1 Introduction

Probability Theory is a mathematical theory whose aim is to quantify uncertainity. We want to make sense of statements such as

- The probability that tomorrow will rain is 0.75.

- The probability that the the next time that I will toss a coin, the result will be heads is 1/2.

The intuitive meaning is clear: in the first example tomorrow we would bring with us an umbrella, while in the second example we are saying that it is equally probable that the next time that I will toss a coin the result will be heads or tails.
We will come back in this introduction to the meaning of probability, and now we focus on the structure of the above examples. We say that E=" Tomorrow will rain" and F=" The next time that I will toss a coin, the result will be heads" are events. Having defined E and F, the above two probability statements can be rewritten as

- The probability of E is 0.75

- The probability of F is 0.5

Using the standard mathematical notation, the two above statements can be rewritten as $\mathbb{P}(E) = 0.75$ and $\mathbb{P}(F) = 0.5$. This is the general structure of probability statements

- The probability of the event E is p, where p is a number between 0 and 1, or, in mathematical notation, $\mathbb{P}(E) = p$.

Concretely, the probability of an event E is a number denoted by $\mathbb{P}(E)$ between 0 and 1 attached to each event. It is intuitive that it quantifies the uncertainity we have on events, in the sense that if $\mathbb{P}(E) = 0.6 < \mathbb{P}(F) = 0.8$, we regard the event F more likely than the event E. We note that the probability is not unique and it depends on the individual who is deciding it and on its state of information. As a simple example, the probability that I would assign to the event E="Real Madrid will win the Champions League this year" would be 1/16 if I only knew that there are 16 teams left and between them there is Real Madrid and had no information of any

kind about football. Since I sometimes watch football, the probability that I assign to the event E is greater than 1/16, say 1/3. It would decrease if I knew that Benzema is not in shape.

A possible definition of the probability of an event E, that is, of the number $\mathbb{P}(E)$ is as the degree of belief an individual(the one assigning the probability) has on the event E.

The first notion that we need to precise is the one of event

1.1 Definition of an event

For us an event something defined by an unambiguous proposition which can be either true or false (but you are uncertain whether they are true or false). By unambiguous we informally mean that a possible bet or insurance based upon it can be decided without question. Examples of events are: " Nadal will win the next Australian Open ", " Democrats will win the next USA elections ", a given shop this year will sell more products than last year... We note that even if events are stated in an unambiguous way, they are usually unknown to you or you are uncertain about them (and a probability will be just a quantitative version of this uncertainty). In order to make a more concrete example to which we will come back later, consider a given shop. You can consider the following events "Tomorrow 1 person will enter to the shop", "Tomorrow 1 woman will enter to the shop", " Tomorrow the number of people that will enter will be less or equal than 30 and more or equal than 5"...

1.2 Operation and relations on events

From two events E and F you can build the events "E and F", "E or F" and not F. In order to make an example, consider again the example of the shop. Define E=" Tomorrow only 1 person will enter in the shop ", F=" Tomorrow it will rain", G="More than 10 people will enter in the shop tomorrow".

- "E and F"= " Tomorrow only one person will enter in the shop and it will be raining ".

- "E or G"= "Tomorrow either more than 10 persons or only one person will enter in the shop".

- "not F "= " Tomorrow it won't be raining".

Given two events E or F we say that E implies F if, given that E occurs, then F occurs. Note that it can very well happen that neither E implies F nor F implies E. As an example, let E, F and G be defined as above. Then E implies "$notG$", E does not imply F and F does not imply G.

Exercise 1:
Convince yourself about the previous statements

1.3   Elementary Events and Sample Space

When we want to do a probabilistic analysis, we don't want to take into account every event. To be more concrete, if we are interested in the result of a coin toss we are interested in the event E="The result of the coin toss is tails" but not on the event "The result of the coin toss is tails and tomorrow it will rain in Rome".
Coming back to the example of the shop, you could be interested in predicting the probability of the number of people that will enter in the shop tomorrow. In this case you can consider events of the form "The total number of people that will enter in the shop tomorrow is more or equal than 10 and less or equal than 100", " Tomorrow 20 people will enter in the shop", ... But you could also want to make a finer analysis by distinguishing the gender of the people entering in the shop tomorrow. In this case you want to to consider events of the form "The number of women entering in the shop tomorrow will be less than 20 and the man entering in the shop tomorrow will be more than 30", and so on.

Definition 1.1. *A choice of events (on which we wish to make a probabilistic analysis) is called (in this notes) a system.*

Once that you fixed a system, that is, once you have chosen the events you want to consider, you can distinguish a special kind of events, the *elementary events*. Those are the events that you cannot subdivide more. By subdivide an event E into F and G we mean that $E = F$ or G.

Definition 1.2. *Fix the events you want to consider. The elementary events are those events which cannot be subdivided into other events you want to*

consider. More precisely, the event E is elementary if E cannot be written as E = "A or B", for A and B events that you want to consider.

For the sake of concreteness, let's come back to the example of the shop. Imagine that you only want to make a probabilistic analysis of the number of people entering in the shop tomorrow, disregarding the gender. We have already seen the events that we want to consider in this case. The event E="The total number of people that will enter in the shop tomorrow is more or equal than 10 and less or equal than 100 " can be subdivided, for instance, into E="F or G", where F="The total number of people that will enter in the shop tomorrow is more or equal than 10 and less or equal than 50", and G="The total number of people that will enter in the shop tomorrow is strictly more than 50 and less or equal than 100", so that it is not elementary. On the contrary, the event H=" Tomorrow precisely 30 people will enter in the shop " cannot be subdivided into other events that we want to take into account. We don't want to consider the event "Tomorrow exactly 30 people will enter in the shop and it will snow". H is an elementary event. If instead we take into account the gender of the people entering, H could be decomposed in smaller events such as L=" Tomorrow exactly 20 man and 10 woman will enter in the shop". In this new system, L will be an elementary event.

Example 1.3 (Two dice rolled)**.** *Two dice have been rolled, and we don't know their result. First assume that you can only look at the sum of the results. Then events are of the form*

- *"The sum is an odd number", "The sum is greater than 8",...*

and the elementary events are

- *"The sum is 2"*

- *" The sum is 3"* ⋮

- *"The sum is 12",*

For example, we don't want to distinguish inside the event " The sum is 6 and Nadal will win the next tennis tournament".

Assume now that we can look at the result of each single die. Now the previosly elementary event " The sum is 3" can be decomposed into the elementary events " The first die gave 1 and the second 2 " and the event "The first die gave 2 and the second 1". In this case the elementary events are

- *" The first die gave 1 and the second 1"*

- *" The first die gave 1 and the second 2" $\vdots$*

- *" The first die gave 1 and the second 6"*

- *" The first die gave 2 and the second 1 " $\vdots$*

- *" The first die gave 6 and the second 6"*

Elementary events are also called outcomes of an experiment. In the shop example, you can think that you are counting the number of people entering through some kind of detector placed above the entrance. The number give to you by the detector is the elementary event. With this interpretation, a system is also called an experiment.

We end this section with an important definition

Definition 1.4 (Sample Space). *Assume you want to do a probabilistic analysis of a system, so that you have chosen the events of which you want to calculate the probability and consequently you have defined the elementary events. The set of elementary events is denoted by Ω and is called the sample space.*

If you like to think at elementary events as the outcomes of the experiments and to probability systems as experiments, then the sample space Ω can be thought as the set of outcomes of the experiment

As we will see, in the text of the exercises it is the sample space which is usually given to you. In the previous example we have recognized the elementary events, so that we just need to place them together to obtain the sample space:

Example 1.5 (Two dice have been rolled). *If we are looking only at the sum of the two dice, the sample space is*

$$\Omega = \{ \text{" The sum is 2", " The sum is 3", ..., The sum is 12"} \},$$

while if we are looking at the result of each single dice

$$\Omega = \{ \text{ " The first die gave 1 and the second 1", " The first die gave 1 and the second 2",} \\ \text{...." The first die gave 6 and the second 6"} \}$$

Example 1.6 (A chess match). *Assume that you are observing a chess match. A full description of the game is given by describing the moves of each pleyer untill the king is killed. Usually this description is given using what is called algebraic notation. An example of a match is given by (It is a Fisher vs. Kasparov match)*
1.d4Nf6 2.c4e6 3.Nc3Bb4 4.Nf3c5 5.e3Nc6 6.Bd3Bxc3+ 7.bxc3d6 8.e4e5 9.d5Ne7
10.Nh4h6 11.f4Ng6 12.Nxg6fxg6 13.fxe5dxe5 14.Be3b6 15.O-OO-O 16.a4a5
17.Rb1Bd7 18.Rb2Rb8 19.Rbf2Qe7 20.Bc2g5 21.Bd2Qe8 22.Be1Qg6 23.Qd3Nh5
24.Rxf8+Rxf8 25.Rxf8+Kxf8 26.Bd1Nf4 27.Qc2Bxa40–1
The above is an elementary event of the system "A chess match between Player1 and Player 2". A game easier than chess will be the object of an exercise at the end of this lecture. Note that describing the event "The white will win" in terms of elementary events is a task too difficult to be made.

1.4   Exercises

Ex. 1 — From the 2019-2020 course " Mathematics and Statistics" given by professors Khovanskaya, Bubilin, Shurov, Filimonov and Sonin at HSE. A six faced die is rolled. Find the elementary events that compose the event

1. "The result is 6"
2. "The result is a number less or equal 2"
3. "The result is an even number"
4. "The result is a number strictly greater than 4"
5. "The result is seven"

Answer (Ex. 1) —   1. It is already an elementary event

2. "The result is a number less or equal than 2"= "The result is 1" or "The result is 2".

3. "The result is a number strictly greater than 4" = "The result is 5" or "The result is 6 ".

4. The event is the impossible event, also denoted by $\emptyset$.

Ex. 2 — A coin is tossed twice. We are interested in the face which the coin shows when it lands: heads or tails. The set of elementary events is $\{HH, HT, TH, TT\}$, where the event HH corresponds to "the first toss gave heads, the second heads", the event HT corresponds... Write the elementary events that compose the events

1. "We obtained two heads".

2. "The first toss gave heads".

3. "We obtained 1 tails "

4. "At least one toss gave tails"

Answer (Ex. 2) —   1. HH.

2. TH, TT.

3. TH, HT.

4. HT, TH, TT.

Ex. 3 — A coin is tossed 4 times. We are interested in the face that it shows when it lands: heads or tails. How many elementary events are there? which elementary events are contained in the events

1. The first result was heads

2. The second result was tails

3. The first result was heads and the second tails

4. All the 4 tosses gave the same result

Answer (Ex. 3) — The elementary events are HHHH, HHHT,HHTH,....,TTTT. There are $2^4 = 16$ of them.

