

Probability and Primality

Abstract

Prime numbers are among the first concepts in number theory that students encounter at school, where they usually work with them only through a limited range of algebraic operations. However, their applications, especially the mathematics behind those applications, receive far less attention during the school years. This is despite the fact that many digital-security systems rely on mathematical structures involving prime numbers, and some related algorithms require computers to generate extremely large primes and assess their validity. Using an educational approach suitable for the high-school level, this article explains the concept of a prime number, a simple method for determining whether a number is prime, how very large prime numbers are generated and their primality assessed, and how such numbers are used in data security. The central theme is the role of probability and its connection to primality testing.

1) Introduction

Probability is commonly introduced in school curricula through coin tosses, dice, and card games [1]. These examples are useful for building a theoretical foundation, but they can create the impression that probability applies only to phenomena that are inherently random. Primality testing for large integers provides an interesting counterexample. A fixed integer is either prime or composite; its status is not random. Nevertheless, an algorithm can use random choices to test it quickly and reduce its probability of error to an arbitrarily small level.

We begin with a review of the familiar concept of a prime number. We then discuss division-based methods for determining whether an integer is prime, the challenge of finding primes with very large numbers of digits, the use of such primes in digital technologies, and the important role of probability both in estimating how many candidates must be searched before a large prime is found and in testing whether a candidate is prime.

The article concludes with a brief remembrance of Farideh Firoozbakht and her remarkably elegant conjecture concerning prime numbers, followed by a note on Paul Erdős and his contributions to the study of primes.

2) Trial Division

A prime number is a number whose only divisors are 1 and the number itself. A number that is not prime is called composite. Suppose we want to determine whether 997 is prime. If a number is composite, it can be written in the form $n = ab$.

If both factors a and b were greater than $\sqrt{n}$, their product would be greater than n . Therefore, if n is composite, at least one of its factors is no greater than $\sqrt{n}$. Consequently, if an integer is not divisible by any integer greater than 1 and no greater

than its square root, it is prime. In practice, these trial divisions can be restricted to prime numbers no greater than the square root of the number being tested [2].

For 997, we have $\sqrt{997} \approx 31.6$. Therefore, it is enough to test whether 997 is divisible by the primes no greater than 31.6, namely 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31.

The number 997 is divisible by none of these primes. Therefore, 997 is prime.

By contrast, for 899 we have $899 = 29 \times 31$. Therefore, 899 is composite.

This method of determining whether an integer is prime is known as trial division.

3. The Number of Primes Below a Given Number

Let $g(x)$ denote the number of primes less than or equal to x . In the worst case when testing the primality of n , the number of required divisions is $\pi(\sqrt{n})$. A fundamental theorem about prime numbers states that, for very large values of x ,

$$\pi(x) \approx \frac{x}{\ln x}.$$

In practical terms, the theorem says that for a very large number, the number of primes below it is approximately $\frac{x}{\ln x}$. This is only an asymptotic approximation and is intended for very large values.

For example, for the prime number $x = 999999999999999999999999999973$, the quantity $\frac{x}{\ln x}$ is approximately 1.498×10^{27} . The exact number of primes below $x = 999999999999999999999999999973$ is 1520698109714272166094258063, which is very close to 1.498×10^{27} [2].

Thus, for a very large number n , we have

$$\pi(\sqrt{n}) \approx \frac{\sqrt{n}}{\ln \sqrt{n}} = \frac{2\sqrt{n}}{\ln n}.$$

In other words, under trial division, the primality of a very large integer n can, in the worst case (when the integer really is prime), be determined by carrying out approximately $\frac{2\sqrt{n}}{\ln n}$ divisions by primes less than or equal to $\sqrt{n}$.

