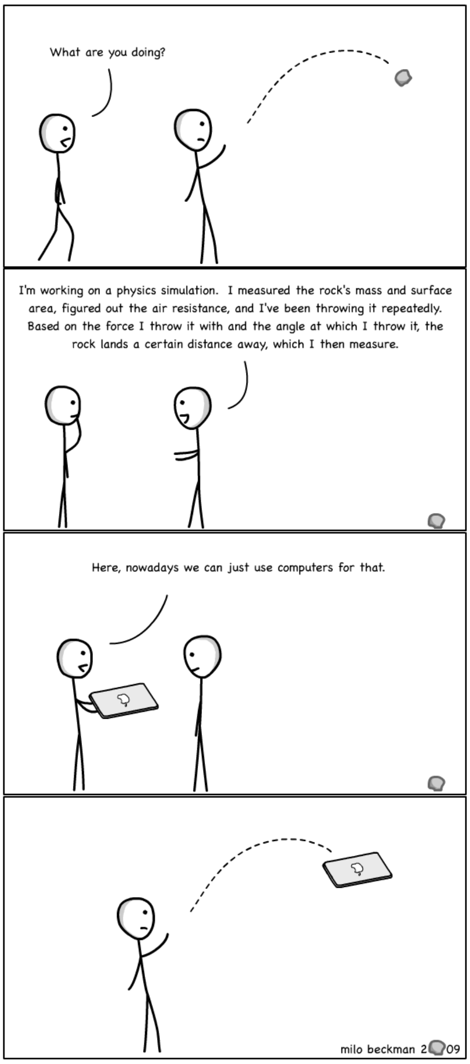
Seconde européenne
Exercices de mathématiques

Chapitre 6
Simulation et échantillonage
Simulating and sampling



Yucca Muffin, by Milo Beckman

At the end of this chapter, you should be able to :
- compute the margin of error and fluctuation interval at 95% confidence for a known probability ;
- use a fluctuation interval to accept or reject an assumption ;
- compute the confidence interval at 95% for a sample ;
- use a confidence interval to estimate a probability ;
- use the calculator to simulate a random experiment.

Aymar de Saint-Seine et Mickaël Védrine
Année scolaire 2010/2011

Margin of error

6.1 Estimation and sampling : an election

In 2008, the American electors had to choose between the Republican John McCain and the Democrat Barack Obama. On Election day, Obama won with 53% of the votes. Of course, this was not known to the candidates before the election. Surveys were organised by both parties to *estimate* the proportion of electors who wanted to vote for each candidate. As it's impossible to gather the opinions of all the electors, surveys are carried over small parts of the population, called *samples*. We will consider that samples are built randomly.

Part A – Point estimate

1. Over a sample of 900 electors, 497 declared that they wanted to vote for Obama. Compute the percentage of potential Obama electors in this sample.
2. The percentage computed for the sample is a *point estimate* of the actual percentage in the whole population. Compute the difference between this point estimate and the true value (known only after the election). What do you think of this estimation ?
3. Ten other surveys were organized over the same period. The size of each sample and the number of potential Obama electors are given in the table below.

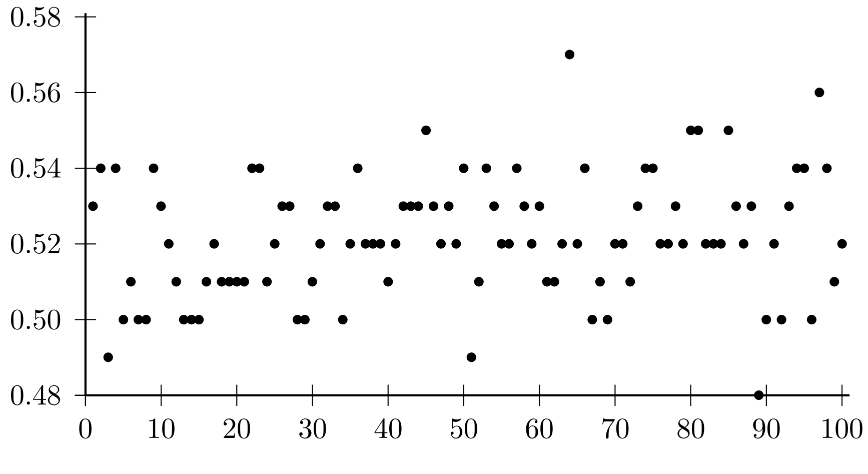
Survey	1	2	3	4	5	6	7	8	9	10
Size	895	873	900	885	899	842	878	900	897	892
Obama electors	462	493	501	437	467	447	468	495	488	478

   a. Compute the point estimate for each survey. Round the answers to 2DP.
   b. How many surveys gave a point estimate equal to the true percentage ?
   c. If McCain had only known about the 4th survey, what could he have deduced ?
   d. What do you think of the point estimate method ?

