

Probability of Two People having the Same Birthday in a Room

Nyan Linn Htut

Abstract — The “birthday paradox” states that the probability of two individuals in a room of 23 people sharing the same birthday is slightly greater than 50%. While this phenomenon is well-established and proven, its presentation often includes vague qualifiers such as “around” or “just a little over.” This paper seeks to derive the exact probability, providing a precise and universally accurate value applicable to all cases. Through detailed mathematical analysis, this study aims to offer a definitive answer to this intriguing probabilistic phenomenon.

Nyan Linn Htut

Hinthar Education, Hinthar International School, Yangon, Myanmar, email: quinc@nekonyan.fun

1 Introduction

The birthday paradox is a well-known problem in probability theory that explores the likelihood of shared birthdays within a group.

There are several variations of this paradox, two of which are frequently analyzed:

“What is the probability that at least two people in a group of n people share the same birthday?” and

“If an individual enters a room with n people, what is the probability that at least one person shares their birthday?”

Both scenarios are similar in their underlying methodology and can be addressed using the same approach. For the purposes of this paper, the focus will be on the first definition.

The traditional proof of the birthday paradox is commonly derived using the complement rule.

Rather than directly calculating the probability of at least one shared birthday, the probability of no shared birthdays is determined first. This result is then subtracted from 1 to find the complement. This method is systematic and provides an intuitive framework for analyzing the problem.

1.1 Traditional Proof

The probability that no two individuals in a group of n people share a birthday can be calculated as follows:

The first individual can have any birthday without restriction, resulting in a probability of 1.

The second individual must have a birthday distinct from the first. This probability is $\frac{364}{365}$, as there are 365 possible days in a year, excluding the birthday of the first individual.

Similarly, the third individual must have a birthday distinct from the first two individuals. The probability of this is $\frac{363}{365}$.

This pattern continues, leading to the general expression for the probability that no two individuals share a birthday in a group of n people:

$$P(n) = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \dots \cdot \frac{365 - n + 1}{365} \quad (1)$$

1.1.1 Complement Rule

To calculate the probability of at least one shared birthday, the complement rule is applied:

$$P'(n) = 1 - P(n) \quad (2)$$

Here, $P'(n)$ represents the probability that at least one shared birthday exists within a group of n people.

This approach is valid because the events “no shared birthdays” and “at least one shared birthday” are mutually exclusive and exhaustive. Therefore, their probabilities sum to 1.

Substituting the expression for $P(n)$ into $P'(n)$ yields:

$$P'(n) = 1 - 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \dots \cdot \frac{365 - n + 1}{365} \quad (3)$$

Simplifying the expression further, the probability of at least one shared birthday is expressed as:

$$P'(n) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - n + 1)}{365^n} \quad (4)$$

Or equivalently:

$$P'(n) = 1 - \frac{365!}{365^n \cdot (365 - n)!} \quad (5)$$

This formula provides the probability of at least one shared birthday in a group of n people.

1.1.2 $n = 23$ Case

When $n = 23$, the calculated probability is approximately 0.5073, which is marginally greater than 50%.

This result is frequently rounded to 50% for simplicity, giving rise to the common assertion that in a group of 23 individuals, the probability of at least two individuals sharing a birthday is “around 50%.”

However, the more precise value is 0.5073, which reflects a slightly higher probability.

1.1.3 Alternative Approaches

Although the complement rule is the most commonly employed method for proving the birthday paradox, alternative approaches are also available.

One such method involves directly calculating the probability of two individuals sharing a birthday. While this approach is more complex, it remains equally valid.

Another alternative uses the pigeonhole principle to provide an intuitive explanation of the paradox. This method conceptualizes “pigeons” as individuals with birthdays and “pigeonholes” as the days of the year, offering a straightforward way to understand the phenomenon.

An alternative method can also employ combinatorics to derive the same result.

The total number of ways to assign birthdays to n people is 365^n .

The number of ways to assign birthdays such that no two individuals share a birthday is $365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - n + 1)$.

Thus, the probability of at least one shared birthday is expressed as follows:

$$P(n) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - n + 1)}{365^n} \quad (6)$$

2 Exact Probability

While the traditional proof is widely utilized, it has several inherent limitations.

The traditional approach assumes a year consists of exactly 365 days, thereby excluding individuals born on February 29 during leap years. While this assumption simplifies the calculations, it does not accurately reflect the real-world distribution of birthdays.

