

The CCK

CLERESTORY

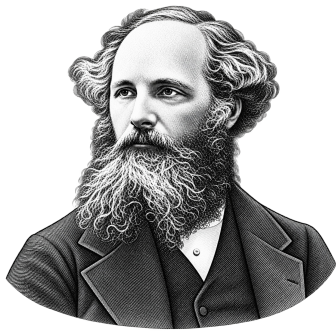


Volume 1, Issue 3, August 2026



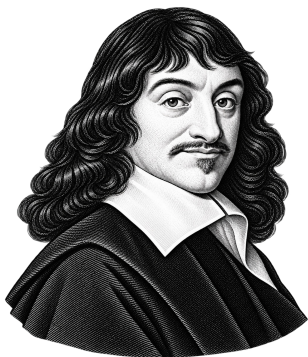
Pierre-Simon Laplace (1749-1827)

“The theory of probabilities is at bottom nothing but common sense reduced to calculus.”



James Clerk Maxwell (1831–1879)

“Therefore the true Logic for this world is the Calculus of Probabilities.”



René Descartes (1596-1650)

“When it is not in our power to determine what is true, we ought to act according to what is most probable.”



Journal Title: The CCK CLERESTORY
Licence Holder: Corridor of Comprehensive Knowledge
Editor-in-Chief: Kamal Shahtalebi
Publisher: Printing and Publishing Unit of the
Corridor of Comprehensive Knowledge
Issue: August 2026
Published Online: September 10, 2026
Frequency: Quarterly
Print ISSN: 2817-2191
Online ISSN: 2817-2205

Acknowledgements

We would like to thank Professor Emeritus Veselin Jungić of Simon Fraser University for his kind words and encouragement regarding *The CCK CLERESTORY*.

We also gratefully acknowledge Dr. Hamid Mala, Associate Professor at University of Isfahan, and Dr. Nooshin Garakyaragh, a Ph.D. graduate of University of Isfahan, for their careful review of selected material in this issue.

Copyright and Publication Notice

© 2026 The CCK CLERESTORY.

Published by the Corridor of Comprehensive Knowledge (CCK).

Except where otherwise noted, original content in this issue may be quoted, reproduced, distributed, or republished, in whole or in part, provided that The CCK CLERESTORY and the original source are clearly credited. Third-party content remains subject to its respective copyright and licensing terms.

Email Address: cck@theccck.com

Phone: +1(778)9999400

Web Address: www.theccck.com

Mailing Address: 1999 Marine Drive, North Vancouver, V7P3J3



Table of Contents

Editor's Note.....	2
The Coin That Challenged Probability.....	3
Teaching Probability with Purpose: Estimation, Strategy, and Surprises	5
The Remarkable Journey of the Monty Hall Problem	11
Bus Waiting Time Problem	3
From a Shared Birthday to a Forged Update (Part I).....	17
Do You Trust the Number?	21
From a Shared Birthday to a Forged Update (Part II).....	23
Probability and Primality.....	27
Of Nods, Eyebrows, and Impossible Probabilities	35
Luck, Numbers, and Live Experiments: Hannah Fry in the 2019 Christmas Lectures	37
The Airplane Seating Problem	43
When One More Toss Matters	45

Editor's Note

Dear Reader,

The central theme of the third issue of *The CCK CLERESTORY* is probability, a subject that brings together a diverse collection of problems, ideas, historical episodes, and applications.

In this issue, 0.5, perhaps the most familiar probability value, commonly associated with obtaining heads in the toss of a fair coin, is itself turned into a challenge for the reader. The everyday experience of waiting for a bus becomes a problem involving random variables and timing. The Monty Hall problem reveals how strongly intuition can conflict with correct probabilistic reasoning, while its remarkable history illustrates the extent to which this deceptively simple problem has engaged human curiosity, from experiments on the feeding behaviour of fish to quantum physics.

We also follow an unexpected path from the birthday problem in probability to the mechanisms behind malware attacks and the breaching of seemingly secure barriers in computer systems. We explore how probability-based algorithms can be used to identify extremely large prime numbers and consider the role these numbers play in modern data transmission. Another contribution presents a distinctive challenge concerning the extent to which we should trust numerical probabilities when making decisions in everyday life.

This issue also revisits the 2019 Royal Institution Christmas Lectures, presented by Professor Hannah Fry, which explored probability and its applications through a series of engaging demonstrations and real-world examples. The issue concludes with intuitive analyses of two classic probability problems, illustrating once again how apparently simple questions can lead to elegant and sometimes unexpected reasoning.

We hope that the material presented in this issue will enrich the teaching and learning of probability and encourage a fresh, practical, and interdisciplinary approach to presenting probability-related concepts in schools.

Sincerely,

Dr. Kamal Shahtalebi
Editor-in-Chief

A handwritten signature in blue ink, consisting of a large, stylized loop followed by a horizontal line and a small flourish.

The Coin That Challenged Probability

On the first day of a probability and statistics class, the teacher walks in holding a coin. After introducing himself, he holds it up for everyone to see.

Teacher: If I toss this coin, what is the probability that it lands heads?

Class: $1/2$.

Teacher: How do you know?

Class: If the coin is fair, heads and tails are equally likely. There are only two possible outcomes, so the probability of heads is $1/2$.

Teacher: But how do you know the coin is fair?

The class pauses for a moment.

Class: We could toss it many times. If we get about the same number of heads and tails, then it is fair.

Teacher: How many times is “many”? Ten? A hundred? Ten thousand?

Class: A lot...

Teacher: Suppose we toss it 1,000 times and get exactly 500 heads. Would you call the coin fair?

Class: Yes.

Teacher: Now suppose we keep going. After 2,000 tosses, we have 1,010 heads. Would you still call the coin fair?

A few students hesitate.

Teacher: Or should we now say that the probability of heads is 1,010 divided by 2,000, or 0.505? And if we continue tossing the coin, should we keep changing the probability whenever the observed proportion changes?

The room grows quieter.

Teacher: If assigning a probability requires tossing the coin an enormous number of times first, then how can we ever assign a probability to a single toss before performing the experiment?

A low murmur spreads through the class.

The teacher looks at the coin again.

Teacher: Now let me ask you something else.

Suppose we toss the coin 2,000 times and heads appears 1,800 times.

Before those tosses, we assigned a probability of $1/2$ to the event “heads.”

Did we make a mathematical mistake?

The class hesitates.



Teacher: What if, instead, we had assigned probability zero to “heads”? Would that have been a mathematical mistake?

Silence.

The teacher smiles and places the coin on the desk.

Teacher: Surprisingly, not necessarily.

As long as the probabilities we assign obey the three axioms of probability, we have not made a mistake *within the mathematical system of probability itself*.

Probabilities cannot be negative. The probability of the entire sample space must be one. And for mutually exclusive events, probabilities must add in the prescribed way.

So mathematics may allow us to assign $1/2$ to heads. It may even allow us, in a suitable probability model, to assign zero.

He pauses.

Teacher: But mathematics is only part of the story.

If the numbers we assign do not reflect the reality we are trying to describe, then our mathematics may be perfectly consistent while our model is poor.

A model that does not match reality will make poor predictions. And a model that makes poor predictions is not very useful, however flawless its mathematics may be.

He picks up the coin again.

Teacher: So perhaps the real question is not simply, “*Is this probability mathematically allowed?*” It is also: “*Is this the right probability for the world we are trying to model?*”

Teaching Probability with Purpose: Estimation, Strategy, and Surprises

Abstract

Probability and statistics play a crucial role in data analysis and decision-making. To foster engagement in high school and early university courses, this article presents three illustrative examples: estimating fish populations using the mark-recapture method, a selection strategy inspired by the secretary problem, and a probability puzzle based on the Monty Hall problem. Each scenario highlights essential probability principles through accessible and thought-provoking contexts.

1) Introduction

Probability and statistics are commonly included in the high school curricula of many countries, typically in the final two years. Moreover, these subjects are core components of undergraduate programs in disciplines such as biology, economics, psychology, engineering, and computer science. This paper introduces three engaging and practical examples of probability applications that can enhance student understanding and interest: (1) estimating populations using the mark-recapture method, (2) optimizing single-choice decisions using the secretary problem, and (3) the counterintuitive outcomes in decision-making illustrated by the Monty Hall problem. The details of these three topics are presented in Sections 2, 3, and 4, respectively.

2) Estimating Fish Populations Using Mark-Recapture

The mark-recapture method is a classical statistical technique used to estimate animal populations [1-2]. For example, in estimating the number of salmon in a lake, a known number are caught, marked, and then released back into the lake. After some time, a sample of salmon is recaptured, and the number of tagged fish in the sample is used to estimate the total population.

