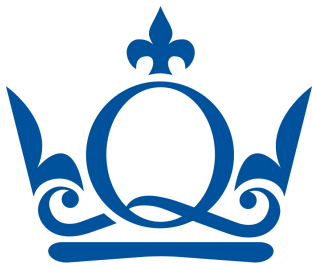

Historical Contributions of Different Cultures to Mathematics

Authors:

Kabiru Abubakari
Zahraa Al-Mahdawy
Asya Bagan
Shaivi Darsi
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 – 2026



Queen Mary
University of London

Contents

1	Introduction	3
2	The Pythagorean Theorem	4
2.1	History	4
2.1.1	Babylonia	4
2.1.2	China	5
2.1.3	India	6
2.1.4	Pythagoras	6
2.2	Different proofs of the Pythagorean Theorem	7
2.2.1	President Garfield's proof	7
2.2.2	Similarity of triangles proof	8
2.2.3	The folding bag proof	9
2.2.4	Miss Ann Condit's proof	9
2.3	Related results	10
2.3.1	Fermat's Last Theorem	10
2.3.2	Archimedes' approximation of π	11
2.3.3	Heron's formula	12
2.3.4	The diagonal of a square	13
3	Irrational Numbers	14
3.1	What are irrational numbers?	14
3.1.1	Why are irrational numbers useful?	14
3.2	$\sqrt{2}$	14
3.2.1	Proof that $\sqrt{2}$ is irrational	14
3.2.2	The Babylonian Clay Tablet YBC-7289 and Irrational Numbers	15
3.2.3	Baudhayana approximation	15
3.3	π	17
3.3.1	Approximations in China	17
3.3.2	Ramanujan's work	18
3.3.3	Applications of π across different cultures, fields, and times	19
3.4	Roots of polynomial equations	21
3.4.1	Al-Khwarizmi completing the square	21
3.4.2	Omar Khayyam cubic equation	22
4	Laws, Limits, and Randomness: A Mathematical History of Probability	25
4.1	Historical Context	25
4.1.1	The Arithmetic Triangle Across Cultures	25
4.1.2	Games of Chance and the Problem of Points	26
4.2	The Law of Large Numbers	29
4.2.1	Early Formulation	29
4.2.2	The Weak Laws of Large Numbers	29
4.2.3	Kolmogorov's Strong Law of Large Numbers (1933)	31
4.3	The Central Limit Theorem	32
4.3.1	First Glimpse of the Bell Curve: De Moivre–Laplace Theorem	32
4.3.2	Generating functions (1749–1827)	33
4.3.3	Shaping the Bell Curve: Gauss and the Least Squares Method	34
4.3.4	Unveiling the Universal Bell Curve: Central Limit Theory	34
4.4	Applications	36
4.4.1	The St. Petersburg Paradox (1713–1738)	36
4.4.2	Brownian Motion as a Scaling Limit of Random Walks	37
4.4.3	Markov Chains	38
4.4.4	Monte Carlo integration	38
4.4.5	Conclusions	39

5	Algorithms for Solving Algebraic Equations	40
5.1	The history of division	40
5.2	Euclid's Algorithm	40
5.2.1	The History of Calendars and Astronomy	40
5.2.2	Greatest Common Divisors	42
5.2.3	Mathematical formulation of Euclid's algorithm	42
5.2.4	Connecting Euclid to the Islamic Golden Age	43
5.2.5	Modern Generalisations	44
5.3	Chinese Remainder Theorem	44
5.3.1	Diophantine equations	44
5.3.2	Some history of Chinese mathematics	45
5.3.3	Sun Zi's remainder problem	45
5.3.4	Indian work on the problem	46
5.3.5	Advances in the 17th-19th centuries	47
5.3.6	Diophantine equations observed in nature	48
6	Calculus: A Global History of Change, Motion, and the Infinite	49
6.1	What is calculus?	49
6.2	The beginnings of calculus: exhaustion, approximation, and curved shapes	50
6.3	Infinitesimals, indivisibles, and tangent problems	52
6.4	Medieval Islamic mathematics and the prehistory of integration	53
6.5	Indian mathematics and infinite series	56
6.6	Newton and Leibniz: two systematic calculi	57
6.7	Calculus, education, and women mathematicians	60
6.8	From calculus to analysis	62
6.9	Modern calculus and new applications	64
6.10	Conclusion: whose calculus?	66
7	The Theory that would not die: Bayesian Probability	68
7.1	Was Bayes trying to calculate the existence of God in the 18th century?	68
7.2	Laplace: Independently Developing and Extending Inverse Probability	69
7.3	Inverse probability in the nineteenth century	72
7.4	So what is Bayes' theorem today?	73
7.5	From inverse probability to competing theories of statistical inference (c. 1900–1954)	74
7.6	Bayesian inference at Bletchley Park: Turing, Good and Naval Enigma (c. 1940–1943)	76
7.7	The computational revolution in Bayesian inference: from Bayes' theorem to algorithms (c. 1950–present)	78
7.7.1	The computational revolution	78
7.7.2	The modelling revolution	79
7.7.3	Bayesian ideas in machine learning and decision-making	80
7.8	Conclusions	81

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

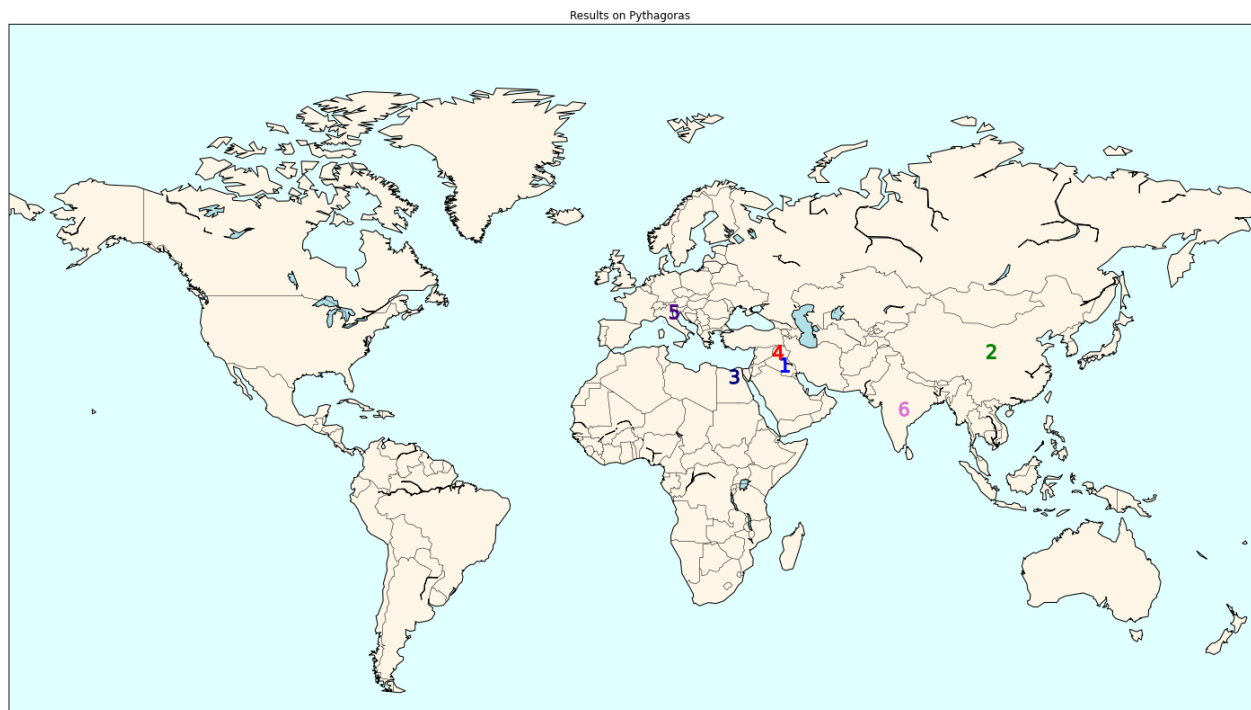
2 The Pythagorean Theorem

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.



1. Babylon: Tablet YBC 7289 and Plimpton 322
2. China: kou-ku theorem
3. Alexandria:
Euclid proof Pythagoras theorem
Archimedes applies PT to find Pi
Pappus proves an extended version of PT
4. Harran mesopotamia: Ibn Qorra generalization of the PT
5. Venice: The first printed edition of the Elements
6. India: Baudhayana states the special case of the PT for the square

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 significant find, *Plimpton 322*, lists Pythagorean triples (sets of integers a, b, c

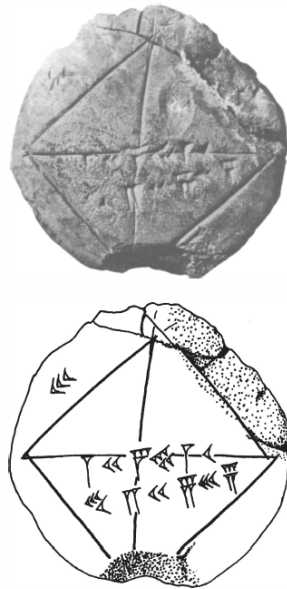


Figure 2: Clay tablet YBC-7289 (Taken from [Maor \[2019\]](#)).

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.

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 known in China as the *kou-ku* theorem.

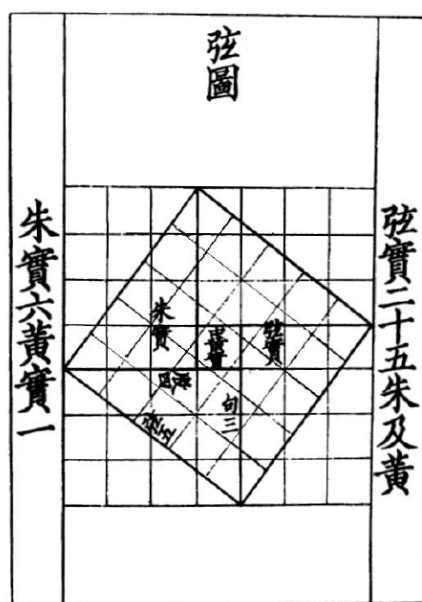


Figure 3: Chiu chang proof (Taken from [Frank J Swetz \[1977\]](#)).

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

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

¹The first edition published in 1927 contains 230 proofs.

area ACB , that is, $\frac{1}{2}c^2 + 2\frac{1}{2}ab = \frac{1}{2}c^2 + ab$. Hence, this yields

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

□

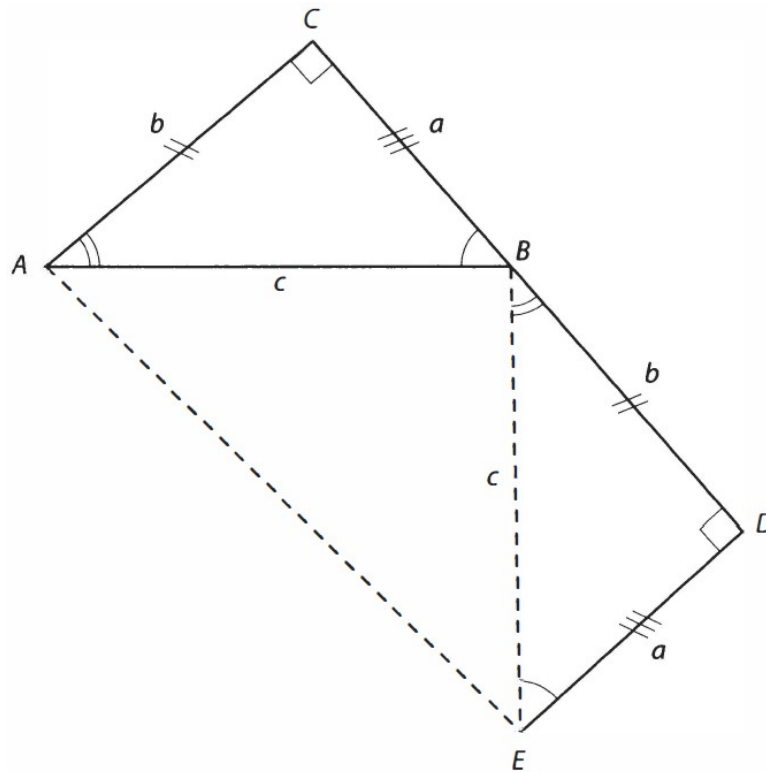


Figure 4: James A. Garfield's proof (1876) (reproduced from Figure 8.8 of Maor [2019]).

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 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

$$A_1 = kc^2, \quad A_2 = ka^2, \quad A_3 = kb^2, \quad (1)$$

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

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

□

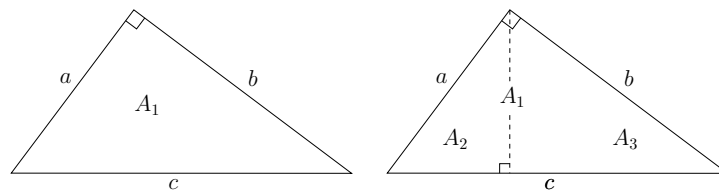


Figure 5: Bisecting a right angle triangle

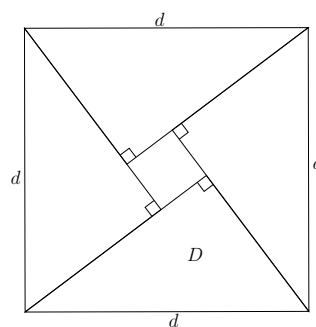


Figure 6: The area of a right-angled triangle with hypotenuse d is proportional to the area of a square of side length d .

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. □

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.

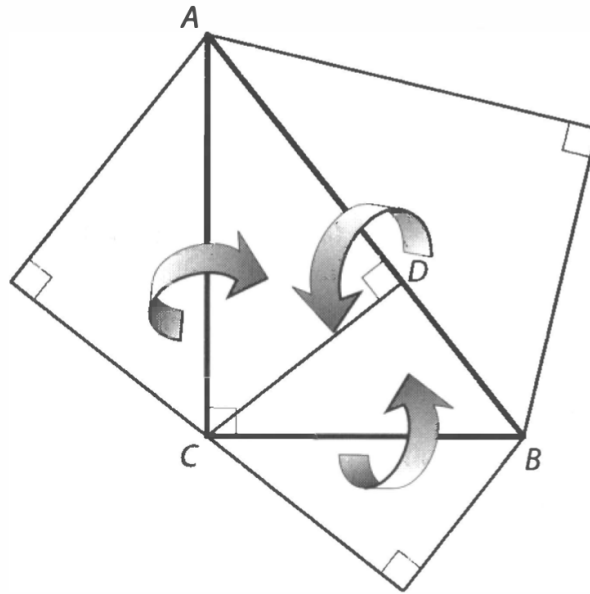


Figure 7: If we choose right triangles as the polygons based on the sides of the original right triangle, we obtain the delightfully simple folding bag proof (reproduced from Figure S4.1 of Maor [2019]).

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, $\angle DCR = 90^\circ - \alpha$, and $\angle CRD = 90^\circ$, 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. □

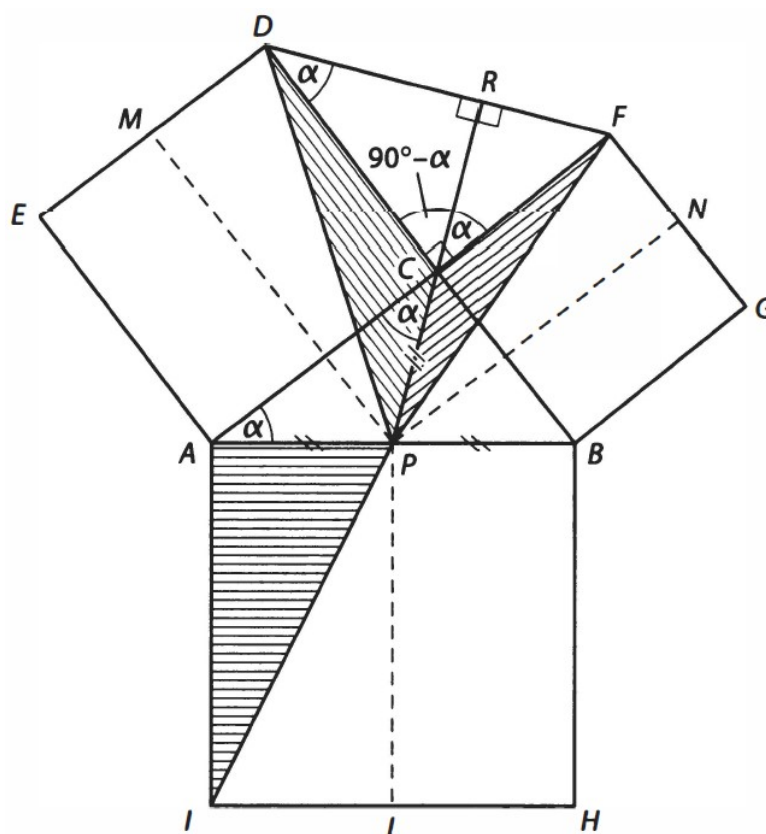
2.3 Related results

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

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

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.

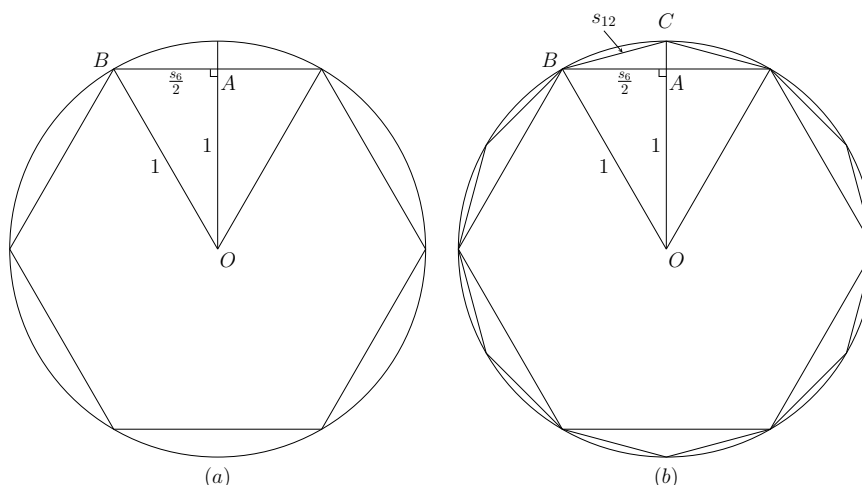


Figure 9: (a): A hexagon of side length s_6 inscribed within a unit circle. (b) A hexagon and dodecagon of side lengths s_6 and s_{12} , respectively, inscribed within a unit circle.

Mathematically,

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

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:

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

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

$$\begin{aligned} m^2 + h^2 &= b^2, \\ n^2 + h^2 &= c^2. \end{aligned}$$

Subtracting the bottom equation from the first, we get

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

After some algebraic manipulation, we obtain

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

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. □

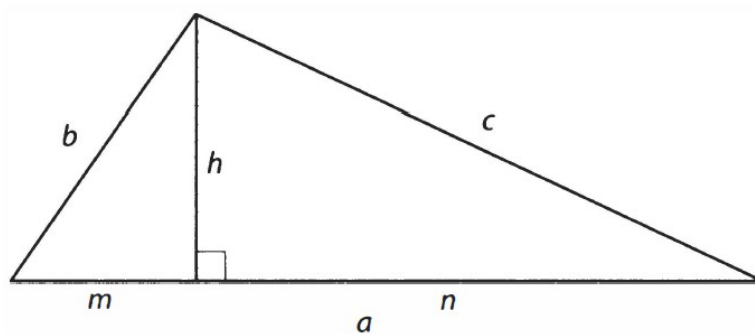


Figure 10: Triangle to prove Heron's formula.

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

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

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:

$$\pi = 3.1415926535897932384626433832795028841971693993751058209 \dots$$

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 5th 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 \[1918\]](#). Let's 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:

$$a^2 = 2b^2. \quad (3)$$

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. $\square$

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 cuneiform (the script used by Mesopotamian scribes and scholars). Its round shape and large writing indicates it was typical tool for Babylonian students *imšukum* 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 \[2009a\]](#). 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 *Sulvasutras*, 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:

$$\frac{1}{3} + \frac{1}{12} - \frac{1}{408} = \frac{577}{408} = \mathbf{1.4142156863}$$

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, 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 ,

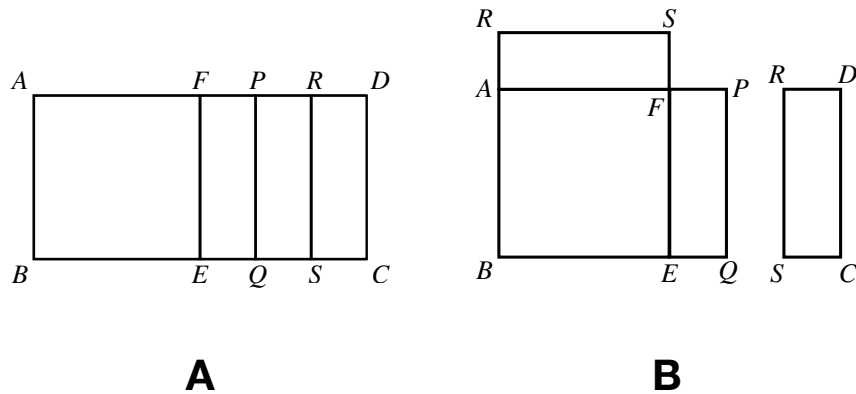


Figure 11: **A** First step of the Baudhayana geometrical approximation of $\sqrt{2}$, in which rectangle $ABCD$ is first halved in two equal squares, then square $FECD$ is divided into three rectangles with the same area. **B** Thanks to a rigid translation, rectangle $PQSR$ moves on one side of square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

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 $\frac{1}{3} + \frac{1}{12} - \frac{1}{408}$. $\square$

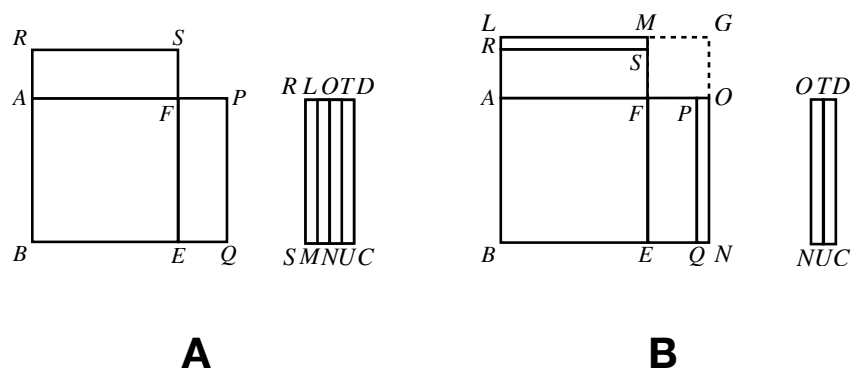


Figure 12: Rectangle $RSCD$ is divided into four equal rectangles (**A**), with two of them then moved to be adjacent to segments RS and PQ respectively (**B**). Reproduced from [Khatri and Tiwari \[2023\]](#).