Part B – Margin of error

Estimation by point estimate is not a very good method. The chances that the sample will yield the true value in the whole population are very small. Furthermore, There may be important differences between the point estimates obtained from different samples. This phenomenon is known as *sampling fluctuation.*

To illustrate this, one hundred surveys were simulated with a computer, each one over a population of 900 people. (To do so, the percentage of Obama electors in the whole population, $p = 0.53$, was used). The scatterplot below shows the percentage of potential Obama electors in each survey.



1. On the scatterplot, show the percentage p of Obama electors in the whole population with a horizontal red line. How many simulated surveys gave that exact value ?

2. **a.** The value $m = \frac{1}{\sqrt{n}}$, where n is the size of a sample, is called the *margin of error at 95% confidence* for that sample. Compute this value to 3DP.
   b. On the graph, show the values $p - m$ and $p + m$ with two horizontal blue lines.
   c. How many surveys gave a point estimate included in the interval $[p-m; p+m]$, called *fluctuation interval at* 95% *confidence* ?
   d. Is the answer to the previous question consistent with the name of the interval ?
3. Compute the margin of error and fluctuation interval at 95% confidence for samples of size $n = 25$, then $n = 100$ and $n = 500$. What do you notice about the margin of error when the size of the sample increases ?

6.2 The French lottery and odd numbers

The principles of the French National Lottery (Loto) are fairly simple. Each player picks six numbers (plus one, that we won't consider in this exercise) between 1 and 49. On lottery day, 49 balls with numbers from 1 to 49 are randomly drawn from a machine. The balls are not put back in the machine, so the same number cannot appear twice in a drawing. The order in which the balls are drawn is irrelevant.
Among the numbers from 1 to 49, there are 25 odd numbers and 24 even numbers. So it seems that the French lottery favours odd numbers. This is what we will study in this exercise.

Part A – Drawing a single number

In this part, we consider the random experiment that consists in drawing a single ball from the 49 in the machine.

1. What is the probability of the drawn number being odd ? Give the result as an irreducible fraction and as an approximate value to 2DP.
2. Fifty samples, each made of $n = 100$ independant drawings of a ball were simulated with a computer. For each sample, the proportion of odd numbers was computed. The results of these fifty samples of size 100 are given below.

0.44	0.52	0.50	0.44	0.51	0.41	0.44	0.40	0.57	0.50
0.51	0.43	0.59	0.46	0.55	0.35	0.55	0.43	0.53	0.53
0.45	0.42	0.47	0.48	0.50	0.45	0.48	0.47	0.46	0.57
0.52	0.55	0.53	0.46	0.45	0.44	0.45	0.48	0.51	0.46
0.55	0.48	0.43	0.51	0.49	0.38	0.52	0.40	0.50	0.46

   a. How many samples showed a proportion equal to the theoretical value to 2DP ?
   b. Compute the fluctuation interval at 95% confidence.
   c. How many samples showed a proportion inside the margin of error ?
   d. Can you find a margin of error at 99% confidence ?

Part B – Drawing six numbers

In this second part, we consider the random experiment that consists of drawing successively six balls, without putting them back in the machine. It can be proven that in each drawing of six numbers, there is an average of 3.0612 odd numbers, so a proportion $q = \frac{3.0612}{6} \approx 0.51$, or approximately 51%.
Fifty samples, each made of $n = 100$ independant drawings of six succesive balls were simulated with a computer. For each sample, the proportion of odd numbers was computed. The results of these fifty samples of size 100 are given below.

0.527	0.475	0.500	0.522	0.558	0.518	0.510	0.518	0.550	0.607
0.515	0.468	0.517	0.485	0.498	0.473	0.505	0.507	0.492	0.498
0.508	0.408	0.563	0.612	0.542	0.497	0.508	0.498	0.500	0.535
0.498	0.508	0.525	0.478	0.517	0.528	0.492	0.487	0.535	0.523
0.512	0.543	0.522	0.482	0.530	0.478	0.508	0.532	0.528	0.527

1. Compute the fluctuation interval at 95% confidence.
2. How many samples showed a proportion inside the interval ?

Part C – Probabilities on the number of odd numbers

The table below shows the probabilities of drawing k odd numbers among the six, for k from 0 to 6. Values have been rounded to 3DP.