For example, suppose $n \approx 10^{100}$ (that is, a particular integer whose magnitude is close to 10^{100} , or, equivalently, a number with about one hundred digits). Then $\sqrt{n} \approx 10^{50}$, and the number of primes up to 10^{50} is approximately

$$\pi(10^{50}) \approx \frac{10^{50}}{\ln(10^{50})} = \frac{10^{50}}{50 \ln 10} \approx 8.69 \times 10^{47}.$$

Accordingly, when testing the primality of n , suppose, very optimistically, that a computer can perform 10^9 large-integer divisions per second. The time required, in seconds, to carry out the divisions over all relevant numbers below 10^{50} would then be approximately

$$\frac{8.69 \times 10^{47}}{10^9} = 8.69 \times 10^{38}$$



Figure 1) Gary L. Miller, computer scientist and originator of the primality test later developed into the Miller–Rabin algorithm. Photo by Andrej Bauer, 2004 (CC BY-SA 2.5 SI).

Converting this value to years gives 2.75×10^{31} . Thus, it would take 2.75×10^{31} years to establish the primality of a number on the order of 10^{100} by this approach.

This leads to two fundamental questions. First, do extremely large prime numbers have practical uses that make it worthwhile to identify them? If the answer is yes, how can such primes be generated and recognized efficiently?

In fact, very large primes are much closer to everyday life than one might imagine: we use systems that depend on them routinely, often without being aware of it. The following section introduces several of their applications, after which we return to the questions of generating primes and testing primality. What is already clear is that trial division is not a practical method for testing extremely large integers.

4. The Role of Prime Numbers in Digital Security

Very large prime numbers have many applications in data transmission in today's digital world. Processes involving such numbers are used in information-transfer and encryption systems [3], [4]. In some applications, the prime involved is a fixed, preselected number. The prime $p = 2^{255} - 19$, which has 77 digits, is used in WhatsApp data encryption [5]. Thus, whenever WhatsApp is used, a particular very large prime plays a role in the information-transmission process. Telegram's Secret Chat service uses a 617-digit prime [6]; writing that number out in full would require six or seven lines of this article.



Figure 2) Michael O. Rabin (1931–2026), pioneer of randomized algorithms and co-creator of the Miller–Rabin primality test. Photo by Andrej Bauer, 2004 (CC BY-SA 2.5 SI).

In addition to applications that rely on primes generated in advance, some applications require prime numbers to be generated during operation. For example, when users connect to certain websites and online stores, special security protocols may be used to help ensure that the connection is being made to the site's authentic server. Some such protocols rely on the random generation of very large prime numbers.

The foundations of prime-number-based cryptography, the methods used in such systems, and the details of how they work are topics for a separate discussion.

5. Generating Prime Numbers

To generate a random prime number of a desired length (for example, a seven-hundred-digit prime), a computer first produces a number of that length that is not divisible by small primes such as 2, 3, 5, 7, 11, and 13. It then applies a primality test such as the Miller–Rabin algorithm [7], [8]. The way this algorithm works is a particularly elegant application of probability. The test has one-sided error: whenever it declares a number composite, that conclusion is certain and the number really is composite. If, however, it returns a prime result, it can be wrong with probability at most $\frac{1}{4}$. If the same candidate is tested a second, third, and subsequent time, each run is independent of the others, and on each run the probability of incorrectly returning a prime result is again at most $\frac{1}{4}$. Consequently, if the algorithm reports a prime result in two consecutive runs, the probability of misclassification is at most $\frac{1}{16}$; after three runs, it is at most $\frac{1}{64}$; and so forth. Thus, if repeated runs of the algorithm, each taking very little time, consistently return a prime result, the candidate is prime with extremely high probability.

If the algorithm determines that the selected candidate is composite, the process continues with a new candidate.

A roughly 300-digit (1024-bit) number is on the order of $x \approx 2^{1024}$. If the search is restricted to odd numbers (without, for example, checking divisibility by 3, 5, 7, and so on), then among integers of this magnitude, approximately $\frac{2}{\ln x}$ will be prime, based on the discussion above. Therefore, assuming that primes are uniformly dispersed for this estimate, a prime should be found after approximately $\frac{\ln x}{2} = \frac{\ln 2^{1024}}{2} = \frac{1024 \ln 2}{2} \approx 355$ candidate searches.

