Historical Contributions of Different Cultures to Mathematics

Authors:

Kabiru Abubakari
Zahraa Al-Mahdawy
Asya Bagan
Jakub Dobrowolski
Claudia Garetto
Akiva Harris
Yuhan Li
Norberto Lucero Azuara
Jordan Marajh
Ivelina Mladenova
Christo Morison
Adam Onus
María Fernanda Pintado Serrano
Silvia Rognone
Robert Sansom

2024 – 2025

Contents

1 Introduction

This year EDI project has been devoted to the contributions of different cultures and populations to important topics in Mathematics. It therefore combines EDI with History of Mathematics and it aims to provide user friendly resources that can be included in our teaching and outreach activities. It is part of a wider project and it is our intention to continue in the next academic years. Keeping in mind the modules that we offer to our undergraduate students, we have decided to start from the following two topics: the Pythagorean Theorem (Chapter 2) and the Irrational Numbers (Chapter 3). I am very grateful to Professor June Barrow-Green (Open University, LSE) who has suggested several papers and books relevant to this project. As always, nothing would be possible, without the dedication, hard-work and enthusiasm of all the PhD students who took part in this project, who this year had the opportunity to collaborate as well with one of our third year graduates and students ambassadors: Zahraa. Kabiru, Zahraa, Huda, Norberto, Jordan, Christo, Adi, Maria, Silvia and Robert: thank you and well done!
I feel very lucky to be able to always count on you!

Claudia Garetto (S&E EDI Lead)
16/09/2024

2 The Pythagorean Theorem

Related topics: Calculus, Geometry

2.1 History

The Pythagorean theorem, a fundamental concept in geometry, has a history reaching back thousands of years. Though named after the Greek philosopher Pythagoras (ca. 570 BCE), evidence shows that the principle was understood and applied long before his time. This section explores the intriguing story of the theorem, following its development from ancient civilizations to its formalization by Pythagoras and beyond.

We look at early uses of Pythagorean triples – sets of three whole numbers that satisfy the equation $a^2 + b^2 = c^2$ in Babylonian, Chinese and Indian mathematics. These examples reveal an intuitive understanding of right triangles, even if the concept was not formally recognized as a theorem. Then, we delve into the evolution of the idea in ancient Greece. Pythagoras, who emphasized the relationship between numbers and geometry, is credited with formally stating and possibly proving the theorem.

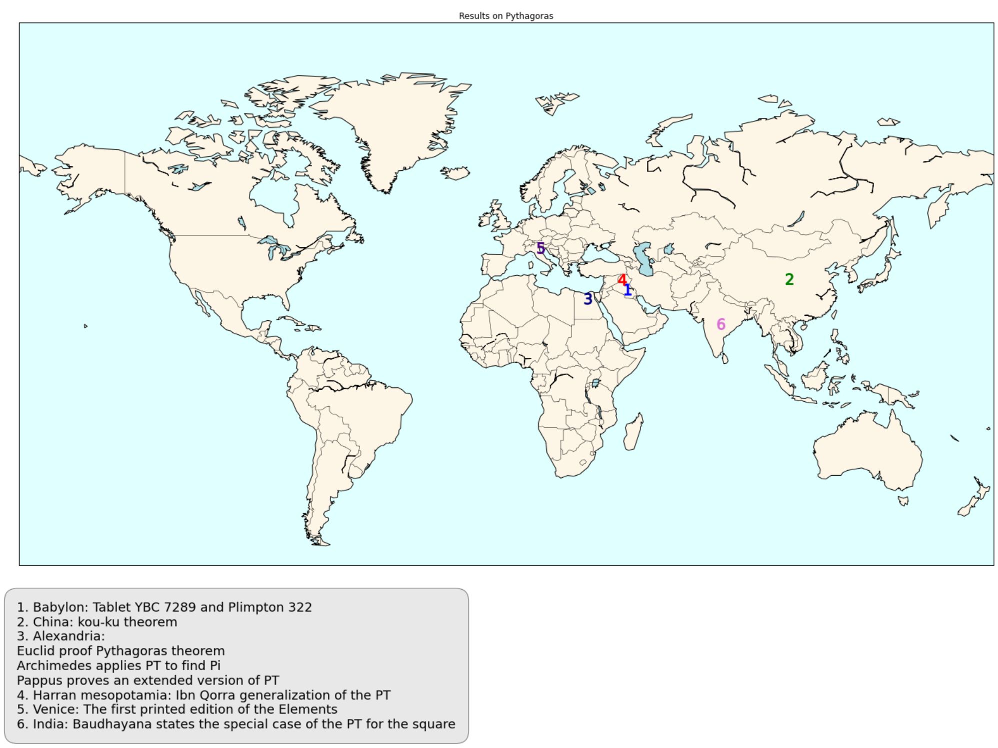


Figure 1: Map of the world with results on Pythagoras

2.1.1 Babylonia

The Fertile Crescent, spanning from the Euphrates to the mountains of Lebanon, birthed Mesopotamia, a prominent ancient civilization in modern-day Iraq Maor [2019]. Discoveries of clay tablets reveal a sophisticated society skilled in commerce, astronomy, arts, and literature. Notably, under the rule of Hammurabi, they created the earliest known legal code, which can be found in the many archaeological clay tablets.

A clay tablet, *YBC*–7289, dating back to the Old Babylonian period (1895-1595 BCE), stands as a testament to the ingenuity of Babylonian mathematicians. Employing a unique number system based on 60, they inscribed an incredibly accurate approximation of the square root of 2 on the tablet. This achievement rivals the precision of modern calculators. Besides that, *YBC*–7289 also hints at the Babylonian's grasp of geometrical concepts, particularly the connection between a square's side and its diagonal, foreshadowing the Pythagorean theorem. The conclusion that Babylonian's knew of the Pythagorean Theorem relies on the fact that they wrote the relationship of length of the diagonal of a square and its side, $d = a\sqrt{2}$. Another

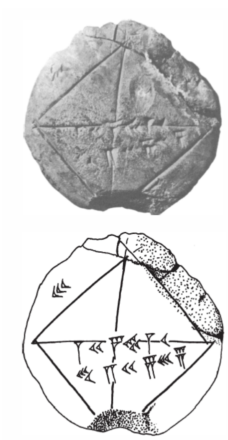


Figure 2: Clay tablet *YBC*–7289 (Taken from Maor [2019]).

significant find, *Plimpton* 322, lists Pythagorean triples (sets of integers a, b, c where $a^2 + b^2 = c^2$), indicating Babylonian knowledge of algebraic methods centuries ahead of their time. Despite missing sections, meticulous reconstruction efforts reveal its content, affirming Babylonian mastery in mathematical theory.

While the Babylonians achievements in identifying Pythagorean triples are undeniably impressive, it's important to recognize that their method applied only to specific right triangles. Developing more general rules based on this concept would require a more formal, mathematical proof, something that wouldn't be achieved until much later.

Figure 3: Chiu chang proof (Taken from Frank J Swetz [1977]).

2.1.2 China

Evidence indicates that Chinese scholars made significant contributions to the understanding of right triangles. Although the Greek origins of the Pythagorean Theorem remain a subject of debate, there is a possibility that the Indian mathematician Bhaskara's work may have been influenced by earlier Chinese mathematics. This influence could be traced back to the ancient Chinese text *The Arithmetical Classic of the Gnomon and the Circular Paths of Heaven*, which is estimated to date to 1100 BCE. This text likely encapsulates the cumulative mathematical knowledge in China up to that period Frank J Swetz [1977].

The *hsuan-thu* proof of the Pythagorean Theorem exemplifies the understanding of this theorem by the Chinese. A notable passage that accompanies the *hsuan-thu* in the *Chou Pei* (the Chinese name for the aforementioned text) is as follows:

> "Let us cut a rectangle (diagonally), and make the width 3 units wide, and the length 4 units long. The diagonal between the two corners will then be 5 units long. Now after drawing a square on this diagonal, circumscribe it by half-rectangles like that which has been left outside, so as to form a square plate. Thus the four outer half-rectangles of width 3, length 4, and diagonal 5, together make two rectangles of area 24; then subtracting 49, the remainder is of area 25. This process is called 'piling up the rectangles' (*chi chu*)."

In addition to the *Chou Pei*, which contains discussions on the application of right triangles, these discussions are obscured by their incorporation into a mystical cosmology. In contrast, subsequent mathematical works are devoid of mystical connotations. Because of the Chinese words for the width and length of a rectangle are *kou* and *hu*, respectively, the Pythagorean Theorem has been know in China as the *kou-ku* theorem.

2.1.3 India

Unlike the isolated development of mathematics in China, India's geographical position exposed it to a diverse array of cultural influences. These cross-pollinations significantly enriched the intellectual landscape, including the realm of mathematics.

As stated in Maor [2019], early Hindu mathematical knowledge is deeply intertwined with religious practices. The *Sulbasutras*, a collection of Vedic texts dating back to at least 600 BCE, provide compelling evidence of this connection. Within these works, we find precise geometric constructions for sacrificial altars, demonstrating a sophisticated grasp of mathematical concepts.

Baudhayana, a prominent author of the *Sulbasutras*, articulated a special case of the Pythagorean theorem: "The rope which is stretched across the diagonal of a square produces an area double the size of the original square." This statement, equivalent to the Pythagorean Theorem for a 45-45-90 degree triangle, is remarkably similar to contemporary understandings.

Katyayana, a later *Sulbasutras* author, extended this concept, stating: "The rope [stretched along the length] of the diagonal of a rectangle makes an [area] which the vertical and horizontal sides make together." This generalization of the Pythagorean Theorem to any right-angled triangle is a testament to the depth of Hindu mathematical knowledge.

Furthermore, the *Sulbasutras* offer detailed instructions for constructing trapezoidal altars, incorporating intricate geometric calculations involving Pythagorean triangles. These practical applications of advanced mathematical concepts underscore the Hindus' mastery of the subject.

The evidence presented in the *Sulbasutras* strongly suggests that the Pythagorean theorem was well-understood in India centuries before Pythagoras, highlighting the rich and independent development of mathematics in this ancient civilization.

2.1.4 Pythagoras

Pythagoras is a mysterious historical figure, often depicted as a wise, bearded philosopher. Much of what we know about him is likely a mix of fact and fiction, with accounts written by historians many years after his time.

Born around 570 BCE on the island of Samos, Pythagoras may have studied with Thales and traveled to Egypt and Persia, absorbing their knowledge. Around 530 BCE, he settled in Croton, Italy, and founded a school that deeply influenced future scholars. The Pythagoreans were a secretive brotherhood, focusing on philosophy, mathematics, and astronomy Maor [2019]. This secrecy, combined with the oral transmission of knowledge, means most of what we know comes from later writers.

Pythagoras made significant discoveries, including in acoustics, where he linked the pitch of sounds to the size of objects. He found that string vibrations relate inversely to length, leading to the concept of musical intervals and harmony. These ideas contributed to the Pythagorean belief in numerical harmony governing the universe, influencing the quadrivium of arithmetic, geometry, music, and astronomy, essential to education in ancient Greece.

In addition to their fascination with sounds, the Pythagoreans were interested in numbers, particularly Pythagorean triples. These are sets of positive integers (a, b, c) that satisfy the equation $a^2 + b^2 = c^2$. As followers of Pythagoras, who is credited with the theorem, they were naturally eager to find right triangles with integer side lengths. However, they quickly realized that while it was easy to pick two sides with integer lengths, the third side would not always be an integer. Despite this challenge, they occasionally discovered Pythagorean triples, a rare event that they reportedly celebrated with great enthusiasm. There is not a clear evidence on how the Pythagoreans found these triangles, but they used the following formula:

$$n^2 + \left(\frac{n^2-1}{2}\right)^2 = \left(\frac{n^2+1}{2}\right)^2$$

to determine that these numbers form a Pythagorean triple for odd values of n.

According to Maor [2019], while the Pythagorean discovery of integer-sided right triangles, known as Pythagorean triples, was significant, it pales in comparison to two events that profoundly influenced the future of mathematics: Pythagoras's proof of his famous theorem and the discovery of irrational numbers—numbers that cannot be expressed as a ratio of two integers. Unfortunately, neither Pythagoras's original proof nor the details of the discovery of irrational numbers have survived, leaving us to rely on later writings and speculative interpretations.

2.2 Different proofs of the Pythagorean Theorem

> *"What is worth proving is worth proving again."*
>
> Attributed to Nick Katz in Ruelle [2023].

It is not uncommon to prove the same mathematical result by two or more different means. There are various theorems known to have a vast amount of proofs, such as the Irrationality of the Square Root of 2, the Fundamental Theorem of Algebra, and the Law of Quadratic Reciprocity, to name a few. With a result as known as the Pythagorean Theorem, it is natural to ask how many proofs exist of this theorem. One possible answer is at least as many mathematics master's students who obtained their degree in the Middle Ages.

According to Elisha Scott Loomis (1852-1940) Loomis [1968], it was a requirement during the Middle Ages to develop an original proof of Pythagorean Theorem before obtaining the Master's degree in mathematics. In the previous section, we learned about the first approaches to this theorem, long before the Greek philosopher walked on this earth. After Pythagoras, a greater diversity of proofs (or demonstrations) followed, from the Middle Ages to the past century. The answer to our initial question is that we don't know. However, we have a lower bound for this value, which is 371. Loomis compiled all-known proofs as of 1940 in the second edition of the book The Pythagorean Proposition.[1]

[1] The first edition published in 1927 contains 230 proofs.

The following paragraphs present four intriguing proofs of the Pythagorean Theorem. These proofs stand out due to their ingenuity or the notable individuals who devised them. Perhaps these examples will inspire the reader to create one of their own.

2.2.1 President Garfield's proof

The first proof we present is by James A. Garfield, the twentieth president of the United States. Throughout his youth, Garfield worked on canal boats, as a carpenter, a part-time teacher, and a janitor Doenecke [2024]. In 1854, at the age of twenty-three, he entered Williams College in Massachusetts, becoming the oldest student there. After graduating, he taught various subjects at the Eclectic Institute in Ohio, mainly classical languages but also mathematics. Garfield later pursued a career in law and politics. In 1876, four years before his presidential appointment, he proposed the proof below during a mathematical discussion with other Members of Congress.

Proof 1. Start with the right triangle ACB in Figure 4, and extend CB to D such that $BD = AC = b$. Draw $DE = CB = a$ perpendicular to BD.

The triangles ACB and BDE are congruent since they have two pairs of equal sides. Thus, $\angle ABC$ and $\angle EBD$ are complementary, and ABE is a right angle.

The area of the trapezoid $ACDE$ is $\frac{1}{2}(b+a) \times (a+b) = \frac{1}{2}(a+b)^2$, which is equal to the area ABE plus twice the area ACB, that is, $\frac{1}{2}c^2 + 2\frac{1}{2}ab = \frac{1}{2}c^2 + ab$. Hence, this yields

$$\frac{(a+b)^2}{2} = \frac{c^2}{2} + ab \implies a^2 + b^2 = c^2$$

□

2.2.2 Similarity of triangles proof

The following proof has been attributed to twelve-year old Einstein Schroeder [1991], but in fact traces back to Legendre and Euclid's second proof Maor [2019]. The proof utilises the similarity of triangles.

Proof 2. Take a right-angled triangle as shown on the left hand side of Figure 5 where A_1 represents the area of the triangle. The right hand side of Figure 5 shows the triangle bisected into two smaller right-angled triangles with areas A_2 and A_3. These two smaller triangles are similar in that their corresponding angles are congruent and their corresponding sides are in proportion. Moreover, they are both similar to the larger triangle.

In Euclidean geometry, the area of a triangle can be related to the length of its hypotenuse in the following manner. For a right-angled triangle with hypotenuse

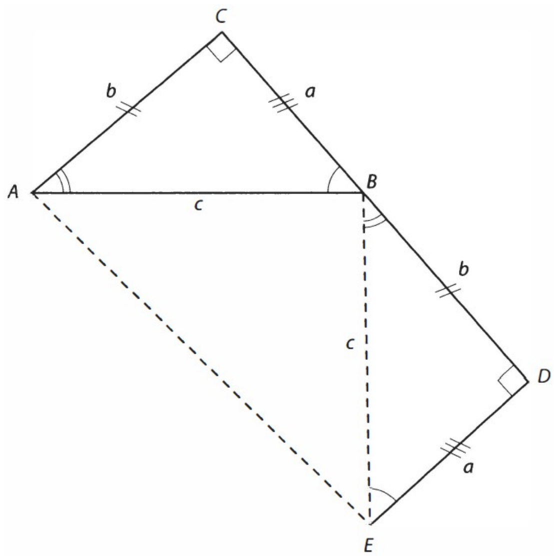


Figure 4: James A. Garfield's proof (1876) (reproduced from Figure 8.8 of Maor [2019]).

d and area D, construct a square with side length equal to the hypotenuse of the triangle as in Figure 6.

Then the area D of the triangle is proportional to the area of the square, $D \propto d^2$. Performing this construction for the three triangles in Figure 5, yields the equations

$$A_1 = kc^2, \qquad A_2 = ka^2, \qquad A_3 = kb^2, \tag{1}$$

where k is a positive number which is the same in each equation due to the similarity of the triangles. Using the fact that the sum of the two smaller triangles is equal to the larger one,

$$A_1 = A_2 + A_3,$$

with the relations (1) obtains

$$kc^2 = ka^2 + kb^2.$$

Dividing through by k produces

$$a^2 + b^2 = c^2.$$

□

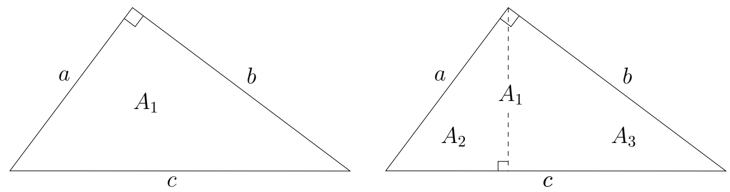


Figure 5: Bisecting a right angle triangle

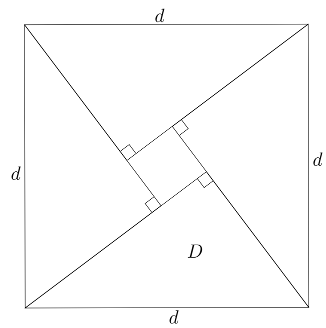


Figure 6: The area of a right-angled triangle with hypotenuse d is proportional to the area of a square of side length d.

2.2.3 The folding bag proof

There is an important generalisation of the Pythagorean theorem that was first elucidated by Euclid. In the geometrical picture, rather than the formula only holding for sums of squares (the squares extending outside the triangle, along each edge), it actually holds for all (similar) shapes. This can be argued for by noticing that areas of similar polygons are in the same ratio as the squares of their corresponding sides, so that if we believe the Pythagorean theorem then the generalisation follows. Alternatively, if we prove the theorem for another set of shapes, then the Pythagorean theorem follows. For example, circles with diameters equal to the side-lengths of the triangle obey the same summation rule. The following proof demonstrates how right triangles themselves can be used, known as the "folding bag proof".

Proof 3. We first note that we can also draw the chosen (similar) shapes, based on the side-lengths of the triangle in question, *inside* the triangle, rather than outside of it, as in the usual geometric picture. By taking internal right triangles as the (similar) polygons, Figure 7 demonstrates that the hypotenuse's triangle (with hypotenuse AB) is the triangle itself ABC, and that the other two triangles (with hypotenuses AC and BC) 'fold' to sum to the same area, proving the theorem. □

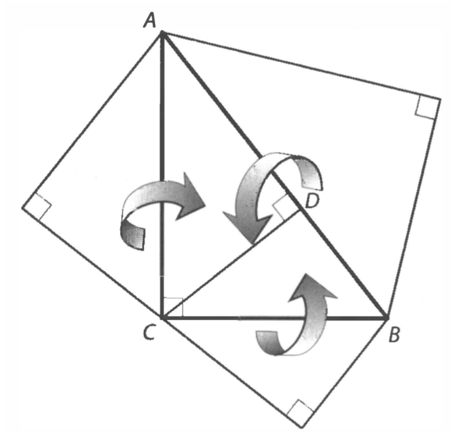


Figure 7: If we choose right triangles as the polygons based on the sides of the original right triangle, we obtain the delightfully simple folding bag proof (reproduced from Figure S4.1 of Maor [2019]).

