AN INTRODUCTION TO STATISTICS USING R

BY

ALESSANDRO GUBBIOTTI

Department of Mathematics

Higher School of Economics

August 18, 2025

Illustration from Alexander Pushkin's "Eugene Onegin."
The poem has been object of the first stattistical study
involving Markov Chains.

0.1 INTRODUCTION

0.1.1 Scope of the Lecture Notes

The first part of the present lecture notes is an introduction to probability theory. Many texts that undertake this task exist and it would be pretensious to think that these notes are particularly better than other existing ones. However, if the notes are oriented towards a mathematical audience, they often omit to explain the reason why objects are introduced and why measure theory is used as a framework to model probability theory. As an example, the notion of sample space is considered as a god given notion, while from a probabilistic point of view it is simply the set of some particular events, the elementary events. The goal of the first part of lecture notes is to fill this gap by providing a mathematically sound introduction (though only in the discrete setting) to probability theory from the point of view of probability theory, rather than from the mesure theoretic point of view. Therefore, the target audience for this part are novice to the field but also those who, having taken a probabity theory course, or, better said, a measure theory course, would like to take this notes as a starting point to rethink at the already known notions from a more probabilistic perspective. The second part consist in a brief and rigorous introduction to statistics. Statistics is a wide subject whose claims are often misinterpreted due to lack of rigor. The aim of these notes is to introduce simple models, and to fully develop them pointing out the critcal points and statements. The target audience are novice to the field and mathematician wishing to see some applications of statistics without delving into the general theory.

0.1.2 Models and claims of empirical sciences

ty˙motivation

One of the pillars of any science is the scientific method. To validate its claims, the measurable outcomes of a scientific theory must be compared with the data that can be collected in the real world. While there are discussions on how to infer information from the collected data, there is no discussion on the fact that a theory must prove its correctness through quantitative considerations.

When dealing with more quantitative, low level sciences such as physics, claims are very precise (think at $f = ma$ ot $f = Gm_1m_2/r^2$). Surprisingly (see [**wigner**]), those kind of claims involve few variables and, up to a certain degree of precision of the measuring instrument, do not involve uncertainty and are universally valid. To validate the formula $f = ma$ it is sufficient to measure the force, the mass and the acceleration of an object and to check whether the relation $f = ma$ holds.
The situation changes completely when dealing with more high level, involved and empirical sciences, such as economics, linguistics or social sciences. As the next example shows, most of the times, constitutive equations such as $f = ma$ - the ultimate goal of any science - have to be abandoned in favour of empirical relations which are true only to some extent and do

involve probability theory.
Indeed, consider a linguistic experiments in which respondents are subjected to a stimulus specifically studied to make them say a word. Suppose that this word can be spelled in two different ways, a formal standard variant and a dialectal variant, and the experiment aims to study the factors determining the choice. What would correspond to a deterministic, fundamental relation, and a complete solution to the research problem would be a set of variables $v_1, v_2, \ldots$ and a function that relates the variables to the choice made by the respondent. Concretely, examples of such variables could be v_1, the age of the respondent, v_2, his/her degree of education, v_3, the city where he/she was born, and other variabls could indicate the context in which the respondent is found when saying the word, his/her family education. A fundamental relation would then be a function f that associates to each possible values the variables $v_1, \ldots,$ can attain the variant used by a respondant for which the variables $v_1, \ldots$, namely

$$\begin{array}{rcl} f : \Omega & \rightarrow & \{\text{Formal}, \text{Dialectal}\} \\ v_1, v_2, \ldots & \rightarrow & f(v_1, v_2, \ldots) = \text{The variant the respondent will use,} \end{array} \quad (1) \quad \boxed{\text{e:impossible}}$$

where Ω is the set of all the possible values for the variables $v_1, v_2, \ldots$.
Even assuming that such relation could in principle be found, it is in practice impossible to obtain and useless, since most of the variables would be unknown to us and the function f would be too complicated. A possible alternative description is given by the following probabilistic model. We loose track of all the variables with the exception of the ones we consider more relevant regard for the study, say v_1, v_2 and v_3. The respondenti will then use the formal spelling with probability $p = p(v_1, v_2, v_3)$, that is.

$$f_{\text{eff}}(v_1, v_2, v_3) = \begin{cases} \text{Formal} & \text{with probability } p = p(v_1, v_2, \\ \text{with probability } 1 - p = 1 - p(v_1, v_2, v_3) \end{cases} \quad (2) \quad \boxed{\text{e:possible}}$$

The difference between the two models is clear. In the first case we can make deterministic claims such as" a respondent for which $v_1 = 39$, $v_2 = 0.3$ (on a scale between 0 and 1 computed in some way), v_3 = Paris, v_4 = Formal,... will use the Dialectal variable". The second model allows to make probabilistic claims such as "If the respondent has an higher degree of education (v_2), it is more probable that he/she will use the formal variant(that is, p is increasing in v_2)". We refer to Section (**??**) for a discussion of causality in a simple effectivce model known as linear regression model. With effective models such as (2), we loose the notion of causality.
Lastly we note that the relation between (1) and (2) is very complicated and is the object of study in many specific models. We simply note here that the probability present in (2) stems up?? from the uncertainty or lack of knowledge we have in the variables $v_4, ..$ and on the relation of f in (1).

0.1.3 The role of Probability and Statistics

The role of probability is used to model the uncertainty and the lack of knowledge and is the framework within which statistics works. In the Introduction **??** we introduce the notion of probability model and we point out that in order to infer information from the real world data we need a probabilistic model. Here we anticipate the main It allows to see the data obained from the real word, say the experiment introduced before where a respondent says a word which has two possible spellings, as an instance between many possible and legit outcomes: other results could have been possible but we observe only one (this is the idea of the sample space in Definition 0.3.3 and Section **??** for more details). Moreover, it The goal of probability is to introduce a framework where data generated in 1a random way (that is, in such a way that other results could have been possible) can be rigorously defined, see Definition **??**. The role of statistics is the converse one, and, once assumed that the real world data is generated as models given by probability theory, to infer from the data information about the model. As an example on what statistics does think about trying to estimate the parameter $p = p(v_1, v_2, v_3)$ in the probabilistic model introduced before.

CONTENTS

Part I

PROBABILITY

0.2 INTRODUCTION

s:intro

Probability Theory is a mathematical framework introduced for quantifying uncertainty

1. The probability that tomorrow will rain is 0.75.
2. The probability that the coin that I am about to toss will show its head is 1/2. e:stat1

The intuitive meaning of the above statements is clear, and we will not delve into the interesting discussion of trying to give a precise meaning to them. Indeed, Probability theory does not discuss the meaning or the veradicity of the above statements, but it rather gives a formal framework to work with them in a consistent way that matches our intuition. For instance, statement 2 is not universally true since one can easily cheat with coin tosses and make heads always appear. A possible approach to take is the *subjective* approach, where the probability of an event can be seen as a subjective evaluation of the degree of trust in the event.
To get an idea of which objects we need to formalise, we now proceed to decompose and to analyse the above statements. Both of them are of the form "The probability of *something* is a number". That *something* in the first case is E=" Tomorrow will rain", and in the second case F=" The coin that I am about to toss will show heads", and it will be called an *event*. Events are the objets to which probability refers and the probability of an event is a number, 0.75 in the first statement and 0.5, in the second. We now make an intermediate step and rewrite the above statements as

- The probability of E is 0.75
- The probability of F is 0.5

Therefore, the probability of an event is a number and the probability is an operation that associates to each event a number. Denoting "The probability of" by $\mathbb{P}()$, the above statements can be rewritten as

- $\mathbb{P}(E) = 0.75$
- $\mathbb{P}(F) = 0.5.$

and $\mathbb{P}$ can be seen as a function that takes in input an event and whose output is a number p between 0 and 1

$$\begin{array}{rcl} \mathbb{P}: \text{Events} & \mapsto & [0,1] \\ E & \to & \mathbb{P}(E) \in [0,1]. \end{array}$$

Therefore a typical probabilisitc statement assumes the form of

- The probability of the event E is p, or, $\mathbb{P}(E) = p$,

where $p \in [0,1]$.