1. HHHH, HHHT,....,HTTT

2. HTHH,HTHT,HTTH,...,TTTT

3. HTHH, HTHT,HTTH,HTTT

4. HHHH, TTTT

Ex. 4 — ([**Ross**] Chapter 2 Exercise 1) A box contains 3 marbles: 1 red, 1 green, 1 blue. Consider the experiment that consists in taking 1 marble from the box and then replacing it in the box and then drawing a second marble. Describe the sample space.

Answer (Ex. 4) — The sample space can be thought as the set of the outcomes of the experiment. In this case the possible outcomes are "The first marble is red and the second marble is red", "The first marble is red and the second is green",... A more efficient way to write an outcome is to write an array with two entries, i.e, (r,g), where the first entry represents the colour of the firs marble, and the second entry represents the colour of the second marble. That is, we write the elementary event "The first marble is red and the second is blue" by (r,b). The sample space becomes

$$\Omega = \{(r,r),(r,g),(r,b),(g,r),(g,g),(g,b),(b,r),(b,g),(b,b)\} \qquad \boxed{\text{ex1}}$$

Ex. 5 — Repeat Exercise 1 when the second marble is drawn without replacing the first marble.

Answer (Ex. 5) — The outcomes (r,r), (g,g) and (b,b) are no longer possible. The sample space is

$$\Omega = \{(r,g),(r,b),(g,r),(g,b),(b,r),(b,g)\} \qquad \boxed{\text{ex2}}$$

Ex. 6 — You toss a coin 3 times. Describe the elementary events.

Answer (Ex. 6) — $\Omega = \{(H,H,H),(H,H,T),....,(T,T,T)\}$, where the event (H,T,H) is "The result of the first toss was head, the result of the second...".

1.5 A graphical way to think about the sample space

A way to visualize the sample space $\Omega = \{\omega_1,...,\omega_n\}$ is by drawing a portion of area in the plane as Ω, and subdivide it in n disjoint regions such as in the picture.

Suppose that now we want to make a finer analysis in our system, and that now each ω_i is further divided into the new elementary events ω_i'. By performing this division on the previous image we obtain the new sample space division of the sample space.

One can think at the point (x, y) of the plane where Ω is pictured as being limiting points that you can obtain from elementary events after making further and further refinements.

1.6   Examples

The aim of the next examples is to introduce set once for all a convenient notation for the examples that we will be treating during the course.

Example 1.7 (A dice has been rolled). *The elementary events are " The result is 1 ",..." The result is 6". We use the shortcut 1=" The result is 1" ,...., 6=" The result is 6". The sample spase becomes $\Omega = \{1, 2, 3, 4, 5, 6\}$. Note that you don't have to consider 1,2,3,4,5,6 as numbers, but as events. Concretely 1+1=2 makes no sense since "The result is 1"+ "The result is 2 " makes no sense.*

Example 1.8 (Two dice rolled). *The sample space is $\Omega = \{(1, 1), (1, 2), ..., (1, 6), (2, 1), ...(6, 6)\} = \{(i, j), i, j = 1, ..., 6\}$, where the (i, j) is a shortcut for the elementary event "the first die gave i and the second j". The event "The sum of the results is 5" corresponds to the subset of Ω $E = \{(1, 4), (2, 3), (3, 2), (4, 1)\} = \{(i, j) \in \Omega, i + j = 8\}$.*

Example 1.9 (A coin toss). *In a coin toss there are only two elementary events: 0=" The result is tails" and 1 = " The result is heads". $\Omega_1 = \{0, 1\}$. We will often use the word succes instead of the word heads.*

Example 1.10 (two, three and n coin tosses). *Assume that we are looking at two coin tosses. Then $\Omega_2 = \{00, 01, 10, 11\} = \{x_1 x_2, x_i \in \{0, 1\} i = 1, 2\}$, where, for example, 01="The first toss gave tails and the second tails". For three tosses, $\Omega_3 = \{000, 001, 010, 011, 100, 101, 110, 111\} = \{x_1 x_2 x_3, x_i \in \{0, 1\} i = 1, 2, 3\}$. The event E= "There has been exactly one succes" is $E = \{001, 010, 100\} = \{x_1 x_2 x_3, x_1 + x_2 + x_3 = 1\}$.*
In general, the sample space of n coin tosses is $\Omega_n = \{x_1 x_2 ... x_n, x_i \in \{0, 1\} i = 1, ..., n\}$, and the event $E_k^n = $ " There have been k successes" is $E_k^n = \{x_1 ... x_n \in \Omega_n, x_1 + ... + x_n = k\}$.

Example 1.11 (finite but undetermined amount of coin tosses). *Suppose that we are observing some coin tosses but we are uncertain also on how many of them there will be. The sample space is $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$ where the in Ω_i have been defined previously.*

Example 1.12 (Infinite coin tosses). *A non-trivial generalization of n coin tosses is the one of an infinite amount of coin tosses. In this case the sample space is $\tilde{\Omega} = \{x_1...x_n..., \ x_i \in \{0,1\}\}$. In this case you know that the amount of times that you flip the coin is infinite.*

Example 1.13 (Sampling). *Suppose that a sample of 100 people is taken in order to estimate how many people smoke. If you are only interested in the number of smokers, disregarding age, gender, etc., the sample space is $\Omega = \{0, 1..., 100\}$, where, for instance, 1 means the event "There is exactly 1 smoker". The event "The majority of people sampled are smokers" is composed by the simple events 51,52,...100.*

1.7 Events as subsets of the sample space

Up to now what we said was an operative way to define your sample space in terms of the events of which you are interested in doing a probabilistic analysis. However, the mathematical theory needs only the sample space as a primitive notion. The other notions are based on Ω, included the one of event. Indeed the previous examples show that an event can be represented as the set of elementary events that compose it, that is, an event is a subset of the sample space. When we roll a dice, the event "The result is odd" can be written as $\{1, 3, 5\}$.

Definition 1.14. *Let Ω be the sample space. An E subset of Ω is said to be an event. We also consider the impossible event $\emptyset$ as an event.*

For example, the sample space of a three faced die is $\Omega = \{1, 2, 3\}$, and the events are $\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$. The event $\{1, 3\}$ can be written as a proposition as "The result is an odd number". Given a set, the set of his subsets is called the power set and is denoted. Using this notion, we could say that the set ov events is the power set of the sample space. A nice exercise is to prove that if Ω has n elements, the power set has 2^n elements.

To every proposition we associate a subset of the sample space, as we did with the event "The resut is an odd number", but the analogy does not end here. The operations "or", "and" "not", can be traduced in terms of $\cup$, $\cap$ and ^{c}. Consider the sample space $\Omega_3 = \{000, 001, ..., 111\}$, and consider the events E="The number of successes? is less than one" and F= " The number of successes is more than 1". Then $E = \{000, 001, 010, 100\}$ and $F = \{001, 010, 100, 011, 101, 110, 111\}$, and the event G=" E and F"=" The number of successes is exactly 1" is simply given by $G = E \cap F = \{000, 001, 010, 100\} \cap \{001, 010, 100, 011, 101, 110, 111\} = \{001, 010, 100\}$. Note also that K=" not E"=" The number of successes is strictly more than 1", and that, indeed, $K = \{011, 101, 110, 111\} = \{000, 001, 010, 100\}^c = \Omega_3 \setminus \{000, 001, 010, 100\}$.

Ex. 7 — ([**Feller**], Chapter 1.8) Let Ω be a sample space and A, B, C be three arbitrary events. Find the expressions for the events that of A, B, C:

1. Only A occurs.
2. Both A and B, but not C occur.
3. All three events occur.
4. At least one occurs.
5. At least two occur.
6. One and no more occurs.
7. Two and no more occur
8. None occurs.
9. Not more than two occur.

Answer (Ex. 7) —   1. $A \setminus (B \cup C) = A \cap B^c \cap C^c$.
2. $(A \cap B) \setminus C = A \cap B \cap C^c$.
3. $A \cap B \cap C$.
4. $A \cup B \cup C$.

ex:set1

5. At least two occur if occur one bertween $A \cap B$, $A \cap C$ or $B \cap C$. Thus $(A \cap B) \cup (A \cap C) \cup (B \cap C)$.
6. It has to happen one between A B and C, that is, it has to happen $A \cup B \cup C$, but it does not have to happen two of them, that is, it cannot happen the event described in 5:
7. $(A \cap B) \cup (A \cap C) \cup (B \cap C)$. Thus the event is $A \cup B \cup C \setminus ((A \cap B) \cup (A \cap C) \cup (B \cap C))$.

8. $((A \cap B) \cup (A \cap C) \cup (B \cap C)) \setminus A \cap B \cap C$ which is the same as $((A \cap B) \cup (A \cap C) \cup (B \cap C)) \cap (A^c \cup B^c \cup C^c)$.

9. $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$.

10. This event happens when the event all three events occur does not happen. Thus the event is $(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$.

2 Exercises

Ex. 8 — ([**Ross**] Chapter 2 Exercise 2) In an experiment, die is rolled continually until a 6 appears, at which point the experiment stops. What is the sample space of this experiment? Let E_n denote the event " The experiment is completed before n rolls". Write it as a subset of the sample space. Describe $(\bigcup_{i=1}^{\infty} E_i)^c$.

Answer (Ex. 8) — The sample space is the space of outputs of the experiment. We encode the event "The first roll gave 4, the second 2, the third 6 and then stops" in the string 426. That is, we encode the output of the experiment in strings of digits 1,2...,5,6, where the digit correspond to the result of the die, and the position of the digit corresponds to when has the digit appeared. The string must stop at after the first 6, since the game stops when the die gives 6.

We have to keep in mind that we can observe events in which we never stop rolling the die, which correspond to infinite strings of 1,2...5, for example 1212121.... or 121233...34453236....

$\Omega = \{6, 16, 26, 36, 46, 56, 116, ...\} \cup \{x_1 x_2 .... x_n ... : x_i \in \{1, 2, 3, 4, 5\}\} = E \cup G.$

The set E is the event that the game stops in a finite number of rolls. The set G is the event that the game goes on without ending. For any reasonable probability, the event G has probability 0, and it is usually neglected. The events E_i can be written as subsets of the state space in the following way: $E_1 = \{6\}$, $E_2 = \{6, 16, 26..., 56\}$, $E_3 = E_2 \cup \{116, 126, ..., 556\}$,... (In a more compact form $E_i = \{x_1...x_n 6 : x_i \in \{1, ..., 5\} n \leq i\}$).
The event

$$\bigcup_{i=1}^{\infty} E_n = E_0 \cup E_1 \cup E_2, ... = "E_0 \ or E_1 \ or .... E_n \ or ..."$$

12

is, by definition of "or", the event that E_n happens for at least one n. That is, is the event "The game ends before n rolls at least one n "="The game ends after n rolls for some n", which coincides with E. The set event E^c corresponds to G.