2.2.4 Miss Ann Condit's proof

The extraordinary Miss Ann Condit, a sixteen-year-old high school student from Indiana (United States), astounded everyone in 1938 with her sophisticated proof presented below. This is the first proof to be devised where all auxiliary lines and triangles originate from the midpoint of the hypotenuse of the given triangle.

Proof 4. Let ABC be a right triangle in Figure 8. Erect the squares $ACDE$, $BCFG$, and $ABHI$, and connect D and F. Let CP be the bisector of AB, and let it meet DF at R.

The triangle ABC can be inscribed in a circle with center P and diameter AB because $\angle ABC$ is a right angle. Thus, $AP = PC$.

The triangles ABC and DFC have two equal sides and share the same right angle between them at C. Hence, the triangles are congruent, and $\angle CDF = \angle BAC = \alpha$. Since ACP is isosceles, $\angle ACP = \alpha$. Therefore, $\angle DCR = 90° - \alpha$, and $\angle CRD = 90°$, so PR is perpendicular to DF.

From P draw PM, PN, and PL to the midpoints of ED,FG and HI, respectively. The area $PFC = \frac{1}{2}(FC \times FN)$, but $FN = \frac{1}{2}FG = \frac{1}{2}FC$, thus area $PFC = \frac{1}{4}(FC \times FC) = \frac{1}{4}BCFG$. Analogously, areas $PDC = \frac{1}{4}ACDE$ and $PAI = \frac{1}{4}ABHI$.

Using the fact that the areas of two triangles with the same base are to each other as their altitudes, we obtain

$$\frac{PDC + PFC}{PAI} = \frac{DR + RF}{AI} = \frac{DF}{AI} = \frac{AB}{AB} = 1.$$

Substituting the expressions for areas PDC, PFC and PAI, we get

$$\frac{ACDE + BCFG}{ABHI} = 1.$$

Therefore, we have proved that the area of the square laying on the hypotenuse is the sum the squares erected over the two smaller sides of the right triangle.

□

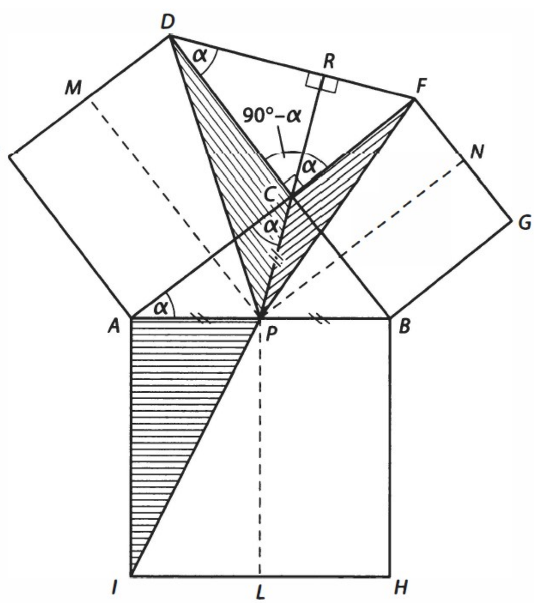


Figure 8: Ann Condit's proof (1938) (reproduced from Figure 8.9 of Maor [2019]).

2.3 Related results

The enduring popularity of the Pythagorean Theorem is a testament to its profound impact on mathematics and its diverse applications across various disciplines. Its attractiveness lies not just in its rich historical background and the multitude of proofs it has inspired, but also in its remarkable appearances in different areas of mathematics and beyond.

In this section, we present some fascinating relationships and applications of the Pythagorean Theorem, illustrating how a simple geometric insight can have far-reaching implications across various domains.

2.3.1 Fermat's Last Theorem

Theorem 2.1. *With $n, x, y, z \in \mathbb{N}$ and $n > 2$, the equation $x^n + y^n = z^n$ has no solutions.*

The theorem stated above, known as Fermat's Last Theorem, was famously accompanied by Pierre de Fermat's note from around 1637:

> *"I have found an admirable proof of this, but the margin is too narrow to contain it."*

This enigmatic note sparked one of the most famous mathematical quests, captivating mathematicians for over 300 years. It was not until 1994 that Andrew Wiles finally proved the theorem. When $n = 2$, we encounter the case of the Pythagorean Theorem, for which there are indeed infinitely many solutions. These solutions, or triples of numbers (x, y, z) are known as the Pythagorean triples, and there exists a collection of formulas for generating such triples.[2]

The simplicity of the statement of Fermat's Last Theorem conceals the profound complexity and depth of the mathematics it encompasses. With the Pythagorean Theorem as a special case, we can appreciate the complex world of mathematics that arises from seemingly simple equations.

2.3.2 Archimedes' approximation of π

The Pythagorean theorem was an essential tool in Archimedes' approximation of π. Born in Sicily, Archimedes (287-212 BCE) is widely regarded as the greatest scientist of the ancient world and is known for his discoveries and inventions across physics, astronomy, engineering and mathematics. By Archimedes' time, civilisations across the world had been attempting to measure the value of π for more than 1000 years. Archimedes is credited with being the first person to devise an algorithm to calculate the value of π. He showed that the value of π lies in the region

$$\frac{223}{71} < \pi < \frac{22}{7}$$

by squeezing a circle between two polygons and increasing the number of sides of the polygons. Calculating the perimeter of these polygons yields an approximation for π.

Briefly, the algorithm works as follows. Begin with a circle of radius 1 and centre O. Inscribe within it a hexagon as in Figure 9(a). The side length s_6 of the hexagon can be simply calculated using the sine rule. The key to Archimedes approximation is that s_{12}, the side length of the dodecagon in Figure 9(b), can be calculated from s_6 by an application of Pythagoras theorem.

Mathematically,

$$s_{12}^2 = \left(\frac{s_6}{2}\right)^2 + (1 - OA)^2.$$

[2]See Weissten [2024], Wikipedia contributors [a] for some methods to generate Pythagorean triples.

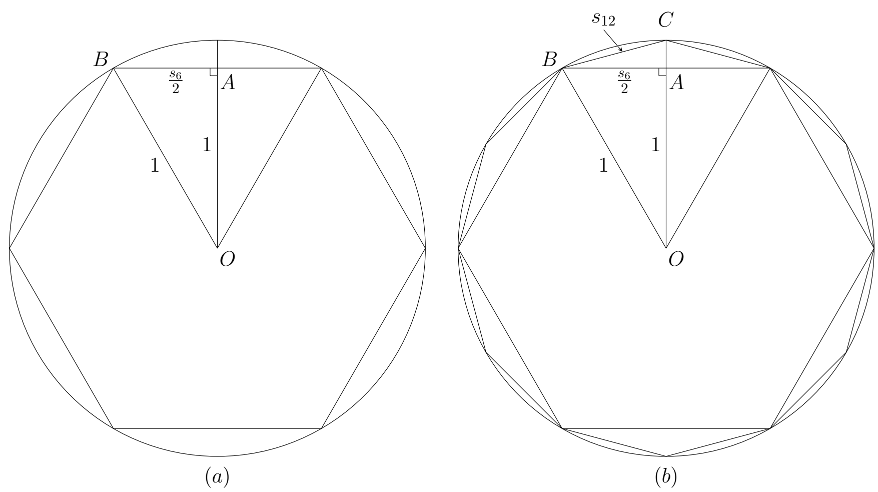


Figure 9: (a): A hexagon of side length s_6 inscribed within a unit circle. (b) A hexagon and dodecagon of side lengths s_6 and s_{12}, respectively, inscribed within a unit circle.

The length OA itself can be calculated from Pythagoras theorem as

$$OA^2 = 1 - \left(\frac{s_6}{2}\right)^2.$$

Thus for s_{12} the following formula is obtained:

$$s_{12}^2 = \left(\frac{s_6}{2}\right)^2 + \left(1 - \sqrt{1 - \left(\frac{s_6}{2}\right)^2}\right)^2.$$

This formula can be used iteratively to find s_{24} and so on. Moreover, there is nothing special about the hexagon and the dodecagon; this algorithm will work for any n-gon.

For Archimedes' approximation above he began with the hexagon and constructed sequentially the dodecagon in Figure 9(b), then a 24-gon all the way up to a 96-gon. To construct the upperbound, he used an analogous algorithm with a unit circle inscribed within polygons. The elegance of the Archimedes' approximation to π lies in the fact that this algorithm can be used to estimate π to *any* desired accuracy.

2.3.3 Heron's formula

Commonly attributed to the Greek mathematician and engineer Heron of Alexandria (ca. 100 BCE - 100 CE), the following formula for the area A of a triangle with sides a, b and c, relies on a double application of the Pythagorean theorem:

$$A = \sqrt{s(s-a)(s-b)(s-c)}, \tag{2}$$

where $s = \frac{1}{2}(a+b+c)$ is the semiperimeter of the triangle. We briefly present the proof of this formula, relying on Figure 10.

Proof 5. First, draw the altitude h from the top vertex to side a, dividing the side into parts m and n. From Pythagoras, we have

$$m^2 + h^2 = b^2,$$
$$n^2 + h^2 = c^2.$$

Subtracting the bottom equation from the first, we get

$$m^2 - n^2 = b^2 - c^2.$$

After some algebraic manipulation, we obtain

$$m = \frac{a^2 + b^2 - c^2}{2a}, \qquad n = \frac{a^2 - b^2 + c^2}{2a},$$

which can be substituted in $h^2 = b^2 - m^2$ to arrive to an expression for the altitude in terms of the semiperimeter:

$$h^2 = \frac{4s(s-a)(s-b)(s-c)}{a^2}.$$

Finally, we take the square root of h, and substitute in the formula for the area of a triangle $A = ah/2$ to arrive to the desired result.

□

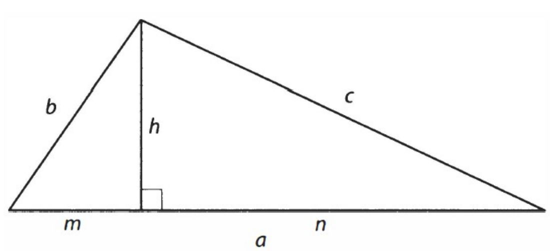


Figure 10: Triangle to prove Heron's formula.

Despite the formula being commonly associated with Heron of Alexandria (and carrying his name), historical accounts suggest otherwise. According to Al-Biruni, an Arabic astronomer who lived around 973-1048, Archimedes is credited with discovering this result. Additionally, an equivalent formula was published in the *Mathematical Treatise in Nine Sections*, authored by the Chinese mathematician Qin Jiushao in 1247 Strick [2022], stated as

$$A = \frac{1}{2}\sqrt{a^2c^2 - \left(\frac{a^2 + c^2 - b^2}{2}\right)^2}.$$

2.3.4 The diagonal of a square

We conclude this chapter by highlighting a special and somewhat obvious case of the Pythagorean Theorem where two sides are of equal length: the diagonal of a square. If a square has a side of length a, then the length of its diagonal d can be found using the formula $d^2 = a^2 + a^2 = 2a^2$, which simplifies to $d = a\sqrt{2}$.

While this result might seem straightforward and innocent, known to the Babylonians for thousands of years, it incited considerable controversy among the ancient Greek philosophers. This is due to the fact that this quantity is not a rational number. In the next chapter, we delve further into irrational numbers, including a detailed exploration of $\sqrt{2}$.

3 Irrational Numbers

Related topics: Analysis, Algebra, Set Theory

3.1 What are irrational numbers?

Irrational numbers are numbers that cannot be written as a ratio of two integers (i.e., they cannot be written as a simple fraction where the numerator and denominator are both integers. When we write out irrational numbers in decimal form, the digits go on forever and do not repeat! Let's take a look at an example:

$$\pi = 3.14159265358979323846264338327950288419716939937510582 09\ldots$$

The numbers after the decimal place never stop and never repeat making π an irrational number!

3.1.1 Why are irrational numbers useful?

Irrational numbers help us solve equations or calculate areas of shape. If you think back to circles, we use the equation πr^2. Another useful irrational number is $\sqrt{2}$ as it helps us understand right-angle triangles.

3.2 $\sqrt{2}$

Throughout history, $\sqrt{2}$ has held profound significance not only in the realm of mathematics but also in various practical and cultural contexts. This enigmatic number, the first known irrational number, has been crucial for understanding and solving numerous problems in geometry, architecture, and beyond. Even more importantly, its discovery challenged ancient mathematical paradigms and opened new avenues for critical mathematical thought.

The square root of 2 has represented a gateway to the concept of irrationality and a cornerstone in the development of number theory. Its value and properties have been studied extensively, driving mathematical innovation and inspiring a multitude of approximations and proofs across different civilizations.

In this section, we begin by presenting the Pythagorean proof of the irrationality of $\sqrt{2}$, then show the ancient Babylonian approximation from the Clay Tablet *YBC–* 7289. We conclude the section with a description of the geometric and elegant approximation by the Indian mathematician Baudhayana, showing how deeply this irrational number has permeated different cultures at different times.

3.2.1 Proof that $\sqrt{2}$ is irrational

As previously mentioned, any number n that can be expressed as a ratio of two integers, such as $n = \frac{a}{b}$, is a rational number.

For a long time, philosophers and mathematicians believed that all numbers were rational. Pythagoras, an ancient Greek philosopher, and his followers held this belief sacredly and anyone who attempted to challenge this belief was considered a heretic.

In the 5^{th} century BCE, Hippasus discovered a way to prove that $\sqrt{2}$ is indeed irrational. According to some traditions, he was thrown overboard and drowned for revealing this discovery to the Pythagorean brotherhood as a punishment for his "sins" Iamblichus [1918]. Lets take a look at this dangerous proof:

In order to prove that $\sqrt{2}$ is an irrational number we will do this by "proof by contradiction". For those who are not familiar with this, proof by contradiction is done by assuming that what we want to prove is not true and then showing that the conclusion is a contradiction. In this example, we are trying to prove the statement "$\sqrt{2}$ is an irrational number", so we start our proof by assuming that "$\sqrt{2}$ is **not** an irrational number" and see that this cannot be true, thus our original statement must be true.

Proof 6. Assume that $\sqrt{2}$ is a rational number. This means we can write $\sqrt{2} = \frac{a}{b}$, where a and b are whole numbers and $b \neq 0$.
Assume that $\frac{a}{b}$ is written in its simplest form, meaning a and b have no common factors other than 1. This means both a and b cannot be even, as otherwise 2 would be a common factor, so one or both must be odd.
Starting from $\sqrt{2} = \frac{a}{b}$, we square both sides to get:

$$2 = \frac{a^2}{b^2}.$$

Multiplying both sides by b^2 gives us:

$$a^2 = 2b^2. \tag{3}$$

Now we can see that a^2 is even as it is twice as much as b^2. Since a^2 is even, a must also be even. (Why? Because only even numbers squared give even results).
Since a is even, we can write a as $2k$ for some integer k. Substituting $2k$ for a into Equation 3, we get:

$$(2k)^2 = 4k^2 = 2b^2.$$

Dividing both sides by 2, we get:

$$2k^2 = b^2.$$

This means that b^2 is also even, and hence b must be even as well (using the same concept as earlier when stating that a is even).
But if both a and b are even, then $\frac{a}{b}$ was not in its simplest form, which contradicts our assumption! Therefore, our initial assumption that $\sqrt{2}$ is a rational number must be false. Thus, $\sqrt{2}$ is an irrational number. □

3.2.2 The Babylonian Clay Tablet *YBC*–7289 and Irrational Numbers

The Babylonian clay tablet *YBC*–7289, dating back to 1800-1600 BCE and originating from modern-day Iraq, is a remarkable artefact in the history of mathematics. This small tablet shows that the ancient Babylonians had an advanced understanding of numbers, especially irrational numbers, long before the Greeks. *YBC*–7289 has scripts that are written by Babylonian students who learned to write and calculate in cuneifrom (the script used by Mesopotamian scribes and scholars). Its round shape and large writing indicates it was typical tool for Babylonian students *imšukkum* or "hand tablet" because they fit comfortably in the students palm.Yale University

The tablet also reveals that the Babylonians understood Pythagorean principles over a thousand years before Pythagoras! They knew the relationship between the sides of a right triangle and its diagonal (the hypotenuse).

YBC–7289 contains an approximation of the square root of 2. It shows a square with its diagonals and cuneiform inscriptions. The Babylonians calculated $\sqrt{2}$ to four decimal places (1.4142), which is incredibly precise for that era. This demonstrates their ability to work with irrational numbers. Additionally, *YBC*–7289 highlights the Babylonians' use of a sexagesimal (base 60) number system, which contributed to their precise calculations. This ancient system is the foundation of our modern time-keeping and angular measurements.

3.2.3 Baudhayana approximation

Baudhayana (ca. 800-740 BCE) was one of the greatest Indian mathematicians, and his contributions permitted a significant advancement towards the understanding of different concepts in mathematics Plofker [2009]. As a distinguished Hindu high priest, he was not only an expert mathematician but also an accomplished architect and astronomer. He introduced several mathematical formulas known as the *Sulva-sutras*, some of which include very elegant proofs Thibaut [1875], like the proof of the Pythagorean theorem, particularly notable for its simplicity, and one of the first approximations of the value of pi, which he approximated as 3.

We will now showcase the approximation method of Baudhayana for the square root of 2 Khatri and Tiwari [2023]. He started with the simple idea of transforming a rectangle into a square, obtaining a value of $\sqrt{2}$ equal to:

$$\frac{1}{3}+\frac{1}{12}-\frac{1}{408}=\frac{577}{408}=\mathbf{1.41421}56863$$

which approximates the correct value of the square root of two up to 5 decimal places. The following steps outline the proof of his approximation.

Proof 7. Consider a rectangle with sides of length $\ell_1 = 1$ and $\ell_2 = 2$ and area $A_r = \ell_1 \cdot \ell_2 = 2$. Baudhayana had the goal to transform this rectangle into a square with the same area, i.e. $A_s = 2$, and so with side $\sqrt{2}$.
With this in mind, divide the rectangle into squares $ABEF$ and $FECD$ with side 1, and then portion $FECD$ into three equal rectangles. Notice that, by doing so,

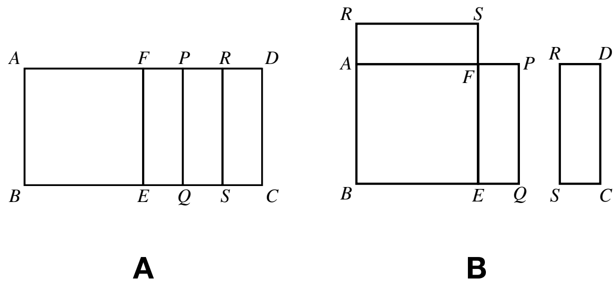


Figure 11: **A** First step of the Baudhayana geometrical approximation of $\sqrt{2}$, in which rectangle $ABCD$ is first halved in two equal squares, then square $FECD$ is divided into three rectangles with the same area. **B** Thanks to a rigid translation, rectangle $PQSR$ moves on one side of square $ABEF$, while rectangle $FEQP$ remains in its initial position. Reproduced from Khatri and Tiwari [2023].

segments $\overline{EQ} = \overline{QS} = \overline{SC} = \frac{1}{3}$, as shown in Figure 11 A.
Now move rectangle $PQSR$ to one side of square $ABEF$ with a rigid translation, while leaving rectangle $FEQP$ in its initial position. Notice that $\overline{BQ} = \overline{BE} + \overline{EQ} = 1 + \frac{1}{3}$ as depicted in Figure 11 B.
Apply the same procedure to the rectangle $RSCD$, dividing it into four equal rectangles, each one of them with the smaller side equal to $\frac{1}{12}$ (see Figure 12 A for reference). Move two of these rectangles to be adjacent to RS and PQ, as illustrated in Figure 12 B. Observe that

$$\overline{BN} = \overline{BE} + \overline{EQ} + \overline{QN} = 1 + \frac{1}{3} + \frac{1}{12}$$

and that the square $EFGD$ has side equal to $\frac{1}{3} + \frac{1}{12}$ and consequently, area $A_e = \left(\frac{1}{3} + \frac{1}{12}\right)^2$.
Finally, fill the area of the square $MFOG$ to complete the square $LBNG$ with two contributions: (i) add the rectangle $ONCD$, with area $2 \cdot \frac{1}{12} \cdot 1 = \frac{1}{6}$, and (ii) subtract two small portions of width x along sides LB and BN (see Figure 13 A for reference), which he called *savisesah*, "the missing part". He obtained:

$$\frac{1}{6} + 2x \cdot \left(1 + \frac{1}{3} + \frac{1}{12}\right) = \left(\frac{1}{3} + \frac{1}{12}\right)^2$$

Solving for x gives $x = \frac{1}{408}$ and so the rectangle has been transformed into a square with side lengths $\frac{1}{3} + \frac{1}{12} - \frac{1}{408}$.

