at least one outcome of the trials is } \alpha\text{"}\right) \\
&= \mathbb{P}\left(\text{"there exists } i \in \{1, 2, \ldots, n\} \text{ such that } \omega_i = \alpha\text{"}\right) \\
&= \mathbb{P}\left(\{(\beta, \beta, \ldots, \beta)\}^c\right) \\
&= 1 - \mathbb{P}\left((\beta, \beta, \ldots, \beta)\right) \\
&= 1 - q^n.
\end{aligned}$$

Examples:

- *The probability of having at least one occurence of H in n flips of a regular coin (or in flipping n regular coins) is* $1 - q^n = 1 - 2^{-n}$.
- *If the couple di parents decides to have n children, the probability of having at least one girl is* $1 - q^n = 1 - 2^{-n}$.
- *The probability to have at least one defective piece on* $n = 100$ *pieces produced by the machine is*

$$1 - q^n = 1 - 0.99^{100} = 63\%.$$

- *The probability to have a success on* $n = 3$ *attempts of IVF (which is the maximum number of attempts paid by the italian national health system) is*

$$1 - q^n = 1 - 0.795^3 = 49\%.$$

Exercise. Find in a Bernoulli process the minimum number *n* of trials such that

$$\mathbb{P}\,(\text{"at least one outcome of the trials is } \alpha\text{"}) \geq C\%$$

as a function of *p* and $C\% \in [0, 1]$.

Exercise. How many times one has to flip a regular coin to be sure with probabilities greater or equal to 50%, 90% and 99% that "H" will appear?

Exercise. How many pieces the machine has to produce to be sure with probabilities greater or equal to 50%, 90% and 99% that at least one defective piece will appear?

- When $np \ll 1$, we have

$$\mathbb{P}\left(\text{"at least one outcome of the trials is } \alpha\text{"}\right) \approx np.$$

In fact, we have the Taylor expansion

$$\begin{aligned} q^n = (1-p)^n =: f(p) &= f(0) + f'(0)p + \frac{1}{2}f''(\xi)p^2 \\ &= 1 - np + \frac{1}{2}n(n-1)(1-\xi)^{n-2}p^2, \end{aligned}$$

where $\xi \in [0, p]$ and

$$0 \leq \frac{1}{2}n(n-1)(1-\xi)^{n-2}p^2 \leq \frac{1}{2}(np)^2.$$

So, if $(np)^2 \ll np$, i.e. $np \ll 1$, then the third term in the Taylor expansion has order of magnitude smaller than the second one and we have $q^n \approx 1 - np$ and then

$$\mathbb{P}\left(\text{"at least one outcome of the trials is } \alpha\text{"}\right) = 1 - q^n \approx 1 - (1 - np) = np.$$

Example. The probability of having a crash with victims in n commercial flights is approximately $np = n \cdot 10^{-7}$ when $n \ll 10^7$.

Exercise. Give an estimate of the probability of taking a commercial flight which will crash with victims during the life of a person.

Exercise. To win the Jackpot (maximum payout) in the italian lottery "Superenalotto", one has to guess exactly a randomly selected subset of $\{1, 2, \ldots, 90\}$ with six elements.

- Determine the probability to win when one plays a system encompassing 1000 combinations (i.e. 1000 different subsets of $\{1, 2, \ldots, 90\}$). The cost of the system is 500 Euros.
- If one plays this system at every weekly draw for fifty years, what is the probability of win at least one time?

- Exercise. Prove that when p is small and np is not large, then

$$\mathbb{P}\left(\text{"at least one outcome of the trials is } \alpha\text{"}\right) \approx 1 - e^{-np}.$$

Use $(1-p)^n = e^{n \log(1-p)}$ and the Taylor expansion of $\log(1-p)$ around $p = 0$.

Exercise. Find the probability of having a crash with victims in $n = 10^7$ commercial flights.

The Bayes' Theorem

In the Bayesian interpretation of the probability, the probabilities of events are interpreted as degrees of belief about the events. These degrees of belief are subjective and they depend on the information one has about the experiment.