0.3 EVENTS AND SAMPLE SPACE

Motivated by the Introduction 0.2 we first introduce the notion of events. We then see that events can be modelled using set theory, after introducing the sample space

0.3.1 Definition of an event

In light of the subjective perspective on Probability Theory we will refer to the person who is evaluating the probabilities of events using "you". Following [**definetti**] we can define events as An event is an unambiguous statement which can be either true or false and about which you are uncertain. Informally, unambiguous statement means that a possible bet or insurance based upon it can be decided without question. Examples of events are: " The next ChatGPT sentence will be incorrect" or " Napoleon was born on March" (to decide this last statement, I should open Wikipedia). Events will be usually denoted by capital letters, $A, B, \ldots$, and the impossible event will be usually denoted by $\emptyset$.

d:event

0.3.2 Operations and relations between events

ss:operations

From two events E and F it is possible to consider new events using the operations "or" "and", and "not", namely, to consider "E and F", "E or F" and " not F". To make a simple example let E= "I will obtain my Ph.D" and F=" Tomorrow it will rain". Then

- "E and F"= " Tomorrow it will rain and I will obtain my Ph.D ".
- "E or G"= "Tomorrow either it will rain or I will take my Ph.D ".
- "not E"= " I will not take my Ph.D".

Given two events E or F we say that E implies F and we write (for reasons that will become clear in the following) $E \subset F$ if, F is true whenever E is. If G = "Tomorrow it will rain heavily", then $G \subset F$. Note that in general

- "E and F" $\subset E$.
- $F \subset$ "E or$F$".

0.3.3 Elementary Events and Sample Space

There are no inherent restrictions on the events one can define and on which you could try to assign a probability. When doing a probabilistic study, you might wish to restrict your attention and consider only particular events. For example, 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".

Example 0.1

You are paid by a shop owner to perform a statistical analysis of the number and the kind of customers visiting his shop. For time and economic constraints, you have to choose what kind of events you are interested in. One possibility is to consider only the number of customers, so that events will be of the form "More than 10 customers will visit the shop", " 3 customers will visit the shop"...
In the case the shop owner is also interested in the gender of the customers, you might wish to consider events that take into account the number of customers and their genders, such as "More than 3 male customers and less than 10 female customers will visit the shop", "2 male and 1 female will visit the shop".

From now on we assume that we have always set beforehand the events that we want to consider. We assume that given two events E and F that we want to consider, then any of the events that can be built from them is an event

d:measurable

that we want to consider [Measurable events] The set of *events we want to consider* $\mathfrak{F}$ is a set of events that is closed under intersection: If $E, F \in \mathfrak{F}$ then also $E \cap F \in \mathfrak{F}$ (if we want to consider E and F we also want to consider "E and F").

Example 0.2: Coin Toss

ex:cointoss

If we toss a coin and we are interested only in the face shown by it, then the events that h we want to consider are those that can be written in terms of theface shown by the coin: consists on the events we can write using the interested is

$$\mathfrak{F} = \{\emptyset, \text{"The coin gave heads"}, \text{"The coin gave tails"}, \\ \text{"The coin gave either H or T"} \}.$$

Once the events we want to consider are fixed and collected in $\mathfrak{F}$ it is possible to define a particular kind of events, the *elementary events* thanks to which we can model events using set theory. Informally, elementary events are the smallest among the events we want to consider, and can be regarded as the outcome of the "experiment" we are observing. It is indeed possible to define a notion of smallness using using the relation "implies" which we have denoted by $\subset$ is Section 0.3.2. An event E is "smaller" than F if $E \subset F$, and the elementary events are simply the smallest events we want

d:elementary

to consider. [Elementary Events] We say that an event $E \in \mathfrak{F}$ is elementary if for any other event F that we want to consider, then F does not imply E. That is, E is elementary if for any $F \in \mathfrak{F}$ with $F \subset E$, then $F = E$.

Ex. 1 — Show that two different elementary events are mutually disjoint, that is, that if E and F are elementary and $E \neq F$, then "E and F" = $\emptyset$.

Answer (Ex. 1) — In light of the assumptions made in Definition 0.3.3, "E and F" $\in \mathfrak{F}$. Since "E and $F \subset E$, if the event "E and F" is possible, it contradicts Definition 0.3.3

The notion of elementary events is fundamental since it allows to see all the previously introduced quantities as already encoded using set theory, with which everyone is familiar

d:sample˙space

[Sample Space] Given $\mathfrak{F}$, the sample space Ω is the set of elementary events. Elements of the sample space, that is, elementary events, are usually denoted with ω. The sample space can be regarded as the set of possible oucomes of the system you are looking at with the degree fo precision that you want to impose. Some examples are in order.

Example 0.3: Sample Space of a Coin Toss

If $\mathfrak{F}$ is the one defined in (0.2), then the elementary events are $\omega_1 =$ " The result is heads" and $\omega_2 =$ " The result is tails", and the sample space is $\Omega = \{\omega_1, \omega_2\}$. Note that we can use a different notation and set $0 = \omega_1$ and $1 = \omega_2$. However, 0 and 1 are must not be mistakes for numbers (numbers are usually associated to unit of measures) . They are symbols indicating elementary events

Example 0.4: a die is rolled

ex:die˙sample

If a die is rolled, we migh wish to look at events of the form "The result of the die is an even number". Elementary events, the ones that cannot be decomposed just by looking at the die, are 1= "The result is 1" , ..., 6= "The result is 6". The sample space is therefore $\Omega = \{1, 2, 3, 4, 5, 6\}$ and, again,the elements 1,...,6 must not be mistaken for numbers, since are elementary events and, for instance, 1 + 3 doesn't make sense.

Example 0.5: A chess match

If you observe a chess match the events you want to consider are those that involve the game alone, such as " The game will end before 10 moves are taken" or "Black wins", while the elementary events are the possible single games. Therefore, the sample space is the set of all possible games. A full description of a game is usually given in terms of the so called algebraic notation. An example of a match is given by (a Fisher vs. Kasparov match) is given by
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 algebraic notation can be seen as a way to encode elementary events and the sample space. In Exercise 11 and 9 examples of sample spaces defined by simple games are given

0.3.4 Exercises

Ex. 2 — 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 imply 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"

Ex. 3 — 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 imply 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"

Ex. 4 — 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

ercise:marbles

Ex. 5 — ([**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.

Ex. 6 — Repeat Exercise 1 when the second marble is drawn without replacing the first marble.

Ex. 7 — You toss a coin 3 times. Describe the elementary events.

0.3.5 Events as Subsets of the Sample Space

ss:subset

[Elementary events favourable to an event] We say that an elementary event $\omega \in \Omega$ is in favour of $E \in \mathfrak{F}$ if ω implies E. The minimality of the elementary events implies that for each $E \in \mathfrak{F}$ and $\omega \in \Omega$ either ω is in favour of E or they are mutually exclusive (that is, "ω and F" $= \emptyset$). We can associate to each event the subset of Ω whose elements are the elementary events that imply E:

$$\begin{array}{rcc} \Phi : \mathfrak{F} & \mapsto & \{\text{subset of } \Omega\} = \{A\,, A \subset \Omega\} \\ E & \to & \{\omega \in \Omega\,, \omega \subset E\} \end{array}$$

e:abstract˙to˙c

ex:die˙subset

Example 0.6: Dice

We have seen in Example 0.4 that the sample space of a dice that is rolled can be identified with $\Omega = \{1,2,3,4,5,6\}$. The elementary events that are favourable to E = "The result is an odd number" are $1,3,5$. Therefore we identify E with the subset of Ω given by $\{1,3,5\}$.

From now on we will consider $\mathfrak{F}$ as a set of subsets of Ω. The fundamental fact of the identification of events with subsets of Ω given by Φ is that the operations "or", "and" and "not" correspond to the operations of $\cup$, $\cap$ and $()^c$, respectively, and that E implies F if and only if E is a subset of F. For instance, in the die example, if E = "The result is an odd number" and F = "The result is greater than 2", then "E and F" = " The result is an odd number greater than 2". The first two events are identified with the subsets of Ω given by $\{1,3,5\}$ and $\{3,4,5,6\}$, while "E and F" is identified with $\{3,5\}$, which is exactly the intersection of the two sets.