Assume 586 salmon are tagged and released. A week later, a sample of 845 salmon is caught, of which 77 are tagged. Under the assumption of a uniform distribution, we apply the ratio:

$$\frac{i}{n} \approx \frac{m}{N} \quad (1)$$

where i , n , m , and N denote the number of tagged fish recaptured in the sample, the total number of fish in the sample, the total number initially tagged, and the total population size, respectively. Thus:

$$\frac{77}{845} \approx \frac{586}{N} \quad (2)$$

and the unknown total population, N , is estimated from (2)

$$N \approx \frac{mn}{i} = \frac{586 \times 845}{77} \approx 6400 \quad (3)$$

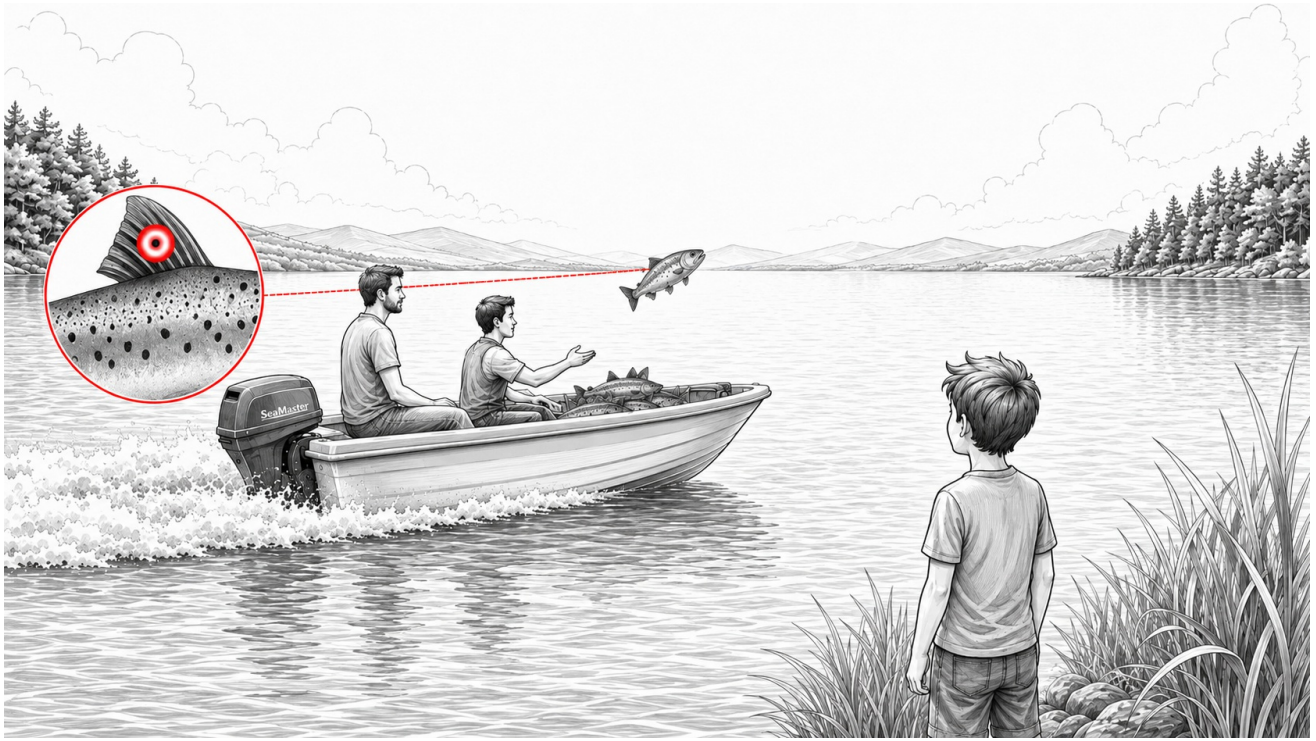


Figure 1) Marked and recaptured: A known number of salmon are caught, marked, and then released back into the lake.

Hereafter, we provide a logical justification for using Equation (1). To estimate the size N of a closed animal population, ecologists catch and mark m animals, then release them. After sufficient time for mixing, a second sample of size n is taken. Let X be the number of marked animals recaptured. Assuming a fixed population and equal capture probability, X follows a hypergeometric distribution. Hence, provided that there are N animals in total, we have:

$$P\{X = i | N\} = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \quad (4)$$

Given the observed value $X = i$, the *maximum likelihood estimate* of N is the value of N that maximizes $P(X = i | N)$. To find it, first note that, after simple algebra

$$\frac{P\{X=i|N\}}{P\{X=i|N-1\}} = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \times \frac{\binom{N-1}{n-i}}{\binom{m}{i} \binom{N-1-m}{n-i}} = \frac{(N-m)(N-n)}{N(N-m-n+i)} \quad (5)$$

Now, the preceding ratio is greater than 1 if and only if

$$(N - m)(N - n) \geq N(N - m - n + i) \quad (6)$$

or, equivalently, if and only if

$$N \leq \frac{mn}{i} \quad (7)$$

Thus, as a function of N , $P\{X = i | N\}$ increases then decreases, peaking at the greatest integer less than or equal to $\frac{mn}{i}$, which serves as the maximum likelihood estimate of N .

3) Optimal Selection Strategy: The Secretary Problem

The secretary problem addresses the challenge of selecting the best option in a sequential process when only one choice is allowed [3-4]. An equivalent and engaging analogy involves a dolphin, Dolphi, observing 100 fish swimming by, one at a time. She can only choose one fish and must decide at the moment she sees it.

To maximize her chance of choosing the largest fish, Dolphi skips the first 37 (about $100/2.71$) and then selects the first fish larger than all previous ones. This strategy provides a success probability of approximately 36.8%.

Mathematically, this optimal stopping rule stems from maximizing the probability of selecting the maximum in a random sequence under known size n . The result is derived using integral calculus and limit analysis. The details of the proof are explained below.

The classic Secretary Problem (or Optimal Stopping Problem) is defined as follows:

- There are n candidates (or options) presented in random order.
- You can observe them one by one, but must make an immediate and final decision upon each.
- You are allowed to select only one candidate.
- The goal is to maximize the probability of choosing the best (i.e., the highest-ranked) candidate.

Suppose the strategy in selecting a candidate from among n candidates (based on the conditions stated in the problem) is as follows:

- 1) Skip the first r candidates.
- 2) Then, select the next candidate who is better than all previous ones.

Based on the above strategy, we determine the optimal value of r that maximizes the probability of selecting the best candidate.

Let the best candidate appear at position k . Following the above steps, you can only select the best candidate if $k > r$ and the best among the first $(k - 1)$ candidates is within the first r observed.

The probability that the best candidate is at position k is:

$$p_k = \frac{1}{n} \quad (8)$$

The probability that the maximum of the first $(k - 1)$ candidates is within the first r is:

$$p_r^{(k-1)} = \frac{r}{k-1} \quad (9)$$

Hence, the total probability of success is given by

$$\sum_{k=r+1}^n p_k p_r^{(k-1)} = \frac{1}{n} \sum_{k=r+1}^n \frac{r}{k-1} \quad (10)$$

Determining the value of r that maximizes Equation (2) is somewhat challenging, but for



Figure 2) Dolphi Makes a Choice.

sufficiently large n , with $x = \frac{r}{n}$, it can be expressed as an integral.

$$\lim_{n \rightarrow +\infty} x \sum_{k=nx+1}^n \frac{1}{k-1} = x \int_x^1 \frac{dt}{t} = -x \ln x \quad (11)$$

The derivative of $-x \ln x$ is equal to $-1 - \ln x$, which becomes zero at $x = \frac{1}{e}$, where e is Euler's number ($e \approx 2.718$) and this value of x maximizes it. Therefore, for sufficiently large n , $x = \frac{r}{n} = \frac{1}{e}$ or $r = \frac{n}{e}$ maximizes the probability of selecting the best candidate, and the maximum probability is approximately $-e^{-1} \ln e^{-1} \approx 0.368$.

This strategy maximizes the probability of selecting the best option.

For $n = 100$, the value of $r = \frac{n}{e}$ will be approximately equal to 37.

4) The Logic of Cups: The Monty Hall Problem

The Monty Hall problem exemplifies a paradox of conditional probability [5]. In its classic form, three doors hide one car and two goats; here, the same logic is represented using cups and a coin. In a practical version of this problem, a player is shown three cups, one of which hides a coin. After selecting one cup, the host reveals an empty cup from the remaining two. The player is offered a choice: **stick or switch**. Although intuitive reasoning suggests a 50-50 chance, mathematical analysis reveals that switching results in a $\frac{2}{3}$ success rate. Initially, the chance that the chosen cup hides the coin is $\frac{1}{3}$, while the combined chance for the other two is $\frac{2}{3}$. Revealing an empty cup does not affect this distribution, so switching targets the higher probability.