Odd numbers	0	1	2	3	4	5	6
Probability	0.010	0.076	0.228	0.333	0.250	0.091	0.013

For each of the following sentences, say if it's true or false. Justify each answer with a computation or an explanation.

1. There are more chances to draw 4 odd numbers or more than 2 odd numbers or less.
2. There are as many chances to draw at least 3 odd numbers than at least 3 even numbers.
3. There are more than 90% chances to draw at least 2 odd numbers.
4. There are as many chances to draw exactly 3 odd numbers than exactly 3 even numbers.
5. There are 50% chances to draw as many odd numbers and even numbers.
6. There are more chances to draw no even number than to draw no odd number.

6.3 Using margins of error to make decisions

Part A – Accepting or rejecting an assumption

It is know that in the French population, 26% are allergic to pollen. In one particular sample of 400 people, 120 suffer from that allergy.

1. Compute the fluctuation interval at 95% confidence.
2. What is the frequency of allergic individuals in this sample ?
3. Would you consider this sample representative of the French population ?

Part B – Parity in French Region councils

After the 2004 regional elections in France, the repartition between women and men in four regional councils was as follows. We consider that these councils are random samples of the local politician population in France.

	Men	Women	Total
Burgundy	32	25	57
Brittany	38	47	85
Rhône-Alpes	81	76	57
Île-de-France	103	106	209

1. Supposing that parity between men and women is real in a regional council, what should be the percentage of women in that council ?
2. Find out the fluctuation interval at 95% confidence for the proportion of women in each council.
3. What do you think of the parity between men and women in the local politician population in France.

Part C – A car factory

In a car factory, a control is done for flaws of the type “grainy spots on the hood”. Normally, 20% of the vehicles present this kind of flaws. While controlling a random sample of 50 vehicles, it is seen that 13 vehicles have it. Should it be a matter of concern ?

Part D – Rodrigo Partida’s case

In 1970, the Mexican-American Rodrigo Partida was sentenced to eight years of prison. He appealed to the judgment contending that he was denied due process and equal protection of law because the grand jury of Hidalgo County, Texas, which indicted him, was unconstitutionally underrepresented by Mexican-Americans. He introduced evidence that in 1970, the total population of Hidalgo County was 181,535 persons of which 143,611, or approximately 79.2% were persons of Spanish language or Spanish surname. Next, he presented evidence showing the composition of the grand jury lists over a period of ten years prior to and including the term of court in which the indictment against him was returned. Of the 870 persons selected for grand jury duty, only 39.0% were Mexican-Americans. If you were a judge in the court of appeals, how would you react to these allegations ?

Confidence interval

6.4 In this exercise, we look again at the US 2008 election. We will introduce a better method of estimation, based on the concept of margin of error. Instead of a simple point estimate, we will build for each sample a *confidence interval* whose diameter depends on the margin of error we allow.
We still note p the percentage of Obama electors in the whole population (so $p = 0.53$). Now, consider a sample fo size n yielding a point estimate f of p. We’ve seen in the previous part that the margin of error at 95% confidence is $m = \frac{1}{\sqrt{n}}$. Indeed, the probability of the point estimate f being in the interval $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ is approximately equal to 95%.

1. Translate the fact that f belongs to that interval with two inequalities.
2. Prove that the fact that f belongs to that interval is equivalent to the fact that p belongs to the interval $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.

The interval $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$ is called a 95% *confidence interval*. Intuitively, this means that, knowing f and not p, we have a 5% risk of being wrong if we consider that p is in the interval. But, as p is fixed, it’s not really correct to talk about probability. Once the confidenc interval is determined, p is either in it or not !

1. Find the 95% confidence intervals for the surveys of exercice 1 part A.
2. How many surveys gave a confidence interval including the real value ?

6.5 The referendum on the European constitution

1. The French referendum on the Treaty establishing a Constitution for Europe was held on 29 May 2005 to decide whether France should ratify the proposed Constitution of the European Union. The question put to voters was : “Do you approve the bill authorising the ratification of the treaty establishing a Constitution for Europe ?”
   Below are given the results of some surveys carried out before the referendum.