We can now model events using set theory and we can depict them using Venn's diagrams, as shown in picture 1

Points in the sample space are events on their own and it might be helpful to consider them as events in a larger and refined sample space Ω. This will be done rigorously is Section **??**

ex:sum˙restriction

Example 0.7: Two dice rolled

Two dice have been rolled and suppose that we can only look ate the sum of tehir results. The sample sspace is therefore $\Omega = \{2,3,4,\ldots,12\}$. If one is allowed to look at each die, then the sample space is $\Omega' = \{11,12,\ldots,66\}$, and the elementary events in Ω are not anymore elementary (apart from 2 and 12) in Ω'. Figure **??** represents this situation, which will be made rigorous in Example **??**

Venn.jpg

f:Venn

Figure 1: Venn's diagram

Another reason for which we would like to consider the points of Ω as events of in a lager sample space Ω' is that the latter can cointain many different sample spaces, as the next example shows. This fact will be again detailed in Section **??** and, in particular, in Subsection **??**.

Example 0.8

x:two˙coin˙sample

The sample space of a coin toss can be identified with $\Omega = \{H, T\}$, and if we toss another coint the sample space is again the same. If we consider a large Ω', depicted in Figure **??**, that contains also the two coin tosses, see Figure **??** and **??**, then we can consider also the sample space of the two coins $\tilde{\Omega} = \{HH, HT, TH, TT\}$, see Figure **??**.
Finally, we can look refine the sample space with other events, see Figure **??**.

Ex. 8 — ([**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.

0.3.6 Examples

Example 0.9: A Coin Toss

If in a coin toss we are interested only in the face the coin shows, as we always are when we speak about coin tosses, then the only two non-tricial events we want to consider are 0=" The result is tails" and 1 = " The result is heads", see also Example 0.2 and 0.6. Therefore

$$\Omega_1 = \{0, 1\},$$

and we will ofter refer to 1 as to a success. The set of events that we want to consider can be represented by

$$\mathfrak{F} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}.$$

Example 0.10: two, three and n coin tosses

ex:n˙coin

The sample space of two coin tosses can be encoded by $\Omega_2 = \{00, 01, 10, 11\} = \{x_1x_2\,,\, x_i \in \{0, 1\}i = 1, 2\}$, where, for example, If we toss twice a coin and we are only interested in the faces they show, then we are interested in events of the form " The first coin shows heads" or "The second coin shows tails". The smallest events we can make by intersecting this kind of events are of the form 01="The first toss gave tails and the second tails". Therefore the sample space is constituted by the set

$$\Omega_2 = \{00, 01, 10, 11\} = \{x_1x_2, x_i \in \{0, 1\}, \text{ for } i = 1, 2\}$$

of two digit string where every digit represents the outcome of the respective coin. Similarly, for 3 coin tosses

$$\Omega_3 = \{000, 001, 010, 011, 100, 101, 110, 111\} = \{x_1x_2x_3\,,\, x_i \in \{0, 1\}i = 1, 2, 3\},$$

and, in the general case of n coin tosses, the sample space is the space of strings of n digit with values in $\{0, 1\}$, namely

$$\Omega_n = \{x_1x_2 \ldots x_n\,, x_i \in \{0, 1\}, i = 1, \ldots, n\}. \tag{3}$$ e:ntoss

Events that we will often use when introducing the Binomial distribution, see Example **??** are the events E_k= "There has been exactly k

successes”, for $k \in \mathbb{N}$, $k < n$. For instance, if $n = 3$ and $k = 1$, is $E_1 = \{001, 010, 100\} = \{x_1x_2x_3\,,\, x_1 + x_2 + x_3 = 1\}$. The particular encoding given by (3) allows us to write in simple terms the condition imposed by E_k:

$$E_k = \{x_1 \dots x_n \in \Omega_n\,, x_1 + \dots + x_n = k\}.$$ e:ntoss˙ksuccesses

Example 0.11: Finite but Undetermined Amount of Coin Tosses

If we are observing a finite amount of coin tosses, but we do not know how many of them there will be, we can define the set of possible outcomes, that is, the sample space, via the following encoding

$$\begin{aligned}\Omega &= \bigcup_{i=1}^{\infty} \Omega_i = \{0, 1, 00, 01, 10, 11, 000, \dots\},\\ &= \{x_1 \dots x_n, n \in \mathbb{N}, x_i \in \{0,1\},\ \text{for } i = 1, \dots n\}\end{aligned}$$ e:finite˙coin

where Ω_i has been defined 3.

ex:infinite˙coin

Example 0.12: Infinite Coin Tosses

If we are observing an infinite amount of coin tosses, then the sample space is the set of strings whose digits denote the result of the corresponding coin.

$$\Omega = \{x_1 ... x_n ...\,,\, x_i \in \{0,1\}\}.$$ e:infinite˙coin

Possible events that we would like to take into account are of the form ”The first toss gave a_1, the second a_2,..., the nth a_n”, where $n \in \mathbb{N}$ and $a_i \in \{0,1\}$ for $i = 1, \dots, n$. This event correspods to the subset of Ω consisting on strings whose first n elements are a_i, for $i = 1, \dots, n$

$$A_{a_1,\dots,a_n} = \{x_1 \dots x_n \dots \in \Omega\ x_1 = a_1, \dots, x_n = a_n\}.$$ e:infinite˙coin˙set

ex:unif

Example 0.13: Random Number Between 0 and 1

By typing on R the command runif, the computer gives us a number X between 0 and 1. runif(1)

```
## [1] 0.7328875
```

Some events in which we are usually interested are " $X < 1/2$ ", " $X \in [1/4, 1/2)$". By intesecating the above events we can see, at least intuitively, that the smallest events one can make are $X = x$. The sample space is therefore composed by the events "$X = a$" for each $a \in [0,1]$, and, by denoting a the event $X = a$,

$$\Omega = \{" X = a", a \in [0,1]\} = [0,1]. \tag{4}$$ e:unif˙1

The set of events we want to consider $\mathfrak{F}$ contains the events of the form $X \in [a, b]$, where $0 \leq a < b \leq 1$, if we are using the notation in the centre of (4), and $[a, b]$ if we are using the other notation for Ω.

Example 0.12 and 0.13 constitute are different from the previous ones. In both cases the spaces are too big to be written as a list of elements, that is, in the form $\Omega = \{\omega_1, \omega_2, \ldots\}$. In jargon, Ω is not discrete. The difficulty is that now the specification of $\mathfrak{F}$ becomes important since, for good mathematical reasons, it is in general not possible to consider each $A \subset \Omega$.

0.3.7 Exercises

exercise:game˙2 **Ex. 9 —** ([**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 $(\bigcup_{i=1}^{\infty} E_i)^c$.

Ex. 10 — ([**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 $E \cap F \cap G$

e:game **Ex. 11 —** ([**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 $(A \cap B)^c$.

0.4 PROBABILITY ON DISCRETE SAMPLE SPACES

0.4.1 Introduction

In this section we are going to formally define what is a probability. As already said in the Introduction 0.2, a probability associates to each $E \in \mathfrak{F}$, the set of events that we want to consider, its probability, $\mathbb{P}(E)$. A probability is therefore a function

$$\begin{aligned} \mathbb{P}: \quad \mathfrak{F} &\mapsto \quad [0,1] \\ E \in \mathfrak{F} &\to \quad \mathbb{P}(E) \end{aligned}$$ e:prob˙def

Not every function from $\mathfrak{F}$ to $[0,1]$ has the right to be called a probability, and some relations must be satisfied. Consider the example of the two events $E =$"Tomorrow will be cloudy" and $F =$"Tomorrow it will rain" and of its associated probabilities $\mathbb{P}(E)$ and $\mathbb{P}(F)$. If F happens, also E happens, and the intuition tells us that the numbers $\mathbb{P}(E)$ and $\mathbb{P}(F)$ have to verify $\mathbb{P}(F) \leq \mathbb{P}(E)$. This monotonicity property, for instance, will be proved in Corollary 0.4.4
In this notes we will not introduce the definition of probability in the full general context. We rather restrict to finite sample space $\Omega = \{\omega_1, \dots, \omega_n\}$, for some $n \in \mathbb{N}$. Up to replacing finite sums with infinite sums, all the definitions and the results below generalise immediately to the discrete infinte setting $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$, where the sample space can be written as a list of elements. Unfortunately, many sample spaces that are used in practice, such as $\Omega = [0,1]$, see Example 0.13 or $\Omega = \{x_1 x_2 \dots, \ x_i \in \{0,1\} \ \text{for } i \in \{0,1\}\ \}$ defined in 0.12, are simply too big to be written as a list of elements. In such cases the theory gets more involved and invokes mathematical results from measure theory. The reason is that the usual requirements that a function $\mathbb{P}: \mathfrak{F} \to [0,1]$, the *Kolmogorov axioms*,, are too strict to allow each $E \subset \Omega$ to be in $\mathfrak{F}$, and one has to restrict to a so-called σ-algebra of subsets of Ω, a notion that this notes carefully avoid to introduce. Nevertheless, in such cases, we will evaluate the probabilities of particular events when dealing with Bernoulli trials, see Definition **??** and with continuous random variables, see Definition **??**.