□

Notice that, in the "proof", Baudhayana left a small empty square at the corner of

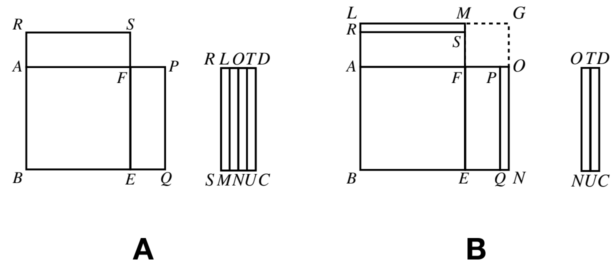


Figure 12: Rectangle $RSCD$ is divided into four equal rectangles (**A**), with two of them then moved to be adjacent to segments RS and PQ respectively (**B**). Reproduced from Khatri and Tiwari [2023].

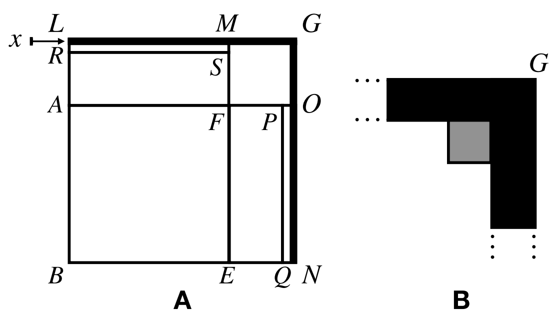


Figure 13: **A** As a final step of his proof, Baudhayana subtracts the *savisesah* (represented in the figure by the two black solid rectangles with shortest side equal to x) from the other contributions to obtain the approximation of square root of 2. **B** The grey square in the figure represents the missing area that the Baudhayana approximation couldn't cover and can lead to the discovery of the irrationality of the square root of 2. Adapted from Khatri and Tiwari [2023].

$MFOG$: in fact, by summing up contributions (i) and (ii), he was just approximating the area of $MFOG$. The missing area is indeed a hint pointing to the irrationality of the square root of two (see Figure 13 B for reference).

3.3 π

Across millennia, the presence of the irrational number π has permeated various fields, from engineering to art, and continues to inspire new generations of scholars, representing one of the most intriguing numbers in mathematics.

Spanning different cultures and regions of the world, π has indeed captivated the minds of mathematicians and scientists, each contributing unique insights and techniques to its understanding. The journey of π in space and time is a testament to humanity's relentless pursuit of precision and mathematical curiosity.

We aim here to explore the elegant π approximations developed by Chinese mathematicians, move on to the profound contributions of Ramanujan, and finally discuss the wide-ranging applications of this fundamental number in various disciplines.

3.3.1 Approximations in China

The *Nine Chapters on the Mathematical Art* (ca. 200 BCE) Dauben [2013] is one of the most influential mathematical texts in Chinese history. Compiled during the Han dynasty, it presents a comprehensive system of practical mathematics used for various purposes, such as land measurement, construction, and trade.

Liu Hui (ca. 225–295 CE) was a prominent Chinese mathematician who provided detailed commentaries on these texts Straffin [1998]. His work demonstrated advanced mathematical techniques and provided more precise approximations of irrational numbers. He developed an elegant algorithm for calculating π, which permitted a better approximation of the irrational number than the ones proposed by his precursors. In fact, before Liu Hui's contributions, the ratio of a circle's circumference to its diameter was often approximated as 3 in China. As two major examples, the mathematician Zhang Heng (ca. 78–139 CE) proposed the value of approximately $\frac{92}{29} \sim 3.1724$ Needham [1959], while Wang Fan (219–257 CE) Schepler [1947], suggested $\pi \sim \frac{142}{45} \sim 3.156$. Despite these empirical approximations providing accuracy up to two decimal places, Liu Hui found them unsatisfactory, criticizing them for being too large.

In his commentary on *Nine Chapters on the Mathematical Art*, Liu Hui observed that π must be greater than 3 since the ratio of the circumference of an inscribed hexagon to the diameter of the circle was three. From this intuition, firstly he detailed an iterative algorithm for calculating π by bisecting polygons, with its approximation lying between 3.141024 and 3.142708, and proposing 3.14 as a sufficiently accurate approximation. He did not stop there, and acknowledging that this estimate was slightly low, Liu Hui later devised a more efficient and iterative algorithm, achieving an approximation of $\pi \sim 3.1416$, accurate to five decimal places, with the only use of a 96-sided polygon Needham [1959].

3.3.2 Ramanujan's work

Srinivasa Ramanujan (22 December 1887 – 26 April 1920) was a brilliant and prodigious Indian mathematician. Though he had almost no formal training in pure mathematics, he made extraordinary contributions to the study of π among other fascinating mathematical discoveries, spanning from number theory to continued fractions and infinite series.

During his brief yet prolific life, Ramanujan provided around 3900 identities and equations Berndt [1997], with a significant number of these entirely new and collected in his famous notebooks (see Figure 14). From the Western perspective, Ramanujan is often seen as a natural genius whose intuition was extraordinary, even

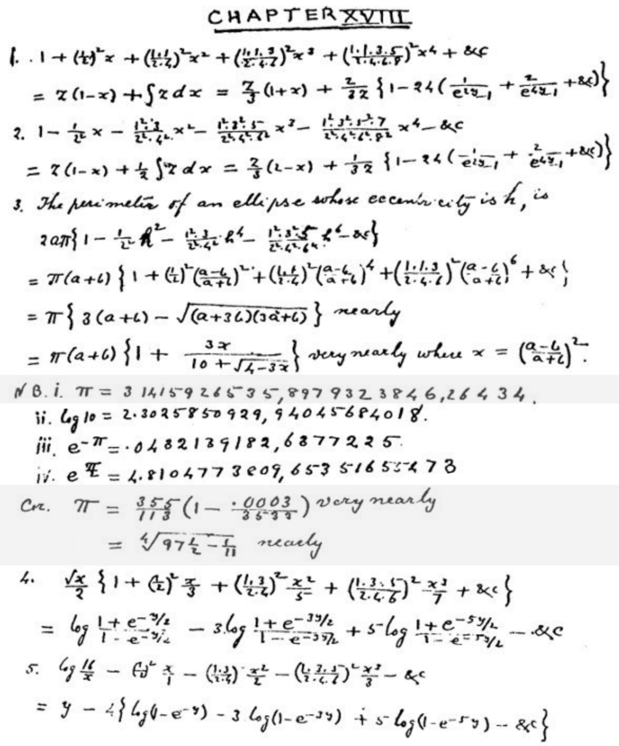


Figure 14: A page from Srinivasa Ramanujan's mathematical notebooks. The highlighted sections refer to his calculations regarding π.

though he lacked formal training. In contrast, the Indian perspective celebrates Ramanujan not only for his genius but more for his mathematics, viewing himself as a cultural and national icon. His story is one of inspiration, depicting how someone from a modest background with limited resources, but deep passion and hard work, could achieve greatness.

Remarkably, the vast majority of his results have been verified as correct, and his groundbreaking formulas, derived from deep insights into complex analysis and modular forms, have greatly influenced subsequent mathematical research on π. For all these reasons, Ramanujan's contributions continue to inspire and challenge mathematicians worldwide, even today.

His work in the early 20^{th} century introduced a series of rapidly converging infinite series that provided unprecedented precision for calculating π Chan et al. [2004]. In fact, unlike previous methods that required a large number of iterations to achieve similar accuracy, Ramanujan's formulas could produce many decimal places of π with remarkable efficiency. He started with expressions that approximate the value of π up to a certain number of decimal places. The first one:

$$\frac{63}{25}\left(\frac{17+15\sqrt{5}}{7+15\sqrt{5}}\right)$$

gives the approximation of $\pi \sim 3.14159265380$, correct up to 9 decimal places. The second one:

$$\frac{12}{\sqrt{130}}\ln\left[\frac{(3+\sqrt{13})(\sqrt{8}+\sqrt{10})}{2}\right]$$

provides the correct value of π up to 14 decimal places; and the third one:

$$\frac{4}{\sqrt{522}}\ln\left[\left(\frac{5+\sqrt{29}}{\sqrt{2}}\right)^3(5\sqrt{29}+11\sqrt{6})\left(\sqrt{\frac{9+3\sqrt{6}}{4}}+\sqrt{\frac{5+3\sqrt{6}}{4}}\right)^6\right]$$

remarkably gives the value of π up to 30 decimal places.

In 1914, the Quarterly Journal of Pure and Applied Mathematics published Ramanujan's work titled "*Modular Equations and Approximations to π*" Berggren et al. [1997]. In this paper, he introduced seventeen different series that converged rapidly to π. Among these, one of the most famous series was the following:

$$\frac{1}{\pi}=\frac{2\sqrt{2}}{9801}\sum_{k=0}^{\infty}\frac{(4k)!(1103+26390k)}{(k!)^4 396^{4k}}$$

Truncating the sum to the first term also gives an approximation which is correct to six decimal places. Truncating after the first two terms gives a value correct to 14 decimal places! This approach laid the groundwork for the fastest-known algorithm: in 1987, mathematician and programmer Bill Gosper utilized this algorithm on a computer, successfully calculating π to approximately 17 million decimal places Arndt and Haenel [2012]. Later on, by building on Ramanujan's methods, mathematicians David and Gregory Chudnovsky developed their variants, which they used to compute π to an astonishing 4 billion decimal places with their homemade parallel computer Chudnovsky and Chudnovsky [1988].

3.3.3 Applications of π across different cultures, fields, and times

The journey of different applications of the number π from ancient to modern times and across many cultures is a clear proof of its timeless significance and versatility.

While much of Mayan literature was lost during the Spanish conquest, the remnants suggest that their approximated value of π was likely more precise than that of their European counterparts at the time. The necessity of a precise π value for their calendar calculations was in fact at the very basis of Mayans' advanced mathematical capabilities and their significant contributions to the understanding and application of π in ancient times. The Mayans indeed developed a highly sophisticated number system, which many historians believe included advanced features unparalleled by other contemporary cultures Powell [2010]. This mathematical system was crucial for their astronomical measurements, which they carried out with remarkable accuracy using simple tools like sticks. The Caracol building in Chichén Itza is a testament to their astronomical knowledge. Often considered a Mayan observatory, many of its windows align with significant astronomical events, such as the setting sun on the spring equinox and specific lunar alignments. This precision in astronomical observations was essential for creating their accurate calendar system, which required a refined value of π (see Figure 15, left).

The application of π for architectural and construction techniques was crucial also in ancient Egypt and Babylon. The Egyptians, renowned for their monumental structures, utilized a value of π in the design of the Great Pyramid of Giza: the ratio of the pyramid's perimeter to its height is indeed approximately 2π, suggesting a profound understanding of the circle's properties Cooper [2011]. This application of π helped achieve not only the pyramid's perfect proportions but also its stability. Similarly, in ancient Babylon, mathematicians and builders employed π in the construction of their ziggurats, which were massive terraced structures serving as temples. As we already discussed in this Chapter, Babylonians had a good approximated value of π that they used in calculations for circular structures and for developing accurate engineering plans, as made evident by the monumental Etemenanki ziggurat in Babylon Neugebauer [1969].

Also in ancient China, the application of π was fundamental to advancements in astronomy and the development of the calendar. In fact, the Chinese calendar, which combined solar and lunar cycles, required precise calculations of the positions of celestial bodies. For this reason, Chinese astronomers relied on their approximated value of π to predict astronomical events such as eclipses, solstices, and equinoxes Needham [1962]. This knowledge was vital for agricultural planning, ceremonial events, and governance. In addition to theoretical advancements, practical applications of π can be seen in ancient Chinese instruments and structures. Chinese astronomical observatories were indeed equipped with sophisticated instruments, like the armillary sphere, which required precise geometric calculations involving π to measure the positions of stars and planets accurately (see Figure 15, middle).

Greek architecture also reflects the practical application of π in architecture. The design and construction of the Parthenon in Athens, for instance, demonstrate an understanding of mathematical principles, including the use of π in circular elements such as columns and decorative friezes.

Islamic architects and artisans applied π in the construction of mosques, madrasas, and palaces, where intricate geometric patterns and designs were essential elements. The precision required to create these complex patterns demanded a thorough understanding of geometric principles, including the use of π. The Alhambra in Spain (see Figure 15, right) and the Selimiye Mosque in Turkey are prime examples of beautiful Islamic architecture that showcase the sophisticated use of π for circular designs and domes. The use of π in Islamic culture extended beyond mathematics and architecture, playing a significant role in the development of geometry, which was crucial for both religious and scientific purposes Hogendijk [1994]. Islamic art often featured elaborate geometric patterns, where the precise calculation of angles and arcs required an implicit understanding of π. These patterns were not only aesthet-

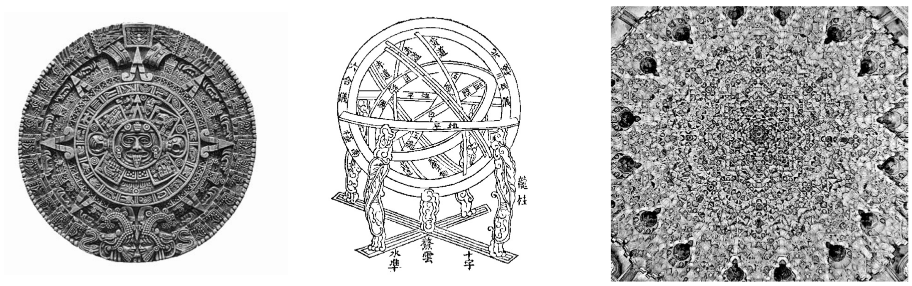


Figure 15: On the left of the figure, Mayans' calendar. An illustration of the ancient Chinese water-powered armillary sphere is shown in the middle of the figure Heng [Han Dynasty]. On the right, a photo of the "Alhambra ceiling" taken by the photographer and artist Jason Priem.

ically pleasing but also held spiritual significance.

In modern times, the application of π has expanded far beyond its ancient uses, becoming a cornerstone of various scientific and engineering fields. In civil engineering, π is essential in the design and construction of bridges, like the Golden Gate Bridge in San Francisco (see Figure 16, left). The calculations for the curvature of arches, the load distribution across circular structures, and the precise measurements required for stability and durability all depend on this irrational number. The value π was fundamental also in the design and navigation of spacecraft. The trajectories of spacecraft, the calculation of orbital paths, and the precise timing of maneuvers all require accurate values of π. NASA's missions, including the Mars rovers and the Voyager probes (see Figure 16, middle), depend on π for their successful journeys through space. The curvature of lenses and mirrors in telescopes and other instruments also uses π to focus light accurately and capture images from distant celestial bodies. In cryptography, π has found a unique application in ensuring secure communications. In fact, the unpredictable nature of π's digits is used to generate pseudo-random numbers, which are critical for encryption algorithms. These algorithms protect sensitive data by making it extremely difficult for unauthorized parties to decode messages without the correct key Kraft and Washington [2018].

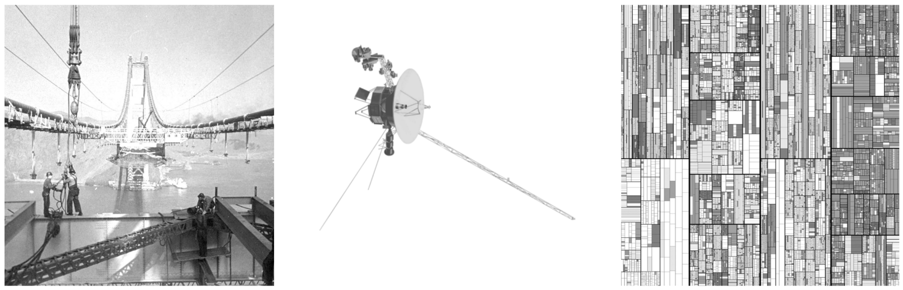


Figure 16: On the left of the figure, a picture of the Golden Gate Bridge in 1930 (The Bancroft Library, *Construction Photographs of the Golden Gate Bridge collection*). Voyager 1 probe is shown in the middle of the figure. On the right, the art "3628 digits of π" from the scientist and artist Martin Krzywinski.

3.4 Roots of polynomial equations

3.4.1 Al-Khwarizmi completing the square

Muhammad ibn Musa al-Khwarizmi was a Persian polymath who lived between (ca. 780 – ca. 850) and came from Khwarazm, hence, his popular name al-Khwarizmi. His influence spreads across different subjects notably in mathematics, astronomy, and geography Wikipedia contributors [c]. His popular work *al-Kitab al-Mukhtasar fi Hisab al-Jabr wal-Muqabalah* (The Compendious Book on Calculation by Completion and Balancing), presented a step-by-step solution to linear and quadratic equations by the method of completing the square. Al-Khwarizmi intended to put together this work hoping that it would provide a simple way to do calculations in matters such as inheritance, lawsuits, trade, and many more dealings between the people of his time that involved calculation or measurement of various kinds of objects Rosen [2009].

> *"When I considered what people generally want in calculating, I found that it always is a number. I also observed that every number is composed of units and that any number may be divided into units."*

When we count from one to ten, each number precedes another number by a unit, then ten is doubled to obtain twenty, and it is tripled to obtain thirty like that up to a hundred, in the same fashion as the units and tens we obtain a thousand, this can be repeated to any large number Rosen [2009].

> *"I observed that numbers which are required in calculating by Completion and Reduction are of three kinds, namely, roots, squares, and simple numbers relative to neither root nor square."*

Al-Khwarizmi identified three kinds of numbers, roots, squares, and simple numbers which are required in his approach (i.e., completing the square). Roots are defined to be any quantity that is to be multiplied by itself, a square is the result when a root is multiplied by itself, and a simple number is any number that is not related to a root or a square. A connection can then be made between the three classes of numbers identified, for instance, a number in one of the classes, say root, could be equal to a number in another class, say square or simple number. Hence, three cases can be established; *"squares are equal to roots"*, *"Squares are equal to numbers"*, and *"roots are equal to numbers."*. The three kinds of numbers can be combined forming three additional cases, that is, *"squares and roots equal to numbers"*, *"squares and numbers equal to roots"* and *"roots and numbers equal to squares"*. Hence, Al-Khwarizmi identified six different cases and demonstrated how to solve for the roots Rosen [2009].