It also assumes that all birthdays are equally likely, which does not align with actual patterns. For instance, certain months are associated with higher birth rates.

Furthermore, the assumption that birthdays are independent events does not always hold in practice. Siblings, for example, may exhibit correlated birthdays due to shared cultural or familial factors. Similarly, specific environments, such as classrooms, may have unique patterns of shared birthdays influenced by enrollment timing.

Finally, the approximation of 0.5073 for $n = 23$ is frequently rounded to 50%, leading to less precise interpretations.

To address these limitations, an exact probability of two individuals sharing a birthday in a group of n people can be derived.

2.1 Accounting for Leap Years

To improve accuracy, modifications can be made to account for leap years and individuals born on February 29.

A leap year includes 366 days, with February 29 as an additional day. To incorporate this, the number of days across four consecutive years can be considered: $365 + 365 + 365 + 366 = 1461$ days, accounting for three non-leap years and one leap year.

In this model, the probabilities are no longer uniformly distributed across the days of the year, as February 29 has a lower likelihood compared to other days.

Under this model, the probability of an individual being born on February 29 is $\frac{1}{1461}$, whereas the probability of being born on any other day (e.g., November 21) is $\frac{\frac{4}{1461}}{\frac{4}{1461}}$.

2.1.1 Case 1 (First person chosen is born on February 29)

The probability that the first individual is born on February 29 is $\frac{1}{1461}$.

The probability that the second individual is born on a day other than February 29 is $\frac{1460}{1461}$.

The third individual must also avoid February 29 and the four days already selected by the second individual. The probability for this is $\frac{1456}{1461}$.

This pattern continues, resulting in the following expression for the probability that no two individuals in a group of n people share a birthday:

$$P_{\text{leap}_1}(n) = \frac{1}{1461} \cdot \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1468 - 4(n-1)}{1461} \cdot \frac{1468 - 4n}{1461} \quad (7)$$

Or equivalently:

$$P_{\text{leap}_1}(n) = \frac{1 \cdot 1460 \cdot 1456 \cdot 1452 \cdot \dots \cdot (1468 - 4(n-1)) \cdot (1468 - 4n)}{1461^n} \quad (8)$$

Factoring out the 4, the expression becomes:

$$P_{\text{leap}_1}(n) = \frac{1 \cdot (4^{n-1})[365 \cdot 364 \cdot 363 \cdot \dots \cdot (367 - (n-1)) \cdot (367 - n)]}{1461^n} \quad (9)$$

Rewriting using factorial notation yields:

$$P_{\text{leap}_1}(n) = \frac{4^{n-1} \cdot 365!}{1461^n \cdot (366 - n)!} \quad (10)$$

2.1.2 Case 2 (Second person chosen is born on February 29)

In this scenario, the first individual can be born on any day except February 29, with a probability of $\frac{1460}{1461}$.

The probability of the second individual being born on February 29 is $\frac{1}{1461}$.

The third individual must be born on any day other than February 29 and also not on the day selected by the first individual. This probability is $\frac{1456}{1461}$.

This pattern mirrors the calculation in Case 1, with the first two terms swapped. The remainder of the computation follows the same structure.

Thus, the probability is expressed as:

$$P_{\text{leap}_2}(n) = \frac{1460}{1461} \cdot \frac{1}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1468 - 4(n-1)}{1461} \cdot \frac{1468 - 4n}{1461} \quad (11)$$

This leads to the same result:

$$P_{\text{leap}_2}(n) = \frac{4^{n-1} \cdot 365!}{1461^n \cdot (366 - n)!} \quad (12)$$

2.1.3 Case 3 (Third person chosen is born on February 29)

This case follows the same general pattern observed in the earlier cases. The probabilities are adjusted based on the position of the individual born on February 29.

The probability that the first person is born on any day except February 29 is $\frac{1460}{1461}$.

The second person must be born on any day other than February 29 and cannot share the same birthday as the first person. This probability is $\frac{1456}{1461}$.

The third person is born on February 29, which has a probability of $\frac{1}{1461}$.