0.4.2 Definition of a probability on a finite sample space

Hereafter we consider a finite sample space $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, for some $n \in \mathbb{N}$, and we consider every event that can be written in terms of the elementary events $\omega_1, \dots, \omega_n$, namely

$$\mathfrak{F} = \{A\, , A \subset \Omega\}$$ e:f˙all

consists in every subset of Ω (including the impossible event $\emptyset$ and the certain event Ω). The particularity of the discrete case we are considering is that in order to define a probability on Ω it is sufficient to define the probabilities of the elementary events [Probability of the elementay events] A ob˙elementary probability for the elementary events ω_i, $i = 1, \ldots, n$ is a collection numbers $p_1, \ldots, p_n$ verifying that $p_i \in [0,1]$ for each $i = 1, \ldots, n$ and

$$p_1 + p_2 + \ldots + p_n = 1 \tag{5}$$

e:normalizatio

The condition (11) has the intuitive meaning that we regard the event in which none of the ω_i happens as impossible, see Proposition 0.4.4.

Example 0.14

ex:definition

The sample space of a rolled die is $\Omega = \{1, 2, 3, 4, 5, 6\}$. Definition 0.4.2 allows for different choices of probabilities on elementary events, but a typical choice is

$$p_1 = \ldots = p_6 = 1/6,$$

e:die˙uni

which corresponds to to a perfectly balanced die or a die about which we do not have any information. A probability corresponding to an unfair die could look like

$$p_1 = p_2 = p_3 = 1/4, \qquad p_4 = p_5 = p_6 = 1/12. \tag{6}$$

e:die˙unfair

The numbers p_i, for $i = 1, \ldots, n$, are the probability of the elementary event ω_i, namely $\mathbb{P}(\{\omega_i\}) := p_i$. We will sometimes, for instance in (7), use the different notation $\mathbb{P}(\{\omega\}) = p_\omega$, which avoids to use the index i with which the elements of Ω are ordered and that, therefore, does not depend on the particular way the elements are ordered. If $\omega = \omega_i$, then the two notations are related by $p_i \equiv p_{\omega_i}$.

Example 0.15

ex:coin˙toss˙notation

This example ilustrates the difference in the two possible notations we have introduced. Consider Ω_2, the sample space of two coin tosses given in Example 0.10. Ordering the elements $\Omega = \{00, 01, 10, 11\}$ by $\omega_1 = 00$, $\omega_2 = 01$, $\omega_3 = 10$ and $\omega_4 = 11$, then a possible probability using the notation introduced in Definition 0.4.2 is given by

$$p_1 = 1/8\,, p_2 = 1/8\,, p_3 = 1/2\,, p_4 = 1/4$$

and the same probability, using the equivalent notation p_ω, is given by

$$p_{00} = 1/8,\ p_{01} = 1/8,\ p_{10} = 1/2,\ p_{11} = 1/4$$

Note that the possible ways of choosing p_ii, for $i = 1, \ldots, n$, are quite arbitrary and the only restrictions involved in the choice is that $p_i \in [0,1]$ and (11), which basically means that there are $n - 1$ numbers (which have to positive and whose sum has to be less than 1) to fix. However, in the discrete setting, this is all the freedom at our disposal, and once the probabilities of the elementary events are assigned, the probability of every event $E \in \mathfrak{F}$ is determined by the sum of the probabilities of the elementary events that compose it, according to the following definition

d:prob Let $p_1, p_2, \ldots, p_n$, $n \in \mathbb{N}$ be a probability for the elementary events, see Definition 0.4.2, and let $A \subset \Omega$ be an event. The probability of A is the sum of the probabilities of the elementary events that compose it

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega \tag{7}$$

e:prob

Using a different notation, any subset A of Ω can be written as

$$A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}, \tag{8}$$

e:notation

for some $k \leq n$ and set of indexes $1 \leq i_1 < i_2 < \ldots < i_k \leq n$. Its probability is then given by

$$\mathbb{P}(A) = p_{\omega_{i_1}} + \ldots + p_{i_k} \equiv p_{i_1} + \ldots + p_{i_k} \tag{9}$$

e:prob2

Example 0.16

ex:dice˙prob

We come back to Example 0.14 in order to see Definition 0.4.2 in action and to get better acquinted with the two different notations (7) and (9). The event "The result is an odd number" corresponds to the subset $\{1, 3, 5\}$ and (7) tells that the two probabilities $\mathbb{P}$ and $\mathbb{Q}$ defined by the choices (**??**) and (6), respectively, evaluate the probability of this event by

$$\mathbb{P}(\{1,3,5\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{3\}) + \mathbb{P}(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

e:die˙prob˙uni

and

$$\mathbb{Q}(\{1,3,5\}) = \mathbb{Q}(\{1\}) + \mathbb{Q}(\{3\}) + \mathbb{Q}(\{5\}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{12} = \frac{7}{12},$$

e:die˙prob˙unfair

respectively. In order to use the notation in (9), we rewrite the set $\{1, 3, 5\}$ as in (8) by setting $k = 3$ and $i_1 = 1$, $i_2 = 3$ and $i_3 = 5$. Observe also that $\mathbb{P}(\Omega) = \mathbb{Q}(\Omega) = 1$ thanks to (11).

We finally collect the definitions of the present section

d:prob3 A *probability* over Ω is a function

$$\begin{array}{rrl} \mathbb{P}: & \mathfrak{F} \to & [0,1] \\ & E \mapsto & \mathbb{P}(E) \end{array}$$

such that $p_i := \mathbb{P}(\{\omega_i\})$ forms a probability on elementary events in the sense of Definition 0.4.2 and (7) (or (9)) holds.

0.4.3 Uniform probability

In the case in which you evaluate each elementary event equally likely to occur, or in the case in which you don't have any knowledge to say that any particular elementary event is more likely to occur, a possible chice is to assign to each elementary event the same probability $p \in [0, 1]$. To determine p we use (11)

$$p_1 + \ldots + p_n = \overbrace{p + \ldots p}^{n \text{ times}} = np = 1,$$ e:uniform1

from which we obtain that $p = 1/n$. To compute the probability of an event composed by k elementary events $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$, with $k \leq n$ and $1 \leq i_1 < \ldots < i_k \leq n$ we use (9) to obtain

$$\mathbb{P}(A) = p_{i_1} + \ldots + p_{i_k} = \overbrace{\frac{1}{n} + \ldots + \frac{1}{n}}^{k\text{times}} = \frac{k}{n}, \tag{10}$$ e:uniform2

which depends only on k, the number of elementary events that compose A. Denoting by $|E|$ the number of elements of the set $E \subset \Omega$, one can rewrite (10) as the number of elements that compose A divided by the total number of events

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

To get acquinted with the notation independent from the ordering used in (7), we repeat the same computations

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} p = p \left(\sum_{\omega \in A} 1 \right) = p|A| = \frac{|A|}{|\Omega|},$$

d:uniform and obtain [Uniform Probability] Let Ω be a finite sample space. The uniform probability is defined by

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|},$$ e:uniform

or each $E \subset \Omega$. Basically, computing the uniform probability reduces to counting configurations. Unfortunately, despite being conceptually simple, it is often computationally unfeasible. The same holds for general discrete probabilities, which can be seen as a way to count objects giving different weights to them. The uniform probability is often mistakenly taken as the definition of probability. This has led to some abuse of notation such as the following

Example 0.17: Extracting Randomly a Person

ex:person Consider extracting a person between 20, and denote $\Omega = \{1, \ldots, 20\}$ the sample space, where i denotes the elementary event "the i-th per-

son is extracted". Usually what is meant by the sentence "a person is extracted randomly" is that a person is selected with uniform probability, namely $\mathbb{P}(\{i\}) = 1/20$. However, from our point of view, a Nobel prize is also chosen randomly among the possible candidates, but there is no reason why the probability should be the uniform one. Another that can be ambiguous are " we choose a person as randomly as possible", by meaning as fairly or as uniform as possible.
As a side note, despite being natural to model the choice of a person with uniform probabiity, when doing a survey or making a poll statisticians struggle to make their choice adhere to this model. When mobile phones were not so widespread, doing surveys by phone call excluded a portion of the population, for instance. See Example **??**.