The six different types of equations

$cx^2 = bx; \quad cx^2 = a; \quad bx = a$

$\therefore x = \frac{b}{c}; \quad x = \sqrt{\frac{a}{c}}; \quad x = \frac{a}{b}$ respectively.

For *"squares are equal to roots"*, *"Squares are equal to numbers"*, and *"roots are equal to numbers."* respectively.

$cx^2 + bx = a; \quad cx^2 + a = bx; \quad cx^2 = bx + a$

$x = \sqrt{\left[\left(\frac{b}{2c}\right)^2 + \frac{a}{c}\right]} - \frac{b}{2c}; \qquad x = \frac{b}{2c} \pm \sqrt{\left[\left(\frac{b}{2c}\right)^2 - \frac{a}{c}\right]}$; and

$x = \sqrt{\left[\left(\frac{b}{2c}\right)^2 + \frac{a}{c}\right]} + \frac{b}{2c}$ respectively.

For *"squares and roots equal to numbers"*, *"squares and numbers equal to roots"* and *"roots and numbers equal to squares"* respectively.

Six different cases are explained, the first three do not require halving the number of roots of the square. However, the last three cases do require halving. This can be explained with a figure where the halving can be made clearer.

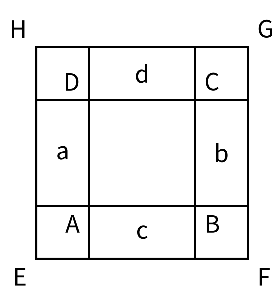


Figure 17: Completing the square EFGH

Demonstration of the case: "a Square and ten Roots are equal to thirty-nine Dirhems"

$$x^2 + 10x = 39 \tag{4}$$

The quadrate ABCD is the square whose root is unknown which we seek to find. Any side of the square ABCD is the root of the square. So, when any of the sides of the square ABCD is multiplied by any number, we obtain the number of the roots which is to be added to the square. In this case, since the number of the roots, which is ten, is combined with the square, we can divide ten by four, we obtain $2\frac{1}{2}$. Each side is combined with $2\frac{1}{2}$, the root of ABCD then becomes the length and $2\frac{1}{2}$ is the width of the rectangles i, j, k and l. However, at the corners of the square ABCD, we are missing smaller square pieces of size $2\frac{1}{2}$ by $2\frac{1}{2}$ and the product is $6\frac{1}{4}$ needed to complete the square EFGH. Therefore, the square EFGH can be completed by adding to it four times $6\frac{1}{4}$ which is 25, i.e the size of smaller pieces of the square. But the quadrate ABCD and ten roots denoted by the rectangles i, j, k, and l equals 39. If we add 25 which is the four square pieces, we obtain the quadrate EFGH which equals 64. We then obtain one side of EFGH which is its root, to be 8. But we are interested in one side of ABCD. Since we know the widths of the rectangles i, j, k, and l, which are the extreme edges of EFGH, we subtract two times the quarter of the ten roots which is 5; the answer is 3 which is one side of the square ABCD, hence the root is 3.

3.4.2 Omar Khayyam cubic equation

Abu'l Fath Omar ibn Ibrahim al-Khayyam, popularly known as Omar Khayyam, is a Persian polymath and scientist born in the town called Nisapur, then capital of the Seljuk empire, who lived between (1048-1131). His contribution to mathematics, astronomy, philosophy, and and poetry is widely recognised.

In modern poetry, Khayyam is popularly known for his "ruba'iyyat"; that is, poems which takes the form of four lines also known as quatrians. In the field of astronomy, he is known to have developed a solar calendar known as the Jalali calendar which became the foundation of a calendar that became known as the Persian calendar Wikipedia contributors [b]. In the field of mathematics he is popularly known for the following works; *Risala fi Sharḥ ma Ashkal min Muṣadarat Kitab Uqlidis* ("Commentary on the Difficulties Concerning the Postulates of Euclid's Elements"), *Risalah fi Qismah Rub' al-Da'irah* ("Treatise On the Division of a Quadrant of a Circle") and *Risalah fi al-Jabr wa'l-Muqabala* ("Treatise on Algebra"). He is also known to have worked on a treatise on binomial theorem, unfortunately this work has been lost. In his "Treatise On the Division of a Quadrant of a Circle" Khayyam postulated;

> *"drop perpendicular from some point on the circumference to one of the radii so that the ratio of the perpendicular to the radius is equal to the the ratio of the two parts of the radius on which the perpendicular falls".* Siadat and Tholen [2021]

Following a specific case led to the equation $x^3 + 200x = 20x^2 + 2000$, for which there is no exact solution, but he provided an approximation Mousavian et al. [2024]. He also presented a general geometric solution to cubic equations, by determining the intersection of two curves (i.e., a circle and a hyperbola).

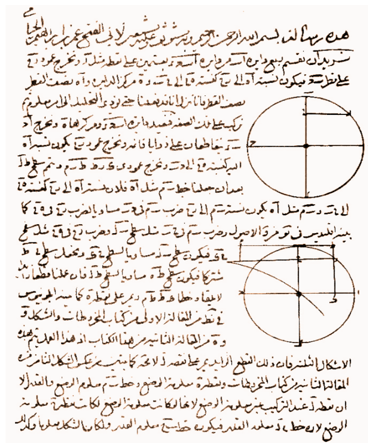


Figure 18: First page of a manuscript kept in Tehran University. Source: Siadat and Tholen [2021]

However, he was not the first to use geometry to provide algebraic solutions, geometric tools have been used to provide solutions to quadratic equations by the Greeks and even the Babylonians. Al-Khawarizmi in the 9^{th} century presented a step-by-step solution to quadratic equations by completing the square, also Thabit ibn Qurra in the late 9^{th} solved quadratic equations using compass geometry of Euclid's elements Mousavian et al. [2024]. Some cubic equations can be solved by taking the cubic root or by first reducing them to a quadratic equation (i.e., dividing the cubic equation by the factor x or x^2). Khayyam identified nineteen types of cubic equations. However, fourteen distinct types of cubic equations cannot be solved by taking the cubic root or reducing it to a quadratic equation.

Fourteen distinct cases

$x^3 = c; \quad x^3 + bx = c; \quad x^3 + c = bx; \quad x^3 + c = ax^2; \quad x^3 + ax^2 = c$

$x^3 = ax^2 + c; \quad x^3 = bx + c; \quad x^3 + ax^2 = bx + c; \quad x^3 + c = ax^2 + bx$

$x^3 = ax^2 + bx + c; \quad x^3 + x^2 + c = bx; \quad x^3 + ax^2 + bx = c; \quad x^3 + bx = ax^2 + c$

$x^3 + bx + c = ax^2$

Four out of the fourteen equations had already been solved by this time, but Khayyam provided general solutions to all the fourteen cases by the use of intersecting conic sections Siadat and Tholen [2021].

Solution for the case where a cube and sides equal to a number

$$x^3 + bx = c \tag{5}$$

Where $b, c > 0$. To solve for x Khayyam identified the intersection of parabola and a semi-circle. Let's go through the solution step-by-step.
Geometric construction of the equation (5), treat all the terms as solid objects; x^3 and ax are 3D with unknown sides x, we assume ax has a square base m^2 and b is rectangular prism with square base m^2 and length n. Now we seek to make x^3 and ax equal to b.

$$\implies x^3 + m^2x = m^2n; \quad \therefore x^3 = m^2n - m^2x \tag{6}$$
$$\implies x^2x = m^2(n - x); \quad \therefore x\colon (n - x) = m^2\colon x^2 \tag{7}$$

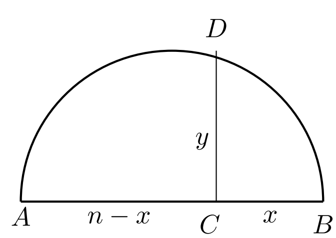
"Theory of Proportions"

In Diagram (a); Let AB denote the diameter of the semi-circle, and CD a perpendicular to the diameter. Then conversely, if $CB\colon CD = CD\colon AC$, then CD is perpendicular to the diameter AB of the semi-circle. Therefore equation (8) is the equation of a circle with diameter n.

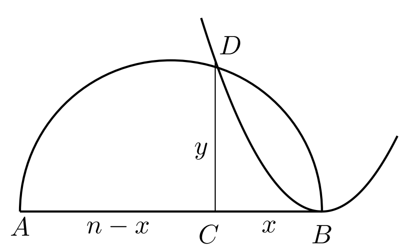
$$\implies x\colon y = y\colon (n - x); \quad \therefore x(n - x) = y^2 \tag{8}$$

$$\implies x\colon (n - x) = y^2\colon (n - x)^2 \tag{9}$$
$$\implies m^2\colon x^2 = y^2\colon (n - x)^2; \quad \therefore m\colon x = y\colon (n - x) \tag{10}$$
$$\implies m\colon x = x\colon y; \quad \therefore my = x^2 \tag{11}$$

The intersection of the equation of the circle (8) and the equation of the parabola (11) gives the solution. It is shown in diagram (b) that the intersection point is D Jeff [2020].



(a)



(b)

4 Laws, Limits, and Randomness: A Mathematical History of Probability

4.1 Historical Context

4.1.1 The Arithmetic Triangle Across Cultures

Pascal's Triangle and Binomial Coefficients

In the 17th century, **Blaise Pascal** studied the properties of an infinite triangular array in which each entry is the sum of the two entries directly above it in the preceding row, as seen in Figure 19. Now known as *Pascal's Triangle*, named after the French mathematician, each row of Pascal's triangle corresponds to the coefficients of the binomial expansion $(a+b)^n$.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

Figure 19: First five rows of Pascal's Triangle.

Each entry in the triangle corresponds to a *binomial coefficient*, which gives the number of ways to choose k elements from a set of n elements, in other words each row corresponds to a value of n in the binomial expansion of $(a+b)^n$, and the k-th entry in the row is the binomial coefficient:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad \text{where } 0 \leq k \leq n$$

where $n!$ denotes the factorial of n.

Although Pascal's name is most commonly associated with the triangle in Western mathematics, the same triangular array had appeared independently in earlier mathematical traditions across India, Persia, and China, centuries before Pascal's formal study.

Early Origins in India: Pingala's Meru-Prastara (300 BCE)

In India 300 BCE, **Pingala**, a famous poet and mathematician, wrote 'Chandahsastra', illustrating the first concept of Pascal's triangle. His book was a technique for linking mathematics with Sanskrit poetry, mapping short and long syllables to binary digits. The study of Sanskrit poetry led to the early ideas of binary number systems, combinatorics, and Fibonacci numbers. Specifically, Pingala's text highlighted the number of different arrangements of short and long syllables, drawing a link to the binomial coefficients and implicitly discovering Pascal's triangle. During the 10th century CE, **Halayudha**, another Indian scholar, proposed that Pingala's final sutra resembles ***'mere-prastara'***, or *'the staircase of mount meru'*, which some argue is the first known evidence of Pascal's triangle.

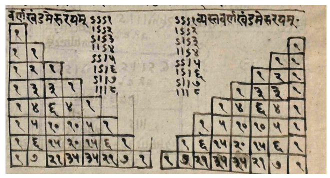


Figure 20: Pingala's *Mere-prastara* 300 BCE

Islamic Golden Age Contributions: Al-Karaji and Omar Khayyam

Omar Khayyam, also a mathematician and poet as well as an astronomer, utilised the binomial theorem to redesign the work of Persian mathematician **Al-Karaji** during 11th century Iran. Al-Karaji laid the foundations for algebra and combinatorics, being one of the first to use binomial coefficients to expand expressions in the form $(a+b)^n$, in a now lost book believed to have been titled *Al-Fakhri*. Developments to the triangular array were made by Khayyam a few decades later, especially in relation to algebraic identities. Their work depicts the well-known triangle that appeared in the Islamic world centuries earlier and how it plays a crucial role in the evolution of algebraic operations and binomial coefficients.

Independent Development in China: Jia Xian to Yang Hui

In ancient China, the triangle was documented by the 11th century mathematician **Jia Xian** who used it to extract square and cube roots to compute binomial coefficients. His texts included a drawn version which was fortunately analyzed by **Yang Hui** in the 13th century before his book became lost. Yang Hui's work *Xiangjie Jiuzhang Suanfa* in 1261, credits Jia Xian and contains a method of obtaining the triangle illustrating how each number is a result of the two above it. In 1303, *Jade Mirror Of The Four Unknown* by **Zhu Shijie** credits his work again, replicating what became known in China as *Yang Hui's Triangle*. These contributions highlight the independent advancements made by Chinese mathematicians and their contributions to the early progressions of algebra and binomial expansions.

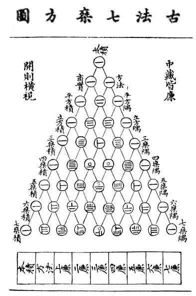


Figure 21: Yanghui triangle ca. 1238–1298.

After appearing independently in several cultures, the triangular array was formally studied by **Blaise Pascal** in his 1665 work, *Treatise on the Arithmetic Triangle*. Drawing on earlier contributions from **Pingala** (India), **Omar Khayyam** (Persia), and **Yang Hui** (China), Pascal popularised the triangle in the Western world, where it became known by his name.

Pascal's Triangle became a fundamental tool in 17$^{\text{th}}$-century mathematics, particularly in the development of probability theory. Its structure provided a systematic way to compute binomial coefficients, enabling early applications of combinatorics to problems involving chance.

One of its first major applications was in solving the ***Problem of Points***–a famous gambling problem discussed in correspondence between Pascal and Fermat. This marked one of the earliest formal uses of combinatorics in probabilistic reasoning.

His work also laid the groundwork for further developments by mathematicians such as **Jacob Bernoulli** and **Abraham De Moivre**, who investigated the behavior of the binomial distribution. Their insights revealed its convergence toward the bell-shaped curve—what we now recognise as the normal distribution—thus setting the stage for the **Central Limit Theorem**.

In this way, the triangle-shaped by many cultures and rediscovered across centuries–served not only as a tool for algebraic expansion, but also as a stepping stone in the mathematical formulation of uncertainty. It ultimately bridged early combinatorics with the birth of modern probability and statistics.

4.1.2 Games of Chance and the Problem of Points

Historical Background of the Problem of Points.

Consider a game between two players, A and B, who agree to compete for a fixed prize pot. The game is played as a sequence of independent rounds– such as flipping a coin – where one player wins each round. The objective is to be the first player to win a fixed number of rounds, say $\mathbf{n}$, in order to claim the entire prize.

Let the probability that player A wins any individual round be p_A, and similarly p_B for player B, where:

$$p_A + p_B = 1.$$

In the classical fair game, both players have equal probability of winning a round:

$$p_A = p_B = \frac{1}{2}.$$

Now suppose the game is interrupted before either player reaches the winning threshold of n rounds. Let the **current scores** be a for A and b for B, with:

$$a < n, \quad b < n.$$

At this point, the central question-known as the **Problem of Points**–is:

How should the prize money be fairly divided between the two players, given the current score and the rules of the game?

This problem was famously solved in the 17th century through the correspondence between Blaise Pascal and Pierre de Fermat, and is considered one of the foundational problems in the development of probability theory.

To resolve the problem fairly, we must consider two core principles: **(1)** what constitutes a fair division of the pot, and **(2)** how likely each player is to win the game if play were to continue from the current score. The *fair share* is then determined by computing the number of all possible future outcomes that would result in victory for each player, weighted by their probabilities of winning each round.

Despite its emerging popularity in the 17th century, the problem of points is a concept that had been studied centuries prior. Italian mathematician **Luca Pacioli** attempted to solve the problem in his 1494 textbook, by dividing the stakes on the basis of the rounds previously won by each player. This result was debunked in the mid-16th century by **Niccolo Tartaglia**. He noticed that if a player were to win the first round, they would obtain the entire prize pot, though a one-point lead in a long game certainly does not imply a winner. He believed that there was no solution where both players would be convinced of the fairness of the outcome.

The problem arose again in 1654 when it was proposed to Pascal by **Chevalier de Mere**. Pascal discussed this problem thoroughly with Fermat in a series of letters, and they came up with different, yet complementary solutions, both of which would lay the foundations of modern probability theory.

Fermat's Method of Enumeration.

Fermat approached the problem by counting all possible future game sequences, introducing **combinatorial analysis**. If Player A **requires** a additional wins and Player B **requires** b, then the game could last at most $a + b - 1$ more rounds. There are 2^{a+b-1} possible outcomes, assuming each round is a fair 50/50 trial.

Fermat created a table of all outcome sequences and identified the number of favorable paths leading to each player's victory. He then split the pot proportionally:

$$\text{Share of player } A = \frac{\text{\# of favorable paths for A}}{2^{a+b-1}}.$$

His solution was counting-based and neglects the idea that these sequences do not each have the *same probability*. However, his method is appreciated when the game favors counts over probabilities or when a fixed number of rounds remain. Despite Fermat's assumption of a uniform probability space and fixed-length pathways, his solution focused on future additional rounds rather than the previous ones, a quality that Pascal carried into his work.

Pascal's Method of Recursion.

Pascal's approach was considered more modern, using *recursive reasoning* based on expected value. It considered all possible pathways, taking the actual probabilities into account, something Fermat did not do. Pascal considered:

> *What if the game were to end after one more round - how would a single round affect the division of the prize pot?*

By examining the probabilities of just the next round, he was able to recursively compute the fair share to each player at any point in the game. He used the *expected values* to calculate the new fair share, for player A, with the **recursive formula**:

$$V(a,b) = \frac{1}{2}V(a-1,b) + \frac{1}{2}V(a,b-1)$$

His solution became easier to compute and more applicable, as in many cases the pathways are not fixed length. Philosophically motivated, Pascal's model focuses on the immediate future to determine what is fair, influencing the calculation of expected payoff in modern probability and game theory.

Through the manipulation of identities involving Pascal's triangle, he further demonstrated the correct **division of stakes** (*Player A : Player B*), is given by:

$$\sum_{k=0}^{b-1} \binom{a+b-1}{k} : \sum_{k=b}^{a+b-1} \binom{a+b-1}{k}$$

The **division of stakes problem** in the 17th century was the first evidence of explicit reasoning for what we now know as the expected value. Pascal's triangle is used as an essential combinatorial tool in calculating the probabilities of the remaining rounds, modeled by the binomial distribution.

It is clear that, although the main motivating force for a solution was made by Pascal, Fermat provided the concept of $a+b-1$ additional rounds. Fermat and Pascal's solution to the problem of points would go on to be considered the foundation of probability theory, the discovery of expected value, and the beginning of game theory. The problem gave early insights into fair division and the mathematical treatment of **uncertainty**, becoming a significant subject in Pascal's work *Treatise on the Arithmetic Triangle*, and shaping the beginning of a systematic approach to probabilistic thinking.

Gambler's Ruin

In the 17th century, games of chance and gambling culture became not only a past time but a grounds for mathematical thinking. One of the most enduring problems in the history of probability is gamblers ruin. Introduced in 1656 in a letter from Pascal to Fermat, the problem states:

> "*Let two men play with three dice … the winner is the first to reach twelve points; what are the relative chances of each player winning?*".

The letter was forwarded onto the Dutch mathematician and scientist **Christiaan Huygens** (1629-1695), who reformulated the problem in his book *On Reasoning in Games of Chance* 1657, the first formal treatise on probability. Through simple coin-toss scenarios, the problem has implications in random walks and finance, capturing the notion that although a game might be fair, risk of ruin is inevitable under certain conditions.

Huygens analysis of expected values revealed that a gambler could lose due to their finite capital and accumulation of variance even in favorable games. Overtime, many versions of the problem surfaced, such as the scenario in which two players start with equal points, transferring them until one is *ruined* by reaching zero points, often applied to a gambler opposing a casino.