If new information now become available, how can we update the degrees of beliefs of the events?

Suppose that the new available information permits to say that a certain event A occurs. We can update the measure of probability to the conditional measure of probability given A.

- Consider an experiment with measure of probability $\mathbb{P}$ and let A be an event. Consider the conditional measure of probability $\mathbb{P}(\cdot|A)$ given A.

  We are interested in the relation between $\mathbb{P}$ called **prior probability** and $\mathbb{P}(\cdot|A)$ called **posterior probability**. We have

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A)}, \; B \text{ event.}$$

  We can also write

$$\mathbb{P}(B|A) \propto \mathbb{P}(A|B)\,\mathbb{P}(B), \; B \text{ event.}$$

  where $\propto$ means proportionality: the constant of proportionality independent of B is $\frac{1}{\mathbb{P}(A)}$. The function

$$B \mapsto \mathbb{P}(A|B), \; B \text{ event}$$

  is called **likelihood** of A. Exercise. Is the likelihood of A a measure of probability for the experiment?

- The Bayes' Theorem is a reformulation of the proportionality relation between the prior probability and the posterior probability.

Theorem

Consider an experiment of sample space Ω*. Let* A *be an event and let* $B_1, B_2, \ldots, B_n$ *disjoint events such that* $\bigcup_{i=1}^{n} B_i = \Omega$*. Given the prior probabilities*

$$\mathbb{P}(B_1), \mathbb{P}(B_2), \ldots, \mathbb{P}(B_n)$$

of the events $B_1, B_2, \ldots, B_n$ *and the likelihood values*

$$\mathbb{P}(A|B_1), \mathbb{P}(A|B_2), \ldots, \mathbb{P}(A|B_n),$$

the posterior probabilities of the events $B_1, B_2, \ldots, B_n$ *are given by*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}, \quad i \in \{1, \ldots, n\}. \ (Bayes'\ Formula)$$

Proof.

We have

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\mathbb{P}(A)},\ i \in \{1, \ldots, n\}.$$

Since

$$1 = \mathbb{P}(\Omega|A) = \mathbb{P}\left(\bigcup_{j=1}^{n} B_j|A\right) = \sum_{i=1}^{n} \mathbb{P}(B_j|A),$$

we obtain

$$\begin{aligned} 1 &= \sum_{j=1}^{n} \mathbb{P}(B_j|A) = \sum_{j=1}^{n} \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\mathbb{P}(A)} \\ &= \frac{1}{\mathbb{P}(A)} \sum_{j=1}^{n} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j) \end{aligned}$$

and then

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j).$$

- A prominent case is when $n = 2$ with $B_1 = B$ and $B_2 = B^c$.

  We obtain

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)}.$$

  and, of course,

$$\mathbb{P}(B^c|A) = 1 - \mathbb{P}(B|A).$$

- *Example. Consider a given disease and let p be the percentage of the individuals with the disease in the population (p is called the **prevalence** of the disease). Suppose* $p = 0.5\%$.

  *Consider a diagnostic test, e.g. a blood test, performed on an individual for detecting the disease. The test has two possible results: **positive**, i.e. the individual has the disease, and **negative**, i.e. the individual has not the disease.*

  Let s_1 *be the percentage of the individuals with the disease having a positive test (*s_1 *is called the of the **sensitivity** of the test). Suppose* $s_1 = 99\%$.

  Let s_2 *be the percentage of the individuals without the disease having a negative test (*s_2 *is called the of the **specificity** of the test). Suppose* $s_2 = 98\%$.

Exercise. What are the percentages of false positive and false negative? Is it better to have a larger number of false negative or a larger number of false positive?

Consider the experiment where the test is performed on a given individual. If the individual has a positive result, she/he will be tested by a more accurate test (but more costly) for the certainty of the disease.

Consider the event

$$B = \text{"the individual has the disease"}$$

We can assume $\mathbb{P}(B) = p = 0.2\%$.