Ex. 9 — ([**Ross**] Chapter 2 Exercise 3) Two dice are thrown. Let E be the event that the sum of the dice is odd, and F the event that at least one of the two dice lands in 1 and let G be the event that the sum is 5. Describe the events $E \cap F$, $E \cup F$, $F \cap G$, $E \cap F^c = E \setminus F$ and $E \cap F \cap G$

Answer (Ex. 9) —

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

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

$$G = \{(1,4), (2,3), (3,2), (4,1)\}$$

$$E \cup F = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,3), (2,5), (3,1), (3,2), (3,4)..(6,5)\}$$

$$E \cap F = \{(1,2), (1,4), (1,6), (2,1), (4,1), (6,1)\}$$

$$F \cap G = \{(1,4), (4,1)\}$$

$$E \cap F \cap G = F \cap G = \{(1,4), (4,1)\}$$

$$E \cap F^c = \{((2,3), (2,5), (3,2), (3,4), (3,6), (4,3), ...(6,5)\}$$

Ex. 10 — ([**Ross**] Chapter 2 Exercise 4) Anton, Barbara and Carlo take turns flipping a coin. The first one to get head wins. The sample space to this experiment can be defined as $\Omega = \{1, 01, 001, 0001, ...\} \cup \{00000...\}$.

1. Interpret the sample space.
2. In terms of the sample space, write the following events: A="Anton wins", B="Barbara wins" and $(A \cap B)^c$.

Answer (Ex. 10) — 1. 1="First toss is heads"(So that Anton wins) 01="The first toss is tails, the second heads "(So that Barbara wins)... The event 00000...="Head never arises, the game never stops".
2. As it is easy to seen, Anton wins whenever the digit 1 appears in the 1st, 4th, 7th.. position. This means that

$$A = \{1, 0001, 0000001, ...\}.$$

13

In the same way, the event that Barbara wins corresponds is

$$B = \{01, 00001, 00000001, ...\}$$

where the digit appears in the after $3 \times k + 1$ zeros, for some $k \in \mathbb{N}$. $(A \cup B)^c = \Omega \setminus (A \cup B) = \{001, 000001, ...\} \cup \{0000...\}$. The event $(A \cup B)^c$ is the event "Carlo wins or the game doesn't end"

3 Probability

Once we choose the events of which we are uncertain about and interest us, we want to quantify this uncertainty by introducing the probability. To each event E we want to assign a number $\mathbb{P}(E) \in [0, 1]$, denoted as the probability of the event E.

Note that when considering two (or more) events, say E and F, the numbers $\mathbb{P}(E)$ and $\mathbb{P}(F)$ are not arbitrary and have to satisfy some relations that ensure that the probability behaves as we expect to behave. Think about the events E=" 5 Star Movement will win next Italy elections "and F=" 5 Star Movement will win next Italy elections and on January 12 2023 it will rain in Rome". Then the numbers $\mathbb{P}(E)$ and $\mathbb{P}(F)$ are not just numbers in $[0, 1]$, but they also need to satisfy $\mathbb{P}(E) \geq \mathbb{P}(F)$, since if F happens then E happens. The relations that $\mathbb{P}(E)$ have to satisy when calculated on different events are the *Kosmogorov Axioms*.

We restrict our attention to finite sample spaces $\Omega = \{\omega_1, ..., \omega_n\}$. In this case we don't need the full generality of the Kolmogorov axioms, and we can define the probability of $E \subset \Omega$, $\mathbb{P}(E)$, in terms of the probability of the elementary events $\mathbb{P}(\omega_i)$. As for the meaning to give to the number $\mathbb{P}(E)$, it can be interpreted as the degree of belief a certain individual has on the event E. However, we don't discuss here how to measure or estimate it, and we imagine those numbers as given.

3.1 Definition on a finite Sample Space

Let $\Omega = \{\omega_1, ..., \omega_n\}$ be a finite sample space. A probability on the elementary events is a set of n numbers $\mathbb{P}(\omega_1)$, $\mathbb{P}(\omega_2)$,...,$\mathbb{P}(\omega_n)$ satisfying

$$\mathbb{P}(\omega_i) \in [0, 1] \text{ for every } i = 1, ..., n \qquad \boxed{\text{e:pos}}$$

and satisfying the followinf *fundamental assumption*

$$\mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_n) = 1. \tag{1} \quad \boxed{\texttt{e:fun}}$$

Example 3.1 (balanced dice). *Assume that we throw a balanced dice. Then the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$, and the hypotheses of the dice being balanced means that there is not favourable outcome: $\mathbb{P}(1) = 1/6$, $\mathbb{P}(2) = 1/6$, $\mathbb{P}(3) = 1/6$, $\mathbb{P}(4) = 1/6$, $\mathbb{P}(5) = 1/6$, $\mathbb{P}(6) = 1/6$.*

Example 3.2 (unbalanced dice). *Assume you throw a dice, so that the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. An possible probability for an unbalanced dice would be $\mathbb{P}(1) = 1/6$, $\mathbb{P}(1) = 1/6$, $\mathbb{P}(3) = 1/12$, $\mathbb{P}(4) = 3/12$, $\mathbb{P}(5) = 1/6$, $\mathbb{P}(6) = 1/6$,*

Definition 3.3. *Let $\Omega = \{\omega_1, ..., \omega_n\}$ be the sample space, and assume that the probabilities of the elementary events $\mathbb{P}(\omega_1),...,\mathbb{P}(\omega_n)$ have been assigned. The probability of an event E is the sum of the probabilities of the elementary events that compose it:*

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega). \tag{2} \quad \boxed{\texttt{d:prob}}$$

In a less obscure way, if $E = \{\omega_1, ..., \omega_k\}$, then $\mathbb{P}(E) = \mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_k)$.

Example 3.4 (The dice). *Let's calculate the probability of the event "The result is an odd number " with the two different probabilities chosen above. The event is $E = \{1, 3, 5\}$, and, in the balanced case, its probability is*

$$\mathbb{P}(E) = \mathbb{P}(1) + \mathbb{P}(3) + \mathbb{P}(5) = 1/6 + 1/6 + 1/6 = 1/2.$$

For the unbalanced case the probability is

$$\mathbb{P}(E) = \mathbb{P}(1) + \mathbb{P}(3) + \mathbb{P}(5) = 1/6 + 1/12 + 1/6 = 5/12.$$

Observe that the fundamental assumption is equivalent ot $\mathbb{P}(\Omega) = 1$, that is, that your experiment will have an outcome for sure. Of course, the probability of the impossible event $\emptyset$ is zero:$\mathbb{P}(\emptyset) = 0$.

3.2   Uniform probablity

When you have no reason to think that one of the outcomes, or elements of the sample space is more likely than the other, then to each ω_i you should assign the same probability $\mathbb{P}(\omega_i) = p$, where $p \in [0,1]$ and for each $i = 1, ..., n$. In this case the fundamental assumption tells us that

$$\mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_n) = p + ... + p = np = 1,$$

so that $p = 1/n$. Therefore each elementary event has the same probability and this probability is $1/n$, where we recall that n is the number of elements of Ω.

Let's consider the event $E = \{\omega_1, ..., \omega_k\}$ (or more generally $E = \{\omega_{i_1}, ..., \omega_{i_k}\}$). The definition (2) tells us that $\mathbb{P}(E) = 1/n + ... + 1/n = k/n$. This can be written in the suggestive form

$$\mathbb{P}(E) = \frac{\text{number of elements of E}}{\text{number of elements of } \Omega}.$$

Usually there is the tendency to think that the probability measure to assign on Ω by default is the uniform one. This common error explains the following common abuse of lenguage:

Example 3.5 (Random extractions). *When you read expressions of the kind "a random person is chosen (or a person is chosen randomly) between a group of 20 people", usually what is really meant is " A random person is chosen between a group of 20 people, in such a way that no outcome is favourable". You are giving the Sample Space $\Omega = \{1, ..., 20\}$, and at the same time you are assigning the uniform probability*

$$\mathbb{P}(1) = ... = \mathbb{P}(20) = \frac{1}{20}. \qquad \boxed{\texttt{e:ru}}$$

A jury trying to choose the new pianist for the orchestra between 20 partici-pants also chooses a random person betweena group of 20, but hopefully each individual has not the same probability.

The next examples show some natural cases where the probability is not the uniform one.

Example 3.6 (A Random Extraction). *In a bowl there are 3 blue balls and 1 red ball. We extract radomly (recall the previous example!) one ball. What*

is the probability that the extracted ball is blue? This is intuivley 3/4. In fact, if we label each one of the blue ball by b_1, b_2 and by b_3, the sample space is $\Omega = \{b_1, b_2, b_3, r\}$. By extracting a ball randomly we mean that we assign the uniform probability measure on Ω. The event b="A blue ball is extracted"= $\{b_1, b_2, b_3\}$ has probability $\mathbb{P}(b) = 3/4$. If our particles are not labelled and we can distinguish them only through their colour, the sample space, space of outcome of the experiment is $\Omega = \{b, r\}$ and the probability is $\mathbb{P}(b) = 3/4$ and $\mathbb{P}(r) = 1/4$.

Example 3.7 (The sum of two dice). *The sample space is $\Omega = \{(1,1),(1,2)....(6,6)\} = \{(i,j) : i,j = 1,...6\}$ and, endow it with the uniform probability (For one dice, the uniform probability is justified by fairness. For two fair dice, the uniform probability is justified by the fairness of each single dice and the independent way we throw the dice, but this will be explained in ??):*

$$\mathbb{P}(i,j) = \frac{1}{|\Omega|} = \frac{1}{36} \qquad \boxed{\text{dice}}$$

The probability of the event $A_3 =$ "The sum of the two dices is 3" $= \{(i,j) \in \Omega : i+j = 3\}$, is

$$\mathbb{P}(A_3) = \frac{|A|}{|\Omega|} = \frac{2}{36} = \frac{1}{18}. \qquad \boxed{\text{a3}}$$

You can check that the probabilities of the events $A_n=$ "The sum is n" $= \{(i,j), i+j = n\}$, for $n = 2,...12$ are

$$\mathbb{P}(A_2) = \mathbb{P}(A_{12}) = \frac{1}{36}$$
$$\mathbb{P}(A_3) = \mathbb{P}(A_{11}) = \frac{1}{18}$$
$$\mathbb{P}(A_4) = \mathbb{P}(A_{10}) = \frac{1}{12}$$
$$\mathbb{P}(A_5) = \mathbb{P}(A_9) = \frac{1}{9}$$
$$\mathbb{P}(A_6) = \mathbb{P}(A_8) = \frac{5}{36}$$
$$\mathbb{P}(A_7) = \frac{1}{6}$$

$\boxed{\text{sum}}$

. If you are only looking at the sum, the sample space is $\tilde{\Omega} = \{A_2, ..., A_12\}$, and the uniform probability is not natural here.