An intuitive way to see this is to imagine a second player who always takes the complement of the first player's choice. If the first player selects Cup 1, the second effectively selects Cups 2 and 3 together and wins whenever the coin is under either cup, which occurs with probability $\frac{2}{3}$. The host's reveal does not alter this probability; it merely identifies the empty cup within the second player's two-cup choice. Thus, the offer to switch effectively asks the first player whether to exchange a $\frac{1}{3}$ chance for the second player's $\frac{2}{3}$ chance. Repeated trials support the superiority of the switching strategy, providing a compelling teaching opportunity to challenge students' assumptions and enhance probabilistic reasoning.



Figure 3) Monty Hall Problem: Stick with the choice or switch?

5) Conclusion

These three scenarios offer intuitive yet rigorous entry points into essential topics in probability. By grounding abstract concepts in engaging contexts, educators can stimulate curiosity and deepen understanding. We encourage teachers to incorporate such examples, supported by classroom discussion and simple simulations, to enrich the teaching of probability.

Table 1) Sample Alignment with the British Columbia Mathematics Curriculum

Article topic	Example in the article	Most relevant B.C. course(s)	B.C. curriculum connection
Sampling, population estimation, and statistical inference	Mark–recapture estimation of a fish population	Statistics 12; supporting connection to Foundations of Mathematics 11	Census versus sample; target population; observational studies; estimators of population parameters; variation and inference. Foundations 11 also includes applications of statistics, data interpretation, confidence intervals, and distributions.
Probabilistic decision-making and strategy	Secretary Problem / optimal selection strategy	Foundations of Mathematics 12	Probabilistic thinking for decisions under uncertainty; strategic reasoning; modelling in situational contexts; using reason to determine winning strategies in puzzles and games.
Calculus-based optimization	Derivation of the optimal stopping point near n/e	Calculus 12	Limits, differentiation, integration, and optimization in contextual situations. This supports the article’s proof, but it is an extension beyond the probability content of Foundations 12.
Conditional probability	Monty Hall Problem: stick or switch	Foundations of Mathematics 12	Odds and probability, including mutually exclusive and non-mutually exclusive events and conditional probability; probabilistic thinking informs decisions involving chance and uncertainty.
Simulation and experimental reasoning	Repeated Monty Hall trials; optional simulations of the other examples	Statistics 12	Use of simulation to build intuition about sampling distributions and inferential concepts; use of software and technology to explore statistical ideas.
Reasoning, modelling, and communication	Comparing intuition with mathematical analysis across all three scenarios	Foundations of Mathematics 11–12; Statistics 12	Reasoning, modelling, problem solving, explaining and justifying decisions, and communicating findings in context.

Source: British Columbia Curriculum, Mathematics 11–12 (Foundations of Mathematics, Statistics, and Calculus).

References

- [1] S. M. Ross, “A First Course in Probability”, Prentice Hall, 8th ed., 2010.
- [2] S. M. Ross, “Introduction to Probability Models” Academic Press, 11th ed., 2014.
- [3] F. Mosteller, “Fifty Challenging Problems in Probability with Solutions”, Dover Publications, 1965.
- [4] T. S. Ferguson, “Who Solved the Secretary Problem?”, Statistical Science, vol. 4, Issue 3, 1989.
- [5] W. Feller, “An Introduction to Probability Theory and its Applications”, vol. 1, Wiley, 1968.

The Remarkable Journey of the Monty Hall Problem

Few probability puzzles have travelled so far. From nineteenth-century paradoxes and television game shows to public controversy, great mathematicians, psychology, animal cognition, game theory, and even quantum physics, the Monty Hall problem has been explored from remarkably diverse perspectives.



1889 — Bertrand's Box Paradox

Joseph Bertrand introduced a three-box puzzle in which observing a gold coin changes the probabilities of the boxes that could have produced it. Its counterintuitive solution anticipates the same conditional-probability structure that later made the Monty Hall problem famous.



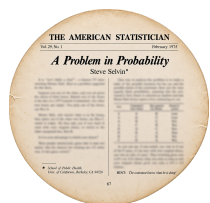
1959 — The Three Prisoners Problem

Martin Gardner presented the Three Prisoners Problem in *Scientific American*. One of three prisoners will be pardoned. Prisoner A asks the guard to name one of the other two who will not be pardoned. Even with only two possibilities left, A's chance remains $1/3$, while the unnamed prisoner's chance becomes $2/3$, just like in the Monty Hall problem.



1963 — Let's Make a Deal

The television game show *Let's Make a Deal*, hosted by Monty Hall, premiered in the United States. The later probability puzzle borrowed the show's setting and Hall's name, although the real program did not consistently follow the exact rules of the mathematical puzzle.



1975 — Steve Selvin

Biostatistician Steve Selvin introduced the puzzle inspired by the television game show *Let's Make a Deal* to the statistics community in *The American Statistician*. In a follow-up later that year, he called it the "Monty Hall problem" and explained why switching doors gives a chance of winning under the standard rules.



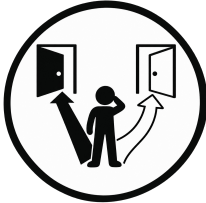
1990 — Marilyn vos Savant

American columnist Marilyn vos Savant, once listed by Guinness for the highest recorded IQ, explained in her *Parade* column that switching gives a chance of winning. Her answer sparked a famous controversy: about 10,000 readers wrote in, including nearly 1,000 PhDs, many insisting the two remaining doors should be equally likely.



Early 1990s — Paul Erdős

The puzzle was presented to Paul Erdős, one of the most prolific and influential mathematicians of the twentieth century, renowned for major contributions to number theory, combinatorics, and probability. He was initially unconvinced by the counterintuitive result, making the episode a striking example of how strongly the Monty Hall problem can conflict with intuition.



1995 — A Tool for Psychology

Researchers began using Monty Hall experimentally to study human decision-making, including status-quo bias, illusion of control, regret, and resistance to switching.



2002 — Quantum Monty Hall

Adrian P. Flitney and Derek Abbott, researchers at the University of Adelaide, published a quantum version of the Monty Hall game in *Physical Review A*. They used the familiar puzzle as a simple model to explore how quantum ideas such as superposition and entanglement can change the information available to players and, in turn, alter the strategies and advantages possible in the game.



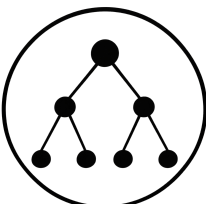
2003 — Why Do We Get It Wrong?

Psychologists Stefan Krauss and Xiao-Tian Wang showed that the Monty Hall problem becomes much easier to solve when probabilities are presented as simple counts of outcomes rather than abstract percentages, and when the problem is viewed from a different perspective. Their study suggested that much of the difficulty lies not in the mathematics itself, but in how the information is represented in our minds.



2010 — Pigeons Learn to Switch

Psychologists Walter T. Herbranson and Julia Schroeder of Whitman College, working in comparative psychology, animal learning, and decision-making, tested pigeons on a Monty Hall-type task. After a losing option was removed, the birds repeatedly chose whether to stay or switch. With extensive practice and immediate reward feedback, switching increased from about 36% to 96%, showing that the pigeons gradually learned the probability-maximizing strategy without explicit mathematical reasoning.



2011 — Game Theory & Modelling

Richard D. Gill, a British-Dutch mathematical statistician and professor at Leiden University, argued that the Monty Hall problem is not just about probability, but also about modelling and game theory. The familiar switching result depends on specific assumptions about the host: that he knows where the prize is, always opens a losing door, and follows a defined rule when he has a choice. Change those rules, and the probabilities can change as well.



2026 — Glass Catfish

Researchers tested glass catfish on a Monty Hall-type dilemma. Unlike the extensively trained pigeons, the fish tended to remain with their initial choice, extending the puzzle into comparative cognition.

Bus Waiting Time Problem



Figure 1) An illustrative bus schedule showing variability in arrival times.

Vancouver's bus network, operated by TransLink, seeks to provide passengers with predictability by scheduling buses at regular intervals. However, traffic and operational conditions introduce tolerances. As a result, passengers' waiting times may deviate from the official schedule, either decreasing or increasing. This creates opportunities for interesting probability examples. Consider the following hypothetical scenario: From 6:01 to 6:11 a.m., Bus 240 Lynn Valley is scheduled to arrive at Capilano Mall, North Vancouver Station, at both 6:01 and 6:11, corresponding to a 10-minute headway. For each of the following cases, determine the probability that Mary, the only passenger, will wait more than 5 minutes for the bus:

- Mary arrives uniformly between 6:01 and 6:11. Bus 240 arrives exactly at 6:01 and 6:11.
- Mary arrives uniformly between 6:01 and 6:16. Bus 240 continues to arrive exactly every 10 minutes starting from 6:01.
- Mary arrives uniformly between 6:01 and 6:11. Bus 240 is scheduled every 10 minutes starting at 6:01, and each bus may arrive up to 1 minute early or late, with its arrival time uniformly distributed over that interval. Assume that Mary's arrival time and the buses' arrival times are independent. If the 6:11 bus arrives early, it remains at the station until 6:11. For this problem, Mary's waiting time is defined as the time from her arrival at the station until her bus departs, including any time she spends seated on the bus before departure.