Dates	Institute	Size	Proportion of « no »
18 and 19 March 2005	Ipsos	860	0.52
25 and 26 March 2005	Ipsos	944	0.54
1er and 2 April 2005	Ipsos	947	0.52
16 and 17 March 2005	CSA	802	0.51
23 March 2005	CSA	856	0.55
1 and 2 April 2005	Louis Harris	1004	0.54
31 March and 1 April 2005	IFOP	868	0.55
24 March 2005	IFOP	817	0.53

   a. Find the 95% confidence interval for each survey.

   b. The result was a victory for the "No" campaign, with 54.67%. A commentator then said that not many surveys had anticipated such a decisive result. What do you think of that opinion ?

2. The United Kingdom referendum was expected to take place in 2006. Following the rejection of the Constitution by voters in France in May 2005 and in the Netherlands in June 2005, the referendum was postponed indefinitely.
ICM research asked 1,000 voters in the third week of May 2005 “If there were a referendum tomorrow, would you vote for Britain to sign up to the European Constitution or not ?” : 57% said no. Find the 95% confidence interval for this survey. If you were a politician, what would you deduce from this ?

6.6 M&Ms

Josh Madison is a 30-something New Yorker who runs a personal website on the internet. In one of his articles, he states the following crucial problem.

I love M&M's. I'm partial to the plain Milk Chocolate variety, but I've been known to have a Peanut from time to time in order to remind myself why I don't like them that much. Often, while eating a pack, I'll wonder how they're made and how the colors are distributed.
After wondering about it a little more, I checked out M&M's web site. According to it, each package of Milk Chocolate M&M's should contain 30% blue, 20% brown, 10% green, 10% orange, 10% red, and 20% yellow M&M's. I checked the next few packages of M&M's that I ate and found that their percentages were not even close to the stated distribution. In my mind, this sort of confirmed my thoughts about how they produce M&M's : When they make M&M's, in any production run, they produce the stated percentage of each color and then just fill the packs off a conveyor line or some other weight based method. This would mean that any single package could be way off from the stated percentage ; but analyze the counts over a large number of packages, and they should converge towards the stated percentages.

Today, we will study this essential problem on a few bags of M&Ms. First, we will consider each bag as a sample.

1. Open your bag of M&Ms and count the number of candies of each color and the total number of candies. Give the results in an absolute frequency table. As soon as you have the results, give them to the teacher who will gather the data for the whole class.

2. **a.** Compute the confidence interval at 95% for each color of your sample.

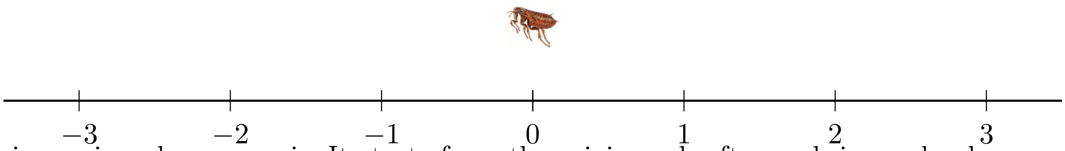
   b. Consider the blue M&Ms. For how many bags was the expected percentage in the confidence interval ? What would you conclude from this result ?

   c. Answer the previous question for the other colors.

3. To get a larger sample, we will now use the total numbers of candies of each color in all the bags in the class. Compute the confidence interval with that sample and compare the experimental results with the expected values. What would you conclude from that ?

Simulations

6.7 Random walks on an axis



A flea is moving along an axis. It starts from the origin and, after each jump, lands one unit to the right or one unit to the left, randomly and with the same probability.
A sequence of jumps is called a *walk*. For example, if the flea is always jumping to the right, the walk will be noted RRRR. If it alternates between right and left, the walk will be noted RLRL.

Part A – Simulations of 4-jumps walks

The “Random” or “Alea” function on your calculator delivers a random decimal number between 0 and 1.

1. Devise a method to simulate a 4-jumps walk using the “Random” function.
2. Simulate 25 walks and note the final position of the flea at the end of each walk.
3. What are the possible final positions on the axis ? Explain why some are impossible.
4. Count the number of walks for each final position and show the counts in a table.
5. Add a row to the previous table with the absolute frequencies for the whole class.
6. Compute the relative frequencies for the whole class.
7. Compute the average final position of the flea at the end of a 4-jumps walk.