De Moivre's Method of Ruin Probabilities

In 1711, De Moivre solved the gambler's ruin problem in a clever and generalisable way by considering the specific nominal value of each point. Let us assume a gambler starts with i points whereas there opponent starts with $N-i$ points. In each round, the gambler will win and gain one point or lose and transfer one point to their opponent, with probabilities p and q respectively, where:

$$q = 1 - p \tag{12}$$

The game ends when one player is *ruined* or gains all N points. Let $P(i)$ denote the probability that the gambler will win the game. The probability that the gambler obtains all N points in a fair or unfair game is:

$$P(i) = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}, & \text{if } p \neq q \\ \\ \dfrac{i}{N}, & \text{if } p = q = 0.5 \end{cases} \qquad \text{for } 0 \leq i \leq N$$

(Note: $P(i)$ is the probability of ruin.)

De Moivre's result echoed Pascal's earlier reasoning, providing identical outcomes when applied to games of chance. By rigorously formalising the problem using recursive methods and probability theory, he laid the foundation for modern stochastic analysis. De Moivre's formulation, as well as the contributions of Pascal and Huygens to the gambler's ruin problem, not only anticipated later developments made by Feller in 1970, but also sparked applications for Markov chains, financial risk modeling, and the theory of random walks.

4.2 The Law of Large Numbers

4.2.1 Early Formulation

Jacob Bernoulli (1655 - 1705) made many contributions to the foundations of mathematics over his life, with his most famous work deriving the first version of the Law of Large Numbers (LLN) in Bernoulli [1713]. His work titled "Ars Conjectandi", published by his nephew, Nicolas Bernoulli, 8 years after his death, is now respected as a greatly significant contribution to the theory of probability. The book contains significant work on expected value and includes much on his theory of permutations and combinations.

The fourth section of this book is credited as being the first version and rigorous proof of the LLN. Once referred to as his 'Golden Theorem', now the Weak Law of Large Numbers or Bernoulli's Theorem, states that the results of observation would approach the theoretical probability as more trials are held. For example, according to the LLN, if you flip a coin multiple times, the proportion of heads will get closer to 0.5 as the number of flips increases. Motivated by a desire to understand the estimation of underlying probabilities, usually in the context of games of chance, Jacob Bernoulli spent 20 years of his life formulating the proof produced in Ars Conjectandi for binary random variables and the LLN.

Bernoulli explains his golden theorem through a theoretical scenario of drawing black and white balls from an urn consecutively, with replacement, in an attempt to estimate the true proportion of black and white balls in the urn. As the amount of balls drawn becomes greater, Bernoulli suggests that the estimation of proportion gets closer to the true ratio of black to white balls. He suggests that a more accurate estimation of the proportion can be achieved through a larger number of trials.

The LLN is applicable to many fields from financial risk assessments to the estimation of probabilities in medical research. Large datasets and simulations are favored in statistical research as they can demonstrate that despite having many fluctuations, a greater number of data points will still result in a relatively reliable average.

4.2.2 The Weak Laws of Large Numbers

Many mathematicians throughout the 19^{th} and 20^{th} centuries deepened Jakob Bernoulli's LLN results, rigorously generalising them and providing modern foundations for probability theory. These developments introduced and clarified fundamental concepts still central to undergraduate probability and measure theory courses today – such as *independence*, *randomness*, and *convergence*.

Poisson's Approximation to the Binomial

In 1837, the French mathematician and physicist **Siméon-Denis Poisson** coined the term **"law of large numbers"** (*la loi des grands nombres*) in his treatise *Recherches sur la probabilité des jugements* [Poisson, 1837]. Mathematically, Poisson worked with a model of independent Bernoulli trials with small probability of success.

Figure 22: Portrait of Siméon-Denis Poisson by E. Marcellot, 1804

Theorem 4.1 (Poisson Limit Theorem, Poisson 1837)**.** *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with success probability p_n, and let $S_n = \sum_{i=1}^{n} X_i \sim$ Binomial(n, p_n). Suppose that*

$$\lim_{n \to \infty} np_n = \lambda > 0.$$

Then, for every $k \in \mathbb{N}_0$, the binomial probability mass function converges to the ***Poisson distribution*** *with parameter λ:*

$$\lim_{n \to \infty} \mathbb{P}(S_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

The proof to this theorem can be found in Chapter 3 of Papoulis [1991].

This is essentially the same result as Bernoulli's LLN, but Poisson was the first to articulate it under the now-standard terminology and with practical framing. This Poisson limit theorem not only served as an approximation to binomial probabilities in practical applications, but also bridged the LLN with the modeling of *rare events* when exact binomial calculations were computationally infeasible. His formulation gave the LLN broader relevance for real-world statistical inference–particularly *when events were individually unlikely but frequent in aggregate*. Poisson was the first mathematician to take LLN from games and theory into real-world applications, along side Belgian-French astronomer, mathematician, statistician and sociologist **Adolphe Quetelet**, such as decision making for jury trials, demographics, and life insurance.

Chebyshev's Generalisation of LLN

A major leap in generality came with the work of the Russian mathematician **Pafnuty Chebyshev**. In 1867, Chebyshev proved a version of the Law of Large Numbers that applied to arbitrary *independent and identically distributed* (i.i.d.) random variables with **finite variance**, moving beyond the Bernoulli/binomial cases of his predecessors. Chebyshev's contribution was not only the generalisation itself, but also the use of what is now known as **Chebyshev's inequality**, given by

$$\mathbb{P}(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \quad \text{for all } k > 0,$$

to bound the probability of large deviations from the mean.

Theorem 4.2 (Chebyshev's Weak Law of Large Numbers, 1867)**.** *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = Var(X_i) < \infty$. Define the sample mean by*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

which implies that

$$\overline{X}_n \xrightarrow{P} \mu \quad \text{as } n \to \infty.$$

The proof of this theorem can be seen in many undergraduate textbooks, but some recommended ones are Grimmett and Welsh [2014], Grinstead and Snell [2012] and Ross [2020].

This result, known as the *Weak Law of Large Numbers*, shows that the sample mean converges to the expected value in probability, under minimal assumptions. Specifically, it requires only that the random variables are *independent, identically distributed*, and possess a *finite variance* – that is, $\mathbb{E}[X_i] = \mu < \infty$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Chebyshev's ideas paved the way for more sophisticated limit theorems in probability, including the Central Limit Theorem, which we will discuss later on, and various strong laws of large numbers developed later by Khinchin, Kolmogorov, and others.

Khinchin's LLN Under Minimal Conditions

A significant advancement in the theory of large numbers came from the Russian mathematician **Aleksandr Khinchin** in 1929, in his *"Ueber das Gesetz der grossen Zahlen"* [Khinchin, 1929]. Building on Chebyshev's earlier work, Khinchin proved that the Weak Law of Large Numbers still holds under a strictly weaker assumption: *it is sufficient for the random variables to have a finite mean – no assumption on the variance is necessary*.

This result greatly broadened the applicability of the LLN to distributions with heavy tails (e.g., Cauchy-type variables truncated to have a finite mean but infinite variance), and it refined our understanding of the minimal conditions under which statistical averages converge.

Theorem 4.3 (Khinchin's Weak Law of Large Numbers, 1929)**.** *Let $X_1, X_2, \ldots$ be independent and i.i.d. with a finite mean $\mu = \mathbb{E}[X_i] < \infty$ (but possibly infinite variance). Define the sample mean by*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then,

$$\overline{X}_n \xrightarrow{P} \mu \quad \text{as } n \to \infty.$$

The proof, which uses truncation techniques and tail bounds instead of Chebyshev's inequality, shows that the convergence in probability of the average of the sample to the expected value requires only the existence of the first moment. The proof of this can be found in many graduate textbooks; see, for example, Billingsley [2013], which discusses weak convergence in detail, and Proschan and Shaw [2018] for background on the various modes of convergence.

Khinchin's theorem thus provides the most general form of the Weak Law of Large Numbers (WLLN) for i.i.d. sequences and forms the theoretical basis for many practical statistical techniques. It is specifically useful in simulation and computational statistics, such as **Monte Carlo integration**, where the sample average $\overline{X}_n$ is used to estimate $\mathbb{E}[X]$. Khinchin's WLLN guarantees that this estimate converges to the true expectation even when the underlying distribution has infinite variance.

In fields like **insurance, finance, and network traffic modeling**, data often exhibit heavy-tailed behavior. Khinchin's WLLN assures practitioners that empirical averages remain valid estimators as long as the mean exists, even if the variance does not. In many cases, this law provides practitioners with the ability to rely on convergence in probability of averages without requiring second-order properties of the data.

4.2.3 Kolmogorov's Strong Law of Large Numbers (1933)

From Borel's Coin Tosses to Cantelli's Lemmas

So far, all versions of the LLN discussed have been "weak" laws, concerning convergence in probability. A stronger mode of convergence is **almost sure (a.s.) convergence**, which guarantees that the sequence of sample averages converges to the expected value with probability one. A result of this form is called a **Strong Law of Large Numbers (SLLN)**. The strong law is strictly stronger than the weak law: *almost sure convergence implies convergence in probability, but not vice versa.*

The first SLLN was proved by **Émile Borel** in 1909 in Émile Borel [1909], for the case of i.i.d. Bernoulli(p) trials. His result showed that with probability one, the relative frequency of successes converges to the probability of success. Borel's approach was number-theoretic and focused on binary expansions of real numbers in $[0, 1]$.

Building on this, **Francesco Cantelli** in 1917, Cantelli [1917], extended the strong law to more *general sequences of random variables under moment conditions*. His proof introduced analytic tools and what would later be formalised as the **Borel–Cantelli lemma**.

Kolmogorov's Axiomatic Breakthrough

The transition from the Weak to the Strong Law of Large Numbers was one of the most important milestones in the formalisation of modern probability theory. The pivotal contribution came from the Russian mathematician **Andrey Kolmogorov** in 1933, whose monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung* laid the axiomatic foundations of probability via measure theory [Kolmogorov, 1933]. Within this framework, Kolmogorov rigorously established a version of the Law of Large Numbers that ensures *almost sure convergence* – that is, convergence of the sample mean for almost every realisation of the sequence of random variables (refer to Proschan and Shaw [2018] for background on the various modes of convergence).

Theorem 4.4 (Kolmogorov's Strong Law of Large Numbers, 1933)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with finite mean $\mu = \mathbb{E}[X_i] < \infty$. Then,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu \quad \text{as } n \to \infty.$$

That is, with probability 1,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu.$$

The proof of this theorem can be found in Section 7.5 of Grimmett and Welsh [2014] and Section 6 of Billingsley [2013]. Some applications of both the WLLN and Strong Law of Large Numbers (SLLN) can be found in Section 7 of Proschan and Shaw [2018].

Kolmogorov's SLLN represents a significant strengthening of the Weak Law: rather than merely guaranteeing that the sample mean converges to the true mean in probability, it asserts that this convergence occurs *almost surely* – that is, for ***almost every sample path***. In other words, deviations from the expected value are not just uncommon; they eventually vanish entirely for nearly all infinite sequences of outcomes.

Developed alongside his groundbreaking ***axiomatic foundation of probability theory***, Kolmogorov's formulation of the SLLN cemented the modern understanding of convergence and formalised probability as a rigorous branch of real analysis. Most remarkably, Kolmogorov proved that for i.i.d. sequences, the SLLN holds *if and only if* the first absolute moment is finite: $\mathbb{E}[|X_1|] < \infty$. This established the minimal condition under which sample averages converge almost surely to the mean-a result that remains a cornerstone of both theoretical probability and its wide-ranging applications.

Summary: From Early Results to Rigorous Foundations

Let $X_1, X_2, \ldots$ be an infinite sequence of identically distributed, Lebesgue integrable random variables. Define the sample mean:

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \quad \text{with } \mathbb{E}[X_1] = \mathbb{E}[X_2] = \cdots = \mu.$$

Weak law of large numbers:

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \Pr\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = 0$$

Strong law of large numbers:

$$\Pr\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1$$

4.3 The Central Limit Theorem

The development of Central Limit Theorem (CLT) was a gradual process, with contributions from many brilliant mathematicians over centuries. The work by Hans Fischer [2010] provides a detailed historical and mathematical explanation of the development of the CLT throughout the years. This section explores the key figures and their contributions to the CLT.

4.3.1 First Glimpse of the Bell Curve: De Moivre–Laplace Theorem

Early mathematicians were fascinated by games of chance; they wanted to predict outcomes and calculate probabilities accurately. One common problem involved binomial distributions, which describe the number of successes in a fixed number of independent trials, n, each with the same probability of success.

For instance, consider flipping a fair coin n times. What is the probability of observing exactly k heads? While exact binomial calculations are feasible for small n, they become increasingly cumbersome for large n. Questions such as *"what is the probability of getting 50 heads out of 100 tosses?"* or *"what happens when we flip a coin a million times?"* demand more efficient tools.

The need for a simpler way to calculate these probabilities drove the work of **Abraham de Moivre** and later **Pierre-Simon Laplace**. They sought a continuous approximation to the discrete binomial distribution that would be easier to work with, especially for large n.

Abraham de Moivre (1667–1754) was from a French family. From 1684 he studied mathematics in Paris, and then left for England at the age of 21, where he lived for the rest of his life, working with gamblers, actuaries, and astronomers (Figure 23.

In his seminal work *The Doctrine of Chances* de Moivre [1738], De Moivre first derived the normal curve as a limit of the binomial distribution. He discovered that as

Figure 23: Portrait of de Moivre by Joseph Highmore, 1736

the number of trials n increases, its shape begins to resemble a smooth, bell-shaped curve. De Moivre's approximation enabled the estimation of binomial probabilities without computing cumbersome factorials.
Building upon this, **Pierre-Simon Laplace** generalised de Moivre's findings in his *Théorie Analytique des Probabilités* Laplace [1812] (Analytical Theory of Probabilities), providing a more rigorous treatment and extending its applicability. This became known as the **De Moivre–Laplace Theorem**.

Theorem 4.5 (De Moivre–Laplace Theorem, 1738)**.** *If k is the number of successes in n independent Bernoulli trials, each with probability of success p, then for large n, the distribution of k can be approximated by a normal distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$:*

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$$

This theorem was revolutionary because it showed that the normal distribution, previously observed in some measurement errors, had a fundamental connection to random events. It provided a powerful tool for approximating binomial probabilities without extensive calculations.

4.3.2 Generating functions (1749–1827)

In the 18th century, the emerging field of modern astronomy was shaped by a growing demand for precision. Laplace, a polymath in mathematics, astronomy, and statistics, aimed to bring analytical rigor to the problem of observational error. He sought a general mathematical framework to quantify and manage uncertainty in scientific measurements.

Laplace's profound contribution to this problem, outlined in his monumental work *Théorie Analytique des Probabilités* Laplace [1812], lay in his sophisticated use of generating functions to analyse the sum of independent errors. While earlier work by de Moivre had shown the normal curve as an approximation for binomial sums, Laplace extended this concept significantly, demonstrating its broader applicability to the aggregation of diverse error sources.

Definition 4.1. *Let X be a discrete random variable taking values in $\mathbb{N}_0$. The* ***probability generating function (pgf)*** *of X is defined as:*

$$G_X(t) = \mathbb{E}[t^X] = \sum_{k=0}^{\infty} \mathbb{P}(X = k)t^k, \quad \text{for } t \in [0, 1].$$

The pgf compactly encodes the entire distribution of X in its Taylor coefficients. In particular, the k-th derivative evaluated at $t = 0$ gives:

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

Laplace recognised that the total error in an astronomical measurement was often the sum of many small, independent errors arising from different sources (e.g., instrument calibration, atmospheric refraction, observer's reaction time). Even if the distribution of each individual error was unknown or complex, he showed that the distribution of their sum would tend towards a normal distribution as the number of error sources increased. This was a profound generalisation of the principles underlying the Central Limit Theorem.

4.3.3 Shaping the Bell Curve: Gauss and the Least Squares Method

In 1801, a small space object named **Ceres** was found, but it quickly disappeared behind the sun. Astronomers had only a few initial measurements of its path, and these measurements were not perfect. The big problem was: *How could they use these few, slightly wrong numbers to figure out where Ceres would go next, so they could find it again?*

Carl Friedrich Gauss saw this important problem and wanted to find a proper way to get the most likely true answer from measurements that had errors. Gauss came up with an answer in 1809. In his work *Theoria Motus Corporum Coelestium*, Gauss fisrt proposed the **Method of Least Squares**, a mathematical approach to estimating unknown quantities from imperfect measurements. The key idea is to choose the values of the parameters, θ, that minimise the sum of squared deviations between the observed data and the values predicted by a model:

$$\min_{\theta} \sum_{i=1}^{n} (y_i - f(x_i; \theta))^2.$$

In other words, suppose that you have many dots on a graph and you want to draw a line that best fits them. "Least squares" tells you to draw the line that makes the total of all the vertical distances from the dots to the line, when each distance is squared, as tiny as possible. This way, bigger mistakes count more, which helps find a better fit.

Gauss didn't stop at proposing a fitting method; he sought to understand the *probabilistic implications* of the Least Squares. He then asked: *if Least Squares is the best way to find answers, what does that tell us about the errors themselves?* He showed that if measurement errors are *independent, symmetrically distributed*, and *small errors are more likely than large ones*, then the only error distribution that makes the least squares estimator $\hat{\theta}$ ($\approx \theta$) the *most likely* (i.e., the *maximum likelihood estimator*) is the **normal distribution**:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

This connection helped solidify the **normal distribution** as a universal model for measurement error and variability in the natural sciences. Combined with Laplace's work on error aggregation, Gauss's contribution helped shape the statistical foundation of the **Central Limit Theorem** and inferential statistics.

4.3.4 Unveiling the Universal Bell Curve: Central Limit Theory

By the late 19th and early 20th centuries, mathematicians knew the normal distribution showed up repeatedly. But their proofs often relied on specific conditions like dealing with coin flips, or assuming errors were perfectly independent and identical. The big question became: *Can we prove the Central Limit Theorem holds true for any collection of independent random events?* The rigour of mathematics drove these mathematicians to explore and precisely *define the minimum conditions* for the "Central Limit Theorem".

Lyapunov's First Big Step (1901)

Alexander Lyapunov, a Russian mathematician, took a major leap in 1901. He gave one of the first proofs for the Central Limit Theorem that worked for a much wider range of random events. He provided one of the first general conditions under which the sum of *independent but not necessarily identically distributed* random variables converges to the normal distribution.

Theorem 4.6 (Lyapunov's Central Limit Theorem, 1901)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with finite means $\mu_i = \mathbb{E}[X_i]$ and finite variances $\sigma_i^2 = \mathbb{V}[X_i]$. Define the total variance:*

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2.$$

Assume that $s_n^2 \to \infty$ as $n \to \infty$, and for some $\delta > 0$, the Lyapunov condition holds:

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0.$$

Then the normalised sum

$$Z_n = \frac{\sum_{i=1}^{n}(X_i - \mu_i)}{s_n}$$

converges in distribution to the standard normal distribution:

$$Z_n \xrightarrow{d} \mathcal{N}(0,1).$$

Lyapunov's condition essentially controls the **size of the "tails"** of the distributions, ensuring that no single variable is too erratic or has extreme influence on the total sum. His theorem was a major generalisation, expanding the reach of the CLT to **sums of unequal, heteroscedastic random variables**. It paved the way for further developments such as the Lindeberg–Feller theorem, which refined and weakened these assumptions even further.

Lindeberg and Feller: Pinpointing the Exact Conditions (1922-1935)

While Lyapunov had already laid a strong foundation for the Central Limit Theorem (CLT), it was **Jarl Waldemar Lindeberg** in 1922, and later **William Feller** in 1935 pushed this further. They discovered conditions that were not only *sufficient* but also *necessary* for the CLT to hold.

The central innovation is the **Lindeberg condition**, a precise way to rule out the possibility that a few extreme values dominate the sum. It ensures that no individual summand has a disproportionately large variance relative to the whole sum.