Solutions:

a) If Mary arrives at the station between 6:01 and 6:06, she will wait more than 5 minutes. The probability of this event, assuming a uniform distribution for Mary's arrival time, is $5/10$.

b) If Mary arrives at the station between 6:01 and 6:06 or between 6:11 and 6:16, she will wait more than 5 minutes. The probability of this event, assuming a uniform distribution for Mary's arrival time, is $\frac{5}{15} + \frac{5}{15} = \frac{2}{3}$.

c) An Intuitive and Formal Solution

Intuitive argument

Because Mary's arrival time is uniformly distributed over the interval from 6:01 to 6:11 a.m., the result can be understood quite intuitively. A simple logical argument shows that the answer to part C is exactly the same as in part A, namely 0.5. A delay in the 6:01 bus works in Mary's favor because it gives her more time to catch that bus. Conversely, a delay in the 6:11 bus works against her by increasing the time she may have to wait. Since the two bus-arrival deviations follow the same distribution, these opposing effects balance on average. An early arrival of the 6:01 bus has no effect because it occurs before Mary's possible arrival interval. An early arrival of the second bus does not change the outcome either: under the stated condition, the bus remains at the station until 6:11, so from Mary's perspective, time spent sitting on the bus before 6:11 is effectively equivalent to waiting until the scheduled departure time. On average, therefore, the situation is equivalent to the case in which both buses arrive and depart exactly on schedule.

Formal derivation

The following derivation goes beyond the level typically expected in high school mathematics.

The same result can also be obtained formally. Taking 6:01 a.m. as time zero, let T_1 denote the arrival time of the first bus, measured in minutes relative to 6:01; let T_2 denote the arrival time of the second bus; and let X denote Mary's arrival time at the station. Because the three arrival times are uniformly distributed over their respective intervals, their probability density functions are:

$$\begin{aligned} f_{T_1}(t) &= \begin{cases} 0.5 & \text{if } -1 < t < 1 \\ 0 & \text{otherwise} \end{cases} \\ f_{T_2}(\tau) &= \begin{cases} 0.5 & \text{if } 9 < \tau < 11 \\ 0 & \text{otherwise} \end{cases} \\ f_X(x) &= \begin{cases} 0.1 & \text{if } 0 < x < 10 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Taking into account that a bus arriving early remains at the station until 6:11, such an early arrival does not affect Mary's waiting time. Assuming the three random variables are independent, the event that Mary waits more than 5 minutes can be divided into four mutually exclusive cases (See Figure 2.).

The desired probability is the sum of the probabilities of these four cases:

$$\begin{aligned} P(T_1 < 0, T_2 < 10, X < 5) &= \int_{-1}^0 \int_9^{10} \int_0^5 f_{T_1}(t) f_{T_2}(\tau) f_X(x) dt d\tau dx \\ &= 0.025 \int_{-1}^0 \int_9^{10} \int_0^5 dx d\tau dt = 0.125 \end{aligned}$$

$$\begin{aligned} P(T_1 < 0, T_2 > 10, X < T_2 - 5) &= \int_{-1}^0 \int_{10}^{11} \int_0^{\tau-5} f_{T_1}(t) f_{T_2}(\tau) f_X(x) dt d\tau dx \\ &= 0.025 \int_{-1}^0 \int_{10}^{11} (\tau - 5) d\tau dt = 0.1375 \end{aligned}$$

$$\begin{aligned} P(T_1 > 0, T_2 < 10, X > T_1, X < 5) &= \int_0^1 \int_9^{10} \int_t^5 f_{T_1}(t) f_{T_2}(\tau) f_X(x) dt d\tau dx \\ &= 0.025 \int_0^1 \int_9^{10} (5 - t) d\tau dt = 0.1125 \end{aligned}$$

$$\begin{aligned} P(T_1 > 0, T_2 > 10, X > T_1, X < T_2 - 5) &= \int_0^1 \int_{10}^{11} \int_t^{\tau-5} f_{T_1}(t) f_{T_2}(\tau) f_X(x) dt d\tau dx \\ &= 0.025 \int_0^1 \int_{10}^{11} (\tau - 5 - t) d\tau dt \\ &= 0.025 \int_0^1 (5.5 - t) dt = 0.125 \end{aligned}$$

By summing the four calculated probabilities, we obtain:

$$0.125 + 0.1375 + 0.1125 + 0.125 = 0.5$$

Thus, the formal calculation agrees with the intuitive argument: the probability that Mary waits more than 5 minutes is 0.5, exactly as in part A.

You can answer the following challenge case.

If Mary arrives at the station between 6:00 and 6:15 (uniformly distributed), the buses arrive uniformly between 1 minute earlier and 1 minute later than scheduled, and each bus always stops for 1 minute upon arrival, what is the probability that Mary waits more than 5 minutes for the bus?

If Mary herself were to solve this problem, she probably would not be happy about the buses arriving earlier or later than scheduled!

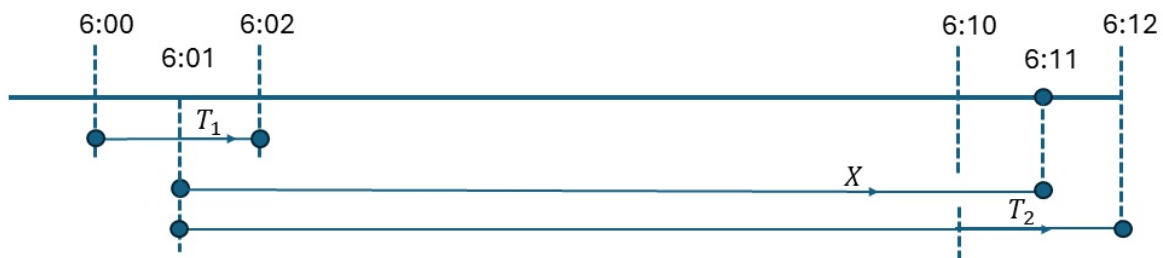


Figure 2) Timing diagram illustrating the arrival times in case c.

Bus vs. Cars

- 1 bus can replace about 40 cars
- Fewer cars → less congestion
- Lower emissions per passenger
- Less road space used



$$= \text{car} \times 40$$

When more people take the bus, we all benefit. Buses reduce the number of cars on the road, which means less congestion and fewer accidents for everyone.

→ Less pollution for cleaner air and a healthier planet.

Public transit also uses less energy per passenger compared to driving alone.

Small choices, big impact.

Question: If p is the probability of an incident during one car trip, what is the probability of at least one incident in 40 separate car trips?

Hint: $P(\text{at least one incident}) = 1 - (1 - p)^{40}$

Choosing public transit is a choice that comes with future benefits.

From a Shared Birthday to a Forged Update (Part I)

Note: In the article that follows, “John,” his office, and the everyday details at the beginning of the story are part of an educational reconstruction and are not attributed to any specific historical victim. However, the core technical chain of the narrative, including Flame's use of an MD5 collision to create a forged certificate, the ability to sign code in a way that chained to Microsoft's trust infrastructure, and propagation within an infected network by disguising a malicious package as a Windows Update, is based on technical reports from Microsoft and Kaspersky [1], [2].

1) A Simple Update

It is a day in the spring of 2012. John is sitting at his desk in the office of an international engineering company in the Middle East. His laptop is running the 32-bit version of Windows 7. Only a few minutes after he starts work, Windows, as on previous occasions, connects to the network to receive updates. Nothing unusual appears on the screen:

Checking for updates → Downloading → Installing → Successfully completed

No security warning appears. No window displays the message "Unknown Publisher.". No program asks John to open a strange file. The update finishes, and he continues working. From John's perspective, everything was normal. And that is precisely the disturbing part of the story.

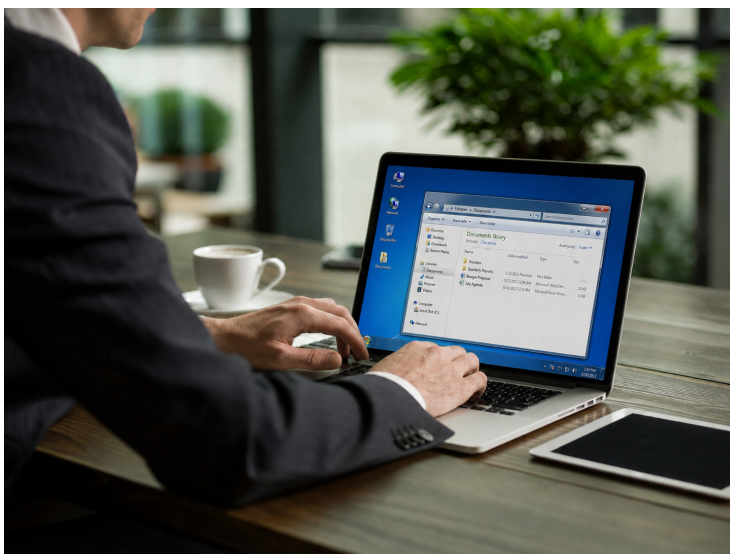


Figure 1) John's laptop was running the 32-bit version of Windows 7, a version that was later found to account for a large proportion of systems infected by the malware known as Flame.