Continuing this pattern, the probability that no two individuals in a group of n people share a birthday is expressed as:

$$P_{\text{leap}_3}(n) = \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1468 - 4(n-1)}{1461} \cdot \frac{1468 - 4n}{1461} \quad (13)$$

In this scenario, the term $\frac{1}{1461}$ appears in the third position, while all other terms remain consistent with the overall pattern.

As the number of individuals in the group increases, the $\frac{1}{1461}$ term continues to shift downward in the sequence. For instance, in case k , the $\frac{1}{1461}$ term appears in the k^{th} position, while all other terms follow the same pattern established in earlier cases.

2.1.4 Case n (Last person chosen is born on February 29)

For the final individual in the group, the $\frac{1}{1461}$ term is positioned last in the sequence. All other terms remain unchanged.

This scenario is represented as:

$$P_{\text{leap}_n}(n) = \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1468 - 4n}{1461} \cdot \frac{1}{1461} \quad (14)$$

The pattern observed in earlier cases persists, with the $\frac{1}{1461}$ term progressively moving to the n^{th} position as the group size increases.

2.1.5 Case 0 (No one is born on February 29)

This case examines the situation where no individual in the group is born on February 29.

The probability that the first individual is not born on February 29 is $\frac{1460}{1461}$.

The second individual must also avoid February 29 and not share the same birthday as the first individual. The probability for this is $\frac{1456}{1461}$.

Since no one is born on February 29, the $\frac{1}{1461}$ term does not appear in this calculation.

The probability of no two individuals in a group of n people sharing a birthday under these conditions is given by:

$$P_{\text{leap}_0}(n) = \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1464 - 4(n-1)}{1461} \cdot \frac{1464 - 4n}{1461} \quad (15)$$

To align with the format used in earlier cases, the final terms can be adjusted:

$$P_{\text{leap}_0}(n) = \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1468 - 4n}{1461} \cdot \frac{1468 - 4(n+1)}{1461} \quad (16)$$

Factoring out the 4, the expression becomes:

$$P_{\text{leap}_0}(n) = \frac{(4^n)[365 \cdot 364 \cdot 362 \cdot \dots \cdot (367 - n) \cdot (367 - (n+1))]}{1461^n} \quad (17)$$

Rewriting using factorial notation yields:

$$\begin{aligned}
 P_{\text{leap}_0}(n) &= \frac{4^n \cdot 365!}{1461^n \cdot (366 - (n + 1))!} \\
 P_{\text{leap}_0}(n) &= \frac{4^n \cdot 365!}{1461^n \cdot (365 - n)!} \\
 P_{\text{leap}_0}(n) &= \frac{4^n \cdot 365!}{1461^n \cdot (366 - n)!} \cdot (366 - n)
 \end{aligned} \tag{18}$$

2.1.6 Combining all Cases

The total probability of no two individuals sharing a birthday in a group of n people, accounting for leap years, is derived by summing the probabilities of all cases.

For cases 1 through n :

$$\begin{aligned}
 P_{\text{leap}}(n) &= n \cdot P_{\text{leap}_1}(n) \\
 P_{\text{leap}}(n) &= n \cdot \frac{4^{n-1} \cdot 365!}{1461^n \cdot (366 - n)!} \\
 P_{\text{leap}}(n) &= \frac{n(4^{n-1}) \cdot 365!}{1461^n \cdot (366 - n)!}
 \end{aligned} \tag{19}$$

For Case 0, where no one is born on February 29:

$$\begin{aligned}
 P(n) &= P_{\text{leap}}(n) + P_{\text{leap}_0}(n) \\
 P(n) &= \frac{n(4^{n-1}) \cdot 365!}{1461^n \cdot (366 - n)!} + \frac{4^n \cdot 365!}{1461^n \cdot (366 - n)!} \cdot (366 - n)
 \end{aligned} \tag{20}$$

Factoring out the common term results in:

$$\begin{aligned}
 P(n) &= \frac{4^n \cdot 365!}{1461^n \cdot (366 - n)!} \cdot (4^{-1}n + 366 - n) \\
 P(n) &= \frac{4^n \cdot 365!}{1461^n \cdot (366 - n)!} \cdot \left(366 - \frac{3}{4}n\right)
 \end{aligned} \tag{21}$$

2.2 Conclusion

The formula derived provides the probability of no two individuals sharing a birthday in a group of n people, accounting for leap years. The complement rule is then applied to calculate the probability of at least one shared birthday:

$$P'(n) = 1 - P(n) \quad (22)$$

Thus, the exact probability of at least two individuals sharing a birthday in a group of n people is given by:

$$P(n) = 1 - \left[\frac{4^n \cdot 365!}{1461^{23} \cdot (366 - n)!} \cdot \left(366 - \frac{3}{4}n \right) \right] \quad (23)$$

This formula incorporates leap year considerations and provides an accurate representation of the likelihood of shared birthdays.