Sometimes probabilities that are not uniform can be seen as arising from a uniform probability space (in the sense of Definition **??**, see Example **??**)

Example 0.18: Sum of two dice

d:two˙dice˙sum

Consider the sample space of two dice

$$\Omega' = \{11, 12...., 66\} = \{x_1x_2 : x_1, x_2 \in \{1, ...6\}\}$$

and endow it with the uniform probability. We will discuss in **??** the reason of this choice, but we anticipate that it is due to the fact that the dice are *fair* and they are rolled *independently*, see Subsectionss:ind?????. Assume we can only look at the sum of the results. This leads to a *coarse grained* sample space, see Example **??** where the elementary events, that is, the elements of the new sample space Ω, are the events described in Ω' by $\{(x_1, x_2) \in \Omega' : x_1 + x_2 = k\}$, and whose probabilities are

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

sum

Therefore, the probability in Ω is not the uniform one.

Example 0.19: Extractions with replacement

ex:replace

Consider the sample space given by extractions with replacements, which are the object of Exercise 5. It makes sense to consider each outcome equally probable, that is

$$p_{(a,b)} = 1/9 \quad \text{for each } (a,b) \in \Omega.$$

e:extraction˙repla

This choice of a probability is a consequence of two properties, the fact that the extractions are *fair* and *independent* , see also ??? for a more general example. The fact that the extractions are independent is sometimes hidden in the extraction procedure: between two extractions we mix the balls.

Example 0.20: Extractions without replacement

ex:no˙replace

If in the extractions from the bowl of Example 0.19 and Exercise 0.20 we do not reinsert the ball the sample space is smaller or, equivalently, we assign $p_{rr} = p_{nn} = p_{gg} = 0$. Nonetheless the symmetry between the colours of the marbles suggests that we might regard each other outcome as equally probable. This is again a consequence of the two following facts, which will be discussed in a more general setting in (**??**). Each single extraction is fair and, since we mix the marbles before drawing again, the second extraction depends on the first one only because of the fact that the composition of the bowl changes.

Ex. 12 — A coin is tossed 4 times. Recall that the state space is $\Omega = \{0000, 0001, ..., 1111\}$, where a 1 means that the associated coin gave heads. Suppose that the probability on Ω is uniform (we will see in Section **??** 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 exactly two heads
3. There has been a head after a tail (that is, we observed a 01)
4. The event is obtained by fixing the first digit to 1 and allowing the other to take arbitrarily the values 0 or 1, that is, $E = \{1x_2x_3x_4,\ x_i \in \{0,1\}\}$. Since $|E| = 2^3$ and $|\Omega| = 2^4$, by formula (**??**) we obtain

$$\mathbb{P}(E) = 2^3/2^4 = \frac{1}{2}.$$

Ex. 13 — We throw two dice, and we suppose that the each output has the same probability (Section 0.5.2 will justify this choice with the hypothesis of independence and fairness of the dice)

1. Calculate the probability of obtaining a 2 and a 4

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

0.4.4 Basic properties

Let Ω be a sample space and $\mathbb{P}$ a probability defined on it. The first property
p1 tells us that there are no possible outcomes outside the ω_i.

$$\mathbb{P}(\Omega) = 1 \tag{11}$$ e:normalization

Proof. It is a direct consequence of (11):

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

□

We now formulate the *additivity* property of the probability. [Additivity]
p:monotone Any two *mutually exclusive* events A B, that is, events that satisfy $A \cap B = \emptyset$, verify

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).i \tag{12}$$ e:add

p:exclusive More generally, Any two events A, $B \subset \Omega$, it holds

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

The proof looks more involved than what really is. The idea is simple and is the following up of the graphycal picture of the events as subset of the sample space using the Venn diagram in Figure **??**. A probability can be seen as a measure of the areas covered by the events, and we can see that $\mathbb{P}(A) + \mathbb{P}(B)$ counts twice the area of $A \cap B$, while only once the area in $A \cup B$ that is not in $A \cap B$.

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 0.4.2,

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

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

and

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

In this way,

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

□

□

Let $A, B \subset \Omega$ be two mutually exclusive events, that is, two events such that $A \cap B = \emptyset$. Then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \tag{13}$$ e:addc

Proof. Use (12) and note that $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$. □

c:monotonicity [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, since if there is a storm there is also bad weather.

Proof. The intuition of the proof is that when we sum over the elementary events in B we are summing over the elementary events in A and the elementary events in $B \setminus A$. Therefore we write B as $A \cup (B \setminus A)$ and using (13) we obtain that $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$. □

c:complementary Let $A \subset \Omega$ be an event. The

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

Proof. Write $\Omega = A \cup A^c$, apply (13) with $B = A^c$ and use (11). □

This property is also 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.

0.4.5 An Implementation Well Modelled by Probability

ss:r˙prob1

In this section we give some possible
Given an n and some probabilities on elementary events $p_1, ..., p_n$, see Definition **??**, the computer is able to generate random outcomes with those probabilities. The following chunk of code, for instance, generates a natural number between 1 and 6 with probabilities proporional to p_aux, that is, with probabilities p. n ¡- 6 The number of elements of our sample space p$_a ux < -c(1,2,3,0.5,1,5) Here I take the$p$_i$ to be positive numbers whose sum needs not to be 1 p ¡- p$_a ux/sum(p_a ux) Now I normalise them so that they sum 1 sample(1:n,1,prob=$

[1] 1

This means that, defining ω_i = "The output is i ", $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $p_1, \ldots, p_n$ is a good model for the above output. Other systems for which the probability models work well that we have introduced before are Example **??** and the random extractions of balls from an urn in Example 0.4.4.

This is to be put later, after introducing inedependent sampling or maybes Extractions with or without replacement. The R command sample allows to implement the model of extractions with or without replacemen.

0.5 CONDITIONAL PROBABILITY

In general, our evaluation of the probabilities changes if our state of information changes. If we observe that the sky is cloudy, our evaluation of the probability that it will rain will become greater. Given an event E, if $\mathbb{P}(\cdot)$ denotes the probability before the knowledge of E, we will denote the new probability, given the knowledge of E, by $\mathbb{P}(\cdot|E)$. With this notations, the previous example reads $\mathbb{P}($"Today it will rain "|"It is cloudy"$) >$ $\mathbb{P}($"It will rain today "$)$. In case the *conditioning* event E can be written in terms of the sample space, that is, in case $E \subset \Omega$, $\mathbb{P}(\cdot|E)$ can be described in simple terms.

0.5.1 Definition and first examples

Let $\Omega = \{\omega_1, ..., \omega_n\}$ be the sample space, $\mathbb{P}$ a probability over Ω, see Definition 0.4.2, and let $B \subset \Omega$ be an event. The probability *conditioned to B* is a new probability denoted by the symbol $\mathbb{P}(\cdot|B)$. Intuitively, it has to satisfy that

$$\mathbb{P}((\{\omega\}|E) = 0 \qquad \text{if } \omega \notin E \qquad \boxed{\text{e:cond1}}$$

and, since we do not have information on which event $\omega \in E$ caused E, we would like to maintain the proportions

$$\mathbb{P}(\{\omega\}|E) = C\mathbb{P}(\{\omega\}) = Cp_\omega, \qquad \boxed{\text{e:cond2}}$$

for some constant C. The constant C is to be determined by (11):

$$1 = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}|E) = \sum_{\omega\in E} C\mathbb{P}(\{\omega\}) = C\mathbb{P}(E) \tag{14} \qquad \boxed{\text{e:cond3}}$$

so that $C = \mathbb{P}(E)$. For the sake of concreteness we rewrite (14) using a different and extended notation. Let $E = \{\omega_{i_1}, \dots, \omega_{i_k}\}$, for some $k \in \mathbb{N}$ and $1 \le i_1 < \dots < i_k \le n$ be an event, then

$$\begin{aligned} 1 &= \mathbb{P}(\{\omega_1, \dots, \omega_n\}|\{\omega_{i_1}, \dots, \omega_{i_k}\}) \\ &= \sum_{i=1}^{n} \mathbb{P}(\{\omega_i\}|\{\omega_{i_1}, \dots, \omega_{i_k}\}) = C\sum_{j=1}^{k} \mathbb{P}(\{\omega_{i_j}\}) = C\mathbb{P}(E) \end{aligned} \qquad \boxed{\text{e:cond4}}$$