6. The Logic Behind the Miller–Rabin Test

Let C be the event that the number is composite, and let W_k be the event that the number passes the k th run of the Miller–Rabin algorithm (that is, on the k th run, the number is reported as prime). From the preceding discussion, $p(W_k|C) \leq \frac{1}{4}$. Assuming independence across repeated runs [7], [8],

$$p(W_1 W_2 \cdots W_n | C) = p(W_1 | C) p(W_2 | C) \cdots p(W_n | C) \leq \frac{1}{4^n}$$

After $N = 64$ runs of the algorithm, the probability that a composite number passes all 64 stages as though it were prime is less than $1/2^{128}$. Thus, after this many Miller–Rabin runs, if the selected number passes every run, one can be confident to an extremely high degree that the number is indeed prime. (There are also supplementary methods that can provide a definitive guarantee of primality.)

7. Conclusion

Primality testing for large integers is a powerful example of probability in computation. In the elementary, school-level method, for an integer n , it is sufficient to test prime divisors up to $\sqrt{n}$, and the number of such tests in the worst case is approximately $\pi(\sqrt{n})$. Using the prime number theorem, for large integers this quantity is estimated by approximately $2\sqrt{n}/\ln n$; for numbers with hundreds of digits, this is still computationally impractical.

Miller–Rabin takes a different approach. Rather than directly searching through all possible factors, it uses fast tests with randomly chosen inputs. A result of "composite" is definitive, whereas a result of "probably prime" has a known upper bound on its error probability. With independent repetition, that bound decreases exponentially.

This illustrates an important point: probability does not always represent uncertainty about the phenomenon itself. Sometimes it is a tool for designing an algorithm that is fast, controllable, and, for practical purposes, highly reliable.



Figure 3) Conceptual illustration inspired by Farideh Firoozbakht's still-open conjecture, adapted from "[Isfahan – 33 Pol – Siyosepol](#)" by Alireza Javaheri (CC BY 3.0).

Farideh Firoozbakht and Her Conjecture

Farideh Firoozbakht (1962–2019) was an Iranian mathematician, researcher, and teacher associated with the University of Isfahan and the Isfahan Mathematics House. She was also a contributor and editor to the Online Encyclopedia of Integer Sequences (OEIS) [9]. According to an autobiographical profile, she entered the University of Isfahan to study pharmacy, but her strong interest in mathematics, especially number theory, led her to change her major to mathematics in her third year; she graduated in mathematics in 1987 [10].

Her name is most closely associated with Firoozbakht's conjecture, which she proposed around 1982 concerning the distribution of prime numbers [10]. If p_n denotes the n th prime in the sequence of primes, Firoozbakht's conjecture states that [11], [12]:

$$\sqrt[n]{p_n} > \sqrt[n+1]{p_{n+1}}$$

In other words, the sequence $\{\sqrt[n]{p_n}\}$ is strictly decreasing. This inequality has been computationally verified for primes below 2^{64} [13]. No generally accepted proof or disproof is currently known. The conjecture is also noteworthy because it implies strong upper bounds on gaps between consecutive primes [14].



Figure 4) Paul Erdős (1913–1996), a prolific mathematician whose work made major contributions to number theory, including the study of prime numbers and probabilistic methods. [Photo by Kmhknh, 1992 \(CC BY 3.0\)](#). *Digitally enhanced.*

Probability Beyond Primality: Erdős–Kac

The connection between prime numbers and probability extends beyond primality testing. A striking example is the Erdős–Kac theorem, established by Paul Erdős and Mark Kac [15]. If $\omega(n)$ denotes the number of distinct prime factors of n , then for integers up to a very large value x , $\omega(n)$ is approximately normally distributed, with both its mean and variance asymptotically equal to $\ln \ln x$, and therefore a standard deviation of approximately $\sqrt{\ln \ln x}$.