2.2.1 $n = 23$ Case

By substituting $n = 23$ into the formula, the resulting probability is:

$$\begin{aligned} P(23) &= 1 - \left[\frac{4^{23} \cdot 365!}{1461^{23} \cdot (366 - 23)!} \cdot \left(366 - \frac{3}{4} \cdot 23 \right) \right] \\ &\approx 0.50686501 \text{ (8 significant figures)} \end{aligned} \quad (24)$$

This value is marginally lower than the traditionally accepted approximation of 0.5073. The discrepancy arises from the inclusion of February 29 in the calculations, which slightly reduces the probability of shared birthdays due to the uneven distribution of birthdays across the calendar.

2.2.2 Graphing

To illustrate the probability of two individuals sharing a birthday within a group of n people, the function $P(n)$ is graphically represented.

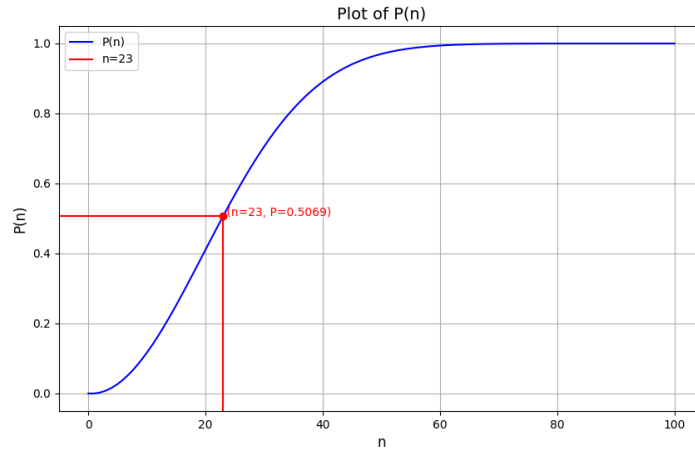


Fig. 1: Probability of Two Individuals Sharing a Birthday in a Group of n People

A comparison with the traditional approximation reveals a minor divergence in probabilities for different values of n .

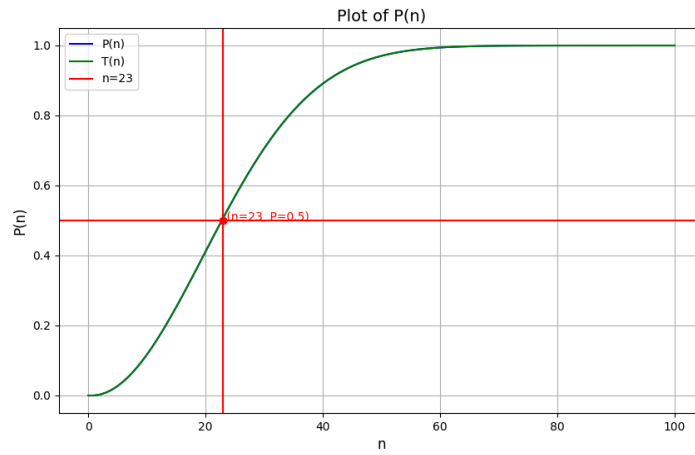


Fig. 2: Comparison of Exact and Approximate Probabilities

Although the difference is not immediately apparent in the graph, it holds significance in precise probabilistic calculations. A zoomed-in view makes the disparity more discernible.