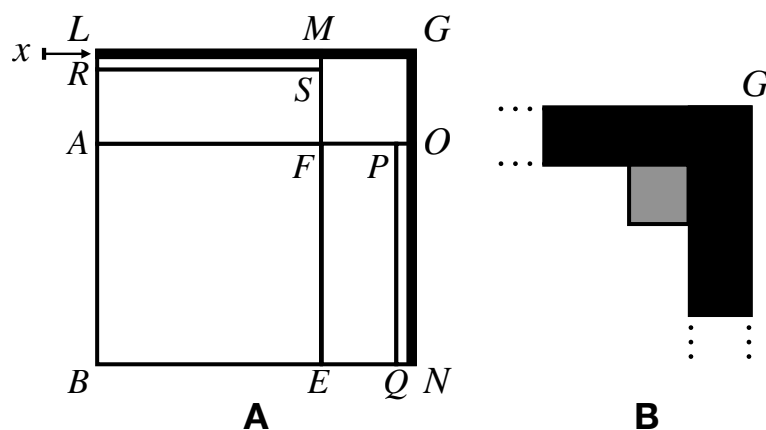


Figure 13: **A** As a final step of his proof, Baudhayana subtracts the *savisesah* (represented in the figure by the two black solid rectangles with shortest side equal to x) from the other contributions to obtain the approximation of square root of 2. **B** The grey square in the figure represents 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 \[2023\]](#).

Notice that, in the “proof”, Baudhayana left a small empty square at the corner of $MFOG$: 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 \[1998a\]](#). 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.

CHAPTER XVIII

1. $1 + \left(\frac{1}{2}\right)^2 x + \left(\frac{1 \cdot 1}{2 \cdot 2}\right)^2 x^2 + \left(\frac{1 \cdot 1 \cdot 3}{2 \cdot 2 \cdot 4}\right)^2 x^3 + \left(\frac{1 \cdot 1 \cdot 3 \cdot 5}{2 \cdot 2 \cdot 4 \cdot 6}\right)^2 x^4 + \dots$
 $= x(1-x) + \int x dx = \frac{x}{2}(1+x) + \frac{x}{2} \left\{ 1 - 24 \left(\frac{1}{e^{1/2}} + \frac{x}{e^{4/2}} + \dots \right) \right\}$
2. $1 - \frac{1}{2}x - \frac{1 \cdot 1}{2 \cdot 2} x^2 - \frac{1 \cdot 1 \cdot 3}{2 \cdot 2 \cdot 4} x^3 - \frac{1 \cdot 1 \cdot 3 \cdot 5}{2 \cdot 2 \cdot 4 \cdot 6} x^4 - \dots$
 $= x(1-x) + \frac{1}{2} \int x dx = \frac{x}{2}(1-x) + \frac{1}{3} \left\{ 1 - 24 \left(\frac{1}{e^{1/2}} + \frac{x}{e^{4/2}} + \dots \right) \right\}$
3. The perimeter of an ellipse whose eccentricity is h , is
 $2a\pi \left\{ 1 - \frac{1}{2}h^2 - \frac{1 \cdot 1}{2 \cdot 2} h^4 - \frac{1 \cdot 1 \cdot 3}{2 \cdot 2 \cdot 4} h^6 - \dots \right\}$
 $= \pi(a+b) \left\{ 1 + \left(\frac{1}{2}\right)^2 \left(\frac{a-b}{a+b}\right)^2 + \left(\frac{1 \cdot 1}{2 \cdot 2}\right)^2 \left(\frac{a-b}{a+b}\right)^4 + \left(\frac{1 \cdot 1 \cdot 3}{2 \cdot 2 \cdot 4}\right)^2 \left(\frac{a-b}{a+b}\right)^6 + \dots \right\}$
 $= \pi \left\{ 3(a+b) - \sqrt{(a+3b)(3a+b)} \right\} \text{ nearly}$
 $= \pi(a+b) \left\{ 1 + \frac{3x}{10 + \sqrt{4-3x}} \right\} \text{ very nearly where } x = \left(\frac{a-b}{a+b}\right)^2.$

N.B. i. $\pi = 3.1415926535897932384626434$
 ii. $\log 10 = 2.302585092994045684018$
 iii. $e^{-\pi} = .04321391826877225$
 iv. $e^{\pi} = 4.810477380965351653473$

Cor. $\pi = \frac{355}{113} \left(1 - \frac{.0003}{355-113} \right) \text{ very nearly}$
 $= \sqrt[4]{97 \frac{1}{2} - \frac{1}{11}} \text{ nearly}$

4. $\frac{\sqrt{x}}{2} \left\{ 1 + \left(\frac{1}{2}\right)^2 \frac{x}{3} + \left(\frac{1 \cdot 1}{2 \cdot 2}\right)^2 \frac{x^2}{5} + \left(\frac{1 \cdot 1 \cdot 3}{2 \cdot 2 \cdot 4}\right)^2 \frac{x^3}{7} + \dots \right\}$
 $= \log \frac{1+e^{-x/2}}{1-e^{-x/2}} - 3 \log \frac{1+e^{-3x/2}}{1-e^{-3x/2}} + 5 \log \frac{1+e^{-5x/2}}{1-e^{-5x/2}} - \dots$
5. $\log \frac{16}{x} - 6 \log \frac{x}{1} - \left(\frac{1 \cdot 1}{2 \cdot 2}\right)^2 \frac{x^2}{2} - \left(\frac{1 \cdot 1 \cdot 3}{2 \cdot 2 \cdot 4}\right)^2 \frac{x^3}{3} - \dots$
 $= 9 - 4 \left\{ \log(1-e^{-x}) - 3 \log(1-e^{-3x}) + 5 \log(1-e^{-5x}) - \dots \right\}$

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

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 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 20th 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.

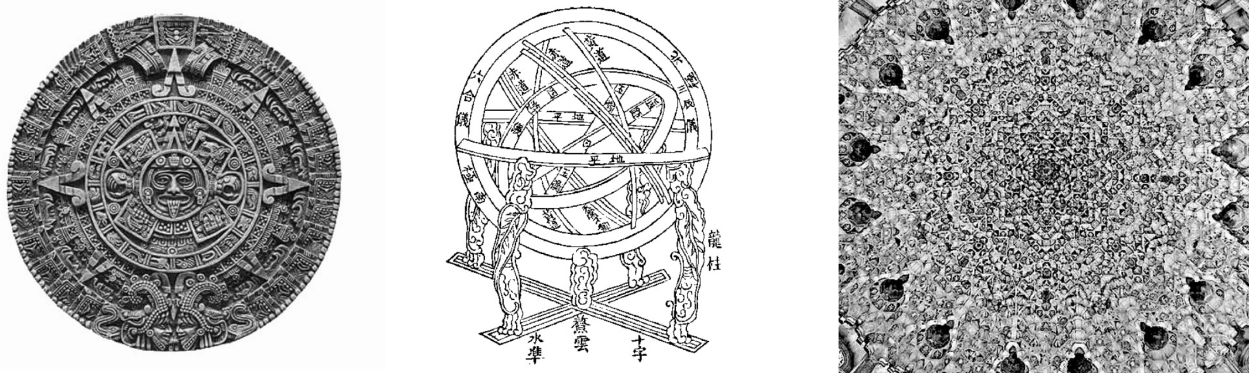


Figure 15: On the left of the figure, Mayans' calendar. An illustration of the ancient Chinese water-powered armillary sphere is shown in the middle of the figure Heng [Han Dynasty]. On the right, a photo of the "Alhambra ceiling" taken by the photographer and artist Jason Priem.

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 aesthetically 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].

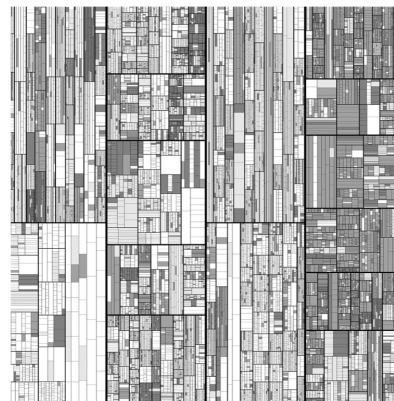
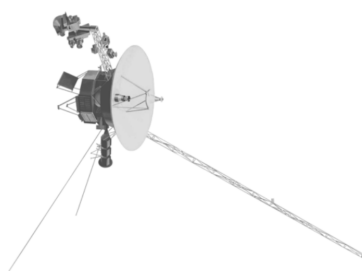


Figure 16: On the left of the figure, a picture of the Golden Gate Bridge in 1930 (The Bancroft Library, *Construction Photographs of the Golden Gate Bridge collection*). Voyager 1 probe is shown in the middle of the figure. On the right, the art “3628 digits of π ” from the scientist and artist Martin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

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

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

For “squares are equal to roots”, “Squares are equal to numbers”, and “roots are equal to numbers.” respectively.

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

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

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

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.

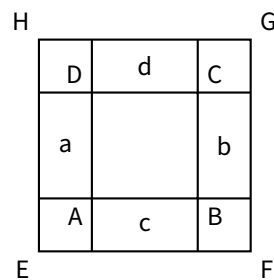


Figure 17: Completing the square EFGH

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

$$x^2 + 10x = 39 \quad (4)$$

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 “rubā'iyyat”; that is, poems which takes the form of four lines also known as quatrains. 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).

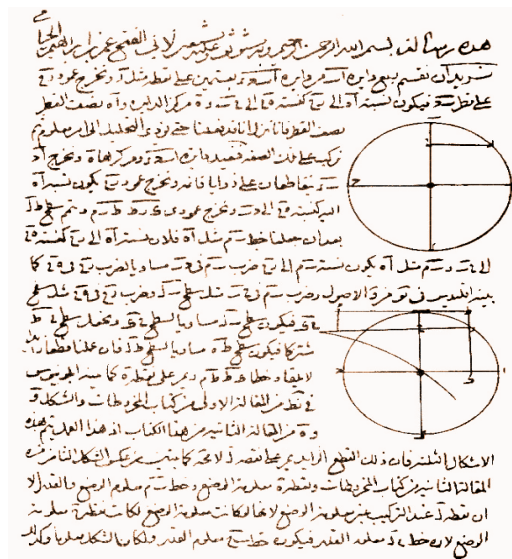


Figure 18: First page of a manuscript kept in Tehran University. Source: [Siadat and Tholen \[2021\]](#)

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 9th century presented a step-by-step solution to quadratic equations by completing the square, also Thabit ibn Qurra in the late 9th 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

$$\begin{aligned} x^3 &= c; & x^3 + bx &= c; & x^3 + c &= bx; & x^3 + c &= ax^2; & x^3 + ax^2 &= c \\ x^3 &= ax^2 + c; & x^3 &= bx + c; & x^3 + ax^2 &= bx + c; & x^3 + c &= ax^2 + bx \\ x^3 &= ax^2 + bx + c; & x^3 + x^2 + c &= bx; & x^3 + ax^2 + bx &= c; & x^3 + bx &= ax^2 + c \\ x^3 + bx + c &= ax^2 \end{aligned}$$

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

$$x^3 + bx = c \quad (5)$$

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 .

$$\Rightarrow x^3 + m^2x = m^2n; \quad \therefore x^3 = m^2n - m^2x \quad (6)$$

$$\Rightarrow x^2x = m^2(n - x); \quad \therefore x : (n - x) = m^2 : x^2 \quad (7)$$

“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 $CB : CD = CD : AC$, then CD is perpendicular to the diameter AB of the semi-circle. Therefore equation (8) is the equation of a circle with diameter n .

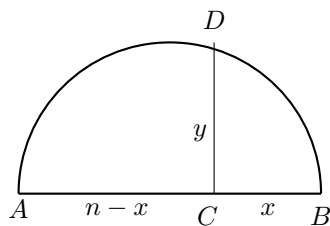
$$\implies x : y = y : (n - x); \quad \therefore x(n - x) = y^2 \quad (8)$$

$$\implies x : (n - x) = y^2 : (n - x)^2 \quad (9)$$

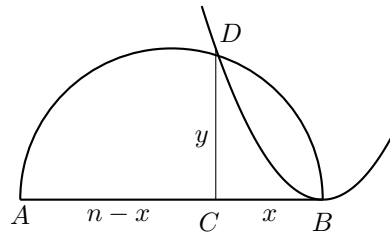
$$\implies m^2 : x^2 = y^2 : (n - x)^2; \quad \therefore m : x = y : (n - x) \quad (10)$$

$$\implies m : x = x : y; \quad \therefore my = x^2 \quad (11)$$

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 $(a + b)^n$.

				1					
			1		1				
		1		2		1			
	1		3		3		1		
1		4		6		4		1	

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.

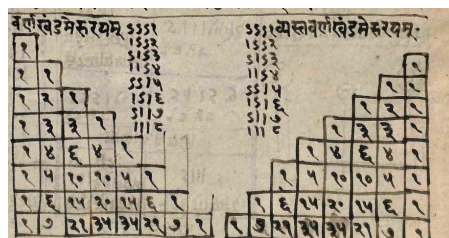


Figure 20: Pingala's *Mere-prastara* 300 BCE

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 re-design 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.

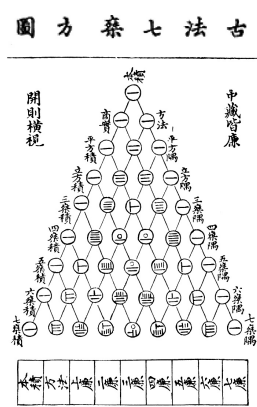


Figure 21: Yanghui triangle ca. 1238–1298.

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 17th-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 *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 their 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:

$$q = 1 - p \quad (12)$$

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:

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

(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 19th and 20th 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* [?]. 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 $X_1, X_2, \dots, X_n$ be independent Bernoulli random variables with success probability p_n , and let $S_n = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p_n)$. Suppose that

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

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

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

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 $X_1, X_2, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \text{Var}(X_i) < \infty$. Define the sample mean by

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

Then, for any $\varepsilon > 0$,

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

which implies that

$$\bar{X}_n \xrightarrow{P} \mu \quad \text{as } n \rightarrow \infty.$$

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 $\text{Var}(X_i) = \sigma^2 < \infty$. 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 $X_1, X_2, \dots$ 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*

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

Then,

$$\bar{X}_n \xrightarrow{P} \mu \text{ as } n \rightarrow \infty.$$

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 $\bar{X}_n$ 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 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, 1933a]. 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 $X_1, X_2, \dots$ be a sequence of i.i.d. random variables with finite mean $\mu = \mathbb{E}[X_i] < \infty$. Then,*

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

That is, with probability 1,

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

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

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

Weak law of large numbers:

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \Pr(|\bar{X}_n - \mu| > \varepsilon) = 0$$

Strong law of large numbers:

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

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 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.



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

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 first 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 $X_1, X_2, \dots, X_n$ 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 $s_n^2 \rightarrow \infty$ as $n \rightarrow \infty$, and for some $\delta > 0$, the Lyapunov condition holds:

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

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:

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

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 $n \rightarrow \infty$, 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 $X_1, X_2, \dots, X_n$ 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 $s_n^2 \rightarrow \infty$ as $n \rightarrow \infty$.

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 $\epsilon > 0$,

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

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:

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

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:

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

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 $X_1, X_2, \dots$ 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

$$S_n = S_0 + X_1 + X_2 + \dots + X_n, \quad n \geq 1.$$

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 $n \rightarrow \infty$, $W_n(t) \xrightarrow{d} B(t)$, 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 $X_1, X_2, \dots$ 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 $X_1, X_2, \dots$ that takes values in a state space, and has the property that the conditional distribution of X_{n+1} given the past, $X_1, \dots, X_n$ 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 $(X_n, X_{n+1}, \dots, X_{n+k})$ 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 astonishing 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 $X = X_0, X_1, X_2, \dots$ be a Harris ergodic Markov chain with invariant probability distribution π and support $\mathcal{X}$. Let f be a function and define $\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(X_i)$. In a general sense, the CLT states that

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

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

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

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

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

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

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

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

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

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

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

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

or equivalently,

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

This result allows us to construct confidence intervals around the estimate $\bar{\mu}_n$ 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.

These limit theorems also provide part of the mathematical foundation for more sophisticated simulation methods. In particular, analogous laws of large numbers and central limit theorems for Markov chains underpin Markov chain Monte Carlo (MCMC), in which dependent samples are generated to explore complicated probability distributions. MCMC became especially important for Bayesian inference after the 1950s; its historical development, beginning with the Metropolis algorithm, is discussed in Section 7.

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 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 ≈ 365.2422 days
- 1 Lunar month ≈ 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

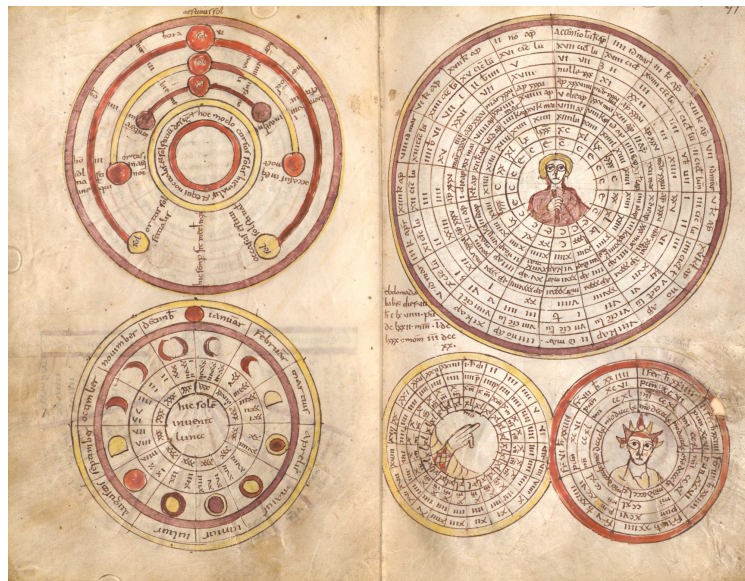


Figure 24: Depiction of the 19 years of the Metonic cycle as a wheel, with the Julian date of the Easter New Moon, from a 9th-century computistic manuscript made in St. Emmeram's Abbey (Clm 14456, fol. 71r)

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

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 $a \leftarrow b, b \leftarrow a \bmod b$, 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:

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

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

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

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

$$252 = 2 \times 105 + 42$$

$$105 = 2 \times 42 + 21$$

$$42 = 2 \times 21 + 0$$

This gives us a sequence 252, 105, 42, 21, 0 and the last non-zero remainder is 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.

```

1 def euclid1(a,b):
2     if b>0:
3         return euclid_algorithm(b, a%b)
4     else:
5         return a
6
7 def euclid2(a,b):
8     while b>0:
9         a, b = b, a%b
10    return a
11
12 euclid3 = lambda a, b: euclid3(b, a%b) if b>0 else a

```

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.



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

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 Peacock* 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 algorithm³. 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 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

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

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:

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = c,$$

where $a_i, c \in \mathbb{Z}$ are fixed integers and x_i are unknowns ($i = 1, \dots, n$). 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,

$$\begin{cases} x + a_iy_i = b_i, & i = 1, \dots, n \end{cases}$$

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 millennium 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 \[2013b\]](#). 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 \[2013b\]](#). 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:

$$x + 3y_1 = 2 \tag{1}$$

$$x + 5y_2 = 3 \tag{2}$$

$$x + 7y_3 = 2 \tag{3}$$

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.

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

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 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:

$$\begin{aligned}d_1 &= 5 \times 7 \times 2 = 3 \times 23 + 1 = 70 \\d_2 &= 7 \times 3 = 5 \times 4 + 1 = 21 \\d_3 &= 3 \times 5 = 2 \times 7 + 1 = 15\end{aligned}$$

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:

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

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 [2013b].

In the study of the kuttaka, the Indian mathematicians made the following important observation about the solvability 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,

$$\begin{aligned}x &= a_1y_1 - b_1 \\x &= a_2y_2 - b_2\end{aligned}$$

we can compare the two equations to obtain $a_1y_1 - b_1 = a_2y_2 - b_2$ which, after rearranging, gives us

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

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.

5.3.5 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 Méziriac** (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:

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

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

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

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

Now reverse Euclid's algorithm:

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

$$\text{From (b)} \quad 3 = 15 - 3 \times 4 \implies 1 = 4 \times 4 - 1 \times 15,$$

$$\text{From (a)} \quad 4 = 64 - 4 \times 15 \implies 1 = 4 \times 64 - 17 \times 15.$$

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

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.

⁵Translation taken from [Datta & Singh \[1962\]](#)

⁶Bachet 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 $ax + by = c$ 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 method⁷ 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 $n_1, \dots, n_r$ be pairwise coprime⁸ natural numbers. Then for any $a_1, \dots, a_r \in \mathbb{N}$, the system of equations*

$$\left\{ x \equiv a_i \pmod{n_i}, \quad i = 1, \dots, r, \right.$$

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

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

$$x = ny = mz$$

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

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

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 $\text{lcm}(m, n)$ 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. \[2000\]](#).

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

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

6 Calculus: A Global History of Change, Motion, and the Infinite

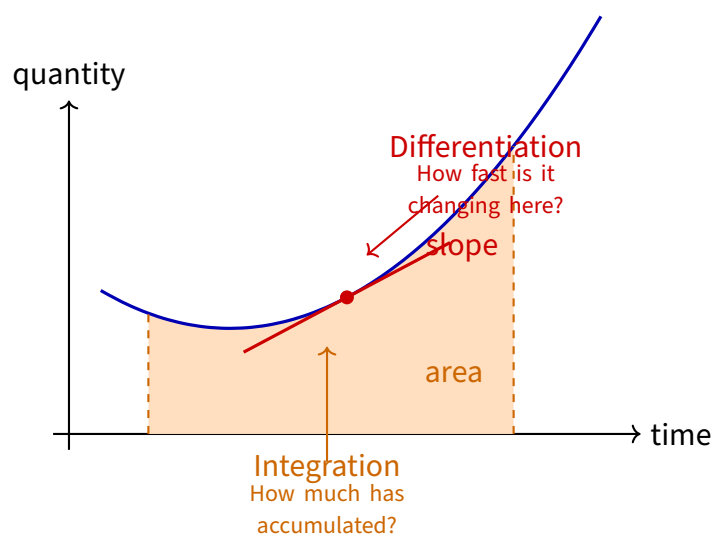


Figure 27: Calculus studies both change and accumulation. Differentiation describes the rate of change at a point, while integration measures the total accumulation over an interval.

6.1 What is calculus?

Two questions at the heart of calculus

Change

$$\left(\frac{ds}{dt}\right)$$

How fast is something changing?

Accumulation

$$\left(\int_a^b v(t) dt\right)$$

How much has built up in total?

A useful way to introduce calculus is to say that it begins with two simple but powerful questions. The first is a question about *change*: if something is moving, growing, cooling, or spreading, how quickly is it changing at a particular moment? The second is a question about *accumulation*: if many small changes happen over time, what is their total effect? These questions are simple enough to appear in everyday life, but they also lead to some of the deepest ideas in mathematics [Strang and Herman \[2016\]](#), [Bressoud \[2019\]](#).