Formally, for every small positive number $\epsilon > 0$, as n gets very large, the following expression must get closer and closer to zero:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E[(X_i - \mu_i)^2 \mathbf{1}_{\{|X_i-\mu_i|>\epsilon s_n\}}] = 0$$

where $s_n^2 = \sum_{i=1}^n \sigma_i^2$ (the total spread of the sum), and $\mathbf{1}_{\{A\}}$ is the indicator function that takes the value 1 if event A is true, and 0 if false. This condition guarantees that the contribution of large deviations vanishes as $n \to \infty$, so the total sum is well-behaved and converges to the normal distribution.

The Feller's theorem is as follows.

Theorem 4.7 (Lindeberg–Feller Central Limit Theorem, 1922–1935)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with finite means $\mu_i = \mathbb{E}[X_i]$ and finite variances $\sigma_i^2 = \mathbb{V}[X_i]$. Define the total variance:*

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2.$$

Assume that $s_n^2 \to \infty$ as $n \to \infty$.
Then the normalised sum

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n}$$

converges in distribution to the standard normal distribution,

$$Z_n \xrightarrow{d} \mathcal{N}(0,1),$$

if and only if *the Lindeberg condition holds: for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu_i)^2 \cdot \mathbf{1}_{\{|X_i-\mu_i|>\epsilon s_n\}}\right] = 0.$$

These rigorous conditions completed the classical theory of the CLT. The result explains why the normal distribution appears so ubiquitously: from physical measurement errors to stock price fluctuations, average behaviours converge to the bell curve, but only under the right assumptions.

Comparing Lyapunov and Lindeberg–Feller Theorems

- **Lyapunov's CLT (1901)** is a powerful and elegant sufficient condition: it works when the random variables are *independent and have uniformly bounded moments* beyond second order (i.e., there exists a $\delta > 0$ such that the $(2 + \delta)$-th moment is finite). It is relatively easy to verify, but is not necessary.
- **Lindeberg–Feller CLT (1922–1935)** provides a *necessary and sufficient* condition for convergence to the normal distribution. It uses the Lindeberg condition – which precisely controls how much "weight" can be in the tails.

Use case: Lyapunov's theorem is easier to apply when moment conditions are known, but Lindeberg-Feller is the ultimate test for convergence: if it fails, the CLT fails. Proofs for both theorems can be found in graduate textbooks, some of which are Billingsley [2013] and Feller [1971].

4.4 Applications

The Limit Theorems play a *central* role in modern mathematical developments. While its classical history is associated with European mathematicians, as noted in the previous section, the contributions and applications from the late 20th century to the present day also include developments by underrepresented groups. In this section, we highlight the impact of the LLN and CLT through various applied fields, and the people behind these results.

Additional code examples of the applications are available on the project's GitHub repository.

4.4.1 The St. Petersburg Paradox (1713–1738)

The **St. Petersburg paradox** is a famous historical problem in probability theory, introduced in 1713 by **Nicolas Bernoulli** and later discussed in depth by his cousin **Daniel Bernoulli** in 1738 in his work *Exposition of a New Theory on the Measurement of Risk* Bernoulli [1738].

The game is defined as follows:

- A player pays an *entry fee* to play a coin-tossing game.
- A fair coin is tossed repeatedly until a head appears.
- If the first head appears on the n-th toss, the player receives a payout of 2^n pounds.

Thus, the random variable X representing the payout satisfies:

$$\mathbb{P}(X = 2^n) = \left(\frac{1}{2}\right)^n, \quad n = 1, 2, 3, \ldots$$

The expected value, the expected payoff, of the game is:

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} 2^n \cdot \left(\frac{1}{2}\right)^n = \sum_{n=1}^{\infty} 1 = \infty$$

The Paradox:

> *What is the fair amount of money that the casino could charge a player to enter the game?*

Most rational individuals would only be willing to pay a modest fee (e.g., £5 or £10) – certainly not an infinite amount. But mathematically, the player should be willing to pay any finite amount to play the game since the expected returns from the game are unlimited. This highlights a disconnect between theoretical expectation and practical decision-making.

Resolution: Bernoulli argued that the *value* of money is not linear. For example, gaining £100 has more significance for a poor individual than for a millionaire. This observation led him to propose a concave **utility function** that models diminishing marginal utility:

$$U(w) = \log(w),$$

where w is a person's wealth.

Suppose a player has initial wealth W and pays an entry fee c to play. If the player wins 2^n, their final wealth is:

$$W' = W - c + 2^n.$$

Assuming the game is free to play ($c = 0$), the utility gain is:

$$U = \log(W + 2^n) - \log(W).$$

The expected utility gain becomes:

$$\Delta\mathbb{E}[U] = \sum_{n=1}^{\infty} \frac{1}{2^n} \left[\log(W + 2^n) - \log(W)\right].$$

Although 2^n grows exponentially, the logarithmic function grows slowly. This keeps the utility increments small as n increases. In fact, for small n, we can approximate the difference:

$$\log(W + 2^n) - \log(W) = \log\left(1 + \frac{2^n}{W}\right) \approx \frac{2^n}{W} \quad \text{for small } n,$$

but this approximation breaks down for large n, and the true utility growth remains bounded. It turns out:

$$\Delta\mathbb{E}[U] \sim \sum_{n=1}^{\infty} \frac{n\log(2)}{2^n} = 2\log(2),$$

a finite value. Bernoulli's formulation resolves the paradox by showing that while the *expected monetary payout* is infinite, the *expected utility* remains finite.

Connection to the LLN and CLT: The St. Petersburg game serves as an early example of a situation where the classical limit theorems of probability theory do not apply:

- The **LLN** requires at least a finite first moment ($\mathbb{E}[X] < \infty$) to guarantee convergence of averages.
- The **CLT** requires a finite second moment ($\mathbb{E}[X^2] < \infty$) to ensure convergence in distribution to the normal law.

In the case of the St. Petersburg game, the extreme values dominate, leading to a **heavy-tailed distribution** that violates the regularity assumptions required by these limit theorems. The paradox serves as an early historical example that motivated the development of more robust theories of risk and utility, well before the formalisation of probability convergence.

4.4.2 Brownian Motion as a Scaling Limit of Random Walks

Brownian motion is a central object in probability theory and mathematical physics. It is a continuous-time stochastic process that models the random movement of particles in a fluid, describing any quantity that undergoes constant, small, random fluctuations. Brownian motion arises as the scaling limit of symmetric random walks, a direct consequence of the CLT. A random walk is a stochastic process described by the successive sum of independent, identically distributed random variables. More formally:

Definition 4.2 (Random walk)**.** *Suppose that $X_1, X_2, \ldots$ is a sequence of $\mathbb{R}^d$-valued i.i.d random variables. A random walk started at $S_0 \in \mathbb{R}^d$ is the sequence $(S_n)_{n\geq 0}$ given by*

$$S_n = S_0 + X_1 + X_2 + \ldots + X_n, \qquad n \geq 1.$$

The quantities (X_n) are the steps of the random walk.

Persi Diaconis, a mathematician of Sephardic Jewish descent, has extensively studied the probabilistic structures of random processes, including convergence to Brownian motion. **Jean-Pierre Kahane**, a French-Jewish mathematician, contributed to harmonic analysis and probabilistic methods on groups, including the behaviour of random walks. Their work reflects how diverse backgrounds have contributed to our deep understanding of the LLN and CLT's manifestation in continuous-time processes like Brownian motion.

The **random walk** is a simple example that beautifully demonstrates both the LLN and the CLT. As we increase the number of steps, the average position stabilises

(LLN), the fluctuations form a bell curve (CLT), and with proper scaling, the walk turns into Brownian motion - a continuous limit of a discrete process:

1. **Random Walks and the LLN.**
   Consider a simple symmetric random walk:

   $$S_n = \sum_{i=1}^{n} X_i, \quad X_i \in \{-1, +1\}, \quad \mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$$

   By the LLN:

   $$\frac{S_n}{n} \xrightarrow{a.s.} 0$$

   The average *position* converges almost surely to zero.

2. **Random Walks and the CLT.**
   By the CLT:

   $$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

   While the average *position* converges to 0, the fluctuations grow like $\sqrt{n}$, forming a bell-shaped distribution.

3. **From Random Walk to Brownian Motion.**
   Define a rescaled process:

   $$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i,$$

   where the floor function $\lfloor nt \rfloor$ gives the greatest integer less than or equal to nt. Then, as $n \to \infty$, $W_n(t) \xrightarrow{d} B(t)$, where $B(t)$ is standard Brownian motion. This convergence is known as **Donsker's Theorem**, a functional version of the CLT.

4.4.3 Markov Chains

A "**Markov chain**" or Markov process is a discrete-time stochastic process on a state space that satisfies the Markov property [Geyer, 1998]. A **stochastic process** is a sequence $X_1, X_2, \ldots$ of random elements of a state space (a state space is a fixed set where the random variables take values). In simple terms, the **Markov property** means that the future is independent of the past given the present state. Hence,

> *A Markov chain is a discrete-time stochastic process $X_1, X_2, \ldots$ that takes values in a state space, and has the property that the conditional distribution of X_{n+1} given the past, $X_1, \ldots, X_n$ depends only on the present state X_n.*

There are two main components of a Markov chain: the initial distribution and the transition probabilities. The former is the marginal distribution of X_1. Meanwhile, the latter refers to the conditional distribution of X_{n+1} given X_n. It is common to assume that the Markov chain has stationary transition probabilities, meaning that there is only one conditional distribution since they are the same for all n. Furthermore, a Markov chain is *stationary* if the joint distribution of $(X_n, X_{n+1}, \ldots, X_{n+1})$ does not change when translated over time, that is, it does not depend on n for each fixed k. Note that a Markov chain having stationary transition probabilities is not necessarily stationary.

Andrey Markov (1856-1922) was a Russian mathematician known for his developments on the theory of stochastic processes, particularly the aforementioned Markov chains. Markov is even more famous for removing the **independence assumption** in the LLN. In 1906, motivated by a public dispute with the Moscow mathematician **Pavel Nekrasov**, Markov introduced what we now call **Markov chains** to show that statistical regularity does not require independent trials. Nekrasov had claimed that independence was necessary for the LLN - a claim Markov viewed as erroneous, especially since Nekrasov was linking it to philosophical arguments about free will.

To refute this, Markov considered a sequence of dependent trials where the probability of each outcome depends on the previous outcome (the simplest Markov chain) and proved that even here, the long-run frequency of states converges to a constant distribution. In other words, he *proved a weak law of large numbers without independence*, establishing what we now recognise as an ergodic theorem for Markov chains, an astonosing achievement, as for over 200 years we had relied on the assumption of independence of trials. This result was groundbreaking: it provided a connection from the LLN theory to the realm of **stochastic processes** and demonstrated that "mixing" (as in a Markov chain with a *unique stationary distribution*) can substitute independence in producing stable averages. In summary. Markov's work laid the groundwork for **ergodic theory**; he even proved a CLT for Markov chains a bit later.

The Central Limit Theorem is for **Harris ergodic Markov chains**; chains that are irreducible and recurrent. That is, the chain can reach any part of the state space and will return to a small subset infinitely often, with probability 1. Let $X = X_0, X_1, X_2, \ldots$ be a Harris ergodic Markov chain with invariant probability distribution π and support $\mathcal{X}$. Let f be a function and define $\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$. In a general sense, the CLT states that

$$\sqrt{n}(\overline{\mu}_n - E_\pi f) \xrightarrow{d} \mathcal{N}(0, \sigma_f^2)$$

as $n \to \infty$, where $\sigma_f^2 = \mathbb{V}\left[f(X_0)\right] + 2\sum_{i=1}^{\infty} cov\left[f(X_0), f(X_i)\right] < \infty$.

A fun fact is that Markov initially used the idea of Markov chains to analyse letter patterns in **Alexander Pushkin**'s poetry *"Eugene Onegin"*, not in finance or physics. The idea was that the probability of the next letter in a word depends on the previous letter. This idea has fuelled the rise of Search Engines like Yahoo and Google, as well as **Natural Language Processing (NLP) algorithms** in modern day, and more specifically has given a foundation for the Neural Network architecture that models current GPTs, like ChatGPT and Gemini.

4.4.4 Monte Carlo integration

A common class of numerical problems in statistical inference is integration problems, where an exact analytical solution is either difficult or impossible to attain. Hence, we can rely on numerical solutions, such as simulations, to calculate the quantities of interest. Monte Carlo is a popular choice as one of those methods, used for the generic problem of evaluating the integral

$$\mathbb{E}_\pi[f(X)] = \int_{\mathcal{X}} f(x)\pi(x)dx, \tag{13}$$

where $\mathcal{X}$ denotes the support of the random variable X (the set where X takes its values), and in most cases it equals the support of the density function π.

The name "Monte Carlo" originates from the famous Monte Carlo Casino in Monaco, reflecting the method's reliance on randomness and chance, similar to gambling – since the technique uses random sampling to estimate deterministic quantities.

To approximate μ, we draw a sample $(X_1, \dots, X_n)$ independently from the density π and use as an approximation to the integral (13) the sample average

$$\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

By the **Strong Law of Large Numbers**, we have

$$\overline{\mu}_n \xrightarrow{\text{a.s.}} \mu = \mathbb{E}_\pi[f(X)] \quad \text{as } n \to \infty,$$

which ensures that the Monte Carlo estimate converges almost surely to the true expected value μ as the sample size increases.

The variance of the approximation can also be estimated from the sample by

$$\sigma_n^2 = \frac{1}{n^2}\sum_{i=1}^{n}[f(X_i) - \overline{\mu}_n]^2.$$

Let $\sigma^2 = \mathbb{V}_\pi(f(X)) < \infty$. Then **Central Limit Theorem** tells us that, for large n,

$$\sqrt{n}(\overline{\mu}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

or equivalently,

$$\frac{\overline{\mu}_n - \mu}{\hat{\sigma}_n/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

This result allows us to construct confidence intervals around the estimate $\overline{\mu}_n$ and quantify the uncertainty of our approximation. Robert and Casella [2019] is a friendly reference for an introduction to Monte Carlo methods and their implementation in the programming language R.

A special case of the general LLN and CLT for Markov chains, which became crucial in the mid-20th century for solving complex numerical problems, is the class of *Markov Chain Monte Carlo (MCMC)* methods. These algorithms are designed to draw samples from complex probability distributions by constructing a Markov chain whose *stationary distribution* matches the *target*. One of the most well-known and widely used examples is the ***Metropolis–Hastings algorithm (MH)***, which laid the groundwork for modern **Bayesian inference** and **probabilistic machine learning**.

Although the Metropolis algorithm was named after **Nicholas Metropolis**, the original 1953 paper Equation of State Calculations by Fast Computing Machines lists **Arianna Wright Rosenbluth**, a physicist and computer scientist, as the primary author of the algorithm's implementation. Rosenbluth programmed the method for one of the earliest digital computers, MANIAC I, and was instrumental in making the method computationally viable.

Another important figure was **Augusta Maria "Mici" Teller**, a Jewish Hungarian-American scientist who also contributed to the development of early computing and the application of statistical physics to numerical simulations.

4.4.5 Some more advanced examples

Bootstrap Resampling

Bootstrapping is a statistical technique that estimates the distribution of an estimator by resampling the data (with replacement). It allows estimation of the sampling distribution of the statistic using random sampling methods, without requiring strong parametric assumptions. Bootstrap simulates the sampling distribution empirically, as opposed to the CLT which uses theoretical properties of the population distribution. When the quantity of interest is the sample mean, the CLT can directly be used to show the consistency of bootstrapping for estimating its distribution.

Lévy's Introduction to Heavy Tails and α-Stable Distributions

The classical CLT assumes finite variance, but many applications, such as finance and physics, exhibit heavy-tailed behaviour. Lévy's work around 1920-1940 on stable distributions extended the CLT to allow for infinite variance, giving rise to the α-stable distributions.

Gopinath Kallianpur (1925–2015) was an Indian-American mathematician and statistician known primarily for his work in stochastic processes, stochastic differential equations (SDEs), and stochastic filtering [Institute of Mathematical Statistics, 2015]. His contributions were crucial in extending the CLT to infinite-dimensional and functional settings, especially within stochastic analysis. For example, in Kallianpur and Mitoma [1992] they prove a central limit theorem for an interacting system of spatially extended neurons. Kallianpur's work allowed generalisations of the CLT in modern probability, particularly in Functional Central Limit Theorems, Gaussian measures on infinite-dimensional spaces, CLT-based asymptotic analysis in filtering, and applications in stochastic PDEs and signal processing.

4.4.6 Conclusions

The Law of Large Numbers and Central Limit Theorem – two fundamental limit theorems – have shaped modern mathematics and its applications. Highlighting the contributions of underrepresented mathematicians gives overdue recognition to those who advanced these theories, applied them in new domains, and promoted equity in mathematical education and research.

5 Algorithms for Solving Algebraic Equations

Related topics: Number Theory, Algebra

5.1 The history of division

Division has been around in many forms, and tracing its history can be long and varied. Interestingly, some of the widespread ideas behind division came from best rational approximations, hinting at the awareness of irrationality well before the Pythagoras school. The history in Egypt details this as they had impossibly long division algorithms for approximations to things they couldn't measure Lumpkin [2002]. Moreover, the Egyptians were clearly aware of zero as a value, by the *nfr* symbol, meaning complete or beautiful. This concept has been around in Indian and Mayan culture also [Joseph, 2010, p. 15], but not as a numeral or object with properties in its own right.

Even more so, the introduction of algorithmic thinking behind division led towards the general ideas of greatest common divisors and the associated domains. The introduction of zero by Islamic mathematicians presented a change in perspective to the application of such structures as well as identity elements in other algebraic structures Baki [1992]. It is said that the influence of the Islamic scholars spread to much of China, giving more familiar number systems to which we use today. There was also a massive, and often underappreciated, contribution to the heliocentric model of our solar system formalised by Copernicus. Specifically, he used theorems of Nasir al-Din al-Tusi and Muayyad al-Din al-Urdi in the mid 1200s (cf. De Revolutionibus and Blåsjö [2014]).

In particular, Indian mathematics, as presented in *The Crest of the Peacock*, is portrayed as a rich and original tradition deeply rooted in astronomy, commerce, ritual practices, and education. Rather than being deductive and axiomatic like the Greek tradition, Indian mathematics was algorithmic, procedural, and often utilitarian in nature.

Indian mathematics evolved through several distinct periods and was well-recorded:

- **The Vedic Period** (c. 1500–500 BCE): Early mathematical ideas appear in ritual texts known as the *Sulbasutras*, where geometry was used for altar constructions and approximations of quantities such as square roots and π were computed.

- **The Jaina and Buddhist Periods**: Mathematicians of the Jaina tradition explored very large numbers, permutations, and combinations. These works show a deep interest in numerical operations and arithmetic logic.

- **The Classical Period** (c. 5th century CE onward): Marked by figures such as *Aryabhata*, who introduced place-value decimal notation, and contributed to algebra and astronomy using systematic and rule-based methods.

A central feature of Indian mathematics, as mentioned above, is its preference for algorithmic and procedural problem-solving methods. Instead of developing mathematics from axioms as in Euclid's *Elements*, Indian mathematicians used step-by-step techniques to compute results. This is evident in several areas:

- Arithmetic operations such as multiplication, square roots, and cube roots were taught using tabulated steps and verbal algorithms.
- Algebraic equations, including indeterminate equations (of the form $ax+by = c$), were solved using recursive techniques. One such technique – the **Kuttaka** method, which we will discuss in Section 5.3.4 – functions similarly to the Euclidean algorithm for computing greatest common divisors (GCDs).
- The *Sulbasutras* contain procedures for constructing right angles and transforming geometric figures, reflecting a numeric and constructionist approach to geometry.