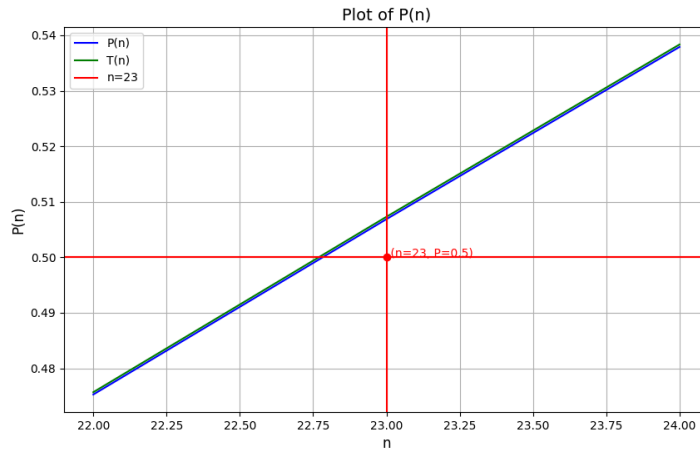


Fig. 3: Zoomed-in Comparison of Exact and Approximate Probabilities

3 Further Analysis

While the exact probability formula provides a more accurate calculation compared to the traditional approximation, there remain areas for further exploration and refinement.

3.1 Birthdate Distribution

The assumption of an even distribution of birthdays simplifies the model but may not accurately reflect real-world patterns.

Empirical data indicates that birth rates vary across months, with September recording the highest birth rates in the United States and February the lowest. [1] These variations are potentially influenced by seasonal factors such as holidays, with February being associated with Valentine's Day and November with Thanksgiving.

Incorporating real-world birthdate distribution data into the model would enhance its accuracy. Such adjustments would involve using demographic or birth registry data to assign probabilities to specific dates.

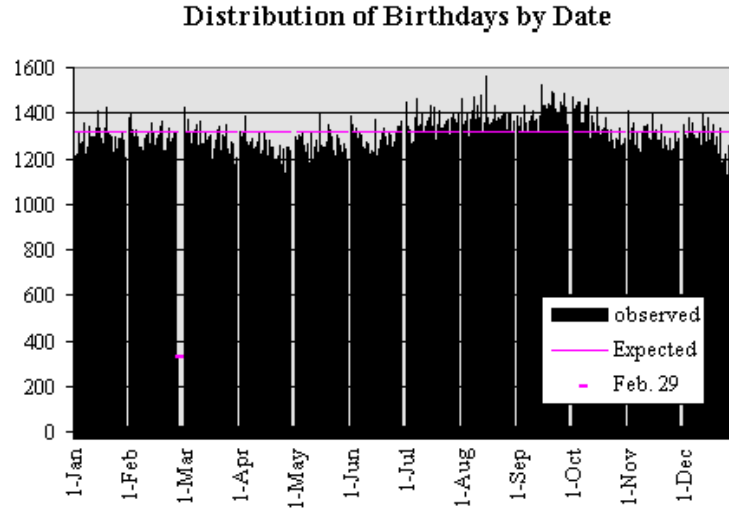


Fig. 4: Birthdate Distribution [2]

However, incorporating weighted distributions significantly complicates the calculations, as the probabilities would no longer remain uniform.

The current calculations are limited to two cases: whether or not a person is born on February 29, with probabilities of $\frac{1}{1461}$ and $\frac{4}{1461}$, respectively. Introducing variable probabilities for each calendar day would necessitate a more complex framework, and this problem is left as an exercise for readers.

3.2 Correlated Birthdays

The assumption of independence in birthdates may not hold in certain contexts. Factors such as cultural practices, family planning, or institutional patterns can introduce correlations.

For instance, siblings are more likely to have correlated birthdays, and student populations may exhibit patterns influenced by academic enrollment cycles.

A potential refinement involves modeling birthdate distributions with a normal distribution, adjusting the mean and standard deviation to reflect correlations. For example, in a classroom setting, a distribution could be represented as:

$$S \sim N(324, 30) \quad (25)$$

Here, S denotes the birthdates of students, with a mean of 324 (corresponding to November 20) and a standard deviation of 30 days. Despite these refinements, challenges such as the February 29 issue remain unresolved.

3.3 Leap Year Adjustments

The model presented is robust and remains applicable regardless of the inclusion of leap years in the calculations. The formula inherently accounts for the number of days in a cycle of four consecutive years. For example, in a sample space from 1975 to 2000, the formula remains valid even though 1975 is not a leap year, while 2000 is.

The sequence 365, 365, 365, 366 is equivalent to 365, 365, 366, 365, 365, 366, 365, 365, and 366, 365, 365, 365.

As such, the choice of the starting year in the sample space does not affect the formula's validity. The model ensures that the cycle includes three non-leap years and one leap year, which holds true for any four consecutive years.