For example, if $(s(t))$ represents the position of an object at time (t) , then the derivative $(s'(t))$ represents its velocity. If $(v(t))$ represents the velocity of the object, then the integral $[\int_a^b v(t), dt]$ represents the displacement of the object between times (a) and (b) . This is one of the reasons calculus became so powerful: it connects local information, such as the speed at one instant, with global information, such as the total distance travelled [Strang and Herman \[2016\]](#), [Edwards \[1979\]](#).

A question for the reader

If we know how fast something is changing at every moment, can we recover how much it changed overall? This question leads to one of the central results of calculus: the *Fundamental Theorem of Calculus*.

Calculus is the branch of mathematics that studies change and accumulation. In simple terms, differentiation helps us describe how fast something is changing, while integration helps us measure how much something has accumulated. These two ideas appear in many familiar situations. For example, if a car is moving, differentiation can be used to describe its speed at a particular moment. If we know the speed of the car over time, integration can be used to calculate the total distance travelled [Strang and Herman \[2016\]](#). Figure 27 gives a simple picture of the two central ideas of calculus: the tangent line represents change at a particular moment, while the shaded region represents accumulation over a period of time.

The word “calculus” comes from the Latin word *calculus*, meaning a small stone used for counting. In the modern mathematical sense, however, calculus is much more than counting. It gives us a language for describing quantities that vary continuously, such as the position of a planet, the growth of a population, the spread of heat, the flow of water, or the changing shape of a curve [Boyer \[1959\]](#), [Edwards \[1979\]](#).

A central idea in calculus is that complicated objects or motions can often be understood by breaking them into very small pieces. If we want to know the area of a curved shape, we can approximate it by many thin rectangles. If we want to understand the motion of an object, we can examine how its position changes over very small intervals of time. This idea of studying the very small in order to understand the whole is one of the reasons calculus became so powerful [Bressoud \[2019\]](#), [Strang and Herman \[2016\]](#).

There are two main parts of elementary calculus:

- **Differential calculus**, which studies rates of change and slopes of curves.
- **Integral calculus**, which studies accumulation, areas, volumes, and total change.

For example, suppose that $s(t)$ represents the position of an object at time (t) . The derivative tells us the instantaneous rate of change of position, which is the velocity of the object. Informally, we can think of this as comparing the change in position with the change in time over smaller and smaller time intervals: $\text{velocity} \approx \frac{\text{change in position}}{\text{change in time}}$. In modern notation, this idea is written using a limit:

$$s'(t) = \lim_{h \rightarrow 0} \frac{s(t+h) - s(t)}{h}.$$

This formula captures the idea of instantaneous change: instead of measuring the average speed over a long time interval, we ask what happens as the time interval becomes extremely small [Strang and Herman \[2016\]](#).

Integration works in the opposite direction. If we know the rate at which something changes, integration allows us to recover the total change. For instance, if $v(t)$ is the velocity of an object, then the distance or displacement over a time interval can be found by adding up all the small contributions from the velocity: $\int_a^b v(t) dt$. This is why integration is often described as a mathematical form of accumulation. It can represent area under a curve, total distance, total growth, total mass, or total probability, depending on the context [Strang and Herman \[2016\]](#), [Bressoud \[2019\]](#).

One of the most important discoveries in calculus is that differentiation and integration are deeply connected. This connection is known as the *Fundamental Theorem of Calculus*.

$$\frac{d}{dx} \left(\int_a^x f(t) dt \right) = f(x).$$

It shows that the problem of finding rates of change and the problem of finding accumulated quantities are, in a precise sense, inverse problems. This is one of the reasons calculus is such a central subject in mathematics, physics, engineering, economics, biology, and many other sciences [Strang and Herman \[2016\]](#), [Edwards \[1979\]](#).

Historically, calculus did not appear suddenly in its modern form. Many of its basic questions are much older than the names and notation we use today. Ancient mathematicians studied areas, volumes, tangents, motion, astronomy, and approximation. Later mathematicians developed methods involving infinitesimals, indivisibles, infinite series, and symbolic notation. Newton and Leibniz created powerful systematic versions of calculus in the seventeenth century, but their work built on a long history of ideas developed across different cultures and periods [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Katz \[2009\]](#), [Cooke \[2013a\]](#).

For this reason, this chapter does not treat calculus only as the achievement of one person, one country, or one moment. Instead, we present calculus as a global and historical development. Its story includes Greek methods of exhaustion, Chinese approximation methods, Islamic work on sums and geometry, Indian work on infinitesimals and infinite series, European developments in symbolic calculus, women mathematicians who helped teach and apply calculus, and modern extensions such as stochastic calculus [Katz \[2009\]](#), [Cooke \[2013a\]](#), [Bressoud \[2019\]](#).

The aim of this chapter is therefore not only to explain what calculus is, but also to show that the history of calculus is a human story. It is a story about people trying to understand change, motion, infinity, and accumulation. It is also a story about communication, notation, education, access, and recognition. By looking at calculus in this wider way, we can better understand both the mathematics itself and the many communities that contributed to its development [Bressoud \[2019\]](#), [Katz \[2009\]](#).

6.2 The beginnings of calculus: exhaustion, approximation, and curved shapes

Whilst calculus as we know it was discovered in the late 17th century, mathematicians had already been ‘doing calculus’ for millennia. We will discuss the **method of exhaustion**, a process for finding the area or volume of curved shapes, which was a precursor to the integral calculus we know today.

The method of exhaustion is generally ascribed to the ancient Greek geometer **Euxodus of Cnidus** (c. 390 BCE - c. 340 BCE), and although an earlier Greek mathematician called Antiphon may have first used the method, Euxodus is credited with perfecting and formalising it [Widgerson \[2019\]](#).

Relatively little is known about Euxodus: he was born in Cnidus in modern-day Turkey, although he would have thought of himself as Greek. Euxodus probably studied under Plato in Plato's 'academy' in Athens; thought to be the first higher education institution in the West, the academy taught subjects including philosophy, mathematics, natural sciences and history. One source reports that during Euxodus's first stay in Athens he was too poor to live near the academy, and so he would walk 11km to attend lectures there [Cooke \[2013b\]](#).

The method of exhaustion works by inscribing a sequence of polygons inside a curved shape, the area of which gets closer and closer to that of the curved shape, in order to show something about the area of the curved shape, usually to find it's area. This same idea can be extended into three dimensions to find volumes.

We will illustrate the method of exhaustion by explaining Euxodus's proof that the ratio between the square of the diameter of two circles is equal to the ratio of their areas, following the exposition in [Widgerson \[2019\]](#). We would find this result trivial, since we can use the formulae for the circumference and area of a circle we are familiar with; however these formulae were only just being discovered by Euxodus and mathematicians of his era.

Proof: Consider any two circles C_1 and C_2 , with areas a_1, a_2 respectively, and diameters d_1, d_2 respectively. We want to show that:

$$\frac{a_1}{a_2} = \left(\frac{d_1}{d_2}\right)^2$$

Suppose otherwise for a contradiction, namely that the above is not an equality, meaning that there exists a circle with diameter d_2 and area x such that:

$$\frac{a_1}{x} < \left(\frac{d_1}{d_2}\right)^2 \quad \text{or} \quad \frac{a_1}{x} > \left(\frac{d_1}{d_2}\right)^2$$

So either $x < a_2$ or $x > a_2$. Let us assume the former.

Notice that for any segment of a circle created by a chord, the triangle created with the vertices being the midpoint of the arc and the two ends of the chord has at least a half the area of the segment (see Figure 28).

This can be shown by constructing a rectangle with one side being the chord of the circle and the other side being equal to the height of the circle. Since the area of the triangle is exactly half that of the rectangle, and the entire segment is contained within the rectangle (by geometric construction) then the triangle must take up at least half of the area of the segment.

We now proceed by inscribing a square within circle C_2 , and then inscribing the circle with a regular octagon by adding a triangle to each side of the square in the manner described in the previous paragraph. Next we inscribe the circle with a regular 16-gon in a similar way and keep on repeating this process. Each time we inscribe a new polygon, the difference in area between the polygon and the circle reduces by at least half by a consequence of the above result. At some point, the difference in area must be less than $a_2 - x$ (we rely of the Archimedean property for this).

Let us call the polygon P_2 and denote its area p_2 . So according to what we have stated above $a_2 - p_2 < a_2 - x$, which means $p_2 > x$.

We repeat this process for the circle C_1 to obtain a similar polygon P_1 with area p_1 . Using the fact that the area of similar polygons is proportional to the square of their height:

$$\frac{p_1}{p_2} = \left(\frac{d_1}{d_2}\right)^2 = \frac{a_1}{x} \quad \implies \quad \frac{x}{p_2} = \frac{a_1}{p_1}$$

But $a_1 > p_1$ since P_1 is contained in C_1 . So by the above equality $x > p_2$, which is a contradiction.

It remains to disprove $x > a_2$. By reversing the roles of the two circles this follows from the previous argument, and we are done. $\square$

Euxodus used this method to show other important results about curved shapes, such as that the ratio of the volume of spheres is proportional to the cube of their diameter, and that the volume of a cone is 1/3 of the volume of a cylinder with the same base and height.

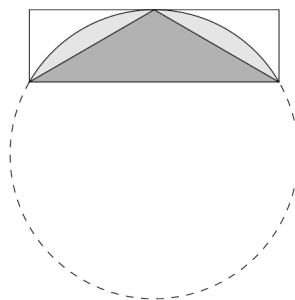


Figure 28: Figure from Widgerson [2019]

Why this matters

Euxodus's method was significant for how it proved and formalised these results. Euxodus related finding the area of a curved shaped to showing that a sequence converges to a limit, in this case the sequence of areas of polygons inscribed in a circle. This formalisation prefigures the eventual formalisation of modern calculus in terms of limits, and calculus's growth into the modern field of analysis in the 19th century.

Archimedes of Syracuse (c. 287 BCE - c. 212 BCE), one of the most well known mathematicians in history, went on to use Euxodus's method of exhaustion to approximate the value of pi as

$$3\frac{10}{71} < \pi < 3\frac{1}{7}$$

by approximating the areas of 96-gons inscribing and circumscribing a circle. In modern calculus this would be similar to evaluating the upper and lower Riemann sums of a curve Widgerson [2019].

However, the ancient Greeks were not the only civilisation to use the method of exhaustion. In the 3rd century CE, the Chinese mathematician Liu Hui used the method of exhaustion to find a slightly tighter bound on the value of pi. Whilst Archimedes circumscribed the circle with a polygon to find an upper bound, using some clever observations Liu Hui approximated pi using only inscribed polygons, going as far as inscribing a circle with a 192-gon Straffin [1998b]. Liu Hui appears to have discovered the method of exhaustion independently from the ancient Greeks. Hence calculus had its beginnings in more than one ancient civilisation.

6.3 Infinitesimals, indivisibles, and tangent problems

Before calculus became a formal subject, mathematicians were already developing methods for solving problems that we would now describe as calculus problems. They wanted to find areas under curves, volumes of curved solids, tangent lines, centres of gravity, and maximum or minimum values. These questions did not yet use the modern language of derivatives and integrals, but they helped prepare the way for the calculus of Newton and Leibniz Boyer [1959], Edwards [1979], Baron [1969].

One of the most important early ideas was the use of quantities that were imagined to be extremely small. These later became known as *infinitesimals*. An infinitesimal is not simply a small number, but a quantity thought of as smaller than any ordinary measurable quantity. Although infinitesimals were not made rigorous in the modern sense until much later, they gave mathematicians a powerful way to reason about curves, motion, and change Boyer [1959], Alexander [2014].

A related idea was the method of *indivisibles*. This method treated a geometric figure as if it were made up of infinitely many very thin slices. For example, an area could be imagined as being composed of many line segments, and a volume as being composed of many plane sections. This approach was developed most famously by Bonaventura Cavalieri in the seventeenth century. Cavalieri's method was not integration in the modern symbolic sense, but it was an important step towards the idea of calculating areas and volumes by summing infinitely many small parts Baron [1969], Edwards [1979]. These ideas were controversial. Some mathematicians found infinitesimals and indivisibles useful, while others worried that they were logically unclear. If a line has no width, how can infinitely many lines form an area? If a plane has no thickness, how can infinitely many planes form a volume? These questions show that the development of calculus was not only about discovering new formulas, but also about debating what mathematical objects really mean Boyer [1959], Alexander [2014].

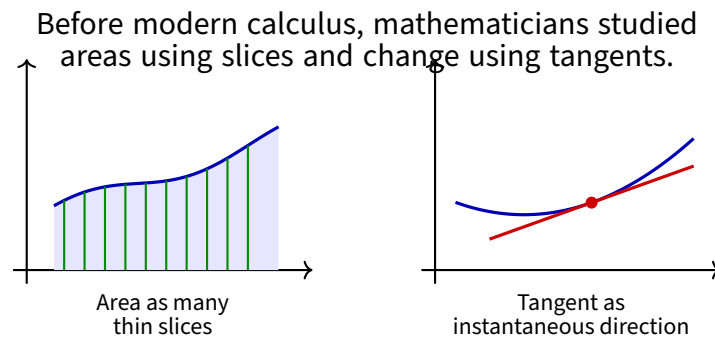


Figure 29: Indivisibles and tangent problems were two important routes towards calculus.

Another important problem was the problem of tangents. Today, we usually understand the tangent to a curve as a line whose slope is given by the derivative. Before derivatives were formally introduced, mathematicians developed different geometric and algebraic methods to find tangent lines. Pierre de Fermat, for example, developed a method called *adequation*, which he used to find tangents and solve maximum and minimum problems. In modern language, some of these methods resemble differentiation, although Fermat did not formulate them using the modern concept of a derivative [Edwards \[1979\]](#), [Baron \[1969\]](#).

René Descartes also contributed to this development by linking algebra and geometry through coordinate methods. This allowed curves to be studied using equations, making it possible to use algebraic techniques to investigate geometric questions. The combination of algebra and geometry was essential for the later development of calculus, because it made curves easier to describe, compare, and manipulate symbolically [Katz \[2009\]](#), [Cooke \[2013a\]](#).

Isaac Barrow, who was Newton's teacher at Cambridge, also played an important role in the prehistory of calculus. His work connected tangent problems with quadrature problems, where quadrature means finding areas. This connection was crucial because it pointed towards what later became the *Fundamental Theorem of Calculus*: the idea that differentiation and integration are inverse processes. Barrow did not present calculus in the same systematic way as Newton or Leibniz, but his work helped bring together two central problems: finding slopes and finding areas [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Dunham \[2005\]](#).

This period is important because it shows that calculus did not emerge from a single moment of inspiration. Instead, it developed through many partial methods, arguments, and debates. Infinitesimals, indivisibles, tangent methods, and maximum–minimum problems all contributed to the creation of calculus. By the time Newton and Leibniz produced their systematic approaches, many of the essential ingredients were already present [Baron \[1969\]](#), [Edwards \[1979\]](#), [Bressoud \[2019\]](#).

6.4 Medieval Islamic mathematics and the prehistory of integration

Between the ancient Greek methods of exhaustion and the seventeenth-century calculus of Newton and Leibniz, mathematicians across the world continued to study problems about area, volume, curves, optics, astronomy, and algebra. One important centre of this work was the medieval Islamic world, where scholars wrote in Arabic, Persian, and other languages, and where mathematical knowledge from Greek, Indian, Persian, and other traditions was translated, studied, criticised, and extended [Joseph \[2011\]](#), [Katz \[2009\]](#), [Robson and Stedall \[2009\]](#).

This is important for the story of calculus because calculus did not appear suddenly. Some of its central ideas grew out of older questions:

How can many small pieces be added together?

How can curved areas and volumes be measured?

How can we locate a maximum or minimum?

These questions were not yet written in the language of derivatives and integrals, but they belong to the long prehistory of calculus.

One of the most important figures for this section is Ibn al-Haytham, also known in Latin as Alhazen. He was born in Basra around 965 and later worked in Egypt. He is especially famous for his work on optics, but he also made important mathematical contributions. In particular, he studied sums of powers, such as



Figure 30: A map from the *Book of Curiosities* (*Kitāb Gharāib al-funūn wa-mulāḥ al-uyūn*), a medieval Arabic cosmographical manuscript. The manuscript was compiled between about 1020 and 1050 and contains important maps and astronomical diagrams. Image source: Bodleian Library, University of Oxford, MS Arab. c. 90.

$$1^k + 2^k + 3^k + \cdots + n^k.$$

These sums may look like arithmetic, but they are closely connected to integration. To calculate the area or volume of a curved object, one can divide it into many thin pieces, add the pieces together, and then imagine what happens as the pieces become smaller and more numerous.

From sums to integrals

A modern integral can be understood as the limit of many small sums:

$$\int_a^b f(x) dx \approx f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x.$$

This is why historical work on sums of powers is part of the background to integral calculus.

Ibn al-Haytham used formulas for sums of powers in connection with volume calculations, including problems involving paraboloids. In modern language, this resembles the idea of adding many slices to find a volume. However, we should describe this carefully. Ibn al-Haytham did not write integral calculus in the notation of Newton or Leibniz. It is better to say that his work forms part of the *prehistory of integral calculus* Katz [1995], [Mathematical Association of America](#) [2014].



Figure 31: (Left) Ibn al-Haytham is best known for optics, but his mathematical work on sums and volumes also belongs to the wider history of calculus. (Right) The structure of the human eye according to Ibn al-Haytham.

A useful way to explain the connection is through a simple comparison. If we want to approximate the area under the curve $y = x^2$ from 0 to n , we might begin by adding square values:

$$1^2 + 2^2 + 3^2 + \cdots + n^2.$$

For higher-dimensional volume problems, higher powers appear:

$$1^4 + 2^4 + 3^4 + \dots + n^4.$$

This does not mean that a sum of powers is already an integral. Rather, it shows that the same basic idea appears again and again in the history of mathematics: complicated curved quantities can be approached by adding many simpler quantities.

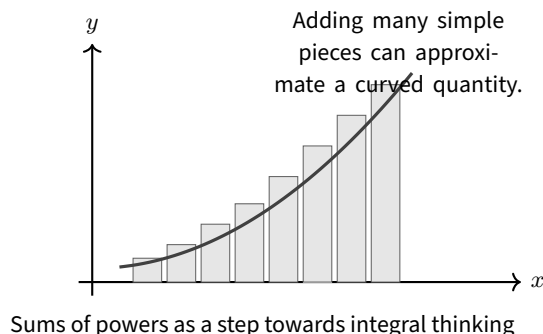


Figure 32: A visual way to connect sums of powers with the later idea of integration.

Another important figure is Sharaf al-Din al-Tusi, a Persian mathematician of the twelfth century. Al-Tusi studied cubic equations and the conditions under which they have positive solutions. In modern notation, one of the examples often discussed can be written in the form

$$x^3 + a = bx,$$

or equivalently,

$$bx - x^3 = a.$$

To understand whether such an equation has a positive solution, it is useful to examine the maximum value of the curve

$$y = bx - x^3.$$

From a modern point of view, we would locate the maximum by differentiating:

$$\frac{dy}{dx} = b - 3x^2.$$

Setting this equal to zero gives

$$b - 3x^2 = 0, \quad x = \sqrt{\frac{b}{3}}.$$

This kind of calculation looks very much like calculus to us. But al-Tusi did not write it in this modern symbolic language. His reasoning belonged to the algebraic and geometric traditions of his own time. For this reason, historians are careful when describing his work. Some emphasise its analytical or derivative-like features, while others stress that it should not simply be translated into modern calculus [Hogendijk \[1989\]](#), [Farès \[1995\]](#).

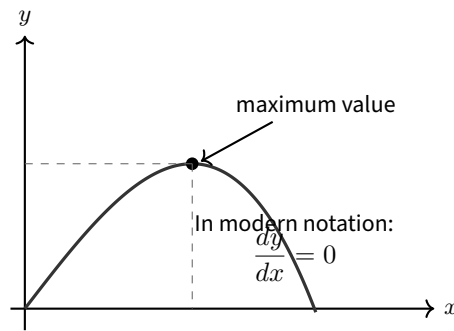
This example is valuable for students because it shows how careful we need to be with history. It would be too simple to say that al-Tusi “had derivatives” in the same way that a modern calculus textbook does. But it would also be unfair to ignore the sophistication of his work. His study of cubic equations shows that mathematicians in the medieval Islamic world were developing powerful methods for analysing curves, equations, and maximum values.

Careful historical language

Instead of saying ‘Ibn al-Haytham invented integration’ or ‘al-Tusi invented derivatives’, it is more accurate to say:

Their work belongs to the prehistory of calculus.

This wording recognises the importance of their ideas without forcing them into modern notation too quickly.



Al-Tusi studied cubic equations and maximum conditions.

Figure 33: A stylised modern visualisation of a maximum problem related to al-Tusi’s study of cubic equations.

The medieval Islamic world therefore forms an important bridge in this chapter. It connects older geometric methods with later algebraic, analytic, and symbolic developments. It also reminds us that the history of calculus is not simply a European story. It is a history of translation, preservation, criticism, extension, and new mathematical invention across cultures.

6.5 Indian mathematics and infinite series

Indian mathematics has an important place in the history of ideas that later became central to calculus. Long before calculus was written in the notation of Newton and Leibniz, Indian mathematicians were studying motion, astronomy, trigonometry, approximation, and extremely small changes. These topics were not only abstract: they were closely connected to practical and scientific questions, especially the need to calculate planetary positions and produce accurate astronomical tables Plofker [2009b], Katz [1995].

One important figure is Bhāskara II, also known as Bhāskarācārya, who lived in the twelfth century. Bhāskara worked on arithmetic, algebra, astronomy, and trigonometry. Some of his arguments involved reasoning with very small changes in quantities. For example, in his astronomical work he considered small variations in trigonometric functions, in a way that can be compared with ideas that later became important in differential calculus Plofker [2009b], Katz [1995].

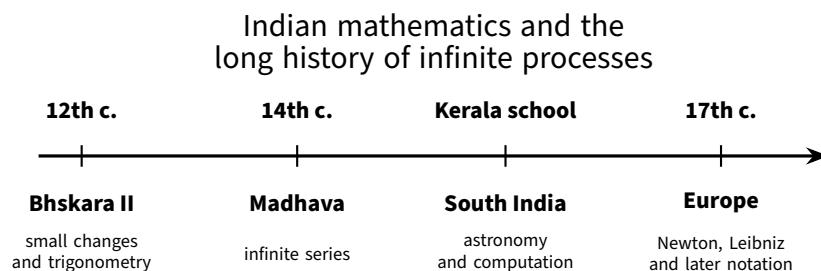


Figure 34: A simplified timeline showing Indian contributions to ideas connected with calculus.

The work of the Kerala school can be made visually striking by showing one of the infinite series in modern notation. For instance,

$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$$

This kind of formula is powerful because it turns a trigonometric quantity into an infinite algebraic process. Even though the Kerala mathematicians did not write the formula using modern notation, their work shows a deep understanding of approximation and infinite procedures Katz [1995], Plofker [2009b].