Part B – An algorithm

A random walk can be described by the algorithm shown on the right-hand side, where the alea function delivers a random number in the interval $[0, 1[$. Parts of the algorithm have been omitted on purpose.

```
begin
    0 → x ;
    1 → i ;
    while i ⩽ 4 do
        if alea < 0.5 then
            ......... → x ;
        else
            ......... → x ;
        end if
        i + 1 → i;
    end while
    Output : x
end
```

1. Explain the functions of the integers x and i in this algorithm.
2. Fill the two incomplete lines.
3. Here are the results of applying the algorithm once. What is the final position of the flea at the end of this walk ?

i		1	2	3	4
alea		0.37	0.01	0.93	0.11
x	0	1	2	1	2

4. Apply the algorithm to create five new walks, using the random function of your calculator and displaying all the steps of the algorithm like in the example of the previous question.
5. How would you change the algorithm to simulate a 30-jumps walk ?

Part C – Probabilistic study

In this part, we will use probabilities to study the situation and compare the theoretical results to the frequencies we found in part A.

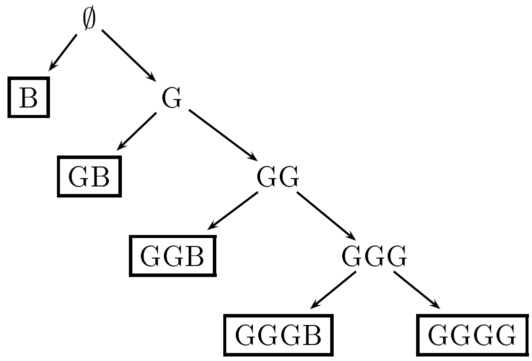
1. Draw a tree to show all the possible 4-jumps walks. At the end of each branch, write the final position of the flea.
2. Use the tree to compute the probability of each final position and give the results in a probability table.
3. Compute the margin of error at 95% confidence for your sample of 25 random walks.
4. For each probability, count in the class how many samples of 25 walks gave a frequency within the margin of error.

6.8 A birth policy

A government has decided to impose a strict birth policy. Births in a family must stop as soon as a boy is born or after the birth of the fourth child.

We consider in this exercise that the probabilities of giving birth to a girl or a boy are equal and that each birth is independant from the previous births in the same family.

This birth policy can be represented as a tree, where the possible families are boxed.



Part A – Simulation and statistical view

1. Do you think that this policy will favour boys or girls ? No justification is expected.
2. Devise a method to simulate the composition of a family with the calculator.
3. Simulate and write down the composition of 100 families. Count the number of children per family and show the results in a table with absolute and relative frequencies.
4. Compute the arithmetic mean m_4 and the median d_4 for the number of children per family in your sample of 100 families.
5. Compute the arithmetic mean M_4 and the median D_4 for all the families in the class.

Part B – An algorithm

This process can be described as an algorithm. The output is then a list of digits, with 0 representing a girl and 1 representing a boy.

1. Explain the functions of the whole numbers x and i in this algorithm.
2. Explain the condition "$x \neq 1$ and $i \leqslant 4$". Does it ensure that the algorithm will always stop ?
3. What is the function of the list L in this algorithm ?
4. Explain the notation $L(i)$.

```
begin
    Clear list L ; 0 → x ; 0 → i ;
    while x ≠ 1 and i ⩽ 4 do
        i + 1 → i;
        if alea < 0.5 then
            0 → x;
        else
            1 → x;
        end if
        x → L(i);
    end while
    Output : L
end
```

5. Here are the results of applying the algorithm once. Apply the algorithm to get 5 families, displaying all the steps of the algorithm like in the example.

i		1	2	3
alea		0.37	0.01	0.93
x	0	0	0	1
L	()	(0)	(0, 0)	(0, 0, 1)

Part C – Probabilistic study

1. Copy the tree at the beginning of the exercise and add the probabilities.
2. Compute the probability of each type of family.
3. Show in a table the possible numbers of children and their probabilities. Are these probabilities consistent with the frequencies found at the end of part A ?
4. Use the table to compute the expected value for the number of children in a family.

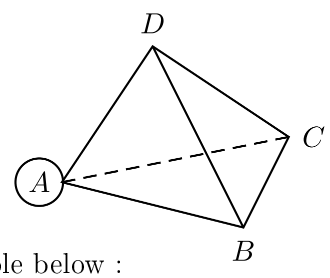
Part D – The percentage of girls

The aim of this part is to study the percentage of girls g induced by this birth policy, and therefore answer the first question of part A. To do so, we will first use the simulations of part A, and then the probabilities of part C.