Exercise 1: A coin is tossed 4 times. Recall that the state space is $\Omega = \{0000, 0001, ..., 1111\}$, where 1 means head. Suppose that the probability on Ω is uniform (we will see in section 5 that this is due to independence of each coin toss). Calculate the probability of the events

1. The first toss gave 1

2. There are two heads

3. There has been a head after a tail (that is, we observed a 01)

Exercise 2: We throw two dice, and we suppose that the each output has the same probability (We will see in Section 6 why this is the case)

1. Calculate the probability of obtaining a 2 and a

2. Calculate the probability of obtain 2 with the first die and 4 with the second

3.3 Basic properties

Let Ω be a sample space and $\mathbb{P}$ a probability defined on it.

p1 **Proposition 3.8.**

$$\mathbb{P}(\Omega) = 1$$ e:normalization

Proof. It is a direct consequence of (1):

$$\mathbb{P}(\Omega) = \mathbb{P}(\{\omega_1, \omega_2, ..., \omega_n\}) = \mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_n) = 1.$$

□

The above property tells us that one outcome in the sample space will happen for sure.

Proposition 3.9. *If A and B are two* mutually exclusive, *that is, $A \cap B = \emptyset$. Then*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \tag{3}$$ e:add

More generally, the following relation holds

18

Proposition 3.10. *Let A, $B \subset \Omega$ be two events.*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \qquad (4) \quad \boxed{\text{e:union}}$$

Before giving the prove, let's see graphically why the above formulas hold. The probability of A is the number of ω_i which are in A, each weighted with weight $\mathbb{P}(\omega_i)$, and the same holds for B. When computing $\mathbb{P}(A) + \mathbb{P}(B)$ we are counting the elements of $A \setminus B$, of $B \setminus A$, and we are counting twice the elements in $A \cap B$. if we subtract $\mathbb{P}(A \cap B)$, the elements of $A \cap B$ are counted only once.

Proof. Let $\Omega = \{\omega_1, ..., \omega_n\}$ be the sample space. Without loss of generality, we assume that $A = \{\omega_1, ..., \omega_s, \omega_{s+1}..., \omega_l\}$, and that $B = \{\omega_1, ..., \omega_s, \omega_{l+1}, ..., \omega_{l+m}\}$. In this way, $A \cap B = \{\omega_1, ..., \omega_s\}$, and $A \cup B = \{\omega_1, ...\omega_{l+m}\}$. Using the definition of probability

$$\mathbb{P}(A) = \mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_s) + \mathbb{P}(\omega_{s+1}) + ... + \mathbb{P}(\omega_l),$$

$$\mathbb{P}(B) = \mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_s) + \mathbb{P}(\omega_{l+1}) + ... + \mathbb{P}(\omega_{l+m})$$

and

$$\mathbb{P}(A \cap B) = \mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_s).$$

In this way,

$$\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_s) + \mathbb{P}(\omega_{s+1}) + ... + \mathbb{P}(\omega_{l+m}) = \mathbb{P}(A \cup B).$$

$\square$

Corollary 3.11 (Monotonicity)**.** *Let A and B be two events $A, B \subset \Omega$. If $A \subset B$, then*

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

The monotonicity is a very intuitive property. The probability that today's weather is bad is greater than the probability that today there is a storm.

Corollary 3.12. *Let $A \subset \Omega$ be an event. The*

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \qquad\qquad \boxed{\text{e:compl}}$$

Proof. Apply the previous proposition with $B = A^c$. □

Also this property is very natural. The probability that I will pass the test is 0,6. Thus, the probability that I will not pass the test is 0.4,

Proof. For the first corollary, just use (**??**) with A and $B \setminus A$. For the second one, use (**??**) with A and A^c.

□

Ex. 11 — You roll three dice and give to the sample space the uniform probability (This choice for the probability will be justified in section **??** by the fact that the dice are rolled independently and each one of them is balanced) A="The first two dice show the same result", B=" The last two dice show the same result". Verify relation (4)

Answer (Ex. 11) — The sample space is $\Omega = \{(1,1,1),(1,1,2)...,(6,6,6)\}$, which contains $6 \times 6 \times 6$ elements. The event $A = \{(i,j,k) : i = j\} = \{(i,i,k) : i,k \in \{1,2,3,4,5,6\}\}$, composed by the 3-uples that whose first 2 coordinates are identical, has 6×6 elements, and the same remains true for the event $B = \{(i,j,j) : i,j \in \{1,2,3,4,5,6\}\}$. You can also check that $A \cup B$" has 66 elements, the ones coming from A plus the ones coming from $B \setminus A$. Since $A \cap B = \{(1,1,1),(2,2,2),...,(6,6,6)\}$ has 6 elements, we can use formula (**??**) to obtain

$$\mathbb{P}(A) = \mathbb{P}(B) = 1/6,$$
$$\mathbb{P}(A \cap B) = 1/36,$$
$$\mathbb{P}(A \cup B) = 7/36.$$

Ex. 12 — (Symmetric Bernoulli Trials) We toss a coin 9 times. The sample space is follows $\Omega = \{x_1...x_{10} : x_i \in \{0,1\} for \ i = 1, ..., 10\}$, the set of possible 9 digit strings with digits in the set $\{0,1\}$ We assign the uniform probability (once again, reason why we do that will be found in the subsection **??**).

1. What is the probability all the coin tosses gave head?
2. What is the probability that only 1 coin toss gave heads?
3. What is the probability that at most 1 coin toss gave heads?
4. What is the probability that at least 8 coin tosses gave heads?

5. What is the probability that less or equal than 4 coin tosses gave heads? (Hint: Pass to the complementary event. How are the probability of the event and of its complementary related?)

Answer (Ex. 12) — The number of elements in Ω is 2^9

1. The event corresponds to the outcome 11...1. Thus the probability is $1/|\Omega| = 1/2^9$

2. The event corresponds to the strings whose digits are all 1 except one, which is 0. We have to calculate the probability of $\{011111111, 101111111, ...\}$. This set has 9 elements, so that applying (**??**), we obtain that the probability is $9/2^9$.

3. Changing the w

4. We observe that this event is the same as F=”There has been at most 1 tails”. This event has the same number of elementary events as the event E, 10. Thus the probability is again $10/2^9$.

5. This computation can be done directly, but we use the following shortcut. Let's denote by H the event in the above question and consider $A^c=$” There are 5 or more heads”=” There are 4 or less tails”. From this last reformulation, we see that A and A^c have the same number of elements (they are obtained one from the other by changing the word heads to the word tails), and thus $\mathbb{P}(A) = \mathbb{P}(A^c)$. Using (**??**), we obtain that $\mathbb{P}(A) = 1 - \mathbb{P}(A)$, so that $\mathbb{P}(A) = 1/2$.

   For this last point we did not use that the probability measure is the uniform, but only the folllowing property of the probability measure: Let A be an event and define the event $\tilde{A}$ the event obtained by A changing the words heads and tails. We used only this property of $\mathbb{P}$: $\mathbb{P}(A) = \mathbb{P}(\tilde{A})$, for every set A and $\tilde{A}$ obtained from it in the way just described. This property is much less restrictive than asking that for the uniform probability measure, and is the right way of asking that heads and tails are symmetric. Check that the probability measure that assigns $\mathbb{P}(111111111) = 0.5 = \mathbb{P}(000000000)$ has this property. I conclude by saying that mathematically $\tilde{A}$ obtained by changing 1 to 0 and 0 to 1 to all its elements.

Ex. 13 — ([**Ross**] Chapter 2 Exercise 10) Sixty percent of students at a

certain school wear neither a ring nor a necklace. Twenty percent wear a ring and 30 percent wear a necklace. If one of the students is chosen randomly (recall what chosen randomly actually means), what is the probability that one of the students is wearing

1. A ring or a necklace
2. A ring and a necklace

Answer (Ex. 13) — The hypotheses mean the following: Denote by n the total number of students, by l the number of those who wear a ring, by k the number of those who wear a necklace and by r the number of those who don't wear neither a ring nor a necklace. Then $l/n = 20/100$, $k/n = 30/100$, $r/n = 60/100$. Chosing a person randomly between a group of n people means that the sample space is $\Omega = \{1, ..., n\}$, where each number denotes a different person, and each person has the same probability of being chosen. Gluing the above information we obtain that, denoting by R="The extracted student wears a ring", N= "The extracted student wears a necklace" and using that "The extracted student wears no ring nor necklace"= $N^c \cap R^c$, that

$$\mathbb{P}(R) = l/n = 20/100,$$
$$\mathbb{P}(N) = k/n = 30/100,$$
$$\mathbb{P}(N^c \cap R^c) = r/n = 60/100.$$

1. Observe that $(R \cup N)^c = R^c \cap N^c$, so that

$$\mathbb{P}(R \cup N) = 1 - \mathbb{P}(R^c \cap N^c) = 1 - 60/100 = 40/100 = 0.4$$

2. By (**??**)

$$mathbbP(R \cap N) = \mathbb{P}(R) + \mathbb{P}(N) - \mathbb{P}(R \cup N) = 20/100 + 30/100 - 40/100 = 0.1.$$

Ex. 14 — ([**Ross**] Chapter 2 Exercise 12) An elementary school is offering 3 lenguage classes, 1 in Spanish, one in French, and one in German. The classes are open to any of the 100 students in the school. There are 28 students in the Spanish class, 26 in the French class, 16 in the German class. There are 12 students who are in both Spanish and French, 4 who are in both French and German, and 6 that are in both Spanish and German. In addition there are 2 students taking all 3 classes.

1. If a student is chosen randomly. What is the probability that he or she is not in any lenguage classes?

2. If a student is chosen randomly, what is the probability that he or she is taking exactly one lenguage class?

3. If two students are chosen randomly, what is the probability that at least 1 is taking a lenguage class?

   To choose a random student means that each student id equally probable to be chosen. Two random choices of students means that we choose one student and then another one from the 99 remaining students independently and with uniform probability. We will see in Section **??** that does this mean, but for the sake of this exercise we need only to use that each one of the $100 \times 99/2$ ways of choosing 2 students out of 100 are all equally likely.