For computations, values such as $\frac{1}{365.25}$ and $\frac{25}{365.25}$ can be used. The former represents all days except February 29, while the latter corresponds to February 29. However, decimals are avoided in the derivation to maintain consistency. This approach assumes an average year length of 365.25 days, where February 29 accounts for 0.25 of a day annually.

3.3.1 Reality isn't always this simple

While it is commonly stated that every four years include a leap year, this is not entirely accurate due to exceptions in the calendar system. Although 2000 was a leap year, it represents an anomaly. The true length of a year is 365.2425 days, necessitating adjustments to account for the discrepancy.

To reconcile this, one day is omitted every 100 years (equivalent to removing 0.01 days annually). As a result, years divisible by 100 are not leap years despite being divisible by 4. For instance, 1700, 1800, and 1900 were divisible by 4 but did not include February 29.

Conversely, an additional day is added every 400 years to account for 0.0025 days annually. Consequently, years divisible by 400 are leap years, even if they are divisible by 100. Examples include 1200, 1600, 2000, and 2400.

To simplify the analysis, the sample space can be limited to individuals born between 1950 and 2025, a range encompassing the typical human lifespan. This restriction ensures that the majority of the sample will fall within this timeframe, minimizing the impact of edge cases such as years like 1900.

3.4 Leap Seconds

The Earth's rotation is not constant; it varies due to factors such as the movement of tectonic plates and shifts in magnetic fields. To address these variations, the International Astronomical Union occasionally decides whether to add a leap second.

When a leap second is added, the clocks progress as follows:

$$\begin{array}{l} 23 : 59 : 58 \\ 23 : 59 : 59 \end{array} \quad (26)$$

Instead of advancing directly to:

$$00 : 00 : 00 \quad (27)$$

the time transitions to:

$$23 : 59 : 60 \quad (28)$$

This adjustment introduces an extra second into the minute, creating 61 seconds instead of the standard 60. Such an irregularity can lead to complications, as systems may either become out of sync or repeat a second.

Incorporating leap seconds into the probability model involves considering 86,401 seconds on December 31, as opposed to the usual 86,400 seconds for other days. This modification slightly increases the probability of being born on December 31. However, the impact is minimal.

Under this adjusted model, the total number of seconds in four consecutive years becomes:

$$(24 \cdot 60 \cdot 60) \cdot 1461 + 4 = 126230404 \quad (29)$$

This adjustment introduces three distinct probabilities for being born on December 31, February 29, and other days. The calculations are further complicated by the large denominator of 126230404.

For extended considerations, such as super leap years, the model must account for 400 consecutive years to include century leap years. The total number of seconds over this span becomes:

$$\begin{aligned} 365 \cdot 400 + 100 - 1 &= 145901 \text{ days} \\ 145901 \cdot 24 \cdot 60 \cdot 60 + 400 &= 12605846800 \text{ seconds} \end{aligned} \quad (30)$$

These calculations present a challenging yet insightful exercise for detailed modeling.

3.5 Approximations

Various simpler approximations are used in analyzing the birthday paradox.

3.5.1 Counting the Number of Pairs

One approximation involves counting the number of pairs of individuals in a group. In a room with n individuals, there are $\frac{n(n-1)}{2}$ unique pairs.

The probability that two people do not share a birthday is $\frac{364}{365}$, so the probability that no pair shares a birthday is expressed as $\left(\frac{364}{365}\right)^{\frac{n(n-1)}{2}}$. This leads to the probability of at least one shared birthday:

$$1 - \left(\frac{364}{365}\right)^{\frac{n(n-1)}{2}} \quad (31)$$

For $n = 23$, this yields a probability close to 0.5005, which aligns with traditional approximations. However, this method assumes independence between pairs, which is not accurate. For example, if two people share a birthday and a third person shares a birthday with one of them, the likelihood of the third person sharing a birthday with the first is no longer $\frac{364}{365}$.