Later, when it became clear that highly important information from John's computer had been stolen, it also became clear that nothing had in fact been normal and that the downloaded file had not come from Microsoft. Yet one question remained: if the received file did not truly come from Microsoft, why did Windows' trust mechanism not stop it?

To answer that question, we must temporarily set John's laptop aside and turn to a problem that, at first glance, seems to have nothing to do with malware, digital certificates, or Windows Update.

2) A Shared Birthday

Suppose that at a party with k invited guests, we are looking for people who share the same birthday. We take the year to have $N = 365$ days and ignore leap years, differences in birth rates across months, and twins. We also assume that each person's birthday is independent of the others and is distributed approximately uniformly over the $N = 365$ days. For $k = 23$ people, the number of possible pairs is:

$$\binom{23}{2} = \frac{23 \times 22}{2} = 253$$

This means that if only 23 people are present at the party, there are 253 different pairs whose birthdays can be compared. As k increases, the number of pairs grows approximately with the square of k , and the probability of finding no pair with the same birthday falls sharply as k increases. To estimate the probability of a collision, meaning that at least two people have the same birthday, we calculate the complementary probability, namely the probability of “no collision.” There is no restriction on the first person's birthday; it can be any day. For the second person to have a different birthday, it must fall on one of the $N - 1$ remaining days. For the third person, $N - 2$ days remain available, and the process continues. Therefore, the probability that no repetition occurs among k selections is:

$$P_{\text{no collision}} = \frac{N}{N} \frac{N-1}{N} \frac{N-2}{N} \dots \frac{N-k+1}{N} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{N}\right)$$

To obtain a simpler expression, we take the natural logarithm of both sides:

$$\ln P_{\text{no collision}} = \ln \prod_{i=0}^{k-1} \left(1 - \frac{i}{N}\right) = \sum_{i=0}^{k-1} \ln \left(1 - \frac{i}{N}\right) \approx -\sum_{i=0}^{k-1} \frac{i}{N} = -\frac{1}{N} \sum_{i=0}^{k-1} i = \frac{-k(k-1)}{2N}$$

The approximation above is based on the approximate equality $\ln(1 - x) \approx -x$ for small

x , with $x = \frac{i}{N}$. Thus: $P_{\text{no collision}} \approx e^{\frac{-k(k-1)}{2N}}$

And consequently:

$$P_{\text{collision}} = 1 - P_{\text{no collision}} \approx 1 - e^{\frac{-k(k-1)}{2N}}$$

If k is not very small, we may also approximate $k - 1$ by k , and therefore:

$$P_{\text{collision}} \approx 1 - e^{-0.5\left(\frac{k}{\sqrt{N}}\right)^2}$$

Now consider a situation in which we have two different groups, Groups A and B, each containing k people. This time, we are interested only in matches in which one person from the first group and one person from the second group were born on the same day. Each of the k people in the first group can be compared with each of the k people in the second group; therefore, the number of possible pairs between the two groups is:

$$k \times k = k^2$$

For each of these pairs, the probability of having the same birthday is $1/N$. For a simple estimate, if we treat these k^2 opportunities for a match as approximately independent, the probability that no cross-group pair shares a birthday is approximately:

$$P_0 \approx (1 - 1/N)^{k^2}$$

Therefore, the probability that at least one person from the first group and one person from the second group share a birthday is given approximately by [3], [4]:

$$P_c \approx 1 - (1 - 1/N)^{k^2}$$

Using the same exponential approximation as in the previous calculation, this relation can be written more simply as:

$$P_c \approx 1 - e^{-\left(\frac{k}{\sqrt{N}}\right)^2}$$

This result shows that in the two-group model, when k is on the order of the square root of N , the collision probability becomes substantial. The important point is that no specific birthday has been targeted in advance; any match between a member of the first group and a member of the second group counts as a solution. This two-group form of the birthday problem is important for understanding the method used in the attack discussed below: there too, two groups of different possibilities are examined, and the goal is to find one pair with the same output. The relation above is approximate because the different pairs are not completely independent; reference [3] examines this approximation and its error for collisions between two sets.

According to the relation above, and as Figure 2 also shows, as k becomes larger relative to N , the cross-group collision probability increases rapidly and approaches 1. For example, if $N = 365$, then for $k = 16$, $k = 24$, and $k = 32$, the value of P_c is approximately 50.4%, 79.4%, and 94.0%, respectively. Thus, even with 16 people in each group, the probability that at least one person from the first group and one person from the second group share a birthday is greater than 50%.

19 The CCK CLERESTORY

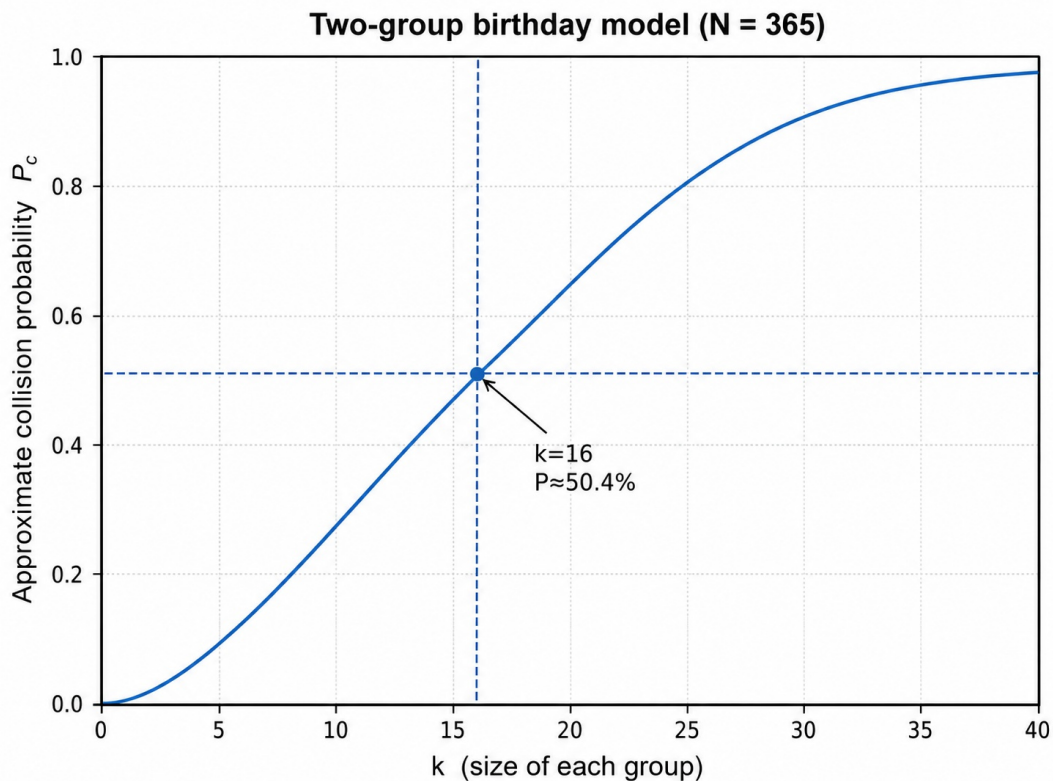


Figure 2) Approximate probability of at least one shared birthday between two groups of k people, as a function of k , for $N = 365$.

From this point on, replace each person's birthday with an identification number, or identifier. In this hypothetical model, the identifier is a number obtained from the person's characteristics using a specified function and is used to identify that person. The identifier is determined not only from the birthday but from various characteristics of the person, using a specified function. The discussion above shows that with a small range of identifiers, here 365 possible forms, we do not need to encounter more than 365 people before two people can have the same identifier; with far fewer people, on the order of the square root of this number, the possibility that two different people have the same identifier becomes substantial. The same idea also held when we considered two distinct groups of people. To reduce the possibility of a forged identifier, we can increase N , the number of possible identifier forms. Even so, by searching among a number of people on the order of $\sqrt{N}$ for a matching pair, two different people with the same identifier can be found.

Part II of this article continues on page 23.

Do You Trust the Number?

A company has spent years developing an **AI-powered emergency braking system for heavy trucks**.

Cameras, radar, predictive algorithms, redundant processors, and independent backup systems work together. The system has undergone extensive simulations, controlled track tests, failure analysis, and rigorous scientific validation.

The engineers finally sign off on an extraordinary estimate:
1/1,000,000,000

One in a billion.

The project director approves the report.

Then he receives an unusual offer.