Therefore $C = \frac{1}{\mathbb{P}(E)}$ and the probability of an ev

that is, we denote the probability of A conditioned to B by $\mathbb{P}(A|B)$. Of course, the new probability has satisfy $\mathbb{P}(\omega|B) = 0$ if $\omega \notin B$. If $\omega \in B$, what should the new value $\mathbb{P}(\omega|B)$ be? Since the new information tells us only that B is true, but it does not give us information about the single ωs in B, the proportions given by $\mathbb{P}$ in B should be maintained by $\mathbb{P}(\cdot|B)$. Therefore we set $\mathbb{P}(\omega|B) = c\mathbb{P}(\omega)$ if $\omega \in B$, for some constant c. c which can be determined by requiring $\mathbb{P}(\cdot|B)$ has to satisfy (**??**). If $B = \{\omega_1, ..., \omega_k\}$, we obtain that

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

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

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

Considering an event $A = \{\omega_1, ..., \omega_s, \omega_{k+1}, ..., \omega_{k+m}\}$, where $s \leq k$ formula (0.4.2) leads to

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

conditional

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

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

that is, is the probability that associates to each $A \subset \Omega$ the value $P(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$.
As shown in Figure **??**, there is an easy interpretation of the conditional probability. Recall from ??? that one can interpret a probability as a way of measuring areas (in the discrete setting as a way to count which gives different weights to different events). If we observe B, our new evaluation of the area doesn't change in E and outside E is updated set zero. Since one of the conditions imposed to a probability is that the total area is 1, see Proposition **??**, to maintain this property we need to divide the new evaluation by E. In this sense we maintain the proportions of our evaluations inside E (since we do not have any information on which event caused E) and we set them to 0 outside, since we know that E happened. Alternative, one can think that we are redistributing the mass of E^c, that is $\mathbb{P}(E^c)$, to E, in such a way that the proportions do not change. One can think at a probability as a distribution of a mass, which in total is 1, over Ω. If one takes the distribution of mass given by $\mathbb{P}$ and sets the mass of B^c to be zero, one has a distribution of mass whose total is not 1. In order to obtain the distribution of mass given by the conditional probability $\mathbb{P}(\cdot|B)$, one has therefore to redistribute the mass it has removed, $\mathbb{P}(B^c)$, in such a way to not change the relative masses of the singletons in B. Thus $\mathbb{P}(\omega|B) = \mathbb{P}(\omega) + \mathbb{P}(\omega)\mathbb{P}(B^c)/\mathbb{P}(B)$. This way of thinking can be useful for the Exercise **??**. Note also that Example (0.5.1) and Exercise (**??**) show that probabilities can vary a lot if conditioned to the an event B. This is schematically shown in Figure (4)
Ex. 14 — We throw two dice and assume that each of numbers is equally probable [a] what is the probability conditioned on "The sum of the two dice

[a] *In other words, on Ω we assume that $\mathbb{P}$ is the uniform one. Why this is the case investigated in Subsection0.5.2*

Conditional˙Probability.png

Figure 3: We forget about B^c and we consider $\mathbb{P}$ restricted to subsets of $\mathbb{B}$. However, to obtain total mass 1 without favouring any of the $\omega \in B$, we multiply by $\frac{1}{\mathbb{P}(B)}$.

f:cond˙prob

Proportion˙conditional.png

fig:prop

Figure 4: $\mathbb{P}(E)$ is much greater that $\mathbb{P}(E|B)$.

is 7"?.

Ex. 15 — You toss a fair coin independently 4 times [1]. 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.

Answer (Ex. 15) — The sample space is $\Omega = \{0000, 0001, 0010, \ldots, 1111\}$. Let $B == \{x_1x_2x_3x_4\,,\, x_1 + x_2 + x_3 + x_4 = 1\} = \{1000, 0100, 0010, 0001\}$ be the event that only one heads result has shown up. The probability of the singleton $x_1x_2x_3x_4$, where $x_i \in \{0,1\}$ for $i = 1,2,3,4$, conditioned to B is given by

$$\mathbb{P}(x_1x_2x_3x_4|B) = 0$$

if $x_1 + x_2 + x_3 + x_4 = 1$ and is equal to

$$\mathbb{P}(x_1x_2x_3x_4|B) = \frac{1/2^4}{4/2^4} = 1/4,$$

which is to say that there is not favourite position in which the heads might appear.

Ex. 16 — Let $\Omega = \{\omega_1, ..., \omega_n\}$ be a sample space and let $\mathbb{P}$ be a probability assigned to it. Let $B \subset \Omega$ be an event, and assume $\mathbb{P}(B) > 0$.

1. Prove that $\mathbb{P}(\cdot|B)$ is a probability on the sample space B.
2. Assume that $\mathbb{P}$ is the uniform probability defined in (**??**) on Ω. Show that $\mathbb{P}(\cdot|B)$ is the uniform probability on the new sample space B.

Answer (Ex. 16) —

1. In order to show that $\mathbb{P}(\cdot|B)$ is a probability measure, we have to show that

   - $\mathbb{P}(\omega|B) \in [0,1]$
   - ((**??**) holds)$\sum_{\omega \in B} \mathbb{P}(\omega|B) = 1$
   - ((0.4.2) holds) For each $F \subset B$, $\mathbb{P}(F|B) = \sum_{\omega \in F} \mathbb{P}(\omega|B)$.

   The first condition is true since, if $\omega \in B$, $\mathbb{P}(\omega) \le \mathbb{P}(B)$, and therefore $\mathbb{P}(\omega|B) = \mathbb{P}(\omega)/\mathbb{P}(B) \le 1$. The second condition, which can be rewritten by Proposition (0.4.4) as

   $$\mathbb{P}(B|B) = 1$$

   is true since

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

[1] The same remark applies here: the probability is the uniform one and in Subsection 0.5.2 we will see why

The last condition is true since

$$\mathbb{P}(F|B) = \frac{\mathbb{P}(F \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(F)}{\mathbb{P}(B)}$$
$$\sum_{\omega \in F} \frac{\mathbb{P}(\omega)}{\mathbb{P}(B)} = \sum_{\omega \in F} \mathbb{P}(\omega|B).$$

2. As for the second question, since we have just proved that $\mathbb{P}$ is a probability measure on B, it suffices to note that every elementary event has the same probability (recall how we have derived formula (**??**): simply by assuming that the probability was equal on each elementary event). In fact $\forall \omega \in B$ $\mathbb{P}(\omega|B) = \mathbb{P}(\omega)/\mathbb{P}(B)$ does not depend on ω, since by assumption $\mathbb{P}$ does not depend on ω.

Ex. 17 — ([**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. 17) — 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

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

Ex. 18 — Consider an urn containg 2 blue marbles, 1 green marble and 1 red marble. Draw randomly (with uniform probability) one of them.

1. What is the probability of extracting a red marble if you know that the extracted marble is not green?
2. We do the following procedure. We draw a marble uniformly. If it is red we draw another marble independently, if not we finish the procedure. We are interested in the colour of the extracted marble: what is the probability of extracting a red marble?
3. Note that it is not necessary to know the probability of drawing in one draw the red ball, that is $\mathbb{P}(r)$. Note also that in the second exercises

gives you a procedure to draw from a bowl where you have removed the red marbles, even if you are drawing from a bowl that has them.