3.5.2 Using the Exponential Function

Another common approximation employs the exponential function. Using a Taylor expansion:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots \quad (32)$$

When x is close to zero, the function simplifies to:

$$e^x \approx 1 + x \quad (33)$$

This method can be applied to the probability functions. For $P_{\text{leap}_k}(n)$:

$$P_{\text{leap}_k}(n) = \frac{1}{1461} \cdot \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1468 - 4(n-1)}{1461} \cdot \frac{1468 - 4n}{1461} \quad (34)$$

Rewriting the terms in the form $1 - x$, the probability becomes:

$$P_{\text{leap}_k}(n) = \frac{1}{1461} \cdot \left(1 - \frac{1}{1461}\right) \cdot \left(1 - \frac{5}{1461}\right) \cdot \left(1 - \frac{9}{1461}\right) \cdot \dots \cdot \left(1 - \frac{4(n-1)-7}{1461}\right) \cdot \left(1 - \frac{4n-7}{1461}\right) \quad (35)$$

Let S represent the sum of terms within the brackets. Similarly, $P_{\text{leap}}(n)$ can be expressed as:

$$P_{\text{leap}}(n) = \frac{n}{1461} \cdot S \quad (36)$$

For $P_{\text{leap}_0}(n)$:

$$P_{\text{leap}_0}(n) = \frac{1460}{1461} \cdot \frac{1456}{1461} \cdot \frac{1452}{1461} \cdot \dots \cdot \frac{1464 - 4(n-1)}{1461} \cdot \frac{1464 - 4n}{1461} \quad (37)$$

Rewriting:

$$P_{\text{leap}_0}(n) = S \cdot \left(1 - \frac{4(n+1)-7}{1461}\right) \quad (38)$$

Combining cases provides the total probability of no shared birthdays in a group of n individuals, accounting for leap years:

$$\begin{aligned} P(n) &= \frac{n}{1461} \cdot S + S \cdot \left(1 - \frac{4(n+1)-7}{1461}\right) \\ P(n) &= S \cdot \left(\frac{n}{1461} + 1 - \frac{4(n+1)-7}{1461}\right) \\ P(n) &= S \cdot \left(\frac{1464 - 3n}{1461}\right) \\ P(n) &= S \cdot \left(1 - \frac{3n-3}{1461}\right) \end{aligned} \quad (39)$$

Approximating S :

$$\begin{aligned}
 S &= \left(1 - \frac{1}{1461}\right) \cdot \left(1 - \frac{5}{1461}\right) \cdot \left(1 - \frac{9}{1461}\right) \cdot \left(1 - \frac{13}{1461}\right) \cdot \dots \\
 &\quad \dots \cdot \left(1 - \frac{4(n-1)-7}{1461}\right) \cdot \left(1 - \frac{4n-7}{1461}\right) \\
 &\approx \left(e^{-\frac{1}{1461}}\right) \cdot \left(e^{-\frac{5}{1461}}\right) \cdot \left(e^{-\frac{9}{1461}}\right) \cdot \left(e^{-\frac{13}{1461}}\right) \cdot \dots \\
 &\quad \dots \cdot \left(e^{-\frac{4(n-1)-7}{1461}}\right) \cdot \left(e^{-\frac{4n-7}{1461}}\right) \\
 &\approx e^{-\frac{1+5+9+13+\dots+(4(n-1)-7)+(4n-7)}{1461}} \\
 &\approx e^{-\frac{2n^2-5n+3}{1461}}
 \end{aligned} \tag{40}$$

Substituting this back into the equation, we get:

$$\begin{aligned}
 P(n) &\approx e^{-\frac{2n^2-5n+3}{1461}} \cdot \left(1 - \frac{3n-3}{1461}\right) \\
 &\approx e^{-\frac{2n^2-5n+3}{1461}} \cdot e^{-\frac{3n-3}{1461}}
 \end{aligned} \tag{41}$$

$$P(n) \approx e^{-\frac{2n^2-8n+6}{1461}} \tag{42}$$

This approximation is notably simpler than the exact formula and provides a reasonable estimate for the probability of no two individuals sharing a birthday in a group of n people.

It is worth noting, however, that the approximation becomes less accurate for larger values of n .

For $n = 23$, substituting into the formula yields:

$$\begin{aligned}
 P(23) &\approx e^{-\frac{2(23)^2-8(23)+6}{1461}} \\
 &\approx e^{-0.6023} \\
 &\approx \mathbf{0.5475} \text{ (4 significant figures)}
 \end{aligned} \tag{43}$$

The calculated value is marginally higher than the traditionally accepted approximation of 0.5073, though it remains a practical estimate.

3.6 Variation of the problem

A related question to the birthday paradox examines:

“What is the probability that at least one person shares your birthday in a room of n people?”