Indian mathematics did not remain isolated. The computational methods developed in India had far-reaching influence:

> "*In the field of mathematics the Islamic world brought together... the remarkable instrument of computation (our number system) that originated in India... These strands were supplemented by a systematic and consistent language of calculation that came to be known by its Arabic name, algebra.*" [Joseph, 2010, p.7]

5.2 Euclid's Algorithm

5.2.1 The History of Calendars and Astronomy

Ancient astronomers faced a timeless challenge: how to align the **solar year**, the **lunar month**, and **planetary cycles**, which do not neatly fit together. For example:

- 1 Solar year $\approx$ 365.2422 days
- 1 Lunar month $\approx$ 29.5306 days

These do not divide evenly, so:

How many lunar months fit in a solar year?

Mathematicians used the **GCD** and **Euclid's algorithm** to find the best rational approximations of such irrational ratios. The Euclidean algorithm finds the GCD of two numbers and, by extension, helps compute the **least common multiple (LCM)** – essential in synchronizing cycles. This principle underlies the famous:

Metonic cycle: 19 solar years ≈ 235 lunar months

Used in the Greek, Babylonian, and Hebrew calendars, this equivalence was found using GCD reasoning to identify common multiples.

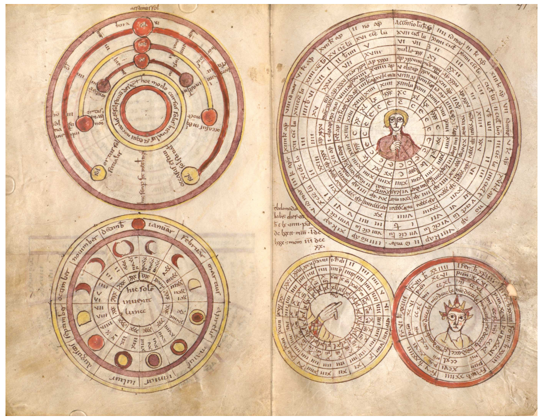


Figure 24: Depiction of the 19 years of the Metonic cycle as a wheel, with the Julian date of the Easter New Moon, from a 9th-century computistic manuscript made in St. Emmeram's Abbey (Clm 14456, fol. 71r)

Indian astronomers like Aryabhata used algorithmic methods akin to Euclid's to solve cyclic equations and align calendars. These methods included the Kuttaka algorithm, which was used to solve equations of the form $ax + by = c$, particularly in astronomical computations such as determining planetary positions and eclipse timings. The Indian use of such algorithms predates the Latin translations of Euclid and was integral to the accuracy of sidereal and lunar calendar systems.

Chinese mathematicians, as discussed in the *Nine Chapters on the Mathematical Art* Dauben [2013], solved calendar and modular congruence problems with methods related to the GCD and remainders. One notable contributor was **Liu Hui** (3rd century CE), who provided detailed commentaries and algorithms resembling the Euclidean algorithm to solve linear congruences and reduce fractions. Later, in the 3rd–5th centuries CE, **Sun Zi** introduced an early version of the Chinese Remainder Theorem – a concept deeply connected to modular arithmetic and Euclid's algorithm – which we will explore in Section 5.3. These tools were crucial in astronomical and calendrical computations.

Islamic scholars translated and refined Greek and Indian works, applying GCD concepts in astronomical tables and eclipse prediction. Notable figures include **Al-Khwarizmi** (9th century), who synthesised Indian arithmetic and Greek geometry in his astronomical tables, and **Al-Biruni** (11th century), who applied number-theoretic methods to lunar calculations and timekeeping. The use of algorithms akin to the Euclidean method was essential in synchronizing lunar and solar calendars, computing least common multiples and periods of planetary orbits, and solving indeterminate equations. Their translation work—especially from the Sanskrit and Greek texts—laid the foundation for later developments in Europe.

> "*Mathematics and astronomy were intertwined in many ancient traditions, with indeterminate analysis and GCD algorithms applied to predict eclipses, align calendars, and track planetary conjunctions.*" Joseph [2010]

All these contributions have helped the development of modern systems, such as:

- Timekeeping (and for example calculating leap-year rules), Harris and Reingold [2004],
- Musical scale systems, Von Zur Gathen and Gerhard [2003],
- Encryption and modular arithmetic in computing, European Information Technologies Certification Academy.

5.2.2 Greatest Common Divisors

The **greatest common divisor (GCD)** of two integers is the largest number that divides both of them exactly.

Example:

$$\gcd(48, 18) = 6$$

because 6 is the largest number that divides both 48 and 18 with no remainder.

Over 2,300 years ago, **Euclid of Alexandria** presented a method to compute the GCD in his monumental text *Elements*. This is now known as the **Euclidean algorithm**, one of the earliest recorded algorithms in history.

The GCD plays a role in:

- Simplifying ratios (e.g., reducing 100:80 to 5:4),
- Solving Diophantine equations (equations with integer solutions),
- Modular arithmetic (used in clocks, calendars, and computing).

The Euclidean algorithm relies on a simple rule: replacing a number by the remainder from dividing it by the other does not change the GCD. In mathematical notation, we represent this rule as follows:

$$\gcd(a, b) = \gcd(b, a \bmod b)$$

where $a \bmod b$ represents the **remainder** when we divide a by b. For example,

$$\gcd(48, 18) = \gcd(18, 12)$$

since dividing 48 by 18 has a remainder of 12. This allows us to reduce the problem step by step until we reach a remainder of zero. Though Euclid formalised the method, the idea of common divisors appeared in other ancient cultures, including the Babylonians and Indians.

5.2.3 Mathematical formulation of Euclid's algorithm

Given two natural numbers a and b (with $a \geq b$), the algorithm proceeds as follows:

1. If $b = 0$, then $\gcd(a, b) = a$. Stop.
2. Otherwise, replace $a \leftarrow b$, $b \leftarrow a \mod b$, and repeat step 1.

This process terminates when the remainder becomes zero, and the last non-zero remainder is the GCD of a and b. That is, if we let $a_0 = a$, $a_1 = b$, and define recursively:

$$a_{n+1} = a_{n-1} \mod a_n \quad \text{for } n \geq 1$$

The sequence $a_0, a_1, a_2, \ldots$ continues until $a_{k+1} = 0$. Then:

$$\gcd(a, b) = a_k$$

For example, we can find the $\gcd(252, 105)$ using repeated division:

$$\begin{aligned} 252 &= 2 \times 105 + 42 \\ 105 &= 2 \times 42 + 21 \\ 42 &= 2 \times 21 + 0 \end{aligned}$$

This gives us a sequence $252, 105, 42, 21, 0$ and the last non-zero remainder is $\boxed{21}$, so

$$\gcd(252, 105) = 21$$

The Euclidean algorithm is very easy to implement in many coding languages. In Python, for instance, the remainder when dividing a by b is written `a%b`, so the Euclidean algorithm can be represented by any of the following recursively defined functions. In each case, we terminate the recursion as soon as our remainder is 0.

```
def euclid1(a,b):
    if b>0:
        return euclid_algorithm(b, a%b)
    else:
        return a

def euclid2(a,b):
    while b>0:
        a, b = b, a%b
    return a

euclid3 = lambda a, b: euclid3(b, a%b) if b>0 else a
```

Figure 25: Portrait of al-Khwārizmī as interpreted by Anreas Strick.

5.2.4 Connecting Euclid to the Islamic Golden Age

During the Islamic Golden Age (8th–13th centuries), the renowned mathematician **Muhammad ibn Mūsā al-Khwārizmī** made foundational contributions to algebra and arithmetic. His role was crucial not only in preserving Greek mathematical knowledge, including that of Euclid, but also in transforming and extending it.

Al-Khwarizmi's most influential works include:

- *Al-Kitāb al-mukhtaṣar fī ḥisāb al-jabr wa'l-muqābala* (The Compendious Book on Calculation by Completion and Balancing), which formalised algebra as an independent discipline and introduced "completing the square" as a method to solve quadratic equations.
- *Algorithmi de numero Indorum*, a Latin translation of his work on Hindu-Arabic numerals, which introduced positional decimal arithmetic to Europe.

Though Al-Khwarizmi did not explicitly modify Euclid's algorithm, his work exemplifies a broader shift toward algorithmic and procedural thinking. The very term *algorithm* derives from the Latinised form of his name (*Algorithmi*).

His algebraic techniques—although expressed rhetorically—paved the way for a more general understanding of problem-solving methods, within which Euclid's algorithm fits as a fundamental archetype. Al-Khwarizmi's emphasis on method and process resonates with the algorithmic spirit of Euclid's original construction.

According to George Gheverghese Joseph's *The Crest of the Peackcock* Joseph [2010], Islamic scholars were not merely passive transmitters of Greek mathematical knowledge. Instead, they acted as active *synthesisers* of Greek, Indian, and Babylonian traditions. The translation and study of Euclid's *Elements* into Arabic, especially during the Abbasid Caliphate under patrons like al-Ma'mūn, allowed for a renewed focus on rigorous method and structure.

This environment enabled a reinterpretation and generalization of classical Greek methods, including Euclidean algorithms, within a new algebraic framework. This recontextualisation would later influence European mathematics through Latin translations in places such as Toledo and Sicily.

The intellectual connection between Euclid and Al-Khwarizmi illustrates a broader narrative: the continuity and transformation of mathematical knowledge across cultures. While Al-Khwarizmi may not have directly altered Euclid's algorithm, his work laid the foundations for algorithmic reasoning and algebraic formalism, setting the stage for the mathematical revolutions of the Renaissance and beyond.

5.2.5 Modern Generalisations

While we are familiar with the notion of GCD's in the integers, in modern mathematics we often want to generalise these concepts when working with more abstract algebraic structures. In doing so, it is common to develop counterparts of Euclid's algorithm.

The first of these one might come across in their education is when working with polynomials and the polynomial division algorithm, which gives us a method to divide one polynomial by another (of smaller degree) and calculate the remainder (of even smaller degree). Remember that Euclid's algorithm works by starting with two numbers a and b, then replacing a with b and b with $a \bmod b$ (or in other words the *remainder* when dividing a by b) until one of the two numbers equals 0. If we now work with polynomials $f(x)$ and $g(x)$, and set $a = f(x)$ and $b = g(x)$ and working with polynomials, then this exact same algorithm will calculate the GCD of the two polynomials $f(x)$ and $g(x)$! For example, we can find that

$$\gcd(x^6 + 2x^5 - 2x^4 - 3x^3 + x^2 + x, x^4 - x^2 + 2x^3 - x - 1) = x - 1$$

using repeated division:

$$\begin{aligned} x^6 + 2x^5 - 2x^4 - 3x^3 + x^2 + x &= (x^2 - 1)(x^4 - x^2 + 2x^3 - x - 1) + (x^2 - 1) \\ x^4 - x^2 + 2x^3 - x - 1 &= (x^2 + 2x)(x^2 - 1) + (x - 1) \\ x^2 - 1 &= (x + 1)(x - 1) + 0. \end{aligned}$$

More generally, we can also work with multivariate polynomials. For example, the function

$$p(x, y, z) = x^3 + 4x^2y + xyz^2 + y^2z + 1$$

is a polynomial in variables x, y and z. With some work, there are generalised division algorithms which can calculate the remainder when dividing one multivariate polynomial by another [Cox et al., 1997, Chapter 5.3]. By the same principle as for single variable polynomials, this means that we can calculate the GCD of any two multivariate polynomials using the Euclidean algorithm[3]. This is a common practice in the field of algebraic geometry, which studies the geometric properties of the solution sets to simultaneous polynomial equations.

More generally still, in modern mathematics we often generalise ideas from the integers and (multivariate) polynomial to study abstract algebraic objects called **rings**. Rings are algebraic structures which still have a notion of addition and multiplication in the same way we can add or multiply two integers or two polynomials. A

[3]However, we can construct many generalised division algorithms which give different remainders, so the GCD may be different for different division algorithms!

lot has been studied about rings and the hierarchy of additional structures they can possess. Among these, rings which possess a division algorithm, and hence also a generalised Euclidean algorithm for calculating the GCD of two terms, are called **Euclidean domains**. Another example of a Euclidean domain is the set of real numbers of the form $a + b\sqrt{3}$ where a and b are integers. On the other hand, the set of even integers forms a ring which is **not** a Euclidean domain, and interestingly neither are sets of multivariate polynomials, even though we can still define a generalised Euclidean algorithm on them. Euclidean domains appear often in the field of number theory when trying to prove that certain algebraic equations have unique solutions, such as proving $x^3 - y^2 = 2$ has the unique positive solution $x = 3, y = 5$.

5.3 Chinese Remainder Theorem

5.3.1 Diophantine equations

As we have mentioned, one of the most well-known problems related to Euclid's algorithm and the GCD comes from the study of **Diophantine equations**. A Diophantine equation can be defined as a polynomial equation with integer coefficients, where we are only looking for integer solutions. The equations are named after **Diophantus of Alexandria**, who discussed them in his *Arithmetica*, written around the 3rd century CE. Diophantine equations have been studied for millennia, an early example being the Pythagorean equation,

$$a^2 + b^2 = c^2.$$

In particular, a linear Diophantine equation is one in which the polynomial is of degree one:

$$a_1x_1 + a_2x_2 + ... + a_nx_n = c,$$

where $a_i, c \in \mathbb{Z}$ are fixed integers and x_i are unknowns ($i = 1, ..., n$). A system of linear Diophantine equations in which there are more unknowns than equations is often called an **indeterminate system**. If such a system of equations has a solution, then there are infinitely many solutions, and we call the system **soluble**. One particular case of an indeterminate system of Diophantine equations has the form,

$$\Big\{x + a_iy_i = b_i, \quad i = 1, ..n$$

where a_i, b_i are given, and x, y_i are unknowns. This system is characterised by the number of unknowns being one greater than the number of equations; if $i = 1$ there are two unknowns and one equation, and each additional equation only adds one unknown, preserving this characteristic of the system. As we will see, regardless of how many equations the system consists of, this type of system can be solved using the same method.

We will examine the development of methods of solving these systems of Diophantine equations, from China and India in the first millennia CE, to western mathematicians in the 19th century.

5.3.2 Some history of Chinese mathematics

Mathematical scholarship in China goes back more than 2000 years. The early Chinese mathematical treatise *Zhou Bi Suan Jing* possibly dates back to before the edict of the first emperor of China, *Qin Shi Huang*, in 231 BCE ordering that many of the books in China be burned Cooke [2013]. Mathematics was highly regarded in China, not only for its utility in applications such as astronomy and calendar calculations, but as an intellectual pursuit in itself Cooke [2013]. The *Sun Zi Suan Jing* or *Mathematical Classic of Sun Zi* writes of Maths that,

> *"If one neglects its study, one will not be able to achieve excellence and thoroughness"* [4]

The author, **Sun Zi**, attributes metaphysical and even religious significance to mathematics Lam, Lay Yong & Ang, Tian Se [2004].

Whilst the *Sun Zi Suan Jing* is difficult to date, historians place it sometime in the 3rd to 5th century CE. The text contains 66 problems illustrating a range of mathematical concepts, from areas including arithmetic, algebra, and number theory Lam, Lay Yong & Ang, Tian Se [2004]. One such problem is the remainder problem, which we discuss in the next section.

5.3.3 Sun Zi's remainder problem

Problem 26 from chapter 3 of the *Sun Zi Suan Jing* is as follows:

> *"Now there are an unknown number of things. If we count by threes, there is a remainder 2; if we count by fives, there is a remainder 3; if we count by sevens there is a remainder 2. Find the number of things."*

This type of problem has come to be known as the remainder problem. In modern algebraic notation, we can see that this problem is equivalent to a system of linear Diophantine equations of the form described above with $i = 3$, namely:

$$x + 3y_1 = 2 \quad (1)$$
$$x + 5y_2 = 3 \quad (2)$$
$$x + 7y_3 = 2 \quad (3)$$

Sun Zi gives the answer as 23. As we discussed above, such an equation will have an infinite solution set, but it is implicit from the answer given that we are looking for the smallest positive integer solution to this system of equations.

Sun Zi presents a method for obtaining this solution, which is well explained in Joseph [2010]. We start by taking the coefficients by y_2 and y_3, which in our case are 5 and 7 respectively. We want to find a quantity d_1 (associated with y_1), which

[4] The translation of *Sun Zi Suan Jing* found in Lam, Lay Yong & Ang, Tian Se [2004] is used whenever the text is quoted.

is the smallest multiple of 5 and 7 that gives a remainder of 1 when divided by the coefficient by y_1 *i.e.* 3. We find that $70 = 2 \times 3 \times 5 = 3 \times 23 + 1$ does the job. We repeat for the other symmetric cases and arrive at the numbers:

$$d_1 = 5 \times 7 \times 2 = 3 \times 23 + 1 = 70$$
$$d_2 = 7 \times 3 = 5 \times 4 + 1 = 21$$
$$d_3 = 3 \times 5 = 2 \times 7 + 1 = 15$$

where $d_1, d_2, d_3,$ are associated to $y_1, y_2, y_3,$ respectively. We then sum the products of the corresponding d and y values as follows:

$$2 \times d_1 + 3 \times d_2 + 2 \times d_3 = 140 + 63 + 30 = 233.$$

This is a solution to our system of equations, but it is not the smallest integer solution; to obtain this minimal solution we subtract multiples of the lowest common multiple (LCM) of the coefficients 2,3 and 5 until we find the lowest integer:

$$\text{lcm}(5,3,2) = 105$$
$$x = 233 - (2 \times 105) = 23.$$

Hence we have obtained 23 as expected.

We could easily extend this method to a system of any finite number of remainder equations by setting i and $j = n \in \mathbb{N}$. Sun Zi did not give any explanation as to why this method worked or if there were any cases when the method failed. Both of these developments would only happen afterwards, but he was the first to propose a method to solve this problem.

5.3.4 Indian work on the problem

The next treatment of remainder problems was by the Indian mathematician and astronomer **Aryabhata I** (476-550 CE), one of the first mathematicians of the classical period of Indian mathematics. As with the Chinese mathematicians, Aryabhata came to the study remainder problems in order to determine cycles in the orbit of planets or calendar periods Joseph [2010].

Figure 26: Portrait of Aryabhata.

The study of remainder problems in India was known as the **Kuttaka**, a word derived from the verb *kutt*, meaning to crush, or to pulverise. This may refer to the way in which this method reduces larger numbers to smaller numbers Joseph [2010], or the reduction in computation time needed to solve remainder problems using this method as opposed to using trial and error Cooke [2013].

In the study of the kuttaka, the Indian mathematicians made the following important observation about the solubility of linear Diophantine equations, well explained by Datta & Singh [1962]. Firstly, note that if we write our system of linear Diophantine equations with two equations in the following form,

$$x = a_1y_1 - b_1$$
$$x = a_2y_2 - b_2$$

we can compare the two equations to obtain

$$a_1y_1 - b_1 = a_2y_2 - b_2$$

which, after rearranging, gives us

$$a_1y_1 - a_2y_2 = c, \quad \text{where} \quad c = b_1 - b_2.$$

Finding an integer value of x, now becomes equivalent to finding an integer solution (y_1, y_2) to a linear Diophantine equation of the form $a_1y_1 - a_2y_2 = c$. The Indian mathematicians of the classical period often solved remainder problems by considering Diophantine equations of this form.