Why this matters

If the story of calculus moves directly from Greek geometry to Newton and Leibniz, then the work of Bhaskara II, Madhava, and the Kerala school disappears. Including them changes the story: calculus becomes part of a wider global history of astronomy, computation, approximation, and infinity.

It is important to describe this carefully. Bhāskara did not have the modern notation of derivatives, nor did he present a general theory of differential calculus in the way it is taught today. However, his work shows that mathematicians in India were already thinking about change, approximation, and the behaviour of functions. In particular, some historians have compared parts of his work to later ideas about rates of change and extrema. This makes Bhāskara a useful example of how calculus-like ideas can appear before calculus exists as a formal subject Boyer [1959], Plofker [2009b].

An even more striking development came from the Kerala school of mathematics and astronomy, which flourished in South India from around the fourteenth century. The figure most closely associated with the beginning of this tradition is Madhava of Sangamagrama. Madhava and later Kerala mathematicians developed infinite series for trigonometric functions, including sine, cosine, and arctangent. They also produced series and approximations connected to (π) Plofker [2009b], Katz [1995], Joseph [2011].

This does not mean that the Kerala school had exactly the same calculus as Newton and Leibniz. Their work was expressed in a different mathematical language, with different aims and methods. It was strongly connected to astronomy and computation, rather than to the European mechanical problems that shaped Newton's calculus. Nevertheless, the Kerala school developed highly sophisticated techniques involving infinite processes, approximation, and correction terms. These are precisely the kinds of ideas that later became central to calculus and analysis Plofker [2009b], Joseph [2011].

This part of the story is important for an inclusive history of mathematics. Many standard histories of calculus move quickly from Greek geometry to seventeenth-century Europe. When this happens, major developments in Indian mathematics can disappear from the story. Including Bhāskara, Madhava, and the Kerala school helps us see that mathematical creativity was not confined to one region or one tradition. It also reminds us that the history of mathematics depends on which texts are preserved, translated, taught, and valued Joseph [2011], Robson and Stedall [2009].

The story of Indian mathematics and infinite series therefore changes how we understand calculus. Instead of seeing calculus as a sudden European invention, we can see it as part of a much longer global history of people trying to understand change, approximation, infinity, and the heavens.

6.6 Newton and Leibniz: two systematic calculi

Isaac Newton and Gottfried Wilhelm Leibniz are often described as the founders of calculus. Their work was revolutionary, but it is important to place them in a wider story. They did not create calculus out of nothing. By the seventeenth century, many mathematicians had already studied tangents, areas, maxima and minima, infinitesimals, indivisibles, and infinite series. Newton and Leibniz brought many of these earlier ideas together and turned them into two powerful and systematic ways of doing calculus Boyer [1959], Edwards [1979], Baron [1969].



Figure 35: Newton (left) and Leibniz (right) developed different forms of calculus in the seventeenth century.

Newton developed his calculus in the language of motion. He called changing quantities *fluents*, because he imagined them as flowing or varying with time. The rate at which a fluent changed was called a *fluxion*. In modern notation, if x is a changing quantity, Newton's fluxion would be written as

$$\dot{x}.$$

This notation is still used today, especially in physics, when a quantity changes with time. For example, if $x(t)$ is the position of a moving object, then

$$\dot{x}(t) = \frac{dx}{dt}$$

represents its velocity, and

$$\ddot{x}(t) = \frac{d^2x}{dt^2}$$

represents its acceleration. This shows why Newton's version of calculus was so closely connected with mechanics, astronomy, and the motion of planets and objects under gravity Boyer [1959], Edwards [1979], Dunham [2005].

Leibniz developed calculus in a different style. Instead of beginning mainly from motion, he focused on differences between quantities. His notation used symbols such as

$$dx, \quad dy, \quad \frac{dy}{dx}, \quad \int.$$

These symbols are still central in calculus today. The expression

$$\frac{dy}{dx}$$

suggests a ratio between a small change in y and a small change in x . The integral sign $\int$ comes from the idea of a sum: adding many small pieces to obtain a total quantity. This notation was one of Leibniz's great achievements. It made calculus easier to write, teach, and extend Boyer [1959], Edwards [1979], Bressoud [2019].

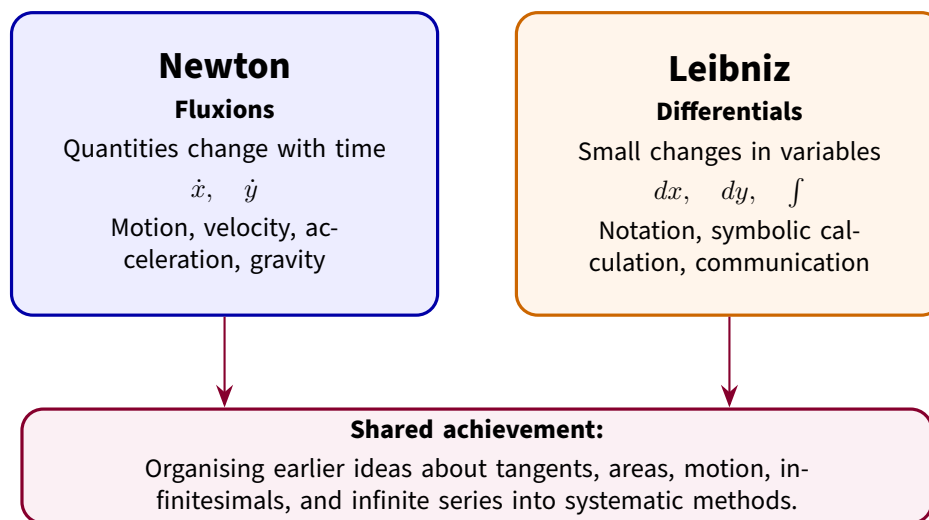


Figure 36: Newton and Leibniz developed different but powerful ways of organising calculus.

The difference between Newton and Leibniz can be seen in a simple example. Suppose a quantity y depends on another quantity x . Leibniz's notation writes the rate of change as

$$\frac{dy}{dx}.$$

Newton's notation is especially natural when x and y change with time. If $x = x(t)$, then the rate of change is written as

$$\dot{x}.$$

Both notations describe change, but they encourage slightly different ways of thinking. Newton's notation emphasises motion through time. Leibniz's notation emphasises relationships between variables. This is one reason both notations survived: each is useful in different contexts.

One of the most important achievements of calculus was the realisation that two old problems were deeply connected. The first was the tangent problem: how can we find the slope of a curve at one point? The second was the quadrature problem: how can we find the area under a curve? In modern language, these are the problems of differentiation and integration.

The Fundamental Theorem of Calculus

Calculus connects two old problems:

finding slopes $\longleftrightarrow$ finding areas.

If

$$F(x) = \int_a^x f(t) dt,$$

then

$$F'(x) = f(x).$$

This means that the rate of change of an accumulated quantity gives back the original function.

This result is known as the *Fundamental Theorem of Calculus*. It explains why differentiation and integration are not separate subjects. Differentiation breaks a changing process down to its rate of change at a moment. Integration adds small pieces together to recover a total quantity. Newton's work made this connection natural through motion and changing quantities, while Leibniz's notation made the relationship easier to write and use [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Baron \[1969\]](#), [Bressoud \[2019\]](#).

For example, if $v(t)$ is the velocity of an object, then its displacement from time a to time b is

$$\int_a^b v(t) dt.$$

If this accumulated displacement is written as

$$S(x) = \int_a^x v(t) dt,$$

then the Fundamental Theorem of Calculus tells us that

$$S'(x) = v(x).$$

In words: if we add up velocity to get distance, then differentiating that distance gives us velocity again. This is one of the reasons calculus became such a powerful language for science.

Newton and Leibniz developed their methods independently, but a famous priority dispute later arose over who had invented calculus first. Newton had developed many of his ideas earlier, but Leibniz published his work first. The dispute became intense and was shaped not only by mathematics, but also by national pride, personal rivalry, publication practices, and the authority of scientific institutions [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Dunham \[2005\]](#).

For this chapter, the priority dispute should not be the main focus. A more interesting question is why these two versions of calculus became so influential. Newton's approach was deeply connected to the mathematical study of nature. Leibniz's approach gave calculus a notation that could be copied, taught, and developed by others. Together, their work helped transform scattered methods into a general mathematical subject [Bressoud \[2019\]](#), [Edwards \[1979\]](#).

This story also helps us think about recognition in mathematics. Who gets credit for an idea often depends on more than who had the idea first. It can depend on who publishes, who has access to universities and academies, whose language becomes dominant, whose notation is adopted, and whose work is preserved and taught. This matters especially in the history of calculus, because many earlier contributions from Greek, Islamic, Indian, Chinese, and European mathematicians formed part of the background from which Newton and Leibniz worked [Katz \[2009\]](#), [Cooke \[2013a\]](#), [Joseph \[2011\]](#), [Robson and Stedall \[2009\]](#).

Newton and Leibniz should therefore not be removed from the story, but they should not be treated as isolated geniuses either. Their achievement was to organise, extend, and communicate powerful methods for studying change and accumulation. Calculus itself belongs to a much broader human story, shaped by many cultures, many problems, and many people across time.

6.7 Calculus, education, and women mathematicians

After Newton and Leibniz systematised calculus, the subject continued to grow rapidly. Mathematicians worked on making calculus more precise, applying it to the natural sciences, and turning it into a subject that could be taught to new generations of students. Calculus was no longer only a tool for a small number of specialists. It became part of textbooks, lectures, correspondence, translations, and scientific education [Robson and Stedall \[2009\]](#), [Osen \[1974\]](#), [Henrion \[1997\]](#).

This part of the story is important because mathematics does not spread by formulas alone. It spreads through people: through teachers, writers, translators, students, and communities. In this section, we focus on three women who were connected with the teaching, circulation, and application of calculus: Maria Gaetana Agnesi, Mary Somerville, and Sophie Germain [Mazzotti \[2007\]](#), [Stenhouse \[2020\]](#), [Bucciarelli and Dworsky \[1980\]](#).



Figure 37: Agnesi, Somerville, and Germain show three different ways women contributed to the history of calculus: through teaching, communication, and application.

Maria Gaetana Agnesi: teaching calculus

Maria Gaetana Agnesi was born in Milan in 1718. Her family was wealthy and provided her with an unusually strong education for a woman of her time. From a young age she became known for her knowledge of languages, philosophy, and mathematics. Her father organised intellectual gatherings where Agnesi could present and discuss advanced topics with scholars and visitors. These opportunities were exceptional, but they still took place in a society where women rarely had full access to universities, academies, or professional mathematical careers [Truesdell \[1989\]](#), [Mazzotti \[2007\]](#), [Mayfield \[2013\]](#).

Agnesi's most important mathematical work was her 1748 textbook,

Instituzioni analitiche ad uso della gioventù italiana.

The title can be translated roughly as *Analytical Institutions for the Use of Italian Youth*. The book presented algebra, analytic geometry, differential calculus, and integral calculus in a systematic way. It was one of the important eighteenth-century texts for organising calculus as a subject that could be studied and taught, rather than only followed through scattered papers, private notes, and correspondence [Agnesi \[1748\]](#), [Mathematical Association of America \[2018\]](#), [Mayfield \[2013\]](#).

Why Agnesi matters

Agnesi's importance does not only lie in a single formula or curve. Her contribution also lies in explanation, organisation, and education. She helped make calculus more accessible as a teachable subject.

Agnesi is also remembered for a curve now known in English as the *Witch of Agnesi*. The name is partly the result of a mistranslation of the Italian word *versiera*. In modern notation, one version of the curve can be written as

$$y = \frac{a^3}{x^2 + a^2}.$$

This curve is useful in the booklet because it gives a visual connection between Agnesi's name and the analytic geometry that was closely connected with calculus [Mathematical Association of America \[2018\]](#), [Katz \[2011\]](#).

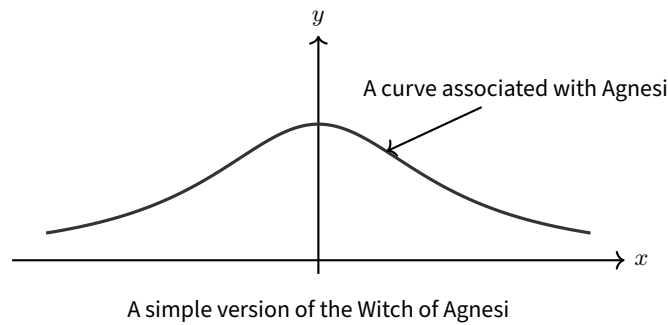


Figure 38: A modern drawing of a curve associated with Maria Gaetana Agnesi.

Agnesi was later offered a chair at the University of Bologna by Pope Benedict XIV, although she did not take up regular teaching duties there. Her story is therefore both inspiring and complicated. It shows that women could sometimes be recognised for mathematical ability, but also that recognition did not necessarily mean full participation in academic life [Mazzotti \[2007\]](#), [Università di Bologna](#), [Mathematical Association of America \[2018\]](#).

Mary Somerville: communicating mathematical physics

Mary Somerville was born in Scotland in 1780. Unlike Agnesi, she did not receive a formal mathematical education from childhood. Much of her learning came through private study. This is important because many women entered mathematics through indirect routes, often outside the institutions that were open to men [O'Connor and Robertson \[c\]](#), [Stenhouse \[2020\]](#).

Somerville became one of the most important scientific writers of the nineteenth century. She helped make advanced mathematical ideas more accessible to English-speaking readers, especially ideas coming from continental Europe. Her work connected British audiences with mathematical physics, astronomy, and the analytical methods associated with calculus [Stenhouse \[2020\]](#), [O'Connor and Robertson \[c\]](#).

A central example is her engagement with the work of Pierre-Simon Laplace. Laplace's celestial mechanics used advanced mathematics to describe the motion of planets and gravitational systems. Somerville's *Mechanism of the Heavens* was based on Laplace's work, but it was not merely a simple translation. She explained and expanded difficult mathematical material for readers who were not already familiar with the continental analytical style [Somerville \[1831\]](#), [O'Connor and Robertson \[c\]](#), [Stenhouse \[2020\]](#).

calculus $\longrightarrow$ mechanics $\longrightarrow$ astronomy

This chain of ideas was central to nineteenth-century mathematical physics. Calculus gave mathematicians tools for describing motion. Mechanics applied those tools to forces and moving bodies. Astronomy used them to study planets, orbits, and the structure of the heavens [Somerville \[1831\]](#), [Stenhouse \[2020\]](#).

Communication is mathematical work

Somerville's story reminds us that explaining mathematics matters. Translation, exposition, and teaching are part of how mathematics grows.

Somerville is important for this chapter because she shows that the history of calculus is also a history of access. Who could study advanced mathematics? Who could read the latest work? Who could translate it into another language or another style? These questions shaped the development of calculus just as much as formal discoveries did [Stenhouse \[2020\]](#), [Robson and Stedall \[2009\]](#).

Sophie Germain: applying calculus to elasticity

Sophie Germain was born in Paris in 1776. She developed her mathematical interests largely through private study, at a time when women were excluded from many formal scientific institutions. To access mathematical education and correspondence, she sometimes used the male name "Monsieur LeBlanc". This detail is often mentioned because it shows the barriers women faced when trying to participate in mathematical communities [O'Connor and Robertson \[a\]](#), [Osen \[1974\]](#).

Germain made important contributions to number theory, but for the history of calculus she is especially relevant because of her work on elasticity. Elasticity studies how objects bend, stretch, and vibrate. These problems require calculus because they involve quantities changing across space and time [Bucciarelli and Dworsky \[1980\]](#), [Timoshenko \[1953\]](#), [O'Connor and Robertson \[a\]](#).

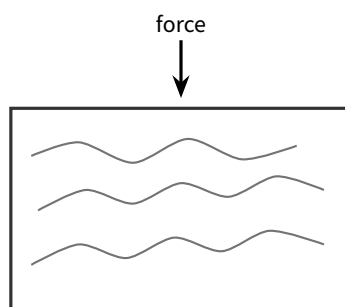
A simplified way to represent this kind of problem is

force $\longrightarrow$ deformation $\longrightarrow$ motion or vibration.

In more advanced mathematical language, the study of vibrating plates involves differential equations. A simplified symbolic form is

$$\nabla^4 w = \text{applied force},$$

where w describes the displacement of the plate. This formula is not meant as a full physical model here, but it gives a sense of the type of mathematics involved: calculus is used to describe how a surface bends and changes [Bucciarelli and Dworsky \[1980\]](#), [Timoshenko \[1953\]](#).



Calculus can describe bending and vibration

Figure 39: A simple visualisation of a vibrating plate, a problem connected with Sophie Germain's work on elasticity.

Germain's work shows another way calculus developed after Newton and Leibniz. It was not only made more rigorous in textbooks; it was also applied to new physical problems. The study of elasticity connected calculus with engineering, sound, vibration, and the behaviour of materials [Bucciarelli and Dworsky \[1980\]](#), [Timoshenko \[1953\]](#).

Together, Agnesi, Somerville, and Germain help us tell a wider story of calculus. Agnesi shows the importance of teaching and organising the subject. Somerville shows the importance of communicating and circulating advanced mathematical ideas. Germain shows the importance of applying calculus to new scientific problems. Their stories also remind us that the history of mathematics is shaped not only by talent, but also by access, education, institutions, and recognition [Robson and Stedall \[2009\]](#), [Osen \[1974\]](#), [Henrion \[1997\]](#).

6.8 From calculus to analysis

By the eighteenth century, calculus had become one of the most powerful tools in mathematics. It could describe the motion of planets, the fall of objects, the bending of materials, the shape of curves, and the behaviour of infinite series. But there was still a problem: mathematicians did not always agree on why calculus worked.

The early calculus of Newton and Leibniz used ideas such as infinitesimals, fluxions, differentials, and infinite sums. These ideas were extremely useful, but they also raised difficult questions.

What needed to be made precise?

- What exactly is an infinitesimal?
- What does it mean for a quantity to approach a limit?
- When does an infinite series really add up to a finite answer?
- When can we trust a derivative or an integral?

One of the great figures of eighteenth-century mathematics was Leonhard Euler. Euler used calculus with extraordinary confidence and creativity. He worked with functions, infinite series, differential equations, mechanics, number theory, and many other areas. A simple rule such as

$$\frac{d}{dx}x^n = nx^{n-1}$$

shows the kind of symbolic power that calculus had developed. Rules like this allowed mathematicians to calculate quickly and solve new problems. However, the success of calculus also made its foundations more urgent: if these methods worked so well, mathematicians needed to understand exactly why they worked [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Dunham \[2005\]](#).

During the nineteenth century, calculus gradually became the more rigorous subject now called *mathematical analysis*. One of the key figures in this change was Augustin-Louis Cauchy. Cauchy placed much more emphasis on the idea of a *limit*. Instead of relying only on intuitive ideas of infinitely small quantities, he tried to explain calculus through quantities becoming closer and closer to a value.

For example, the derivative can be written as

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

This expression says that the slope of a curve at a point is obtained by looking at the slopes of secant lines over smaller and smaller intervals. The language of limits made calculus more precise and helped mathematicians explain what was happening behind the calculation [Grabiner \[1981\]](#), [Bressoud \[2019\]](#).

Later in the nineteenth century, Karl Weierstrass pushed this precision even further. He promoted a more careful approach to limits, continuity, and derivatives. In modern notation, continuity can be expressed using the famous ε - δ language:

$$\forall \varepsilon > 0, \quad \exists \delta > 0 \quad \text{such that} \quad |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon.$$

This looks technical, but the idea is simple. If we want $f(x)$ to be very close to $f(a)$, then we can make that happen by choosing x close enough to a . This gave mathematicians a way to discuss smoothness and change without depending on vague pictures or unclear infinitesimals [Grabiner \[1981\]](#), [Dunham \[2005\]](#).

The integral was also made more precise. Bernhard Riemann developed a way of defining integration by dividing an interval into small pieces, forming sums, and then taking a limit. In modern notation, this idea is often written as

$$\int_a^b f(x) dx = \lim_{\max \Delta x_i \rightarrow 0} \sum_i f(x_i^*) \Delta x_i.$$

This formula captures a familiar idea: the area under a curve can be approximated by many thin rectangles. As the rectangles become thinner, the approximation becomes closer to the true area. This connects modern integration back to the much older ideas of exhaustion, approximation, and indivisibles discussed earlier in this chapter [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Dunham \[2005\]](#).

At the beginning of the twentieth century, Henri Lebesgue introduced a more general theory of integration. The Riemann integral works well for many functions, especially the ones usually seen in elementary calculus. Lebesgue's theory allowed mathematicians to integrate more complicated functions and became essential in modern analysis, probability, Fourier analysis, and later areas of mathematics. In modern notation, a Lebesgue integral is often written as

$$\int f d\mu,$$

where μ represents a measure. This notation reflects a new way of thinking about integration: not only as area under a curve, but as a more general method for measuring accumulated quantities [Dunham \[2005\]](#), [Bressoud \[2019\]](#).

This development was not only mathematical. It was also shaped by culture, language, notation, and institutions. In much of continental Europe, Leibniz's differential notation was widely adopted and became a flexible language for calculation and teaching. In Britain, Newton's prestige and the priority dispute between Newton and Leibniz contributed to the continued use of fluxional notation for a long time. This affected how calculus was taught and how easily British mathematicians communicated with developments on the continent [Boyer \[1959\]](#), [Edwards \[1979\]](#), [Cooke \[2013a\]](#).

In the nineteenth century, British mathematicians such as Charles Babbage, George Peacock, and John Herschel helped promote the use of continental notation and methods through the Analytical Society at Cambridge. This was not just a change of symbols. It was also a change in mathematical culture: a movement towards engaging more directly with European analysis and its newer techniques [Grattan-Guinness \[1980\]](#), [Bottazzini \[1986\]](#).

American mathematics also developed through international exchange. In the nineteenth century, many American students and scholars looked to European universities, especially in Germany and France, for advanced mathematical training. Ideas about rigorous analysis, research seminars, journals, and professional mathematical communities travelled through these networks. This reminds us that the development of analysis was not only about individuals proving theorems, but also about how mathematics was taught, circulated, translated, and institutionalised [Parshall and Rowe \[1994\]](#), [Robson and Stedall \[2009\]](#).

A wider lesson

The move from calculus to analysis shows that mathematics is not born in its final form. Ideas can be useful before they are fully understood. Over time, mathematicians refine definitions, repair arguments, create new notation, and build stronger foundations. The history of analysis is therefore also a history of communication, education, and changing standards of proof.

This section also connects back to the inclusive aim of the chapter. When we ask how calculus became analysis, we are not only asking about formulas. We are also asking who had access to universities, books, translations, journals, and mathematical communities. The development of analysis depended on people, institutions, languages, and networks. This is another reason why the history of calculus should not be told as the achievement of only one person or one country.

6.9 Modern calculus and new applications

Calculus did not stop developing after Newton, Leibniz, Cauchy, Weierstrass, Riemann, or Lebesgue. The same basic questions are still alive today:

How does something change? How can we measure its total effect?