1. Count the number and percentage of girls in your 100 simulated families.
2. From the various values found in the class, what could be the theoretical value of this percentage ?
3. Compute the 95% confidence interval for the percentage of girls in your sample.
4. Of all the confidence intervals in the class, how many include the possible value we found in question 2 ? Does it validate or invalidate this hypothesis ?
5. Use part C to find the true value of the percentage g.

6.9 Random walks on a tetrahedron

An ant is walking on the edges of a tetrahedron $ABCD$, starting from vertex A. When it gets to a vertex, it chooses randomly the next edge it will walk on. The aim of this exercise is to study the time it will take for the ant to go back to vertex A, assuming that it walks along one edge in exactly 1 minute.



A walk will be noted as a sucession of vertices, as in the example below :

$$A \to D \to B \to C \to A.$$

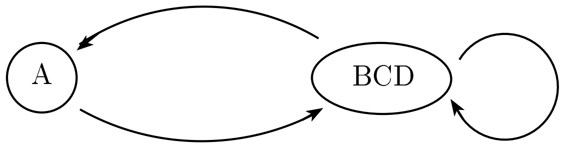
A walk will always start from A and stop as soon as the ant comes back to A.

Part A – Simulations

1. Devise a method to simulate a random walk.
2. Simulate 25 random walks and count the duration of each one. Gather the data in a table with the absolute frequency of each duration (from 1 to 20 minutes).
3. Explain the value in the column for 1 minute.
4. Is the duration necessarily less than 20 minutes ?
5. Find out the minimum, maximum, range, mean and median of this data.
6. Carry out the previous computations for all the simulated walks in the class.

Part B – Probabilistic study

In this part, we will gather the vertices B, C, D, for which the walk doesn't end in a single outcome noted BCD. Therefore, from vertex A, the only possibility is to go to BCD, while from BCD it's possible to go to A or stay in BCD.



1. Illustrate this new way of seeing the problem with a more simple graph.
2. Starting from BCD, compute the probabilities to go to A and to stay in BCD. Use these results to put the right probabilities on the arrows of the previous graph.
3. Show in a probability tree the first four steps of this process.
4. Compute the probabilities of a 2-minutes, a 3-minutes and a 4-minutes walk.
5. Without adding a level to the tree, conjecture a value for the probability of a 5-minutes walk. Deduce the probability of a walk lasting 5 minutes or less.

Part C – Estimation of a percentage

In this part, we will find an estimation of the percentage p of walks lasting 5 minutes or less.

1. Compute the percentage of walks lasting 5 minutes or less in your 25 simulated walks.
2. Compute the 95% confidence interval for this percentage in your sample.
3. Does this interval validate the value conjectured in part B ?
4. Compute the 95% confidence interval for this percentage in the sample made of all the simulated walks in the class.
5. Does this interval validate the value conjectured in part B ?

Homework

6.10 Galileo Galilei (15 February 1564 - 8 January 1642) was an Italian physicist, mathematician, astronomer and philosopher, and withouth any doubt one of the greatest minds of all times. He was born, the same year as Shakespeare, in the Italian town of Pisa. He was sent to the University of Pisa to study medicine, but in 1589 became professor of mathematics (at the age of 25) through the favours of Ferdinando dei Medici, the Grand Duke of Tuscany. During this period, Galileo was "ordered" by the Grand Duke of Tuscany to explain a paradox arising in the experiment of tossing three dice.

> *Why, although there were an equal number of 6 partitions of the numbers 9 and 10, did experience state that the chance of throwing a total 9 with three fair dice was less than that of throwing a total of 10 ?*

Part A – Simulating 1000 throws

Use a spreadsheet (OpenOffice Calc or Microsoft Excel, for example) to simulate 1000 throws of a fair die, and count the frequency of each result. To do so, you may use the following functions :

`=ALEA.ENTRE.BORNES(1;N)` **:** Produces a random integer between 1 and N.

`=SOMME(A1:A7)` **:** Computes the sum of the numbers in cells A1 to A7, including all cells in between.

`=NB.SI(A1:A7;k)` **:** Counts how many times the value k occurs in the cells A1 to A7.

Give the results of your 1000 simulations in an absolute frequency table. Does it confirm the Duke's statement ?

Part B – Galileo's solution

Below is an extract from Galileo's answer to the Great Duke of Tuscany, translated by E. H. Thorne.