Answer (Ex. 14) — s

1. The event of the first point is $(S \cup F \cup G)^c =$. We use formula (**??**) with $A = S \cup F$ and $B = G$, so that

$$\mathbb{P}(S \cup F \cup G) = \mathbb{P}(S \cup F) + \mathbb{P}(G) - \mathbb{P}((S \cup F) \cap G)$$
$$= \mathbb{P}(S \cup F) + \mathbb{P}(G) - \mathbb{P}((S \cap G) \cup (F \cap G)).$$

   Using twice more formula (**??**), first with $A = S$ and $B = G$, then with $A = F \cap G$ and $B = S \cap G$, we obtain that

$$\mathbb{P}(S \cup F \cup G) = \mathbb{P}(S) + \mathbb{P}(F) - \mathbb{P}(S \cap F) + \mathbb{P}(G) - \mathbb{P}((S \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(F \cap S \cap G))$$
$$= 28/100 + 26/100 - 12/100 + 16/100 - 6/100 - 4/100 + 2/100 = 50/100.$$

2. We know from the previous point the probability that the student doesn't follow any course. If we calculate the probability that the student follows at least two courses $(S \cap F) \cup (S \cap G) \cup (F \cap G)$. Once again we iteratively use formula (**??**) first with $A = (S \cap F)$ and $B = (S \cap G) \cup (F \cap G)$ to obtain

$$\mathbb{P}((S \cap F) \cup (S \cap G) \cup (F \cap G)) = \mathbb{P}(S \cap F)$$
$$+ \mathbb{P}((S \cap G) \cup (F \cap G)) - \mathbb{P}(S \cap F \cap G).$$

Once again we use (**??**) with $A = S \cap G$ and $B = S \cap G$ to obtain

$$\mathbb{P}((S \cap F) \cup (S \cap G) \cup (F \cap G)) = \mathbb{P}(S \cap F)$$
$$+ \mathbb{P}(S \cap G + \mathbb{P}(F \cap G)) - 2\mathbb{P}(S \cap F \cap G) = 12/100 + 4/100 + 6/100 - 2 \times 2/100 = 18/1$$

Since the event that the student takes exactly one course is the complementary of the event that the student takes no course or more (or equal) than two courses, we obtain that

$$\mathbb{P}(\text{" The student takes exactly 1 course "}) = 1 - (50/100 + 18/100) = 32/100.$$

3. Denote by E_1 the event that the first student chosen will follow a lenguage class and by E_2 the same event for the second student. The event of which we want to compute the probability is $E_1 \cup E_2$. Noting that we already know the probability of the events E_1 and E_2, which is the same as the one of the event $S \cup G \cup F$, we need to compute the probability of $E_1 \cap E_2$. Since we are extracting two students, the sample space is now $\Omega = \{(1,2), (1,3), ...(99, 100)\} = \{(i,j), i \neq j, i, j = 1, 2..., 100\}$, and the probability is the uniform one. The number of elements of Ω is $100^2 - 100$ (I have to subtract the pairs where I choose twice the same student, inammissible choice) How many pairs of students that are both in a lenguage class? We now that there are 50 students that are at least in a lenguage class. The number of pairs that I can form with students is therefore $50^2 - 50$ (I have to subtract the pairs in which the same student is chosen). Therefore $\mathbb{P}(E_1 \cap E_2) = 2450/9900$ and $\mathbb{P}(E_1 \cap E_2) = 0.5 + 0.5 - 2450/9900$.

Ex. 15 — ([**Ross**] Chapter 2 Exercise 17)
If 8 rooks (castles) are randomly (We assume that we are using the uniform probability on the sample space. Why this is done will be clear in Section 7) placed in a chessboard, compute the probability that none of the rooks can capture any of the others. That is, compute the probability that no row or file contains more than one rook. (Hint: to count the number of ways you can place the rooks, start by placing the first and count how many places are left). Hint: start with a 2x2 or 3x3 chessboard.

Answer (Ex. 15) — We will count all the possible ways to place 8 (distinguishable) rooks in a chessboard. First place the first rook: you 64 possibil-

ities. Then place the other one: you have 64-1=63 possibilities (64 places minus the one which is already occupied). For the third you have 64-2=62 possibilities.... For the 8-th one you have 64-(8-1)=57 possibilities. In total, the number of ways you can place the rooks is $64 \times 63 \times ... \times 57$. The number of configurations in which the rooks can be placed in such a way that any of them can capture any of the other can be counted in a similar way. First place the first tower: you have 64 possibilities. Then place the second: you have 7×7 possibilities (the whole chessboard minus the row and the column where the first rook is). For the third you have 6×6 possibilities... Using formula (**??**), the solution is

$$\frac{\text{number of ways of placing 8 rooks as required}}{\text{number of possible ways of placing 8 rooks}} \frac{8 \times 8 \times 7 \times 7 \times 6 \times 6... \times 1 \times 1}{64 \times 63 \times ... \times 57} = \frac{(8!)^2}{64 \times 63 \times ... \times 5}$$

Note that are implicitly using the following sample space : number the 8 rooks from 1 to 8 and the 64 places in the chessboard from 1 to 64. Denote x_1 the place of the first rook, by x_2 the place of the second, and so on. Since two different rooks are in different places $x_i \neq x_j$ if $i \neq j$. A configuration of rooks is $(x_1, ..., x_8)$, where $x_i \in \{1, ..., 64\}$ and $x_i \neq x_j$ if $i \neq j$. Thus, the state space of placing 8 rooks in a chessboard is

$$\Omega = \{(x_1, ..., x_8) : x_i \neq x_j \text{ if } i \neq j\},$$

and we have proved is that the number of elements in Ω is $|\Omega| = 64 \times 63 \times, ... \times 57$

4   Conditional probability

The probability of an event depends on the state of information that an individual has on a system. We can ask ourselves how does the probability change when we know that an event A has happened. For example, the probability of the event "A.C. Milan wins the Champions League", decreases if the event "Ibrahimovich gets injured" happens. Let $\Omega = \{\omega_1, ..., \omega_n\}$ be the sample space and assume that the event $B = \{\omega_1, ..., \omega_k\}$ has been observed. We denote the new probability *conditioned to B* by the symbol $\mathbb{P}(\cdot|A)$, that is, we denote the probability of A conditioned to B by $\mathbb{P}(A|B)$. Of course, the new probability has to satisfy $\mathbb{P}(\omega|B) = 0$ if $\omega \notin B$. If $\omega \in B$, what should

25

the new value $\mathbb{P}(\omega|B)$ be? Since the new informtion tells us only that B has happened, but it does not give us information about the single ωs in B, the proportions given by $\mathbb{P}$ in B should be manteined by $\mathbb{P}(\cdot|B)$. We therefore set $\mathbb{P}(\omega|B) = c\mathbb{P}(\omega)$ if $\omega \in B$, for some constant c. c can be determined by requiring $P(\cdot|B)$ to datisfy (1). If $B = \{\omega_1, ..., \omega_k\}$, we obtain that

$$1 = \mathbb{P}(\omega|B) + ... + \mathbb{P}(\omega_k|B) + \mathbb{P}(\omega_{k+1}|B) + ... + \mathbb{P}(\omega_n|B)$$
$$= c\mathbb{P}(\omega_1) + ... + c\mathbb{P}(\omega_k) = c(\mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_k)) = c\mathbb{P}(B).$$

Therefore $c = 1/\mathbb{P}(B)$ and

$$\mathbb{P}(\omega|B) = \begin{cases} \frac{\mathbb{P}(\omega)}{\mathbb{P}(B)} & \text{if } \omega \in B \\ 0 & \text{if } \omega \notin B, \end{cases} \qquad \boxed{\texttt{e:condi}}$$

Now consider an event $A = \{\omega_1, ..., \omega_s, \omega_{k+1}, ..., \omega_{k+m}\}$, where $s \leq k$. (2) gives

$$\mathbb{P}(A|B) = \frac{1}{\mathbb{P}(B)}(\mathbb{P}(\omega_1) + ... + \mathbb{P}(\omega_s) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

$\boxed{\texttt{conditional}}$ **Definition 4.1.** *Let Ω be the sample space and let $\mathbb{P}$ be a probability. Consider an event $B \subset \Omega$ such that $\mathbb{P}(B) > 0$. The conditional probability $\mathbb{P}(\cdot|B)$ given B is the probability measure defined by*

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

for each $A \subset \Omega$.

Ex. 16 — If you throw two dice (we regard each outcome as equally probable and the reason will be investigated in **??**), what is the probability conditioned on "The sum of the two dice is 7"?.

Answer (Ex. 16) — The sample space is $\Omega = \{(i,j),\ i,j \in \{1,2,3,4,5,6\}\}$, and the event corresponds to $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$. We see that $\mathbb{P}(\omega|A) = \frac{1/36}{6/36} = 1/6$ if $i + j = 7$, 0 otherwise. That is,all the outcomes in A are equally probable.

Ex. 17 — You toss a fair coin 4 times. Show that, knowing that only one head has been observed, the head is equally probable to have appeared in the 1st,2nd,3rd and 4th position.

Ex. 18 — Show that $\mathbb{P}(\cdot|A)$ is a probability on the new sample space A. Show that if $\mathbb{P}$ is the uniform probability in Ω, then $\mathbb{P}(\cdot|A)$ is the uniform probability in A. That is, show that $\mathbb{P}(\omega|A) = 1/\#$elements of A, for every $\omega \in A$.

Ex. 19 — ([**Feller**], Chapter V.8) Three dice are rolled. If no two show the same face, what is the probability that one is an ace?