In modern mathematics, these questions appear in many new settings: vibrating materials, moving planets, spreading diseases, changing climates, financial markets, biological populations, and random motion. This is one of the most important messages of this chapter: calculus is not only a historical achievement, but a living subject that continues to grow [Bressoud \[2019\]](#), [Edwards \[1979\]](#).

One important modern development is the study of *differential equations*. A differential equation is an equation involving a quantity and its rate of change. For example, a very simple model of growth can be written as

$$\frac{dN}{dt} = rN,$$

where $N(t)$ is the size of a population and r is its growth rate. This equation says that the rate of change of the population depends on the current population size. Differential equations are used to model motion, heat, waves, disease spread, population dynamics, and many other processes that evolve over time [Boyce et al. \[2017\]](#), [Tenenbaum and Pollard \[1985\]](#).

Calculus also became central to physics and engineering through the study of materials. The theory of elasticity, for example, studies how objects bend, stretch, and vibrate. This connects with the work of Sophie Germain, who studied vibrating plates in the early nineteenth century. Her work belongs to the history of calculus and differential equations, but also to the history of exclusion: Germain worked at a time when women were largely excluded from universities and scientific academies. Her story shows that mathematical talent often existed even where formal opportunities were denied [O'Connor and Robertson \[a\]](#), [Timoshenko \[1953\]](#).

Calculus as a tool for modelling

A modern mathematical model often begins with a rate of change:

$$\frac{d}{dt}(\text{quantity}) = \text{rule for how it changes.}$$

This simple idea appears in physics, biology, engineering, economics, medicine, climate science, and many other fields.

Another major modern direction is *numerical calculus*. In many real problems, exact answers are impossible or too difficult to find by hand. Instead, mathematicians and scientists use algorithms to approximate derivatives, integrals,

and solutions of differential equations. This is one reason calculus is so important in computing. Numerical methods allow us to simulate weather, design bridges, model epidemics, study galaxies, and solve engineering problems [Burden and Faires \[2011\]](#), [Sauer \[2017\]](#).

Calculus is also essential in *optimisation*, the study of finding the best possible value of something. Optimisation asks questions such as: What is the shortest route? What is the strongest design? What is the lowest cost? What is the best fit to the data? These questions often depend on derivatives, because derivatives tell us how a quantity changes when we adjust an input. Today, optimisation is used in engineering, economics, data science, machine learning, medicine, and many other areas [Nocedal and Wright \[2006\]](#).

A particularly important twentieth-century development was *stochastic calculus*, a form of calculus designed for randomness. Ordinary calculus works very well when a path is smooth enough to have a clear tangent. But many real processes are not smooth. The motion of small particles in a fluid, fluctuations in financial markets, and noisy biological systems can move in irregular and unpredictable ways. To study this kind of motion, mathematicians needed new tools [Itô \[1951\]](#), [Øksendal \[2003\]](#).

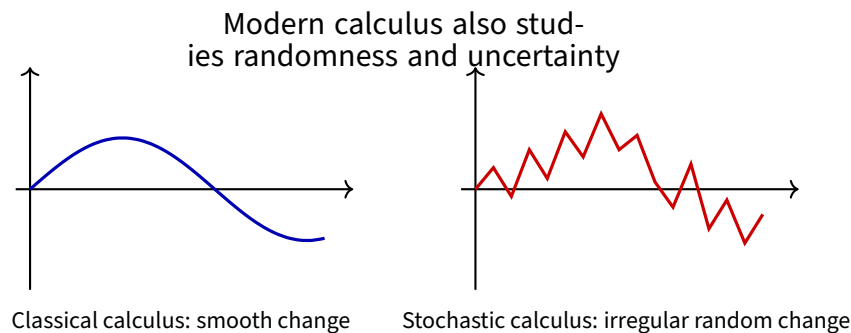


Figure 41: Classical calculus often studies smooth change, while stochastic calculus studies irregular change affected by randomness.

Figure 41 shows the basic difference. A smooth curve has a clear direction at each point, so ordinary derivatives make sense. A random path, such as Brownian motion, is much rougher. It changes so irregularly that the usual idea of a tangent line breaks down.

One of the central figures in this new kind of calculus was the Japanese mathematician Kiyosi Itô. Itô developed a theory of stochastic integrals and stochastic differential equations, now known as *Itô calculus*. A stochastic differential equation is like an ordinary differential equation, but it includes a random term. A simple example is

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dW_t.$$

Here X_t is the changing quantity, dt represents a small change in time, and dW_t represents a small random change coming from Brownian motion. The function μ is often called the *drift*; it describes the predictable part of the motion. The function σ is often called the *diffusion*; it describes the size of the random fluctuations [Itô \[1951\]](#), [Itô and McKean \[1965\]](#), [Øksendal \[2003\]](#).

One surprising feature of Itô calculus is that the usual rules of calculus have to be modified. In ordinary calculus, very small squared quantities are usually ignored. But for Brownian motion, the random change dW_t is so irregular that its square behaves like dt :

$$(dW_t)^2 = dt.$$

This leads to a new version of the chain rule, called *Itô's formula*. If X_t satisfies

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dW_t,$$

then for a smooth function f ,

$$df(X_t) = f'(X_t) dX_t + \frac{1}{2} f''(X_t) \sigma^2(X_t, t) dt.$$

Equivalently,

$$df(X_t) = \left[\mu(X_t, t) f'(X_t) + \frac{1}{2} \sigma^2(X_t, t) f''(X_t) \right] dt + \sigma(X_t, t) f'(X_t) dW_t.$$



Figure 42: Kiyosi Itô developed stochastic calculus, extending calculus to random processes.

The extra term

$$\frac{1}{2} f''(X_t) \sigma^2(X_t, t) dt$$

is what makes Itô calculus different from ordinary calculus. It appears because random paths fluctuate too much for the usual chain rule to work without correction Øksendal [2003], Mikosch [1998].

What changed with Itô calculus?

Classical calculus studies smooth change:

$$dx = \text{predictable change.}$$

Itô calculus studies change with randomness:

$$dX_t = \text{predictable change} + \text{random fluctuation.}$$

This made it possible to use calculus in systems affected by uncertainty.

Itô's work is now used in many fields. In physics, it helps describe random motion and diffusion. In biology, it can be used in models of noisy population dynamics and systems affected by random environments. In finance, it became central to mathematical models of asset prices and option pricing. In probability theory, it opened a new way to study stochastic processes Mikosch [1998], Øksendal [2003], Black and Scholes [1973].

Including Itô is important for the inclusive aim of this chapter. Standard histories of calculus often focus mainly on Europe from the seventeenth to nineteenth centuries. Itô's work reminds us that the history of calculus continued into the twentieth century and became increasingly international. Modern calculus was shaped not only by classical European analysis, but also by new questions, new applications, and mathematicians working across the world.

Modern calculus therefore shows that mathematics is never finished. The same questions that motivated earlier mathematicians are still present, but they now appear in new forms. How does something change? How can we measure total accumulation? How can we predict motion? What happens when change is random, noisy, or too complex for classical methods? From elasticity to optimisation, from numerical simulations to stochastic processes, calculus continues to grow as a global language for understanding the world.

6.10 Conclusion: whose calculus?

Calculus is often introduced through a small number of famous names, especially Newton and Leibniz. Their contributions were extremely important, but the history of calculus is much wider than the story of two individuals. Calculus was not created by one person, in one country, or at one single moment. It grew gradually from many attempts to understand change, motion, area, volume, infinity, and accumulation Boyer [1959], Edwards [1979], Bressoud [2019].

When we look at the longer history, the story becomes richer and more global. Greek mathematicians developed methods of exhaustion to study areas and volumes. Chinese mathematicians used approximation methods to study

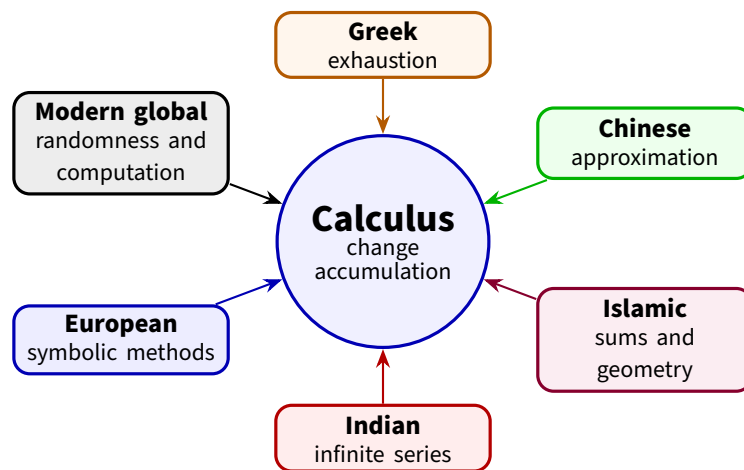


Figure 43: A more inclusive history places calculus inside a wider network of cultures, problems, and mathematical traditions.

curved shapes and numerical problems. Mathematicians in the medieval Islamic world worked on summation, geometry, optics, and algebraic problems that belong to the prehistory of integration and optimisation. Indian mathematicians, including Bhāskara II and the Kerala school, studied small changes, trigonometric approximation, and infinite series. European mathematicians later developed symbolic methods, infinitesimals, fluxions, differentials, and more rigorous foundations. Women mathematicians such as Maria Gaetana Agnesi, Mary Somerville, and Sophie Germain contributed to the teaching, circulation, and application of calculus despite serious barriers to education and recognition. In the twentieth century, mathematicians such as Kiyosi Itô extended calculus into new areas such as stochastic processes and random motion [Katz \[1995\]](#), [Plofker \[2009b\]](#), [Joseph \[2011\]](#), [Robson and Stedall \[2009\]](#), [Stenhouse \[2020\]](#), [O'Connor and Robertson \[a,b\]](#).

This raises an important question: whose calculus do we usually learn about? If the history is told only as a European story beginning with Newton and Leibniz, then many people and traditions disappear from view. We lose sight of the mathematicians who worked in different languages, in different cultures, and with different forms of notation. We also risk forgetting those who were excluded from universities, academies, publications, or professional recognition because of gender, geography, class, race, religion, or language.

A more inclusive history does not remove Newton and Leibniz from the story. Instead, it places them in a wider picture. Their achievement was to organise and systematise powerful methods for studying change and accumulation. However, these methods built on earlier ideas and were later expanded, taught, criticised, refined, and applied by many others. Calculus should therefore be seen as a shared human achievement, shaped by communication, translation, debate, exclusion, rediscovery, and collaboration [Katz \[2009\]](#), [Cooke \[2013a\]](#), [Robson and Stedall \[2009\]](#).

The history of calculus also teaches us something important about mathematics itself. Mathematics is not fixed from the beginning. Ideas develop over time. Notation changes. Definitions become more precise. Methods that once seemed unclear can later become rigorous. Problems from astronomy, physics, engineering, biology, finance, and everyday life can lead to new mathematics. In this way, calculus is not only a collection of formulas, but a living subject that continues to grow [Bressoud \[2019\]](#), [Dunham \[2005\]](#).

The final message of this chapter is that calculus belongs to a long international history of curiosity. It is a history of people trying to understand how things move, how quantities change, how small pieces add up to a whole, and how infinity can be used to solve real problems. By telling this wider story, we can help students see that mathematics does not belong to only one culture or one group of people. Calculus, like mathematics itself, belongs to everyone.

7 The Theory that would not die: Bayesian Probability

The title of this chapter is inspired by Sharon Bertsch McGrayne's *The Theory That Would Not Die* (2011) [McGrayne \[2011\]](#).

What do guiding NASA's **Apollo missions** to the Moon, locating the wreckage of **Air France Flight 447**⁹ in the Atlantic Ocean (2011) [Stone et al. \[2014\]](#), [Cornell University INFO 2040 Blog \[2015\]](#), and improving **Google DeepMind's AlphaGo** before its historic victory over Lee Sedol (March 2016) [Silver et al. \[2016b\]](#) have in common?

The answer lies in **Bayesian statistics**. From recursive state estimation in spacecraft navigation, through Bayesian search theory, to Bayesian optimisation in artificial intelligence, the underlying question for these problems is similar: "*How should uncertain beliefs be revised as new evidence arrives?*".

This chapter follows the developments of the 18th-century theorem and how it evolved into one of the most powerful computational tools in modern applied mathematics.

7.1 Was Bayes trying to calculate the existence of God in the 18th century?

The intellectual setting.

Thomas Bayes (1702 - 1761) lived during the **Enlightenment**, an intellectual movement of the seventeenth and eighteenth centuries in which European philosophers, mathematicians, and scientists increasingly emphasised reason, observation, and empirical evidence as ways of understanding both the natural and social world. Questions that had traditionally belonged largely to philosophy or theology—*How strong is the evidence for a claim? How certain can we be about an event we did not observe directly?*—were increasingly being examined via empirical observations adding to the developments in mathematics and probability [Hacking \[1975\]](#), [Stigler \[1986a\]](#). Probability itself had developed rapidly from seventeenth-century work on games of chance, particularly through Pascal, Fermat, Huygens, and later Jacob Bernoulli and Abraham de Moivre, see chapter 4. By Bayes' lifetime, probability was therefore beginning to offer a mathematical language not only for games of chance, but for reasoning about *uncertain evidence*.

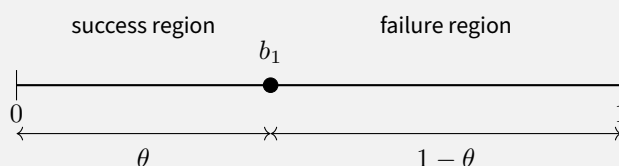
In 1748, the Scottish philosopher **David Hume** (1711 - 1776) wrote in his second book [Hume \[1748\]](#), his ideas about human nature. In particular, the idea that we are more influenced by our feelings than by reason. In *Section X, "Of Miracles"*, he questioned whether reason and testimony could justify belief in miraculous religious claims. Hume wrote:

"The Christian Religion [...] cannot be believed by any reasonable person without [a miracle]. Mere reason is insufficient to convince us of its veracity." [Hume \[1748\]](#)

His arguments raised a question that has an unmistakably probabilistic character: *How should our confidence in a claim change when new evidence is presented?* It was around the same time of Hume's book publication that **Thomas Bayes** investigated what would later be called an **inverse-probability problem**. Bayes did not publish his work during his lifetime. Following his death in 1761, his friend **Richard Price** (1723 – 1791), a British moral philosopher, Non-conformist minister and mathematician, discovered the manuscript, edited it, and forwarded it to the Royal Society, where it appeared in 1763 as *An Essay towards Solving a Problem in the Doctrine of Chances* [Bayes \[1763\]](#). Rather than assuming that the probability of an event was known and predicting what observations should follow, Bayes asked the reverse question: *given repeated observations, what can be inferred about the unknown probability that produced them?*

Bayes' square-table thought experiment with modern notation.

Bayes imagined a "**square table or plane**"^a on which balls fall uniformly at random [Bayes \[1763\]](#). After rescaling the table to have length 1, let $\theta \in (0, 1)$ denote the position of the first ball, measured as the proportion of the table lying to its left. In other words, θ measures only the distance from the left-hand edge of the table to the point where the first ball lands, expressed as a proportion of the total table length. The vertical position of the ball is ignored.



⁹A useful contrast is the search for Malaysia Airlines Flight MH370. Bayesian methods were also used to estimate probable flight trajectories and crash locations from sparse radar and satellite observations [Davey et al. \[2016\]](#). Unlike Air France Flight 447, however, these estimated search regions did not lead to recovery of the aircraft's main wreckage. For Air France 447, investigators faced a comparatively constrained search-location problem, whereas for MH370 they first had to infer an unobserved flight trajectory from sparse and indirect measurements before estimating where the aircraft may have entered the ocean.

Thus, before observing any subsequent throws, we can assign the unknown horizontal position θ of the first ball a **uniform (prior) distribution**,

$$\theta \sim \text{Unif}(0, 1) \equiv \text{Beta}(1, 1).$$

A subsequent ball b_2 is called **success** if it lands to the left of the first ball. Conditional on θ ,

$$X_i \mid \theta \sim \text{Bernoulli}(\theta), \quad P(X_i = 1 \mid \theta) = \theta.$$

where $X_i = 1$ if the i th ball lands to the left of b_1 , and $X_i = 0$ otherwise. After n independent throws, let $S = \sum_{i=1}^n X_i$ denote the total number of **successful throws**. Then the number of balls that land to the left of the initial ball b_1 is modelled by

$$S \mid \theta \sim \text{Binomial}(n, \theta).$$

If we observe exactly $S = s$ successful throws, the probability of this outcome, given θ , is

$$p(s \mid \theta) = \binom{n}{s} \theta^s (1 - \theta)^{n-s}.$$

Note that this conditional density function is viewed as a function of the unknown parameter θ , and it is called the **likelihood**. Since the factor $\binom{n}{s}$ does not depend on θ , we may write

$$p(s \mid \theta) \propto \theta^s (1 - \theta)^{n-s}.$$

Combining the likelihood with the uniform prior using Bayes' rule gives

$$p(\theta \mid s) \propto p(s \mid \theta)p(\theta).$$

Since the uniform prior satisfies $p(\theta) = 1, 0 < \theta < 1$, we obtain

$$p(\theta \mid s) \propto \binom{n}{s} \theta^s (1 - \theta)^{n-s} \times 1 \propto \theta^s (1 - \theta)^{n-s}.$$

Although this has the same mathematical shape as the likelihood, this occurs because the prior is uniform and therefore contributes no preference towards any particular value of θ . The observed data now determine how the posterior probability is distributed across the possible values of θ . After normalisation,

$$\theta \mid S = s \sim \text{Beta}(s + 1, n - s + 1).$$

The **prior** represents uncertainty about the unknown position before observing the additional throws; the **likelihood** records how compatible each possible value of θ is with the observations; and the **posterior** describes the remaining uncertainty about θ after the data have been observed.

^aFor more detailed mathematical and visual explanations of Bayes' square-table experiment, see Begley [Begley \[2020\]](#) and the Cheenta Probability Series [Cheenta Academy](#). Accessible video demonstrations are also available here [bay, Fry and Parker](#).

But was Hume the reason? This remains historically uncertain. One interpretation is that Bayes developed inverse probability partly in response to Hume's arguments concerning miracles: perhaps sufficiently strong accumulated evidence could overcome the initially small probability assigned to an extraordinary claim. **Stephen Stigler** (1941), a professor in Statistics, argues that a connection between Bayes, Price, and contemporary theological debates is plausible [Stigler \[2013\]](#). While, statisticians [Diniz and Bellhouse](#) argue that Bayes' interest was primarily mathematical, arising from problems encountered in de Moivre's *Doctrine of Chances*, rather than from an attempt to refute Hume [Diniz and Bellhouse \[2022\]](#). Crucially, there is **no surviving statement from Bayes confirming that refuting Hume or proving the existence of God was his intention**.

7.2 Laplace: Independently Developing and Extending Inverse Probability

Thomas Bayes' essay of 1763 had solved a particular inverse problem: after observing a sequence of successes and failures, how could one infer the unknown probability that generated them? Through his square-table construction, Bayes effectively treated the unknown probability as uniformly distributed over $(0, 1)$ and then updated this distribution using the observed data.

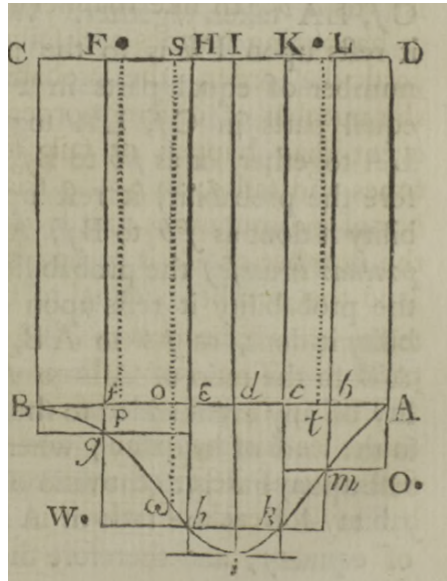


Figure 44: A modern illustration of Bayes' square-table thought experiment. Bayes described a "square table or plane", rather than specifically a billiard or pool table, see [Bayes' original 1763 essay](#).

Pierre-Simon Laplace (1749–1827) independently developed the idea in a more general form. In his 1774 memoir, *Mémoire sur la probabilité des causes par les événements* ("Memoir on the Probability of Causes from Events"), Laplace distinguished between predicting an uncertain event when its cause (a different event) is known and the reverse problem in which the event has been observed but its cause remains unknown [Laplace \[1774\]](#), [Stigler \[1986b\]](#). As **Sharon Bertsch McGrayne** summarises it in *The Theory That Would Not Die* (2011),

"the probability of a cause (given an event) is proportional to the probability of the event (given its cause)."
[McGrayne \[2011\]](#)

Laplace did not state the general principle using modern conditional-probability notation. Instead Laplace's statement appears in Article II of the 1774 memoir, shown in Figure 45. He explains that if an observed event can have arisen from several possible causes, then the probabilities assigned to those causes after observing the event should be proportional to the probabilities with which each cause would have produced that event [Laplace \[1774\]](#). Under the assumption that the possible causes are initially equally probable, the principle can be written in modern notation as

$$P(C_i | E) \propto P(E | C_i).$$

PRINCIPE. — Si un événement peut être produit par un nombre n de causes différentes, les probabilités de l'existence de ces causes prises de l'événement sont entre elles comme les probabilités de l'événement prises de ces causes, et la probabilité de l'existence de chacune d'elles est égale à la probabilité de l'événement prise de cette cause, divisée par la somme de toutes les probabilités de l'événement prises de chacune de ces causes.

Figure 45: Laplace's statement of his inverse-probability principle in Article II of *Mémoire sur la probabilité des causes par les événements* (1774); see the [original facsimile](#).

Laplace's urn problem: from past observations to future probability, see also [Laplace \[1774\]](#) for more details.

Laplace illustrated his principle of inverse probability using an urn containing an infinite number of white and black tickets in an **unknown proportion**.

Suppose that p white tickets and q black tickets have already been observed, and let x denote the **unknown proportion of white tickets in the urn**.

If the **true** proportion were x , then in modern notation

$$P(\text{white ticket} | x) = x, \quad P(\text{black ticket} | x) = 1 - x.$$

Therefore, after observing p white tickets and q black tickets, Laplace associated each possible value of x with the

quantity

$$x^p(1-x)^q.$$

Thus, $x^p(1-x)^q$ measures how compatible a particular value of x is with the observations already made.