> *The fact that in a dice-game certain numbers are more advantageous than others has a very obvious reason, i.e. that some are more easily and more frequently made than others, which depends on their being able to be made up with more variety of numbers. Thus a 3 and an 18, which are throws which can only be made in one way with 3 numbers (that is, the latter with 6.6.6 and the former with 1.1.1, and in no other way), are more difficult to make than, e.g. 6 or 7, which can be made up in several ways, that is, a 6 with 1.2.3 and with 2.2.2 and with 1.1.4, and a 7 with 1.1.5, 1.2.4, 1.3.3, and 2.2.3. Nevertheless, although 9 and 12 can be made up in as many ways as 10 and 11, and therefore they should be considered as being of equal utility to these, yet it is known that long observation has made dice-players consider 10 and 11 to be more advantageous than 9 and 12. And it is clear that 9 and 10 can be made up by an equal diversity of numbers (and this is also true of 12 and 11). [...]*
>
> *Three special points must be noted for a clear understanding of what follows. The first is that that sum of the points of 3 dice, which is composed of 3 equal numbers, can only be produced by one single throw of the dice : and thus a 3 can only be produced by the three ace-faces, and a 6, if it is to be made up of 3 twos, can only be made by a single throw. Secondly : the sum which is made up of 3 numbers, of which two are the same and the third different, can be produced by three throws : as e.g., a 4 which is made up of a 2 and of two aces, can be produced by three different throws ; that is, when the first die shows 2 and the second and third show the ace, or the second die a 2 and the first and third the ace ; or the third a 2 and the first and second the ace. And so e.g., an 8, when it is made up of 3.3.2, can be produced also in three ways : i.e. when the first die shows 2 and the others 3 each, or when the second die shows 2 and the first and third 3, or finally when the third shows 2 and the first and second 3. Thirdly the sum of points which is made up of three different numbers can be produced in six ways. As for example, an 8 which is made up of 1.3.4. can be made with six different throws : first, when the first die shows 1, the second 3 and the third 4 ; second, when the first die still shows 1, but the second 4 and the third 3 ; third, when the second die shows 1, and the first 3 and the third 4 ; fourth, when the second still shows 1, and the first 4 and the third 3 ; fifth, when the third die shows 1, the first 3, and the second 4 ; sixth, when the third shows 1, the first 4 and the second 3. Therefore, we have so far declared these three fundamental points ; first, that the triples, that is the sum of three-dice throws, which are made up of three equal numbers, can only be produced in one way ; second, that the triples which are made up of two equal numbers and the third different, are produced in three ways ; third, that those triples which are made up of three different numbers are produced in six ways.*

1. In the first paragraph, Galileo explains that there is only one way to make a 3, but several to make a 6.

a. Find in the text how many ways there are to make a 6.

b. In the same manner, find out all the ways to make the numbers 9 and 10.

c. What statement in Galileo's text is then proved ?

2. Explain the second paragraph with the example of the different ways to make a 6, and deduce how many different throws actually make this value.

3. Compute in the same way the number of different throws that make a 9 and those that make a 10.

4. Write the conclusion of Galileo's paper, explaining to the Great Duke of Tuscany the solution to this problem. (Minimum 50 words, maximum 150 words.)

Part C – The Passe-Dix game

Passe-dix, also called passage in English, is a game of chance using dice. It is played with three dice. There is always a banker, and the number of players is unlimited. To win, a player must throw a point above ten (or pass ten – whence the name of the game).

1. Use Galileo's method to compute the number of throws that can give each possible sum. Give the results in an absolute frequency table.

2. Add to the previous table the probabilities of each sum, given as an irreducible fraction and an approximate value to 3 DP.

3. Compute the relative frequencies in the 1000 simulated throws of part A and compare them to the probabilities. Are the values very different ?

4. Compute the probability of "passing ten". Is it a fair game ?

Last year's test

A random race is going on between a hare [1] and a tortoise [2]. It is played by rolling repeatedly a four-sided fair die.

- If a 4 turns up, the hare directly reaches the finish line and wins.
- If a 1, a 2 or a 3 turns up, the tortoise moves towards the finish line and the die is rolled again. The tortoise wins after moving four times.

Part A – Simulation and statistical study

1. Devise a method to simulate the game with a calculator.

2. A sample of 140 races have been simulated. Below are given the winners for each of these races, where the letter H stands for the hare and T for the tortoise.