Answer (Ex. 18) — 1. The sample space is $\Omega = \{r, b, g\}$,$\mathbb{P}(r) = 1/4$, $\mathbb{P}(b) = 1/2$ and $\mathbb{P}(g) = 1/4$. Thus

$$\mathbb{P}(g^c|r^c) = \frac{\mathbb{P}(b)}{\mathbb{P}(b,g)} = \frac{2}{3}$$

2. Now the sample space is more complicated, and resembles the game of Exercise (11)

$$\Omega = \{r...rc_n\,,\, n \in \mathbb{N}\,, c_n \in \{b,g\}\} \cup \{rr...\}$$

The probability that we place in Ω is the following: The infinite sequence has probability 0: $\mathbb{P}\,(r...)$, while the finite sequences have probability

$$\mathbb{P}(rr...rc_n) = \mathbb{P}(rr...rc_n|\text{"first extraction r"})\mathbb{P}(r) = \mathbb{P}(r...c_{n-1})\mathbb{P}(r)$$
$$= ... = \mathbb{P}(r)^{n-1}\mathbb{P}(c_n)$$

(It will be maybe more clear once the geometric random variable has been obtained) and we take $\mathbb{P}(c)$ =(number of balls of colour c)/(total number of balls). The event I draw a blue ball is the event $\{b, rb, rrb, ...\}$ which has probability

$$\mathbb{P}(\text{"I draw a blue ball"}) = \mathbb{P}(b) + \mathbb{P}(b)\mathbb{P}(r) + \mathbb{P}(b)\mathbb{P}(r)^2 + \dots$$
$$= \mathbb{P}(b)\left(\sum_{i=0}^{+\infty}\mathbb{P}(r)^i\right) = \frac{\mathbb{P}(b)}{1-\mathbb{P}(r)} = \frac{\mathbb{P}(b)}{\mathbb{P}(b)+\mathbb{P}(g)} = 2/3$$

3. The above procedure can be used to draw on a bowl of. In a more realistic situation, maybe you want to interview families with children, but you don't have at disposal the "bowl", that is, the list with of the families of two children. This exercises tells that it suffices to contact each family untill a family with children is contacted. In more complex situations (i.e, in montecarlo simulations), a similar method is used to get to know the conditional probabilities

Example 0.21: 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

Judicial˙defence.png

fig:defence **Figure 5:** Defense argument: The proportion between the two areas is small

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 .

e:Judicial **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 some data that showed that only 1/10000 men beating their wives end up killing them, thus claiming that there where no sufficient clues to judge him guilty. 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. Here much more is known: the wife actually died.
Denoting the events

$$A = \text{"The husband beated the wife "}$$

$$B = \text{"The woman has been assassinated by his husband "}$$

$$C = \text{"The woman has been assassinated by a person different from her husband "}$$

The claim of the defence is depicted in Figure 5 We wish to compute $\mathbb{P}(B|A\cap(B\cup C))$, which, which is depicted in Figure 6.

$$\begin{aligned}\mathbb{P}(B|A\cap(B\cup C)) &= \frac{\mathbb{P}(B\cap A\cap(B\cup C)}{\mathbb{P}(A\cap(B\cup C))} = \frac{\mathbb{P}(B\cap A)}{\mathbb{P}((B\cup C)\cap A))} \\ &= \frac{\mathbb{P}(B\cap A)}{\mathbb{P}(B\cap A)+\mathbb{P}(C\cap A)} = \frac{1}{1+\frac{\mathbb{P}(C\cap A)}{\mathbb{P}(B\cap A)}}.\end{aligned} \tag{15}$$

e:proportions

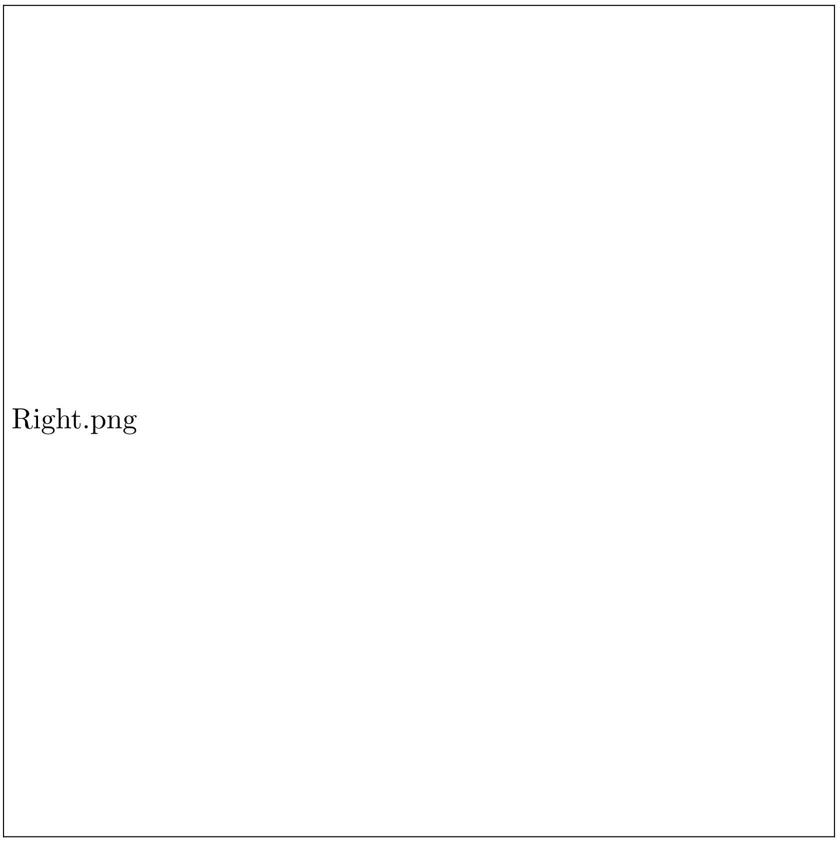

fig:right

Figure 6: What we need to compute: the proportion of the blue area with respect to the total area. Why this is what we need to compute is shown in (15)

Thus it is not $\mathbb{P}(B|A)$ that matters, but only its relative value $\mathbb{P}(B|A)/\mathbb{P}(C|A) = \mathbb{P}(C \cap A)/\mathbb{P}(B \cap A)$, as shown in Figure 6. We assume that C and A are independent, so that $\mathbb{P}(C|A) = \mathbb{P}(C)$ and (15) becomes

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

We estimate the $\mathbb{P}(C)$ by the probability that a woman gets killed: $\mathbb{P}(C) \leq \mathbb{P}(B \cup C)$ which they estimated by $\mathbb{P}(B \cup C) = 1/100000$. Therefore

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

Plugging in this value, one obtains that

$$\mathbb{P}(B|A \cap (B \cup C)) \geq 10/11$$

At the end, the man has been judged guilty.

0.5.2 Independence of Events

ss:independence

In view of the definition of condiitonal probability in Definition **??** the following definition is natural to say that the events A and B are independent if $\mathbb{P}(A) = \mathbb{P}(A|B)$. In a more symmetric form Two events $A, B \subset \Omega$ are *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{e:ind}$$

It is important to notice that the word "independent" can be missleading, since it could be wrongly associated to the absence of a causal relation between A and B. Example **??** shows that there are events not related by any causal relation that are dependent and that, conversely, there are independent events that are causal. The events A and B are independent if from the knowledge of one of them you cannot infer anything on the evaluation of the probability of the other one: the probability remains the same. For this reason the independence property is sometimes stochastic independence. Another way to see that the notions of causality and independence are unrelated is by noting that the notion of independence depends from the probability $\mathbb{P}$, while the notion of causality does not.

Note that the notion of independence is unrelated with the notion of causality. The notion of causality is a notion that takes into account only the relation between events, while the notion of independence takes into account the notion of probability. Independence means that, given $\mathbb{P}$, that is, having an evaluation of the probabilities of the events, we can infer no information from the realization of B that causes to change the evaluation of the probability of A. Thus is a probability dependent hypothesis and, for instance, Exercise (**??**) shows that coin tosses are not independent if we don't know exactly the parameter of the coin, since we can infer some information useful to determine the coin parameter by tossing the coin, which, in turn, will change the probability of the next toss. It is the same situation of (26), which makes clear the striking fact that the number of shark attacks and the number of ice creams sold are numbers with a strong dependence, even if, of course, not related by any causal relation.

Example 0.22: Two independent uniform samplings

nple˙fundamental

Consider a bowl where 3 marbles, 1 red, 1 blue and 1 green, are placed and draw two marbles with replacement. The state space of the system is

$$\Omega = \{(r,r), (r,b), ..., (g,g)\},$$

We want to prove that the following are equivalent:

- The probability measure on Ω is the uniform one.