He then asked a slightly different question from Bayes: *given these past observations, what is the probability that the next ticket drawn will be white?*

For any fixed value of x , that probability is simply also just x , as the ratio shouldn't change. However, since x is unknown, **Laplace considered all possible values of x between 0 and 1**, weighting each according to $x^p(1-x)^q$. He denoted the resulting "entire probability" by E :

$$\begin{aligned} P(x \mid p \text{ white}, q \text{ black}) &= \frac{P(p \text{ white}, q \text{ black} \mid x) P(x)}{P(p \text{ white}, q \text{ black})} = \frac{x^p(1-x)^q \times 1}{\int_0^1 x^p(1-x)^q dx}, \\ E = P(\text{white next} \mid p \text{ white}, q \text{ black}) &= \int_0^1 P(\text{white next} \mid x) P(x \mid p \text{ white}, q \text{ black}) dx, \\ &= \frac{\int_0^1 x^{p+1}(1-x)^q dx}{\int_0^1 x^p(1-x)^q dx}, \end{aligned} \quad (4)$$

because we assume a uniform prior density on the continuous *proportion variable* x given by $P(x) = 1$ when $0 \leq x \leq 1$. The denominator ensures that the weights assigned to all possible values of x are properly normalised, while the additional factor x in the numerator accounts for the probability of drawing a white ticket for each possible value of x .

Laplace evaluates the integral over a few pages, and gives the answer

$$E = \frac{p+1}{p+q+2}$$

so E is the probability, inferred from the $p+q$ observations already made, that the *next* ticket drawn from the urn will be white **Laplace [1774]**.

McGrayne illustrates the difference in scope between Bayes and Laplace with the analogy

"If Bayes and Price asked whether today's puddles could tell us whether it had rained yesterday and whether it might rain tomorrow, Laplace asked how much rain might fall and then repeatedly revised that estimate as additional information became available." **McGrayne [2011]**

The important extension was the possibility of **repeated updating: as new observations arrived, the probability assigned to an unknown quantity could be revised again**.

Laplace's work on probability soon developed alongside his work in astronomy and celestial mechanics. In Chapter 2, "The Man Who Did Everything," McGrayne describes how eighteenth-century astronomers possessed observations extending over centuries, although the measurements differed greatly in precision and sometimes appeared inconsistent with Newtonian celestial mechanics.¹⁰ One particularly troubling example concerned the apparent long-term motions of Jupiter and Saturn. As McGrayne writes,

"...eventually, they thought, Jupiter would smash into the sun, and Saturn would spin off into space." **McGrayne [2011]**

In his 1787 work **Laplace [1787, 1788]** on celestial mechanics, Laplace used gravitational theory and perturbation methods to show that, despite their apparent changes in motion, Jupiter would not 'smash into the Sun' and Saturn would not 'spin off into space'. Some of Laplace's related work on probability and large-sample approximations is discussed in Chapter 4.

¹⁰For examples of astronomical observations available before Laplace, see Flamsteed's *Historia Coelestis Britannica* **Flamsteed [1725]**. For the older European tradition of planetary tables, see Chabás and Goldstein **Chabás and Goldstein [2012]**.

7.3 Inverse probability in the nineteenth century

The nineteenth century was a period of rapid expansion in both probability theory and the use of mathematics to analyse empirical observations. As discussed further in Chapter 4, Laplace developed **generating functions** and **normal approximations**; Siméon-Denis Poisson (1781 - 1840) extended the **theory of repeated trials** and introduced the expression **law of large numbers**; Irénée-Jules Bienaymé (1796 - 1878) contributed to **probability inequalities** and **error theory**; and Pafnuty Chebyshev (1821 - 1894) later established increasingly general results concerning **averages, inequalities and convergence** Poisson [1837] Bienaymé [1853] Chebyshev [1867]

At the same time, probability and statistical methods were being applied to an expanding range of empirical data. It was beginning to form what we would now recognise as **statistical methodology**. Astronomers and geodesists continued to estimate unknown quantities from noisy measurements using methods associated with Laplace (1749 - 1827), Legendre (1752 - 1833) and Gauss (1777 - 1855), while Adolphe Quetelet (1796 - 1874) transferred ideas about averages, variation and the normal law from astronomy to large collections of demographic and social observations Stigler [1986a], Gauss [1809], Quetelet [1835]. McGrayne also writes

" **James Clerk Maxwell** (1831 - 1879), learned about probability from Adolphe Quetelet (1796 - 1874), not from Laplace, and adopted frequency-based methods for statistical mechanics and the kinetic theory of gases.", McGrayne [2011]

In modern language, the nineteenth century was beginning to face an early version of the **"big data" problem**: scientists and governments were collecting increasingly large amounts of *astronomical, demographic and social observations*, creating a need for mathematical methods that could identify patterns, quantify uncertainty and estimate unknown quantities from data. During the time inverse probability remained in use, particularly in applied problems of estimation, but it was no longer the only framework through which mathematicians and scientists were learning from uncertain observations.

The theory's biggest difficulty was concerned with the probabilities assigned to unknown causes or parameters. Bayes' construction and many of Laplace's calculations began by treating possible values as **initially equally probable**. This assumption continued discussions in the 20th century, under the name of *principle of indifference*¹¹. For a symmetric problem such as throwing an idealised fair die, assigning equal probabilities may appear natural. For an unknown planetary mass parameter, population parameter, or other continuous quantity, however, there is no obvious reason why a lack of prior knowledge should imply that we assign equal initial probabilities to the model parameters. George Boole (1815 - 1864), John Venn (1834 - 1923) and Joseph Bertrand (1822 - 1900) criticised different aspects of this reasoning Boole [1854], Venn [1866], Bertrand [1889]. In the chapter **"Many doubts, few defenders"** McGrayne [2011] McGrayne discusses many arguments noted during the century about Laplace's reasoning.

While John Maynard Keynes, see *A Treatise on Probability* (1921), Chapter XXX, "Statistical Inference", around pp. 382 states

"...it has been very widely accepted, — by Augustus De Morgan (1806 - 1871), by William Stanley Jevons (1835 - 1882), by Rudolf Hermann Lotze (1817-1881), by Emanuel Czuber (1851 - 1925), and by Professor Pearson, — to name some representative writers of successive schools and periods." Keynes [1921], McGrayne [2011] McGrayne [2011]

The controversy was therefore increasingly about the *interpretation* of inverse probability rather than the validity of conditional probability itself: **could an unknown but fixed quantity legitimately be assigned a probability distribution, and how should that distribution be chosen?** These questions remained unresolved at the end of the nineteenth century. In the early twentieth century they became much sharper as **alternative theories of statistical inference** were developed. R. A. Fisher (1890 - 1926) would reject inverse probability as a general foundation for parameter inference and develop **likelihood-based methods** that did not require prior probabilities Fisher [1922a], while other mathematicians—including Harold Jeffreys (1891 - 1989), Dorothy Wrinch (1894 - 1976)¹² and Bruno de Finetti (1906 - 1985)—would instead attempt to place probabilistic inference on stronger foundations. The resulting disagreement marks the beginning of the much more explicit twentieth-century debate over how statistical inference should be performed.

¹¹The expression "principle of indifference" was used prominently by Keynes in his *Treatise on Probability* (1921); related arguments had long been discussed under the older expression "principle of insufficient reason".

¹²**Dorothy Wrinch** (1894-1976) was a mathematician and philosopher of science, and **the first woman to receive a Doctor of Science (DSc) degree from Oxford University (awarded in 1929)**. She collaborated with Harold Jeffreys on probability and scientific inference between 1919 and 1923. Their work has been identified as an important conceptual precursor to Bayesian hypothesis testing and the later Bayes factor; Jeffreys subsequently acknowledged Wrinch's substantial contribution to the foundations of his work on scientific inference. *An extract from Jeffreys' book on Probability Wrinch and Jeffreys [1919, 1921], Jeffreys [1973]*

7.4 So what is Bayes' theorem today?

An excellent introductory Probability book is written by Tijms' *Understanding Probability* Tijms [2012]. **Chapter 8** develops conditional probability, the law of total probability and Bayes' rule before introducing Bayesian statistics for the discrete case. **Chapter 13** later extends these ideas to continuous random variables and continuous Bayesian models. For readers already comfortable with these probability and calculus concepts, Gelman et al.'s *Bayesian Data Analysis* provides a natural next step Gelman et al. [2013]. A visual and intuitive explanation of Bayes' theorem, is provided in several **YouTube video introductions**, including 3Blue1Brown's [geometric explanation of Bayes' theorem](#), as well as further examples by [RiskByNumbers](#) and [Brandon Rohrer](#). 3Blue1Brown also explores Bayes' rule through the [medical test paradox](#), illustrating visually why base rates and prior probabilities matter when interpreting evidence.

Modern introductions to Bayes' theorem usually begin with the **discrete case**¹³. Suppose that the events $B_1, \dots, B_k$ represent all possible hypotheses, explanations or parameter values under consideration. For example, in Bayes' experiment, the unknown quantity was **the position θ of the first ball** on the table, while in Laplace's urn problem it was the actual **unknown proportion x of white tickets**. These are assumed to be **mutually exclusive**, so that no two can occur simultaneously, and **exhaustive**, so that one of them must occur.

Let A denote some observed evidence, or collected data. In Bayes' experiment, this could be the observation that, among n subsequent balls, s fell to the left of the first ball and $n - s$ to its right. In Laplace's urn problem, the corresponding evidence was the observation of p white tickets and q black tickets.

Two results from probability theory are needed. First, **conditional probability** is defined by

$$P(B_i | A) = \frac{P(B_i \cap A)}{P(A)}.$$

Second, the **law of total probability** says that the probability of observing A can be found by considering all the different hypotheses under which A could have occurred:

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

This can be visualised geometrically by dividing the event A into the mutually exclusive pieces contributed by each possible hypothesis, as shown in Figure 46.

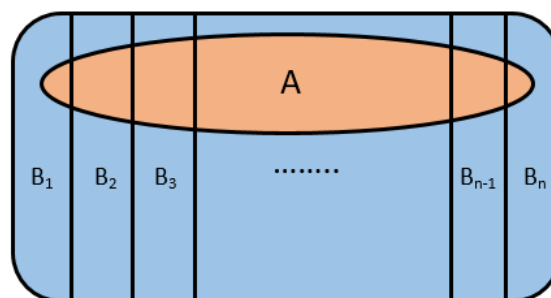


Figure 46: Geometric illustration of the law of total probability. The sample space is partitioned into mutually exclusive events $B_1, \dots, B_n$, and the probability of A is obtained by adding its intersections with each part of the partition. Illustration source: [Blog by Authors: Armin Nikzad and Fatima Mosallaei](#).

Since the intersection can also be obtained by $P(B_i | A)P(A) = P(B_i \cap A) = P(A \cap B_i) = P(A | B_i)P(B_i)$, substituting the law of total probability into conditional probability gives the modern discrete form of **Bayes' theorem**:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{j=1}^k P(A | B_j)P(B_j)}. \quad (5)$$

Further details on conditional probability and the law of total probability can be found in Tijms [2012].

¹³For an intuitive geometric interpretation of the simplest two-event form of Bayes' theorem, see [Wikipedia](#) and [Geometric Bayes](#) by Janosh Riebesell. Both represent the sample space geometrically, with B as a hypothesis and A as the observed evidence, so that the relevant conditional probabilities can be interpreted as areas within a square or rectangle.

In words, Bayes' theorem begins with probabilities assigned to the different possible hypotheses B_i , asks how compatible the observed evidence A is with each of them, and then uses this information to revise their probabilities after the evidence has been observed. In modern Bayesian terminology this is often summarised as

$$\boxed{\text{posterior} \propto \text{likelihood} \times \text{prior}}. \quad (6)$$

The same idea extends naturally to **continuous** problems. In the discrete case there is a finite or countable list of possibilities, $B_1, B_2, \dots, B_k$. In a continuous problem, however, **an unknown quantity may take any value within an interval**. For example, an unknown probability might take any value between 0 and 1, **rather than being restricted to a list of possible values**. Such an unknown quantity is commonly represented by a parameter such as θ .

To connect the discrete and continuous forms, let B_i represent the event that an unknown parameter Θ takes the value θ_i , suppose that the hypotheses B_i correspond to possible values of an unknown parameter Θ , so that $B_i = \{\Theta = \theta_i\}$, and let the observed evidence A correspond to observing the data x . Then the discrete form of Bayes' theorem, may equivalently be written in parameter notation as

$$P(\Theta = \theta_i | x) = \frac{P(x | \theta_i) P(\Theta = \theta_i)}{\sum_{j=1}^k P(x | \theta_j) P(\Theta = \theta_j)}.$$

Here the possible parameter values $\theta_1, \dots, \theta_k$ still form a finite or countable collection. If Θ is instead continuous, these individual probability masses are replaced by probability densities, and the sum over the possible values θ_j becomes an integral over the continuum of possible values of θ . This is the same mathematical transition from summation to integration discussed in the Calculus chapter (see Section 6).

The continuous form of Bayes' theorem is therefore

$$\boxed{\begin{array}{c} \text{posterior density} \\ \underbrace{p(\theta | x)} = \frac{\overbrace{p(x | \theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior density}}}{\underbrace{\int p(x | u) p(u) du}_{\text{marginal density } p(x) \text{ (evidence / normalising constant)}}} \end{array}}. \quad (7)$$

Here u is a dummy variable of integration. The denominator sums, or in the continuous case integrates, over all possible values of the unknown parameter.

7.5 From inverse probability to competing theories of statistical inference (c. 1900–1954)

A new foundation for probability (1900–1933). At the beginning of the twentieth century, probability itself was being reconstructed on firmer mathematical foundations. David Hilbert (1862–1943) called for its axiomatization in his sixth problem of 1900; Émile Borel brought measure-theoretic ideas into probability, drawing on the emerging theory of measure and integration developed by Borel and Henri Lebesgue, while Richard von Mises (1883–1953) developed an explicitly frequency-based interpretation. The mathematical development by Andrey Kolmogorov's (1903–1987) *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933), which defined probability axiomatically as a measure on a space of events Hilbert [1900], Émile Borel [1909], von Mises [1919], Kolmogorov [1933b].

But the question of inverse probability persisted:

What exactly may be assigned a probability?

If θ is a fixed quantity whose true value we do not know, can we assign probabilities to the different values that θ might take?

During the following decades of the twentieth century, several different answers were developed.

Fisher: likelihood instead of inverse probability (1922–1935). Ronald Aylmer Fisher (1890–1962) developed a non-Bayesian framework of statistical inference in which **likelihood** provided a way of comparing possible values of an unknown parameter without assigning probabilities to those values. He then developed **maximum likelihood estimation** and introduced concepts for assessing how much information about a parameter could be extracted from data, including **sufficiency**, **efficiency** and **statistical information**. In parallel, he developed **significance testing** and **fiducial inference** as an alternative way of making inferential statements about unknown parameters without introducing Bayesian prior probabilities Fisher [1922b, 1925, 1930, 1935], Aldrich [1997]. While his rejection of inverse probability was explicit. In 1930 he described it as a doctrine that appeared to

“a succession of sound writers to be *fundamentally false and devoid of foundation*” [Fisher, 1930, p. 528].

At Rothamsted, Fisher worked closely with his first statistical assistant, **Winifred Alice Mackenzie (1896–1954)**, who co-authored studies on rainfall, bacterial counts and crop variation, including an important 1923 paper connected with the early development of analysis of variance Fisher and Mackenzie [1923], Taylor and Aldrich [2022]. Historians have additionally documented the **substantial workforce of women employed as computing assistants at Rothamsted**, whose computational labour was rarely acknowledged in conventional accounts of the development of statistics Parolini [2015].

Neyman and Pearson: hypothesis testing and error control (1928–1937). Jerzy Neyman (1894–1981) and Egon Pearson¹⁴ (1895–1980) developed another non-Bayesian approach to statistical inference. Their work between 1928 and 1933 treated hypothesis testing as a decision between a **null hypothesis** H_0 and an **alternative hypothesis** H_1 Neyman and Pearson [1928, 1933]. Instead of assigning initial probabilities to the hypotheses themselves, they asked how often a testing rule would make the wrong decision if the same experiment were performed many times under the same conditions:

$$\underbrace{P(\text{reject } H_0 \mid H_0 \text{ is true})}_{\text{Type I error } (\alpha)} \quad \text{and} \quad \underbrace{P(\text{do not reject } H_0 \mid H_1 \text{ is true})}_{\text{Type II error } (\beta)}. \quad (8)$$

The **power**¹⁵ of the test is the probability of correctly rejecting H_0 when H_1 is true:

$$P(\text{reject } H_0 \mid H_1 \text{ is true}) = 1 - \beta. \quad (9)$$

Neyman applied the same idea to his 1937 theory of **confidence intervals** Neyman [1937]. Fisher and Neyman–Pearson are now often grouped together as “frequentist”, although their interpretations of statistical testing differed substantially. Later textbooks on Statistics often combined Fisherian p-values/significance testing with Neyman–Pearson terminology and error-control machinery, despite disagreements between their founders Lehmann [2011], Perezgonzalez [2015].

Wrinch, Haldane and Jeffreys: probabilities of hypotheses (1919–1946). Meanwhile, Dorothy Wrinch (1894–1976) and Harold Jeffreys (1891–1989) were developing almost the opposite framework. Between 1919 and 1923 they treated probability as a logic for reasoning about scientific propositions Wrinch and Jeffreys [1919, 1921, 1923]. In 1921 they wrote that

“an inference drawn from a simple scientific law may have a very high probability, not far from unity” [Wrinch and Jeffreys, 1921, p. 380].

J. B. S. Haldane (1892–1964) developed an early Bayesian approach to hypothesis testing. He introduced a prior in which an exact hypothesis, such as “there is no effect” ($\theta = 0$), could itself be assigned some probability, while the remaining probability was spread continuously over possible alternatives Haldane [1932], Etz and Wagenmakers [2017]. Jeffreys then developed Bayesian significance tests in 1935 and systematically in *Theory of Probability* (1939) Jeffreys [1935, 1939].

Jeffreys also returned to the old difficulty of choosing a prior. His 1939 book developed invariant approaches to prior choice, and his 1946 paper gave the general Fisher-information rule $\pi_J(\theta) \propto \sqrt{\det I(\theta)}$, which is invariant under smooth one-to-one reparameterisations Jeffreys [1939, 1946].

¹⁴Karl Pearson founded the world’s first university statistics department was founded in 1911 at University College London (UCL). His student **Florence Nightingale David**, who would publish her work under the nickname F.N. David, later became the first woman professor in the department. Her research focused heavily on classic frequentist tools like goodness-of-fit tests, the behaviour of correlation coefficients, and tests for randomness.

¹⁵Dr Nic has an excellent YouTube video explaining [Type 1 and Type 2 errors](#) in more detail.

Wald and Savage: inference becomes decision-making (1939–1954). Abraham Wald¹⁶ (1902–1950) reframed the problem at hand. From 1939 onward he treated estimation and hypothesis testing as instances of a more general **decision problem**: after observing data, a statistician must choose some action a . This action is evaluated through a **loss function** $L(a, \theta)$, which measures the cost of taking action a when the true but unknown, model parameter is θ Wald [1939, 1950].

Using a Bayesian approach, first we update the probability distribution of the unknown parameter θ after observing the data, and then chooses the action with the smallest **expected loss under that posterior distribution** Wald [1950], Savage [1954]. In other words, the preferred action is the one expected to perform the best given what is now known about θ . Leonard J. Savage (1917–1971) developed this connection further in *The Foundations of Statistics* (1954), giving subjective Bayesian probability a decision-theoretic foundation based on personal probabilities and expected utility Savage [1954].

Wald's work on sequential analysis was an important precursor to Richard Bellman's (1920–1984) development of **dynamic programming**. Bellman described his new theory as "*intimately related*" to Wald's sequential analysis Bellman [1952], while his broader development of dynamic programming also drew on work on sequential decision problems and multi-stage optimisation Bellman [1952, 1957]. Dynamic programming later became foundational to optimal control, Markov decision processes and reinforcement learning¹⁷.

7.6 Bayesian inference at Bletchley Park: Turing, Good and Naval Enigma (c. 1940–1943)

Although inverse probability had already found applications in astronomy, geodesy, population and mortality statistics, insurance and actuarial science during the nineteenth and early twentieth centuries McGrayne [2011], Hacking [1975], Franklin [2001], the Second World War produced one of the most remarkable practical applications of probabilistic and Bayesian reasoning.

In *Part II: Second World War Era of The Theory That Would Not Die*, McGrayne devotes considerable attention to the problems faced during the war McGrayne [2011]. Importantly, her account emphasises that British cryptanalytic work at Bletchley Park did not begin from nothing. The Polish Cipher Bureau had already made substantial progress against German Enigma during the 1930s. She writes,

"By early 1938 the Poles were reading 75% of Germany's army and air force messages." McGrayne [2011].

The mathematicians Marian Rejewski, Jerzy Różycki and Henryk Zygalski had developed mathematical, mechanical and paper-based methods for attacking the German Enigma system. Shortly before the outbreak of war in 1939, the Polish cryptanalysts shared their discoveries with British and French intelligence, providing an important foundation for the subsequent work at Bletchley Park McGrayne [2011], Hodges [2014], Hinsley and Stripp [1993].

The increasingly mechanical character of cryptography also changed the type of expertise required by the British Government. Alastair Denniston, head of the Government Code and Cypher School, deliberately began recruiting mathematicians from British universities, including Alan Turing (1912–1954), Gordon Welchman (1906–1985) and Max Newman (1897–1984). Turing subsequently took on the particularly difficult problem of German Naval Enigma and became a central figure in Hut 8. Irving John Jack" Good (1916–2009), originally trained as a pure mathematician, joined Bletchley Park in 1941 and later described himself as Turing's main statistical assistant" during the war Good [1979].

This was nevertheless **a much broader collaborative cryptanalytic effort rather than the work of Turing alone**. Different German communications systems were being attacked by different groups, and some were already yielding results during the early months of the war. McGrayne notes that

"By January [1940] the English were reading German air force messages." McGrayne [2011].

Progress against Naval Enigma, however, presented a substantially more difficult problem. The German Navy employed Enigma with more sophisticated operating procedures than the other German services; GCHQ's historical account consequently describes it as *the hardest cryptanalytic problem facing Bletchley Park*. It was this problem that Turing chose to attack and within which probabilistic reasoning, Banburismus and the weighing of evidence would become particularly important.

¹⁶ Jordan Ellenberg, a professor of mathematics, famously opens with Abraham Wald and the missing bullet hole in his 2014 book *How Not to Be Wrong: The Power of Mathematical Thinking*. You can also find his talk at *The Royal Institution* on YouTube.

¹⁷ Reinforcement learning later became a central component of several major artificial-intelligence systems developed by DeepMind (now Google DeepMind). AlphaGo combined deep neural networks, tree search, supervised learning from expert human games and reinforcement learning through self-play to achieve professional-level and subsequently superhuman performance in Go Silver et al. [2016a]

Women at Bletchley Park. Women played a significant role both in the mathematical work at Bletchley Park and in putting its cryptanalytic methods into practice. The mathematician [Joan Clarke](#) (1917–1996), who worked alongside Turing in Hut 8, became particularly skilled at the process *Banburismus*¹⁸.

The contribution of women extended far beyond Hut 8. Around **75% of the Bletchley Park workforce were women**¹⁹, in roles including cryptanalysis, linguistics, traffic analysis, interception, clerical work and the operation of cryptanalytic machinery [GCHQ \[2017\]](#). Members of the Women's Royal Naval Service (*Wrens*), for example, operated the Bombe machines used to test possible Enigma settings. Among the notable women were [Mavis Batey](#), whose work against Enigma later proved important to Allied preparations for D-Day, and [Charlotte “Betty” Webb](#), who worked with German intercepts and later with Japanese communications.