A fully loaded tractor-trailer will accelerate directly toward him. He will stand in the centre of the test track. The AI braking system must bring the truck to a complete stop **exactly half a metre in front of him**.

If he agrees to stand there, he receives **\$1,000,000**.

He knows the system.

He knows the tests.

He knows the mathematics.

Would he do it?

More importantly:

Would you?

Still hesitating?

Tomorrow, you may get into your car and drive 20 kilometres without giving the risk much thought.

In Canada, the 2023 road-fatality rate was **4.5 deaths per billion vehicle-kilometres travelled**. As a rough population-level comparison, 20 kilometres of vehicle travel corresponds to a fatality exposure on the order of **one in 11 million**, about **90 times larger** than the one-in-a-billion failure probability in our fictional truck test. Canada is only one reference point. Road-fatality risk is similar in many developed countries and substantially higher in some parts of the world.





Yet tomorrow you will probably start your car again.
No million-dollar reward.
No team of engineers standing beside you.
No report promising that the chance of catastrophe is only one in a billion.
You simply drive.
So why are you still unwilling to stand in front of the truck?
Is it because you do not trust the **one-in-a-billion estimate**?
Because choosing a risk feels different from simply living with one?
Because driving gives you a sense of control?
Or is there something about **forty tonnes of steel coming directly toward you** that overwhelms what the mathematics is telling you?
Perhaps the real question is not whether we understand probability.
It is whether we trust a number when everything we see and feel is telling us to run.

From a Shared Birthday to a Forged Update (Part II)

To connect this discussion with the installation of a forged and malicious program on John's laptop, suppose the identifiers of all members of Group A have been approved by a trusted authority.

If one of them has a package bearing that person's name and unique identifying mark, the guard can verify that the package belongs to that person. Because the identifier is trusted, the guard treats the package as coming from a trusted source and allows it to enter.

The members of Group B are not trusted. However, if a reasonable number of them are compared with the same number from Group A, one Group B identifier may match one from Group A, since both are produced by the same function and mechanism. In that case, the authority's approval for the Group A identifier also appears valid for the matching Group B identifier. The guard therefore mistakenly treats that person as trusted. The Group B person can then mark a malicious package with the name and unique identifying mark associated with that identity. The guard correctly determines that the package belongs to the holder of the identifier, but because the identifier itself has been mistakenly accepted as trustworthy, the package is also allowed to enter.

3) The Connection Between the Birthday Problem and Forged Programs

When deciding whether to trust certain programs, Windows also checks their digital certificates, which can be viewed as “electronic credentials” [5]. A trusted certificate, together with a valid link to the program, allowed Windows to identify the program as code from a trusted source. In the analogy above, the certificate corresponds to a person, the program to that person's package, Windows to the guard, and the laptop to the building. In one of Microsoft's certificate-issuing systems in 2012, MD5 was used to produce a 128-bit code from a certificate's information. This number was not itself “valid” or “invalid”; during certificate issuance, a trusted authority confirmed it through a digital-signature mechanism.

We would normally expect different certificates to produce different codes. However, MD5 weaknesses and the collision principle explained earlier made it possible to create two different certificates with the same MD5 code [1], [6].

The attackers exploited this weakness. They considered at the same time one ordinary certificate structure that could be issued through a legitimate Microsoft path and one forged structure with the characteristics they wanted. Then, using weaknesses in MD5, they examined and modified different versions of both until they obtained two different certificates with the same MD5 output. (In cryptography, this type of collision is called a chosen-prefix collision [7].) The problem, therefore, was not to find another certificate for a valid code that had already been fixed in advance; both sides could be varied from the beginning, and the goal was to reach a common output. The ordinary certificate was issued and approved through the genuine Microsoft process. The attackers did not need to know the final result of Microsoft's approval in advance; they only needed to predict the content of the certificate that would be issued with sufficient accuracy. Because the ordinary and forged certificates had the same MD5 value, the same approval that Microsoft issued for the ordinary certificate also appeared valid for the forged certificate. Windows therefore mistakenly regarded the forged certificate as trusted [1]. The attackers then associated Flame with this forged certificate. Information accompanying the program allowed Windows to check that Flame really was connected to the holder of that certificate. This part of the check was correct; the mistake had occurred earlier, because Windows had already accepted the forged certificate itself as trustworthy. As a result, Flame could also appear to be a program associated with a trusted source.

The attack went further. The attackers did not need to install Flame manually on computers one by one. If Flame was already installed on one computer in a network, that computer could help spread the malware to some other computers on the same network. For example, when another computer on that network sent a Windows Update request, the infected system could place itself in the communication path and send a seemingly legitimate malicious package instead of Microsoft's genuine response. In this way, Flame could spread within an infected network from one system to others [2].

The connection between this story and the birthday problem lies in the stage of finding two different certificates with the same output. As we saw, the rapid growth in the number of possible pairs means that a collision can become possible much sooner than one might expect by looking only at the total number of possible states. This probabilistic principle is the basis of the Birthday Attack, whose name should now be clear from the previous section.

The real Flame attack, however, was more complex than a simple birthday search. In addition to this principle, it relied on specific weaknesses in MD5, more advanced techniques, and many additional details to create the collision, meaning the same code, and make the attack succeed [1], [7].

The harmful effects of the attack on individuals and organizations were extensive. For more information about MD5 and its role in this attack, see [1], [6], and [7]. More information about the measures Microsoft took after discovering the attack, including revoking the abused certificates and strengthening the Windows Update mechanism, is provided in [5]. Kaspersky's analysis reports Flame's espionage capabilities, including information collection, screenshot capture, keystroke logging, and audio recording [8]. This last point makes it possible to imagine how John's presence with his laptop in a highly important meeting could have allowed information from the meeting to leak outside. It also helps explain why companies and organizations can be highly cautious about the use of electronic communication devices in important business meetings.

Table 1) Alignment of the Topic with the Curriculum, Using British Columbia (BC) as an Example

Alignment with this article	Relevant curriculum components	Course / Grade
The birthday problem, counting pairs, complementary probability, approximation of collision probability, and interpretation of the role of probability in a real-world problem.	Probabilistic thinking; combinatorics; chance and probability; analysis and application of mathematical ideas through reasoning and technology [9]	Foundations of Mathematics 12
Simulation of the birthday problem, graphing collision probability, the idea of search, and converting a real security problem into a computational model.	Algorithms and computational thinking; random-number generation; modeling mathematical problems; estimating theoretical probability through simulation and linking theoretical and experimental probability [10]	Computer Science 11
Examining the assumptions of the birthday model, comparing theoretical probability with simulation, and interpreting the limitations of a probabilistic model.	Statistical thinking; probabilistic models; understanding variability; use of technology and simulation to strengthen statistical intuition [11]	Statistics 12
Comparing the size of the search space, understanding the difference between the total number of possible states and the order of effort required to find a collision, and linking computational efficiency to security.	Analysis of algorithmic efficiency and Big-O; working with large data sets; modeling mathematical problems and estimating theoretical probability through simulation [12]	Computer Science 12

References

- [1] Microsoft Security Research & Defense, “Flame malware collision attack explained,” Microsoft Security Response Center, Jun. 6, 2012. <https://www.microsoft.com/en-us/msrc/blog/2012/06/flame-malware-collision-attack-explained>.
- [2] A. Gostev, “‘Gadget’ in the middle: Flame malware spreading vector identified,” Kaspersky Securelist, Jun. 4, 2012. <https://securelist.com/gadget-in-the-middle-flame-malware-spreading-vector-identified/33081/>.
- [3] M. C. Wendl, “Collision probability between sets of random variables,” *Statistics & Probability Letters*, vol. 64, no. 3, pp. 249-254, Sep. 2003, doi: 10.1016/S0167-7152(03)00168-8.
- [4] K. Nishimura and M. Sibuya, “Probability to meet in the middle,” *Journal of Cryptology*, vol. 2, no. 1, pp. 13-22, 1990, doi: 10.1007/BF02252867.
- [5] Microsoft Security Response Center, “Microsoft Security Advisory 2718704: Unauthorized Digital Certificates Could Allow Spoofing,” Jun. 3, 2012, updated Jun. 13, 2012. <https://learn.microsoft.com/en-us/security-updates/securityadvisories/2012/2718704>.
- [6] R. L. Rivest, “The MD5 Message-Digest Algorithm,” RFC 1321, Apr. 1992. <https://www.rfc-editor.org/rfc/rfc1321>.
- [7] M. Stevens, A. Sotirov, J. Appelbaum, A. Lenstra, D. Molnar, D. A. Osvik, and B. de Weger, “Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate,” in *Advances in Cryptology - CRYPTO 2009*, LNCS 5677, Berlin, Germany: Springer, 2009, pp. 55-69, doi: 10.1007/978-3-642-03356-8_4.
- [8] A. Gostev, “The Flame: Questions and Answers,” Kaspersky Securelist, May 28, 2012. <https://securelist.com/the-flame-questions-and-answers/34344/>.
- [9] Government of British Columbia, “Foundations of Mathematics 12,” B.C. Curriculum. <https://curriculum.gov.bc.ca/curriculum/mathematics/12/foundations-of-mathematics>.
- [10] Government of British Columbia, “Computer Science 11,” B.C. Curriculum. <https://curriculum.gov.bc.ca/curriculum/mathematics/11/computer-science>.
- [11] Government of British Columbia, “Statistics 12,” B.C. Curriculum. <https://curriculum.gov.bc.ca/curriculum/mathematics/12/statistics>.
- [12] Government of British Columbia, “Computer Science 12,” B.C. Curriculum. <https://curriculum.gov.bc.ca/curriculum/mathematics/12/computer-science>.