The Indian mathematician **Bhaskara I** wrote the first major commentary on the *Aryabhatiya* in 629 CE Joseph [2010]. In his discussion of solving $a_1y_1 - a_2y_2 = c$, he writes:

> "*The dividend and divisor will become prime to each other on being divided by the residue of their mutual division. The operation of the pulveriser (kuttaka) should be considered in relation to them.*" [5]

In this context, the dividend and the divisor referred to are a_1 and a_2, and the "residue of their mutual division" refers to their greatest common divisor (GCD). Bhaskara is stating that in order to find solutions to our equation, we must divide both sides by $d = \gcd(a_1, a_2)$. By definition of the GCD, a_1 and a_2 are divisible by d; and c must also be divisible by d, since we are seeking integer solutions. This effectively gives a condition for when the equation $a_1y_1 - a_2y_2 = c$ has an integer solution, namely when d divides c. In this case Bhaskara proposes an algorithm to solve for y_1 and y_2. This is equivalent to stating the following:

Proposition 5.1. *For all $c \in \mathbb{N}$ and given $a, b \in \mathbb{Z}$, the equation $ar + bs = c$ is soluble in integers r and s if and only if $\gcd(a, b)$ divides c.*

This shows a significant advance in the approach to solving systems of indeterminate equations. Sun Zi was simply proposing a method which he knew worked for some equations; Bhaskara gives us a way of telling if an equation is soluble without us needing to try and solve it first.

[5] Translation taken from Datta & Singh [1962]

5.3.5 Advances in the 17th-19th centuries

The study of remainder problems by Chinese and Indian mathematicians continued for many centuries, but for the sake of brevity we will now consider the approach taken to this problem by European mathematicians in modern times. The developments in this field in the 17th-19th centuries are well explained in an article by Bullynck [2008], which we draw on below.

The first major work on this problem in the Europe was done by the French mathematician **Claude Gaspar Bachet de Mériziac** (1581-1638). He is well-known for producing a translation of Diophantus's work *Arithmetica*[6], and for writing books of recreational mathematics. In his works, Bachet began to offer generalised remainder problems with solutions and proofs, making the link to Diophantine equations. Bachet, who was unaware of the earlier Chinese and Indian work on remainder problems, is credited with proving the following result:

Theorem 5.1. *Let $a, b \in \mathbb{Z}$. Then there exist $r, s \in \mathbb{Z}$ such that $ar + bs = \gcd(a, b)$.*

This theorem is often called **Bézout's identity**, after the French mathematician **Étienne Bézout** (1730-1783), who proved that the identity holds for polynomials.

Bachet realised that one could use Euclid's algorithm to solve the Diophantine equation $mr + ns = 1$ for integers r and s, where m and n are given integers such that $\gcd(m, n) = 1$. One first uses Euclid's algorithm to find $\gcd(m, n)$, and then reverses the process used to find the integers r and s. This is best illustrated by an example:

$$64r + 15s = 1 \qquad \text{We check } \gcd(64, 15) = 1.$$

$$(a)\;\; 64 = 4 \times 15 + 4 \quad \Longrightarrow \gcd(64, 15) = \gcd(15, 4),$$

$$(b)\;\; 15 = 3 \times 4 + 3 \quad \Longrightarrow \gcd(15, 4) = \gcd(4, 3),$$

$$(c)\;\; 4 = 1 \times 3 + 1 \quad \Longrightarrow \gcd(4, 3) = \gcd(3, 1) = 1.$$

Now reverse Euclid's algorithm:

From (c) $\quad 1 = 4 - 1 \times 3,$

From (b) $\quad 3 = 15 - 3 \times 4 \quad \Longrightarrow 1 = 4 \times 4 - 1 \times 15,$

From (a) $\quad 4 = 64 - 4 \times 15 \quad \Longrightarrow 1 = 4 \times 64 - 17 \times 15.$

$$\therefore (r, s) = (4, -17).$$

If we multiply the equation $mr + ns = 1$ by any integer, we obtain the identity $ar + bs = \gcd(a, b)$ from above. This gives us the outline of a proof for Bézout's identity for integers.

[6]Bachet may have based his work on earlier translations of *Arithmetica* Gillispie [1970], but Bachet's interest in Diophantine equations is of note; also Fermat wrote his famous 'Last Theorem' as a marginal note in a copy of *Bachet's* translation.

There is some debate as to whether Bachet was indeed the first to prove this identity. Whilst the established position, as mentioned above, is that he proved Bézout's identity, it has been recently suggested by number theorist Andrew Granville that this identity was, in fact, known to Euclid Granville [2024].

Bachet's work is significant because it gave a generalised method for solving a system of two indeterminate Diophantine equations (as we showed above, solving $ar + bs = c$ is equivalent to solving a system of two such equations) although such a method had already been developed by Chinese and Indian mathematicians. Bachet also began the movement away from thinking of "remainder problems" to thinking of equations and systems of equations instead, which would be important to further work on the subject.

Bachet did not give a general method for solving a system of n indeterminate Diophantine equations, nor a condition for when such a system is soluble. It would take about a century until the renowned mathematician **Leonhard Euler** (1707-1783) would provide this. In a paper published in 1740, Euler proved that one could generalise Bachet's method[7] for to any number of indeterminate equations, and gave a condition for such a system to be soluble Euler [1740]. The result came to be known as the **Chinese Remainder Theorem** (CRT).

In modern mathematics texts, the CRT is stated using modular arithmetic, a notation (see Section 5.2.2) developed by **Carl Friedrich Gauss** (1777-1855). The CRT is stated in Gauss's classic text on number theory, *Disquisitiones Arithmeticae*, published in 1801, as follows:

Theorem 5.2 (CRT: Gauss's version)**.** *Let $n_1, \ldots, n_r$ be pairwise coprime[8] natural numbers. Then for any $a_1, \ldots, a_r \in \mathbb{N}$, the system of equations*

$$\left\{ x \equiv a_i \mod n_i, \quad i = 1, \ldots, r, \right.$$

has an integer solution x, and this solution is unique modulo $n_1 \times \cdots \times n_r$.

The proof of this theorem builds on Bachet's method of reversing Euclid's algorithm, and gives a general method for finding x.

Finally, we answer the natural question of how the Chinese Remainder Theorem (also called **Sun Zi's Theorem**) got its name. According to Dickson [1920], this originated with a mathematically literate missionary named Alexander Wylie during the late Qing dynasty (mid 19th century). Wylie wrote an article in the *North China Herald* in which he correctly credited the first work on this theorem to the Chinese mathematicians, over 1000 years before Bachet, Euler and Gauss.

5.3.6 Diophantine equations observed in nature

We now consider an equation of the form

$$x = ny = mz$$

[7] Although, it appears that Euler was not aware of Bachet's work

[8] That is, for any two distinct $n_i, n_j \in \{n_1, \ldots, n_r\}$, we have $\gcd(n_i, n_j) = 1$.

where $n, m \in \mathbb{Z}$ are fixed and x, y, z are variables. This system can be solved relatively easily, and the smallest positive integer solution is given by

$$x_0 = \mathrm{lcm}(|m|, |n|).$$

This simple identity has been observed in several species of cicadas, where it has been empirically "optimised" by evolutionary processes. All of the so-called "periodical" cicadas live for either 13 or 17 years and spend most of their lives underground. In spring, towards the end of their lifetime , they emerge synchronously, reproduce, and die within two months.

A natural question to ask is where do the numbers 13 and 17 come from. One might notice that both of those numbers are primes, which turns out to be very significant in this situation. Suppose that the life cycle is 12 years long instead. It does not seem like there is a big change; however, this means that the predators that also lead a periodic life cycle could synchronise much more easily. If their life cycle is 1, 2, 3, 4, 6, or 12 years, they can prey on the cicadas every 12 years, when the cicadas emerge overground, and drive them to extinction. A tiny change from 12 to 13 now means that there are only two relevant cycles: 1 and 13, which makes the cicadas are much safer from predators. For all we know, thousands of years ago we could have had many more species of periodical cicadas that went extinct over time. Only those that evolved and changed their life cycles to "optimal" periods could survive.

Another aspect that is relevant in the discussion is that there can be a situation where two species of cicadas of different life cycles n and m occupy the same territory. How often will they emerge in the same spring? We answered this at the beginning of the section. They will overlap every $\mathrm{lcm}(m, n)$ many years, and in that season there might not be enough food for all of them. Notice that if $n = 12$ and $m = 18$, then the overlap is every 36 years as opposed to 221 years, when $n = 13$ and $m = 17$.

For more information on the subject, see Goles et al. [2000].

References

Jörg Arndt and Christoph Haenel. *Pi-unleashed*. Springer Science & Business Media, 2012.

Adnan Baki. Al khwarizmi's contributions to the science of mathematics: Al kitab al jabr wa'l muqabalah. *Journal of Islamic Academy of Sciences*, 5(3):225–228, 1992.

L. Berggren, J. Borwein, and P. Borwein. *Modular equations and approximations to π*. Springer, New York, 1997.

Bruce C. Berndt. *Ramanujan's Notebooks.* Springer Science & Business, 1997.

Daniel Bernoulli. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5:175–192, 1738. Translated as "Exposition of a New Theory on the Measurement of Risk" in *Econometrica*, 1954, vol. 22(1), pp. 23–36.

Jacob Bernoulli. *Ars Conjectandi*. Thurnisius, 1713. First formal statement of the Law of Large Numbers.

Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

Viktor Blåsjö. A critique of the arguments for maragha influence on copernicus. *Journal for the History of Astronomy*, 45(2):183–195, 2014.

Maarten Bullynck. Modular arithmetic before c.f. gauss: Systematizations and discussions on remainder problems in 18th-century germany. *Historica Mathematica 36*, 2008.

Francesco Paolo Cantelli. Sulla legge forte dei grandi numeri. *Rendiconti della Reale Accademia dei Lincei*, 26:39–45, 1917. Strong law under moment conditions.

Heng Huat Chan, Song Heng Chan, and Zhiguo Liu. Domb's numbers and ramanujan–sato type series for $1/\pi$. *Advances in Mathematics*, 186(2):396–410, 2004.

David Chudnovsky and Gregory Chudnovsky. *Approximation and complex multiplication according to Ramanujan*. Ramanujan revisited: proceedings of the centenary conference, 1988.

Roger L. Cooke. *The History of Mathematics: A Brief Course*. John Wiley & Sons, third edition, 2013.

Leon Cooper. Did egyptian scribes have an algorithmic means for determining the circumference of a circle? *Historia Mathematica*, 38:455–484, 2011.

David Cox, John Little, Donal O'shea, and Moss Sweedler. *Ideals, varieties, and algorithms*. Springer, 1997.

Datta & Singh. *Histroy of Hindu Mathematics: A Source Book*. Asia Publishing House, single volume edition edition, 1962.

J. W. Dauben. *"Jiu zhang suan shu" (Nine Chapters on the Art of Mathematics) – An Appraisal of the Text, its Editions, and Translations*, volume 23. Sudhoffs Archives, 2013. doi: 10.25162/sudhoff-2013-0017.

Abraham de Moivre. *The Doctrine of Chances: Or, A Method of Calculating the Probabilities of Events in Play*. Millar, London, 1738. 2nd edition.

Leonard Eugene Dickson. *History of the Theory of Numbers, Volume II: Diophantine Analysis*. Carnegie Institute of Washington, 1920.

Justus Doenecke. James A. Garfield – Miller Center of Public Affairs, University of Virginia. [Online; accessed 24-July-2024], 2024. URL https://millercenter.org/president/garfield.

Leonhard Euler. Solutio problematis arithmetici de inveniendo numero, qui per datos numeros divisus relinquat data residua. *Commentarii academiae scientiarum Petropolitanae*, pages 46–66, 1740.

European Information Technologies Certification Academy. How does the euclidean algorithm work to find the greatest common divisor (gcd) of two integers, and why is it important in cryptographic protocols? [Online; accessed 12-August-2025], 2024. URL https://tinyurl.com/fndsd64f.

William Feller. *An Introduction to Probability Theory and Its Applications, Volume II*. John Wiley & Sons, New York, 2nd edition, 1971. ISBN 978-0-471-25709-1.

Hans Fischer. *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. Springer, New York, 2 edition, 2010.

T.I. Kao Frank J Swetz. *Was Pythagoras Chinese? An Examination of Right Triangle Theory in Ancient China*, volume 40. The Pennsylvania State University Press, 1977.

Charles Geyer. *Markov Chain Monte Carlo Lecture Notes*. University of Minnesota Twin Cities, 1998. URL https://www.stat.umn.edu/geyer/f05/8931/n1998.pdf.

Charles Coulston Gillispie. *Dictionary of Scientific Biography, Volume I*. Charles Scribner's Sons, 1970.

Eric Goles, Schulz Oliver, and Markus Mario. Prime number selection of cycles in a predator-prey model. *Complexity*, 6(4):33–38, 2000.

Andrew Granville. It is not "bézout's identity", 2024. URL https://arxiv.org/abs/2406.15642.

Geoffrey Grimmett and Dominic JA Welsh. *Probability: an introduction*. Oxford University Press, 2014.

Charles Miller Grinstead and James Laurie Snell. *Introduction to probability*. American Mathematical Soc., 2012.

Mitchell A Harris and Edward M Reingold. Line drawing, leap years, and euclid. *ACM Computing Surveys (CSUR)*, 36(1):68–80, 2004.

Z. Heng. Water-powered armillary sphere. The Ancient Chinese Technology, Han Dynasty. URL https://learning.hku.hk/ccch9051/group-1/items/show/21. [Online; accessed: 2024-08-13].

I. Hogendijk. Pure mathematics in islamic civilization. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, 1994.

Iamblichus. *The life of Pythagoras*. Krotona ; Hollywood, Calif. : Theosophical Pub. House, 1918.

Institute of Mathematical Statistics. Obituary: Gopinath kallianpur 1925–2015. Online; accessed 2025-07-17, 2015. URL https://imstat.org/2015/08/28/obituary-gopinath-kallianpur-1925-2015/.

Suzuki Jeff. Al khayyami and the cubic: Setup. [Online; accessed: 2024-09-13], 2020. URL https://www.youtube.com/watch?v=d553qUklUEY.

George Gheverghese Joseph. *The Crest of the Peacock: Non-European Roots of Mathematics*. Princeton University Press, 2010.

Gopinath Kallianpur and Itaru Mitoma. A segal-langevin type stochastic differential equation on a space of generalized functionals. *Canadian Journal of Mathematics*, 44(3):524–552, 1992. doi: 10.4153/CJM-1992-034-8.

V. Khatri and P. Tiwari. Square root of 2 - extension of ancient indian vedic method. *Preprint*, 2023.

Aleksandr Khinchin. Ueber das gesetz der grossen zahlen. *Mathematische Annalen*, 101:282–285, 1929. WLLN under minimal assumptions (only finite expectation).

Andrey Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, 1933. Foundations of measure-theoretic probability; formalizes SLLN and CLT.

J. Kraft and L. Washington. *An Introduction to Number Theory with Cryptography*. Chapman and Hall/CRC, 2018.

Lam, Lay Yong & Ang, Tian Se. *Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra in Ancient China*. World Scientific, revised edition, 2004.

Pierre-Simon Laplace. *Théorie Analytique des Probabilités*. Courcier, Paris, 1812.

E. S. Loomis. *The Pythagorean Proposition*. Classics in mathematics education. National Council of Teachers of Mathematics, 1968. ISBN 9780873530361. URL https://books.google.co.uk/books?id=V_o-AAAAIAAJ.

Beatrice Lumpkin. The zero concept in ancient egypt. In *Proceedings of the International Seminar and Colloquium on 1500 Years of Aryabhateeyam*, pages 161–67, 2002.

Eli Maor. *The Pythagorean theorem: a 4,000-year history*, volume 65. Princeton University Press, 2019.

Seyed N. Mousavian, Suzanne Sumner, Mehdi Aminrazavi, and Glen Van Brummelen. Umar Khayyam. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2024 edition, 2024.

J. Needham. *Science and Civilization in China: Volume 3, Mathematics and the Sciences of the Heavens and the Earth.*, volume 3. Cambridge University Press, 1959.

Joseph Needham. Astronomy in classical china. *Quarterly Journal of the Royal Astronomical Society*, 3, 1962.

Otto Neugebauer. *The Exact Sciences in Antiquity*. Courier Corporation, 1969.

A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. Electrical Engineering Series. McGraw-Hill, 1991. ISBN 9780070484771. URL https://books.google.co.uk/books?id=4IwQAQAAIAAJ.

K. Plofker. *Mathematics in India*. Princeton University Press, 2009.

Siméon-Denis Poisson. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Bachelier, 1837. Introduced the term "law of large numbers" and generalized Bernoulli's theorem.

Christopher Powell. *The Shapes of Sacred Space: A Proposed System of Geometry Used to Lay Out and Design Maya Art and rchitecture and Some Implications Concerning Maya Cosmology*. PhD thesis, The University of Texas at Austin, 2010.

Michael A Proschan and Pamela A Shaw. *Essentials of probability theory for statisticians*. Chapman and Hall/CRC, 2018.

Christian Robert and George Casella. Introducing monte carlo methods with r. *SpringerLink*, 2019. doi: https://doi.org/10.1007-978-1-4419-1576-4. URL https://link.springer.com/book/10.1007/978-1-4419-1576-4.

Frederic Rosen. *The Compendious Book on Calculation by Completion and Balancing, al-Khwārizmī. English Translation, Retrieved*. London Printed for the Oriental Translation Fund and sold by J. Murray, 2009.

Sheldon M Ross. *A first course in probability*. Pearson Harlow, UK, 2020.

David Ruelle. The nature of properly human mathematics. *Advances in continuous and discrete models*, 2023(1), Mar 2023. doi: https://doi.org/10.1186/s13662-023-03761-9. URL https://advancesincontinuousanddiscretemodels.springeropen.com/articles/10.1186/s13662-023-03761-9#citeas.

H. C. Schepler. *The Chronology of Pi*, volume 23. Mathematics Magazine, 1947. URL https://search.worldcat.org/it/title/51205246.

M.R. Schroeder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W.H. Freeman, 1991. ISBN 9780671742171. URL https://books.google.co.uk/books?id=_KBIMwEACAAJ.

M. Vali Siadat and Alana Tholen. Omar khayyam: Geometric algebra and cubic equations. *Math Horizons*, 28(1):12–15, 2021. doi: 10.1080/10724117.2020.1770495. URL https://doi.org/10.1080/10724117.2020.1770495.

P. D. Straffin. *Liu Hui and the First Golden Age of Chinese Mathematics*, volume 71. Mathematics Magazine, 1998.

Heinz Klaus Strick. Qin Jiushao: Der Unberechenbare. *Spektrum.de*, Jun 2022. URL https://www.spektrum.de/wissen/qin-jiushao-der-unberechenbare/2035786. Translated to English by O'Connor, J. (2022).

G. Thibaut. *The Śulvasútra of Baudháyana, with the commentary by Dvárakánáthayajvan*, volume 9. The Pandit, 1875.

Joachim Von Zur Gathen and Jürgen Gerhard. *Modern computer algebra*. Cambridge university press, 2003.

E. W. Weissten. Pythagorean triple. [Online; accessed 2024-07-10], 2024. URL https://mathworld.wolfram.com/PythagoreanTriple.html.

Wikipedia contributors. Formulas for generating pythagorean triples. [Online; accessed 2024-07-10], 2024a. URL https://en.wikipedia.org/w/index.php?title=Formulas_for_generating_Pythagorean_triples&oldid=1222796995.

Wikipedia contributors. "omar khayyam". [Online: Accessed 2024-09-09], 2024b. URL https://en.wikipedia.org/w/index.php?title=Omar_Khayyam&oldid=1244754401.

Wikipedia contributors. Al-khwarizmi. [Accessed; 2024-09-09], 2024c. URL https://en.wikipedia.org/w/index.php?title=Al-Khwarizmi&oldid=1244778352.

Yale University. A 3d-print of ancient history: one of the most famous mathematical texts from mesopotamia | institute for the preservation of cultural heritage. Online: Accessed 2024-07-19, 2016. URL https://ipch.yale.edu/news-events/3d-print-ancient-history-one-most-famous-mathematical-texts-mesopotamia.

Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–271, 1909. First strong law of large numbers under specific conditions.