Answer (Ex. 19) — The sample space is $\Omega = \{(i,j,k), i,j,k \in \{1,....6\}\}$, and the event "No two dice show the same face is $A = \{i,j,k) \in \Omega, i \neq j, j \neq k, i \neq k\}$. The event B="The sum of the two dice is 7" is such that $A \cap B$ is the disjoint union of the events"A and the 1 is in 1st position", "A and the 1 is in 2nd", "A and the 1 is in 3rd poosition". This union is disjoint since if A happens, a 1 in the 1st position implies that there is no 1 neither in the 2nd nor in the 3rd. In formulas $A \cap B = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$, where $B_1 = \{(1,j,k) \in \Omega\}$, $B_2 = \{(i,1,k) \in \Omega\}$ and $B_3 = \{i,j,1) \in \Omega\}$. Thus

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B_1)}{\mathbb{P}(A)}$$
$$+ \frac{\mathbb{P}(A \cap B_2)}{\mathbb{P}(A)} + \frac{\mathbb{P}(A \cap B_3)}{\mathbb{P}(A)} = 3\frac{\mathbb{P}(A \cap B_1)}{\mathbb{P}(A)}.$$

The event $A \cap B_1 = \{(1,j,k), j \neq k\}$ is composed by $6 \times 5 = 30$ elements. The event A is composed by $6 \times 5 \times 4$ elements. Thus

$$\mathbb{P}(B|A) = \frac{6 \times 5/216}{6 \times 5 \times 4/216} = \frac{6 \times 5}{6 \times 5 \times 4} = 1/4$$

Example 4.2 (Sampling interpretation of the Conditional Probability). *Suppose that the Italian population is composed by N individuals, that N_B of them are taller than 175 cm and that N_A of them come from Sardinia. Let the events that a person chosen at random (i.e. with uniform probability) is taller than 175 cm and a comes from Sardinia be A and B, respectively. Then $\mathbb{P}(A) = N_A/N$ and $\mathbb{P}(B|A) = N_{AB}/N_A$, where N_{AB} is the number of people coming from Sardinia being taller than 175 cm. The interpretation of the probability measure $\mathbb{P}(\cdot|A)$ is clear: we are extracting a random person from the subpopulation N_A (which is a population on its own right), the one composed by people from Sardinia .*

4.1 Independence of events

Intuitively, two events $A, B \subset \Omega$ are independent if the realization of one of them does not influence the probability of the other: $\mathbb{P}(B|A) = \mathbb{P}(B)$. This can be rewritten in a more symmetric form as

Definition 4.3. *Let $A, B \subset \Omega$ be two events. A and B are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

For 3 events A, B and C to be independent mean that they are pairwise independent $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, and that they all satisfy $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. An example three events pairwise independent but not independent is given in one of the following exercises.

Ex. 20 — You toss a fair coin 4 time. Assume that every outcome is equally probable (we will rigorously see why). Show that the events "The fourth toss gave heads" and the event "The first three tosses gave head" are independent.

Ex. 21 — You draw from a bowl that contains three marbles, one red, one blu, and one green. You replace the marble and you draw another. $\Omega = \{(r, r), ..(g, g)\}$, and we use the uniform probability on Ω (we will see rigourously why in **??**). Are the events the firs marble is red and the second is green independent?

Ex. 22 — We draw two marbles from a bowl, without replacing the ones that we have drawn. We can keep the same sample space Ω as before and consider $\mathbb{P}(r, r) = \mathbb{P}(b, b) = \mathbb{P}(g, g) = 0$. In the other outcomes we place the uniform probability: $\mathbb{P}(r, g) = - - - = \mathbb{P}(b, g) = 1/6$ (this is due to independence and will be explained in **??**). Are the events "The first marble is r" and "the second marble is g" independent?

Ex. 23 — $\Omega = \{1, 2, 3, 4\}$ and $\mathbb{P}(i) = 1/4$ for every $i \in \Omega$. Show that $A = \{1, 2\}$ $B = \{1, 3\}$ and $C = \{2, 3\}$ are pairwise independent but not independent.

4.2 Total probability Formula

Usually when trying to estimate the probability of an event, such us, "Tomorrow I will go to the cinema", we do reasonings of the kind i" if it rains I will go with probability q, if it is good weather I will with probability p".

The rigourous mathematical analog is given by the total probability formula. We start with a particular case

Proposition 4.4. *Let $B \subset \Omega$ be an event. Then*

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \qquad (5) \quad \boxed{\texttt{e:totaleasy}}$$

Proof.

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap (B \cup B^c)) = \mathbb{P}((A \cap B) \cup (A \cap B^c)).$$

Since $A \cap B$ and $A \cap B^c$ are disjoint, we can use the (3) and the above expression becomes

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

which is what we wanted to prove. $\qquad\qquad\qquad\qquad\qquad\square$

The above formula has an easy interpretation. We are summing the proportion of A in B, but we have tonormilize by the total mass of B, which is $\mathbb{P}(B)$, and the proportion of A in B^c, weighted $\mathbb{P}(B^c)$. Let's see an example

Example 4.5. *We toss a coin. If the coin gives heads we draw a ball from a bowl that contains a blue ball, a red ball and a white ball. If the con gives tails, we draw a ball from a bowl that contains a blue ball or a red ball. What is the probability of drawing a red ball?*
Let's call B="The coin gave heads". We know that $\mathbb{P}(B) = \mathbb{P}(B^c) = 1/2$. We also know that the event r=" We draw a red ball" has the following probabilities conditioned on B:

$$\mathbb{P}(r|B) = 1/3, \ \ \mathbb{P}(r|B^c) = 1/2.$$

We can apply formula (5) to obtain

$$\mathbb{P}(r) = \mathbb{P}(r|B)\mathbb{P}(B) + \mathbb{P}(r|B^c)\mathbb{P}(B^c) = 1/2 \cdot 1/3 + 1/2 \cdot 1/2 = 5/12.$$

We now give the total probability formula in full generality. The proof is almost the same, but we need to introduce the concept of partition of the sample space Ω:

Definition 4.6. *The events $B_1,...,B_n$ are said to be mutually disjoint if $B_i \cap B_j = \emptyset$ whenever $i \neq j$. A partition is a set $\{B_1, ..., B_n\}$ of mutually disjoint events B_i, such that $\Omega = \bigcup_{i=1}^{n} B_i$.*

Examples of partitions are

- In 4 coin tosses, a partition of $\Omega = \{0000, 0001, ..., 1111\}$ in two events is, for example, $\{B_1, B_2\}$, where $B_1=$"The first toss gave tail" and $B_2 = (B_1)^c=$"The first toss was head".

- In the roll of a die, an example of partition of $\Omega = \{1, 2, 3, 4, 5, 6\}$ is $\{B_1, B_2\}$, where $B_1 = \{2, 4, 6\}$ ="The result is an even number", and $B_1 = \{1, 3, 5\}$ ="The result is an odd number".

- If we roll two dice, a partition is $\{B_2, ..., B_{12}\}$, where B_i is the event "The sum of the dice is i".

- We observe some coin tosses but we don't know how many of them. $\Omega = \cup_n \Omega_n$, where $\Omega_n=$" There have been n tosses "= $\{x_1...x_n, x_i \in \{0, 1\}\}$. The Ω_n form a partition.

Note that if a partition consists of two events B_1 and B_2, then $B_2 = (B_1)^c$.

Proposition 4.7 (Total probability formula). *Let $\{B_1, ..., B_n\}$ be a partition of Ω. Then*

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + ... + \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

We omit the prove, which is similat to the previous.
 At the moment 1/1000 has covid. The test has a 5% of possibility of giving a false positive: You don't have covid but the test is positive (and if you don't have covid, it always testes negative).

Ex. 24 — Ex. 25 — 1. You make the test and you are tested positive. What is the probability that you have covid

2. You have some sympthoms and you think that the probability that you have covid is 50%. You get tested and

Answer (Ex. 25) — Let the sample space be $\{(C, P), (C, N), (NC, P), (NC, N)\}$, where C tells you that you have covid and NC that you dont. The second variable is the result of the test.

1. For the first question, the data that is given to you is, denoting A="You tested positive", B="you have covid", is that $mathbbP(B) = 1/1000$, that $\mathbb{P}(A|B) = 1$ and that $\mathbb{P}(A|B^c) = 5/100$. You are interested in $\mathbb{P}(B|A)$

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

   The only information that is left is $\mathbb{P}(A)$. We use the total probability formula

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

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

2. Doing the same computations, now using $\mathbb{P}(B) = 50/100$, shows that the probability now is much higher.

   The reason is that even if the test has a good presition, on 1000 people the positive tests come 1 from a person that effectively has covid, while the other 999 x 5% $\sim$ 50 come from of a false positive. Around 1/51 is the proportion of the people with covid in the population of

   Put an image here!!

4.3   Some paradoxes.

With paradoxes we actually mean facts that are counterintuitive at a first glance.

The Mounty Hall problem Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

One is tempted to say that it makes no difference, but actually this is incorrect, and it is convenient to change the door. This fact is explained once

we note that the without changing the door, we win with probability $1/3$. Changing the door, we win if and only if we have chosen at the beginning the wrong door, and this happens with probability $2/3$. We are going to see this in a more mathematical way.

We will call Strategy 1 the strategy of playing the game without changing your door. We will call Strategy 2 the game played choosing to change your door.

Suppose that we choose the door 1. The sample space is

$$\Omega = \{(G_1, G_2, C), (G_1, C, G_2), (G_2, G_1, C), (G_2, C, G_1), (C, G_1, G_2), (C, G_2, G_1)\}$$

but we actually can solve the problem without specifically refer to Ω. If we play by Strategy 1, then the event "We win"="We choose the car at the beginning", which has probability $1/3$

If we play by Strategy 2, then the event A=" We win"=" The second choice is a car". The second choice is a car is contained in the event B="The first choice is a goat". Here the important fact is that, conditioned on B, the probability of finding a car is 1, since the host has opened the other door. We write

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

Exercise: Repeat the computations when the number of doors is n (you can assume that you chose the first

1. The host opens all the doors that are .

2. The host opens n-2 doors which have a goat behind.

Solution Let A,B be the events defined above. With strategy 1, $\mathbb{P}(A) = 1/n$. With Strategy 2, we can say that, since $A \subset B$,

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

Clearly $\mathbb{P}(B) = (n-1)/n$. Now, knowing that B has happened, we need to find the car between k doors, where k is equal to $k = n - 2$ if the host opens only 1 door or $k = 1$, if the host opens $n - 2$ doors. When we do this choice, independently from B, we have the probability of $1/k$ to find a car. Thus

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

1. The first case corresponds to $k = 1$. Thus $\mathbb{P}(A) = (n-1)/n$

2. The second case corresponds to $k = n - 2$. Thus $\mathbb{P}(B) = \frac{1}{n-2}\frac{(n-1)}{n}$.

You can verify that in both cases Strategy 2 is the best strategy.

Boy or Girl Paradox This example shows that is important to clearly state the conditioning event.

1. What is the probability that the second child of a couple is a boy, knowing that the firs one is a boy?

2. We know that one of the child is a boy. What is the probability that both children are males?

3. We have chosen a child randomly (and independently on the gender, of course) and have seen that he is male. What is the probability that both children are boys? (Note that the first poin is a particular case of this one, when the chosen child is surely the first one).