Thus, most large integers have a number of distinct prime factors close to $\ln \ln x$, while substantially smaller or larger values become progressively less common, producing the familiar bell-shaped pattern [15], [16]. The theorem became a cornerstone of probabilistic number theory, and its underlying ideas have since been extended to the statistical study of other arithmetic structures [18].

With Erdős’s name now part of the discussion, it is also worth noting his intriguing connection with the Monty Hall problem, discussed in the first article of this issue. According to mathematician Andrew Vázsonyi, Erdős initially did not accept that switching doors increases the probability of winning from $1/3$ to $2/3$. A computer simulation convinced him that the result was correct, although he continued to seek a satisfactory explanation of why it worked [17].

For a mathematician of Erdős’s stature, the episode serves above all as a reminder that even when mathematics is exact, probabilistic intuition can be remarkably deceptive.

References

- [1] S. M. Ross, *A First Course in Probability*, 8th ed., Prentice Hall, 2010.
- [2] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th ed., Oxford University Press, 2008.
- [3] National Institute of Standards and Technology, *Digital Signature Standard (DSS)*, FIPS PUB 186-5, 2023.
- [4] A. Langley, M. Hamburg, and S. Turner, “Elliptic Curves for Security,” RFC 7748, 2016.
- [5] WhatsApp, *WhatsApp Encryption Overview / Security Whitepaper*.
- [6] Telegram, “End-to-End Encryption, Secret Chats,” Telegram API Documentation.
- [7] G. L. Miller, “Riemann’s Hypothesis and Tests for Primality,” *Journal of Computer and System Sciences*, vol. 13, no. 3, pp. 300–317, 1976.
- [8] M. O. Rabin, “Probabilistic Algorithm for Testing Primality,” *Journal of Number Theory*, vol. 12, no. 1, pp. 128–138, 1980.
- [9] OEIS Foundation, “User: Farideh Firoozbakht,” OeisWiki. Accessed Aug. 16, 2026.
- [10] C. Rivera (ed.), “Farideh Firoozbakht,” *The Puzzlers, Prime Puzzles & Problems Connection*, updated Jan. 10, 2020.
- [11] P. Ribenboim, *The Little Book of Bigger Primes*, 2nd ed., Springer, 2004, p. 185.
- [12] A. Kourbatov, “Verification of the Firoozbakht Conjecture for Primes up to Four Quintillion,” *International Mathematical Forum*, vol. 10, no. 6, pp. 283–288, 2015, doi:10.12988/imf.2015.5322.
- [13] M. Visser, “Verifying the Firoozbakht, Nicholson, and Farhadian Conjectures up to the 81st Maximal Prime Gap,” *Mathematics*, vol. 7, no. 8, Art. 691, 2019, doi:10.3390/math7080691.
- [14] A. Kourbatov, “Upper Bounds for Prime Gaps Related to Firoozbakht’s Conjecture,” *Journal of Integer Sequences*, vol. 18, Art. 15.11.2, 2015.
- [15] P. Erdős and M. Kac, “The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions,” *American Journal of Mathematics*, vol. 62, no. 1/4, pp. 738–742, 1940, doi:10.2307/2371483.
- [16] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, 3rd ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, 2015.
- [17] A. Vázsonyi, “Which Door Has the Cadillac?,” *Decision Line*, vol. 30, no. 1, pp. 17–19, Dec. 1998–Jan. 1999.
- [18] Y.-R. Liu, “A Generalization of the Erdős–Kac Theorem and its Applications,” *Canadian Mathematical Bulletin*, vol. 47, no. 4, pp. 589–606, 2004, doi:10.4153/CMB-2004-057-4.