H – H – H – T – H – H – H – T – T – H – H – T – H – H – H – H – H – T – T – T
T – T – H – T – T – T – T – H – H – H – T – T – T – H – T – T – H – H – H – T
T – H – H – H – H – T – T – T – T – H – H – H – T – H – H – H – H – H – H – H
T – T – T – H – H – H – H – H – T – T – T – T – H – T – H – H – H – T – T – H
T – T – H – T – H – H – T – H – T – H – T – T – T – H – T – H – H – H – T – T
T – H – H – T – H – H – T – T – H – T – T – H – H – H – H – T – T – H – H – H
T – T – H – T – H – H – H – H – H – T – H – H – H – H – H – H – H – H – H – H

3. Copy the table below and fill it out with the absolute and relative frequencies.

1. hare : *lièvre*
2. tortoise : *tortue*

Winner	Hare	Tortoise
Absolute frequencies		
Relative frequencies		

4. Compute the confidence interval at 95% for the percentage of races won by the hare.
5. Explain in a few words what a confidence interval at 95% confidence is.
6. What is the range[3] of this confidence interval ? How could we lower it ?
7. According to this sample, do you think that these rules are to the advantage of the hare or the tortoise ?

Part B – An algorithm

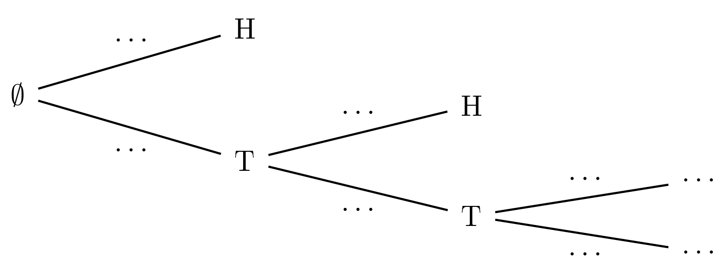
This game can be represented as the following algorithm.

```
Input : 0 → T ; O → H ;
begin
    while H ≠ 1 and T ≠ 4 do
        if alea < 0.25 then
            1 → H ;
        else
            T + 1 → T ;
        end if
    end while
    if ... then
        Output : "The hare wins."
    else
        Output : "The tortoise wins."
    end if
end
```

8. Explain the functions of the numbers H and T in this algorithm.
9. Explain the condition "$H \neq 1$ and $T \neq 4$". Does it ensure that the algorithm will always stop ?
10. What should be entered in the last "If" instruction so that the output will be correct ?

Part C – A probabilistic view

11. Copy and fill out the following probability tree :



12. At the end of each branch of the tree, write down the winner and the probability of this situation.
13. Deduce that the probability of the hare winning is $\dfrac{175}{256}$.
14. According to the following result do you think that these rules are to the advantage of the hare or the tortoise ?
15. The probability computed in question 13 is not included in the confidence interval from part A. Does it mean that something is wrong with one of these computations ?

3. range of an interval : *amplitude*

Glossary

English	French	Explanation
Survey	Sondage	A method for collecting quantitative information about items in a population.
Sample	Échantillon	A subset of a population selected for measurement, observation or questioning, to provide statistical information about the population.
Sampling	Échantillonage	The process or technique of obtaining a representative sample.
Margin of error	Marge d'erreur	An expression of the lack of precision in the results obtained from a sample.
Fluctuation interval	Intervalle de fluctuation	For a certain proportion of samples, the interval where the parameter studied should be.
Estimate (verb)	Estimer	To calculate roughly, often from imperfect data.
Estimate	Estimation	A rough calculation or guess.
Estimation	Estimation	The process of making an estimate.
Point estimate	Estimation ponctuelle	A single value computed from sample data, used as a "best guess" for an unknown population parameter.
Confidence interval	Intervalle de confiance	A particular kind of interval estimate of a population parameter.
Simulate (verb)	Simuler	To model, replicate, duplicate the behavior, appearance or properties of a system or environment
Simulation	Simulation	Something which simulates a system or environment in order to predict actual behaviour.

Aw, people can come up with statistics to prove anything, Kent. Forfty percent of all people know that. (Homer Simpson)

Lottery : A tax on people who are bad at math. (Anonymous)

Do not put your faith in what statistics say until you have carefully considered what they do not say. (William W. Watt)

He uses statistics as a drunken man uses lampposts - for support rather than for illumination. (Andrew Lang)