For the first question, look at the sample space $\Omega = \{(M, M), (M, F), (F, M), (F, F)\}$ and endow it with the uniform probability (in section **??** we will see that this is due to the fact that we regard rhe gender of the two child as independent, and the probability of each single child to be either a boy or a girl $1/2$). The event A="The first child is male" is given by $\{(M, M), (M, F)\}$. We are interested in

$$\mathbb{P}((M, M)|A) = \mathbb{P}(M, M)/\mathbb{P}(A) = \frac{1/4}{2/4} = 1/2.$$

Maybe counterintuitively, the answer is $1/3$. But this is clear since the conditioning event A is $\{(M, M), (M, F), (F, M)\}$, and

$$\mathbb{P}((M, M)|A) = \mathbb{P}(M, M)/\mathbb{P}(A) = \frac{1/4}{3/4} = 1/3.$$

Using a frequentist approach, if one had the list of al the couples with two children, one of which male, approximately $1/3$ would have two male childeren. In the last case, the probability is $1/2$! Let's take as state space $\Omega = \{MM1, MM2, MF1, MF2, FM1, FM2, FF1, FF2\}$, where the last number

means that you have observed the gender of the child 1 or child 2. The event A="you observed a male child"=$\{MM1, MM2, MF1, FM2\}$, while the event "The couple has two male children" is $B = \{MM1, MM2\}$. Since we assume that the choice is done independently, we have that if we denote 1=" I observed the first child" $\mathbb{P}(MM1|1) = \mathbb{P}(MF1|1) = \mathbb{P}(FF1|1) = 1/4$. Similarly for the event $2 =$ "I observed the second child ". We assuem that $\mathbb{P}(1) = p$ and we compute

$$\mathbb{P}(A) = \mathbb{P}(\{MM1, MF1\}|1)\mathbb{P}(1) + \mathbb{P}(\{MM2, FM2\}|2)\mathbb{P}(2) = p/2 + (1-p)/2 = 1/2,$$

and

$$\mathbb{P}(\{MM1, MM2\}) = \mathbb{P}(MM1|1)\mathbb{P}(1) + \mathbb{P}(MM2|2)\mathbb{P}(2) = 1/4.$$

$$\mathbb{P}(\{MM1, MM2\}|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A) = 1/2.$$

This apparent paradox is due to the fact that the event A is different from the event B="There is at least one male child". If you repeat the same calculations with B instead of A, you should obtain 1/3.

Ex. 26 — 3 prisoners, (A,B,C) are condemned to death. They decide to save one of them, but the name will be given only some minutes before the executions. A asks to a guard if he will be saved, but gets no answer. A asks then which one between B and C will be executed, and the guard says C. At this point A says."Now only one between me and B will be executed; I have 50% of chances to survive ". Is it correct his reasoning?

Answer (Ex. 26) — Is the same situation as the one in the above example: you are not asking if C will or not will be executed. The guard is telling you which one between B and C will be executed, and you don't know why he has chosen C. If you have no information on why the guard would tell you about the execution of C instead f the execution of B, wen both have to be executed, then the probability that A is executed is still 1/3. To see this consider the sample space$\{(AB, B), (AC, C), (BC, B), (BC, C)\}$, where the first component tells who will be executed and the second who the guard says it will be. Denote the events that the guard tells that B will be executed and the durd tells that C will be executed by GB and GC respectively. We assume that when $BC=$"both B and C will be executed" happens, the guard tells B or C according to some probability $\mathbb{P}(GC|AB)$. We are looking for

the probability of the event $A = \{(AC, C), (AB, B)\}$ that prisioners B and C will be excecuted, given GC. We use the total probability formula to get

$$\mathbb{P}(GC) = \mathbb{P}(GC|AC)\mathbb{P}(AC) + \mathbb{P}(GC|BC)\mathbb{P}(BC) = 1/3 + 1/3\mathbb{P}(GC|BC),$$

and that

$$\mathbb{P}((AC, C)) = \mathbb{P}(GC|AC)\mathbb{P}(AC) = 1/3,$$

so that, finally

$$\mathbb{P}(A|GC) = \mathbb{P}(AC|GC) = \frac{\mathbb{P}((AC, C))}{\mathbb{P}(GC)} =$$
$$= \frac{1/3}{1/3 + 1/3\mathbb{P}(GC|BC)}.$$

$\mathbb{P}(GC|BC) = 1/2$, the above value is 2/3 and the event A is indipendent from the event GC.

Ex. 27 — There are three cards one of them has both faces black, one with both red faces, and the last one with one black face and the other face red. You draw a card and observe a red face. What is the probability that the other face is red?

Answer (Ex. 27) — Denote by A the event "I observe a red face". Clearly $\mathbb{P}(A) = 1/2$. Let B="The other side is red" and observe that $A \cap B=$ "I draw the red red card". $\mathbb{P}(A \cap B) = 1/3$.

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

If you want you can introduce the sample space of $\Omega = \{(RR, R), (RB, R), (RB, B), (BB, B)\}$, where the first component denotes the extracted card and the second one the shown face, andyou can repeat the same computations.

   A judicial case A woman had been killed, and the principal suspect was his husband. During the investigations, the police discovered that the husband had beaten his wife more than once. His lawyer used the data that from the man that beat their wives, only 1/10000 end killing her. Fortunately, the prosecutors pointed a fallacy in the defense argument. 1/10000 gives an

estimate of the probability that a man kills his wife knowing that he beats her. Now we have to calculate the probability that a mans kills his wife knowing that he beat her and that she has been killed.
Denoting the events

$A = $ "The woman has been beaten by his husband "

$B = $"The woman has been assassinated by his husband "

$C = $"The woman has been assassinated by a person different from his husband "

Note that the event "The woman has been assassinated" corresponds to $B \cup C$. We wish to compute $p = \mathbb{P}(B|A \cap (B \cup C))$. Noting that $B \cap C = \emptyset$,

$$p = \mathbb{P}(B|A \cap (B \cup C)) = \frac{\mathbb{P}(B \cap A \cap (B \cup C))}{\mathbb{P}(A \cap (B \cup C))} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}((B \cup C) \cap A))}$$
$$= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B \cap A) + \mathbb{P}(C \cap A)} = \frac{\mathbb{P}(B|A)}{\mathbb{P}(B|A) + \mathbb{P}(C|A)} = \frac{1}{1 + \frac{\mathbb{P}(C|A)}{\mathbb{P}(B|A)}}.$$

Thus it is not $\mathbb{P}(B|A)$ that matters, but only its relative value $\mathbb{P}(B|A)/\mathbb{P}(C|A)$. Assuming that $\mathbb{P}(C|A) = \mathbb{P}(C) \leq \mathbb{P}(B \cup C)$ (in the first equality we are assuming that the probability of a woman being killed by another person is independent from the fact that his husband beats her. the third inequality), they found the estimate for $\mathbb{P}(B \cup C) = 1/100000$. Thus

$$\frac{\mathbb{P}(C|A)}{\mathbb{P}(B|A)} \leq 1/10.$$

Plugging in this value, one obtains that

$$p \geq 10/11$$

At the end, the man has been judged guilty.

5 Two experiments

Suppose that we have that we observe two experiments, one with sample space Ω_1 and the other one with Ω_2. We want to consider both experiments at the same time.

5.1   Example

The main example to have in mind for this section is the following. Imagine we want to toss two coins, each one of them in a fair way. This means that the probability of $A_1 =$" The first coin gives heads" and $B_1=$ "The second coin gives heads" has to verify $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$, that is, if we look at one of the two coins, it just makes a normal flip. The state space of the experiment is $\Omega = \{0,1\} \times \{0,1\} = \{00, 01, 10, 11\}$. As we just said, we know that "The first toss gave heads" $= \{10, 11\}$ verifies $\mathbb{P}(\{10, 11\}) = 1/2$, and that " The second toss gave heads" $= \{01, 11\}$ has probability $\mathbb{P}(\{01, 11\}) = 1/2$. However this is not sufficient to determine the probabilities of a joint realization such as $\mathbb{P}(01)$. To determine $\mathbb{P}$ one needs information on how the two experiment are performed jointly, and there is not a unique way to do this, as we are about to show.

Suppose we glue together the two coins in such a way that they both give heads or both give tails. In this case the probability on Ω is

$$\mathbb{P}(11) = \mathbb{P}(00) = 1/2, \quad \mathbb{P}(10) = \mathbb{P}(01) = 0.$$

In the same way, if we glue the two coins in such a way that, when one gives heads, the other one gives tails, we have that

$$\mathbb{P}(10) = \mathbb{P}(01) = 1/2, \quad \mathbb{P}(00) = \mathbb{P}(11) = 0.$$

Ex. 28 — Denote $A_1=$" The first result is 1", $A_0=$" The first result is 0", and $B_i=$" The first result is i", for i=0,1. Show that for the two probabilities introduced above the following holds

$$\mathbb{P}_1(A_i) = \mathbb{P}_2(A_i) = 1/2.$$

Show that the events A_i and B_j are not independent.

Answer (Ex. 28) — Since $A_1 = \{(1,1), (1,0)\}$,

$$\mathbb{P}_1(A_1) = \mathbb{P}_1(1,1) + \mathbb{P}_1(1,0) = 1/2 + 0 = 1/2,$$

and

$$\mathbb{P}_2(A_1) = \mathbb{P}_2(1,1) + \mathbb{P}_2(1,0) = 0 + 1/2 = 1/2.$$

The other cases are similar. In order to see that the events A_i and B_j are not independent, simply verify that

$$\mathbb{P}_1(A_i|B_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } \neq j, \end{cases}$$

and that

$$\mathbb{P}_2(A_i|B_j) = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } \neq j. \end{cases}$$

Another way to perform the two con tosses is the following: we throw a coin, observe the result and then, independently we throw the other one. The way we can see this is through the following formula, valid for $i, j \in \{0, 1\}$:

$$\mathbb{P}((i,j)) = \mathbb{P}(A_i \cap B_j) = \mathbb{P}(B_j|A_i)\mathbb{P}(A_i). \tag{6} \quad \boxed{\texttt{e:cond1}}$$

Now we use the hypothesis of independence: $\mathbb{P}(A_i|B_j) = \mathbb{P}(A_i) = 1/2$, so that above (6) becomes

$$\mathbb{P}(i,j) = \mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j) = 1/4.$$

Formula (6) can be interpreted in this way: we throw the first coin, and then the second, where the probability of the second toss depends on the first one, and holds for every probability. In the two cases where the two coins are bounded to have the same result, we have that

$$\mathbb{P}(A_1|B_1) = \mathbb{P}(A_2|B_2) = 1$$
$$\mathbb{P}(A_1|B_2) = \mathbb{P}(A_2|B_1) = 0.$$

5.2 Sample space of two experiments

We generalise what we have said in the previous example. There is a standard way of defining the sample space of an experiment which is the union of two experiments. Let $\Omega_1 = \{\omega_1, ..., \omega_n\}$ be the sample of the first experiment and $\Omega_2 = \{\eta_1, ..., \eta_m\}$ be the sample space of the second experiment. The event "ω_1 happened" is no longer elementary when we consider the two experiments together: for example, "ω_1 and η_3 happened " is contained in "ω_1 happened". The elementary events are those of the form "ω_i and η_j happened", and we will denote those events by (ω_i, η_j). Thus, the sample space of the two experiments is:

Definition 5.1. *The sample space of the experiment union of the two experiments $\Omega_1 = \{\omega_1, ..., \omega_n\}$ and $\Omega_2 = \{\eta_1, ..., \eta_m\}$ is $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_i, \eta_j)\,,\ i = 1, .., n, j = 1, .., m\}$*

Ex. 29 — How many elements does Ω have?