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 Kmhkmh, 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.

Of Nods, Eyebrows, and Impossible Probabilities

I had taught probability and statistics a few times before, but this semester felt different. One student always sat in the middle of the back row, watching me with an unusually intense look. He seemed a little older than the others, maybe even older than me.

Whenever I finished explaining a concept, he would slowly nod in approval. Somehow, that little nod felt like a standing ovation. But if I made a mistake and corrected myself, all it took was one raised eyebrow from him to make me wish I could disappear. The way he watched the class, I was convinced he understood probability and statistics better than anyone else in the room.

Then came the midterm.

One by one, students came to my office to go over their grades. Eventually, he walked in. The nodder. The eyebrow-raiser.

I did not even need to ask his name. I simply started looking for the paper with the highest score.

Then he told me his name.

I found his exam.

Zero.

There was a huge X at the top of the page, perhaps because writing the actual score would have been too painful.

Trying to keep a straight face, I asked if he had any questions. He pointed to one problem and wanted to know why he had received no credit.

I looked at his solution and said, "Well... you calculated the probability as -3.5 ."

For the next ten minutes, he passionately explained that it was only a small calculation mistake. What he had really meant to write was 3.5 .

Apparently, in the world of approving nods and intimidating eyebrows, a probability of 3.5 was perfectly reasonable. The only real problem was that annoying minus sign.



Luck, Numbers, and Live Experiments: Hannah Fry in the 2019 Christmas Lectures

For more than two centuries, the Royal Institution Christmas Lectures have brought science to life through live demonstrations, hands-on experiments, and audience participation. In 2019, British mathematician and science communicator Dr. Hannah Fry presented the three-part series *Secrets and Lies: The Hidden Power of Maths*. Its first episode, *How to Get Lucky?*, explored the role of probability in what we often call luck.

We have included this program for several reasons. First, it is a successful example of science communication that explores probability from different angles, especially through practical applications. Second, it shows the value of bringing together experts from different fields. Contributions from mathematics, psychology, geology, epidemiology, and sport demonstrate how an interdisciplinary approach can make scientific ideas more vivid, understandable, and engaging for students. This is especially important because today's students will help shape tomorrow's world, and connecting science with young people is an essential part of science education. Third, the examples in the program provide a useful starting point for briefly exploring other practical applications of probability in everyday life and across science.

The episode features creative experiments and demonstrations with guests from several disciplines. These include modelling how crowds leave a stadium using colour-coded groups, forecasting volcanic eruptions, simulating the spread of disease, and exploring how random our supposedly random choices really are.

The episode is energetic, full of ideas, and accessible to a wide audience. One of its central messages is that probability is not simply about predicting the future, but about understanding what might happen and why. For anyone who enjoys science through demonstrations, experiments, and real-world examples, Fry's 2019 Christmas Lectures are well worth watching.

At the beginning of the program, as Fry walks toward the lecture theatre, a series of seemingly random events takes place. She passes safely through the path of a swinging pendulum, ice suddenly bursts into flame, and she catches an egg thrown toward her. These moments echo the unpredictable situations we encounter in everyday life, when we may think of ourselves as either lucky or unlucky (See Figure 1.).



Figure 1) The host entering the lecture hall; passing through a swinging pendulum, ice catching fire, catching a thrown egg.

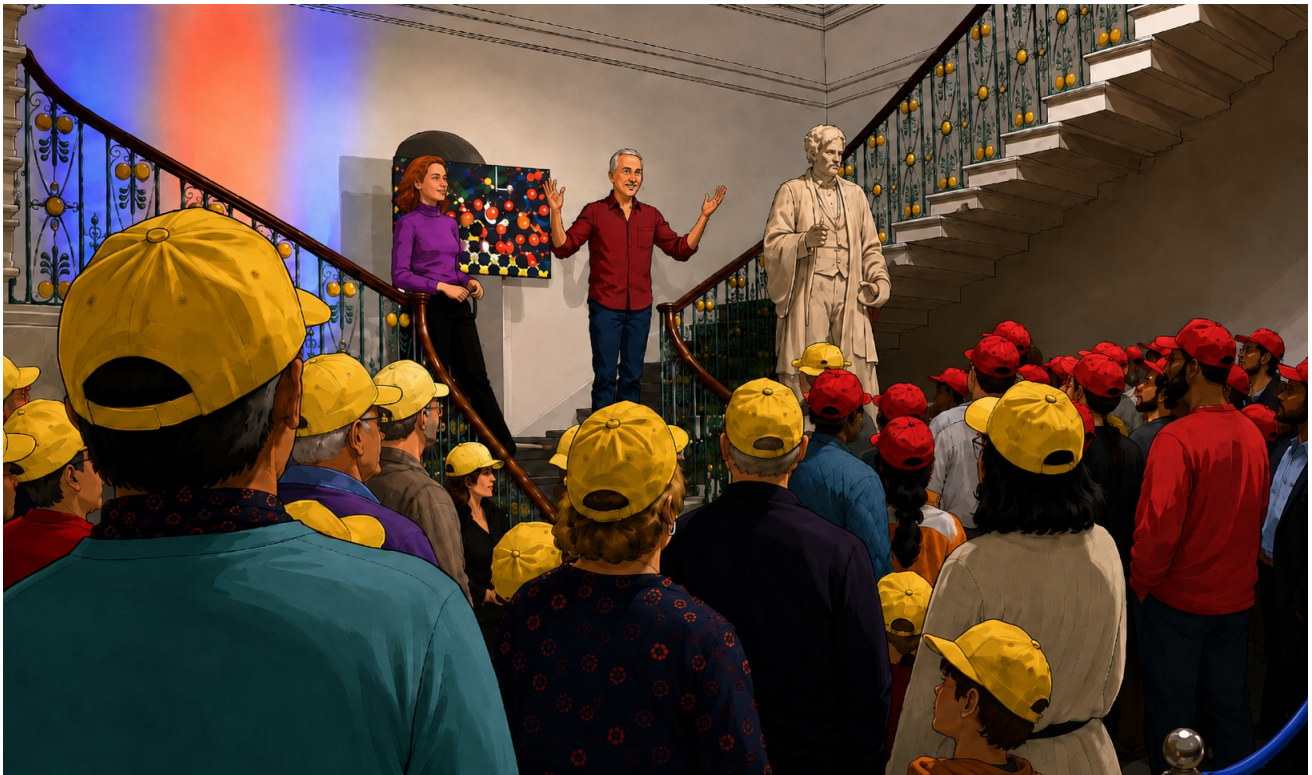


Figure 2) Random motion in a crowded space.

The lecture then moves from these playful examples of chance to situations where probability is used to understand real-world behaviour. Below, we look at one of these examples and a few closely related applications.

Modelling Movement in Crowds: With the help of people wearing red and yellow hats outside the lecture hall, the program demonstrates how people move through crowded spaces. Mathematician and science communicator Matt Parker explains how data from these movements can be used to study crowd behaviour and help design safer exits in places such as stadiums and theatres. Several cameras record the participants while a large screen shows their movements live (See Figure 2.).

Similar probability-based models are used in other crowded environments, including airports, railway stations, concert venues, and large public events. Since the exact path of every person cannot be predicted, planners study likely patterns of movement and congestion to identify where bottlenecks may form and how people are likely to respond to different routes.

The same approach is especially important in emergency evacuation planning. Engineers can simulate scenarios such as blocked exits or unusually dense crowds and estimate how they may affect evacuation times. In this way, probability helps turn uncertain human behaviour into useful information for making public spaces safer.

39 The CCK CLERESTORY



Figure 3) Simulated volcanic eruption.

Natural Disasters and Probability: Geologist Christopher Jackson joins Fry with a model volcano to show how probability helps scientists assess natural events such as volcanic eruptions and rainfall. Rather than predicting exactly when an event will occur, scientists combine observations, historical data, and mathematical models to estimate how likely different outcomes are (See Figure 3.).

The same approach is used for other natural hazards. Meteorologists estimate the probability of heavy rainfall, storms, and flooding, while seismologists use past earthquakes and geological data to assess earthquake risk in different regions. These forecasts cannot remove uncertainty, but they can help communities prepare for events that are difficult to predict with certainty.