This problem differs from the classic paradox by focusing on the likelihood of finding at least one person who shares a specific individual’s birthday, rather than determining whether any two people share a birthday.

Unlike the general scenario, this variation fixes one birthday (e.g., the reader’s) and evaluates the probabilities for other individuals in the room relative to that date. This fixed reference point changes the nature of the calculations.

3.6.1 Case 1: You are born on February 29

If the individual’s birthday is February 29, the probability that any given person does not share this birthday is $\frac{1460}{1461}$.

For n individuals in the room, the probability that no one shares this birthday is $\left(\frac{1460}{1461}\right)^n$.

The complementary probability, representing at least one individual sharing the birthday, is $1 - \left(\frac{1460}{1461}\right)^n$.

3.6.2 Case 2: You are not born on February 29

For individuals not born on February 29, the probability that any other person does not share the specific birthday is $\frac{1457}{1461}$.

This includes the possibility of being born on February 29.

For n individuals in the room, the probability that no one shares the specified birthday is $\left(\frac{1457}{1461}\right)^n$. The complementary probability is $1 - \left(\frac{1457}{1461}\right)^n$.

3.6.3 Combining the cases

To account for both cases, the total probability of at least one person sharing a specific birthday in a group of n individuals can be expressed as:

$$P(n) = \frac{1}{1461} \left(1 - \left(\frac{1460}{1461} \right)^n \right) + \frac{1460}{1461} \left(1 - \left(\frac{1457}{1461} \right)^n \right) \quad (44)$$

The terms are weighted by the probabilities of the individual being born on February 29 or any other date.

For $n = 23$, substituting into the formula gives $P(23) \approx 0.061078$, or approximately 6%.

The probability of finding at least one person who shares your birthday is significantly lower than the probability of two individuals in a group sharing the same birthday.

This difference arises because the probability of two people sharing a birthday considers all possible pairs within the group, which may not include you specifically. For instance, the shared birthday could occur between any two individuals among the n others in the room.

In contrast, the probability of someone sharing your birthday focuses exclusively on one specific event—your birthday being matched. This distinction explains why the likelihood of you sharing a birthday with someone is notably lower than the broader scenario of any two people sharing a birthday.

3.6.4 Follow-up Question

A follow-up to this variation asks how many individuals are required in a room to achieve a 50% chance of finding at least one person who shares a specific birthday.

Setting $P(n) = 0.5$ and solving for n , we find:

$$n = 252.996 \quad (45)$$

Rounding to the nearest whole number, approximately 253 individuals are needed.

This contrasts sharply with the 23 individuals required in the classic problem for a 50% chance of any two individuals sharing a birthday.

3.7 Conclusion

The birthday paradox serves as a compelling illustration of probability theory's often counterintuitive outcomes, especially in scenarios involving large sample spaces. This research has delved into both the traditional and extended interpretations of the paradox, highlighting its mathematical underpinnings and practical implications. By integrating refinements such as leap year adjustments, non-uniform birthdate distributions, and cultural variations, this study bridges the gap between theoretical models and real-world complexities.

For $n = 23$, the calculated probability of approximately 0.5073 illustrates the rapid increase in likelihood of shared birthdays with group size, while the adjusted probability $P(23) \approx 0.50686501$ demonstrates the impact of accounting for leap years. Furthermore, the analysis of variations, such as the probability of at least one individual sharing a specific birthday, underscores the importance of contextual framing in probability assessments.

Through rigorous mathematical modeling, graphical comparisons, and nuanced assumptions, this research advances our understanding of the birthday paradox. The findings not only enhance the accuracy of probability estimations but also highlight broader applications in fields where probabilistic reasoning is essential.

Ultimately, the study reaffirms the paradox's relevance as a foundational concept in probability theory while offering practical insights for its application in diverse domains. Future research could further refine these models by exploring additional real-world variables, thereby expanding the scope of this classic probabilistic phenomenon.

Bibliography

1. Randall, W.: A statistical analysis of the annual pattern in births in the USA, 1967 to 1976. *Journal of Interdisciplinary Cycle Research*. 18, 179–191 (1987). <https://doi.org/10.1080/09291018709359944>
2. Murphy, R.: The Birthday Problem, <https://www.panix.com/~murphy/bday.html>