Answer (Ex. 29) — $n \times m$. To see this, note that for each $i = 1, .., n$ there are m different elementary events: $(i, 1), ...(i, m)$. Since i varies from 1 to n, we sum m for a total of n times.

Example 5.2 (Two dice). *The state space of the roll of two dice $\Omega = \{(1,1), (1,2), ..., (6,6)\}$ can be thought as the state space of the union of two experiments which consist in rolling a die. That is $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$*

5.3   Joint probability

The complete probabilistic description of the two experiments is given by a probability on Ω

Definition 5.3 (Joint probability). *The joint probability of the two experiments is a probability $\mathbb{P}$ on $\Omega = \Omega_1 \times \Omega_2$, the state space of the two experiments.*

The knowledge of the probability of the joint realization of the two experiment is given by the joint probability.

We now show that, knowing the joint probability, we know the probability $\mathbb{P}_1$ on Ω_1 and $\mathbb{P}_2$ on Ω_2, of each experiment separately. To see this, notice that the event ω_i="ω_i happened" is $\{\omega_i\} \times \Omega_2 = \{(\omega_i, \eta_1), ..., (\omega_i, \eta_m)\}$, and the event $\eta_j = "\eta_j$ happened" is the event $\Omega_1 \times \{\eta_j\} = \{(\omega_1, \eta_j), ..., (\omega_n, \eta_j)\}$.
Thus

$$\mathbb{P}(\omega_i) = \mathbb{P}(\omega_i, \eta_1) + ... + \mathbb{P}(\omega_i, \eta_m)$$
$$\mathbb{P}(\eta_j) = \mathbb{P}(\omega_1, \eta_j) + ... + \mathbb{P}(\omega_n, \eta_j).$$

The probability on the first and the second experiment are called, respectively, the first and the second marginal

Definition 5.4. *Let $\Omega = \Omega_1 \times \Omega_2$ and let $\mathbb{P}$ be a probability of Ω (the joint probability of the two experiment. The first marginal $\mathbb{P}_1$ is the probability on Ω_1 defined by*

$$\mathbb{P}_1(\omega_i) = \mathbb{P}(\omega_i) = \mathbb{P}(\omega_i, \eta_1) + ... + \mathbb{P}(\omega_i, \eta_m). \qquad \boxed{\texttt{d:marg1}}$$

The second marginal is the probability $\mathbb{P}_2$ on Ω_2 defined by

$$\mathbb{P}_2(\eta_j) = \mathbb{P}(\eta_j) = \mathbb{P}(\omega_1, \eta_j) + ... + \mathbb{P}(\omega_n, \eta_j). \qquad \boxed{\texttt{d:marg2}}$$

(Note the abuse of notation of confusing the elementary event $\omega \in \Omega_1$ with the event $\omega \subset \Omega$.)

Ex. 30 — On $\Omega = \{0, 1\} \times \{0, 1\}$ with probability $\mathbb{P}$ defined by

$$\mathbb{P}(0, 0) = (1 - q)(1 - p) , \quad \mathbb{P}(0, 1) = (1 - p)q,$$
$$\mathbb{P}(1, 0) = p(1 - q), \quad \mathbb{P}(1, 1) = pq,$$

find the marginals of $\mathbb{P}$.

Ex. 31 — Look at the examples at the beginning of this section ($\mathbb{Q}(00) = \mathbb{Q}(11) = 1/2$, and $\mathbb{P}(01) = \mathbb{P}(10) = 1/2$ for the glued coins, $\mathbb{P}(ij) = 1/4$ for all $ij \in \Omega$ for the independent coins) and check that they have the same marginals.

Ex. 32 — Consider on $\Omega = \{(r, r), (r, b), (r, g), (b, r), (b, b), (b, g), (g, r), (g, b), (g, g)\} = \{(r, b, g)\} \times \{(r, b, g)\}$ the probability

$$\mathbb{P}((b, b)) = \mathbb{P}((g, g)) = \mathbb{P}((r, r)) = 0,$$

and $1/6$ for the other elementary events. Calculate the marginal over Ω_2, $\mathbb{P}_2$. Repeat the exercise with the uniform probability $\mathbb{Q}$ on Ω.

As we saw in the example of the previous section, the marginals is not enough to recover $\mathbb{P}$ (Since different $\mathbb{P}$ can have the same marginals). We can recover $\mathbb{P}$ if know the probability of the second experiment once a given a realization of the first experiment (or viceversa). To see this

$$\mathbb{P}(\omega_i, \eta_j) = \mathbb{P}(\eta_j|\omega_i)\mathbb{P}(\omega_i) = \mathbb{P}(\eta_j|\omega_i)\mathbb{P}_1(\omega_i), \qquad (7) \quad \boxed{\texttt{e:p}}$$

where we recall that $\mathbb{P}(\eta_j|\omega_i) = \mathbb{P}(\omega_i, \eta_j)/\mathbb{P}(\omega_i).$ is a probability on Ω_2 for each fixed i. Thus, in order to recover $\mathbb{P}$ it suffices to know the probability

of the first experiment $\mathbb{P}_1$ and of the conditional probabilities on the second experiment $\mathbb{P}(\cdot|\omega_i)$, for $i = 1, ..., n$. The interpretation of (7) is that first you observe the first experiment, and then we look at the second knowing what happened in the first. Usually are the probabilities $\mathbb{P}_1$ and $\mathbb{P}(\cdot|\omega_i)$ which are given, and $\mathbb{P}$ (7).

Ex. 33 — Repeat the previous exercise, but this time calculate the conditional probabilities $\mathbb{P}_{2|1}$.

5.4 Independent experiments

When you regard the experiments as independent, then the probabilities $\mathbb{P}(\eta_j|\omega_i) = \mathbb{P}_2(\eta_j)$. In this case $\mathbb{P}(\omega_i, \eta_j) = \mathbb{P}_1(\omega_i)\mathbb{P}(\eta_j)$, so that $\mathbb{P}$ is determined by the probability of the two experiments. However you have to keep in mind that independence has to be regarded as an hypotheses on the probability $\mathbb{P}(\eta_j|\omega_i)$ and you cannot tell that it is true a priori: it is a property of the probability that you are assignign in Ω.

Definition 5.5. *We say that the two experiments are independent if*

$$\mathbb{P}(\omega_i, \eta_j) = \mathbb{P}("\omega_i happens")\mathbb{P}("\eta_j" happens) = \mathbb{P}_1(\omega_i)\mathbb{P}_2(\eta_j). \qquad (8) \quad \boxed{\texttt{e:indexp}}$$

The relation (8) is called independence because is equivalent to $\mathbb{P}("\eta_j\text{happens}|"\omega_i \text{ happens}") = \mathbb{P}("\eta_j happens")$, that is, the probability that η_j happens does not change if we observe ω_j.

Example 5.6. *the sample space for the roll of a die is $\Omega = \{1, ...6\}$. It is natural to put the uniform probability to each roll of the die, that is $\mathbb{P}(1) = ... = \mathbb{P}(6) = 1/6$. Why did we use the uniform probability on the outcomes of the roll of two dice? In this case $\Omega = \{(1, 1), (1, 2), ..., (6, 6)\}$. The fact is that the two rolls are independent:*

$$\mathbb{P}(i, j) = \mathbb{P}(i)\mathbb{P}(j) = 1/36 \quad \text{for every } i, j = 1, .., 6. \qquad \boxed{\texttt{e:uniform}}$$

This is precisely the uniform probability on Ω.

Ex. 34 — Assume that two experiments $\Omega_1 = \{\omega_1, ..., \omega_n\}$ $\Omega_2 = \{\eta_1, ..., \eta_m\}$ have the uniform probability (let $\mathbb{P}_i$ denote the uniform probability of Ω_i). Show that, if they are independent, then the joint experiment $\Omega = \Omega_1 \times \Omega_2$ have the uniform probability.

Answer (Ex. 34) —

$$\mathbb{P}(\omega_i, \eta_j) = \mathbb{P}_1(\omega_i)\mathbb{P}_2(\eta_j) = \frac{1}{n}\frac{1}{m}$$

Since the number of elements in Ω is $m \cdot n$, this is precisely the uniform probability.

5.5   Extractions

We are now able to discuss rigourously the probability distribution of two extractions.

Ex. 35 — In an urn there are 4 red balls and 2 blue balls. We draw a ball, we reinsert it and then, independently from the first draw, we draw another ball. What is the probability of drawing balls of different color?

Ex. 36 — There are 8 lottery tickets, and each one of them is winning with probability $1/8$ *independently* from the others. You buy 3 tickets. What is the probability that at least one of them is winning. Hint: consider the complementary event and use the independence.

Ex. 37 — There are 8 lottery tickets, 4 of which are winning. You buy 3 tickets. What is the probability that at least one of them is winning. Hint: consider the complementary event and define the number

$$X_i = \begin{cases} 1 & \text{if i-th first ticket is losing} \\ 0 & \text{if the i-th ticket is winning.} \end{cases}$$

for $i = 1, 2, 3$. Using formula (**??**) you should be able to calculate $\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1)$.

Answer (Ex. 37) — We need to compute the probability of the event A=" At least one ticket is winning". Consider the event A^c=" All the tickets are losing". We can proceed as follows: number your tickets from 1 to r. Denote

$$X_i = \begin{cases} 1 & \text{if i-th first ticket is losing} \\ 0 & \text{if the i-th ticket is winning.} \end{cases}$$

The event A^c is the event $(X_1 = 1, ..., X_r = r)$. We are therefore in the situation of the Bernoulli trials and we can use formula (**??**) and write

$$\mathbb{P}(X_1 = 1, ..., X_r = 1) = \mathbb{P}(X_r = 1|X_{r-1} = 1, ..., X_1 = 1)...\mathbb{P}(X_2 = 1|X_1 = 1)\mathbb{P}(X_1 = 1).$$

We now need to observe that the $\mathbb{P}(X_k = 1|X_{k-1} = 1...X_1 = 1) = (n - m - (k-1))/n - (k-1)$, since knowing that the tickets 1,...k-1 are losing, there are left $n - (k-1)$ tickets, of which $n - m - (k-1)$ are losing. Thus

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{(n-m)(n-m-1)...(n-m-(r-1))}{n(n-1)...(n-(r-1)).}$$