Now suppose that the event

$$A = \text{"the individual has positive test"}$$

occurs. How can we update the probability of the event B*?*

By using the likelihood values

$$\mathbb{P}(A|B) = s_1 = 99\%,\ \mathbb{P}(A|B^c) = 1 - \mathbb{P}(A^c|B^c) = 1 - s_2 = 2\%,$$

Bayes' theorem says that

$$\begin{aligned}\mathbb{P}(B|A) &= \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)} \\ &= \frac{s_1 p}{s_1 p + (1 - s_2)(1 - p)} \\ &= \frac{99\% \cdot 0.5\%}{99\% \cdot 0.5\% + 2\% \cdot 99.5\%} = 19.9\%.\end{aligned}$$

So, with the additional information that the individual has positive test, the probability that she/he has the disease passes from 0.2% *to* 19.9%.

Exercise. Fixed a percentage $C\%$, which relation have to satisfy s_1 and s_2 in order to have $\mathbb{P}(B|A) \geq C\%$. Describe in the plane the region given by the points (s_1, s_2) satisfying this relation.

Exercise. For the disease AIDS, find on internet the prevalence of the disease and the sensitivity and the specificity of the tests used for detecting it. Then compute $\mathbb{P}(B|A)$.

Exercise. Describe the sample space Ω for this experiment and find the probabilities of the elementary events.

Exercise. Determine the probability that the individual has not the disease given that she/he has negative test.

- *Example An insurance company divided people into three classes* $1, 2$ *and* 3 *of increasing possibility to have an accident.*

  It assumes that the percentages of people in classes $1, 2$ *and* 3 *are* p_1, p_2 *and* p_3, *respectively.*

  Moreover, it also assumes that the probabilities for people in classes 1 *and* 2 *and* 3 *to have an accident in 1-year period are* q_1, q_2 *and* q_3, *respectively, with* $q_1 < q_2 < q_3$.

  Suppose

$$p_1 = 80\%,\ p_2 = 18\%,\ p_3 = 2\%$$

  and

$$q_1 = 5\%,\ q_2 = 10\%,\ q_3 = 20\%.$$

Consider the experiment where a new policyholder is monitored in the first year.

Consider the events

$$B_i = \text{"the policyholder is in class } i\text{"},\ i \in \{1, 2, 3\}.$$

We can assume

$$\mathbb{P}(B_i) = p_i,\ i \in \{1, 2, 3\}.$$

Now suppose that the event

$$A = \text{"the policyholder has an accident"}$$

occurs. How can we update the probability of the events B_i, $i \in \{1, 2, 3\}$?

By using the likelihood values

$$\mathbb{P}(A|B_i) = q_i,\ i \in \{1, 2, 3\}$$

Bayes' theorem says that

$$\mathbb{P}(B_1|A) = \frac{q_1 p_1}{q_1 p_1 + q_2 p_2 + q_3 p_3} = \frac{5\% \cdot 80\%}{5\% \cdot 80\% + 10\% \cdot 18\% + 20\% \cdot 2\%} = 65\%$$
$$\mathbb{P}(B_2|A) = \frac{q_2 p_2}{q_1 p_1 + q_2 p_2 + q_3 p_3} = \frac{10\% \cdot 18\%}{5\% \cdot 80\% + 10\% \cdot 18\% + 20\% \cdot 2\%} = 29\%$$
$$\mathbb{P}(B_3|A) = \frac{q_3 p_3}{q_1 p_1 + q_2 p_2 + q_3 p_3} = \frac{20\% \cdot 2\%}{5\% \cdot 80\% + 10\% \cdot 18\% + 20\% \cdot 2\%} = 6\%.$$

So, with the additional information that the new policyholder has an accident in the first year, the probabilities that she/he stays in the classes $1, 2$ *and* 3 *passes from* 80%, 18% *and* 2% *to* 65%, 29% *and* 6%*, respectively.*

Exercise. Describe the sample space Ω *for this experiment and find the probabilities of the elementary events.*

Exercise. Suppose that the new policyholder is monitored for another year and she/he has again an accident. What are the new probabilities for staying in classes 1, 2 *and* 3*?*