- In each single extraction you can draw each marble with the same probability and the extractions are independent

To see what the second condition means precisely, let's assume that $\mathbb{P}$ is uniform and define the events

$$F_c = \text{"The colour of the first marble is c"} = \{(c,r),(c,b),(c,g)\},$$

for $c = r, b, g$, and

$$S_c = \text{"The colour of the second marble is c"} = \{(r,c),(b,c),(g,c)\},$$

for $c = r, g, b$. By saying that each extraction is a "random" extraction (that is, is uniform) we mean that

$$\begin{aligned} \mathbb{P}(F_c) &= 1/3 \\ \mathbb{P}(S_c) &= 1/3 \end{aligned} \tag{16}$$ 1

for each $c = b, g, r$, which is easily seen to hold true. By the fact that the extractions are independent we mean that for each c and c' in $\{r, b, g\}$

$$\mathbb{P}(F_c \cap S_{c'}) = \mathbb{P}(F_c)\mathbb{P}(S_{c'}). \tag{17}$$ 2

The event on the left hand side is the singleton $\{(c, c')\}$, so that its probability is $1/9$. The events on the right hand side have probability 1/3, so that the above equality holds. Let's assume that now (16) and (17) hold. Then

$$\begin{aligned} \mathbb{P}(c, c') &= \mathbb{P}(F_c \cap S_{c'}) \\ &= \mathbb{P}(F_c)\mathbb{P}(S_{c'}) = 1/3 \times 1/3 = 1/9, \end{aligned}$$

so that $\mathbb{P}$ is uniform. This example will be generalised in Section **??** to the case where the two extraction procedure can be performed jointly in an arbitrary way. Examples (0.23) and Exercise (**??**) and (**??**) will show examples in which you perform two dependent extractions and how this can lead to surprising results.

Example 0.23: Mounty Hall Problem

MountyHall

In a game show called the Mounty Hall, the participants had to choose between three doors. Behind 1 door there was a car while behind the other two there was a goat. The participant had to guess which door contained the car. After the first choice, the host disclosed a door which had behind a goat and then asked the participant whether or not he would like to change the choice of the door. The result is at a first sight surprising. Let's call Strategy 1 the strategy in which we don't change the door, and Strategy 2 the strategy in

which we decide to change the door. Then $\mathbb{P}(\text{"win"}) = 1/3$ if we play following Strategy 1, and $\mathbb{P}(\text{"win"}) = 2/3$ if we play by Strategy 2
To prove the above fact, assume that the participant choses the door 1 and that he plays following Strategy 2. The result is clear if one observes that the participant wins if and only if he chooses a goat at the beginning. **Exercise:** Consider the Mounty Hall problem with n doors and k cars. What is the probability of winning under Strategy 2?

Solution The sample space should encode the doors, the thing that is behind the door, the first door chosen by the participant and its second choice. However, we can give a reduced description since we are interested only on the choices the participant makes. The sample space can be taken to be $\Omega = \{CC, GG, GC, CG\}$ where the first digit denotes what is there behind the first door and similarly for the second. Let's denote the event "The first choice is a car" by F, and the event "The second choice is a car" = "win" by S. Therefore

$$\mathbb{P}(S) = \mathbb{P}(S|F)\mathbb{P}(F) + \mathbb{P}(S|F^c)\mathbb{P}(F^c),$$

but, given F, after the goat is revealed, the participant chooses between n-2 doors with k-1 cars. Given F^c, then, after the host reveals the goat, has to choose between $n-2$ doors and k cars. Therefore

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}(S|F)\mathbb{P}(F) + \mathbb{P}(S|F^c)\mathbb{P}(F^c)\\ &= \frac{k-1}{n-2}\frac{k}{n} + \frac{k}{n-2}\frac{n-k}{n}\end{aligned}$$

So, it is not diffiucult to understand the Mounty Hall problem: what happens is that the different extractions, the first door and the second one are dependent.

The next "paradox" has interest on its own and shows that it is important to clearly state the conditioning event

Example 0.24: Boy or Girl Paradox

Consider a family with two children and consider their genders. The state space is $\Omega = \{MM, MF, FM, FF\}$ and consider the uniform probability on it. [b]

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

2. Assume that we get to know that one of the children 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 (see Example 0.22). The event A="The first child is male" is given by $\{(M, M), (M, F)\}$. A basic computation shows that

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

As for the second question, maybe counterintuitively, the answer is $1/3$. In this case the conditioning event A is $\{(M, M), (M, F), (F, M)\}$, and

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

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

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

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

Even if the second point of the example might come as a surprise, if one interprets the probability with a wrong frequentist definition, if one were to obtain the list of all the families with two children such that one is a male, then approximately 1/3 of the hamilies in those lists will have 2 male children. The procedure 3 would have the following equivalent in the frequentist approximation. From the list of all the families with two children, you toss an unfair coin to decide which children to take into account and to check. So there is a

Ex. 19 — 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. 20 — In the setting of Example 0.22 you draw two marbles without replacement.

1. What are the probabilities of each single extraction?
2. Are extractions independent?

ex:3indep **Ex. 21 —** $\Omega = \{1,2,3,4\}$ 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.

Ex. 22 — 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?

Ex. 23 — 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?
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 0.4.4 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 $1/3$.

You can interpret this example as the fact that is important to know how to get the information and it is important to specify the conditioning event rigorously.

Three events A,B and C are independent if they are pairwise independent

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B), \\
\mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C), \\
\mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C),
\end{aligned}$$

and

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

An example three events that are pairwise independent but not independent is given in Exercise 21.

0.5.3 Sampling Interpretation of Conditional Probability

ss:r˙cond

We now come back to the example of Subsection **??**, where $\Omega = \{\omega_1, \ldots, \omega_6\}$ and we assume that we observe $E = \{\omega_1, \omega_4, \omega_5, \omega_6\}$. We can define the conditional probability by setting $p_2 = p_3 = 0$ and normalising the thuis obtained vector so that the sum is 1.
Looking at Example **??**, if we know how to sample from $\mathbb{P}$, and indeed we were able to do it in Subsection **??**, we can sample from the $\mathbb{P}()$

0.5.4 Total probability Formula

When speaking about events many times one reasons by cases in the following sense. Assume you don't know whwther to go out this night, then, in order to think at the event "I will go out this night" one thinks at this event first in the case it rains, then in the case it is good weather. Since the probability is simply a quantitative version of the logic we use when speaking about events, we might hope to reproduce the same reasoning at the quantitative level of probabilities. This is indeed true and it is done in the Total Probability Formula. [Total Probability Formula for two conditioning events] Let $B \subset \Omega$ be an event such that $\mathbb{P}(B) \in (0,1)$. Then

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \tag{18}$$

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 (12) to develope the above expression into

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

which is what we wanted to prove. □

Ex. 24 — We toss a fair 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?

Answer (Ex. 24) — 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 (18) 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: 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". Another partition would be $\{C_0, C_1, C_2, C_3, C_4\}$ where C_i = "There have been i heads".

- 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 there will be. Thus Ω is the one given by (**??**) and it can be partitioned into $\Omega = \cup_n \Omega_n$, where Ω_n=" There have been n tosses "$= \{x_1...x_n\ , x_i \in \{0, 1\}\}$.

In other words a partition is a family of mutually disjoint events (or hypotheses) that together they include all the possible eventualities, conditions intuitive and necessary to make reasonings of the form " If B_1, then and you can reason separately on each element of the partition, both at the level of sets and at the level of probabilities thanks to the next Note that if a partition consists of two events B_1 and B_2, then $B_2 = (B_1)^c$.
[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 similar to the previous and we remark that, up to some basic knowledge of series, one could take a partition to be an infinite family of the form $\{B_i\}_{i \in \mathbb{N}}$, and the total probability formula would still be true.
Ex. 25 — Consider an infection that is contagious and which is suffers by 1 person out of 1000. There is a test used to determine whether a person has the infection and the test has a 5% of possibility of giving a false positive: You don't have the infection but the test is positive, while the probability of having a false negative is zero.

1. If you are tested positive, what is the probability of actually having the infection?
2. You have some symptoms and the doctors estimate around 50% of probability that you have the infection. What is the probability that you have the infection provided that the test gives a positive result?

ex:indep

Ex. 26 — We have two urns, one containing 3 red balls, 1 blue ball, and one white ball, the other one containing 3 blue balls, 1 red ball and 1 white ball. We toss a fair coin to choose one of the two urns.

1. We draw a ball from the chosen urn. What is the probability that you extract a red ball? (Hint: Use the total probability formula)
2. Now we extract a ball, reinsert it in the urn, and draw again another ball from the same urn. Are the event "The first ball extracted is red" and the event "The second ball extracted is red" independent?

0.5.5 Odds and Bayes Formula

Part II

STATISTICS