Probability is therefore important in disaster planning. By estimating the likelihood and possible consequences of different events, scientists and authorities can make better decisions about warnings, evacuations, infrastructure, and emergency preparedness.



Figure 4) Simulating the spread of infectious diseases.

Disease Spread and Risk Modeling: Dr. Rosalind Eggo explains how probability is used to model the spread of disease. In an interactive game, a symbolic “zombie germ” spreads among masked and unmasked participants, showing in a simple and memorable way how protection can reduce transmission. Rather than tracing every individual case with certainty, probability helps scientists estimate how infections are likely to spread through a group (See Figure 4.).

The same approach is widely used in epidemiology. Researchers study how often people come into contact, how easily an infection passes from one person to another, and how measures such as masks, vaccination, or isolation can change the outcome. These models do not predict exactly who will become ill, but they help show the likely pattern of spread and the level of risk in a population.

Probability is therefore an important tool in public health planning. It helps scientists and health authorities make better decisions about prevention, warnings, hospital preparedness, and disease control. In this way, mathematical modelling turns uncertainty into useful guidance for protecting communities.

41 The CCK CLERESTORY



Figure 5) Analyzing football using statistical models.

Football Data Analytics: A photograph of the Liverpool FC team is shown, and in a playful reveal the audience learns that six people in the image are mathematicians working with the club's data analytics team. One of them, Tim Waskett, explains how mathematical models estimate the likelihood of a scoring opportunity becoming a goal. Two young players then recreate situations on stage while data and heatmaps are shown on screen.

Similar methods are widely used in professional sport. Analysts study passing patterns, player positioning, shot locations, and movement to estimate probabilities and compare tactical choices. These models cannot predict exactly what will happen in a match, but they can reveal patterns that are difficult to see through observation alone. Probability therefore helps teams evaluate performance, assess risk, and make more informed tactical decisions.

The full program can be watched on YouTube here:

https://www.youtube.com/watch?v=_q4DrUHKC0Q



The Airplane Seating Problem



Note: Part C of the *Bus Waiting Time Problem* showed how an intuitive guess may point to the correct answer, which is then established mathematically. We follow the same approach here and in the next example. Curiously, all three problems lead to the same answer: **0.5**.

1) Problem

Consider n passengers boarding a plane in a fixed order, each with an assigned seat. Number the passengers $1, \dots, n$ in boarding order. Passenger 1 has lost their seat number and chooses one of the n seats uniformly at random. Each later passenger takes their assigned seat if it is available; otherwise, they choose uniformly from the remaining empty seats. What is the probability that passenger n , the last to board, sits in their assigned seat?

2) Intuitive solution

Only two seats ultimately determine the outcome: the assigned seats of the first and last passengers. If passenger 1 chooses their own seat, the last passenger succeeds; if passenger 1 chooses the last passenger's seat, the last passenger fails. Suppose instead that passenger 1 chooses the seat assigned to an intermediate passenger k , where $2 \leq k \leq n - 1$. Passengers $2, \dots, k - 1$ then sit in their own seats. When passenger k boards, their seat is occupied. For the remaining process, relabel passenger 1's still-empty assigned seat as passenger k 's effective assigned seat. This is only a relabelling and changes no probabilities: passenger k must now choose uniformly among the remaining empty seats, exactly as passenger 1 did. Thus the problem has reproduced itself with only $n - k + 1$ passengers. Repeating this reduction leaves the first and last passengers' seats as the two symmetric terminating possibilities. Hence the last passenger succeeds with probability $\frac{1}{2}$.

Formal proof by strong induction. Let W_n denote the event that the last passenger sits in their assigned seat, and let $p_n = P(W_n)$.

For the base case, $p_2 = \frac{1}{2}$. Assume that $p_m = \frac{1}{2}$ for every $2 \leq m < n$.

Partition the first passenger's choice into three mutually exclusive cases: E_1 , their own seat; E_2 , the last passenger's seat; and E_3 , any of the other $n - 2$ seats. Therefore,

$$P(E_1) = P(E_2) = \frac{1}{n}, \quad P(E_3) = \frac{n-2}{n}.$$

The corresponding conditional success probabilities are

$$P(W_n | E_1) = 1, \quad P(W_n | E_2) = 0, \quad P(W_n | E_3) = \frac{1}{2}.$$

To justify the last equality, fix k with $2 \leq k \leq n - 1$, and let F_k be the event that passenger 1 chooses the seat assigned to passenger k . Then passengers $2, \dots, k - 1$ sit in their own seats. When passenger k boards, the remaining passengers are $k, \dots, n$ and the empty seats are passenger 1's seat together with seats $k + 1, \dots, n$. In the reduced problem, relabel passenger 1's empty seat as passenger k 's effective assigned seat, while leaving all other remaining seat labels, in particular the last passenger's seat, unchanged. This purely notational relabelling makes passenger k mathematically equivalent to a new first passenger who has lost the number of that effective seat and must choose uniformly among the remaining seats. The residual process is therefore isomorphic to the original problem with $m = n - k + 1 < n$ passengers. By the induction hypothesis, $P(W_n | F_k) = p_{n-k+1} = \frac{1}{2}$. Since this holds for every intermediate k , it follows that $P(W_n | E_3) = \frac{1}{2}$. By the law of total probability,

$$p_n = \frac{1}{n} \cdot 1 + \frac{1}{n} \cdot 0 + \frac{n-2}{n} \cdot \frac{1}{2} = \frac{1}{2}.$$

Thus, for every $n \geq 2$, the probability that the last passenger sits in their assigned seat is $\frac{1}{2}$.

Historical Note: The problem appears to have originated in a Danish version published in 2001 in *FAMØS*, a mathematics magazine produced by students at the University of Copenhagen. In that version, the characters were *nisse*, traditional Danish folklore figures associated with Christmas, choosing beds rather than airline passengers. Solutions and a generalization followed in 2002. The now-familiar airplane version became popular soon afterward as the *Lost Boarding Pass Problem*.



When One More Toss Matters



In many mathematical problems, especially in probability, an intuitive grasp of the solution may seem evident, yet transforming this intuition into a rigorous mathematical argument can be far from straightforward. The following problem from [1] illustrates this point:

Suppose player A tosses $n + 1$ fair coins and player B tosses n fair coins. Let $P(\text{win})$ denote the probability that A obtains more heads than B. Show that $P(\text{win}) = 0.5$.

Hint: Condition on the outcome after both have flipped n coins. (Three scenarios are possible.).

Intuitively, the result seems plausible. Over the first n tosses, the two players are in symmetric positions; the only difference is that A has one additional toss. However, in mathematics, intuition is only the starting point. To formalize this reasoning, set aside one of A's tosses and compare A's remaining tosses with B's tosses. Let the events $\mathcal{R}$, $\mathcal{S}$, and $\mathcal{T}$ represent, respectively, the cases in which:

- A has more heads than B ($\mathcal{R}$),
- B has more heads than A ($\mathcal{S}$),
- both have an equal number of heads ($\mathcal{T}$).

By symmetry, $P(\mathcal{R}) = P(\mathcal{S})$, and since these three events are mutually exclusive and exhaustive, $2P(\mathcal{R}) + P(\mathcal{T}) = 1$.

Let $\mathcal{W}$ denote the event that A's extra toss results in a head. We now condition the probability of A winning on the three cases above.

- If A already leads after n flips, $\mathcal{W}$ is irrelevant. So: $P(\text{win} \mid \mathcal{R}) = 1$
- If B leads, $\mathcal{W}$ cannot reverse the outcome. So: $P(\text{win} \mid \mathcal{S}) = 0$
- If tied, A wins only if $\mathcal{W}$ occurs. Hence: $P(\text{win} \mid \mathcal{T}) = 0.5$

Thus, the total probability is:

$$\begin{aligned} P(\text{win}) &= P(\text{win} \mid \mathcal{R})P(\mathcal{R}) + P(\text{win} \mid \mathcal{S})P(\mathcal{S}) + P(\text{win} \mid \mathcal{T})P(\mathcal{T}) \\ &= 1 \times P(\mathcal{R}) + 0 \times P(\mathcal{S}) + 0.5 \times P(\mathcal{T}) = 0.5(2P(\mathcal{R}) + P(\mathcal{T})) = 0.5 \times 1 = 0.5 \end{aligned}$$

[1] S. M. Ross, "A First Course in Probability", Prentice Hall, 8th ed., 2010.

The CCK CLERESTORY is an interdisciplinary educational and scientific journal dedicated to curriculum-aligned teaching and learning. Each issue explores a central theme through problems and educational content drawn from mathematics, physics, chemistry, history, and other disciplines. Developed with guidance from subject-matter experts, the journal connects school-based knowledge with real-world contexts and encourages meaningful links across disciplines. Its aim is to strengthen students' motivation, deepen conceptual understanding, and foster curiosity and critical thinking.