For further historical accounts, see Philippa Lacey's interview with Sir Dermot Turing, [The Hidden Women Code-breakers of World War II](#). For the wider international story of women in wartime cryptanalysis, see Liza Mundy's book *Code Girls: The Untold Story of the American Women Code Breakers of World War II* (2017) and her [Talks at Google presentation](#).



Figure 47: Women in Bletchley Park, (date of image: unknown); Source: [Wikipedia](#).

The problem: Naval Enigma and Bayesian hypothesis testing

The Enigma machine encrypted each letter through a plugboard and a sequence of rotating wired rotors. For more information, Alexander's post-war history and Mahon's *History of Hut Eight* are excellent sources for Naval Enigma [Alexander \[1945\]](#), [Mahon \[1945\]](#). Because the rotors moved as the machine was used, the substitution continually changed. Knowing how an Enigma machine worked was therefore not enough: the cryptanalysts had to infer its **unknown configuration** from intercepted messages and other evidence.

For Naval Enigma, the difficulty could be formulated partly as a problem of **hypothesis testing**, more specifically [Bayesian Hypothesis Testing](#). Cryptanalysts compared pairs of intercepted messages by shifting one ciphertext relative to another. A proposed alignment represented a hypothesis that portions of the two messages had been enciphered at corresponding Enigma positions. If an alignment produced an unusually large number or pattern of letter coincidences, this provided evidence that the two portions were in *depth* — that is, that corresponding letters had passed through the same sequence of Enigma substitutions. Probable depths could then be combined to infer relationships between message settings and to eliminate unlikely rotor orders, greatly reducing the subsequent search [Turing \[1941\]](#), [Alexander \[1945\]](#), [Zabell \[2012\]](#).

Let H_1 and H_0 denote two competing hypotheses about the cipher configuration, and let E denote the observed evidence—for example, the pattern of coincidences produced by a particular message alignment. In odds form, Bayes' theorem gives

$$\underbrace{\frac{P(H_1 | E)}{P(H_0 | E)}}_{\text{posterior odds}} = \underbrace{\frac{P(H_1)}{P(H_0)}}_{\text{prior odds}} \underbrace{\frac{P(E | H_1)}{P(E | H_0)}}_{\text{Bayes factor (likelihood ratio)}} \quad (10)$$

¹⁸Banburismus was a statistical cryptanalytic procedure developed by Turing for Naval Enigma. Evidence from intercepted messages was accumulated and expressed through logarithmic weights of evidence, or decibans, in order to rank possible rotor orders and reduce the amount of Bombe processing required [Good \[1979\]](#), [Zabell \[2012\]](#).

¹⁹Source: [Wikipedia](#)



Figure 48: A German Enigma machine. [Source: Wikipedia.](#)

2.1. Vigenère

The factor principle can be applied to the solutions of a Vigenère problem with great effect. I will assume here that the period of the cipher has already been determined. Probability theory may be applied to this part of the problem also, but that is not so elementary. Suppose our cipher, written out in its correct period is¹

D K Q H S H Z N M P
R C V X U H T E A Q
X H P U E P P S B K
T W U J A G D Y O J
T H W C Y D Z H G A
P Z K O X O E Y A E
B O K B U B P I K R
W A C E J P H L P
T U Z Y F H L R Y C

FIGURE 1. Vigenère problem.
(It is only by chance that it makes a rectangular array.)

Let us try to find the key for the first column, and for the moment let us only take into account the evidence afforded by the first letter D. Let us first consider the key B. The factor principle tells us

$$\text{Odds in favour of key B} = \frac{\text{A priori odds in favour of key B}}{\text{Probability of getting D in cipher if key is B}} \times \frac{\text{Probability of getting key D in cipher if key is not B}}{\text{Probability of getting key D in cipher if key is B}}$$

Now the *a priori* odds in favour of key B may be taken as 1/25. The probability of getting D in the cipher with the key B is just the probability of getting C in the clear which (using the count on 1000 letters in Fig. 2) is 0.021. If however the key is not B we can have any letter other than C in the clear, and the probability is $(1 - 0.021)/25$. Using the evidence of the D then the odds in favour of the key B are

$$\frac{1}{25} \times \left(\frac{25 \times 0.021}{1 - 0.021} \right).$$

Figure 49: Turing's Vigenère cipher example in *The Applications of Probability to Cryptography*. [Read the original text.](#)

Thus the **prior odds** are multiplied by a **likelihood ratio** measuring how strongly the observed evidence favours one hypothesis over the other. Turing described this idea in his wartime manuscript *The Applications of Probability to Cryptography*, calling it the *factor principle* [Turing \[1941\]](#), [Zabell \[2012\]](#).

Because the calculations required the multiplication of many such factors, Turing expressed the resulting weight of evidence on a logarithmic scale using the **deciban** [Turing \[1941\]](#), [Good \[1979\]](#),

$$D = 10 \log_{10} \left(\frac{P(E | H_1)}{P(E | H_0)} \right). \quad (11)$$

Ten decibans multiplied the odds by 10, while one deciban changed them by a factor of $10^{0.1} \approx 1.26$. These ideas formed part of Turing's statistical procedure known as *Banburismus*, which allowed the accumulation of evidence *for or against candidate rotor relationships* and thereby eliminate unlikely rotor orders before committing scarce Bombe time [Good \[1979\]](#), [Alexander \[1945\]](#), [Zabell \[2012\]](#).

The importance of **Turing's contribution** was therefore not the invention of Bayesian hypothesis testing itself, which had earlier roots in the work of [Wrinch](#), [Haldane](#) and [Jeffreys](#) [Wrinch and Jeffreys \[1919, 1921, 1923\]](#), [Haldane \[1932\]](#), [Jeffreys \[1935\]](#), [Etz and Wagenmakers \[2017\]](#). His important contribution was to **make these ideas practical**. As new clues were obtained, the cryptanalysts could increase or decrease their confidence in different possible Enigma settings, gradually narrowing down the possibilities. Good later documented and further developed the wartime statistical ideas into the public literature here [Good \[1950, 1979\]](#).

7.7 The computational revolution in Bayesian inference: from Bayes' theorem to algorithms (c. 1950–present)

Post-1950 history of Bayesian statistics can be viewed as three connected developments:

Computation	: how can difficult posterior distributions be calculated?
Modelling	: what kinds of unknown objects can carry probability distributions?
Learning and decisions	: how can uncertainty be used for prediction and action?

7.7.1 The computational revolution

The history of Bayesian computation also includes several women whose contributions are easily hidden by the names later given to the algorithms. The original 1953 Metropolis paper was co-authored by the physicists [Arianna Rosenbluth](#) and [Augusta Teller](#) [Metropolis et al. \[1953\]](#). Rosenbluth played a particularly important role in turning the proposed

method into a working program on the early MANIAC computer at Los Alamos; later historical accounts have emphasised that her computational contribution was for many years under-recognised [Gubernatis \[2005\]](#).

Several decades later, the statistician [Nicky Best](#) became an important figure in making Bayesian computation accessible to applied researchers. From 1996 she led work at Imperial College expanding [WinBUGS](#), which allowed complicated Bayesian models to be fitted using MCMC without users having to construct every sampling algorithm themselves [Lunn et al. \[2000, 2009\]](#).

For readers who want to study these methods in detail today, Robert and Casella's *Monte Carlo Statistical Methods* gives a standard route into MCMC [Robert and Casella \[2004\]](#), Chopin and Papaspiliopoulos provide an introduction to particle filters and particle MCMC [Chopin and Papaspiliopoulos \[2020\]](#), and Murphy's *Probabilistic Machine Learning: An Introduction* gives an introduction to approximate inference, including variational ideas, in the wider machine-learning setting [Murphy \[2022\]](#).

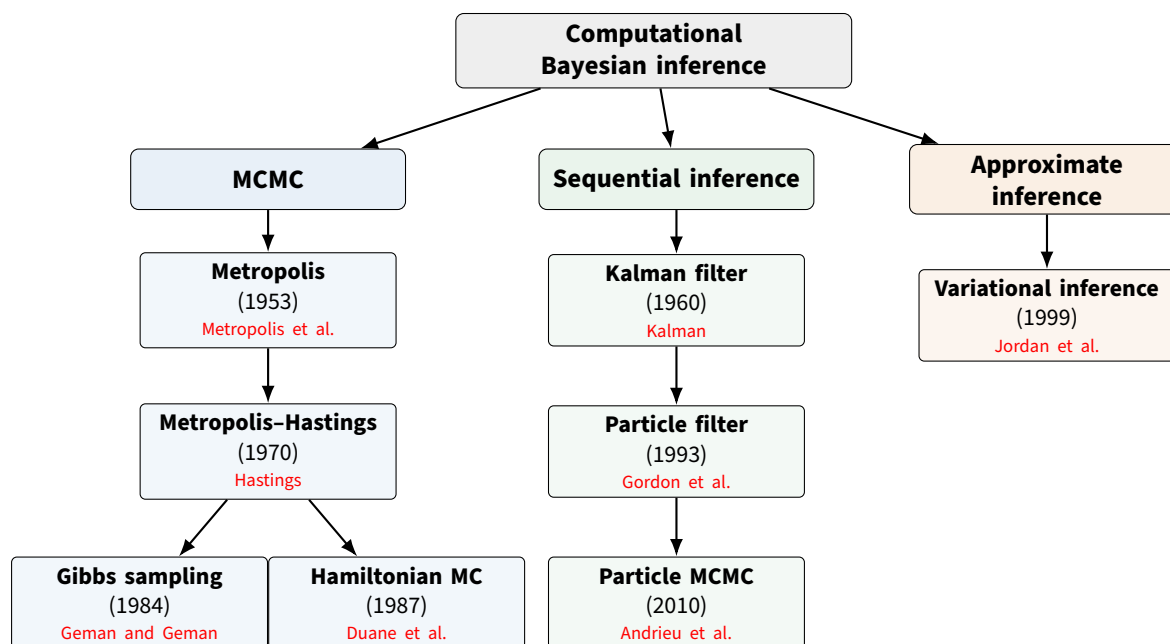


Figure 50: Selected developments in Bayesian computation after 1950.

7.7.2 The modelling revolution

Women also made important contributions to the increasingly flexible models shown in Figure 51. In 1970 the statistician [Grace Wahba](#), together with George Kimeldorf, established an important correspondence between Bayesian estimation of random functions and smoothing splines [Kimeldorf and Wahba \[1970\]](#). Wahba's later work on smoothing, regularisation and reproducing-kernel methods became an important part of the mathematical background linking Bayesian function estimation with modern kernel methods.

The biostatistician [Sylvia Richardson](#) likewise contributed to the development of flexible Bayesian mixture models. Her 1997 paper with Peter Green developed a fully Bayesian analysis in which even the number of mixture components was unknown, using variable-dimension MCMC [Richardson and Green \[1997\]](#).

To study this area in detail today, Gelman et al.'s *Bayesian Data Analysis* and Hoff's *A First Course in Bayesian Statistical Methods* are strong starting points for hierarchical and multilevel modelling [Gelman et al. \[2013\]](#), [Hoff \[2009\]](#); Hjort, Holmes, Müller and Walker's *Bayesian Nonparametrics* gives a focused route into Dirichlet-process and related models [Hjort et al. \[2010\]](#); and Rasmussen and Williams' *Gaussian Processes for Machine Learning* is the standard reference for the function-space side of the subject [Rasmussen and Williams \[2006\]](#).

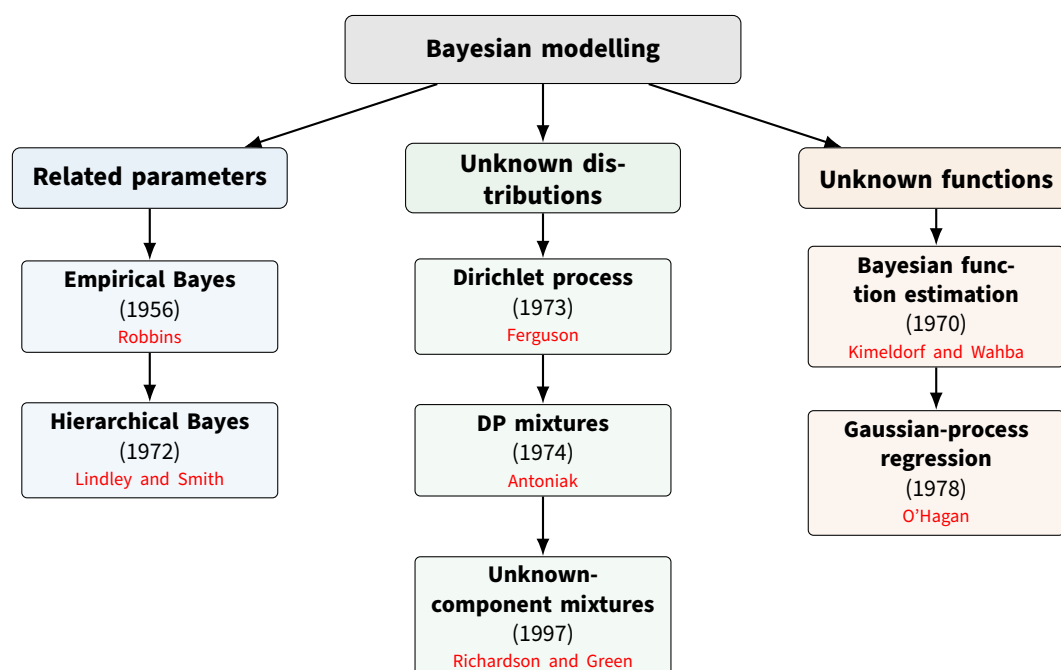


Figure 51: Selected developments in Bayesian modelling after 1950. Bayesian uncertainty expanded from parameters to distributions and functions.

7.7.3 Bayesian ideas in machine learning and decision-making

An important figure in the history of Bayesian decision theory was the African-American mathematician [David Blackwell \(1919–2010\)](#). Blackwell made fundamental contributions to probability, game theory, information theory and Bayesian inference, and his 1954 book with Meyer Girshick, *Theory of Games and Statistical Decisions*, helped develop Bayesian approaches to sequential decision problems [Blackwell and Girshick \[1954\]](#). His career is remarkable as he pursued this work while facing racial discrimination in American academia.

The later development of Bayesian machine learning was similarly international. The Israeli-born computer scientist [Daphne Koller](#) became a major contributor to probabilistic graphical models and their application to complex problems in artificial intelligence [Koller and Friedman \[2009\]](#), while the Iranian-born researcher [Zoubin Ghahramani](#), also Vice President of Research at Google DeepMind, contributed to variational inference, probabilistic machine learning and Bayesian nonparametric methods [Jordan et al. \[1999\]](#). Bayesian decision-making also involved women in major applications: **Colleen Keller** was one of the analysts and authors of the Bayesian search analysis used in locating the wreckage of Air France Flight AF447 [Stone et al. \[2014\]](#).

Koller and Friedman's *Probabilistic Graphical Models* is a standard text for structured probabilistic models [Koller and Friedman \[2009\]](#), Murphy's *Probabilistic Machine Learning: An Introduction* gives a broad modern route through probabilistic prediction and learning [Murphy \[2022\]](#), Berger's *Statistical Decision Theory and Bayesian Analysis* remains a strong reference for the decision-theoretic side [Berger \[1985\]](#), and Garnett's *Bayesian Optimization* is a focused modern text on sequential design and optimisation [Garnett \[2023\]](#).

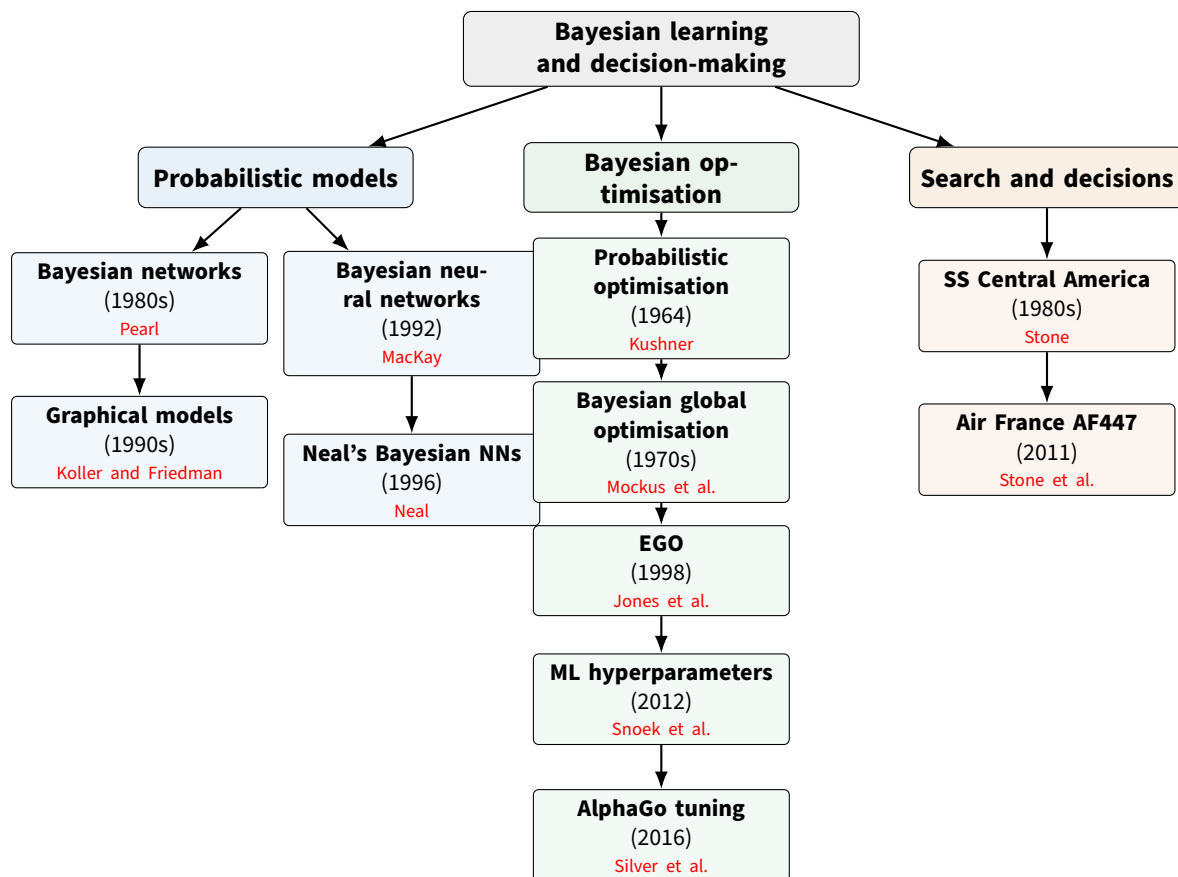


Figure 52: Selected developments and applications of Bayesian ideas in machine learning and decision-making.

Bayesian methods today. Bayesian research continues to develop rather than forming a finished theory. For readers interested in current work, the [International Society for Bayesian Analysis \(ISBA\)](#) maintains conferences, workshops and specialist research communities, while its journal *Bayesian Analysis* publishes current methodological and applied research. Active university groups include [UCL Statistical Science](#) and the [Bayesian Computational Statistics and Modeling group at KAUST](#), which develops scalable Bayesian methods including R-INLA, spatial and spatio-temporal models and approximate inference. KAUST also hosts a [Bayesian Deep Learning group](#), illustrating the continuing interaction between Bayesian inference, Gaussian processes and modern neural networks.

7.8 Conclusions

From Bayes and Laplace to modern computation and machine learning, Bayesian inference has developed into a general framework for **learning from evidence under uncertainty**.

Bayesian ideas have also returned to a much older question: *how do humans reason under uncertainty?* The **psychologist Daniel Kahneman** (1934–2024), whose work with Amos Tversky transformed the study of judgement and decision-making, explicitly describes Bayes’ rule as the normative rule for combining prior beliefs with the diagnosticity of new evidence, while noting that *intuitive judgement* often fails to do this correctly [[Kahneman, 2011](#), p. 154].

More recently, the psychologist and **neuroscientist Lisa Feldman Barrett** (b. 1963) has connected Bayesian ideas with theories of perception and learning. Barrett describes the brain as a predictive system that uses past experience to form hypotheses about incoming sensory information and revises those predictions in light of new evidence [[Barrett 2017](#)]. In the scientific notes accompanying *How Emotions Are Made*, she explicitly connects this predictive-coding framework with Bayes’ theorem and Bayesian inference.

Thus, Bayesian reasoning returns to its original problem: **How should we change what we believe when new evidence arrives?**

References

- Bayes' theorem and the billiard table. YouTube video. URL <https://www.youtube.com/watch?v=rwtDBhD6Mq0>.
- Maria Gaetana Agnesi. *Instituzioni analitiche ad uso della gioventù italiana*. Nella Regia-Ducal Corte, Milan, 1748.
- John Aldrich. R. A. Fisher and the making of maximum likelihood 1912–1922. *Statistical Science*, 12(3):162–176, 1997. doi: 10.1214/ss/1030037906.
- Amir Alexander. *Infinitesimal: How a Dangerous Mathematical Theory Shaped the Modern World*. Scientific American/-Farrar, Straus and Giroux, New York, 2014.
- C. H. O'D. Alexander. Cryptographic history of work on the german naval enigma, 1945. URL <https://www.ellsbury.com/gne/gne-000.htm>. UK National Archives, Kew, reference HW 25/1. Wartime/post-war historical report, c. 1945.
- Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010. doi: 10.1111/j.1467-9868.2009.00736.x.
- Charles E. Antoniak. Mixtures of dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974. doi: 10.1214/AOS/1176342871.
- 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.
- Margaret E. Baron. *The Origins of the Infinitesimal Calculus*. Pergamon Press, Oxford, 1969.
- Lisa Feldman Barrett. *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt, Boston, 2017.
- Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763. doi: 10.1098/rstl.1763.0053.
- Tom Begley. Bayesian billiards, December 2020. URL <https://www.tcbegley.com/blog/posts/bayesian-billiards>.
- Richard Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the United States of America*, 38(8):716–719, 1952. doi: 10.1073/pnas.38.8.716.
- Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.
- James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2 edition, 1985. doi: 10.1007/978-1-4757-4286-2.
- 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.
- Joseph Bertrand. *Calcul des probabilités*. Gauthier-Villars, Paris, 1889.
- Irénée-Jules Bienaymé. Considérations à l'appui de la découverte de laplace sur la loi de probabilité dans la méthode des moindres carrés. *Comptes Rendus de l'Académie des Sciences*, 37:309–324, 1853.
- Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3): 637–654, 1973. doi: 10.1086/260062.
- David Blackwell and Meyer A. Girshick. *Theory of Games and Statistical Decisions*. John Wiley and Sons, New York, 1954.

- Viktor Blåsjö. A critique of the arguments for maragha influence on copernicus. *Journal for the History of Astronomy*, 45(2):183–195, 2014.
- George Boole. *An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*. Walton and Maberly, London, 1854. URL https://archive.org/details/bub_gb_pFsmwHjdHSsC.
- Émile Borel. *Éléments de la théorie des probabilités*. A. Hermann, Paris, 1909.
- Umberto Bottazzini. *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*. Springer, New York, 1986.
- William E. Boyce, Richard C. DiPrima, and Douglas B. Meade. *Elementary Differential Equations and Boundary Value Problems*. Wiley, Hoboken, NJ, 11 edition, 2017.
- Carl B. Boyer. *The History of the Calculus and Its Conceptual Development*. Dover Publications, New York, 1959.
- David M. Bressoud. *Calculus Reordered: A History of the Big Ideas*. Princeton University Press, Princeton, 2019.
- Louis L. Bucciarelli and Nancy Dworsky. *Sophie Germain: An Essay in the History of the Theory of Elasticity*. D. Reidel Publishing Company, Dordrecht, 1980.
- Maarten Bullynck. Modular arithmetic before c.f. gauss: Systematizations and discussions on remainder problems in 18th-century germany. *Historica Mathematica* 36, 2008.
- Richard L. Burden and J. Douglas Faires. *Numerical Analysis*. Brooks/Cole, Boston, 9 edition, 2011.
- Francesco Paolo Cantelli. Sulla legge forte dei grandi numeri. *Rendiconti della Reale Accademia dei Lincei*, 26:39–45, 1917. Strong law under moment conditions.
- José Chabás and Bernard R. Goldstein. *A Survey of European Astronomical Tables in the Late Middle Ages, volume 2 of Time, Astronomy, and Calendars: Texts and Studies*. Brill, Leiden and Boston, 2012. ISBN 9789004230583. doi: 10.1163/9789004230590. URL <https://brill.com/display/title/21745>.
- 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.
- Pafnuty L. Chebyshev. Des valeurs moyennes. *Journal de Mathématiques Pures et Appliquées*, 12:177–184, 1867.
- Cheenta Academy. Bayes and the billiard table. URL <https://cheenta.com/bayes-and-the-billiard-table-cheenta-probability-series/>.
- Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Springer, Cham, 2020. doi: 10.1007/978-3-030-47845-2.
- 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, Hoboken, NJ, 3 edition, 2013a.
- Roger L. Cooke. *The History of Mathematics: A Brief Course*. John Wiley & Sons, third edition, 2013b.
- Leon Cooper. Did egyptian scribes have an algorithmic means for determining the circumference of a circle? *Historia Mathematica*, 38:455–484, 2011.
- Cornell University INFO 2040 Blog. Bayes' theorem and the search for missing planes. <https://blogs.cornell.edu/info2040/2015/11/04/bayes-theorem-and-the-the-search-for-missing-planes/>, 2015.
- David Cox, John Little, Donal O'shea, and Moss Sweedler. *Ideals, varieties, and algorithms*. Springer, 1997.
- Datta & Singh. *Histry 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.
- Sam Davey, Neil Gordon, Ian Holland, Mark Rutten, and Jason Williams. *Bayesian Methods in the Search for MH370*. Springer Nature, 2016.

- 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.
- Marcio Alves Diniz and David Richard Bellhouse. Mr. bayes and the classics: A suggested interpretation. *Revista Brasileira de História da Matemática*, 22(43):1–16, 2022.
- 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>.
- Anthony Duane, Brian D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987. doi: 10.1016/0370-2693(87)91197-X.
- William Dunham. *The Calculus Gallery: Masterpieces from Newton to Lebesgue*. Princeton University Press, Princeton, 2005.
- C. H. Edwards. *The Historical Development of the Calculus*. Springer, New York, 1979. doi: 10.1007/978-1-4612-6230-5.
- Alexander Etz and Eric-Jan Wagenmakers. J. B. S. Haldane’s contribution to the bayes factor hypothesis test. *Statistical Science*, 32(2):313–329, 2017. doi: 10.1214/16-STS599.
- Leonhard Euler. Solutio problematis arithmetici de inveniendō 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>.
- Nicolas Farès. Le calcul du maximum et la “dérivée” selon sharaf al-din al-tusi. *Arabic Sciences and Philosophy*, 5(2): 219–237, 1995.
- 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.
- Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973. doi: 10.1214/aos/1176342360.
- Hans Fischer. *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. Springer, New York, 2 edition, 2010.
- Ronald A. Fisher. *On the Mathematical Foundations of Theoretical Statistics*, volume 222. 1922a. doi: 10.1098/rsta.1922.0009.
- Ronald A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 222:309–368, 1922b. doi: 10.1098/rsta.1922.0009.
- Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925.
- Ronald A. Fisher. Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26(4):528–535, 1930. doi: 10.1017/S0305004100016297.
- Ronald A. Fisher. The fiducial argument in statistical inference. *Annals of Eugenics*, 6(4):391–398, 1935. doi: 10.1111/j.1469-1809.1935.tb02120.x.
- Ronald A. Fisher and Winifred A. Mackenzie. Studies in crop variation. II. the manurial response of different potato varieties. *The Journal of Agricultural Science*, 13(3):311–320, 1923. doi: 10.1017/S0021859600003592.
- John Flamsteed. *Historia Coelestis Britannica*. H. Meere, London, 1725. Contains astronomical observations made primarily at Greenwich, including observations of planets and stellar positions.
- 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.

- James Franklin. *The Science of Conjecture: Evidence and Probability before Pascal*. Johns Hopkins University Press, 2001.
- Hannah Fry and Matt Parker. Bayes' theorem. YouTube video. URL <https://www.youtube.com/watch?v=7GgLSnQ48os>.
- Roman Garnett. *Bayesian Optimization*. Cambridge University Press, Cambridge, 2023.
- Carl Friedrich Gauss. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Friedrich Perthes and I. H. Besser, Hamburg, 1809. URL <https://books.google.com/books?id=ORUOAAAAQAAJ>.
- GCHQ. Celebrating 100 years of women's service in the uk armed forces, 2017. URL <https://www.gchq.gov.uk/information/centenary-womens-service-uk-armed-forces>.
- Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, third edition, 2013.
- Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984. doi: 10.1109/TPAMI.1984.4767596.
- 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.
- I. J. Good. *Probability and the Weighing of Evidence*. Charles Griffin and Company, London, 1950.
- I. J. Good. Studies in the history of probability and statistics. xxxvii. a. m. turing's statistical work in world war ii. *Biometrika*, 66(2):393–396, 1979. doi: 10.1093/biomet/66.2.393.
- N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F: Radar and Signal Processing*, 140(2):107–113, 1993. doi: 10.1049/ip-f-2.1993.0015.
- Judith V. Grabiner. *The Origins of Cauchy's Rigorous Calculus*. MIT Press, Cambridge, MA, 1981.
- Andrew Granville. It is not "bézout's identity", 2024. URL <https://arxiv.org/abs/2406.15642>.
- Ivor Grattan-Guinness. *From the Calculus to Set Theory, 1630–1910: An Introductory History*. Duckworth, London, 1980.
- 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.
- James E. Gubernatis. Marshall Rosenbluth and the Metropolis algorithm. *Physics of Plasmas*, 12(5):057303, 2005. doi: 10.1063/1.1887186.
- Ian Hacking. *The Emergence of Probability*. Cambridge University Press, 1975.
- J. B. S. Haldane. A note on inverse probability. *Proceedings of the Cambridge Philosophical Society*, 28(1):55–61, 1932. doi: 10.1017/S0305004100010495.
- Mitchell A Harris and Edward M Reingold. Line drawing, leap years, and euclid. *ACM Computing Surveys (CSUR)*, 36(1): 68–80, 2004.
- W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: 10.1093/biomet/57.1.97.
- 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].
- Claudia Henrion. *Women in Mathematics: The Addition of Difference*. Indiana University Press, Bloomington, 1997.

-
- David Hilbert. Mathematische probleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 253–297, 1900.
- F. H. Hinsley and Alan Stripp, editors. *Codebreakers: The Inside Story of Bletchley Park*. Oxford University Press, 1993.
- Nils Lid Hjort, Chris Holmes, Peter Müller, and Stephen G. Walker, editors. *Bayesian Nonparametrics*. Cambridge University Press, Cambridge, 2010. doi: 10.1017/CBO9780511802478.
- Andrew Hodges. *Alan Turing: The Enigma*. Princeton University Press, 2014.
- Peter D Hoff. *A first course in Bayesian statistical methods*, volume 580. Springer, 2009.
- I. Hogendijk. Pure mathematics in islamic civilization. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, 1994.
- Jan P. Hogendijk. Sharaf al-din al-tusi on the number of positive roots of cubic equations. *Historia Mathematica*, 16 (1):69–85, 1989.
- David Hume. *An Enquiry Concerning Human Understanding*. A. Millar, London, 1748.
- Iamblichus. *The life of Pythagoras*. Krotona ; Hollywood, Calif. : Theosophical Pub. House, 1918.
- Kiyosi Itô. *On Stochastic Differential Equations*, volume 4 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence, RI, 1951.
- Kiyosi Itô and Henry P. McKean. *Diffusion Processes and Their Sample Paths*. Springer, Berlin, 1965.
- Suzuki Jeff. Al khayyami and the cubic: Setup. [Online; accessed: 2024-09-13], 2020. URL <https://www.youtube.com/watch?v=d553qUklUEY>.
- Harold Jeffreys. Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophical Society*, 31(2):203–222, 1935. doi: 10.1017/S030500410001330X.
- Harold Jeffreys. *Theory of Probability*. Clarendon Press, Oxford, 1939.
- Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A*, 186(1007):453–461, 1946. doi: 10.1098/rspa.1946.0056.
- Harold Jeffreys. *Scientific Inference*. Cambridge University Press, Cambridge, 3 edition, 1973.
- Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998. doi: 10.1023/A:1008306431147.
- Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999. doi: 10.1023/A:1007665907178. URL <https://mlanthology.org/mlj/1999/jordan1999mlj-introduction/>.
- George Gheverghese Joseph. *The Crest of the Peacock: Non-European Roots of Mathematics*. Princeton University Press, 2010.
- George Gheverghese Joseph. *The Crest of the Peacock: Non-European Roots of Mathematics*. Princeton University Press, Princeton and Oxford, 3 edition, 2011.
- Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, 2011.
- R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960. doi: 10.1115/1.3662552.
- Victor J. Katz. Ideas of calculus in islam and india. *Mathematics Magazine*, 68(3):163–174, 1995. doi: 10.2307/2691411.
- Victor J. Katz. *A History of Mathematics: An Introduction*. Addison-Wesley, Boston, 3 edition, 2009.
- Victor J. Katz. Mathematical treasures: Maria agnesi’s analytical institutions, 2011. URL <https://old.maa.org/press/periodicals/convergence/mathematical-treasures-maria-agnesis-analytical-institutions>.
- John Maynard Keynes. *A Treatise on Probability*. Macmillan and Co., London, 1921.

- 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).
- George S. Kimeldorf and Grace Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970. doi: 10.1214/aoms/1177697089.
- Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, Massachusetts, 2009.
- Andrey Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, 1933a. Foundations of measure-theoretic probability; formalizes SLLN and CLT.
- Andrey N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, volume 2 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Julius Springer, Berlin, 1933b.
- J. Kraft and L. Washington. *An Introduction to Number Theory with Cryptography*. Chapman and Hall/CRC, 2018.
- Harold J. Kushner. A new method of locating the maximum point of an arbitrary multippeak curve in the presence of noise. *Journal of Basic Engineering*, 86(1):97–106, 1964. doi: 10.1115/1.3653121.
- 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. Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences de Paris*, pages 621–656, 1774. URL https://fr.wikisource.org/wiki/Page:Laplace_-_%C5%92uvres_compl%C3%A8tes,_Gauthier-Villars,_1878,_tome_8.djvu/41. Reprinted in *Oeuvres complètes de Laplace*, Tome VIII, Gauthier-Villars, Paris, 1878.
- Pierre-Simon Laplace. Mémoire sur les inégalités séculaires des planètes et des satellites. *Mémoires de l'Académie royale des sciences de Paris*, pages 1–50, 1787. Volume for the year 1784.
- Pierre-Simon Laplace. Théorie de jupiter et de saturne. *Mémoires de l'Académie royale des sciences de Paris*, pages 33–160, 1788. Volume for the year 1785.
- Pierre-Simon Laplace. *Théorie Analytique des Probabilités*. Courcier, Paris, 1812.
- Erich L. Lehmann. *Fisher, Neyman, and the Creation of Classical Statistics*. Springer, New York, 2011. doi: 10.1007/978-1-4419-9500-1.
- D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(1):1–41, 1972. doi: 10.1111/j.2517-6161.1972.tb00885.x.
- 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.
- David Lunn, David Spiegelhalter, Andrew Thomas, and Nicky Best. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067, 2009. doi: 10.1002/sim.3680.
- David J. Lunn, Andrew Thomas, Nicky Best, and David Spiegelhalter. WinBUGS — a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337, 2000. doi: 10.1023/A:1008929526011.
- David J. C. MacKay. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3):448–472, 1992. doi: 10.1162/neco.1992.4.3.448.
- A. P. Mahon. The history of hut eight, 1939–1945, 1945. URL <https://discovery.nationalarchives.gov.uk/details/r/C6136632>. UK National Archives, Kew, reference HW 25/2.
- Eli Maor. *The Pythagorean theorem: a 4,000-year history*, volume 65. Princeton University Press, 2019.

- Mathematical Association of America. Sums of powers of positive integers: Abu ali al-hasan ibn al-hasan ibn al-haytham, 2014. URL <https://old.maa.org/press/periodicals/convergence/sums-of-powers-of-positive-integers-abu-ali-al-hasan-ibn-al-hasan-ibn-al-haytham-965-1039-egypt>.
- Mathematical Association of America. Mathematical treasure: Maria agnesi's analytical institutions in english, 2018. URL <https://old.maa.org/press/periodicals/convergence/mathematical-treasure-maria-agnesi-s-analytical-institutions-in-english>.
- Betty Mayfield. Women and mathematics in the time of euler. *The College Mathematics Journal*, 2013.
- Massimo Mazzotti. *The World of Maria Gaetana Agnesi, Mathematician of God*. Johns Hopkins University Press, Baltimore, 2007.
- Sharon Bertsch McGrayne. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, New Haven, 2011.
- Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114.
- Thomas Mikosch. *Elementary Stochastic Calculus with Finance in View*. World Scientific, Singapore, 1998.
- Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. The application of Bayesian methods for seeking the extremum. In L. C. W. Dixon and G. P. Szegö, editors, *Towards Global Optimization, Volume 2*, pages 117–129. North-Holland, Amsterdam, 1978.
- 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.
- Kevin P. Murphy. *Probabilistic Machine Learning: An Introduction*. MIT Press, Cambridge, Massachusetts, 2022.
- Radford M. Neal. *Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes in Statistics*. Springer, New York, 1996. doi: 10.1007/978-1-4612-0745-0.
- 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.
- Jerzy Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A*, 236:333–380, 1937. doi: 10.1098/rsta.1937.0005.
- Jerzy Neyman and Egon S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference. part i. *Biometrika*, 20A(1–2):175–240, 1928. doi: 10.1093/biomet/20A.1-2.175.
- Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231:289–337, 1933. doi: 10.1098/rsta.1933.0009.
- Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, 2 edition, 2006.
- John J. O'Connor and Edmund F. Robertson. Marie-sophie germain, a. URL <https://mathshistory.st-andrews.ac.uk/Biographies/Germain/>.
- John J. O'Connor and Edmund F. Robertson. Kiyosi itô, b. URL <https://mathshistory.st-andrews.ac.uk/Biographies/Ito/>.
- John J. O'Connor and Edmund F. Robertson. Mary somerville, c. URL <https://mathshistory.st-andrews.ac.uk/Biographies/Somerville/>.
- Anthony O'Hagan. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1):1–42, 1978. doi: 10.1111/j.2517-6161.1978.tb01643.x.

- Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 6 edition, 2003.
- Lynn M. Osen. *Women in Mathematics*. MIT Press, Cambridge, MA, 1974.
- A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. Electrical Engineering Series. McGraw-Hill, 1991. ISBN 9780070484771. URL <https://books.google.co.uk/books?id=4IwQAQAIAAJ>.
- Giuditta Parolini. From computing girls to data processors: Women assistants in the rothamsted statistics department. In Valérie Schafer and Benjamin G. Thierry, editors, *Connecting Women: Women, Gender and ICT in Europe in the Nineteenth and Twentieth Century*, pages 103–117. Springer International Publishing, Cham, 2015. doi: 10.1007/978-3-319-20837-4_7.
- Karen Hunger Parshall and David E. Rowe. *The Emergence of the American Mathematical Research Community, 1876–1900: J. J. Sylvester, Felix Klein, and E. H. Moore*. American Mathematical Society and London Mathematical Society, Providence, RI, 1994.
- Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.
- Jose D. Perezgonzalez. Fisher, neyman–pearson or NHST? a tutorial for teaching data testing. *Frontiers in Psychology*, 6:223, 2015. doi: 10.3389/fpsyg.2015.00223.
- K. Plofker. *Mathematics in India*. Princeton University Press, 2009a.
- Kim Plofker. *Mathematics in India*. Princeton University Press, Princeton and Oxford, 2009b.
- S. D. Poisson. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Bachelier, 1837.
- Christopher Powell. *The Shapes of Sacred Space: A Proposed System of Geometry Used to Lay Out and Design Maya Art and Architecture 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.
- Adolphe Quetelet. *Sur l’homme et le développement de ses facultés, ou Essai de physique sociale*. Bachelier, Paris, 1835. URL <https://archive.org/details/surhommeetled00quet>.
- Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Massachusetts, 2006. doi: 10.7551/mitpress/3206.001.0001.
- Sylvia Richardson and Peter J. Green. On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):731–792, 1997. doi: 10.1111/1467-9868.00095.
- Herbert Robbins. An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume I*, pages 157–163, Berkeley, California, 1956. University of California Press. doi: 10.1525/9780520313880-015.
- 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>.
- Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, 2nd edition, 2004. doi: 10.1007/978-1-4757-4145-2.
- Eleanor Robson and Jacqueline Stedall, editors. *The Oxford Handbook of the History of Mathematics*. Oxford University Press, Oxford, 2009.
- 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>.

- Timothy Sauer. *Numerical Analysis*. Pearson, Boston, 3 edition, 2017.
- Leonard J. Savage. *The Foundations of Statistics*. John Wiley & Sons, New York, 1954.
- 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>.
- David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016a. doi: 10.1038/nature16961.
- David Silver, Aja Huang, Chris J Maddison, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016b.
- Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25*, 2012.
- Mary Somerville. *The Mechanism of the Heavens*. John Murray, London, 1831.
- Brigitte Stenhouse. Mary somerville’s early contributions to the circulation of differential calculus. *Historia Mathematica*, 51:1–25, 2020. doi: 10.1016/j.hm.2019.12.001.
- Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, 1986a.
- Stephen M. Stigler. Laplace’s 1774 memoir on inverse probability. *Statistical Science*, 1(3):359–378, 1986b. doi: 10.1214/ss/1177013620. URL <https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/Laplaces-1774-Memoir-on-Inverse-Probability/10.1214/ss/1177013620.full>.
- Stephen M. Stigler. The true title of bayes’s essay. *Statistical Science*, 28(3):283–288, 2013. doi: 10.1214/13-STS438.
- Lawrence D. Stone. Search for the SS central america: Mathematical treasure hunting. *Interfaces*, 22(1):32–54, 1992. doi: 10.1287/inte.22.1.32.
- Lawrence D Stone, Carl M Keller, Thomas M Kratzke, and Johan Strumpf. Search for the wreckage of air france flight 447. *Statistical Science*, 29(1):69–80, 2014.
- P. D. Straffin. *Liu Hui and the First Golden Age of Chinese Mathematics*, volume 71. Mathematics Magazine, 1998a.
- Phillip D. Straffin. Liu hui and the first golden age of chinese mathematics. *Mathematics Magazine*, 1998b.
- Gilbert Strang and Edwin “Jed” Herman. *Calculus Volume 1*. OpenStax, Rice University, Houston, TX, 2016. URL <https://openstax.org/details/books/calculus-volume-1>.
- 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).
- Ursula Winifred Taylor and John Aldrich. Statistician, missionary, mother. *Significance*, 19(5):35–37, 2022. doi: 10.1111/1740-9713.01689.
- Morris Tenenbaum and Harry Pollard. *Ordinary Differential Equations*. Dover Publications, New York, 1985.
- G. Thibaut. *The Śūlasūtra of Baudhāyana, with the commentary by Dvārakānāthayajvan*, volume 9. The Pandit, 1875.
- Henk Tijms. *Understanding Probability*. Cambridge University Press, Cambridge, 3 edition, 2012. ISBN 9781107658561.
- Stephen P. Timoshenko. *History of Strength of Materials*. McGraw-Hill, New York, 1953.

-
- C. Truesdell. Maria gaetana agnesi. *Archive for History of Exact Sciences*, 1989.
- Alan M. Turing. The applications of probability to cryptography, 1941. URL <https://arxiv.org/abs/1505.04714>. Wartime manuscript, c. 1941. UK National Archives, reference HW 25/37. Typeset edition prepared by Ian Taylor, 2015.
- Università di Bologna. Progetto maria gaetana agnesi. URL <https://site.unibo.it/progetto-agnesi/it/il-progetto>.
- John Venn. *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to its Application to Moral and Social Science*. Macmillan and Co., London and Cambridge, 1866.
- Richard von Mises. Grundlagen der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919. doi: 10.1007/BF01203155.
- Joachim Von Zur Gathen and Jürgen Gerhard. *Modern computer algebra*. Cambridge university press, 2003.
- Abraham Wald. Contributions to the theory of statistical estimation and testing hypotheses. *The Annals of Mathematical Statistics*, 10(4):299–326, 1939.
- Abraham Wald. *Statistical Decision Functions*. John Wiley & Sons, New York, 1950.
- E. W. Weissten. Pythagorean triple. [Online; accessed 2024-07-10], 2024. URL <https://mathworld.wolfram.com/PythagoreanTriple.html>.
- Yuval Wigderson. Euxodus, the most important maethematican you’ve never heard of. *Stanford Student Analysis Seminar*, 2019.
- 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>.
- Dorothy Wrinch and Harold Jeffreys. On some aspects of the theory of probability. *Philosophical Magazine*, 38(228): 715–731, 1919. doi: 10.1080/14786441208636005.
- Dorothy Wrinch and Harold Jeffreys. On certain fundamental principles of scientific inquiry. *Philosophical Magazine*, 42(249):369–390, 1921. doi: 10.1080/14786442108633773.
- Dorothy Wrinch and Harold Jeffreys. On certain fundamental principles of scientific inquiry. second paper. *Philosophical Magazine*, 45(266):368–374, 1923. doi: 10.1080/14786442308634125.
- 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>.
- Sandy Zabell. Commentary on alan m. turing: The applications of probability to cryptography. *Cryptologia*, 36(3): 191–214, 2012. doi: 10.1080/01611194.2012.697811.
- 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.